IJARIIT INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume3, Issue2)
Available online at www.ijariit.com

Optimize Workflow Scheduling Using Hybrid Ant Colony Optimization (ACO) & Particle Swarm Optimization (PSO) Algorithm in Cloud Environment

Misha Goyal
Lala Lajpat Rai Institute of Engineering & Technology, Moga
mishagoyal29@gmail.com

Mehak Aggarwal
Lala Lajpat Rai Institute of Engineering & Technology, Moga
mehakaggarwal@llriet.ac.in

***Abstract:** Those days are gone when storing and accessing of data were done on computer's hard drive. Now with innovation in technology and with the great success of Internet computing resources have become more economical, more powerful and more ubiquitously available than ever before. This technological trend of the 21st century has given birth to the realization of a new computing model called Cloud Computing. This Computing is not only about the hard drives were storing and accessing can be done but it is latest computing paradigm and it offers tremendous opportunities to solve the large-scale scientific problem. To fully exploit the applications of cloud, various challenges need to be addressed where scheduling is one among them. Although catholic research has been done on Workflow Scheduling, there are very few edges tailored for Cloud environments. For some basic principles of Cloud such as elasticity and heterogeneity existing work fails to meet optimal solution. Therefore our work focuses on the scheduling strategies for scientific workflow on IaaS cloud. We present an algorithm based on the meta- heuristic optimization technique where the best of two algorithms Ant colony Optimization (ACO) and Particle Swarm Optimization (PSO) are merged to optimize locally and globally which minimizes the overall workflow time (makespan) and reduces the cost. Our heuristic is evaluated using CloudSim and several well-known scientific workflows of different sizes. The results show that our approach performs better when compared to PSO algorithm.*

***Keywords:** PSO, ACO, Cloud, Workflow.*

I. INTRODUCTION

With progression in innovation, preparing and capacity furthermore with the accomplishment of the Internet, registering assets have ended up less expensive, more capable and more universally accessible than any time in recent memory. This mechanical pattern has brought forth the acknowledgment of another registering model called cloud computing, in which assets (e.g., CPU and capacity) are given as general utilities that can be rented and discharged by clients through the Internet in an on-demand. In a cloud computing environment, the customary part of the service provider has isolated into two: the infrastructure suppliers who oversee cloud stages and rent assets as per a user based estimating the model, and administration suppliers, who rent assets from one or numerous framework suppliers to serve the end clients. Substantial organizations, for example, Google, Amazon, and Microsoft endeavour to give all the more effective, dependable and cost-proficient cloud stages, and business undertakings try to reshape their plans of action to pick up advantage from this new worldview. [28]Cloud computing can be distributed into three service models: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).

Need of Scheduling in clouds

Unlike Grids, Scalability, flexibility reliability of Cloud resources allows real- time processing of resources to meet application requirement. At lower cost services of cloud such as compute, storage, and bandwidth are available. Normally undertakings are scheduled by client prerequisites. New planning methodologies should be proposed to defeat the issues postured by system properties in the middle of client and assets. New booking methodologies might utilize a percentage of the customary planning ideas to consolidation them together with some system mindful techniques to give answers for better and more effective employment booking. The customary path for booking in distributed computing was to utilize the immediate assignments of clients as the overhead application base. The problem in that scheduling was there is no association between the overhead application base and the way that different tasks cause overhead costs of resources in Cloud systems which may incur the cost of Cloud. That is why there is need of scheduling in Cloud Environment so that parallel processing of the complex application can be done efficiently [22].

Scheduling in Cloud Computing

In the cloud, computing scheduling is the process of plotting tasks onto resources and the systems (e.g. CPU time, bandwidth and memory) efficiently. In cloud computing, many complex applications require parallel processing to execute the jobs effectively. Due to the communication and synchronization between corresponding processes, there is a decrease in utilization of CPU resources. Therefore it is necessary for a data center to achieve the utilization of nodes while maintaining the level of responsiveness of parallel jobs [23]. Due to the availability of vast data on the internet and growing number of user's day to day, it almost impossible to assign the various tasks manually to the virtual machines[20]. Hence, to allocate the resources to each job effectively, scheduling plays an important role in cloud computing. Thus various scheduling algorithms are proposed so that they can help in achieving the order of jobs in such a way that balance between improving the performance, cost, makespan, load balancing and more over the quality of service can be improved. For proper scheduling, many task parameters need to be considered which is an essential aspect in successful working of the cloud. The accessible resources should be utilized efficiently without affecting the service limits of the cloud.

1. Resource discovering and filtering Data-canter Broker locates the resources present in the web system and collects status information related to them.

2. Resource selection Based on the specific parameter of task and resources target resource is selected.

3. Task submissions To the selected resources task are submitted.

II. LITERATURE REVIEW

A deep study is performed for workflow scheduling and resource provisioning in IaaS Cloud environment and there are a number of strategies proposed wherein scheduling of workflows which might be started out at any time and the QoS necessities are taken into account and able to boom the scheduling fulfillment rate extensively.

Ahmad et al. (2016) presented new hybrid genetic algorithm (HGA) for workflow scheduling. The proposed algorithm seeds the Heterogeneous Earliest Finish Time (HEFT) based schedule in the initial population that guides the algorithm to reach an optimal (makespan) schedule in fewer generations. Rigorous search with two fold crossover and mutation operators cover the large problem space and enhances the HGA performance. The proposed algorithm optimizes workflow schedule length with an additional feature of load balancing that ensures the optimized resource utilization.

Verma et al. (2015) presented a Budget and Deadline Constrained heuristic based upon Heterogeneous Earliest Finish Time (HEFT) to schedule workflow tasks over the available cloud resources. In the first step, the task assigned the highest priority which is calculated by the upward rank is selected, and various cost and time affecting parameter are calculated for the selected task. The second step assigns the task to the best resource which will minimize the EFT and execution cost.

Kalra et al. (2015) presented a broad review and comparative analysis of the three meta heuristic techniques(GA, PSO and ACO) already written previously along with two novel methods (BAT and League championship algorithm)for workflow scheduling in distributed computing (grid and cloud). A change in transitional operators, a high-quality seed for initial population and a hybrid algorithm, enhance the quality and the convergence speed of the optimal solution.

Alkhanak et al. (2015) investigated and analyzed various cost aware challenges of WFS in cloud computing such as Quality of service, performance, system functionality and system architecture. They also discussed various WFS cost aware approaches from the available pools of alternatives.

Rodriguez et al. (2014) have used PSO Meta heuristic technique to optimize the scientific workflow on IaaS cloud. The authors have incorporated the basic features of IaaS cloud such as unlimited heterogeneous resources, flexibility, dynamic provisioning of resources, VMs boot time as well as leased time and performance variation of VMs. The idea was to combine the resource provisioning along with the scheduling strategy with an objective to minimize the execution cost while staying in the defined deadline.

Netjinda et al. (2014) focused on optimizing the value of buying infrastructure-as-a-service cloud competencies to attain clinical work goes with the flow execution in the unique closing dates. Authors considered the quantity of purchased times, example types, buying options, and venture scheduling as constraints in an optimization technique. PSO augmented with a variable community seeks approach turned into used to discover the superior solution. Results display promising performance from the views of the total fee and fitness convergence when in comparison with other trendy algorithms.

Abrishami et al. (2013) have proposed two workflow scheduling algorithms: a one-phase algorithm which is called IaaS Cloud Partial Critical Paths (IC-PCP), and a two-phase algorithm which is called IaaS Cloud Partial Critical Paths with Deadline Distribution (IC-PCPD2). IC-PCP schedules the workflow in one phase by scheduling each partial critical path on a single instance of a computation service, while IC-PCPD2 is a two-phase algorithm which first distributes the overall deadline on the workflow tasks, and then schedules each task based on its sub deadline. Both algorithms have a polynomial time complexity which makes them suitable options for scheduling large workflows. Both algorithms have a promising performance but IC-PCP performing better than IC-PCPD2.

Malawski et al. (2012) mentioned, developed, and assessed algorithms primarily based on static and dynamic techniques for each project scheduling and aid provisioning and evaluated thru simulation using a set of scientific workflow ensembles with a wide variety of budget and closing date parameters, considering uncertainties in assignment runtime estimations, provisioning delays, and failures.

Xue et al. (2012) proposed a QoS-based totally hybrid particle swarm optimization (GHPSO) to schedule packages to cloud sources. In GHPSO, crossover and mutation of a genetic set of rules are embedded into the particle swarm optimization set of rules (PSO), in order that it could play a position within the discrete hassle. A hill hiking algorithm was additionally brought into the PSO that allows you to improve the nearby seek ability and to hold the variety of the populace. The simulations effects show that the GHPSO achieves higher performance than fashionable particle swarm algorithm used in reduce costs inside a given execution time.

Pandey et al. (2010) defined that user programs may incur large information retrieval and execution costs when they may be scheduled taking into account only the 'execution time'. Further to optimizing execution time, the cost springing up from statistics transfers among resources in addition to execution expenses need to be taken into consideration. They offered a particle swarm optimization (PSO) based heuristic to timetable applications to cloud sources that take into account each computation cost and statistics transmission price and evaluate with current 'Best Resource Selection' (BRS) set of rules.

Chen et al. (2009) proposed an ant colony optimization (ACO) algorithm to schedule massive-scale workflows with numerous QoS parameters. This algorithm permits customers to specify their QoS options as well as outline the minimal QoS thresholds for software. The goal of this set of rules was to find a solution that meets all QoS constraints and optimizes the user-preferred QoS parameter. The author designed seven new heuristics for the ACO technique and proposed an adaptive scheme that lets in synthetic ants to select heuristics based on pheromone values.

III. WORK METHODOLOGY

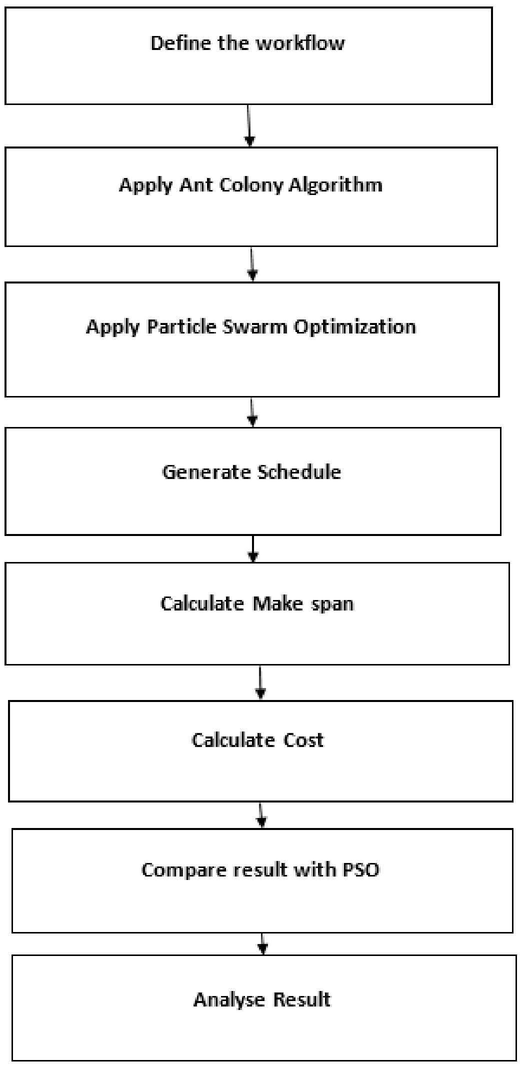
Proposed Methodology

- **Define the workflow** Different workflows from different scientific areas are chosen and simulated in a cloud environment.
- **Apply Ant Colony Optimization Algorithm:** 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.
- **Apply Particle Swarm Optimization Algorithm:** Output ACO will be given as input to PSO for finding the best path**.** PSO is a problematic optimization technique in which the most basic concept is that of the 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.
- **Schedule Generation** Construct a schedule by converting a particle's position.
- **Evaluate Makespan:** Makespan is evaluated to meet the given deadline.
- **Evaluate Cost** For each Workflow cost in terms of money is calculated.
- **Compare the result** Simulate and compare the result of proposed algorithm with PSO and ACO.

Proposed Algorithm

Input: Workflows like Cybershake, Ligo, Genome, Montage
Output: Optimal value of Scheduling the task

1. Input Workflow X
2. Topological Sort (X)
3. For I ϵ X
   {
   Provide VM randomly i ++ :
   i ++ :
   }
4. Initialize ACO parameter
   Xt^{best} Time taken by VM
5. For i ϵ X
   {
   Determine the Value C_{xt}^{best}
   Save X_t^{best}
   }
   If $f(X_t^{best})^{new} \leq f(X_t^{best})^{old}$ then $X = (X_t^{best})^{new}$
   Else $X^{min} = (X_t^{best})^{old}$ .eq. 1
6. If X_t^{best} is not converge
7. Initialize
   For i ϵ X
   {
   Initialize local velocity X^t
   Initialize global velocity Y^t
   }
   For each task
   {
   Update velocity
   Update VM
   For i = 1,2,3…………X



If ($X^t < X_t^{best}$)
$X_1^t = X_t^{best}$
}
For each task
{
If ($Y^t < Y_t^{best}$)
$Y_1^t = Y_t^{best}$
}

8. Compare local and global best to eq. 1
9. If $X_t^{ACO} < Y_t^{PSO}$
{
Then set threshold X_t^{ACO}
}
Else
Y_t^{PSO}
10. VM update or migration according to Y_t^{PSO}
11. Terminate the process

IV. EXPERIMENTAL SETUP AND RESULT ANALYSIS

The proposed algorithm is implemented on HP machine with Intel core2 duo processor, 400 GB HDD, and 3 GB RAM on LINUX platform as it is difficult to run proposed algorithm on real systems because access to the infrastructure incurs a very high cost. Thus the proposed algorithm is implemented in a simulated environment using Cloudsim 3.0.3 simulation toolkit. Cloudsim is programmed in JAVA and the Cloudsim classes are extended to implement the proposed work.

Parameter Setting

To analyze and compare the proposed algorithm with PSO algorithm for scientific workflows on cloud [14] few parameters have been set to a pre-defined value. Various parameters of data center, virtual machines and cloudlets have been set to the different values. A number of iterations of different scientific workflows is analyzed and results are compared to two different parameters (i) cost (ii) time delay with PSO algorithm on IAAS cloud computing environment.

Expected Outcome

Table 1 Genome

	Table 1 Genome Workflow			
VM	Average Cost(PSO)	Average MakeSpan(PSO)	Average Cost(ACO_PSO)	Average MakeSpan(ACO_PSO)
two	148.98	43434.998	43.23	711.2482
four	618.57	63804.7918	105.35	23771.5127
six	843.36	94768.4994	515.7	35311.0717
eight	923.37	116380.9196	448.32	37130.2134
ten	632.28	163289.5874	434.55	58144.186
twelve	2111.13	208559.28	557.87	79042.3512
eighteen	2930.28	275656.083	1113.83	110300.8237
twenty five	2174.79	392256.841	1434.97	166505.8034
thirty	3148.62	478408.0798	1533.94	191834.0959

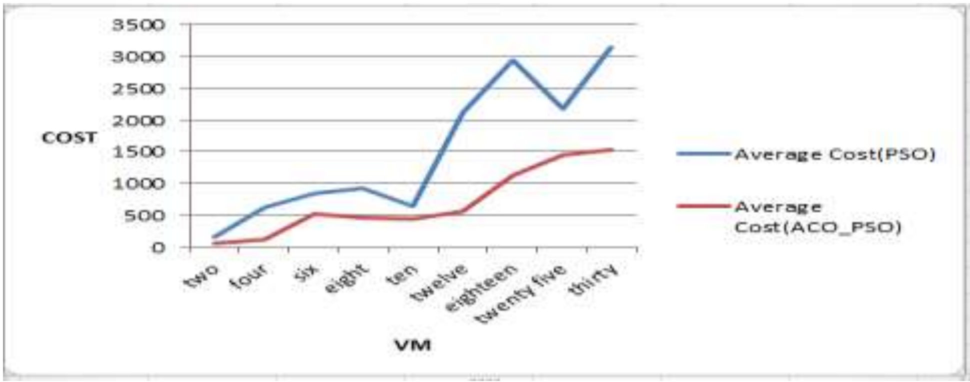


Figure 1 Cost of Genome Workflow

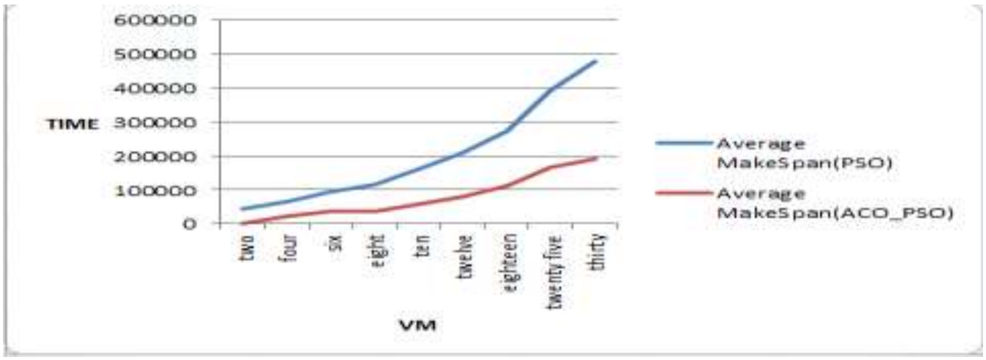


Figure 2 Makespan of Genome Workflow

TABLE 2 Cybershake

	Table 2 CyberShake Workflow			
VM	Average Cost(PSO)	Average MakeSpan(PSO)	Average Cost(ACO_PSO)	Average MakeSpan(ACO_PSO)
two	3.18	264.4448	1.6	13.656
four	7.8	875.3528	2.1	478.1314
six	14.67	1610.1014	6.5	524.9606
eight	18.15	2041.2912	8.1	1135.1083
ten	16.71	2638.0596	10	1462.4828
twelve	26.82	3055.3436	15	1655.3824
eighteen	31.41	3916.8388	15.59	2916.5683
twenty fi	63.63	4421.4986	24.13	3774.9163
thirty	69.48	5763.0504	26.28	4601.9896

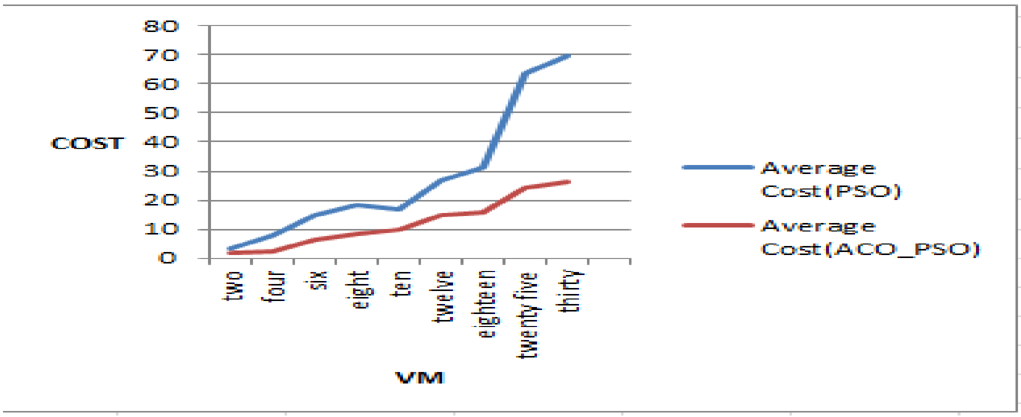


Figure 3 Cost of Cybershake Workflow

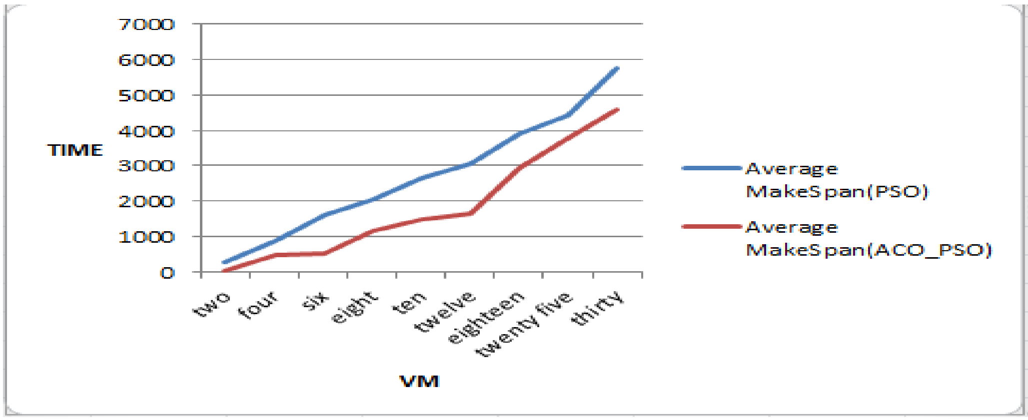


Figure 4 Makespan of Cybershake Workflow

Table 3 LIGO

	Table 3 LIGO Workflow			
VM	Average Cost(PSO)	Average MakeSpan(PSO)	Average Cost(ACO_PSO)	Average MakeSpan(ACO_PSO)
two	28.53	3083.9904	12	710.6529
four	24.75	5049.356	15.1	1862.3855
six	75.78	8679.1234	32.9	3655.9573
eight	101.76	11238.7914	33.87	4251.1423
ten	104.82	13411.725	56.55	5443.2358
twelve	116.01	15795.5116	58.41	6172.0119
eighteen	397.62	23974.4658	119.49	10160.8733
twenty fiv	405.9	32741.8312	122.18	13065.4069
thirty	690.15	38695.7594	210.6	18442

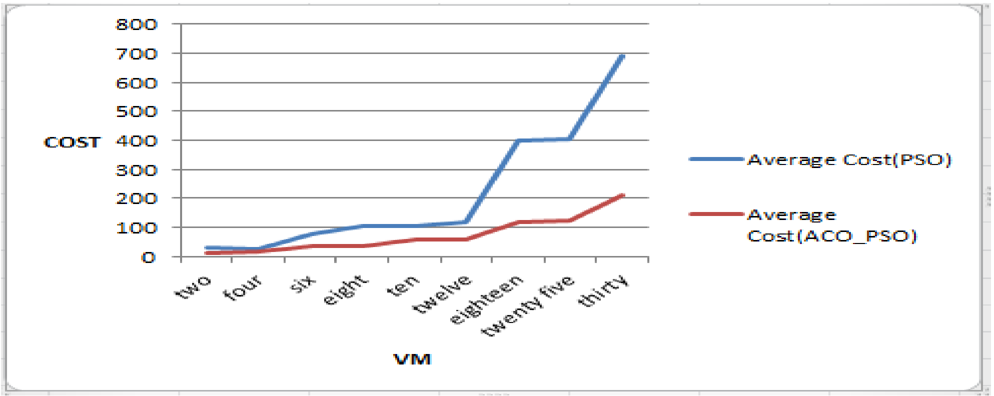


Figure 5 Cost of Ligo Workflow

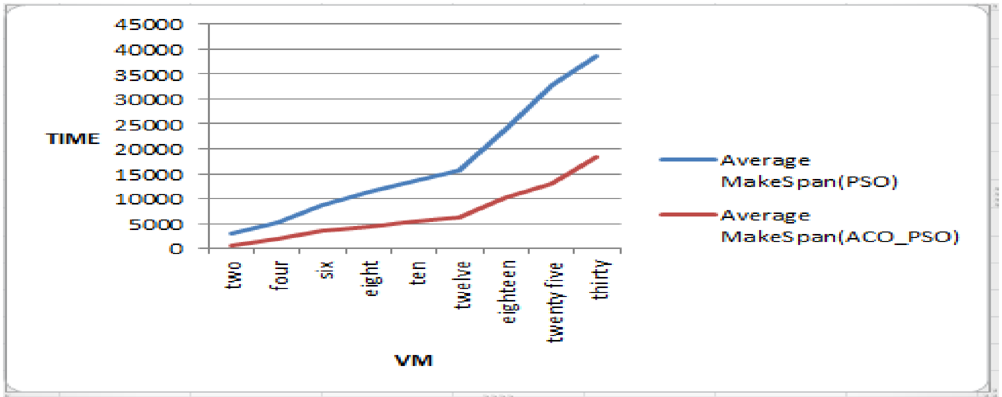


Figure 6 Makespan of Ligo Workflow

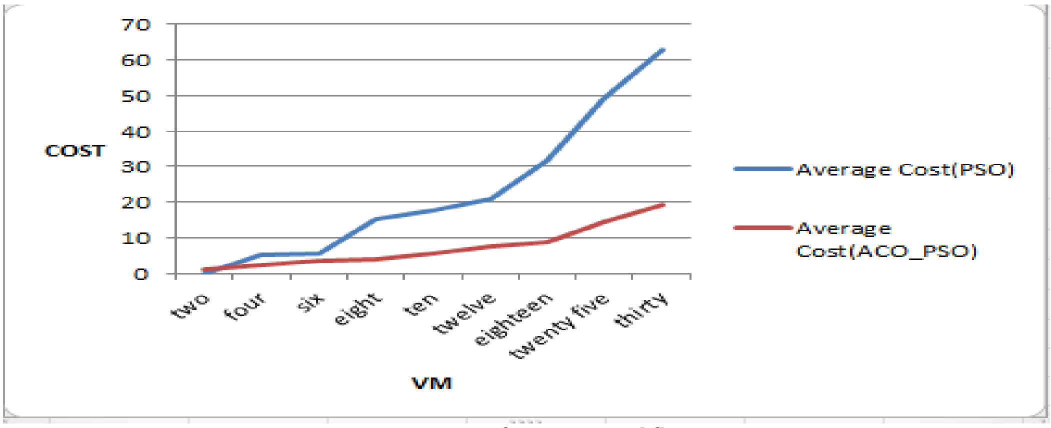


Figure 7 Cost of Montage Workflow

Table 4 MONTAGE

	Table 4 Montage Workflow			
VM	Average Cost(PSO)	Average MakeSpan(PSO)	Average Cost(ACO_PSO)	Average MakeSpan(ACO_PSO)
two	0	0	1.1	3.918
four	4.86	642.4224	2.1	258.1832
six	5.52	784.07	3.6	194.0196
eight	14.97	1380.5446	4	625.8844
ten	17.67	1748.8282	5.5	940.2801
twelve	20.76	1997.1848	7.5	1319.0793
eighteen	31.65	2670.6732	8.5	1475.1686
twenty fiv	48.99	4417.6592	14.5	2330.4413
thirty	62.67	5098.9976	19	2707.9055

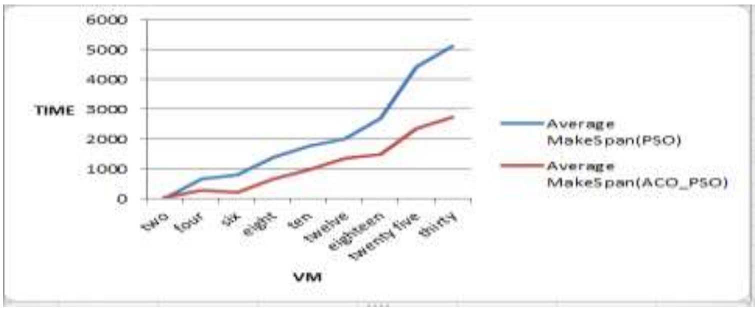


Figure 7 Make span of Montage Workflow

CONCLUSIONS

In this thesis, we presented a scheduling strategy for executing scientific workflows on IaaS Clouds. The scenario was modelled as an optimization problem which aims to minimize the overall execution cost while reducing the make span and the problem was solved using the hybrid of ACO and PSO. The experiments were conducted by simulating four well-known workflows (Cybershake, Ligo, Genome, and Montage) on Clouds, which shows that our solution has an overall healthier performance than other state-of-the-art algorithms. The worthy results are achieved because PSO (particle swarm optimization) play important role in global optimization and ACO (ant colony optimization) optimize locally and we have merged the two algorithms by taking the best out of them. With the proposed approach in most of the workflows, we are able to produce lower cost efficient schedule meanwhile also reducing the time delay.

As future work, we would like to explore various options for the selection of the preliminary resource pool as it has a major impact on the performance of the algorithm. We would also like to research with different optimization approaches such as genetic algorithms and compare their performance with PSO, ACO, and Hybrid of both these algorithm. Another future work is to study the data transfer cost between data centers so that VMs can be deployed in different locations. Finally, we wish to implement our approach in a workflow engine so that it can be utilized for deploying applications in real life environments.

REFERENCES

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

[2] Alkhanak,E.N; Lee,SP. and Khan,S.(2015), "Cost-aware challenges for workflow scheduling approaches in cloud computing environments: Taxonomy and opportunities", Future Generation Computer Systems, vol. 50 pp. 3-21.

[3] Ahmad, S.G.; Liew,C.S; Munir,E.U.; Ang,T.F. and Khan,S.U.(2016), "A hybrid genetic algorithm for optimization of scheduling workflow applications in heterogeneous computing systems", Journal of Parallel and Distributed Computing, Elsevier, vol. 87, pp. 80-90.

[4] Bala,A. and Chana,I.(2011) "A Survey of Various Workflow Scheduling Algorithms in cloud Cloud Environment", 2nd National Conference on Information and Communication Technology.

[5] Bilgaiyan,S.; Sagnika,S. and Das,M.(2014), "An Analysis of Task Scheduling in Cloud Computing using Evolutionary and Swarm-based Algorithms", International Journal of Computer Applications vol. 89, no. 2.

[6] Bittencourt, L. and Madeira,E.(2011), "HCOC: A Cost Optimization Algorithm for Workflow Scheduling in Hybrid Clouds", Journal of Internet Services and Applications, vol.2, no. 3, pp 207-227.

[7] Byun,Eun-Kyu; Kee,Y.S; Kim,J.S and Maeng,S.(2011), "Cost optimized provisioning of elastic resources for application workflows", ELSEVIER Future Generation Computer Systems, vol. 27, no. 8, pp. 1011-1026.

[8] Chen,Wei-Neng and Zhang,Jun(2009), "An ant colony optimization approach to a grid workflow scheduling problem with various QOS requirements", IEEE Transactions Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 39, no.1 pp. 29-43.

[9] Delavar, A. Ghorbannia and Aryan,Y.(2014), "HSGA: a hybrid heuristic algorithm for workflow scheduling in cloud systems", Cluster computing, Springer, vol. 17, no. 1, pp. 129-137.

[10] Dillon, Tharam; Wu, Chen and Elizabeth,C.(2010), "Cloud computing: issues and challenges",24th IEEE International Conference on Advanced Information Networking and Applications (AINA), pp. 27-33.

[11] Kalra,M. and Singh,S.(2015), "A review of Metaheuristic Scheduling Techniques in Cloud Computing", Egyptian Informatics Journal 16, no. 3 pp. 275-295.

[12] Malawski,Maciej; Juve,G.; Deelman,E. and Nabrzyski,J.(2012), "Cost-and deadline-constrained provisioning for scientific workflow ensembles in IaaS clouds", Proceedings of the International Conference on High-Performance Computing, Networking, Storage and Analysis, pp. 22.

[13] Mao,Ming and Humphrey,M.(2011), "Auto-Scaling to Minimize Cost and Meet Application Deadlines in Cloud Workflows", International Conference High-Performance Computing, Networking, Storage and Analysis (SC), pp. 1-12.

[14] Mell, Peter and Grance ,T.(2011),"The NIST definition of cloud computing", NIST Special Publication 800-145.

[15] Nancharaiah,B. and Chandra Mohan,B.(2014), "The performance of a hybrid routing intelligent algorithm in a mobile ad hoc network", Computers & Electrical Engineering vol. 40, no. 4 pp. 1255-1264.

[16] Netjinda,Nuttapong; Sirinaovakul,B. and Achalakul,T.(2014), "Cost optimal scheduling in IaaS for dependent workload with particle swarm optimization", Journal of Supercomputing, pp. 1-25.

[17] Pandey,Suraj;Wu,L.; Mayura,S. and Buyya,R.(2010),"A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments", 24th IEEE International Conference on Advanced Information Networking and Applications (AINA), pp. 400-407.

[18] Rahman, Mustafizur; Li,Xiaorong and Palit, Henry (2011), "Hybrid Heuristic for Scheduling Data Analytics Workflow Applications in Hybrid Cloud Environment", IEEE International Parallel & Distributed Processing Symposium, pp. 966-974.

[19] Rodriguez Sossa, M. and Buyya,R.(2014) ,"Deadline based Resource Provisioning and Scheduling Algorithm for Scientific Workflows on Clouds", IEEE Transactions On Cloud Computing, vol. 2, no. 2, pp.222-235.

[20] Sajid, Mohammad, and Z. Raza, "Cloud Computing: Issues and Challenges," in International conference on Cloud, Big Data and Trust, pp. 35 – 41, Nov.2013.

[21] Sridhar, M. and Babu, G.R.M ," Hybrid Particle Swarm Optimization scheduling for cloud computing" in IEEE International Advance Computing Conference (IACC), pp. 1196- 1200, June 2015.

[22] Tilak,S. and Patil,D.(2012), " A Survey of Various Algorithms in Cloud Environment", International Journal of Engineering Inventions, vol.1 no. 2,pp. 36-39.

[23] Tripathy,L. and Patra,R. R(2014), "Scheduling in Cloud Computing", International Journal on Cloud Computing: Services and Architecture vol.4, no.5, pp. 21-27.

[24] Verma,A. and Kaushal,Sakshi(2014), "Bi-Criteria Priority based Particle Swarm Optimization workflow scheduling algorithm for cloud", Engineering and Computational Sciences (RAECS), pp. 1-6.

[25] Verma,A. and Kaushal,S.(2015), "Cost-Time Efficient Scheduling Plan for Executing Workflows in the Cloud", Journal of Grid Computing, Springer, vol. 13, no. 4, pp. 495-506.

[26] Xu, Meng; Cui, Lizhen; Wang,H. and Yanbing,B.(2009), "A multiple QOS constrained scheduling strategy of multiple workflows for cloud computing", IEEE International Symposium Parallel and Distributed Processing with Applications, pp. 629-634.

[27] Xue,Sheng-Jun and Wu, Wu(2012) , "Scheduling workflow in cloud computing based on hybrid particle swarm algorithm ", TELKOMNIKA Indonesian Journal Of Electrical Engineering, vol. 10, no. 7 pp. 1560-1566.

[28] Zhang,Qi; Cheng ,Lu and Boutaba,R.(2010), "Cloud computing: state-of-the-art and research challenges" ,Journal of Internet Services and Applications, vol. 1, pp. 7-18.

[29] Zhan, Shaobin and Huo, Hongying(2012), "Improved PSO-based task scheduling algorithm in cloud computing", Journal of Information & Computational Science vol. 9, no. 13, pp. 3821-3829.

[30] Zhu, Linan;Li,Qingshui and Lingna,H.(2012), "Study on cloud computing resource scheduling strategy based on the ant colony optimization algorithm", International Journal of Computer Science Issues, vol.9, no. 5.

