A new whale optimizer for workflow scheduling in cloud computing environment

Sounder Rajan Thennarasu, M. Selvam & K. Srihari

Journal of Ambient Intelligence and Humanized Computing

ISSN 1868-5137

J Ambient Intell Human Comput
DOI 10.1007/s12652-020-01678-9

ONLINE FIRST

Springer

Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-020-01678-9

ORIGINAL RESEARCH

A new whale optimizer for workflow scheduling in cloud computing environment

Sounder Rajan Thennarasu[1] · M. Selvam[2] · K. Srihari[3]

Received: 16 October 2019 / Accepted: 2 January 2020

Abstract
Cloud computing environments enable real time applications on virtualized resources that can be provisioned dynamically. It is one of the efficient platform service which permits to enable the various applications based on cloud infrastructure. Nowadays workflow systems become an easy and efficient task for the development of scientific applications. Efficient workflow scheduling algorithms are employed to improve the resource utilization by enhancing the cloud computing performance and to meet the users' requirements. Many scheduling algorithms have been proposed but they are not optimal to incorporate benefits of cloud computing. In this paper a new framework are introduced as whale optimizer algorithm (WOA) which mimics the social behaviour of humpback whales and aims to maximize the work completion for meeting QoS constraints such as deadline and budget. This proposed method outperforms well when compared with other techniques and measured in terms of makespan, deadline and it is applicable for real time applications.

Keywords Wireless communication · Service and semantic computing · Autonomic computing · Whale optimizer · Makespan · Bubble-net search mechanism

1 Introduction

Cloud computing is the most recent and rising pattern in information technology field. It offers utility-based IT administrations to client over the internet (Rehman et al. 2013). The quick development of virtualization has made cloud computing an inventive stage to handle logical and scientific issues. These issues are essentially tackled through the cloud computing worldview without purchasing any infrastructure (e.g. computational resources, arrange, capacity). Nonetheless, the end-users pay for whatever they utilize. Cloud computing gives a pool of virtualized resources, including computing power, storage and programming applications over the internet in view of users request.

Workflow scheduling is one of the challenging constraint in cloud environment (Bardsri and Hashemi 2012). A good workflow scheduling algorithm should restrict the computation time and cost of workflow application. A workflow is as sequence of connected instructions. Scheduling of workflows is an issue of finding a right execution sequence for workflow activities (Kaur et al. 2011). The vulnerability about learning of parameters like number of processing resources accessible with their speed and capacity, the bandwidth variations, accessibility of assets requires service providers as well as service users to be progressively concerned for guaranteeing minimum Quality of Service (QoS). The workflow management system is charged in the cloud and permits the cloud service providers to enhance the efficiency by way of flexible resource allocation, scalability, reduce node communication loss, fault tolerance in the cloud and, as a result, greatly minimize the cost of operation. There is also a need to address the failure and improve the effectiveness of scheduling algorithm (Shi and Dongarra 2006). This complication causes the workflow applications, networks environments, security and scheduling algorithms are more complexity.

✉ Sounder Rajan Thennarasu
ksrsound2009@yahoo.com

M. Selvam
ammselvam@gmail.com

K. Srihari
harionto@gmail.com

[1] Department of Cse, Al-Ameen Engineering College, Erode, India

[2] Department of Cse, HKBKCE, Bangalore, India

[3] Department of Cse, SNSCE, Coimbatore, India

Published online: 11 January 2020

By employing the workflows process in cloud computing which enable us to achieve the benefits of cloud for workflow scheduling constraints. Many workflow scheduling approaches have been suggested by different researchers but it suffers from certain limitations. In this paper a new framework was introduced as whale optimization (Watkins and Schevill 1979) to reduce the workflow complexity among large database.

A data center sometimes called server farm is a centralized repository for the storage, management and dissemination of data and information. Virtual machine is an emulation of a particular computer system. Virtual machines operates based on the computer architecture and functions of a real or hypothetical computer. Their implementation may involve specialized hardware, software or a combination of both. The data center broker component randomly selects the data center irrespective of their heterogeneity in hardware, software configuration and pricing schemes for usage. Then the broker maps the workflows to all the created virtual machines in a circular fashion without considering the processing elements (PEs) required by the tasks. Mohammed et al. (2019) cloud performance is monitored by service level agreement. The rest of the study is systematized as follows. Section 2 deals with the survey about the workflow scheduling methods. Section 3 deals with the proposed methodology based workflow scheduling. Section 4 reveals about the mathematical model of whale optimizer algorithm-based workflow scheduling. Section 5, gives the result and performance analysis. Finally, the overall proposed method concludes in Sect. 6.

2 Literature survey

Efficient scheduling of workflow applications can be achieved by weighted directed acyclic graphs (DAG). Shi and Dongarra (2006) exhibited a novel rundown scheduling based algorithm. In the beginning stage, it considers the impact of percentage of capable processors (PCP) and assigns task node weights accordingly. Next in order, when selecting the processor, the algorithm fine-tunes the effective earliest finish time. The algorithm is assessed utilizing an extensive arrangement of generated task graphs and the outcomes demonstrate that each characteristic of the SDC algorithm improves the schedule length. Examination contemplate demonstrates that SDC calculation algorithm superior to related work in general (Durillo and Prodan 2014) discussed about Multi-Objective Heterogeneous Earliest Finish Time (MOHEFT), a Pareto-relied content skeleton heuristic which provides the stakeholders with an assorted manner of tradeoff optimal solutions from which the one that perfectly coordinates the customers' needs could be physically chosen MOHEFT assembles an intermediate workflow schedules for parallel in each progression rather than a single one. To guarantee the nature of the tradeoff arrangements, MOHEFT utilizes predominance connections and a metric called crowding distance to ensure their decent variety in an Amazon-based commercial cloud. The outcomes show that the approach can compute solutions of higher quality than SPEA2*.

Workflow scheduling is one of the actual problems in cloud frameworks. An efficient scheduling algorithm must limit the execution time and cost of workflow process application alongside QoS necessities of the client. Singh and Singh (2013) consider deadline as the significant requirement and propose a score based deadline compelled workflow scheduling algorithm that executes workflow inside reasonable cost while meeting client characterized deadline limitation. Zhan and Huo (2012) suggested the enhanced particle swarm optimization method in resource framing technique of the cloud computing. This paper combines PSO and scheduling algorithm with characteristics of job scheduling to get the mixed scheduling algorithm. Scheduling algorithm (SA) is easy to sink into local optima with serial search. As a whole, genetic algorithm and scheduling algorithm spends more time as the number of tasks increase. Ant colony optimization algorithm performs task slowly in the beginning, but later its time increase in minimum than GA and SA algorithm with enhanced positive feedback (Mirjalili and Lewis 2016) brings new-age environment-impact multi objective optimization algorithm, named whale optimization algorithm. Komaki and Kayvanfar (2015) addresses gray wolf optimizer (GWO) which is motivated by living and hunting manners of wolves is suggested to which local search methods are used to improve the nature of the suggested algorithm. Experimental outcomes describe that the suggested dispatching rules are genuinely great. Contrasting GWO with other famous algorithms that GWO has better functioning. The proposed GWO as a basic and powerful algorithm could be connected to different issues. Elghamrawy and Hassanien (2019) proposed new algorithm known as hybrid genetic whale optimization algorithm (GWOA) to optimize the spectrum utilization. The GWOA algorithm adds the crossover and mutation operations with WOA algorithm to attain the balance between exploration and exploitation phases, to obtain the best results based on a fitness function.

Naseri and Jafari Navimipour (2019) proposed a new technology for well-organized service composition in the cloud. The particle swarm optimization (PSO) algorithm is employed for selecting the best services based on fitness function and the agent-based technique is taken to arrange services by choosing the QoS parameters.

An Adaptive Niche Hierarchy Genetic Algorithm (ANHGA) is proposed by Ye et al. (2011). The algorithm relies on the versatile mutation operator and crossover operator that modifies the crossover rate and intermittence of variation of every person, and embraces the slope of the person to select their mutation value.

Tawfeek et al. (2015) discussed a cloud task scheduling algorithm relying on ant colony optimization (ACO) which is random optimization search was simulated availing the Cloudsim toolkit package and compared with various scheduling algorithms like FCFS and round-robin. The functioning outcomes are compared to First Come First Served (FCFS) and Round Robin (RR) revealed that the ACO algorithm provides better outcomes.

3 Research methodology

Cloud computing offers a shared pool of resources such as data storage space, networks, computer processing power and specialized corporate and user applications. There is some uncertain in allocating and scheduling the workload in cloud environment and it can be reduced by generating the effective workflow scheduling algorithm according to the client's requirements. A new framework of bio-inspired optimization algorithm based on hunting process of bubble net humpback whales named as whale optimization to reduce the workflow complexity in large database. It is implement with certain adaptive techniques to reduce the execution time for complex problems.

3.1 Whale optimization algorithm

A new technology was introduced known as whale optimization algorithm (WOA) which is multi objective and parameter free optimization algorithm relied on the hunting manners of humpback whales. Unlike grey wolf optimization, the hunting behavior of WOA is the finest search optimizer to hunt the prey by spiral behavior of bubble-net mechanism of humpback whales. The world's biggest mammals are whales in which adult whales are 30 m long and 180 t weight. Whales can never sleep since it has to breathe from deep ocean's surface so it is known as predators and half of whose brain only sleeps.

According to existing method whales are similar to humans because it has some similar cells in brain area called spindle cells which are responsible for judgment, feelings and social behaviors. Figure 1 reveals their special hunting technique of the humpback whales. This type of searching behavior named as bubble-net feeding, it only applicable for humpback whales and this method for which humpback whales are chosen to hunt school of krill or small fishes under bottom surface by forming bubbles net as circle shape or '9'-shaped path.

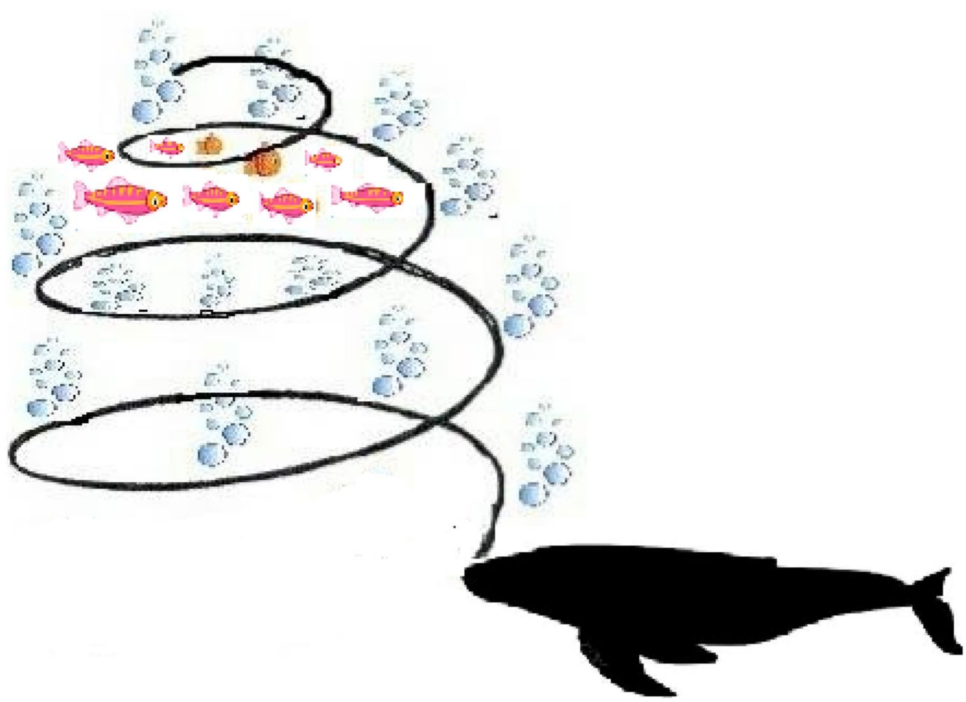

Fig. 1 Humpback whales bubble-net feeding behavior

4 Mathematical model and optimization algorithm

In this section firstly the mathematical structure of enclosing victim, spiral bubble-net feeding maneuver and hunting victim is provided and secondly the proposed algorithm is employed.

4.1 Encircling prey

Initially humpback whales identify the prey location and surround them. In the search space, the optimal design for unknown location is found by the best whale as the target prey or which is close to the optimum design. After the search done by the best search agent, the rest of the search agents would update their locations with respect to the best search agent which is indicated by

$$\vec{D} = \left| C \cdot \vec{X^*}(t) - \vec{X}(t) \right| \tag{1}$$

$$\vec{X}(t+1) = \vec{X^*}(t) - \vec{A} \cdot \vec{D}. \tag{2}$$

From the above equations, the prevailing iteration is denoted by *t,* the coefficient vectors are $\vec{A}\, an \vec{A}$ and $CC\, X^* X^*$ indicates the place vector of the best solution, *X represents* the position vector, | | are stated for absolute attribute for element-by-element multiplication. For each iteration $X^{\,*}$ must be updated for a better solution. The vectors A and C are calculated as follows

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a}\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \tag{3}$$

$$\vec{C} = 2 \cdot \vec{r}\vec{C} = 2 \cdot \vec{r} \tag{4}$$

where a is linearly decreased from 2 to 0 for every repetition (both exploitation and investigation phases) and r denotes as random vector in [0, 1].

4.2 Bubble-net attacking method (exploitation phase)

The two methods such as shrinking and spiral are employed for describing the mathematically model of humpback whales bubble-net behavior, as follows.

4.2.1 Shrinking encircling mechanism

The shrinking behavior is realized by decreasing the value in the Eq. (3) with that the A range value also decreased with respect to a. [− a, a] can be described as a attribute in the gap where a decreased from 2 to 0 for each iterations. Random attribute for A in the interval [− 1, 1], can be set by which it is possible to find the modern upgraded position of a seeking agent betwixt the naive position of the agent and the best agent position as revealed in Fig. 2.

4.2.2 Spiral updating position

A spiral equation for updating the location is generated between the location of whale and location of victim to mimic the spiral shaped humpback whales. They are indicated as

$$\vec{X}(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X^*}(t) \tag{5}$$

where $D' = \left|\vec{X^*}(t) + \vec{X}(t)\right| D' = \left|\vec{X^*}(t) + \vec{X}(t)\right|$ denotes the gap between the ith whale position to the prey (best solution), b is constant for stating the spiral shape, l is a random number between the interval [− 1, 1] for step-by-step multiplication. In the shrinking circle the humpback whales are swim around the prey beside a spiral-shaped path. The probability of 50% of spiral mechanism is assumed to estimate the hunting behavior for updating the whale's position for optimization by choosing either the shrinking model or the spiral model. The hunting behavior based mathematical model is

$$\vec{X}(t+1) = \begin{cases} \vec{X^*}(t) - A \cdot D & if\ p < 0.5 \\ D' \cdot e^{bl} \cdot \cos(2\pi l) & if\ p \geq 0.5 \end{cases}. \tag{6}$$

In which p denotes the random number between the interval [0, 1].

4.2.3 Search for prey (exploration phase)

The bubble net humpback whales seek for prey randomly rely on the position which is based on same approach in which the A vector can be used prey searching (exploration). If values are greater than 1 or less than − 1, it forces the prey seeking agent to go out of the way from a target whale.

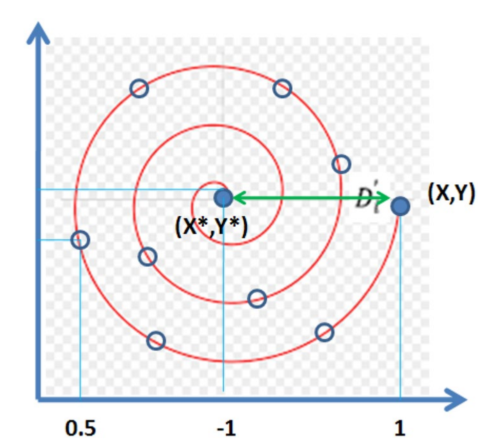

Fig. 2 Bubble-net search mechanism implemented in WOA

In exploration phase the positions of the search agents are updated (either the best solution obtained or the randomly chosen search agent) for each repetition and the search agent parameter is minimized from 2 to 0 to present exploration and exploitation, respectively. If $|A| < 1$ the best solutions are selected for updating the search agent position and for $|A| > 1$, the random search agents are selected. Depending on the value of p, WOA is able to switch between either a spiral or circular movement and the WOA algorithm is ended by the termination criterion.

From theoretical behavior WOA can be denotes as a global optimizer because it includes both exploration/exploitation ability. The exploration is the potential of the algorithm to explore for new solutions far from the current best solution in the search space. Exploitation is to search the neighboring search area nearby the current solution. Also, the proposed hunting mechanism describes a search space as best solution and allows other search agents to exploit the current best agent inside that domain. The search vector A allows the WOA algorithm to smoothly transit between exploration and exploitation, by decreasing A, in which some iterations are dedicated to exploration ($|A| \geq 1$) and the remaining dedicated to exploitation ($|A| < 1$). The WOA includes only two main internal parameters to be adjusted for search agents (A and C).

4.3 Whale optimization for workflow scheduling

The pseudo code of whale optimization in cloud environment for workflow scheduling are given below.

The scheduling whale optimizer algorithm is implemented to solve workflow scheduling problem in cloud environment. Workflow (B) is given to the system as input data randomly. Then the cycle elimination (B_O) is used to remove the path that violates the topological order from the existing paths. Then the flow of work is generated as input to the WOA (I). The WOA schedules the workflow by optimizing the remaining path.

Symbol explanation table.

Symbols	Explanation
B	Nodes, edges
X	Best search
A	Attribute
t	Time
N	Nodes

WOA optimizer will calculate the fitness value of each search agent. Each value is scored and either accepted or rejected before considering it for the next generation. WOA updates the position of the current search agent after achieving maximum number of iterations. Update the fitness by calculating of all search agents. Then, the workflow is submitted and the process repeats itself until next scheduling event occurs. Then $Nodes_{Migr}$ is used where it consists of all the executing tasks that need migration because of performance contract violation or failures. The running workflow with time instance t is defined (B_t). Finally reschedule is made for workflow with time instance t.

Step-1Input data: workflow: B = (Nodes, Edges);

Step-2Cycleelimination: B_O = (Nodes, Edges – $Edges_{Queued}$)

Step-3Schedule**:**I = Whale optimization (B_O);

a. Initialize the grey wolf population X_i (i = 1, 2, ..., n)

b. Compute the fitness of each search agent

c. X*=the best search agent

d. **while** (t < Max number of iteration) **for** each search agent

e. Update *a, A, C, l* and *p*

f. If1 (p<0.5)

g. If 2 (|A| < 1)

h. *Change current search agent position by the Eq.(1)*

i. Else if2 (|A| ≥1)

j. *Select a random search agent* $\overrightarrow{X_{rand}}$

k. *Change current search agent position by the Eq.(8)*

l. End if2

m. Else if 1 (p ≥0.5)

n. *Change current search position by the Eq. (5)*

o. End if1

p. Check if any search agent goes beyond the search space and calculate the fitness value

q. Update X* if there is better solution

r. t=t+1

s. end while

t. return X^*

Step-4 submit workflow:execute(B,)
Step-5 repeat
Step-6 t = sleep until next scheduling event
Step-7 select tasks for migration:
Step-8 $Nodes_{Migr}$ = {N ∈ Nodes | state (N, t) = failed ∨

Step- 9 state (N, t) = running ∧PC(N, t) >f_N};

Step-10Bt = generate static DAG (B, I, t, $Nodes_{Migr}$)

Step-11 cancel (N), ∀ N ∈$Nodes_{Migr}$;

Step-12 **reschedule:**I = Whale optimizer (B_t)

Step-13 Until state(N,t) = completed, ∀ N ∈ Nodes ∧

succ(N) = φ.

5 Experimental results

The performances of proposed whale optimizer for workflow scheduling are evaluated by Hadoop under Linux environment. The whale optimizer was tested to reduce the workflow scheduling complexity and measured in terms of makespan, scheduling time, deadline hit and resource utilization are used with respect to processing capacity and memory space such as 128×8, 256×16, 512×32 and 1024×64. These measures are allowed to evaluate the exploitation capability of the investigated meta-heuristic algorithms.

5.1 Scheduling model

Since 1970s the process of automatic process developed in office environment led to the development of workflow which might be involved in achieving the easier commercial activities. Sequence of stages will be considered in workflow, that streamlines and reduces the complicatedness in the process of implementation along with administration of workflow.

Tool for explaining, producing along with handling the implementation of workflows was performed with the help of workflow administration framework. Explanation in the manner of activities to processing arrangement has to be incorporated when performing the formulating process of workflow.

Illustration of workflows in the manner of directed acyclic graph. Directed acyclic graph demonstrated in Fig. 3.

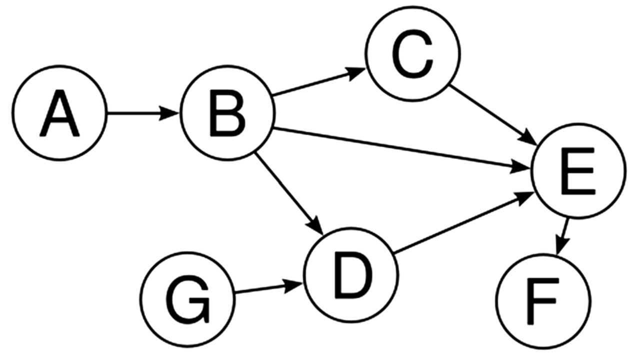

Fig. 3 Directed acyclic graph

Here, the task B might initialize execution subsequent to finishing the activity A. Similarly, activity E can get executed only after the completion of tasks B, C and D. And finally, task F can be executed only if all other tasks complete their execution.

Huge quantity of reserve utilization along with minimization of processing duration are considered as prominent motivation existed with the framework. Utilization of one activity with respect to simulated processing equipment will be allocated by prototype providing the guarantee which during the instance all computation will be carried over by one simulated processing equipment.

Although, processing equipment will be received from the similar infrastructure which contains various scheduling and processing equipment without flaws. Allocating tasks, monitoring, implementation, failed task can be re-run and remove the errors are the performance which is responsible for master node. This master node is used to group the statistics belonging to equipment which might be taking part in storage space, capacity along with bandwidth.

Implementation of the task is responsible for slave node which is assigned by master node. Then, the suitable computing node is identified by using slave node. Data is distributed through these nodes. It considers the work load on each node. Some of the node may be always busy and other some node may be idle. The nodes are computationally equal by using this proposed system.

This proposed WOA method was compared with existing genetic algorithm, particle swarm optimization, ant colony optimization and grey wolf optimization to show the better results. Table 1 shows the comparison of WOA workflow scheduling algorithm with respect to different memory space.

Table 1 Comparison of whale optimizer

Scheduling algorithm	Resource matrix	Makespan (s)	Dead line hit (%)	Resource utilization (%)
GA	128×8	59	61	49
PSO		56	65	53
ACO		52	71	57
GWO		47	73	61
WOA		43	75	69
GA	256×16	57	69	71
PSO		51	72	76
ACO		43	76	79
GWO		39	82	81
WOA		37	84	85
GA	512×32	47	79	75
PSO		41	82	79
ACO		35	85	81
GWO		31	86	86
WOA		29	89	89
GA	1024×64	41	81	81
PSO		35	83	84
ACO		30	86	86
GWO		25	89	88
WOA		21	91	92

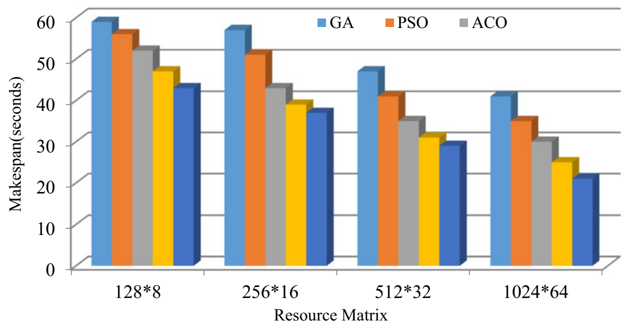

Fig. 4 Comparisons for make span

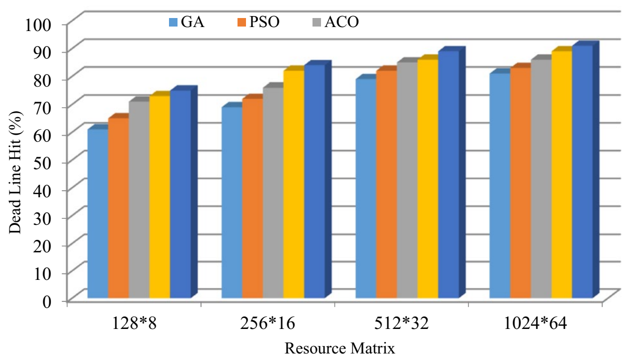

Fig. 5 Comparisons for dead line

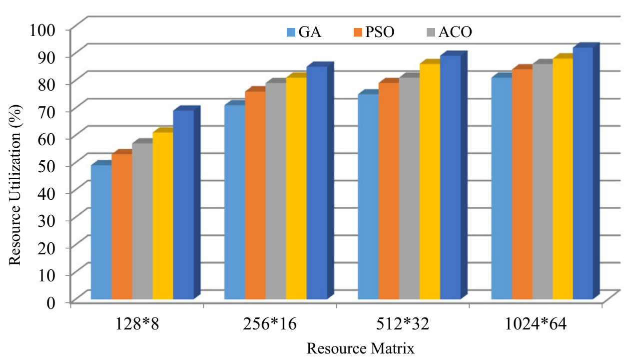

Fig. 6 Comparisons for resource utilization

Figure 4 shows the comparison of makespan for proposed whale optimizer based workflow schedule. The proposed method of WOA has less makespan than other existing technique.

The comparison for dead line is shows in Fig. 5. The proposed WOA has achieved efficient deadline with high rate when compare to existing technique.

The resource utilization is evaluated for proposed WOA are shown in Fig. 6. The proposed WOA exactly balances the load in available resource. Compared to existing technique, the resource utilization is balanced in WOA for workflow scheduling.

6 Conclusion

Nowadays cloud computing is an emerging technique in IT based technologies, in which the workflow scheduling systems are considered to enable the cloud infrastructure to support large scale real time applications such as E-business and E-science. In cloud computing managing the scheduling resources are quite complex, so it necessary to introduce a new framework for analysis the scheduling algorithm. The proposed whale optimizer is employed to improve the workflow scheduling constraints and balances the load among resources by efficient tasks distribution in cloud environment. The performances of proposed WOA are evaluated using makespan, deadline hit and resource utilization parameter as a Multi-objective Optimization Problem (MOP) for the cloud environments. It shows that proposed WOA obtain better results when compared to other existing techniques and it is applicable for real time applications.

References

Bardsiri AK, Hashemi SM (2012) A review of workflow scheduling in cloud computing environment. Int J Comput Sci Manag Res 1(3):348–351

Durillo JJ, Prodan R (2014) Multi-objective workflow scheduling in Amazon EC2. Clust Comput 17(2):169–189

Elghamrawy SM, Hassanien AE (2019) GWOA: a hybrid genetic whale optimization algorithm for combating attacks in cognitive radio network. J Ambient Intell Hum Comput 10:4345–4360. https://doi.org/10.1007/s12652-018-1112-9

Kaur N, Aulakh TS, Cheema RS (2011) Comparison of workflow scheduling algorithms in cloud computing. Int J Adv Comput Sci Appl 2(10):89

Komaki GM, Kayvanfar V (2015) Grey Wolf Optimizer algorithm for the two-stage assembly flow shop scheduling problem with release time. J Comput Sci 31(8):109–120

Mirjalili S, Lewis A (2016) The whale optimization algorithm. Adv Eng Softw 95:51–67

Mohammed A, El-Bahnasawy N, Sadi S, Wagdi M (2019) On the design of reactive approach with flexible checkpoint interval to tolerate faults in cloud computing systems. J Ambient Intell Hum Comput 10:4567–4577. https://doi.org/10.1007/s12652-018-1139-y

Naseri A, Jafari Navimipour N (2019) A new agent-based method for QoS-aware cloud service composition using particle swarm optimization algorithm. J Ambient Intell Hum Comput 10:1851–1864. https://doi.org/10.1007/s12652-018-0773-8

Rahman M, Hassan R, Ranjan R, Buyya R (2013) Adaptive workflow scheduling for dynamic grid and cloud computing environment. Concurr Comput Pract Exp 25(13):1816–1842

Shi Z, Dongarra JJ (2006) Scheduling workflow applications on processors with different capabilities. Future Gener Comput Syst 22(6):665–675

Singh R, Singh S (2013) Score based deadline constrained workflow scheduling algorithm for cloud systems. Int J Cloud Comput Serv Archit (IJCCSA) 3(6):89
Tawfeek M, El-Sisi A, Keshk A, Torkey F (2015) Cloud task scheduling based on ant colony optimization. Int Arab J Inf Technol 12(2):129–137
Watkins WA, Schevill WE (1979) Aerial observation of feeding behavior in four baleen whales: *Eubalaena glacialis*, *Balaenoptera borealis*, *Megaptera novaeangliae*, and *Balaenoptera physalus*. J Mammal 60(1):155–163
Ye F, Qi W, Xiao J (2011) Research of niching genetic algorithms for optimization in electromagnetics. Proced Eng 16:383–389
Zhan S, Huo H (2012) Improved PSO-based task scheduling algorithm in cloud computing. J Inf Comput Sci 9(13):3821–3829