Resource Management For Workflows In Cloud Computing Environment Based On Group Technology Approach

Saleh Shahdi-Pashaki
Department of Industrial Engineering
Shomal University
Amol, Iran
Miladshahdi@gmail.com

Ehsan Teymourian
Department of Industrial Engineering
Mazandaran University of Science and Technology
Babol, Iran
ehsan.teymorian@gmail.com

Vahid Kayvanfar
Department of Industrial Engineering
Amirkabir University of Technology
Tehran, Iran
v.kayvanfar@aut.ac.ir

Zahra Booyavi
Department of Industrial Engineering
University of Science and Culture
Tehran, Iran
sh.booyavi@gmail.com

Koorosh Mashhadi Alizaheh
Department of Industrial Engineering
Sadjad University of Technology
Mashhad, Iran
koorosh.farvahar@gmail.com

***Abstract*— Data transferring and load balancing are some of factors which make the resource allocation problem more challenging in cloud computing (CC) environment. So due to the lack of exact solution for this problem, this paper presents a new mathematical model based on group technology to allocate the virtual machines (VMs) to workflows with the aim of controlling the number of data transferring and server load variation simultaneously. Group technology is a resource allocation technique in industrial environments with proven ability to optimize some measures (e.g. part movements, resource utilization) also a genetic algorithm (GA) is designed to find near optimal solution for larger problems with no optimum solution.**

Keywords-Cloud computing; Group technology; Virtual machine; Workflow; Mathematical model; Genetic algorithm

I. INTRODUCTION

Today, advances in information technology have increased the demand for computing power greatly; users need to do their computational tasks without the need for expensive hardware and software at any time or place. Cloud computing (CC) is a new computational technology tries to deliver computing as a service similar to traditional public utility services (i.e. water, electricity, gas, and telephony). It is obtained from development of several technologies including: hardware, internet technologies, distributed computing and systems management. This model brings benefits to both consumers (people and companies that do not have their own infrastructure) and service providers (SPs) that encourage them to invest in this technology. Some of the mentioned benefits are listed below [1]:

i. Reduce capital costs;
ii. Reduce IT management efforts;
iii. Accelerate technology deployment;
iv. Accelerate business innovation;
v. Accelerate return on investment (ROI).

Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) are main services of CC that are available to users based on pay as you go model to execute their computational tasks [1, 2]. Fig. 1 indicates the layers of service in CC. SPs and users usually have different goals which the most important ones are user satisfaction in cloud quality of services (QoS) and increment of resource utilization rate (from the SPs perspective). The efficiency of the CC environment can be improved if the goals like these be considered in resource allocation strategies [3].

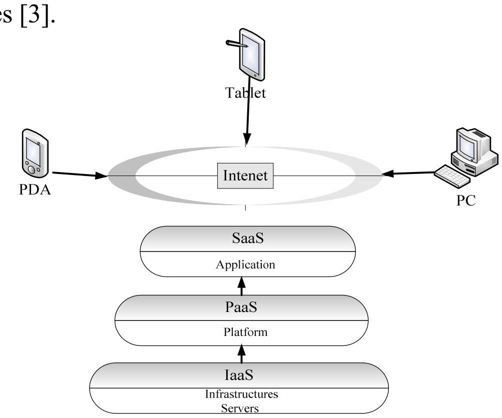

Figure 1. Service layers of cloud computing

There are many different kinds of computing tasks. Some of them are very simple (e.g. sending a comment on social networks) but the majority of tasks in cloud systems consist of multiple communicated subtasks (e.g. scientific workflows). An online hotel booking is illustrated as a workflow in Fig. 2.

The SPs may face different challenges when allocating computational resources to data-intensive workflow (e.g. montage workflow) due to the limited capacity of resources and servers communication constraints. The load balancing of system is one of the challenges which can directly impact on user satisfaction and resource efficiency. The load must be well distributed on servers during the workflow execution to decrease the data transfer time and dependency imbalance [4]. Data transfer is a bad factor which makes system slowdown and increase power consumption.

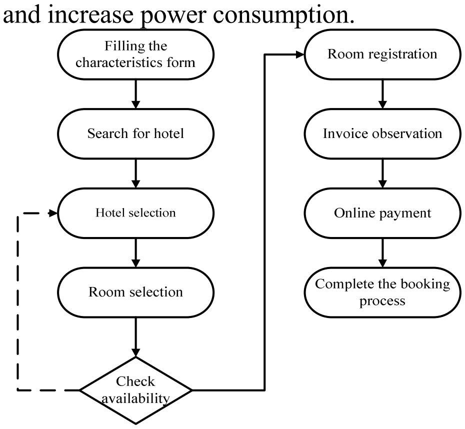

Figure 2. A sample of workflow (online hotel booking)

A solution for resource allocation problem can be ideal when the transferring of data between servers is minimal. In this case, the SPs must try, as far as possible, to allocate the subtasks of a workflow to a server to avoid unnecessary transfers. But this is not always possible in complex cloud environments; a server rejects the requests when it violates its capacity limit. So, some subtasks of a workflow must be transferred to other servers. It leads inevitably to increase the completion time of tasks. Along with these the number of virtual machines is an important problem that affects on system performance. A large number of virtual machines can cause interface in their performance and reduce the user's consent. Therefore they need to have a strategy to control all these simultaneously.

Group Technology (GT) is a manufacturing technique which prevents unnecessary movements (transfers) and increases the production efficiency in manufacturing systems. GT was first proposed by Mitrofanov [5], and then developed by Burbidge [6]. The main objective of GT is grouping the similar parts and machine to increase the flexibility. Cellular manufacturing system (CMS) is an application of GT for implantation of the modern manufacturing systems (e.g. flexible manufacturing system and just in time). Cell formation (CF) is a component of CMS that organizes independent cells by specifying the machine and part groups. The multiple proven abilities of GT to allocate the resources in CMS are as follow:

- Ability to reduce the total execution cost and time;
- Ability to providing a suitable management on VMs;
- Ability to improving resources utilization;

The items mentioned above as well as the structural similarities of cloud systems and CMS are the most important reasons to suggest GT for resource management in CC. The servers, virtual machines and tasks in CC have similar functions to cells, machine and component respectively. So in this paper, a new mathematical model based group technology is proposed to improve the efficiency of cloud by minimizing the transfer between the servers, number of created VMs and server load variation. Also a genetic algorithm is designed to solve the large scale problems and speed up the decision-making.

The remainder of the paper is structured as follows: a literature review on resource management in cloud computing and group technology is presented in Section 2. Problem modeling and proposed genetic algorithm so as to tackle the considering problem are given in Section 3 and 4, respectively. Section 5 is devoted to computational results, and finally conclusion and future remarks are summarized in section 6.

II. RELATED WORK

Many studies have been conducted in the context of resource allocation in CC with different polices. These studies have tried to offer various approaches. Some of them employ heuristics [7, 8, 9, 10, 11, 12] and others give met-heuristics approaches [13, 14].

Beside these, various researches have been done specifically in the context of workflow resource management. An approach presented by Varalakshmi *et al.* [15] to handle the workflows with the aim of meeting the user-preferred QoS parameters. Shi *et al.* [16] designed a resource allocation mechanism for handle market-oriented workflows. The purpose of their method is meeting the different consumers' demands. A multi-objective optimization approach presented by Szabo Shi *et al* [17] to allocate the scientific workflows with consideration of transfer and execution time. The results showed that their approach can outperform the current state-of-the art approaches. Chen and Deelman [18] Proposed an approach to allocate the resources to large-scale scientific workflows in order to reduce make span and resource cost of workflows. They showed that the heuristics and genetic algorithms is a proper candidate to solve this problem. A PSO based approach for workflow management proposed by Pandey *et al*. [19] to control the transfer cost and distribute the load on resources well. Likewise, we attempt to find an exact solution for workflow management in such a way that the number of transfers will be minimized along with fair distribution of loads. Moreover, our approach determines the optimal number of virtual machines in the system which has not been studied before.

Furthermore, various studies have been conducted on GT. They focused on proposing an efficient mathematical model as well as heuristic approaches. a multi-objective mathematical model to create machine and part groups and also single-objective mathematical model to minimize the intra-cell and inter-cell movements presented by Chan *et al.* [20]. Nair and Narendran [21] offered an algorithm to specify machine groups and part families relying on production sequence data. A genetic algorithm was proposed to solve a CFP considering several objectives

including movements cost, cell load variation and exceptional elements (EEs) by Zhao and Wu [22]. Mahdavi *et al.* [23] presented a unified mathematical model to solve the CFP and the cell layout problem (CLP) simultaneously minimizing backtracking against the forward movements. Tavakkoli *et al.* [24] presented an integer-linear programming to minimize the inter-cell movement and machine costs. They developed a simulated annealing (SA) for solving it. A two-phase solution approach has been studied by Mahdavi *et al.* [25]. They consider multifunctional machines and developed a mathematical model to group the machines with the aim of minimizing dissimilarity of machines in cells. Tavakkoli *et al.* [26] solved a dynamic cell formation problem whit GA, SA, and tabu search (TS) and stated that with the improvement of genetic operators the probability to achieve the optimal solution be increased.
According to the studies reviewed, the main objectives of GT are as follow:

- Mapping components to machines and machines to cells
- Minimizing number of inter/intra-cell moves
- Improving resource utilization through cell load balancing
- Increasing the flexibility of system

The above mentioned objectives can be interpreted in CC too, then considering them in a resource management framework could improve the performance of cloud environment. So, according to surveys conducted in the literature; the most contributions of our work are as follow:

- Introduce a new resource management approach (GT approach) for workflow management.
- Develop a mathematical model (exact solution) to control the number of transfers, the load of CC system and determine optimal number of VMs simultaneously for small problems.
- Design a genetic algorithm for larger problems with no optimum solutions.

III. Problem Modeling

In this research, a CC environment is considered with K servers and M virtual machine types which are available during the time period. Each server and machine type has a certain capacity including: CPU, storage capacity and memory which are known and constant during the time period. There are also j workflows with pre-defined demand, processing time and QoS (e.g. CPU, storage capacity and memory).The VMs and tasks must be distributed among K serves and M VM types regarding their capacity so that the transfers between servers and the load variance of servers are minimized. The optimal number of each VM type (N_{pk}) is obtained by model. A resource allocation problem in CC environment is presented in Fig.3.

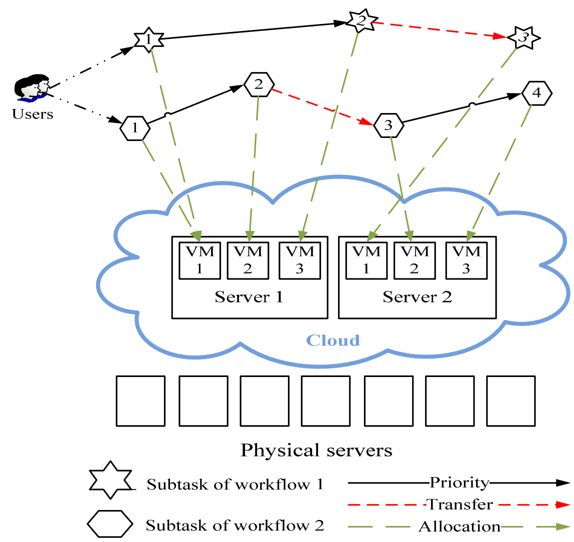

Figure 3. A simple case of resource allocation in cc

A. Notations.

The notations used in problem formulation are as follow:

B. Indices

j task, $j = 1,\dots, t$;

i subtask, $i = 1,\dots, s$;

p virtual machine type, $p = 1,\dots, m$;

k server, $k =1,\dots, c$;

C. Input parameters

W_{jpk} Workload on machine type p by task j in server k;

A_{jpk} Average intra-server processing time for task j in server k;

t_{ijp} processing time for subtask i of task j on machine p;

a_{ijp} 1 if subtask i of task j will require machine type p,0 otherwise;

L_p The maximum amount of QOS that machine type p can provide;

C_k The maximum amount of QOS that server k can provide;

T_p The total time which VM p is available;

D. Decision variables:

X_{ijpk} 1 if subtask i of task j is done on machine type p in server k,0 otherwise;

Y_{jpk} 1 if subtask i of task j assigned to server k,0 otherwise;

N_{pk} Number of machine type p assigned to server k;

$$W_{jpk} = \frac{\sum_{i}^{s} D_j . t_{ijp} . X_{ijpk}}{T_p} \quad (1)$$

$$A_{jpk} = \frac{\sum_{p}^{m} W_{jpk} . N_{pk}}{\sum_{p}^{m} N_{pk}} \quad (2)$$

E. Mathematical model formulation.

The objective functions of proposed model that given in (3) is composed of three criteria (F_1, F_2 and F_3) according to (4-6).

$$F = \alpha F_1 + \beta F_2 + \gamma F_3 \quad (3)$$

$$F_1 = \sum_{j}^{s} \sum_{p}^{m} \sum_{k}^{c} \left(W_{jpk} - A_{jpk} \right)^2 \quad (4)$$

$$F_2 = \sum_{i}^{s} \sum_{j}^{t} \sum_{p}^{m} \sum_{k}^{c} Y_{ijk} \left(1 - Y_{i+1jk} \right) \quad (5)$$

$$F_3 = \sum_{p}^{m} \sum_{k}^{c} N_{pk} \quad (6)$$

Which α, β and γ are the user-defined preferences. The servers load variation is given in (4), where W_{jpk} is obtained from (1) and implies workload on machine p by task j in server k. A_{jpk} is obtained from (2) and computes the average processing time for task j in server k. (5) indicates the number of tasks transferring between servers. Also, the number of virtual machine is identified in (6).

Now, the model formulation can be presented as following:

$$\text{Min}\, F = F_1 + F_2 + F_3 \quad (9)$$

s.t

$$X_{ijpk} = a_{ijp} Y_{ijk} \qquad \forall i,j,p,k \quad (10)$$

$$X_{ijpk} \leq N_{pk} \qquad \forall i,j,p,k \quad (11)$$

$$\sum_{p}^{m} \sum_{k}^{c} X_{ijpk} = 1 \qquad \forall i,j,p,k \quad (12)$$

$$\sum_{i}^{s} \sum_{j}^{t} D_j . t_{ijp} . X_{ijpk} \leq L_p . N_{pk} \qquad \forall i,j,p,k \quad (13)$$

$$\sum_{p}^{m} L_p . N_{pk} \leq C_k \qquad \forall p,k \quad (14)$$

$$Y_{ijk}, X_{ijpk} \in \{0,1\} \qquad \forall i,j,p,k \quad (15)$$

$$N_{pk} \geq 0, integer \qquad \forall p,k \quad (16)$$

The objective function in (7) is composed of three terms presented in (4 – 6) with identical preferences. Constraints (8) shows the relation between two variables (X_{ijpk} and Y_{jpk}) and state that task i of task j is performed on VM p in server k (X_{ijpk} =1) if (Y_{ijk} =1) and (a_{ijp} =1) simultaneously. Constraints (9) and (10) ensure that each subtask is assigned just to one VM in one server that required VMs are existent in it. Constraints (11) and (12) guarantees that virtual machine and server's capacity are not exceeded. Constraints (13) and (14) specify the decision variables.

IV. THE PROPOSED GENETIC ALGORITHM

After running the model with the Lingo 9 software it was found that the proposed model can solve problems up to 5 × 5 × 3 (tasks × VM type × servers) sizes. So it is necessary to use approximate solution methods for larger problems. Genetic algorithm (GA) is a heuristic search technique which proposed by Holland [27] for achieving optimal or near-optimal solutions in complicated problems. This algorithm uses from biological techniques like crossover, mutation and natural selection to create better solutions. GA was proposed in different fields such as designing of artificial intelligence programs (game-playing program) by Bagley [28] and simulation biology processes by Rosenberg [29]. Here, a genetic algorithm is offered for solving the studying model according to Tavakkoli *et al.* [26]. The factors intended for designing the GA are described in following.

A. Chromosome structure

Chromosome structure (Fig.4) which composed of three genes (i.e. *X, Y, and N*) is described below:

$$\left|\begin{array}{cccc|cccc|cccc} X_{11} & X_{12} & \cdots & X_{1j} & Y_{11} & Y_{12} & \cdots & Y_{1j} & N_{11} & N_{12} & \cdots & N_{1k} \\ X_{21} & X_{22} & \cdots & X_{2j} & Y_{21} & Y_{22} & \cdots & Y_{2j} & N_{21} & N_{22} & \cdots & N_{1k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{i1} & X_{i2} & \cdots & X_{ij} & Y_{i1} & Y_{i2} & \cdots & Y_{ij} & N_{p1} & N_{p2} & \cdots & N_{pk} \end{array}\right|$$

Figure 4. Chromosome

Two gens [X] and [Y] are i (subtask)-by-j (task) matrices related to assignment of subtask to VMs and servers and their elements are initialized from 1 to p and 1 to k respectively. Also matrix [N] is related to the number of VMs in each server. The elements of matrix [N] are positive integer values. Worth noting that matrix [X] is constant and obtained by a_{ijp}.

B. Generation of initial population

For generating an initial population, first matrix [N] initialized randomly considering the capacity constraints of servers. Then matrix [Y] is initialized randomly with regard to the possibility of task execution.

C. Fitness value

The fitness value is a criterion to measure the quality of a chromosome. In this paper it is calculated by (17) and (18).

$$Fitness\ function = \sum_{i} w_i f_i \quad (17)$$

$$f_i = \frac{F_i}{\sum_i F_i} \quad (18)$$

Where W_i is a user-defined weight, F_i is the value of objective function i and f_i is normalized value of F_i.

D. Selection strategies

A good selection strategy should be able to select the best chromosomes as for creating new generation. in this study the fitness of chromosomes in current generation are first normalized by (19) and then the chromosomes, which their normalized fitness are less or equal to zero, are selected as a mating pool.

$$Z_i^g = \frac{Fitness\left(X_i^g\right) - \mu_g}{\delta_g} \tag{19}$$

Where Z_i^g is normalized fitness of chromosome i and X_i^g is the fitness of g^{th} generation, also as μ_g and δ_g are mean and standard deviation of fitness values for generation g.

E. Operators

Due to the matrix structure of the chromosomes the operators should be adapted with them, so the modified operators are as follows:

Districted crossover- A block of parent's genes (*Y* or *N*) is selected and then all the selected elements are replaced together (in crossover) and changed (in mutation).

Simultaneous districted crossover- A block of parent's genes *Y* and *N* is selected simultaneously and their elements are replaced together (in crossover) or changed (in mutation). The crossover and mutation operators with probability of 0.7 and 0.4 are applied respectively.

F. Repairing

An infeasible chromosome will be created when the capacity of VMs and servers exceed the permitted limit so, genes should be changed so that surplus capacity be modified. The repair procedure is presented in Fig. 5.

Repairing

Servers repair
If the servers have been changed, **then**
While equipped capacity of machine is less than total capacity in all servers
Do Rearrange components based on their possibility of doing them on machines;
End
VMs repair
Else If the machine section have been changed, **then**
While equipped capacity is less total capacity of servers than zero
Do Rearrange the machines;
End
End

Figure 5. VM and Server Repairing

G. Stopping criteria

The process of GA stops when it reaches to maximum number of generations or current time exceeds the specific time. The pseudo-code of the proposed GA is given in Fig.5

Genetic algorithm

Initialize parameters *K*,*G*;
Generate *K* feasible chromosomes X_i^g, X_j^g, …, X_k^g for the first population;
Evaluate the solutions;
For $g \leftarrow 1$ to G**,, Do**
Select two chromosomes X_i^g, X_j^g as parents through the given selection strategy;
Procedure *crossover operator*;
Procedure *mutation operator*;
Procedure *repairing*;
Evaluate the solutions;
Replace the *new population* X_i^{g+1}, X_j^{g+1}, …, X_k^{g+1};
End

Figure 6. Pseudo-code of proposed GA

V. Computational Results

To prove the validity of our proposed approaches, we compare the mathematical model and GA with First-come, first-served (FCFS) algorithm. With FCFS algorithm, tasks are assigned the resources in the order they request them. several small and large examples are randomly designed with $t \times m \times k$ (tasks× VM type× servers) demotions. Details of VMs presented in Table I in accordance with Amazon EC2. The maximum number of tasks and servers for small and large problems are (t=5, m=5, k=3) and (t=20, m=15 k=5) respectively.

Table I VMs information from Amazon EC2

VM type	VM1	VM2	VM3
Memory (GB)	15	17.1	7
CPU (EC2 units)	8	6.5	20
Storage (GB)	1690	420	1690

Examples are solved as optimum (Mathematical model) with Lingo 9 software and near optimum (GA and FCFS) with MATLAB.R2010 programming language on a PC Pentium (R) Dual-core CPU with 2.50 GHz and 2GB RAM, using Win 7 as operating system. The results are presented in Table II.

Table II Computational results for small and large sized problems

Problem	Lingo	GA	FCFS	Gap (%)		
				Lingo And GA	***Lingo And FCFS***	***GA And FCFS***
Small-sized problems						
S.S.P _01	148.36	148.36	242.23	2.1	63.27	63.27
S.S.P _02	152.92	152.92	237.15	2.3	55.08	55.08
S.S.P _03	152.98	152.98	250.01	1.4	63.42	63.42
S.S.P _04	148.18	148.18	245.57	1.2	65.72	65.72
S.S.P _05	154.35	154.35	243.6	1.7	57.82	57.82
S.S.P _06	151.66	151.66	244.42	2.5	61.16	61.16

S.S.P _07	148.96	148.96	230	1.5	54.40	54.40
S.S.P _08	152.42	152.42	241.12	1.7	58.19	58.19
S.S.P _09	148.13	148.13	221.54	1.4	49.55	49.55
S.S.P _10	152.45	152.45	222.36	3.6	45.85	45.85
Avg.	151.04	151.04	237.8	1.94	57.45	57.45
large-sized problems						
L.S.P_01	907.02	895.28	1162	1.3	27.89	29.79
L.S.P_02	920.84	877.42	1157	4.7	27.99	31.86
L.S.P_03	924.30	902.16	1167	2.3	30.15	29.35
L.S.P_04	909.39	898.09	1204	1.2	33.05	34.06
L.S.P_05	910.04	890.48	1173	2.1	30.54	31.72
L.S.P_06	910.85	877.28	1197	3.6	31.08	36.44
L.S.P_07	921.22	897.69	1201	2.5	30.26	33.78
L.S.P_08	900.91	876.94	1159	2.6	32.19	32.16
L.S.P_09	906.53	879.96	1163	2.9	29.72	32.16
L.S.P_10	928.39	899.05	1213	3.1	28.93	32.62
Avg.	913.494	889.435	1179.6	2.63	30.18	32.39

Table II is related to the comparison of Lingo, GA and FCFS for small and large sized problems. Column 1presents the problem IDs (S.S.Ps and L.S.Ps).the optimum solutions obtained by Lingo are resented in column 2. It is worth mentioning that for large sized problems (which cannot be solved optimally in reasonable time) column 2 presents the objective bound of lingo obtained in 2 hours. Columns 3 and 4 present the solutions of GA and FCFS in 10 independent runs. Also the differences between GA and Lingo in different runtime are presented in columns 5-7. Fig. 7 and Fig.8 show the average normalized objective values which obtained by Lingo, GA and FCFS for the small and large sized problems. As it can be seen the Lingo and GA Have better performance compared to FCFS.

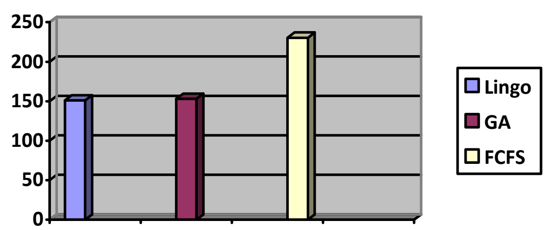

Figure 7: Performance analysis of GA and Lingo with FCFS small problems

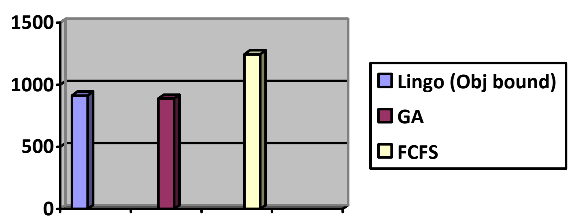

Figure 8: Performance analysis of GA and Lingo with FCFS for large problems

VI. CONCLUSION

Data transferring and load balancing are significant challenges which must be considered in workflow management. For this purpose, this paper presented a new mathematical model based on group technology to handle these Challenges. Group technology is a resource allocation technique in industrial environments. The objectives of model are controlling the data transferring and server load variation as well as determining the optimal number of virtual machines simultaneously. Also a genetic algorithm has been proposed to solve the large scale problems that no optimum solution exists for them.

REFERENCES

[1] R. Buyya, J. Brober, A. Goscinski, Cloud computing principles and paradigms. Hoboken, New Jersey, 2011.

[2] S. Selvarani, G.S. Sadhasivam, "Improved cost-based algorithm for task scheduling in cloud computing," 2010 IEEE International Conference on Computational Intelligence and Computing Research. Dec. 2010 ,pp. 1-5, doi: 10.1109/ICCIC.2010.5705847.

[3] Y. Fang, F. Wang, G. Junwei, "A task scheduling algorithm based on load balancing in cloud computing," Lecture Notes in Computer Science, vol.6318,Oct.2010, pp. 271-277. doi: 10.1007/978-3-642-16515-3_34.

[4] W. Chen, E. Deelman, R. Sakellariou, "Imbalance Optimization in Scientific Workflows," Proceedings of the 27th international ACM conference on International conference on supercomputing, June. 2013, pp 461- 462. doi: 10.1145/2464996.2467270.

[5] S.P. Mitrofanov, The scientific principles of group technology, National Lending Library Translation. Boston Spa, York's,1966.

[6] J.L. Burbidge, "Production flow analysis", Production Engineer, Vol.50, May.1971, pp.139-152, doi: 10.1049/tpe:19710022.

[7] D. Ergu, G. Kou, Y. Peng, S. Yong, S. Yu, "The analytic hierarchy process: task scheduling and resource allocation in cloud computing environment", The Journal of Supercomputing, Vol.64, June 2013, pp. 835-848.doi: 10.1007/s11227-011-0625-1.

[8] Xu, C. Zhao, E. Hu, B. Hu, "Job scheduling algorithm based on Berger model in cloud environment, Advances in Engineering Software," Vol.42 , July 2011,pp. 419-425, doi: http://dx.doi.org/10.1016/j.advengsoft.2011.03.007.

[9] G. Wei, A.V. Vasilakos, Y. Zheng, N. Xiong, "A game-theoretic method of fair resource allocation for cloud computing services," The Journal of Supercomputing, Vol.54, November 2010, pp. 252-269,doi: 10.1007/s11227-009-0318-1.

[10] L. Li, "An optimistic differentiated service job scheduling system for cloud computing service users and providers," Proc.Third International Conference.Multimedia and Ubiquitous Engineering (MUE '09), IEEE Press , June 2009, pp. 295-299, doi: 10.1109/MUE.2009.58 .

[11] K. Liu, H. Jin, J. Chen, X. Liu, D. Yuan, Y. Yang, "A compromised-time-cost scheduling algorithm in swin DeW-C for instance-intensive cost-constrained workflows on a cloud computing platform," International Journal of High Performance Computing Applications, Vol.24, May 2010, pp.445-456, doi: 10.1177/1094342010369114.

[12] Beloglazov, J. Abawajy, B. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing,", Future generation computer., Vol.28, May 2012, pp.755-768.doi: http://dx.doi.org/10.1016/j.future.2011.04.017.

[13] N. Netjinda, B. Sirinaovakul, T. Achalakul, "Cost optimization in cloud provisioning using particle swarm optimization," Proc. 2012 9th International Conference on. Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), May 2012, pp.1-4, doi: 10.1109/ECTICon.2012.6254298.

[14] Zhong, K. Tao, X. Zhang, "An approach to optimized resource scheduling Algorithm for Open-Source Cloud Systems," Proc. The Fifth Annual. China Grid Conference (China Grid), July 2010, pp.124-129.doi: 10.1109/ChinaGrid.2010.37.

[15] P. Varalakshmi, A. Ramaswamy, A. Balasubramanian, P. Vijaykumar. "An Optimal Workflow Based Scheduling and Resource Allocation in Cloud," First International Conference on Advances in Computing and Communications. July. 2011, pp 411-420. doi: 10.1007/978-3-642-22709-7_41.

[16] X. Shi, Y. Zhao, " Dynamic Resource Scheduling and Workflow Management in Cloud Computing," Web Information Systems Engineering – WISE 2010 Workshops. December, 2010, pp 440-448. doi: 10.1007/978-3-642-24396-7_35.

[17] C. Szabo, Q. Z. Sheng, T. Kroeger, Y. Zhang, J. Yu, "Science in the Cloud: Allocation and Execution of Data-Intensive Scientific Workflows," Journal of Grid Computing, October. 2013, doi: 10.1007/s10723-013-9282-3.

[18] W. Chen, E. Deelman, "Integration of Workflow Partitioning and Resource Provisioning," IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), May. 2012, pp 764 – 768, doi: 10.1109/CCGrid.2012.57.

[19] S. Pandey, L. Wu, G. S. Mayura, R. Buyya, , "A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments. 24th IEEE International Conference on Advanced Information Networking and Applications," April. 2010, pp 400-407, doi: 10.1109/AINA.2010.31.

[20] F.T.S. Chan, K.W. Lau, L.Y. Chan, V. H. Y. Lo, " Cell formation problem with consideration of both intracellular and intercellular movements," International Journal of Production Research, Vol.46, Mar 2008,pp. 2589-2620 doi: 10.1080/00207540500478843.

[21] G. J. Nair, T. Narendran, "CASE: a clustering algorithm for cell formation with sequence data," International Journal of Production Research, Vol.36, Nov 2010 , pp. 157-180, doi: 10.1080/002075498193985.

[22] C. Zhao, Z. Wu, "A genetic algorithm for manufacturing cell formation with multiple routes and multiple objectives," International Journal of Production Research, Vol. 38, Nov 2010, pp. 385-395, doi: 10.1080/002075400189473.

[23] Mahdavi, E. Teymourian, N. Tahami Baher, V. Kayvanfar, "An integrated model for solving cell formation and cell layout problem simultaneously considering new situations," Journal of Manufacturing Systems, Vol. 32, October 2013, pp.655-663,doi: http://dx.doi.org/10.1016/j.jmsy.2013.02.003

[24] R. Tavakkoli-Moghaddam, N. Safaei, F. Sassani, "A new solution for a dynamic cell formation problem with alternative routing and machine costs using simulated annealing," Journal of the Operational Research Society, Vol. 59, June 2007 pp. 443-454, doi:10.1057/palgrave.jors.2602436.

[25] Mahdavi, N. Tahami Baher, E. Teymourian, "A new cell formation problem with the consideration of multifunctional machines and in-route machines dissimilarity - A two-phase solution approach," Proc. IEEE 17Th International Conference. Industrial Engineering and Engineering Management (IE&EM), Oct 2010, pp. 475-479, doi: 10.1109/ICIEEM.2010.5646569.

[26] R. Tavakkoli-Moghaddam, M.B. Aryanezhad, N. Safaei, A. Azaron, "Solving a dynamic cell formation problem using metaheuristics,"Applied Mathematics and Computation 170 (2005) 761-780.doi: 10.1016/j.amc.2004.12.021.

[27] H.J. Holland, Adaptation in Natural and Artificial Systems. Ann Arbor, MI,1992.

[28] J.D. Bagley, The behavior of adaptive systems which employ genetic and correlation algorithms. Dissertation, University of Michigan.1976.

[29] R.S. Rosenberg, Simulation of genetic population with biochemical properties. Dissertation, University of Michigan. 1967.

Biographies

Saleh shahdi-pashaki received his B.S. and M.Sc degrees in Industrial Engineering from Mazandaran University of Science and Technology, Babol, Iran and Shomal University, Amol, Iran, respectively. His graduate thesis and research interest are about resource management and tasks scheduling in cloud computing.

Ehsan Teymourian received B.S. and M.Sc. degree from the Industrial Engineering department of Mazandaran University of Science and Technology (MUST). His research interests lie in cloud computing, logistic, production planning & scheduling, group technology, optimization, and operations research. He has published several journal and conference papers over the recent few years.

Vahid Kayvanfar is a Ph.D. candidate of Industrial Engineering at Amirkabir University of Technology (Tehran Polytechnic). His research interests mostly include supply chain management, production planning and scheduling in different areas, soft computing and applied operations research. He has published several journal and conference papers in the above-mentioned areas. He has also performed several practical projects in different organizations and taught courses such as operations research, project control and management, engineering statistics, inventory control, multiple criteria decision making (MCDM), etc.

Koorosh M.Alizadeh is studying Industrial engineering at Sadjad university of Technology in Mashhad, Iran. His research interest fields are around saving environment, HSE, operations research, customer relationship management (CRM) and philosophy of science and history. He was also worked as general network manager and manager assistant in department coffee shops of Ketab in where he live.