

¹ Mehdi Salehi babadi¹ Mohammad Ebrahim Shiri² Mohammad Reza Moazami Goudarzi³ Hamid Haj Seyyed Javadi

Multi-objective Scheduling of Tasks in Cloud Computing Using Fire Fly and Bee Colony Algorithms



Abstract: - The problem of allocating hardware and processing resources in cloud computing is considered in this research. Here, a virtual bee is assigned to any task that dynamically seeks the most appropriate computing resources. Each available computing source also has a virtual fire fly that its brightness is proportional to the amount of processing unit and the amount of computing residue that examines the computing source. The problem of distribution of computing resources among existing tasks will be investigated using the greedy technique of calculating the location of each task in general. The current method has many advantages compared to the random distribution method, but to complement the other topics, researchers can offer other tactics of distribution of computing resources simultaneously with the method proposed in this study and compare their performance with the implementation of both simulations. The simulation results show that the performance criteria of the cloud computing system such as time and computing resources have been greatly improved.

Keywords: Cloud Computing, Fir Fly Algorithm, Resource Allocation, Bee Colony Algorithm.

I. INTRODUCTION

Today, the applications of information technology have become very much and cover many areas. The IT specialist has a variety of tasks, from installing applications to designing complex computer networks and databases. With the dramatic increase in the variety of information technology equipment and services, the management of services provided in this area has also faced many challenges. management of investigating problems and requests, equipment and resource management in relation to technical support services and their allocation to users, as well as monitoring, control and scheduling in this field are among the issues that force IT managers to provide useful and efficient tools. These tools include information technology management software that can help managers, experts, and technicians in this regard (Ramachandra et al. 2017).

Sharing the "consumable and intangible" computing power between several tenants can improve productivity; because of this way, servers no longer remain idle for any reason (which significantly reduces costs while increasing the speed of production and development of applications). One side effect of this approach is that computers are used to a greater extent because cloud computing customers do not need to calculate and determine the maximum for their maximum load (Shishira et al. 2017).

The purpose of this study is to provide a method for allocating computing resources between input tasks in the cloud computer using a combination of bee colonies and fire fly algorithms to provide a new multi-objective meta-innovative method. Fire Fly Algorithm (Florence and Shanthi 2014) considers each existing option as a fire fly that the amount of brightness and attractiveness of it is proportional to the parameters in the problem.

The artificial bee algorithm was introduced in 2006 by Karaboga to optimize poly-nomial functions (Karaboga and Basturk 2008). In the bee algorithm, a food source indicates a possible solution to the problem. In this algorithm, the colony includes three types of bees: worker bees, scout bees, and outlooker bees. The task of the

¹Department of Computer Engineering, Borujerd Branch, Islamic Azad University, Borujerd, Iran, Email: msalehib@hotmail.com, shiri@aut.ac.ir

²Department of Mathematics, Borujerd Branch, Islamic Azad University, Borujerd, Iran, Email: mrmoazamig@gmail.com

³Department of Mathematics and Computer Science, Shahed University, Tehran, Iran, Email: h.s.javadi@shahed.ac.ir

*Corresponding author: Mohammad Ebrahim Shiri, Email: shiri@aut.ac.ir

worker bee is to search for finding new food sources. The outlooker bee is responsible for evaluating the solutions obtained by the worker bees. In each repetition of the algorithm, a number of worker bees that have found the best solutions are transformed into scout bees by outlooker bees. For scout bees, the recruitment process takes place, meaning that each scout bee is assigned a number of worker bees to search near the food source related to the scout bee. This process continues until the optimal answer is found.

II. A REVIEW OF PREVIOUS RESEARCH

So far, the division of functions and tasks in cloud computing has been much discussed (Zhan and Huo 2012), in Kaur and Verma (2012) a task scheduling algorithm in the cloud computing environment using a genetic algorithm designed called SGA. The optimization parameters in this article are the time and cost of performing the tasks. Unlike the standard genetic algorithm, the initial population in this algorithm is not randomly selected, and the results of the two SCFP and LCFP algorithms are used to determine the initial population.

In Kumar and Verma (2012), they used an improved genetic algorithm to schedule independent tasks in the cloud computing environment, which is a new version of the genetic algorithm for tasks scheduling in the cloud. The proposed algorithm is a combination of the Min-Min, Max-Min scheduling methods and the standard genetic algorithm (Braun et al. 2001). The obtained results indicate a reduction in completion time compared to the standard genetic algorithm.

In Braun et al. (2001), an optimization algorithm was provided for tasks scheduling in the cloud computing environment which consists of two parts. After applying the selection operators, recombination and mutation of the genetic algorithm, simulated refrigeration is used (Gan et al. 2010), in which case the new generation will be closer to the optimal answer. Also in this article, the service quality parameter consists of five parameters: completion time, bandwidth, cost, distance and reliability, which are assigned different weights to each of these parameters according to the type of task.

In Moschakis and Karatza (2015), the balance of load in cloud computing was examined using mathematics of Geodesic path finding which suggested a random climbing method for assigning tasks to virtual machines. The purpose of this method is to assign tasks to less loaded resources in order to improve response time to tasks. A table called the index table is also used to determine the current state of virtual machines.

In Mondal et al. (2012), they proposed an improved algorithm based on the Max-Min method for scheduling tasks in cloud computing called MMST. To solve the two main problems of the traditional Max-Min method (Atashpaz-Gargari and Lucas 2007), i.e long waiting times for smaller tasks and using more resources when tasks are low, a new parameter called maximum waiting time is used.

In Ge and Wei (2010), tasks scheduling based on the genetic algorithm was invented for the cloud computing system, which used the genetic algorithm (Vijindra and Shenai 2012) to minimize the time of completing tasks in the cloud computing environment. The matrix that the predicted time to perform each task is on each resource is used to calculate the completion time. Also, a parameter called ready [i] is used, which indicates when the computing i completes the previous task allocated in the previous scheduling, which will also affect the source load in finding the optimal sequence. This means that the work loading of resources will be directly affected at the time of completion.

In Xu et al. (2011), the tasks scheduling algorithm was designed based on the burger model in the cloud environment, which for the first time, the problem of scheduling in cloud with a model based on the distribution of social wealth called the burger model was investigated. In this method, two parameters of service quality called completion time and bandwidth are considered. For all tasks, a set of candidate virtual machines can be selected that can perform that task. For each candidate virtual machine, the Euclidean distance is measured by the current efficiency from the expected performance (the efficiency characteristics of each task are initially given by the user) and then the virtual machine with the lowest Euclidean distance performs that task.

In Chang et al. (2009) an ant algorithm for symmetrical tasks scheduling in the tour was presented. The goal is to allocate the optimal resource to each task according to the characteristics of the resource and the task. Each work is considered as an ant and the weight of the resources as a pheromone. In other words, the more the source weight is heavier, the pheromone is more. First, the pheromone matrix is formed for the tasks and resources, and then the

maximum pheromone is selected at each stage, and the corresponding task with that pheromone is assigned to its corresponding source. Of course, there are two types of local and global updates for pheromones. The results indicate the optimal amount of completion and balance time.

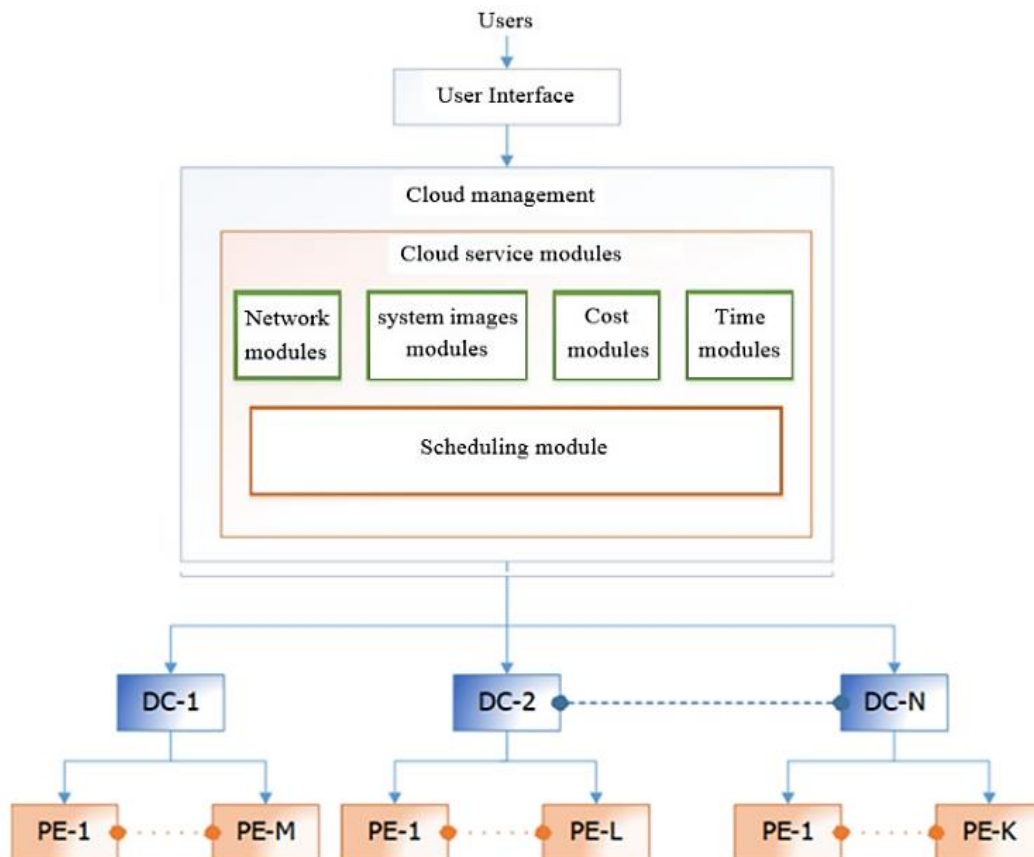
In Wang and Ai (2013), the task scheduling was described based on the ant colony optimization in the cloud computing environment, which used the ant colony algorithm to optimize the task scheduling in the cloud computing environment (Mouradian et al. 2018; Kalra and Singh 2015). This algorithm can bring the makepan close to the optimal value in the tasks sent to the cloud environment. The initial value of the pheromone is proportional to the ability of the virtual machines, and each ant first randomly assigns a virtual machine to each task.

In Wang et al. (2013) the resource scheduling model is designed based on particle swarm optimization algorithm, according to the characteristics of cloud computing resources, time constraints, cost and user needs.

According to the study, the division of computing tasks and resources, both in cloud systems and in multi-core computing, seems to have been analyzed using meta-heuristic algorithms. According to this issue, in the present study, the combination of two colonial algorithms of bee and fire fly has been used to solve this problem.

III. RESEARCH THEORY

Cloud computing is multi-part systems consisting of a network module, operating system image module, cost module, verification module and the like, and a cloud computing resource management system manages how to distribute the input tasks between these parts. Tasks entered by users are located in the DC's data center. Each data center divides the user's work into several "subsets" and makes them available to the computing resources (PE). The scheduling module will provide assigning the right task to the right source at the right time, within the cloud. In Figure 1, 'DC' represents a data center and 'PE' represents the CPU elements.



In this model, a new task is considered as a set of user tasks, the complex calculations of which are performed using cloud resources. Suppose UserJob= (U1, U2, ..., UN) is a set of user programs that enter (execute-demand) at a given time. Each UserJob (Ui) is shown with a double <ai, di>, which ai indicates the time of entry and di indicates the time period of the UserJob. If a task is not completed on time, it is considered an incomplete and failed task and must be re-entered into a new line of scheduling. In the scheduling process, the user's tasks are assigned to the available data centers (D1, D2, ..., DM) where $M \leq N$, the number of data centers may be less than the number of tasks required. Each data center (Di) is represented by a double <Ci, mi>. In this dual Ci, the cost of running tasks in the data center per unit of time and mi is the number of PEs available to perform user tasks. Each data center has a number of processing elements {PE1, PE2, ..., PEk to execute user task. Each computing element with PES characteristic means computing speed.

These tasks are located in one of the existing data synthesis called D, which has a certain number of PE computing units, and now these tasks must be distributed among these computing units. Each U has a required computing volume, P, and an allowable time, T. Each PE also has a PEj computing speed and a computing cost.

If the execution time of Tk in PEj is shown with tk, then the end time can be expressed as follows: this computing time is proportional to the difference in the computing volume required by Pk in PEj.

$$P_{k_{\text{remaining}}} = P_{K_{\text{initial}}} - \tau_k * PE_j \quad \leq \tau_k = \frac{P_k}{PE_j} \quad (1)$$

This means that if $P_k > PE_j$, then even though the Tk that has been used to the Uk task computing is reduced from the required computing value, some computing remains.

Naturally, we want this computing time to be less than the allowable time for the entire Usertask task, TiM. Assuming the parallel distribution of Uk tasks among PEj computing, the completion time of the U task is equal to the time of completion of the longest tasks.

$$\text{Makespan} = \max\{\text{Finish}(\tau_k)\} \quad (2)$$

If we consider the unit computing cost per PC to be equal to PCj, then the energy consumed to perform the entire U task is equal to:

$$E_U = \sum_{k=1}^n ((\tau_k * PC_j)) \quad (3)$$

That is equal to the cost spent for UserJob's task on the D-data synthesis.

$$E_U = \sum_{k=1}^n (\tau_k * PC_j) \quad (4)$$

Here, a new relationship must be established to determine the productivity of the task distribution algorithm. This productivity can be obtained by comparing the amount of computing that D could do with U at the total time with the amount of computing that has actually done. That's mean:

$$\text{eff} = \frac{\sum_{k=1}^n (\tau_k * PE_{jk})}{\text{Makespan} * \sum_{j=1}^J PE_j}^{-1} \quad (5)$$

This equation is equal to the adverse sum of the time that each computing unit has been dealing with a particular task divided by total time of task U multiplied by the total computing power. Therefore, the more the total task time is less and the computing power is distributed among the tasks more optimal, the efficiency of the algorithm increase.

The SLR parameter is also obtained from this equation:

$$\text{SLR} = \frac{\max\{\text{Finish}(\tau_k)\}}{\min\{\text{Finish}(\tau_k)\}} \quad (6)$$

Now suppose that each task U corresponds to the subtasks of T and has a bee with greedy choice. In other words, in each m period of the performance of the computing resource allocation algorithm, we have a number of residual Pk from different Uk, each of which has a bee with greedy choice. To continue, we must define a selection power for each of the available PKs, which are the capacity of the bee allocated to it, and we consider it equal to the amount of computing left from the Uk task. In short, each remaining or new task in each period, according to its

chosen power, selects the strongest of the available computing resources, and this process will continue until all tasks are completed.

Fire Fly Algorithm is an algorithm based on the population of the members of the set. In the fire fly community, communication is done through light production patterns, studies show (Yang 2009) that three laws are governed on this communication method:

- A. More light means more gravity, and the lighter member will move toward the brighter member.
- B. The amount of gravity also has an inverse relationship with the distance between the two fire flies.
- C- Members who are not attracted to other members will have a random movement around to be exposed to each other's light again.

To use this algorithm, the amount of light intensity can be considered equal to the value of the target function at this point. The attraction between the two members can be modeled by the following formula: If d is the Euclidean distance, α is a random coefficient, B_0 is the maximum gravity and 2 is the adsorption coefficient, for the gravity between the two members:

$$B = B_0 e^{\gamma d^2} + \alpha \quad (7)$$

According to the theory presented in the reference (Gholizadeh 2014), the obtained path using this law of absorption will always be a member of the Pareto front and will be close to the best solution in the acceptable range. Each cloud computing can be considered as a fire fly, and outlooker bees representing input tasks at each stage of the algorithm repetition must choose a path between all other fire flies (computing) so that all computing and the tasks in them are checked and the computing that best fits the amount of computing left over from the current task and the previous task in which it is less fit to be selected. In equation 6, the value of absorption coefficient and random coefficient are equal to 1, and B_0 is equal to the computing power of the cloud computing or PE_j , and d is the difference between the computing power of the computing, i.e, PE_j and the remaining computing value of T_k task is the equation (1). If we show the periods of assigning tasks with m , the selection power of each bee for each cloud computing is equal to:

$$POW_k = PE_j e^{(P_{k_{initial}} - \sum_{m=1}^M \tau_k(m) * PE_j(m))} \quad (8)$$

That is, the selection power of each bee for its own task is proportional to the initial amount of computing required for that task minus the computing power allocated to that task at different times. However, each repetition of the resource allocation algorithm that has the highest U_k selection power allocates the highest PE_j . However, the energy consumption ratio is also rewritten as follows:

$$E_U = \sum_{m=1}^M \sum_{k=1}^n (\tau_{km} * PC_{jm}) \quad (9)$$

And the efficiency of the algorithm is equal to:

$$eff = \frac{\sum_{m=1}^M \sum_{k=1}^n (\tau_k * PE_{jm})}{Makespan * \sum_{j=1}^J PE_j}^{-1} \quad (10)$$

Input passive include the number of tasks available and the computing volume of each, the number of computing and computing power, and the computing cost of each and the output of the problem of the total time of computing, total energy consumed and the efficiency of the algorithm that compared in both cases of distribution with fire fly and bee algorithm with the greedy choice and the constant random distribution. In a fixed random distribution at the beginning of the task, each of the tasks is randomly assigned a computing source and this allocation no longer changes.

A remained point is that according to the wide range of servers and computing resources in cloud systems, the probability that the number of tasks assigned is greater than the number of computing is very small, which is usually $J \geq K$.

IV. SIMULATION AND RESULTS

After introducing the generalities of the subject, examining the technical terms and reviewing similar works, we were able to present a new theory for the distribution of cloud computing resources among the existing tasks. It's time to look at the results and judge how accurate they are and how effective the technique is.

We perform the simulation by considering a 12-subtask computing on a cloud computing with 20 hardware resources. Naturally, a virtual machine is defined simultaneously to perform each task, but how having each virtual machine from hardware resources is set by the algorithm provided.

The first step is to determine the number of hardware resources with different computing power. The reason for the random selection is that due to the use of cloud servers from discrete computing resources in a very large area of the Internet and the disconnection and connection of many of these resources, the existing computing power cannot always be assured.

The required computing values in each sub-task are also unknown and are randomly assigned to a reasonable range.

Also, the cost of computing does not only depend on the computing power and also depends on other parameters such as the distance and the position of the computing, etc., so this issue must be determined randomly.

After determining the tasks and hardware resources available for each cloud processor, a characteristic of fire fly is selected and a bee is determined for each task and their performance is such that the location of each sub-task may depend on the selection power of the relevant bee that it may change during program execution that the next table shows this issue (**Table. 2**).

Performing the whole task with the current method to divide the computing resources and tasks in this example took 48 hours, i.e, the 12 parallel automation was changed 48 times.

As we have said, the computing of all tasks by existing cloud resources is once introduced by the method and once by the allocation of random computing resources. Now let's compare the output parameters between these two modes (**Table 1**).

Table 1. System performance parameters in the first simulation.

SLR	Efficiency of computing	Values of energy consumption	Computing time	Parameter of system	Sources management method
45	20	450 Mega Jouls	8 periods	Proposed method	
220	18	3.6 Giga Jouls	220 periods	random distribution of sources	

As it can be seen, the use of the proposed algorithm in this study has reduced the computing time by more than 4 times, the use of the proposed method compared to the random allocation method has reduced the energy consumption by 8 times, improving all operating system parameters in the use of fire fly algorithm with the greedy bee selector is very impressive and clear compared to the random allocation of resources. It seems that the proposed method can have a significant impact on improving cloud computing services.

In order to evaluate the performance of the algorithm in different conditions, we repeat the simulation once again in the 45 sub-task mode with a maximum load of 4000 computing units and distributed among 65 hardware resources with a computing capacity of 1000 units and computing cost of 10,000 repeating units. The results are as follows (**Table 2**).

Table 2. System performance parameters in the second example.

SLR	Efficiency of computing	Values of energy consumption	Computing time	Parameter of system	Sources management method
6	25	6 Mega Jouls	8 periods	Proposed method	
1350	40	18 Giga Jouls	1400 periods	random distribution of sources	

It seems that in the case of random distribution, some computing sources have been working much longer than other computing sources. This issue has also been effective in the following diagram.

The remained point is the time allowed for computing that due to the speed limit and the number computing, this time limit cannot be considered, while from the users' point of view, prolonging the time of performing tasks is undesirable, but at the same time, the user cannot expect that any task that is expected from the cloud system must be performed in the time set by the user, in any case, the allowable computing time is an independent parameter and it can be assumed to be any value and it cannot be expected from the computing distribution system to performed a very large task using small computing in a short time.

In this project, the problem of distribution of computing resources between existing tasks was investigated using the greedy technique of computing location of each task in general. Here are some suggestions for next studies:

- Applying input tasks in specific modes, it can consider infinitely different modes for the sequence of computing tasks and the performance of the system in this mode can be examined.
- Considering two parallel Usertask to make the computing busy that are idle before completing all under-tasks.
- Comparison of other techniques other than random distribution of tasks to be compared with this method.
- Considering unforeseen factors such as unavailability of some computing resources.

The current method has many advantages compared to the random distribution method, but to complement the other topics, researchers can offer other tactics of distribution of computing resources simultaneously with the method proposed in this study and compare their performance with the implementation of both simulations.

V. Conflict of interest

The authors declare that they have no conflict of interest.




REFERENCES

- [1] Atashpaz-Gargari E, Lucas C (2007) Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition. In: 2007 IEEE Congress on Evolutionary Computation, pp 4661–4667
- [2] Braun TD, Jay Siegel H, Bölöni L, et al (2001) A Comparison of Eleven Static Heuristics for Mapping a Class of Independent Tasks onto Heterogeneous Distributed Computing Systems. *J Parallel Distrib Comput* 61(6):810–837
- [3] Chang RS, Chang JS, Lin PS (2009) An ant algorithm for balanced job scheduling in grids. *Futur Gener Comput Syst* 25(1):20–27
- [4] Florence P, Shanthi V (2014) A load balancing model using firefly algorithm in cloud computing. *J Comput Sci* 10:1156–1165. <https://doi.org/10.3844/jcssp.2014.1156.1165>
- [5] Gan G, Huang T, Gao S (2010) Genetic simulated annealing algorithm for task scheduling based on cloud computing environment. In: 2010 International Conference on Intelligent Computing and Integrated Systems, pp 60–63
- [6] Ge Y, Wei G (2010) GA-Based Task Scheduler for the Cloud Computing Systems. In: 2010 International Conference on Web Information Systems and Mining, pp 181–186
- [7] Gholizadeh S (2014) Optimal Design Of Truss Structures By Improved Multi-Objective Firefly And Bat Algorithms. *Int J Optim Civil Eng* 4(3):415–431
- [8] Kalra M, Singh S (2015) A review of metaheuristic scheduling techniques in cloud computing. *Egypt Inform J* 16(3):275–295
- [9] Karaboga D, Basturk B (2008) On the performance of artificial bee colony (ABC) algorithm. *Appl Soft Comput* 8(1):687–697
- [10] Kaur S, Verma A (2012) An Efficient Approach to Genetic Algorithm for Task Scheduling in Cloud Computing Environment. *Int J Inf Technol Comput Sci* 4(10):74–79
- [11] Kumar P, Verma A (2012) Scheduling using improved genetic algorithm in cloud computing for independent tasks. In: Proceedings of the International Conference on Advances in Computing, Communications and Informatics - ICACCI '12, p 137
- [12] Mondal B, Dasgupta K, Dutta P (2012) Load Balancing in Cloud Computing using Stochastic Hill Climbing-A Soft Computing Approach. *Procedia Technol* 4:783–789
- [13] Moschakis IA, Karatza HD (2015) Multi-criteria scheduling of Bag-of-Tasks applications on heterogeneous interlinked clouds with simulated annealing. *J Syst Softw* 101:1–14
- [14] Mouradian C, Naboulsi D, Yangui S, et al (2018) A Comprehensive Survey on Fog Computing: State-of-the-Art and Research Challenges. *IEEE Commun Surv Tutor* 20(1):416–464




- [15] Ramachandra G, Iftikhar M, Khan FA (2017) A Comprehensive Survey on Security in Cloud Computing. *Procedia Comput Sci* 110:465–472
- [16] Shishira SR, Kandasamy A, Chandrasekaran K (2017) Workload scheduling in cloud: A comprehensive survey and future research directions. In: 2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence, pp 269–275
- [17] Vijindra S, Shenai S (2012) Survey on Scheduling Issues in Cloud Computing. *Procedia Eng* 38:2881–2888
- [18] Wang L, Ai, L (2013) Task scheduling policy based on ant colony optimization in cloud computing environment. In: LISS 2012 - Proceedings of 2nd International Conference on Logistics, Informatics and Service Science, pp 953–957
- [19] Wang Y, Wang J, Wang C, Song X (2013) Research on Resource Scheduling of Cloud Based on Improved Particle Swarm Optimization Algorithm. Springer, Berlin, Heidelberg, pp 118–125
- [20] Xu B, Zhao C, Hu E, Hu B (2011) Job scheduling algorithm based on Berger model in cloud environment. *Adv Eng Softw* 42(7):419–425
- [21] Yang XS (2009) Firefly algorithms for multimodal optimization, In: Stochastic Algorithms: Foundations and Applications (Eds O. Watanabe and T. Zeugmann), SAGA 2009, Lecture Notes in Computer Science, 5792, Springer-Verlag, Berlin, pp 169-178
- [22] Zhan S, Huo H (2012) Improved PSO-based task scheduling algorithm in cloud computing. *J Inf Comput Sci* 9:3821–3829

BIOGRAPHIES OF AUTHORS






Mehdi Salehi Babadi    received his M.Sc. degree in software Engineering from Department of Computer Engineering, Borujerd Branch, Islamic Azad University, Borujerd, Iran, in 2012. Currently, He is a Ph.D. student in the Department of Computer Engineering at Borujerd Branch, Islamic Azad University, Borujerd, Iran since 2015. His Ph.D. advisor is Professor Mohammad Ebrahim Shiri. His research interests include social network, Cellular automata, IoT, cloud computing, cryptography and security. He can be contacted at email: msalehib@hotmail.com.






Mohammad Ebrahim Shiri    received his Ph.D. degree in Computer Science from the Department of Computer Science, university of Montreal, Montreal, Canada in 2000. Currently, he is an Assistant Professor in the Department of Computer Science at Amirkabir University of Technology in Tehran, Iran. His research interests include Machine learning, E-learning, Database, Network security and Cloud Computing. He can be contacted email: shiri@aut.ac.ir.



Mohammad Reza Moazami Goudarzi    received his Ph.D. degree in the Department of Mathematics, Islamic Azad University, Tehran, Iran in 2011. Currently, he is an assistant professor in the Department of Mathematics at Borujerd Branch, Islamic Azad University, Borujerd, Iran. His research areas include operation research and data envelopment analysis. He can be contacted at email: mrmoazamig@gmail.com.



Hamid Haj Seyyed Javadi    received the M.Sc. and Ph.D. degrees in Amirkabir University of Technology, Tehran, Iran in 1996 and 2003 respectively. He has been working as a fulltime faculty member and Professor in the Department of Mathematics and Computer Science at Shahed University, Tehran, Iran. His research interests are IoT, computer algebra, cryptography and security. He can be contacted at email: h.s.javadi@shahed.ac.ir.