**Bulletin of Electrical Engineering and Informatics**
Vol. 10, No. 4, August 2021, pp. 2320~2326
ISSN: 2302-9285, DOI: 10.11591/eei.v10i4.2947 

# Novel load balancing approach based on ant colony optimization technique in cloud computing

**Yasameen A. Ghani Alyouzbaki, Muaayed F. Al-Rawi**
Computer Engineering Department, College of Engineering, Mustansiriyah University, Baghdad, Iraq

**Article Info**

***Article history:***

Received Mar 5, 2021
Revised May 26, 2021
Accepted Jun 17, 2021

***Keywords:***

Cloud computing
CloudSim
Load balancing
Virtual machine

**ABSTRACT**

The cloud is the framework in which communication is connected with virtual machines, data centers, hosts, and brokers. The broker searches for a highly reliable cloudlet virtual machine for execution. Vulnerability can occur in the network because of which framework gets overburden. A research strategy is introduced in this article to expand the fault tolerance of the framework. The proposed approach improvement depends on the algorithm of ant colony optimization (ACO) that can choose the better virtual machine on which is to migrate the cloudlet to reduce the execution time and energy consumption. The efficiency of the proposed approach simulated in terms of execution time, energy consumption and examined with CloudSim. The introduction is provided in this article with a detailed description of cloud computing and, in addition, green cloud computing with its models. This article also discussed the virtual machine (VM) in more depth in the introduction section, which allows cloud service providers to supervise cloud resources competently while dispensing with the need for human oversight. Then the article submitted and explained the related works with their discussion and then it explained the novel proposed load balancing based on ACO technique and concluded that the execution time and energy consumption of the proposed technique is better than the three-threshold energy saving algorithm (TESA) technique that is commonly used in cloud load balancing.



***Corresponding Author:***

Muaayed F. Al-Rawi
Department of Computer Engineering
Mustansiriyah University
Palestine Street; PO Box: 14022, Baghdad, Iraq
Email: muaayed@uomustansiriyah.edu.iq

## 1. INTRODUCTION

Cloud computing is a computing model. In this the frameworks are huge in numbers that are linked in public and private networks. The purpose for utilizing cloud computing is to give a foundation to applications that should be dynamically scalable that has been utilized for saving information and files. The creation of cloud computing has decreases to be too expensive degree alongside it decreases the time needed for application hosting, delivery, and content storage [1]-[3]. When all is said in done three kinds of services are offer by cloud providers for example infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). The organization needs to pay the cloud provider based on utilization of resources on computation premise every one of these reasons power an organization to pull in towards cloud computing. To make guarantee that organization will consistently be accessible to clients when needed can be meeting utilizing the services of cloud computing [4]-[6]. Cloud computing is a profoundly scalable and practical foundation for running number of applications, for example, Web applications, and ventures. Nonetheless, there is one major basic issue in cloud computing which have been arising because of its

developing interest which has definitely expanded the utilization of energy in data centers. The issue of intense usage not just builds the activity cost which decreases the benefit of cloud providers however it additionally influences the climate as the intense usage of energy prompts high outflow of carbon. Henceforth, energy-efficient solutions are needed to limit the effect of cloud computing on the climate [7], [8]. The target of making cloud which is climate neighborly can be accomplished by the utilization of green cloud computing. There are number of issues in green distributed computing. Out of the multitude of issues the virtual machine (VM) is the most imperative to be considered. By expanding the cloud resource use level with the utilization of virtualization innovation cloud activity cost get decrease to much degree [9], [10]. Yet, in the event that the utilization of virtualization isn't done appropriately in cloud data centers than the performance of cloud can be debase an excess of degree. VM migration is a strategy that helps cloud service providers to proficiently oversee cloud resources while dispensing with the need of human oversight. VM migration strategy move current-hosted outstanding burden starting with one server then onto the next by either utilizing live or non-live migration design [11], [12].

In comparison with non-live migration, live migration doesn't suspend application services preceding VM migration process. VM migration empowers cloud operators to accomplish different resource management objectives, for example, real time server, fault management, and green computing. There is a linked among the power being devoured by data centers and the handling power components. The quantity of actual hosts' accessible on-position likewise assumes a significant job in computing this power exhaustion. Subsequently, the power being consumed by cloud data centers is decreased by including dynamic VMs. It is imperative to compute the quantity of working hours on work days and the quantity of requests produced for the cloud data center. The quantity of VMs that are to be assigned and made to run over the actual machines and hosts is additionally essential to compute [13], [14]. There is no necessity of big number of VMs to be designated on the actual machines by the data center on days when traffic is less. Subsequently, the quantity of active hosts can be decreased effectively through mobility and integration of VMs inside the data center of the frameworks. There are two assortments of threshold needed to control the allocation of VMs to different hosts [15], [16]. The over-burden on actual servers and sending the newly created VMs to different hosts is managed by one threshold known as the upper threshold. Moreover, the hosts that use low-level and are the sources of squandering resources are made do with the assistance of lower threshold. There is a slight diverse among the green dynamic VM allocators and others. The hosts are controlled in outstanding burden that is shut to the upper threshold of the hosts inside the frameworks that are used inside these green dynamic VM allocators [17]-[19]. The different in center and low working servers are solidified with the assistance of this activity and rests of the servers are closed down. This entire process is alluded to as green cloud VM the management process.

## 2. STATEMENT OF PROBLEM AND METHOD Of SOLUTION

Despite the benefits, cloud users face a number of challenges. Security and resource ownership, shared resources and software licenses, availability/uptime, regulation/compliance, and product/service availability are only a few of them. In addition to the challenges already listed, there is one more. The energy issue has been described as a big issue with the cloud data centers' consumption, moreover the time execution in the cloud, therefore this article done a novel approach that depends on ACO technique to reduce cloud energy consumption and execution time comparing with TESA technique that is commonly used in cloud load balancing and that will be shown in section 4 and 5.

## 3. RELATED WORKS WITH DISCUSSION

Rao and Babu [20] has researched the issue of power decreasing for the client terminals by application offloading in multi-cell multi-client orthogonal frequency-division multiple access (OFDMA) mobile cloud computing networks where several practical bonds, for example, interference level on each channel, capacity limitation, interference level on each channel, and maximum tolerable delay for each client. By the utilization of D.C. estimate mixed integer nonlinear issue is besides changed over into a convex structure. Simulation results of this novel proposed approach shows that by using the J-PAD approach, in comparison with baselines, extensive power saving could be accomplished for example about 30% for delays in excess of 100 ms. Still there is require to enhance power saving and require to decrease the delay even to more degree.

Goudarzi [21], the authors have suggested that cloud processing is an extraordinary platform utilizing which corporate world is thoroughly hoping to diminish their consumption on resources like programming, platform tools, and framework. Cloud computing provides a cost effective foundation with large scalability and performance. The cloud service providers give the types of services through a huge data center. These data centers are kept up as numerous clusters and each cluster is a gathering of a few actual

machines. These actual machines will be virtualized to make virtual machines and the customers access the services through these virtual machines. There will be increment in utilization of power and carbon emission by increment demand on the data center.

Arianyan [22], the authors have suggested managing contemplating different methods, models, and algorithms, for efficient green cloud computing by utilizing virtualization procedures. There are various strategies which are identified with power saving which can likewise help in the efficiency of the frameworks based on sever and network included. All such procedures are to be concentrated here to introduce an investigation on the existing techniques in this article. The network devices, for example central processing unit (CPU), servers, and switches are the ones which consume the most elevated power. To plan modern approaches there is still research being conveyed. New procedures with enhance energy efficiency are being developed which likewise incorporate the quality of service (QoS), service level agreement (SLA) and VM combination in these frameworks. They didn't chip away at the proportion of computation and power which assist in using the resources in better path alongside least consumption of energy.

Gu *et al.* [23] has proposed solidification as a new method for energy saving in cloud data centers. One of the significant downsides of current investigations on solidification solution is that they concentrate just around one criterion and disregard different ones. In light of altered insightful analytic hierarchy process (AHP) procedure this study proposed a new multi goal combination solution. The three goals have been considered in this article, for example, energy consumption, SLA violation, and number of migrations in decision cycle. The comparisons are made among different methodologies and their outcomes are assessed as far as simulation parameters. There is decreasing in the energy consumption inside the outcomes accomplished through proposed strategy. By actualizing the proposed strategy in real cloud foundation, the management items and the experiments are led in this research.

Zhou *et al.* [24] has suggested that for research use, virtual machine combination is the best solution found. At the moment of power consumption for each VM is known, more power can be stored here. There are various modeling strategies proposed here to compute the power consumption as it isn't effortlessly determined in direct way. A tree regression-based technique is proposed in this article, which assists in computing the power being consumed by the VMs on same hosts of the frameworks. The dataset will be apportioned according to the benefits of this strategy. Here, each dataset is a simple modeling subset for the other. In different applications that run on VMs, the accuracy accomplished by applying this proposed technique is around 98% according to the trial results. The accuracy of individual VMs is anyway not computed in this article.

Islam *et al.* [25], the authors suggested a novel technique to increase the energy performance of large-scale data centers, the three-threshold energy saving algorithm (TESA) virtual machine deployment algorithm, which relies on the linear relationship between energy consumption and the use of processor resources. Hosts in data centers are divided into four groups in TESA, as indicated by load, i.e. host with light load, host with proper load, host with middle load and host with heavy load. VMs on a lightly loaded host or VMs on a highly loaded host are migrated to another host with the proper load by characterizing TESA. This article will be compared with this present work because it's commonly used in cloud load balancing.

## 4. RESEARCH METHOD AND PROPOSED METHOD

Green cloud computing is the energy-efficient cloud computing methodology for storing and processing information from clouds. The fault event is a big green cloud computing issue that reduces its performance. For the assignment and execution of the cloud, meta-heuristic methodology is used in this article. The ACO technique for task execution and task assignment will be used in this exploration. The ACO technique is performed to move the task from one virtual machine to another if machine over-burdening occurs. In the three stages depicted below, the ACO technique operates:

- Defining initial population: the initial population is provided as an input in the first stage, which is the execution time and failure rate of each machine. The initial population is used for choosing the most efficient cloudlet execution machine.
- Updating the pheromone: the probability of failure of each machine is determined on the basis of the initial population in the second stage of the method. In (1) was used to measure the probability [26].

$$\rho_{i,j} = \frac{(\tau_{i,j}^{\alpha})(\eta_{i,j}^{\beta})}{\Sigma(\tau_{i,j}^{\alpha})(\eta_{i,j}^{\beta})} \quad (1)$$

where $\tau_{i,j}$ is the total pheromone deposited on path $i$, $j$, $\eta_{i,j}$ is the heuristic value of path $i$, $j$, $\alpha$ and $\beta$ are parameters that control the relative importance of the pheromone trail versus a heuristic value.

– Choosing of the best pheromone: The machine with the minimum probability of failure is chose as the best machine for task migration in the last phase. The (2) and (3) below were used to choose the best machine [26].

$$\tau_{i,j} = (1-\rho)\tau_{i,j} + \Delta\tau_{i,j} \quad (2)$$

$$\Delta\tau_{i,j}^{k} = \begin{cases} \frac{1}{L_k} \; if \; ant \; is \; the \; chances \; of \; failure \\ 0 \; otherwise \end{cases} \quad (3)$$

Where $L_k$ is the objective function taken by the *k-th* iteration. The flow chart of proposed load balancing approach is shown in Figure 1.

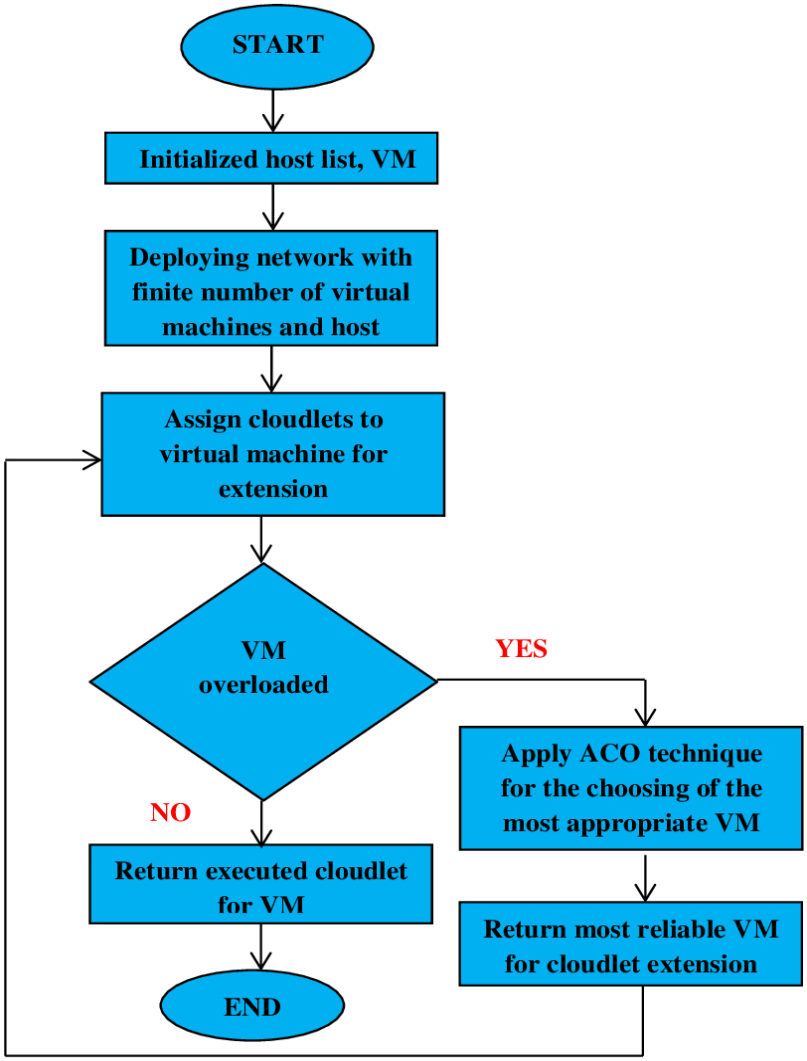


Figure 1. Flow chart of proposed load balancing approach

## 5. RESULTS AND DISCUSSION

A progression has been made of CloudSim software based PC production tests are done on the proposed and TESA techniques because the TESA technique is commoly used in cloud load balancing. Figure 2 and Figure 3 revealed that the empirical results for execution time and energy consumption are better for the proposed technique than the TESA technique. To calculate the execution time of the cloudlet (task or job) in each virtual machine, VM (CPU), the proposed technique assumed that only have one

cloudlet per VM (cloudlet.length/total million instructions per second (MIPS) of VM) and then create a numbers of cloudlets (tasks) equal to 10 increament till 80 by CloudSim 3.0.3 under Windows 7 with execuation time from 0 to 200 second and give the same task length to all tasks. Then, after running the code, the calculation is done by (4) [27] for each cloudlet (cl) for both TESA and proposed techniques to estimate the execution time that has been shown in Figure 2.

$$\text{ExecutionTime}=\text{cl.getFinishTime ()} - \text{cl. getExecStartTime ()} \quad (4)$$

Power consumption is directly related to the CPU utilization of VM. By using CloudSim 3.0.3, energy consumption depends only on host's CPU load. In (5) shows how CloudSim gets the host power consumption P (u), depending on its CPU utilization u [27]:

$$P(u) = k.P_{max} + (1-k).P_{max}.u \quad (5)$$

The maximum power of the host computer is indicated by $P_{max}$ (i.e. when CPU utilization is equal to 100%). The ratio of maximum to idle power (i.e. when CPU is not used) is denoted by k., here k=0.7 [27]. The host CPU utilization is represented by the input parameter of the power function P in (5), (noted u). The workload on the VM (CPU) and the performance of the host CPU evaluate this consumption. The CPU workload of a provided VM is the number of all Cloudlets CPU workloads listed in MI (Million Instructions). The only CPU efficiency metric used by CloudSim is million instructions per second (MIPS). Energy is defined as the product of power and time. The CloudSim 3.0.3 solves the energy consumption by using the linear interpolation of (5) [27]. Figure 3 revealed the energy consumption for both techniques that estimated by (5) and simulated it by CloudSim 3.0.3 code for each cloudlet (VM or CPU).

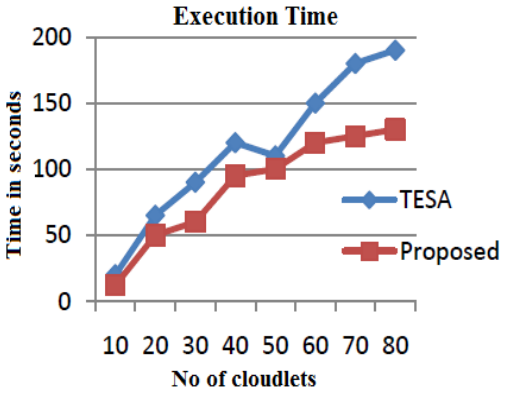


Figure 2. Execution time

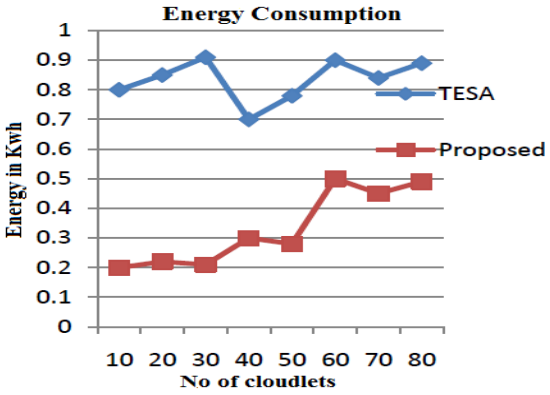


Figure 3. Energy consumption

## 6. CONCLUSION

The migration of the virtual machine is the process that is applied to migrate one virtual machine's cloudlets to another. In this study, it is revealed that execution time and space consumption are raised at a constant rate due to virtual machine overloading. In this article, the ACO technique is utilized, which will migrate the virtual machine task that is overloaded for effective execution to another virtual machine. In CloudSim, the efficiency of the suggested approach is examined and it is analyzed that after virtual machine migration, execution time and space usage are decreased.

## REFERENCES

[1] R. Kaur and G. Kaur, "Proactive Scheduling in Cloud Computing," *Bulletin of Electrical Engineering and Informtics*, vol. 6, no. 2, June 2017, pp. 174-180, doi: 10.11591/eei.v6i2.64.

[2] I. Odun-Ayo, Toro-Abasi Williams, M. Odusami, and J. Yahaya, "A Systematic Mapping Study of Performance Analysis and Modelling of Cloud Systems and Applications," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 2, April 2021, pp. 1839-1848, doi: 10.11591/ijece.v11i2.pp1839-1848.

[3] M. Wu, Ting-Jie Lu, Fei-Yang Ling, J. Sun, and Hui-Ying Du, "Research on the architecture of Internet of Things," *2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, 2010, pp. V5-484-V5-487, doi: 10.1109/ICACTE.2010.5579493.

[4] D. Jiang, P. Zhang, Z. Lv, and H. Song, "Energy-Efficient Multi-Constraint Routing Algorithm with Load Balancing for Smart City Applications," in *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1437-1447, December 2016, doi: 10.1109/JIOT.2016.2613111.

[5] D. Mishra and S. De, "Energy Harvesting and Sustainable M2M Communication in 5G Mobile Technologies," *Book series on Modeling and Optimization in Science and Technologies*, Springer, vol. 8, pp. 99-125, April 2016, doi: 10.1007/978-3-319-30913-2_6.

[6] J. Chen, Y. Gong, M. Fiorani, and S. Aleksic, "Optical interconnects at the top of the rack for energy-efficient data centers," in *IEEE Communications Magazine*, vol. 53, no. 8, pp. 140-148, August 2015, doi: 10.1109/MCOM.2015.7180521.

[7] E. Caron, F. Desprez, and A. Muresan, "Forecasting for Cloud computing on-demand resources based on pattern matching," *INRIA, Research Report* RR-7217, July 2010.

[8] M. Maurer, I. Brandic, and R. Sakellariou, "Enacting SLAs in Clouds Using Rules," in *Proceedings of the 17th International Conference on Parallel Processing, Euro-Par'11-Volume Part I*. Berlin, Heidelberg: Springer-Verlag, vol 6852, pp. 455-466, 2011, doi: 10.1007/978-3-642-23400-2_42.

[9] A. Naskos, *et al*., "Dependable Horizontal Scaling Based on Probabilistic Model Checking," *2015 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, 2015, pp. 31-40, doi: 10.1109/CCGrid.2015.91.

[10] H. C. Lim, S. Babu, and J. S. Chase, "Automated Control for Elastic Storage," in *Proceedings of the 7th International Conference on Autonomic Computing, ICAC '10*, New York, NY, USA: ACM, June 2010, pp. 1-10, doi: 10.1145/1809049.1809051.

[11] J. Huang, C. Li, and J. Yu, "Resource prediction based on double exponential smoothing in cloud computing," *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, 2012, pp. 2056-2060, doi: 10.1109/CECNet.2012.6201461.

[12] H. Mi, H. Wang, G. Yin, Y. Zhou, D. Shi, and L. Yuan, "Online Self-Reconfiguration with Performance Guarantee for Energy-Efficient Large-Scale Cloud Computing Data Centers," *2010 IEEE International Conference on Services Computing*, 2010, pp. 514-521, doi: 10.1109/SCC.2010.69.

[13] Z. Gong, X. Gu and J. Wilkes, "PRESS: PRedictive Elastic ReSource Scaling for cloud systems," *2010 International Conference on Network and Service Management*, 2010, pp. 9-16, doi: 10.1109/CNSM.2010.5691343.

[14] G. A. Moreno, J. Camara, D. Garlan, and B. Schmerl, "Proactive Self-adaptation Under Uncertainty: A Probabilistic Model Checking Approach," in *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering, ESEC/FSE 2015*, New York, NY, USA: ACM, August 2015, pp. 1-12, doi: 10.1145/2786805.2786853.

[15] A. Najjar, X. Serpaggi, C. Gravier, and O. Boissier, "Multi-agent Negotiation for User-centric Elasticity Management in the Cloud," *2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing*, 2013, pp. 357-362, doi: 10.1109/UCC.2013.104.

[16] X. Dutreilh, A. Moreau, J. Malenfant, N. Rivierre, and I. Truck, "From Data Center Resource Allocation to Control Theory and Back," *2010 IEEE 3rd International Conference on Cloud Computing*, 2010, pp. 410-417, doi: 10.1109/CLOUD.2010.55.

[17] A. Ali-Eldin, J. Tordsson, and E. Elmroth, "An adaptive hybrid elasticity controller for cloud infrastructures," *2012 IEEE Network Operations and Management Symposium*, 2012, pp. 204-212, doi: 10.1109/NOMS.2012.6211900.

[18] G. Tesauro, N. K. Jong, R. Das, and M. N. Bennani, "A Hybrid Reinforcement Learning Approach to Autonomic Resource Allocation," *2006 IEEE International Conference on Autonomic Computing*, 2006, pp. 65-73, doi: 10.1109/ICAC.2006.1662383.

[19] M. Masoudi, B. Khamidehi, and C. Cavdar, "Green Cloud Computing for Multi Cell Networks," *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1-6, doi: 10.1109/WCNC.2017.7925672.

[20] G. J. Rao and G. S. Babu, "Energy analysis of task scheduling algorithms in green cloud," *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 2017, pp. 302-305, doi: 10.1109/ICIMIA.2017.7975624.

[21] Z. Goudarzi and A. Faraahi, "Effective load balancing in cloud computing," *Int. J. Intell. Inf. Syst*., vol. 3, no. 6, pp. 34-42, 2017.

[22] E. Arianyan, "Multi objective consolidation of virtual machines for green computing in Cloud data centers," *2016 8th International Symposium on Telecommunications (IST)*, 2016, pp. 654-659, doi: 10.1109/ISTEL.2016.7881903.

[23] C. Gu, P. Shi, S. Shi, H. Huang, and X. Jia, "A Tree Regression-Based Approach for VM Power Metering," in *IEEE Access*, vol. 3, pp. 610-621, 2015, doi: 10.1109/ACCESS.2015.2430276.

[24] Z. Zhou, Zhi-Gang Hu, T. Song, and Jun-Yang Yu, "A novel virtual machine deployment algorithm with energy efficiency in cloud computing," *Journal of Central South University*, vol. 22, no. 3, pp. 974-983, March 2015, doi: 10.1007/s11771-015-2608-5.

[25] S. Islam, J. Keung, K. Lee, and A. Liu, "Empirical Prediction Models for Adaptive Resource Provisioning in the Cloud," *Future Generation Computer Systems*, vol. 28, no. 1, pp. 155-162, January 2019, doi: 10.1016/j.future.2011.05.027.

[26] M. R. Jalali and A. Afshar, "Ant Colony Optimization Algorithm (ACO); A New Heuristic Approach for Engineering Optimization," *WSEAS Transactions on Information Science and Applications*, Lisbon, Portugal, vol. 2, no. 5, pp. 606-610, June 2005.
[27] H. Ouarnoughi, J. Boukhobza, F. Singhoff, and S. Rubini, "Integrating I/Os in Cloudsim for Performance and Energy Estimation," *ACM SIGOPS Operating Systems Review*, vol. 50, no. 2, pp. 27-36, December 2016, doi: 10.1145/3041710.3041715.

## BIOGRAPHIES OF AUTHORS

**Yasameen A. Ghani Alyouzbaki** was born in Iraq, 1969. She received B.Sc. degree in Electrical Engineering from Mustansiryah University, Iraq, in 1993, and M.Sc. degree in Communications and Electronics Engineering from Mustansiryah University, Baghdad, in 2009. And Ph.D. in Communications and Electronics Engineering from Politehnica University, Romania in 2017. She had worked as electrical engineer for several years at Faculty of Engineering, Iraq. Currently, she is working as a Lecturer in Mustansiriyah University, Faculty of Engineering, Computer Engineering Department, Iraq. His research interests include computer communication networks, digital communications, image processing and biomedical image processing, e-mail: yasaalyozbaki@uomustansiriyah.edu.iq.

**Muaayed F. AL-Rawi** was born in Iraq, 1971. He received B.Sc. degree in Electrical and Nuclear Engineering from Baghdad University, Iraq, in 1992, and M.Sc. degree in Communication and Electronics Engineering from Jordan University of Science and Technology, Jordan, in 1999. He had worked as nuclear and electrical engineer for several years at Iraqi Atomic Energy Organization, Iraq. Currently he is with Department of Computer Engineering, Mustansiryia University, Iraq. His research interests include computer communication networks, digital communications, analogue and digital electronics, digital signal processing, and biomedical engineering. E-mail, muaayed@uomustansiriyah.edu.iq.