

VM placement with effective energy management in cloud using optimal VM allocation framework (OAAF)

Susmita J A Nair¹, T R Gopalakrishnan Nair²

¹India Networks and Security Research Group, Bharathiar University, India

^{1,2}Rajarajeswari Group of Institutions, India

Article Info

Article history:

Received Oct 25, 2019

Revised Nov 9, 2019

Accepted Dec 7, 2019

Keywords:

Energy consumption

Live migration

VM placement

ABSTRACT

The performance of servers at the data centers is affected when the servers are overloaded. To overcome this problem, the workload at the overloaded servers has to be redistributed to other servers which is possible with live VM migration. Live migration plays a crucial role in handling the overload at the data centers without service interruption. However, live migration also incurs some performance loss and energy overhead. The energy consumption at the data centers is a matter of utmost concern both in terms of economy and ecology. In this paper we are proposing a novel approach to find the most suitable server for VM placement. We have introduced an Optimal VM Allocation Framework (OAAF) in which the hosts at the source requests the destination for their available slots. Based on the response from the available servers, the utilization factor is calculated and the selection of appropriate destination for VM placement is done. Simulations carried out have shown 10% improvement in energy saving.

Copyright © 2020 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Susmita J A Nair

Networks and Security Research Group,
Bharathiar University, Coimbatore, India.

Email: susmitajanair@gmail.com

1. INTRODUCTION

In cloud computing the IaaS layer provides high computation speed and large storage capacity for an enterprise. Virtualization in servers helps the data center to function as a sophisticated cloud architecture [1, 2]. A cluster of host forms the core of the datacenters. At the datacenters, the hosts may experience more workload. To maintain the efficiency and performance of hosts, the workload has to be transferred to hosts or datacenters with lesser workload. This is achieved through VM Migration. VM migration requires extensive IO operations which results in high energy consumption. The other factors which contributes to the energy consumption are: an idle server, power consumption by server, data transmission etc. The data center manager aims at the ways to develop VM Placement Policy for energy reduction and maximize resource utilization and return of investment [1-4]. An energy and cost efficient migration can be achieved by properly choosing the VM to be migrated and considering the factors like the destination where the VM has to be copied, bandwidth availability, downtime and migration time [5]. The self-managing techniques implemented at the data centers for dynamic allocation of resources results in reduction of efficiency [6].

Recent studies show that even the idle servers consume 69-97 % of the total energy consumed by a fully utilized server [7]. Even the shutting down process of a physical machine for a specified time and restarting it after elapse of the inert period does not significantly contribute towards an acceptable energy saving procedure, as the switch on-switch off cycle leads to several steady state or quiescent state violations calling for enhanced power requirements. In this paper we are concentrating on the energy consumption during live migration. The energy, E can be calculated as,

$$E=P * T$$

where P represent the power consumption and T is time taken through the process of live migration.

On an average, the major sections of energy consumption at the data centers are infrastructure, server, cooling etc., as shown in Figure 1 [8, 9].

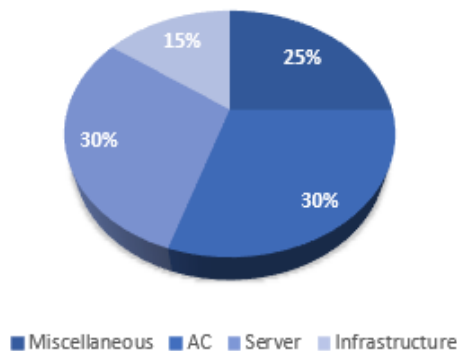


Figure 1. Average energy consumption at the data centers

2. RELATED WORK

Various studies have been conducted to estimate and reduce the energy consumption at the data centers during live VM Migration. In [3], the author has developed a model by applying linear regression on the available data. The model helps to estimate the cost of energy consumed during live VM Migration with 90% accuracy. In an incessant VM demand state, static allocation is considered to be the better option whereas for sporadic VM requests, dynamic allocation is more efficient [7]. In [8], the authors have suggested integer quadratic program for power optimization at the data center. This method, however, is not suitable for systems with heavy workload.

Li et al. in [10] have proposed the multi-agent method for VM consolidation to balance the workload at the physical machines by VM allocation. However, the best results of this approach is attributed to the proper VM selection process. The Iterative Weighted Linear Regression method (IWLR) has been used to predict the host utilization and select the most suitable host for migrating the VMs [11].

The traditional resource allocation methods viz, First Fit, Worst Fit and Best fit can be applied to the VM consolidation problem. It can be considered as an online bin packing optimization problem where the randomly arriving VM requests (objects) must be accommodated into the available PMs(bins). The requirement of a new PM arises when the allocation methods fail to find a suitable PM for allocating the VM [12]. The heuristics-based dynamic approaches does not guarantee optimal performance, especially in the worst cases [13].

In [14, 15], the author has analyzed the dynamic part of power consumption by a physical machine. The difference between the energy consumed before migration and the energy consumed during migration gives the energy overhead incurred during live migration.

$$\text{Energy}_{\text{overhead}} = \text{Energy}_{\text{beforemig}} - \text{Energy}_{\text{duringmig}}$$

In [16], the authors made an observation that approximately 20 seconds are required for the source and destination servers to balance the level of power consumption with the number of allocated VMs. In [17], the authors have suggested Medium Fit method to balance between the overload threshold and underload threshold. The power efficiency of a physical server is calculated as $\text{Power}_{\text{efficiency}} = \text{CPU}_{\text{total}} - \text{Max}_{\text{power}}$

In [18], the authors have used the correlation of coefficient to determine the PM that has to be chosen for VM Placement. The PM with value close to -1 is chosen as the appropriate PM. The case of multiple PMs with the same co-efficient is not considered here. In [19, 20], the authors have computed the total energy consumption, TE, in the time duration between t_0 and t_1 as follows.

$$TE = \int_{t_0}^{t_1} \text{Power}_{\text{util}}(t) dt \quad (1)$$

The CPU utilization can be enhanced to minimize the energy consumption. In [21], the authors mention power usage effectiveness (PUE) as a factor to measure the energy efficiency.

$$PUE_i = \frac{power_{total}}{power_{IT}} \tag{2}$$

$PUE \geq 1$. The lesser the value of PUE, the data center will be more energy efficient.

3. OPTIMAL VM ALLOCATION FRAMEWORK

In [22, 23], the authors have used the following four categories of VM consolidation.

- a) The overloaded host detection.
- b) The Selection of VM to be migrated
- c) The selection of destination for VM Placement
- d) The underloaded host detection.

The VM migration comprise of mainly a preparatory work at the source and the Host Acceptance work at the destination. Once the workload at the source datacenter reaches a threshold, say 80%, the VMs and the workload from the particular datacenter has to be redistributed to another datacenter. The second step is the selection of appropriate VM for migration. The VM that is selected for migration has to be placed at a destination server. The destination is a server with moderate workload, which can accommodate the migrated VM. Not only does the overloaded servers affects the performance, even running the datacenters with very less workload also affects the performance. This the VMs from the underloaded host also need to be migrated so that they can be shut down to save energy.

Load balancing at the IaaS level can be effectively done by predicting the future resource requirements of the new tasks and allocating the resources at the physical machine accordingly [24]. Based on the predicted load the set of physical machines is sorted in descending order. The last PM in the sorted list, say PM1 which is the most underloaded server, is chosen. Further in PM1, the VMs with the minimum load is selected for migration. Suitable PM is searched from the sorted list which can accommodate VMs. The process is repeated until all the VMs from underloaded servers are migrated to other physical machines [25].

The migration process is based on the three statistics obtained from source and destination servers. a) Request statistics from Source (HostRequestStatistics), b) Reply statistics from Destination (DestinationReplyStatistics) and c) the statistics of Energy Efficiency (EnergyEfficiencyStatistics).

The Minimum Migration Time (MMT) policy for VM Placement is found to be efficient [20, 25]. In MMT, the VM to be migrated, say vm , is selected based on the (16). In the equation $Mem_{util}(vm)$ represents the memory utilization of the VM, vm , selected for migration and $Mem_{util}(q)$ represents the quantity of memory utilization of other VMs on host H. The notation $unusedbw_H$ is used to denote the network bandwidth unused at host H.

$$vm \in H | \forall q \in H, \frac{Mem_{util}(vm)}{unusedbw_H} \leq \frac{Mem_{util}(q)}{unusedbw_H} \tag{3}$$

3.1. Migration Request Protocol

The analysis has to be done for the request and reply statistics. The statistics is based on the spatial and temporal conditions. The spatial conditions include the resource listing and capacity of utilization. The temporal conditions are done on the basis of past, present and future migrations(prediction).

For minimizing the energy involved in migration, the number of repetitions, i.e, the transfer attempts have to be minimized. For that solid maximum successful transfer only should take place. The smaller VMs can be offloaded to another server. The total time for setup and data content transfer has to be minimum. In order to save maximum energy, the most effective way is to reduce the number of active servers [26]. The major energy consumption areas are Computation energy, Migration energy, Switching energy and the overhead energy [27].

$$Total_{energy} = Comp_{energy} + Mig_{energy} + Switch_{energy} + Overhead_{energy} \tag{4}$$

The objective is to minimize the number of operational servers (S) with reduced energy usage. For ‘n’ physical machines the sum of energy consumption, ep , should be minimum. While migrating the resource requirement of virtual machines, VR , should be less than the physical resources, PR , available. The Virtual Machine Placement problem can be formulated as follows.

$$Minimize S = \sum_{i=1}^n ep_i$$

Subject to,

$$\sum_{j=1}^m V_j R \leq PR$$

Algorithm 1: OptimalSlotSelectionAlgorithm

Input: $H \triangleq \{h_1, h_2, h_3, \dots, h_n\}$ represent n hosts

Output: Allocate the VMs with proper load balancing

1. Track the overloaded hosts.
2. For each host, h_i in the hostList H,
3. Calculate resource utilization

$$U_i = \frac{1}{h_i} \sum_{k=1}^n R_k(t) \quad (5)$$

4. if the utilization is above threshold, i.e.,
 $U_i > 80\%$ then
5. OverloadHostList.add(h_i)
6. Endif
7. Endfor
8. Get the VM Request Statistics
9. Use the HostRequestStatistics
10. $VM_a \leftarrow$ Select a VM from Overload Host
11. SelectVMList.add(VM_a)
12. Choose the destination servers for VM allocation.
 $D =$ DestinationReplyStatistics()
13. Allocate the tasks to the most suitable destination server, with minimum energy consumption using the EnergyEfficientAlgorithm().

Algorithm 2: HostRequestStatistics

Input : $OH \triangleq \{oh_1, oh_2, oh_3, \dots, oh_r\}$ represents the Overloaded hosts.

Output: The list of Underloaded Hosts where migrated VMs can be placed.

1. Sort the hosts in descending order based on the CPU utilization.
2. Predict the Host for migration based on Iterative Weighted Linear Regression method.
3. The simple regression line is shown below.
 $Y = a + bX$
 The regression co-efficients a and b are derived using least square method.
4. The tricube weight function can be used to predict the K values of host utilization.
 $W = (1 - (p)^3)^3$
 where p is the difference between the current and the last observation.
5. If $W < 1$, UH=add the host to underloaded host list
6. Return UH.

Algorithm 3: DestinationReplyStatistics

Input : $UH \triangleq \{uh_1, uh_2, uh_3, \dots, uh_v\}$ represents the Overloaded hosts.

Output: Server with correlation co-efficient ≈ -1 .

For hosts in the UH list, choose

$D \triangleq \{d_1, d_2, d_3, \dots, d_m\}$ for space allocation.

1. Select the VM with Minimum Migration time.
2. Calculate γ , the correlation co-efficient as shown below.

$$\gamma = \frac{\sum_{i=1}^m (e - \bar{e})(r - \bar{r})}{\sqrt{\sum_{i=1}^m (e - \bar{e})^2 \cdot \sum_{i=1}^m (r - \bar{r})^2}} \quad (6)$$

where, \bar{e} is the mean of the estimated resource demands and \bar{r} is the mean of the residual.

3. Choose the servers from D , with $\gamma \approx -1$.
4. Return D

Algorithm 4: EnergyEfficiencyEstimation

Input : Power utilization by various components at the data center.

Output: Server with minimum Energy consumption.

1. Determine the Euclidean distance for the resource request and the resource utilization of the selected servers.
2. Choose the server with maximum distance so that the servers are fully utilized.
3. Obtain the power usage values while the ith server is inactive and also when it is completely utilized.
4. Calculate the power consumption for the migration process with M memory size and the allotted bandwidth B. Assume the migration starts at time t₁ and ends at time t₂.

$$Migr_{energy} = \int_{t_1}^{t_2} Power_{src} * \frac{M}{B} + \int_{t_1}^{t_2} Power_{src} * \frac{M}{B} \tag{7}$$

5. The maximum power usage by the physical machine and the CPU utilization are the parameters that affects the power consumption at the server [28-30].

$$Power_{cu} = n . Max_{power} + (1 - n) . Max_{power} . cu \\ = Max_{power} . (0.7 + 0.3.cu) \tag{8}$$

Here n represents the fraction of power that the inactive server consumes, cu denotes the CPU utilization. The variations in workload changes the CPU utilization over time, t.

6. The computation energy by servers can be computed as

$$Comp_{energy} = \sum_{i=0}^n Power_{cu} * time \tag{9}$$

where n denotes the number of servers currently in use.

7. The Overhead energy is the difference between the power consumed during migration and the idle power consumption at the source and target servers.

$$OH_{energy} = Migr_{energy} - Idle_{energy} \tag{10}$$

8. Compute the overall energy consumption as the sum of energy computation cost, energy for migration, energy for server switching and the energy overhead as

$$Total_{energy} = Comp_{energy} + Migr_{energy} + OH_{energy} \tag{11}$$

4. RESULT AND ANALYSIS

The simulation set up has been done in Cloudsim 3.0.3, by randomly selecting the values for the number of VMs to be migrated with the settings shown in Table-1. The features of the VM is similar to the features of AmazonEC2 instance types, but with single core VM. The Minimum Migration Time (MMT) policy is used in the model which selects the VM with minimum migration time requirement [31].

Table 1. Experimental Settings

Parameters	Configuration
Server Types	HPProLiantML110G4(2X1800MIPS) HPProLiantML110G5(2X2660MIPS)
Number of Servers	800 diverse servers;400 of each host type
Workloads	PlanetLab (10 days of traces)
Overload decision	Iterative Weighted Local Regression

The increase in number of VMs increases the computation energy, migration energy and overhead energy. This in turn increases the total energy consumption. Figure 2 shows the variation of energy with the variation in number of VMs.

The energy consumption for a particular run with and without OVAF is shown in the Figure 3. The result of implementation of OVAF has shown an improvement of approximately 10% in energy consumption as shown in Figure 4.

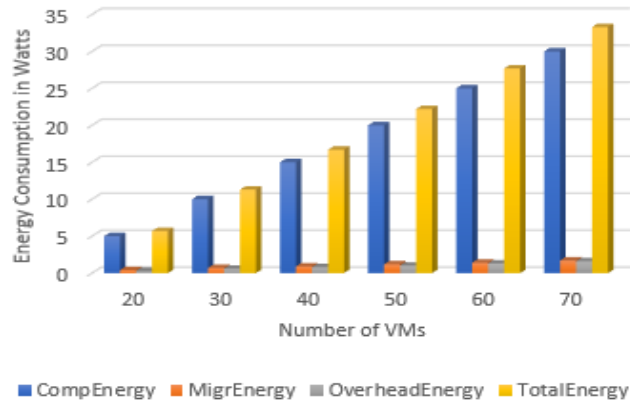


Figure 2. Graph representing the variation in energy consumption with the increasing number of VM



Figure 3. Energy Consumption with and without OVAF

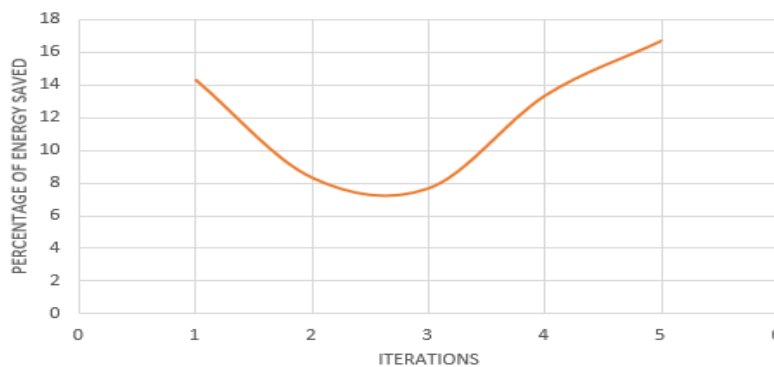


Figure 4. Improvement in energy consumption

Factors such as hardware, bandwidth etc., affect the premigration and postmigration scenario. Our methodology helps in energy savings to a certain extent which can be complemented by the following methods, but at a certain cost (i) An improved bandwidth of the migration channel which is an important parameter that decides the efficiency of migration process. (ii) Using the finest set of processors, memory and storage for migration activities can also help in energy saving.

5. CONCLUSION

The upsurge in the number of data centers is a major threat to the environment. The energy consumption at the data centers are static and dynamic in nature. Energy is consumed during VM Migration and even while the servers are idle. The energy consumption can be reduced with the proper selection of hosts during migration. The selection of host is done by using the most appropriate methods like Iterative Weighted Local Regression (IWLR) and with the Euclidean methods. Our experiment has shown that with the implementation of our algorithm the energy consumed during VM migration can be brought down to a certain extent.

REFERENCES

- [1] Susmita J A Nair, T.R. Gopalakrishnan Nair, "Resource Extension Techniques in Enterprise Architectures through Cloud Connectivity", ICICT-2016, Aug 26-27.
- [2] Susmita J A Nair, T.R. Gopalakrishnan Nair, "Energy Efficiency Estimation in Cloud During VM Migration" ICTRCET'18, May 17-18, 2018, Bangalore, India.
- [3] Haikun Liu, Cheng-Zhong Xu, Hai Jin, Jiayu Gong, Xiaofei Liao, "Performance and Energy Modeling for Live Migration of Virtual Machines", HPDC'11, June 8-11, 2011, San Jose, California, USA.
- [4] Shahram Jamali, Sepideh Malektaji & Morteza Analoui, "An imperialist competitive algorithm for virtual machine placement in cloud computing", *Journal of Experimental & Theoretical Artificial Intelligence*, DOI: 10.1080/0952813X.2016.1212101.
- [5] N. Kumara, S. Saxena "Migration Performance of Cloud Applications - A Quantitative Analysis", International Conference on Advanced Computing Technologies and Applications (ICACTA2015), *Procedia Computer Science* 45 (2015) 823 – 831
- [6] Danilo Ardagna, Barbara Panicucci, Marco Trubian, and Li Zhang, "Energy-Aware Autonomic Resource Allocation in Multitier Virtualized Environments," *IEEE Transactions On Services Computing*, Vol. 5, No. 1, January-March 2012
- [7] Andreas Wolke, Martin Bichler, and Thomas Setzer, "Planning vs. Dynamic Control: Resource Allocation in Corporate Clouds," *IEEE Transactions On Cloud Computing*, Vol. 4, No. 3, July-September 2016.
- [8] S. Vakili et al "Energy Efficient Resource Allocation in Cloud Computing Environments", *IEEE Access, Special section on future networks: architectures, protocols, and applications*, December 1, 2016.
- [9] Miyuru Dayarathna, YonggangWen, Rui Fan, "Data Center Energy Consumption Modeling: A Survey", *IEEE Communications Surveys & Tutorials*, Vol. 18, No. 1, First Quarter 2016.
- [10] Hongjian Li Guofeng Zhu1 Chengyuan Cui1 Hong Tang1 Yusheng Dou Chen He, "Energy-efficient migration and consolidation algorithm of virtual machines in data centers for cloud computing", *Computing*, Mar2016, Vol. 98 Issue 3, p303-317. 15p
- [11] M Ali, et.al, "Energy-Efficient Algorithms for Dynamic Virtual Machine Consolidation in Cloud Data Centers", *IEEE Access, Special Section On Green Cloud And Fog Computing: Energy Efficient And Sustainable Infrastructures, Protocols And Applications*, June 1, 2017.
- [12] Mehdi Dabbagh, Bechir Hamdaoui, Mohsen Guizani, and Ammar Rayes, "Energy-Efficient Cloud Computing: Prediction, Consolidation, and Overcommitment", *IEEE Network*, March/April 2015.
- [13] A. Beloglazov, R. Buyya, "Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers", Wiley InterScience(www.interscience.wiley.com). DOI: 10.1002/cpe.1867.
- [14] Anja Strunk, "A Lightweight Model for Estimating Energy Cost of Live Migration of Virtual Machines", 2013 IEEE Sixth International Conference on Cloud Computing, 978-0-7695-5028-2/13 © 2013 IEEE DOI 10.1109/CLOUD.2013.17
- [15] Anja Strunk "Costs of Virtual Machine Live Migration: A Survey", 2012 IEEE Eighth World Congress on Services, 978- DOI 10.1109/Services.2012.23
- [16] Mar Callau-Zori, et.al, "An experiment-driven energy consumption model for virtual machine management systems." *Sustainable Computing: Informatics and Systems*, Elsevier, 2018, 18, pp.163-174. 10.1016/j.suscom.2017.11.001.hal-01632962
- [17] Moges and Abebe *Journal of Cloud Computing: Advances, Systems and Applications* (2019) 8:2 <https://doi.org/10.1186/s13677-019-0126-y> "Energy-aware VM placement algorithms for the OpenStack Neat consolidation framework".
- [18] H. Shen, L.Chen, "CompVM: A Complementary VM Allocation Mechanism for Cloud Systems", *IEEE/ACM Transactions On Networking*, March 2018, DOI: 10.1109/TNET.2018.2822627.
- [19] Wenhong Tian T. "On minimizing total energy consumption in the scheduling of virtual machine reservations", *Journal of Network and Computer Applications*, Volume 113, 1 July 2018, Pages 64-74
- [20] Y. Liu et al." EQVC Method in Cloud Environment", *IEEE Access* Vol 6 2018, DOI:10.1109/ACCESS.2018.2835670.
- [21] Khosravi A., Garg S.K., Buyya R., "Energy and Carbon-Efficient Placement of Virtual Machines in Distributed Cloud Data Centers". In: Wolf F., Mohr B., and Mey D. (eds) Euro-Par 2013 Parallel Processing. Euro-Par 2013. *Lecture Notes in Computer Science*, vol 8097. Springer, Berlin, Heidelberg.

- [22] Abdelkhalik Mosa and Norman W.Paton, "Optimizing virtual machine placement for energy and SLA in clouds using utility functions", *Journal of Cloud Computing: Advances, Systems and Applications (2016)* 5:17 DOI:10.1186/s13677-016-0067-7.
- [23] A. Beloglazov, R. Buyya, "Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers", Wiley Inter Science (www.interscience.wiley.com). DOI: 10.1002/cpe.1867.
- [24] Q. Z. Ullah et.al, "Adaptive Resource Utilization Prediction System for IaaS Cloud", *Hindawi Comp. Intelligence and Neuroscience Vol 2017*, Article ID 4873459, <https://doi.org/10.1155/2017/4873459>.
- [25] Q. Huang et.al, "Prediction-based Dynamic Resource Scheduling for Virtualized Cloud Systems", *Journal Of Networks*, Vol. 9, No. 2, February 2014.
- [26] X. Liu et.al, "An Energy Efficient Ant Colony System for Virtual Machine Placement in Cloud Computing", *IEEE Transactions On Evolutionary Computation*, Vol. 22, No. 1, February 2018.
- [27] Atefeh Khosravi, et.al, "Online virtual machine migration for renewable energy usage maximization in geographically distributed cloud data centers", wileyonlinelibrary.com/journal/cpe, Copyright©2017 John Wiley & Sons Ltd. <https://doi.org/10.1002/cpe.4125>.
- [28] Anton Beloglazov and Rajkumar Buyya, "Adaptive Threshold-Based Approach for Energy- Efficient Consolidation of Virtual Machines in Cloud Data Centers".
- [29] Wanyuan Wang, et.al, "Multiagent-Based Resource Allocation for Energy Minimization in Cloud Computing Systems", *IEEE Transactions On Systems, Man, And Cybernetics: Systems*, Vol. 47, No. 2, February 2017
- [30] Xijia Zhou, Kenli Li, Chubo Liu 1, Keqin Li, "An Experience-Based Scheme for Energy-SLA Balance in Cloud Data Centers", *IEEE Access*, Feb 2019, DOI: 10.1109/ACCESS.2019.2899101.
- [31] Susmita J A Nair, T R Gopalakrishnan Nair, "Performance Degradation Assessment and VM Placement Policy in Cloud," *International Journal of Electrical and Computer Engineering (IJECE)* Vol. 9, No. 6, December 2019, pp. 4961-4969

BIOGRAPHIES OF AUTHORS



Susmita J A Nair holds a graduate degree in computer science and a post graduate degree in Computer Applications. She is currently pursuing her Ph.D. in Computer Science from Bharathiar University, Coimbatore, India. Her area of interest is Cloud Computing.



Dr. T.R. Gopalakrishnan Nair, a fellow of Institution of Engineers, has 34 years of experience in professional field spread over Research, Industry and Education. Currently, he is the Rector for Rajarajeswari Group of Institutions in India. He was the ARAMCO Endowed Chair in Technology in PM University, KSA. He holds degrees M.Tech. (I.I.Sc., India) and a Ph.D. in Computer Science. His areas of interest include Advanced networks, Cognitive Systems and Multidisciplinary studies including Brain and physical systems. He is a senior member of IEEE, ACM and few other professional bodies.