IJCSE International Journal of Computer Sciences and Engineering Open Access

Survey Paper Vol.-7, Special Issue-16, May 2019 E-ISSN: 2347-2693

# Dynamic Threshold Based Load Balancing and Server Consolidation in Cloud

**Geetha Megharaj[1*], Mohan G. Kabadi[2]**

[1]Computer Science and Engineering, Sri Krishna Institute of Technology, Bangalore, India
[2]Computer Science and Engineering, Presidency University, Bangalore, India

[*]*Corresponding Author: geethagvit@yahoo.com, Tel.: 080-23514539*

DOI: https://doi.org/10.26438/ijcse/v7si16.117121 | Available online at: www.ijcseonline.org

***Abstract*** - High power consumption in Cloud Data center leads to high emissions of carbon which is unsuitable for environment. Energy consumption in the data center can be reduced by balancing load among active physical nodes and minimizing the number of active servers which are lightly loaded. Static lower and upper thresholds are not suitable for dynamically changing resource usage of physical machines. Dynamic Threshold based load balancing algorithms are proposed: i)Upper threshold dynamically varied based on CPU utilization and the Lower Threshold is predefined. ii) The average utilization of all the machines in the datacenter is used to define Upper Threshold. System is monitored at regular intervals and whenever server load goes above the Upper Threshold or lower than Lower Threshold, system identified as in imbalance state and Virtual Machine migration is initiated for load balance and for server consolidation. Simulation results shows the proposed schemes can improve resource utilization and energy.

***Keywords—*** VM Migration, Load Balancing, Server consolidation, Dynamic Threshold

## I. INTRODUCTION

Large amount of energy utilization in Cloud Data Center results in high operational cost and carbon dioxide (CO2) emissions. As per the statistics, 3% of total electricity production of globe is consumed by data centers producing nearly 200 million metric tons of CO2 [1] and this percentage has been increasing over the years. Significant research has been done to make data centers more environment friendly, to reduce consumption of energy.

Virtualization technology helps in optimized use of power in Data Centers by allowing multiple Virtual Machines (VMs) to run on single physical machine and turning off underutilized physical machines in the process of server consolidation. The performance of data center might degrade as the number of Virtual Machines may increase dynamically which leads to overutilization of CPU or increase in energy consumption. To overcome this, the process of server consolidation need to be initiated to perform live migration of virtual machines and reduce the number of active hosts.

Normally Lower Threshold defines underloaded machine and Upper Threshold defines overloaded machine. But these static thresholds are not suitable for cloud environment since resource requirement in it changes dynamically. So in this paper we have proposed dynamic threshold based VM migration for load balancing and server consolidation. If load on the host is below the lower threshold all VM running on that host are moved to the suitable target host and the source host is shutdown. This process leads to server consolidation. If load on the host is greater than the upper threshold then host is identified to be overloaded and some VMs are migrated to reduce load on overloaded host. VM migration [2, 3] techniques are used in the load balancing.

The rest of this paper is organized as follows. Section II describes related work. In Section III the proposed dynamic threshold algorithms are described. In Section IV simulation setup and results of algorithms are explained and Section V Conclusion is presented.

## II. RELATED WORK

Nathuji and Schwan [4] have shown the energy advantages in Data Center by applying dynamic Virtual Machine Consolidation(VMC) and they concluded that energy consumption is minimized and observed 34% of improvement in power consumption after VMs are consolidated on few machines.

The authors in [5] have used fixed or static threshold to identify overloaded and underloaded machines. If CPU utilization of physical machine reduced below the lower utilization threshold, all the virtual machines are migrated from this physical machine to suitable physical machine and the underutilized machine is turned off for optimal use of energy. On the other hand if the CPU utilization exceeds the upper utilization threshold some virtual machines are selected for migrating to avoid performance degradation. The policies used for VM selection are Minimization of Migrations (MM), Highest Potential Growth (HPG) and Random Choice (RC). The simulation results showed the flexibility of the proposed algorithms.

Cloud is dynamic computing environment in which workload keeps changing. Therefore setting static threshold is not suitable. Therefore, the work done in [5] is continued by authors in [6] in which they suggested a system which changes its behaviors based on patterns of workload. A technique based on energy efficient dynamic threshold for consolidation of VMs is proposed which adjusts resource utilization threshold dynamically and guarantees the Service Level Agreements (SLA).

In [7] authors have proposed technique for autoadjustment of utilization threshold, Minimization of Migration policy for VM selection and modified best fit algorithm for VM placement and demonstrated that their algorithm performs better compared to static threshold policies.

Extensive contributions have been made to achieve SC through VM migration technique. Various techniques for SC in virtualized data center has been discussed in [8]. In [9], two VM migration techniques namely–Hybrid and Dynamic Round Robin (DRR) was presented. Two states were defined in the solution framework called–retiring and non-retiring. If VM contains limited number of active VMs which are about to finish their task, then, the PM is in retiring state, else, it is in non-retiring state. The retiring PMs will not accept new tasks, and the active VMs are migrated to suitable PMs. Both, Hybrid and DRR exhibit excellent performance with respect to reducing power consumption in CCs.

Most of the VM migration techniques for SC are modeled through Bin Packing Problem (BPP), which is NP-complete. An approximation scheme based on First Fit Decreasing algorithm was proposed in [10] to effectively migrate VMs. Each bin is considered as a PM, and the highest priority PMs are subjected to VM migration.

The Magnet scheme proposed in [11], performs selection of suitable subsets of available PMs which can guarantee the expected performance levels. The PMs outside the selected subset are shutdown.

In [12] the virtual Machine task migration technique is extended to address the Server Consolidation issue and static CPU utilization and power utilization threshold used to identify overloaded hosts. In [13] an approach for load balancing based on dynamic threshold is proposed. When the host load is greater than the upper threshold some virtual machine are migrated to reduce load on machine. They consider the average CPU utilization of the host to decide the threshold. Dynamic Threshold is used in this approach but server consolidation is not done .

In [14], J. M. Galloway has proposed power aware load balancing technique. If CPU utilization of all hosts is greater than 75%, a new virtual machine is instantiated on the compute node having the lowest utilization. Otherwise, the new virtual machine (VM) was booted on the compute node with the highest utilization (if it can accommodate the size of the VM). Shivani Gupta [15] proposed a load balancing scheme with dynamic lower and upper threshold and demonstrate that it increases the resource utilization compared to the traditional VM migration algorithm.

## III. PROPOSED ALGORITHMS

Cloud environment consists of data center with number of Physical Machines (PM). Each physical machine can run number of VM. When user demands for the resources VMM create a VM and assign to the user. VMM is a main part of the virtualization, which handle all VM related task. So VM creation, deletion and scheduling all are done by the VMM. It is also responsible for the monitoring of the resources such as CPU, RAM used by the VM and PM. Resource utilization can be increased by the virtualization but for the proper utilization of the resource an efficient load balancing approach is required that take the decision according to the situation.

Resource utilization of host is sum of utilizations of all the virtual machines running on the host and it varies dynamically. Two dynamic threshold based approaches for load balancing are proposed and for the ease of reference, the proposed algorithms are denoted as STAT-VAR and DYN-AVG.

### *A. HOST UTILIZATION*

Host CPU, Memory and Bandwidth utilizations can be calculated as in (1), (2) and (3) respectively. Host utilization can be expressed as sum of utilization of CPU, Memory and Bandwidth as in (4). Equation (5) gives average resource utilization of a host.

$$Host\_CPU\_Utilization = U_{cpu} = \frac{MIPS\_Allocated}{Total\_MIPS} \tag{1}$$

$$Host\_Mem\_Utilization = U_{mem} = \frac{Mem\_Allocated}{Total\_Memory} \tag{2}$$

$$Host\_BW\_\ Utilization{=}U_{bw} = \frac{BW\_Allocated}{Total\_BW} \tag{3}$$

$$Host\_Utilization = Ut = U_{cpu} + U_{mem} + U_{bw} \tag{4}$$

$$\textit{Average\_Utilization} = \frac{1}{n} \sum_{i=0}^{n} Ut \qquad (5)$$

where n is the number of Hosts in Datacenter

### *B. STAT-VAR*

In STAT-VAR initially Upper threshold is initialized to 95% and at regular intervals utilization of all the PMs are calculated. If the resource utilization of a machine goes beyond the upper utilization threshold then some VMs are migrated to reduce load on overloaded PM and if none of the PM is overloaded then Upper threshold is reduced by 10%.

Lower threshold plays a significant role in the load balancing approach. The number of active servers are increased if the lower threshold has minimum value and number of migrations are more if lower threshold takes high value. So we have considered a suitable static value for the lower threshold. When the resource utilization of a physical machine is below the lower threshold, then all VM running on that host are migrated for server consolidation. Algorithm 1 shows steps performed by STAT-VAR approach for Load Balancing.

**Algorithm 1**

```
i.   Initialize Upper Utilization Threshold (UT) and
         Lower Utilization Threshold (LT)
ii.  Repeat following steps (iii-v) periodically

iii. For each Host in the Datacenter
         {
         Calculate Host_Utilization
         If Host_Utilization < LT then
             Initiate Server Consolidation
          Else
             If Host_Utilization > UT then
              Initiate Load Balancing
          }

iv.  If none of the Host has utilization above UT
         Reduce utilization by 10%
         Goto step (ii).
```

### *C. DYN-AVG*

In DYN-AVG Lower Threshold is set to 15% and if CPU utilization of host is greater than average CPU utilization , it is considered as overloaded and VMs from overloaded are migrated to suitable PMs to reduce load. Further if the resource utilization of a PM is below the lower threshold, then all the VMs running in that PM are migrated for server consolidation.

The average CPU utilization is equal to the sum of CPU utilization divided by the total number of hosts. The working of DYN-AVG is explained in Algorithm 2.

**Algorithm 2**

```
i.   Initialize Lower Utilization Threshold (LT)
ii.  Calculate Average_Utilization

iii. UT = Average_Utilization

iv.  For each Host in the Datacenter

         a. If Host_Utilization > UT then
            Initiate Load Balancing
         b. If Host_Utilization <  LT then
            Initiate Server Consolidation

v        Goto step (ii)
```

### *D. VIRTUAL MACHINE SELECTION AND PLACEMENT*

When host is overloaded some virtual machines are to be migrated to suitable hosts and this migration affects the system performance. The two parameters that are to be considered for VM migrations are Down time and the total migration time. if small virtual machine is chosen for migration then the numbers migrations are more to reduce load on host and VM down time might increase if large virtual machine chosen. In the proposed algorithms, Minimization of Migrations (MM) policy [5] is used. It selects least number of VMs for migration to bring the resource utilization below upper utilization threshold. The extended version of Best Fit Decreasing algorithm [5] is applied for VM placement. It arranges all VMs in decreasing order of their present CPU utilizations. Further it allocates each VM to a machine that provides the minimum increase of power consumption due to this allocation.

## IV. SIMULATION SETUP AND PERFORMANCE EVALUATION

The simulation is carried out using CloudSim [16] simulator. CloudSim is a cloud simulation toolkit used to simulate various components of cloud system. Power models for resources of data center , different VM allocation and VM selection policies are supported. It also helps in generating various types of workloads. The comparative evaluation of the proposed algorithms (STAT-VAR, DYN-AVG) is performed with energy efficient dynamic threshold Load Balancing approach [15], which is referred as EE-DTLB.

A Data Center with 50 heterogeneous servers and VMs is simulated and each node is modeled as dual core with performance 2000, 2500 Million Instructions Per Second (MIPS), 4 GB of RAM, 100 Mbps network bandwidth and 1 GB of memory. The user's requests are provisioned with 50 heterogeneous VMs. Each VM simulated having one CPU core with 500, 1000, 2000 or 2500 MIPS, 1 GB of RAM, 100 Mbps bandwidth and 1 GB memory. Energy and Power consumption metrics are chosen to evaluate the proposed algorithms. For the experimental purpose, the number of heterogeneous tasks varies from 100 to 1000.

According to the model of power consumption, at zero percentage of utilization host consumes 175 W of power and when it is utilized to highest it consumes 250 W [16]. The proposed policies simulated by changing the values of Lower and Upper utilization thresholds to determine the value of threshold that has high influence on energy consumption of the host. The comparison analysis of energy consumption and various values of utilization threshold is done. The results show that energy consumption is highly influenced by lower utilization threshold than higher utilization threshold. Higher value of lower utilization threshold eliminates low utilization of resources and leads to high energy saving. But it might result in increase of number of migrations.

All experiment conducted for two different scenarios. In first case lower threshold varied from 10 to 30 and upper threshold initialized to 95% in case of STAT-VAR and in case of DYN-AVG and EE-DTLB varied dynamically based on CPU Utilization. In second case . initial Upper utilization threshold is set to 95% in the case of STAT-VAR , whereas in DYN-AVG and EE-DTLB varied dynamically based on CPU utilization and Lower Threshold is set to 15% in all three algorithms and Number of tasks submitted varied in number from 100 to 1000.

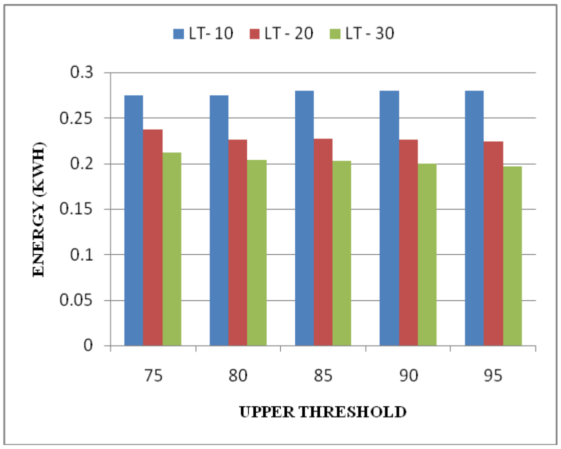

*Figure 1. Upper Utilization Threshold VS Energy*

First experiment conducted to study performance of STAT-VAR algorithm with respect to Energy consumption when Lower Utilization Threshold varied from 10 – 30. It has been observed that energy consumption in the datacenter is reduced for highest value of Lower threshold as number of host shutdowns increase which is indicated in Figure 1.

Second experiment conducted to analyze Energy consumptions of the proposed algorithms STAT-VAR and DYN-AVG and results compared with EE-DTLB [14] scheduling algorithm. Figure 2. shows energy consumptions in the data center to execute tasks varied from 100 to 1000. Graphs clearly indicate that energy consumption significantly improved with using our algorithms when compared to EE-DTLB.

In next experiment Lower utilization Threshold varied from 10 to 30 and Energy consumptions of the proposed approaches STAT-VAR and DYN-AVG is determined and compared with that of EE-DTLB and the results are shown in Figure 3. Results shows that the proposed approaches exhibit improvement when compared with EE-DTLB algorithm.

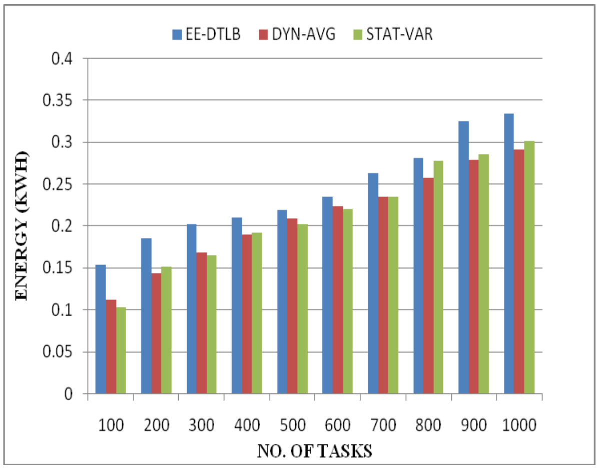

*Figure 2. No. of Tasks VS Energy*

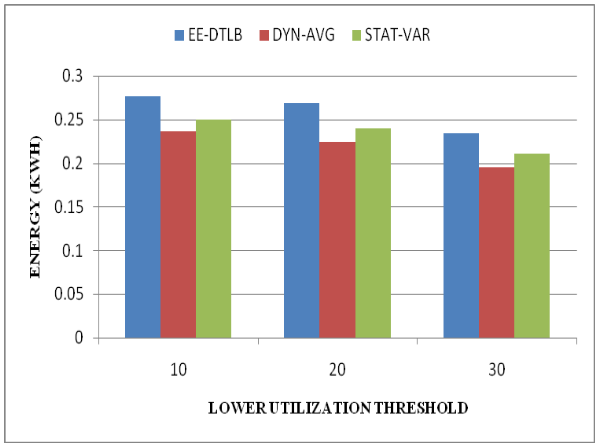

*Figure 3. Lower Utilization Threshold VS Energy*

## V. CONCLUSION

Load management is very important mechanism in cloud for efficient usage of resources in Data Center. Resource requirement and allocation of resources is done dynamically in cloud data center. Hence static threshold values are not suitable in cloud data center.

We have proposed two different dynamic threshold based approaches for Load Balancing and Virtual Machine

Consolidation. The results of simulation shows that the proposed policies exhibit good performance when compared to the existing policies with regard to Energy.

## REFERENCES

[1]. Datacenter Dynamics 2012, *Global Census*.

[2] H. Jin et al., "*Live virtual machine migration with adaptive memory compression*", Proceeding of the IEEE international conference on cluster computing, pp. 1-10, 2012.

[2]. H. Jin et al., "*Live migration of virtual machine based on full system trace and replay*", Proceeding of the 18th ACM 2009.

[3]. R. Nathuji and K. Schwan, "*Virtual power: Coordinated power management in virtualized enterprise systems*", Proceeding of the ACM SIGOPS Symp. on Op.Sys. Principles, ACM press, pp.265-278, Dec. 2007.

[4]. A. Beloglazov, J. Abawajy, and R. Buyya, "*Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing*", Future Generation Computer Systems, pp.755- 768, May. 2012.

[5]. A. Beloglazov and R.Buyya. "*Adaptive threshold-based approach for energy-efficient consolidation of virtual machines in cloud data centers*", Proceedings of the 8th international workshop on middleware for grids, clouds and e-science ACM, pp. 4, 2010.

[6]. A. Beloglazov and R. Buyya, "*Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers*", Concurrency and Computation: Practice and Experience(CCPE), Wiley Press, New York, USA, vol. 24, no. 13, pp. 1397-1420, Sep. 2012

[7]. Amir Varasteh and Maziar Goudarzi, "*Server Consolidation Techniques in Virtualized Data Centers: A Survey*", IEEE Systems, VOL. 11, NO. 2, June 2017.

[8]. C.C. Lin, P.Liu and J.J Wu, "*Energy-efficient Virtual Machine provision Algorithms for cloud systems*", Utility and Cloud Computing (UCC), 2011 Fourth IEEE International Conference, 2011,81-88.

[10]. Shingo Takeda and Toshinori Takemura, "*A Rank based VM Consolidation Method for Power Saving in Datacenters*", IPSJ Online Transactions 3(2):88-96, January 2010.

[11]. Liting Hu, Hai Jin, Xianjie Xiong and Haikun Liu, "*Magnet: A novel scheduling policy for power reduction in cluster with virtual machines*", 2008 IEEE International Conference on Cluster Computing, 13-22.

[12]. Geetha Megharaj, Mohan G. Kabadi, "*Server Consolidation through Virtual Machine Task Migration to achieve Green Cloud*", International Journal of Computer Science and Information Security (IJCSIS), Vol. 16, No. 3, March 2018.

[13]. A Jain et al., "*A Threshold Band Based Model for Automatic Load Balancing in Cloud Environment*", in proc. of IEEE International Conference on Cloud Computing in Emerging Markets, pp 1-7, 2013.

[14]. J. M. Galloway, K. L. Smith, and S. S. Vrbsky, "*Power aware load balancing for cloud computing*", Proc. the World Congress on Engineering and Computer Science, 2011, 19-21.

[15]. Shivani Gupta, Damodar Tiwari, Shailendra Singh, "*Energy Efficient Dynamic Threshold Based Load Balancing Technique in Cloud Computing Environment*", International Journal of Computer Science and Information Technologies, Vol. 6 (2) , 2015, 1023-1026

[16]. R. N. Calheiros, R. Buyya, and A. Beloglazov, *"CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms*", Software: Practice and Experience, Wiley Press, pp.23-50, Jan. 2011.

## Authors Profile

*Mrs. Geetha Megharaj* pursed Bachelor Engineering from Karnataka University of Dharwad, 1991, Master of Science from BITS, Pilani in year 2002 and Master of Technology from Visvesvaraya Technological University, Belagavi. She is currently pursuing Ph.D. and currently working as Associate Professor in Department of Computer Science and Engineering, Sri Krishna Institute of Technology, Bangalore. She is a life member of ISTE, Member of Institution of Engineers since 2006. Her main research work focuses on Cloud Computing, Distributed Computing, Parallel Computing, Machine Learning and Data Science. She has 22 years of teaching experience and 4 years of Research Experience.

*Dr. K. G. Mohan* received Bachelor Degree in Electrical Engineering from University of Mysore in 1984 and Master of Technology in Power and Energy System from KREC (Mangalore University) during 1995. Received Ph.D with the specialization of Computer Architecture from Anna University in the year 2007. He has published and presented more than 30 research papers in reputed International Journals and Conferences. His area of research includes Low Power Architecture design, Wireless Sensor Networks, IoT, Cloud Computing, Network Security and Cryptography. He has 30 years of teaching experience and 4 years of Research Experience.

.