wjert, 2018, Vol. 4, Issue 1, 368-375. *Review Article* **ISSN 2454-695X**

World Journal of Engineering Research and Technology
WJERT
www.wjert.org
**SJIF Impact Factor: 4.326**

# A REVIEW ON ENERGY AWARE RESOURCE MANAGEMENT THROUGH DECENTRALIZED AND BALANCED WORKLOAD IN CLOUD COMPUTING

**Dr. S. Karthikeyan[1] and Dr. S. Sundararajan[2]***

[1]Head of Major – IT Security, Department of IT, College of Applied Sciences.

[2]Associate Professor, SNS College of Technology.

**Article Received on 20/11/2017** **Article Revised on 11/12/2017** **Article Accepted on 01/01/2018**

***Corresponding Author**
**Dr. S. Sundararajan**
Associate Professor, SNS College of Technology.

## ABSTRACT

Cloud computing is one of the emerging trends in the computing technologies which change the way the people using IT resources for business and other purposes. The demand for energy efficient resource management techniques in cloud computing is increasing dramatically due to its growth in financial, business, healthcare, governance, social and web applications. One of the main concerns in this technology is to reduce the cost of hardware, software, power and maintenance. To fulfil the higher demand, services providers are increasing large scale data centres that consume high volume of electric power which makes the negative impact in nature. In this paper, we present the comprehensive review of energy aware techniques available for data centres which makes optimum allocation of resources and selection algorithms for virtual machines in the cloud.

**KEYWORD:** Cloud computing, Virtualization, migration of virtual Machines, Energy-aware systems, Centralization/decentralization.

## INTRODUCTION

The growth of cloud computing and virtualization in the last decade is tremendous and the need arises in all most all the business areas in the world. Cloud services are span in large number in infrastructure, platform and application levels. Cloud computing paradigm utilizes virtualization technology and provides the ability to provision resources on-demand on the pay-per use basis.[1] It eliminates the many issues raised during the maintenance of own computing infrastructure by the industries and other business that increases the risks of security also. This technology replaces tradition computing infrastructure by ensuring energy efficient computing through the following characteristics.

- Optimal and effective utilization of available resources.
- Mobility of Virtual machines leads to location independent.
- Scaling up and down according to the demands.
- Effective resource management using open source tools like Hadoop.
- Elimination of redundancy in the data storage.

But in enterprise level, scalable cloud physical computing infrastructure consumes huge amounts of electrical power and emits heats, which increases operational costs and carbon footprints to the environment. Hence there is a demand to introduce or implement energy conserving techniques on the cloud which reduces both cost and carbon footprint. Cloud service providers built large data centres with heavy volume of resource to ensure the availability of resources for services at anytime and anywhere. In those data centres, workload getsessentially shared between a number of provisioned virtualmachine (VM) instances. But still the majorityof cloud data centres utilize only a fraction oftheir available resources, which makes a considerable amount of power consumption is lost due to both over-provisioned and underutilized idle resources.[2] Dynamic VM consolidation helps to continuously strives for reducing the energy consumption of the data centre by encapsulating the running VM instances to as few physical machines (PMs) as possible.

General energy saving approaches in the cloud infrastructure covers energy conservation in both network and computing nodes.[3] One of the important features of cloud computing influences the effective energy management is load balancing among computing nodes in the data center. This module tries to distribute the workload evenly over the computing nodes of the system.

**LITERATURE REVIEW**

An approach to a distributed and decentralized architecture of the resource management system for Cloud data centres proposed comprising heterogeneous physical nodes by Beloglazov et al.[4] Each node has a multi-core CPU, considerable amount of RAM and network bandwidth. The system architecture is built with major elements like dispatcher, global and local managers. The local managers are reside in each physical node of the provisioned network as a part of a Virtual Machine Monitor (VMM) and are responsible for observing current utilization of the node's resources and its thermal state. They also report information about the utilization of resources and VMs chosen to migrate to global managers. Each global manager is attached with some set of nodes and processes data obtained from their local managers. A three stage allocation policy is introduced for VM placement. In the first stage utilization of resources are ensured by reallocating VMs which minimize the number of physical machines. In the second stage, communicating VMs are placed in a way minimizing the overhead of data transfer over network and third stage optimizes cooling system operation which reduces a significant amount of energy. The nodes' thermal state are monitored continuously using sensors gives an opportunity to recognize overheating nodes and reallocate workload that allow the natural cooling.

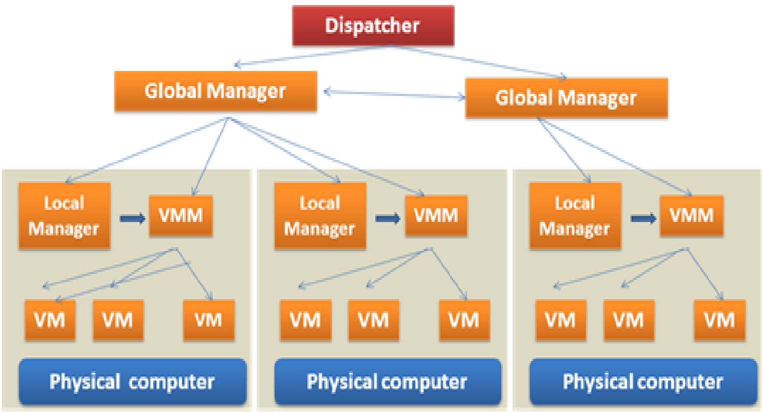


**Fig. 1: System Architecture.**

Pantazoglou et al[6] introduced highly scalable hypercube overlay network that contains physical nodes of the data centre, which used to host the VM instances effectively in self-organized manner. Each node in the network is allowed to act autonomously to manage its workload through VM instances. This topology has several advantages over the other structural models. The numbers of connections that need to be maintained between physical machines are reduced and number of messages exchanged is proportional to the number of open connections exists between nodes. This structure can easily recover from frequent node departures, which result from switching off underutilized nodes upon VM consolidation.

Mastroianni et al.[8] proposed ecoCloud in which initial placement and migration of VM instances are driven by probabilistic process based on the computing and memory utilization. In ecoCloud, VMs are consolidated using two types of probabilistic procedures, for the assignment and the migration of VM. When user request is made, it is transmitted to the data centre manager, which selects an appropriate VM and assigns the VM to one of the available servers during the assignment process.

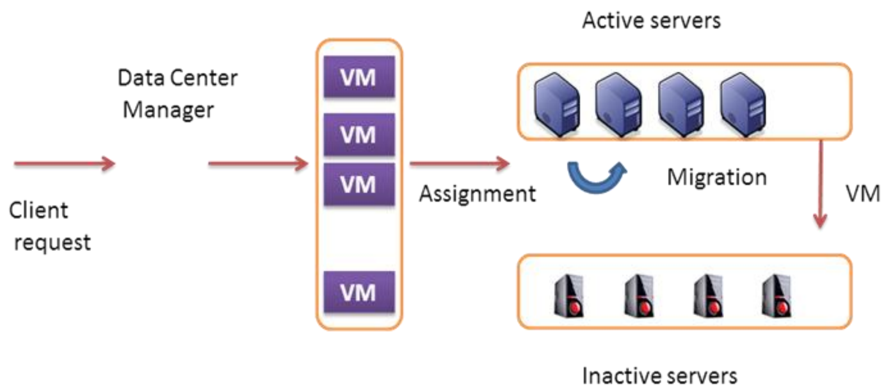


**Fig. 2: Assignment and Migration of VM in Data centre.**

The workload is continuously monitored and a migration of a VM to another server is requested with high probability either when server resources is under-utilized or over-utilized.

Three metrics related to power, performance and network-traffic proposed by Claudio etal[5] to improve the energy efficiency of the communication systems in the cloud network. A higher degree of connectivity increases network capacity that makes topology fault tolerant. Network Power Usage Effectiveness (NPUE) is a metric calculated using power consumption of individual equipment against total consumption. A metric, Average Server Degree Connectivity (ASDC) is used to assess the degree by considering number of servers and number of links that connect servers with others in the network. Management and Monitoring Traffic Ratio (MMTR) is introduced that helps to unveil traffic overhead for network management and can be computed on internal and external traffic. All these metrics are need to be considered for building the cloud computing infrastructure for making use electric energy efficiently.

One or more physical controller nodes in a typical data centre manage overall cloud operating system and number of compute nodes. Compute nodes of the data centre can be organized in a hypercube structure, so that each one of them is directly connected to at most n neighbours, while the maximum number of compute nodes is $N = 2^n$. The data centre is able to initially place VM instances to its compute nodes in a completely decentralized manner, by leveraging the hypercube topology.[7] When the client makes request for provision of a new VM instance, an active compute node is selected by one of the data centre controller nodes, by performing a random walk within the hypercube. The overutilized compute node will start a partial VM migration process in order to shift part of its current workload to one or more nodes in the hypercube, whose current status is ok and idle.

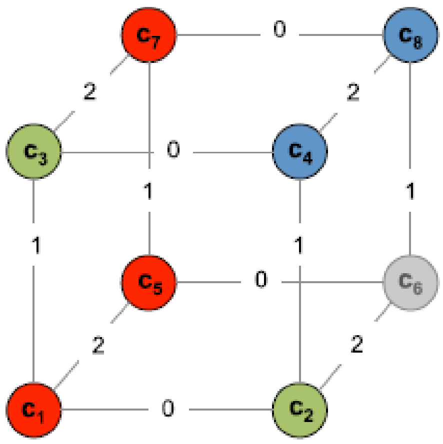


**Fig. 3: Three dimensional hypercube.**

With a wide range of distributed load balancing algorithms, which rely on live VM migration to shift workload between nodes, one can minimize the active resources of the data centre, and avoid overloading of compute nodes for conserving electricity energy. The author proposed algorithms for initial VM placement, partial and full VM migration.

The workflow scheduling under cloud computing is considered as a time-cost optimization problem and Quality of Service (QoS)-based hybrid particle swarm optimization (GHPSO) is used to schedule applications to cloud resources.[14] In this approach, constraints between workflow tasks can be modelled as a directed acyclic graph. The solution has three parts: first, initialize the particles, second, use the crossover and mutation operation to improve PSO, so that it can be applied to this discrete problem, third, make the climbing operation after the end of each iteration to improve the local search ability.

By using Particle Swarm optimization [PSO], dynamic load balance can be applied with decentralized load balancer.[9] PSO is a swarm based heuristic optimization algorithm used to identify the optimal path of total solution space. When the load is applied on specific VM, it moves along all the virtual machines and identifies the optimal machine to put the load the optimal V.M which has less load and available. PSO algorithm works on the following three stages: 1. initialize population of particles with random position and velocities; 2. Evaluate fitness function value of each particle in the space; 3. Compare current particle's fitness function value with other particles and find pbest value.

Improved Particle Swarm optimization approach proposed by Jun et al[13] considers success rate and reliability for calculating fitness function value. The algorithm starts with initialize the position of particles in the solution space. Then the performance of each particle is calculated based on the proposed fitness function and compared with global best value. Then velocity of the best article is changed and moved to a new position.

PSO is combined with Genetic Algorithm (GA) in Hybrid GAPSO algorithm.[10,11] To get a better result, this proposed methodology introduced the enhancement, selection, crossover, and mutation operators. GA operations are performed on the enhanced elite achieved by PSO which compares and improves fitness value of two enhanced elite selected randomly. With crossover operator, an arbitrary contiguous region can be searched, provided there is enough diversity among feasible parent solutions. Mutation is a another variable dependent random mutation where a solution is created near the parent solution space with uniform probability distribution.

Awad et al[12] proposed a mathematical model using Load Balancing Mutation a particle swarm optimization (LBMPSO) based schedule and allocation that takes into account of reliability, execution time, transmission time, make span, round trip time, transmission cost and load balancing between tasks and virtual machine. Three mathematical models proposed for task scheduling that contains objective function with some constraints. Expected execution time for task is calculated in first object function and expected transmission is evaluated using second function. Round trip is the total time required for both sending and receiving which would be calculated in the third object function. Based on round trip time, the task is allocated to VM. PSO is used to balance the load between the VMs by moving the task based on pbestvalue.

## CONCLUSION

This paper presented the various performance issues encountered by the cloud computing technology and how to address those issues using available methodologies. Green computing is the essential one today to save our plant and energy conservation is one development among the initiatives. Various energy aware resource management techniques have been reviewed for making better workflow management, task scheduling, load balancing and cloud infrastructure development. After study, it is seen that optimization of network parameters and infrastructure parameters gives better result in energy conservation. It is expected more hybrid optimization techniques in this series will come in future for addressing the issues faced by cloud service providers for better cost management.

**REFERENCES**

1. L. A. Barroso and U. H€olzle: The case for energy-proportional computing. IEEE Computing, 2007; 40(12): 33–37.
2. L. Wang and S. U. Khan: Review of performance metricsfor green data centres: a taxonomy study. The Journal of Supercomputing, 2013; 63(3): 639–656.
3. Dang Minh Quan, Robert Basmadjian Hermann De Meer: Energy efficient resource allocation strategy for cloud data centres. In Proc. of the 26th Int'l Symposium on Computer and Information Sciences (ISCIS), Springer-Verlag, 2011; 133-142,
4. Anton Beloglazov, Rajkumar Buyya: "Energy Efficient Resource Management in Virtualized Cloud Data Centres. 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, 2010; (7).
5. Claudio Fiandrino, Dzmitry Kliazovich, Pascal Bouvry Albert Y. Zomaya: Performance and Energy Efficiency Metrics for Communication Systems of Cloud Computing Data Centres. IEEE 98th International Conference on Cloud computing, 2015.
6. Shuang Chen, Yanzhi Wang, Massoud Pedram: Resource Allocation Optimization in a Data Centre with Energy Storage Devices. 40th Annual Conference of IEEE, IECON, 2014.
7. Michael Pantazoglou, Gavriil Tzortzakis, and Alex Delis: Decentralized and Energy-Efficient Workload Management in Enterprise Clouds. IEEE transactions on cloud computing, 2016; 4(2).
8. C. Mastroianni, M. Meo, and G. Papuzzo: "Probabilistic consolidation of virtual machines in self-organizing cloud data centres. IEEE Trans. Cloud Computing, 2013; 1(2): 215–228.
9. Ginni Bansal, Amanpreet Kaur: "Enhancement of Dynamic Load Balancing Using Particle Swarm Algorithm in Cloud Environment. IOSR Journal of Computer Engineering, 2016; 18(4): 06-10.

10. M. Sridhar and G. Rama Mohan Babuf: Hybrid Particle Swarm Optimization scheduling for Cloud Computing. IEEE International Advance Computing Conference, 2015; 15: 1196-1200.
11. M. Sridhar: “Hybrid Genetic Swarm Scheduling for Cloud Computing. Global Journal of Computer Science and Technology: B Cloud and Distributed, 2015; 15(3).
12. A.I. Awad, N.A. El-Hefnawy, H. M. Abdel kader: Enhanced Particle Swarm Optimization for Task Scheduling in Cloud Computing Environments. International Conference on Communications, management, and Information technology, 2015.
13. R. JeminaPriyadarsini, L. Arockiam: “An Improved Particle Swarm Optimization Algorithm for Meta Task Scheduling In Cloud Environment. International Journal of Computer Science Trends and Technology, 2015; 3(4): 108-112.
14. Jun Xue S, Wu W: Scheduling Workflow in Cloud Computing Based on Hybrid Particle Swarm Algorithm. TELKOMNIKA Indonesian Journal of Electrical Engineering, 2012; 10(7): 1560-1566.