A Survey on Energy Aware Resource Allocation Techniques in Cloud

Pavithra.B.[1], Ranjana.R.[2]
PG Student, Sri Sairam Engineering College, Chennai[1]
Assistant Professor, Department of IT, Sri Sairam Engineering College, Chennai[2]

Abstract — **A candid observation of cloud computing is that it is an internet-based computing, that consists of enormous groups of remotely located servers which are networked to allow the centralized storage and online access to computer services and resources. It has emerged as the prominent driver for distributed and shared computing. It is embraced by researchers, practitioners and service providers across all industries around the globe. Clouds mainly focus to maximize the effectiveness of shared resources which are not only pooled and shared by multiple tenants but can be dynamically reallocated as per demand. The Resource Allocation Strategy, coupled with energy aware data centers is all about amalgamating several cloud service provider activities for allocating scarce resources efficiently within the limit of cloud environment so as to meet the increasing demands of the cloud users and with the goal to minimize the energy consumption in large cloud data centers. This paper gives an overview of a comparative study on the various existing resource scheduling techniques in cloud computing environment.**

Keywords:* *Cloud computing, Resource Allocation Strategy (RAS), scheduling, energy efficiency.

I. INTRODUCTION

The Cloud Paradigm

Cloud Computing can be simply defined as computing in remote location or location independent with shared and dynamic resource availability on demand. The paramount motive behind more organizations moving to cloud is the mitigation in cost and dynamic provisioning of resources. It is a model for broad network access for enabling ubiquitous, convenient approach to a shared pool of computing resources. Cloud computing is an attractive computing model since it allows for the provision of resources on-demand.

In the cloud computing domain, the allocation and reallocation of resources dynamically is the prime focus for accommodating unpredictable demands and, eventually, contribute to high return on investment. Hence, Cloud Computing is making our business application more mobile and collaborative. The consumption of energy associated with the resources allocation should be taken into account. Resource allocation is the key technology of cloud computing domain, which utilizes the computing resources like bandwidth, energy, delay and so on in the network to facilitate the execution of cumbersome tasks that require large-scale computation.

A Resource Allocation Strategy (RAS) in Cloud Computing can be understood as any mechanism that aims to assure the application's requirements are attended to precisely by the provider's infrastructure.

Cloud providers offer these computing resources as measured services for their clients in a pay-as-you-go fashion. Cloud clients, also called as tenants, request the amount of resources needed to perform certain jobs, to the cloud providers. Upon receiving a client or tenant request, the cloud provider, with the help of virtualization, creates several virtual machine (VMs) on a physical machine (PM) or server and allocates the requested resources to it and thereby reduces the amount of hardware and execution time. The objective of this paper is to focus on various existing resource allocation techniques in cloud computing environment and thereby providing a comparative study.

II. Significance of Resource Allocation

Resource allocation is a mechanism that has been implemented in many computing areas, such as operating systems, grid computing, and datacenter management. Resource allocation involves scheduling of activities and assigning the available resources in an economic way and applies optimal algorithms to efficiently allocate physical and/or virtual resources to developers' applications, thus minimizing the operational cost of the cloud environment. The hardware and software resources are allocated to the cloud applications on-demand basis. In case of scalable computing, virtual machines (VMs) are rented, which isolates the physical hardware to create specific dedicated resources.

The order and time of allocation of resources are also considered as an input for an optimal RAS. It requires the type and amount of resources needed by each application in order to complete a user job. From the cloud user's perspective, for optimal RAS,

ISSN: 2278 – 7798

the application requirement and Service Level Agreement (SLA) are major inputs to RAS.

III. Advantages of Resource Allocation in Cloud

1) The vital benefit of resource allocation is that the users neither need to install software nor hardware to access develop and host the applications over the internet and scale resources based on demand.

2) The location and medium are not restricted. Our applications and data can be easily accessed anywhere in the world, on any system, which corresponds to on-demand self service and ubiquitous network access.

3) The user does not need to shell out excess money on hardware and software systems, as these are available as a service in the Infrastructure-as-a-Service (IaaS) cloud model.

4) Cloud providers, as well as multiple tenants can share their resources over the internet during resource scarcity, which forms part of the resource pooling characteristic of cloud computing domain.

IV. Limitations of Resource Allocation in Cloud: Some Open Issues

1) Since users do not hold ownership over the resources but only rent resources from remote servers for their purpose, they do not have control over their resources. Hence users or clients are popularly called tenants and not owners.

2) Migration problem occurs, when the users wants to switch to some other cloud provider for the better storage of their data. It is not easy to transfer enormous amount of data from one provider to the other.

3) In public cloud domain, the clients' data are prone to hacking or phishing attacks. Since the servers on cloud are interlinked, it is easy for malware to spread. Hence security issues in cloud are the major limitations to resource allocation strategy.

4) Peripheral devices like printers or scanners might not work well with cloud. Many of them require software to be installed locally and require constant internet connection to use and access devices and resources, even in transit.

5) More and deeper enlightenment is required for allocating and properly managing resources in cloud, since all knowledge about the working of the cloud mainly depends upon the cloud service provider (CSP).

V. The Resource Strategy

The main intention of resource allocation strategy is to maximize the profits of both the customers or tenants and the cloud service providers (CSP) in a large datacenter by balancing the demand and supply in the market.

The amount of energy consumed, cost incurred to provide services over the cloud, amount of execution time, are the major causes of concern in resource allocation strategy (RAS) and improvising the optimal scheduling of tasks helps in minimizing these parameters. The aim of any scheduling algorithm is to meet user demand in an economic way and the resource strategy takes into account the user requirements to properly allocate resources and thereby avoid resource overload.

VI. Analysis of Existing Resource Allocation Techniques in Cloud

S.No	Paper Title	Techniques Used	Metrics Considered	Advantage	Disadvantage
1	Energy efficient scheduling of virtual machines in cloud with deadline constraint. [1]	Energy efficient scheduling algorithm, EEVS is used and can support DVFS well.	Number of virtual machines and Performance – Power ratio.	Reduces the total energy consumed by the cloud. Higher optimal performance power ratio.	It does not suit for I/O-intensive or network-intensive VMs. The execution time and power consumption are ignored.
2	Real-Time Tasks Oriented Energy-Aware Scheduling in Virtualized Clouds. [2]	Rolling-horizon scheduling called Energy-Aware Rolling-Horizon scheduling algorithm or EARH is used.	Task count and Task arrival rate, Task Deadlines.	Virtualization technique increases resource utilization and reduces energy consumption.	The maximum amount of CPU cycles assigned to a VM that runs a task must be updated dynamically.

ISSN: 2278 – 7798

3	Optimized task scheduling and resource allocation on cloud computing environment using Improved Differential Evolution Algorithm (IDEA). [3]	IDEA Combines Taguchi method and DEA.	Cost versus Time.	High effectiveness and easy optimization.	The processing time of each subtask is resource dependent. Pre-emption is not allowed.
4	An Energy-Aware Fault Tolerant Scheduling Framework for Soft Error Resilient Cloud Computing Systems. [4]	The fault tolerant cloud scheduling framework is composed of two phases: Static scheduling and dynamic scheduling.	Slack, Application Index and Replication factor considered.	CSP to achieve high error coverage and fault tolerance confidence while minimizing global energy costs under user deadline constraints.	Does not guarantee to execute within deadlines. Cannot guarantee high compatibility among more than two VMs on the same machine.
5	More than bin packing: Dynamic resource allocation strategies in cloud data centers. [5]	Static and dynamic allocation. Bin packing heuristics	Time and Active Server Count.	Increases resource utilization. Demand-based placement controllers with a reallocation controller appear to be the most energy-efficient solution.	Reservation-based controllers have more migration and thereby high overload.
6	Efficient Multi-Tenant Virtual Machine Allocation in Cloud Data Centers. [6]	Uses Internet Data Centers (IDCs). LP-MKP Algorithm (Layered Progressive Multiple Knapsack Problem)	Maximum idle resources (Greedy), maximum available resources in the information tree and network diameters of similar tenant requests.	LP-MKP is significantly superior to the Greedy algorithm and better than the heuristic allocation algorithm MinTree and, it guarantees the fairness of resource allocation for similar tenant requests.	Obtaining the optimization goal is a tedious task. Integrating the LP-MKP algorithm into open-source cloud computing platforms, such as Open Stack and Cloud Stack need to be considered.
7	Energy-Efficient Resource Allocation and Provisioning Framework for Cloud Data Centers. [7]	Data clustering (k-Means clustering), Workload prediction (Best Fit Decreasing) and Power management.	Sum of Squared Distances (SSD), Number of Clusters, and Execution time. Average CPU Utilization and Time.	This system is evaluated using real traces from Google cloud cluster. Achieves significant energy savings and high utilization that are very close to the optimal case.	Need to test the framework on other cloud traces too. Must work to improve the workload prediction module in case of overhead evaluation of regular daily trends requests.
8	Resource Allocation Optimization in a Data Center with Energy Storage Devices. [8]	Convex optimization techniques	Relation between the cost function and the maximum Charging/discharging rate, ESD capacity.	ESD management algorithm and the server consolidation have significant effects on reducing the total cost.	Analysis of the power hierarchy in a data center and the incorporation of more complex battery models need to be addressed.
9	Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment [9]	Skewness Algorithm, Server usage and resource allocation status. Hotspot Migration and Green Computing concept.	A set of overloaded resources in server and hot threshold for resource, along with temperature of a hot spot.	Achieves both overload avoidance and proper utilization of servers. Saves the energy using the green computing concept.	The evaluation of resource allocation status is based on the predicted future resource demands of VMs, hence prediction need to be efficient to comply with real time requests.
10	A green energy-efficient scheduling algorithm using the DVFS technique for cloud datacenters. [10]	Green energy-efficient scheduling algorithm, with extension of DVFS method and priority job scheduling	Number of Jobs versus Energy consumption and Execution time.	Satisfies the minimum resource requirement of a job and prevent the excess use of resources and increases resource utilization.	Servers chosen for a job have to satisfy the two proposed inequalities in this model. The system architecture is complicated to implement in real time cloud environment and heterogeneous servers.
11	Quality of Service Based Efficient Resource Allocation in Cloud Computing. [11]	Energy Aware Best Fit Decreasing (EABFD) algorithm.	Number of VM migrations, Percentage of SLA violations and Energy consumption versus Policy.	Energy consumption was reduced significantly and optimization of QoS is done by applying the EABFD with MAD RS Policy.	This model is implemented only in Cloud Sim toolkit. It needs to be extended for real time implementation and in other cloud simulators.
12	Energy Aware Resource Allocation in Cloud Datacenter. [12]	VMs placement and VMs allocation policies using Modified Best Fit Decreasing (MBFD) and Non Power Aware policy.	Number of VMs and Energy consumed.	This proposed solution delivers both reliability and sustainability, contributing to our goals of optimizing energy utilization and reducing carbon emission.	There is more complexity of the migration algorithm that needs to be addressed while implementing.

ISSN: 2278 – 7798

13	Performance and Energy Efficiency Metrics for Communication Systems of Cloud Computing Data Centers.[13]	Proposes three metrics namely power-related metrics, performance-related metrics and network traffic-related metrics.	Communication Network Energy Efficiency(CNEE) and Network Power Usage Effectiveness (PUEE), Inter-Server Communication Latency (ISCL),	Analyzes end-to-end error rates at bit and packet levels to assure network performance and the desired level of QoS and helps detecting hardware faults.	The presented set of metrics needs to be standardized for performing evaluation in operational data centers.
14	Towards Energy-Efficient Cloud Computing: Prediction, Consolidation, and Over commitment. [14]	Workload prediction, VM placement and Workload Consolidation, and Resource Over commitment.	Time versus number of Requests and Saved Energy.	Resource over commitment has great potential for reducing cloud center energy consumption and solves under-utilization issues.	This requires the development of sophisticated resource management techniques that enables to reduce energy. One major problem with over commitment is PM overload, which need to be addressed.
15	Energy-aware Load Balancing and Application Scaling for the Cloud Ecosystem. [15]	Energy-aware Scaling Algorithm with Load Balancing and energy-optimal operation regime	High-cost versus low-cost application scaling. Number of servers versus regime of operation.	Idle and lightly-loaded servers are switched to one of the sleep states to save energy. Attempts to maximize the number of servers operating in this regime.	Need to evaluate the overhead. This requires balancing the computational efficiency and SLA violations.

VII. CONCLUSION AND FUTURE WORK

This paper proposes a comparative analysis of various existing resource scheduling techniques in cloud computing environment, taking into consideration the energy awareness for optimal performance of cloud data centers and achieve automated provisioning of resources. These techniques focus on various parameters such as execution time, number of VMs, energy consumed, CPU utilization, cost, available resources and number of requests. An evaluation shows that dynamic resource allocation with energy aware scheduling is the recent growing demand of cloud providers in maximizing their profit and satisfying more number of users, with less response time, and thereby meeting the Service Level Agreements (SLA). Thus, cloud computing enables organization to reduce total cost of ownership and maximize the return on investment (ROI) on IT infrastructure on computing resource services and data storage. The future research work aims to implement one of the optimal resource allocation algorithm coupled with energy aware scheduling in a real cloud simulator and thereby obtain the experimental results based on the scenario and metrics to be considered.

REFERENCES

[1] Youwei Ding, Xiaolin Qin, Liang Liu, Taochun Wang, “Energy efficient scheduling of virtual machines in cloud with deadline constraint”, Science Direct 2015.

[2] Xiaomin Zhu, Laurence T. Yang, Huangke Chen Ji Wang, Shu Yin and Xiao cheng Liu, “Real-Time Tasks Oriented Energy-Aware Scheduling in Virtualized Clouds”, IEEE 2014.

[3] Jinn-Tsong Tsai Jia-Cen Fang, Jyh-Horng Chou “Optimized task scheduling and resource allocation on cloud computing environment uses Improved Differential Evolution Algorithm (IDEA)”, Science Direct 2014.

[4] Yue Gao Ming Hsieh, Gupta, S.K., Yanzhi Wang “An Energy-Aware Fault Tolerant Scheduling Framework for Soft Error Resilient Cloud Computing Systems”, IEEE 2014.

[5] Youwei Ding, Xiaolin Qin, Liang Liu, Taochun Wang “More than bin packing: Dynamic resource allocation strategies in cloud data centers”, Science Direct 2015.

[6] Jiaxin Li , Dongsheng Li, Yuming Ye, and Xicheng Lu, “Efficient Multi-Tenant Virtual Machine Allocation in Cloud Data Centers”, IEEE 2015.

[7] Mehiar Dabbagh, Bechir Hamdaoui, Mohsen Guizaniy and Ammar Rayes, “ Energy-Efficient Resource Allocation and Provisioning Framework for Cloud Data Centers”, IEEE 2015.

[8] Shuang Chen, Yanzhi Wang, Massoud Pedram, “Resource Allocation Optimization in a Data Center with Energy Storage Devices”, IEEE 2015.

[9] Zhen Xiao, Weijia Song , Qi Chen, “Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment”, IEEE 2013.

[10] Chia Ming Wu, Ruay Shiung Chang, HsinYu Chan, “A green energy-efficient scheduling algorithm using the DVFS technique for cloud datacenters”,ScienceDirect2014.

[11] Riddhi Patel, Hitul Patel, Sanjay Patel, “Quality of Service Based Efficient Resource Allocation in Cloud Computing”, IJTRE 2015.

[12] Manasa H.B, Anirban Basu, “Energy Aware Resource Allocation in Cloud Datacenter”, IJEAT 2013.

[13] ClaudioFiandrino, Dzmitry Kliazovich, Pascal Bouvry Albert Y. Zomaya, “Performance and Energy Efficiency Metrics for Communication Systems of Cloud Computing Data Centers”, IEEE 2015.

[14] Mehiar Dabbagh, Bechir Hamdaoui, Mohsen Guizani, Ammar Rayes, “Towards Energy-Efficient Cloud Computing: Prediction, Consolidation, and Over commitment”, IEEE 2015.

[15] Ashkan Paya and Dan C. Marinescu:, “Energy-aware Load Balancing and Application Scaling for the Cloud Ecosystem”, IEEE 2015.

ISSN: 2278 – 7798

Authors Profile

Pavithra.B. [1], PG student, currently pursuing M.E. CSE (Networks), Sri Sairam Engineering College, affliated to Anna University, Chennai. She has completed B.Tech Information Technology in the year 2013, from Easwari Engineering College,Chennai.She is an IEEE member with research areas of interest in Cloud Computing, Computer Networks, WSN and MANET .

Ranjana.R.[2], Assistant Professor, Department of Information Technology, Sri Sairam Engineering College, Chennai. Has completed B.E in Computer Science and Engineering from Madras University in the year 2000. Completed M.S.(By Research) form Anna University, Chennai. Research areas include Computer Networks, Cloud Computing and Distributed Systems.

ISSN: 2278 – 7798