Harvest: An International Multidisciplinary and Multilingual Research Journal

E-ISSN: 2582-9866 Volume III Issue I January-March 2023

Energy Efficient Strategy for Request and Server Consolidation Schemes in Cloud Environment

[1*]**C. Senthilkumar,** *Assistant Professor, Department of Computer Science and Engineering, Erode Sengunthar Engineering College, Thuduppathi, Perundurai, Erode-638057.*

[2]**Dr. R. Kalaivani,** *Professor, Department of Electronics and Communication Engineering, Erode Sengunthar Engineering College, Thuduppathi, Perundurai, Erode-638057.*

[3]**A. Rajesh,** *Assistant Professor, Department of Computer Science and Engineering, Erode Sengunthar Engineering College, Thuduppathi, Perundurai, Erode-638057.*

Abstract

An important issue of energy efficiency in the cloud environment is to perform more jobs while consuming less amount of power. Virtual machine consolidation remains the most deployed strategy to manage both performance and energy consumption. Most existing energy efficiency techniques save energy against the cost of performance degradation. Consolidation techniques leverage thresholds to detect overloaded and under-loaded hosts that could be vacated to achieve the optimal balance between host utilization and energy consumption. In this research, we propose an energy-efficient strategy (EES) to consolidate virtual machines in the cloud environment with an aim of reducing energy consumption while completing more tasks with the highest throughput. Many power management strategies have been proposed for enterprise servers based on dynamic voltage and frequency scaling (DVFS), but those solutions cannot further reduce the energy consumption of a server when the server processor is already at the lowest DVFS level and the server utilization is still low (e.g., 10 percent or lower). To achieve improved energy efficiency, request batching can be conducted to group received requests into batches and put the processor into sleep between the batches. And it is challenging to perform request batching on a virtualized server because different virtual machines on the same server may have different workload intensities. Hence, putting the shared processor to sleep may severely impact the application performance of all the virtual machines.
Keywords: Energy efficiency; Data Center; Cloud Computing; virtual machine consolidation, Virtual Batching.

Introduction

Cloud computing provides seamless services using virtualization technology over the Internet to serve the Quality of Service (QoS)-driven end users' requirements. There are different models of cloud computing; Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (Saas), etc… In order to provide various services like SaaS, PaaS, and IaaS, cloud server's data centers are kept active, which have large electrical consumption. Due to improper utilization of resources, optimizing the servers' energy consumption becomes a significant challenge for service vendors from environmental and economic perspectives as well. The challenge to provide services with a low energy consumption profile opens up a new dimension of optimized server use for intelligent management of resources (such as CPU/disk/memory), with reduced power consumption through server consolidation. This enables fewer active physical servers to provide the required services without compromising the QoS.

Consolidation techniques aim to consolidate both applications and services in fewer resources. It is one of the beneficial features of virtualized data centers. Consolidation schemes should optimize resource utilization in such a way as to avoid violating the service-level agreements (i.e., SLA) and performance degradation. One way to reduce energy consumption is virtual machine consolidation in which virtual machines are periodically reallocated to minimize the number of active servers, consequently decreasing the whole cloud data center's energy consumption. Virtual machine (VM) consolidation makes use of live migration, which preserves application performance requirements

while delivering services to customers with short downtime. The consolidation problem can be divided into four steps:

- Detection of overloaded servers that may violate the SLA.
- Detection of under-load servers that have to be vacated in order to minimize the number of active hosts.
- Virtual machine selection aims to select the convenient VMs to migrate from overloaded hosts.
- Virtual machine placement policy that selects hosting servers for the appropriately selected virtual machines.

Once both overloaded and under-loaded servers are detected, VM consolidation techniques should try to evacuate some of their hosted virtual machines to meet SLA and minimize the overall data center's energy consumption. The selected virtual machines to be vacated have to be placed in convenient servers in such a way as to respond to the consolidation's objectives. Usually, hosts where CPU utilization exceeds an upper threshold value are considered overloaded hosts, while hosts with utilization less than a lower threshold value will be considered under-loaded. Hosts that keep utilization between upper and lower thresholds are considered in a normal state.

Actually, to achieve workload consolidation hosts are classified into three categories based on their utilization, namely overloaded hosts, under-loaded hosts, and normal hosts. Usually, hosts where CPU utilization exceeds an upper threshold value are considered overloaded hosts, while hosts with utilization less than a lower threshold value will be considered under-loaded. Hosts that keep utilization between upper and lower thresholds are considered in a normal state. To get relevant papers, different digital research papers sources were investigated. In order to determine the components that were pertinent to the literature review, more than 20 journal papers were evaluated by applying various inclusion and exclusion criteria. We frequently used terminology like "energy-efficient strategy (EES) to consolidate virtual machines," "reducing energy consumption and enhancing efficiency in Server Consolidation," and "Server Consolidation schemes," among others. This study looks at 28 peer-reviewed articles. This research will be useful for future researchers.

Literature

This section discusses the previous works done on reducing energy consumption and enhancing efficiency. Most of the existing VM consolidation techniques with consideration of managing energy consumption make use of static or dynamic lower threshold values based on the current utilization of servers to detect under-loaded ones without considering the overall data centers utilization workload. Actually, these techniques may fall in some drawbacks like placing a VM on a target node, which was prospective to be evacuated in near future, which increase the number of active hosts per data center and lead to more VMs migration in the future. This problem has been presented in [1]. The proposed solution takes into consideration the total data center workload for detecting under-loaded servers. However, data center hosts are heterogeneous. They may include high and low-performance servers. High-performance servers produce less heat in comparison with low-performance servers while having more workload and consuming less power.

In recent years, due to the increasing use of cloud computing services and following the welcome of customers to these services, cloud computing service providers have increased the number and volume of data centers greed to consume more energy [2], and this issue is very costly. Heavy operations have followed. The service quality assurance set out in the SLA, which is regulated between customers and providers, is essential for the cloud computing environment; Therefore, cloud service providers tend to strike a balance between energy and efficiency, and in order to reduce operating costs, they must reduce energy consumption to the extent that it does not disrupt or reduce the quality of service.

Gill et al. [3] proposed a taxonomy for sustainable cloud computing and introduced a taxonomy for VM consolidation-based algorithms without focusing on energy efficiency. Mann et al. [4] introduced a survey on VM allocation in cloud data centers from problem modeling and optimization algorithms perspectives. Ahmad et al. [5] conducted a survey on VM migration and server consolidation

framework for cloud data centers, in which the commonalities and differences of investigated VM migration algorithms are highlighted.

In their research Zhen Xiao, Weijia Song, and Qi Chen [6] introduced a dynamic resource allocation that aims to prevent the PM from being overloaded at the same time minimizing the number of active PMs. to accomplish their purpose, they monitored the overall status of the data center. They also implemented a prediction algorithm that captured the future resources usage and based on this prediction make a decision on how to place the VMs based on the data collected. In our approach, there is no intention of implementing a prediction algorithm but we keep monitoring the existing PMs to determine the overloaded PMs.

All the above works considered a tradeoff between energy consumption (host operating, VM migration, network communication) and QoS (VM computing performance). However, the applications require data centers to provide higher QoS of computing power. Melhem et al. [7] proposed an overload host detection algorithm based on the Markov prediction model and a VM placement algorithm based on the physical host that has received the migrated VM. Their work drastically reduced data center SLA violations, allowing data centers to support applications better. However, this method caused the data center's energy consumption to rise sharply, which is contrary to the design goal of the VM consolidation method. Therefore, we propose a new method that aims to increase the energy efficiency of data centers while meeting the computing resource requirements of applications.

Dahsti et al. [8] developed a solution to address the requirements provided by both service providers and users of these technologies. As a result of the research, a one-of-a-kind PaaS service for organizing client errands was developed. In the cloud, excessive energy usage and an energy performance tradeoff may occur if the specifications of the physical machine and the user's expectations are incompatible, resulting in lower provider profitability. Using the PSO, energy efficiency is increased without compromising service quality. The final aim of these strategies was to reallocate the relocated virtual machines inside the whole host.

A new multi-objective strategy, based on double thresholds and the ACO algorithm, was presented by Xiao et al. [9]. It uses two levels of CPU usage as cut-offs in order to determine whether or not the PM is overloaded. When a host becomes overburdened or underutilized, virtual machine consolidation is initiated. In consolidating VMs, the ACO algorithm uses several selection strategies that consider the loads of the PMs to choose VMs from both the overloaded as well as the destination PMs. In addition to minimizing migrations, performance degradation, service level agreement (SLA) violations, and overall energy consumption, the suggested approach effectively uses available computing and storage resources.

Fatima et al. [10] have proposed a new hybrid algorithm, LMOGWO. The suggested method took design cues from grey wolves, modeling itself after the animals' hunting and pack-leading techniques. It also came with a storage archive for secondary options. The top three answers are the alpha wolves who rally the pack to pounce on their victim. Alpha, beta, and delta wolves represent the pack's top three leaders, while omega wolves represent the remaining members who have found a solution. The suggested method determines the wolf step size based on the levy flight. Finally, the suggested method is put through its paces using nine industry-standard benchmark functions.

Al-Moalmi et al. [11] proposed that the overhead caused by the dynamic placement of VMs may be caused by the time that is spent migrating. As a result, they have thought of using a static approach in order to distribute virtual machines. Their primary objective is to unify the container-to-virtual-machine placement and the virtual-machine placement-to-physical-machine into a single optimization challenge. The time needed to resolve the positioning challenges may be drastically cut down by using this method. To cut down on the time and energy spent on VM creation, WOA has been used to optimize this issue.

In this paper [12], they present two energy-aware techniques for VM consolidation. The proposed schemes consider energy consumption by RAM along with CPU. Moreover, the schemes utilize a threshold mechanism in order to keep some resources free to tackle the increased resource demands at run time. They subdivide the resource allocation problem into two components: (a) host selection and

(b) placement. The proposed techniques take into account a PM's capacity and energy consumption while placing VMs on a server. Moreover, consideration of a server's capacity while placing VMs on the host improves resource management which is indicated as an improvement in our energy graphs. Their main contributions are to present a detailed analysis of the selected energy-efficient resource management techniques using cloud environments, two new energy-efficient SLA-aware resource management techniques namely MaxCap and RemCap are proposed to optimize energy and handle SLA violations by balancing the network load, proposed algorithms aim to improve performance in terms of energy efficiency, SLA violations, and address performance degradation due to migrations, and also presents the time and space complexity analysis of the proposed techniques.

In this section, they have reviewed and highlighted the related work on CDCs for implementing energy-efficient data centers. During the last decade, cloud computing solutions have been widely applied for sharing resources and services with better efficiency, speed, availability, reliability, and security. The problem comes when the demand for these cloud services from business enterprises increases. This leads toward expanding the currently existing data centers without looking into the aspect of huge power consumption by the server machines and other facilities which ultimately contribute to generating carbon emissions hazardous to environmental sustainability [13].

Energy consumption becomes a critical concern for large-scale data centers. This Paper [14], explores the energy consumption patterns of infrastructure-as-a-service cloud environments under various synthetic and real application workloads. For each scenario, they investigated the power overhead triggered by different types of virtual machines, the impact of the virtual cluster size on the energy efficiency of hosting infrastructure, and the tradeoff between the performance and energy consumption of MapReduce virtual clusters through typical cloud applications. They have performed two types of experiments. First, carried out a set of component tests to analyze the impact of virtualized workloads on the power consumption of each node. In the second, they focused on three typical MapReduce applications to estimate the effect of increasing the capacity of a virtual cluster on performance 23 and energy consumption. This paper has concluded that, by considering the entire cloud, the evaluations provide valuable insights into the cloud computing potential to save energy.

In [15], developed an energy-efficient heuristic algorithm to minimize energy consumption in the cloud computing environment. To analyze the developed algorithm, they have assumed a centralized cloud is hosted on a data center that is composed of a large number of heterogeneous servers. The energy consumed by a set of the virtual machine for executing the task run on the resource. The energy consumed by the resource is proportional to the processor associated with the resource. The resource allocated to a task must sufficiently provide the resource usage for this task, if the resources are not provided the task is put in a waiting queue. Simulation experiments are conducted to know the performance of a heuristic-based task consolidation algorithms to optimize energy consumption. Concluded that the MaxMaxUtil algorithm is preferred over others to optimize the energy consumption in the cloud computing system

Paper [16], analyzed the impact of VM size and network bandwidth on VM migration time and energy consumption of the source system. To analyze they have considered Kernel-based Virtual Machine (KVM) hypervisor and Virt Manager to perform VM live migration on Ubuntu 14.04 Linux machines in various conditions. The VM was migrated six times between two machine hosts for a fixed value of bandwidth. Then the VM size is increased to 2GB and migration time and power consumption are measured for different bandwidth values. The average migration time and average energy consumption are calculated using the obtained results. The parameters considered here are the VM size, migration time, and energy consumption of the source machine. This paper concluded that live migration can reduce energy consumption and migration time of subsystems by selecting VM with the least memory size for migration and increased network bandwidth.

In [17], presents power consumption evaluation on the effects of live migration of VMs. They have first considered the practical approach to evaluate the power consumption of virtual machines and then estimate the power cost of VM migration both for the original physical server that starts the migration and the destination server that accepts the transfer. In this work, they have considered two aspects that mainly dedicate to the power cost of the server, the processor frequency, and CPU

utilization percentage, and have conducted two experiments. The first one is to verify that server power cost can be represented by CPU usage, specifically, directly proportional to CPU usage. This conclusion drawn is, in the source server, as CPU Utilization increases power impact of live migration falls and the time cost is not affected by the CPU for both source and destination. The second is to get the power consumption of the server in each processor frequency, which also verifies that in a fixed frequency, the power consumption can be represented by CPU utilization percentage. In the end, they evaluated the power consumption caused by live migration. The paper concluded that VM migration is key to realizing VM-based resource reservation and power reduction.

In [18] presents novel techniques, models, algorithms, and, software for distributed dynamic consolidation of Virtual machines in cloud data centers main goal of this work is to improve the utilization of resources and reduce energy consumption under the workload-independent quality of service constraints. To achieve this, they proposed a distributed approach to energy-efficient dynamic VM consolidation and several novel heuristics which lead to a great reduction in energy consumption. The paper also focuses on the heuristics for distributed dynamic VM consolidation under host overload, host under-load, VM selection, and VM placement conditions. This paper brings the idea of switching off the ideal nodes when not n use and allocation of virtual machines to other hosts when it is overloaded which I have implemented in this work. This also gives a brief idea of power and energy models and problems of high power and energy consumption.

In [19], considered possible energy savings of wireless access networks through the development of an integer linear programming (ILP) model based on energy-efficient network management. This paper throws in an idea about the heuristic algorithm that ensures the minimization of network energy consumption in a reasonable amount of time. They developed their own heuristic algorithm that mainly focuses on the minimization of energy consumption in WLANs. The algorithm is composed of two phases: The first phase is the Greedy approach to build up a feasible solution. The second phase is local search (LS) which starts with an initial solution and iteratively moves to the best candidate within the current neighborhood. A comparison of optimization results and computational time of the heuristic approach with those of the ILP model is done. The presenters concluded that heuristic algorithms can be valuable alternatives to exact algorithms due to the possibility of offering good solutions in a reasonable amount time of time. All the above-cited techniques are designed to deliver energy-efficient resource allocation, ultimately minimizing the overall cost of cloud DCs.

Conclusion

The main concern of cloud computing data centers is to reduce energy consumption and consequently reduce operating costs and increase the profitability of such centers. With the expansion of cloud computing applications, the demand for data processing and application storage has grown excessively. Cloud computing services provide customers with a large number of computing resources and storage space, which effectively promotes the further development of other industries. The major focus of the research project is on energy-aware server consolidation approaches and algorithms. The Virtual Batching scheme is improved to manage resources with load balancing mechanisms, to examine optimization mechanisms to manage relative response times, and Resource levels and application needs are incorporated in the allocation process. Virtual Batching dynamically allocates the CPU resource such that all the VMs can have approximately the same performance level relative to their allowed peak values.

Reference

[1] Patel N, Patel H (2017) Energy efficient strategy for placement of virtual machines selected from underloaded servers in compute Cloud. J King Saud Univ Comput Inf Sci. doi: https://doi.org/10.1016/j.jksuci.2017.11.003

[2] Gao Y, Guan H, Qi Z, Song T, Huan F, Liu L. Service level agreement based energy-efficient resource management in cloud data centers. Computers & Electrical Engineering. 2014;40(5):1621-33.

[3] S. S. Gill and R. Buyya, "A taxonomy and future directions for sustainable cloud computing: 360 degree view," ACM Computing Surveys, vol. 51, no. 5, pp. 104:1–104:33, Dec. 2019.
[4] Z. A. Mann, "Allocation of virtual machines in cloud data centers—a ´ survey of problem models and optimization algorithms," Acm Computing Surveys (CSUR), vol. 48, no. 1, p. 11, 2015.
[5] R. W. Ahmad, A. Gani, S. H. A. Hamid, M. Shiraz, A. Yousafzai, and F. Xia, "A survey on virtual machine migration and server consolidation frameworks for cloud data centers," Journal of network and computer applications, vol. 52, pp. 11–25, 2015.
[6] Zhen Xiao, Weijia Song and Qi Chen, "Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment," IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 6, pp. 1107-1117, June 2013.
[7] S.B. Melhem, A. Agarwal, N. Goel, M. Zaman, Markov prediction model for host load detection and vm placement in live migration. IEEE Access. 6 (2018), pp.7190-7205.
[8] Dashti, S.E.; Rahmani, A.M. Dynamic VMs placement for energy efficiency by PSO in cloud computing. J. Exp. Theor. Artif. Intell. 2016, 28, 97–112. [CrossRef]
[9] Xiao, H.; Hu, Z.; Li, K. Multi-objective VM consolidation based on thresholds and ant colony system in cloud computing. IEEE Access 2019, 7, 53441–53453. [CrossRef]
[10] Fatima, A.; Javaid, N.; Anjum Butt, A.; Sultana, T.; Hussain, W.; Bilal, M.; Hashmi, M.; Akbar, M.; Ilahi, M. An enhanced multi-objective gray wolf optimization for virtual machine placement in cloud data centers. Electronics 2019, 8, 218. [CrossRef]
[11] Al-Moalmi, A.; Luo, J.; Salah, A.; Li, K.; Yin, L. A whale optimization system for energy efficient container placement in data centers. Expert Syst. Appl. 2021, 164, 113719. [CrossRef]
[12] B. Gul *et al.*, "CPU and RAM energy-based SLA-aware workload consolidation techniques for clouds," *IEEE Access*, vol. 8, pp. 62990–63003, 2020.
[13] Zhang W, Qi Q, Deng J. Building intelligent transportation cloud data center based on SOA. *Int J Ambient Comput Intel* 2017; 8(2): 1–11.
[14] A. Carpen-Amarie and A. Cecile Orgerie, "Experimental Study on the Energy Consumption in IaaS Cloud Environments," IEEE 6th International Conference on Utility and Cloud Computing, 2013.
[15] "Energy Efficient Heuristic Resource Allocation for Cloud Computing (PDF Download Available)." [Online]. Available: 46 https://www.researchgate.net/publication/260438724_Energy_Efficient_Heurist ic_Resource_Allocation_for_Cloud_Computing. [Accessed: 18-Apr-2016].
[16] I. S. Dhanoa and S. S. Khurmi, "Analyzing energy consumption during VM live migration," in 2015 International Conference on Computing, Communication Automation (ICCCA), 2015, pp. 584–588.
[17] Q. Huang, F. Gao, R. Wang, and Z. Qi, "Power Consumption of Virtual Machine Live Migration in Clouds," in 2011 Third International Conference on Communications and Mobile Computing (CMC), 2011, pp. 122–125.
[18] A. Beloglazov, "Energy-Efficient Management of Virtual Machines in Data Centers for Cloud Computing," PhDThesis, THE UNIVERSITY OF MELBOURNE, 2013. http://hdl.handle.net/11343/38198
[19] J. Lorincz and M. Bogarelli, "Heuristic Approach for Optimized Energy Savings in Wireless Access Networks," FESB-Split, University of Split, Croatia, DEI, Politecnico di Milano, Italy, 2011.