Virtual Machine consolidation policy for power usage management in cloud data centers

Ulysse Rugwiro[1, a], Gu Chunhua [2, b]

[1]School of Information Science and Engineering East China University of Science and Technology

130 Meilong Road, Xu hui District, Shanghai 200237 P. R. China

[2]School of Information Science and Engineering East China University of Science and Technology

130 Meilong Road, Xu hui District, Shanghai 200237 P. R. China

[a]Email:ulysserugwiro@live.com, [b]Email: chgu@ecust.edu.cn

Keywords: Virtual Machine consolidation, Energy efficiency, Cloud computing, virtual machine migration

Abstract. In Today's world, IT technologies are growing day by day so the need of computing and storage are growing with it. The increasing of cloud services demands requires more computing resources to fulfill the end user's requirements. So, energy consumption by cloud computing resources is also increasing day by day and become a key problem in cloud environment. In cloud computing, data centers consume huge amount of energy and also emit carbon dioxide in the environment. For energy optimization, energy efficient resource management is required. In this paper a novel approach is presented to manage the power usage of virtual machines by introducing a dynamic programming algorithm that would be responsible for the selection of the best virtual machines that would cater for the migration from an overloaded physical machine. In order to identify the overloaded physical machines, underloaded physical machines and how many of them have no tasks; the ratio of occupied resources is calculated using MIPS, RAM and Bandwidth. Optimally selection of Virtual Machines and place them on appropriate host lead to minimize energy consumption.

Introduction

The concept of cloud computing and the related environments have seen the introduction of new computing models through the shift of the location of computational infrastructure to the internet. This has also seen the reduction of the costs related with the management of software and hardware resources. It is critical to note that the cloud model utilizes virtualization technology to efficiently consolidate virtual machines into physical machines. This has the effect of enhancing the utilizations of physical machines remotely. However, research has revealed that the average usage of physical severs in numerous cloud data centers has not yet reached optimal levels. There is a need for new approaches which would improve the level of usage of physical servers through the use of the cloud computing technologies. An introduction of a new approach that would consolidate VMs dynamically for there to be an optimal utilization of PMs. A dynamic programming algorithm that would be responsible for the selection of the best VMs that would cater for the migration from an overloaded PM. A consideration of the migration overhead of a virtual machine [1] is also needed.

A VM consolidation approach makes sure that both PMs and their VMs are monitored periodically. This will allow cloud providers to meet the SLAs and deriver good services as well as saving the energy consumed by minimizing the number of active PMs. whenever a PM is considered to be overloaded, its VMs will be reallocated with live migration according to [2], and we will focus on two problems. Firstly, for the selected VM, where to migrate. Choosing the right PM for the VM selected for Migration or the new VM is another challenge that might influence the energy consumed by the system and the quality of VM consolidation as well. Secondly, Deciding which and when a PM should be switch off/on. Consolidating VM and associating it with a dynamic method to switch the power states of node addresses the challenge of the amount of power consumed by removing servers that are in Idle states.it is critical to determine when and which PM should be reactivated to manage the increase demand of resources, or deactivated to save energy. Our purpose is to find the best VM for migration from an overloaded PM by introducing a resource utilization formula. We will be able to get the resource ratio for each PM. We then introduce a new dynamic programming algorithm for picking the best VMs for migration from an overloaded PM. Finally, an algorithm for selecting the best destination for the selected VM that is a candidate for migration based on the obtained results.

Related Work

Researches have been conducted in the area of virtual machine placement in cloud. The main objective was to exploit the available resources while avoiding the risk of having the degradation of VM performance. In this regards, different algorithms with different objectives have been proposed such as reducing the number of running PMs [3] [4] [5] [6].

Konstantinos et al [9]. Viewed consolidation as a bin packing problem with variable bin sizes calling for power-aware dynamic placement of the application. Live migration of the VMs would then be applied to move them to a new host at each time frame. However, they left out the trust model and the fact that user may provide the hints that are not compatible with the cloud infrastructure.

Lee et al [8] also viewed the task as a bin packing problem, basing their work on a linear power model. In their research, they presented two energy-aware task consolidation heuristics that were aimed at maximizing resource utilization.

Sharifi et al [7]. Proposed an energy-aware heterogeneous scheduling algorithm with performance. Objective functions were applied in terms of fitness metric and a set of VMs were placed on a set of PMs with the view of reducing the total power consumption in the data center.

Anton Beloglazov et al [2] introduced an algorithm for detecting the overloaded PM not to forget proves its advantage. They basically allowed system admins to be able to set the service quality goals in terms of Output Text Format (OTF) parameters. In our approach to detect overloaded PMs we use a heuristic algorithm. We will also propose algorithms for picking best VM from an overloaded PM for migration. Furthermore, we will also select the perfect destination for migrated VMs.

In their research Zhen Xiao, Weijia Song and Qi Chen [10] introduced a dynamic resource allocation that aim 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 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 intent of implementing a prediction algorithm but we keep monitoring the existing PMs to determine the overloaded PMs.

The Proposed Method

The proposed method, comparing to the previous ones ([3], [11], [12] and [13]), has the following advantages:

Previous works have been focusing on CPU utilization which has caused them to be unsuitable for real scenario, we introduce the use of memory utilization. By using memory utilization in choosing the under-loaded hosts, we are able to reduce the number of VM migrations, SLA violations and energy consumption. With live migration, the whole VM memory should be moved to a destination host. Using memory and CPU utilization, the proposed method can reduce the number of VM migration, the network traffic, the number of migrations and make sure that the SLA is not violated.

The occupied resource weight ratio is defined as the ratio of the sum of resource weights of all virtual machines over the sum of available resource weights of the running physical machines. Based on this factor we can identify which PM have no task, which PM is overloaded and under loaded. Next, three different factors identified such as if PM have no task then turn it off. If a PM is under-loaded, migrate the VMs to a PM which has low computing task. If a PM is overloaded, migrate one or more VMs from it to other PM or wake up a device on stand-by. PM is treated as overloaded if gross occupied resource weight ratio is more than upper threshold n value. If it is less than lower threshold value m, then that machine is treated as under loaded virtual machine. This way we can optimize the energy consumption by virtual machine. To compare the resources usage, Eq. 1, Eq. 2 and Eq. 3 show how we intent to get the values of each parameter utilization and make decision based on the results and comparing them to the existing approach that have only one parameter (CPU).

$$\text{ramUtilisation} = \frac{\sum_{i=0}^{n} \text{RAMvmi}}{\text{RAMpm}} \quad . \tag{1}$$

Where **ramUtilization = Utilization of RAM (Memory)**

$$\text{mipsUtilisation} = \frac{\sum_{i=0}^{n} \text{MIPSvmi}}{\text{MIPSpm}} \quad . \tag{2}$$

Where **mipsUtilization = Utilization of MIPS (CPU)**

$$\text{bwUtilisation} = \frac{\sum_{i=0}^{n} \text{BWvmi}}{\text{BWpm}} \quad . \tag{3}$$

Where **bwUtilization = Utilization of Bandwidth (Network)**

$$\text{TotalUtilization} = \frac{\left(\frac{\sum_{i=0}^{n} \text{RAM vmi}}{\text{RAMpm}}\right) * \left(\frac{\sum_{i=0}^{n} \text{MIPS vmi}}{\text{MIPSpm}}\right) * \left(\frac{\sum_{i=0}^{n} \text{BWvmi}}{\text{BWpm}}\right)}{3} \quad . \tag{4}$$

Where **TotalUtilization ← (ramUtilization * mipsUtilization * bwUtilization) / 3**

Selecting Best VMs for Migration.

Overloaded PM have impact on the quality of service provided because when the resource capacity is completely utilized, it is likely that the applications cannot have the sufficient amounts of resources [3]. In contemplation of detecting the overloaded and under loaded PM we set threshold for resources in a PM. Whenever the utilization reaches the threshold, the PM is considered as an overloaded. This call for the attention to the PM and some of its VMs should be migrated away. The same principle will be applied, if the utilization of all resources is below the threshold, the PM is considered as under-loaded, this indicates that the PM is underutilized; therefore, its VMs should be migrated and turn it off to save energy. The algorithm is shown in below.

Algorithm 2: Selecting VM for Migration

U[ram] define utilization of Ram, U[mips] define utilization of MIPS and U[bw] define utilization of bandwidth.

```
VMtoMigrate ← NULL
Vmid ← NULL
Max ← 0
If (ramUtilization > mipsUtilization)
    If (ramUtilization > bwUtilization)
        For each VM vi in overloaded PM
            If (Max < u[ram])
                Max ← u[ram]
                vm ← vi
            End If
        End For
        VMtoMigrate ← vm
    End If
End If
Else If (mipsUtilization > bwUtilization)
    For each VM vi in overloaded PM
        If (Max < u[mips])
            Max ← u[mips]
            vm ← vi
        End If
    End For
    VMtoMigrate ← vm
End Else if
Else
    For each VM vi in overloaded PM
```

```
If (Max < u[bw])
        Max ← u[bw]
        vm ← vi
End If
                End For
                VMtoMigrate ← vm
        End Else
```

Selecting the Best PMs.

In this section an algorithm for VM placement will be presented. VM placement have been the hot topic for recent research [14] [15]. When a VM is initially selected for migration, we ensure that the best destination is select based on VM characteristics. For example, to save energy a data centers may consolidate VMs on servers which have the high level of resources utilization. However, this policy might have impact on the QoS violations since VMs don't have access to sufficient resources. Also, to make sure that VMs have access to the sufficient resources a data center might consolidate VMs on servers which have low level of resources utilization and delivered services. We will follow energy efficient policy where a PMs have to be one with a highest level of resource utilization to reduce the running PMs and save the overall energy consumption of the data center. We then consolidate VMs in the PMs which have sufficient resources to make sure that the QoS is delivered. For each host calculate RAM (Memory) utilization, MIPS (CPU) utilization and Bandwidth (Network) utilization. Then calculate total utilization of Physical Machine (PM) based on equation (1). Here, we find the ratio by first multiplying RAM utilization, MIPS utilization and Bandwidth utilization and then divide resultant value with 3 (as we have taken 3 parameters). Now if total utilization of PM is greater than n then respective host will be overloaded and will store host id in overloaded host list

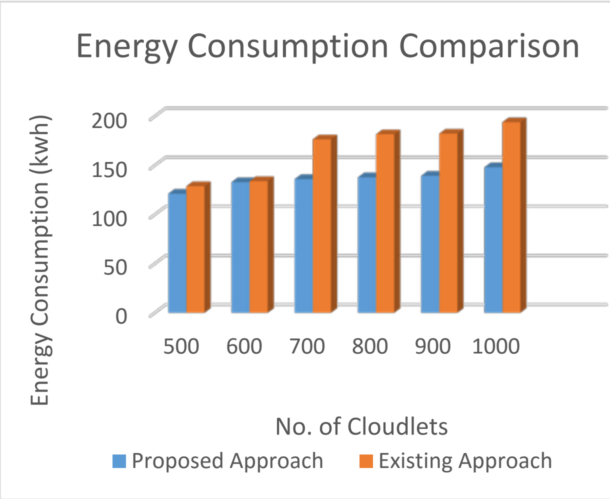
Simulation Results and Analysis

Given simulation results, there are improvements in reducing Energy Consumption. Considering the existing approach which uses one parameter in terms of MIPS. With the proposed approach, we considered three parameters (MIPS, RAM and Bandwidth). Add-on, if we compare both approaches, existing approach with the proposed approach, a clear distinctions of Energy consumption can be identified when increasing No. of cloudlets; Energy consumption is gradually increasing for the existing approach.

Table.1. comparison of energy consumption and SLA violation for the proposed algorithm and the existing algorithm

	MIPS, RAM and BW			**MIPS**		
No. of cloudlet	***Energy Consumption (kwh)***	***Average SLA violation***	***Overall SLA violation***	***MIPS (Energy Consumption) (kwh)***	***Average SLA violation***	***Overall SLA violation***
500	121.17	77.69	58.75	128.75	77.43	58.75
600	132.93	81.01	61.25	133.94	61.25	80.41
700	136.07	63.05	78.7	176.49	71.09	63.09
800	137.87	26.36	65.35	181.63	72.32	64.47
900	139.37	32.24	68.79	182.28	55.18	65.83
1000	148.08	37.7	70.94	193.94	97.45	67.56
	threshold=0.6			threshold=0.9		

Comparing the Energy consumption of both approaches, we can see that the proposed approach have a very low Energy consumption. And as per cloud computing cost calculation, energy consumption is directly proportional to total cost. Hence this approach will also reduce the total cost during execution. Fig.1 is the graphical representation of the energy consumed for the proposed algorithm and the existing algorithm.



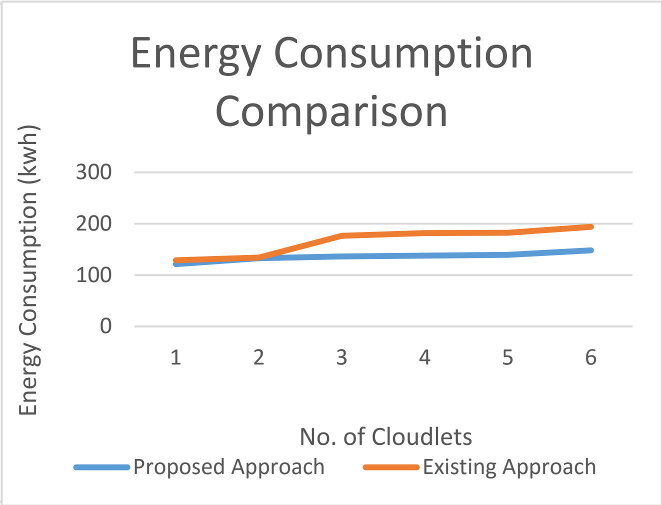


Fig.1. Energy consumption comparison for the proposes approach and the existing approach

Given simulation results showing No. of VMs migrated during execution of cloudlets. If we focus on both scenarios, then in the proposed approach the No. of migrated VMs are quite low compared to existing model.

These simulation results conclude that if No. of migrated VMs are small then it will increase QoS (Quality of Service) for cloudlets. Also increases the system performance by minimizing swapping time and to find suitable Physical Machine. The Table.2. provide the number of Migrated VMs given the number of Cloudlet.

Table.2. comparison of number of VM migrated for the existing approach and the proposed approach

	MIPS, RAM and BW		**MIPS**	
No. of cloudlet	***Energy Consumption (kwh)***	***No. of VM Migrations***	***MIPS (Energy Consumption) (kwh)***	***No. of VM Migrations***
500	121.17	61	128.75	1230
600	132.93	59	133.94	1051
700	136.07	51	176.49	341
800	137.87	44	181.63	246
900	139.37	47	182.28	210
1000	148.08	44	193.94	1691
	threshold=0.6		threshold=0.9	

Data centers have many reason to consolidate VM from one PM to another. The migration can be for maintenance reasons, balancing performance needs, consolidating VMs into fewer PM during the non-peak hours to conserve resources. In the proposed approach we aim to conserver the energy and execute the move as quickly as possible to minimize the impact to end-users and make sure the SLA is not violated.

The Fig.2. illustrate the number migrated VM given the number of cloudlet. Series1 indicate the proposed approach while series 2 indicate the existing approach.

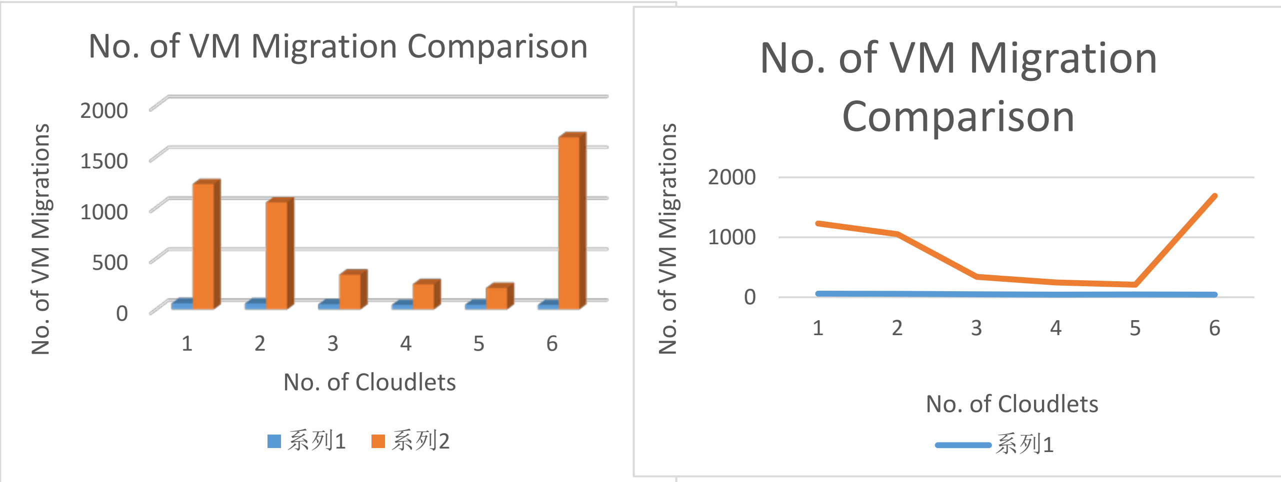


Fig.2. comparison of migrated VM for the existing approach and the proposed approach

Conclusion

In this paper, we have implemented VM migration, VM allocation and PM selection algorithms for resource management in cloud datacenters using three resources, that is; CPU, memory and Bandwidth. This was to address the problem of VM consolidation and energy efficiency in cloud environment. We defined an algorithm to select the VM for migration based on the resources utilization (RAM, CPU, Bandwidth) from an overloaded PM. Additionally we presented an algorithm to allocate the best destination for the intended VMs. The presented simulation results show an improvement in terms of reducing the overall energy consumption for the cloud data center. It is imperative to say that with the right allocation and consolidation of VM, a data center improves its efficiency in energy consumption and at the same time ensures that the utilization of the physical machines is enhanced.

References

[1] Asyabi, Esmail, and Mohsen Sharifi. "A New Approach for Dynamic Virtual Machine Consolidation in Cloud Data Centers." (2015), pp 1 – 6.

[2] Anton Beloglazov, Rajkumar Buyya, "Managing Overloaded PMs for Dynamic Consolidation of Virtual Machines in Cloud Data Centers under Quality of Service Constraints," IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 7, pp. 1366-1379, July 2013.

[3] Anton Beloglazov, Rajkumar Buyya, "Managing Overloaded PMs for Dynamic Consolidation of Virtual Machines in Cloud Data Centers under Quality of Service Constraints," IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 7, pp. 1366-1379, July 2013.

[4] Wenting Wang, Haopeng Chen and Xi Chen, "An Availability-aware Approach to Resource Placement of Dynamic Scaling in Clouds", IEEE Fifth International Conference on Cloud Computing, 2012, Honolulu, USA.

[5] Konstantinos Tsakalozos, Mema Roussopoulos, and Alex Delis, "Hint-based Execution of Workloads in Clouds with Nefeli", ieee transactions on parallel and distributed systems, 2012.

[6] Nicolo Maria Calcavecchia, Ofer Biran, Erez Hadad and Yosef Moatti, "VM Placement Strategies for Cloud Scenarios," cloud, pp.852-859, 2012 IEEE Fifth International Conference on Cloud Computing, 2012.

[7] Asyabi, Esmail, and Mohsen Sharifi. "A New Approach for Dynamic Virtual Machine Consolidation in Cloud Data Centers." (2015), pp 1 – 6.

[8] Lee, S., et al. Validating heuristics for virtual machines consolidation. Microsoft Research. MSR-TR-2011-9, 2011, pp. 1 – 14.

[9] Konstantinos Tsakalozos, Mema Roussopoulos, and Alex Delis, “Hint-based Execution of Workloads in Clouds with Nefeli”, ieee transactions on parallel and distributed systems, 2012.

[10] 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.

[11] 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, May 2011.

[12] A. Verma, P. Ahuja, and A. Neogi, “pMapper: power and migration cost aware application placement in virtualized systems,” Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware, pp. 243-264, Dec. 2008.

[13] M. S. Mozafari, A. Horri, G. Dastghibyfard, and G. hossein Dastghaibyfard, “A Novel Method for Energy Efficiency in Cloud Environment,” in Congress on Electrical, Computer and Information Technology, 2013, no. 1, p. 6.

[14] Wesam Dawoud, Ibrahim Takouna, and Christoph Meinel, “Elastic Virtual Machine for Fine-grained Cloud Resource Provisioning”, Springer 4th International Conference on Computing and Communication Systems, 2012.

[15] Marco Guazzone, Cosimo Anglano, and Massimo Canonico, “Exploiting VM Migration for the Automated Powerand Performance Management of Green Cloud Computing Systems”, springer 1st International Workshop on Energy-Efficient Data Centres (E2DC 2012), Madrid, Spain, May 2012