TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 5212-5216

Power-Aware Virtual Machine Migration for Resource Allocation in Cloud

[1]Mohammad Sulaiman Hyder, [2]Vijeth J, [3]S Sushma, [4]Mohammed Khaled Bawzir, [5]Supreeth S

[1,2,3,4,5]School of Computing and Information Technology, REVA University, Bangalore

Article Info
Volume 83
Page Number: 5212-5216
Publication Issue:
May-June 2020

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *16 May 2020*

Abstract

With the extensive growing demand for cloud computing, virtualization plays an important role to provide services to the end users. Due to the increased use of cloud, managing and running multiple VMs on cloud is becoming a difficult task. Therefore, it is important to solve the problem using efficient technique. The task includes reducing the energy cost which can be achieved by reducing the power consumption. Reducing the power consumption decreases the carbon emission that leads to green cloud computing. Our main objective is to decrease power consumption and reduce SLA violations. This objective is achieved by using CPU stipulation for VM selection, modified local regression for VM migration, adaptive utilization threshold base and non-threshold base algorithm for host selection.

Keywords: *Quality of service, Virtual Machine, Migration, Resource Allocation, Cloud*

1. Introduction

Cloud computing is an impressive technology which is replacing the traditional server based technology. Cloud comprises pool of resources and are allocated based on demand. Resources can be shared over the internet giving access to large number of users. It is used for various reasons which includes network, storage and memory. Cloud computing provides three service models as follows. Software as a service (SAAS), Platform as a service (PAAS) and Infrastructure as a service (IAAS). It also provides deployment models categorized as public, private, community and hybrid. With the growing demand for cloud computing, there is an enormous increase in operational cost and energy consumption, making an adverse impact on environmental conditions. Reducing the power consumption of data centers is a challenging task as there is rapid and huge incoming data which needs to be processed faster by large servers within the given time constraints. The problem is addressed by maintaining the energy consumption with efficient processing and utilization. Otherwise, tremendous amount of energy will be consumed by the data centers. The main objective of the present wok is to promote an eco-friendly environment by reducing energy consumption. It also focuses on managing Quality of service (QoS) to reduce SLA violations by developing efficient policies and algorithms. The rest of the paper comprises of the following. Section 2 depicts about related work, followed by in Section 3 explains architecture of virtual machine migration, Section 4 explains VM Selection method. Section 5 explains VM migration method. Section 6 is about performance metric, Section 7 is about experimental setup and simulation results. Section 8 concludes the paper with Section 9as references.

2. Related Work

Reduction in the Power consumed by all the servers running full time at the data centers and to reduce the Service Level Agreement (SLA) Violation as minimal as possible to tackle the growth of Data. The Servers runs full time as the users do not appreciate down time of the server due to which the power consumed every hour is really high.

Cardosa et al [1] built up a VM placement that had utilized the CPU portion that highlights least, greatest

Published by: The Mattingley Publishing Co., Inc.

and shares which are available in present day hypervisors. These were the highlights which were centered on CPU. The calculation meant to make balance between virtual machine execution and energy usage. There were four strategies which were accommodated for VM position making full use of the natural highlights of virtualization innovation.

Keller et al [2] were developing based on the variations of the First Fit heuristic. It was to address the VM movement issue, they indicated that the request for Virtual machines or the host's relocation influences the exhibition measurements of server farms like SLA violations and energy usage.

Zhen et al [3] acquainted skewness as a quantifiable measure to measure irregularity for multi-dimensional asset use of a server. He introduced a framework that uses virtualization innovation to distribute datacenter assets progressively dependent on requests of the applications. The framework was adjusted to green processing by improving the quantity of servers being used. By limiting the skewness, various kinds of workloads were joined.

Srikantaiah et al [4] had explored the impact of execution decrease because of high use of various resources when combining the workload. They had contemplated the issue of request scheduling for multiple layered web-based applications in virtualized frameworks to limit energy usage while meeting execution prerequisites.

Speitkamp and Bichler [5] proposed a static server solidification approach that examine the data numerically to portray varieties of real-world workload traces.

Gandhi et al [6] introduced a queuing theoretic model which allows the prediction of the mean response time as a function of the power-frequency relationship, arrival rate, and peak power budget as they had examined the impact of various factors on mean response time. They considered the issue of allocating an available power budget among servers in a virtualized server's datacenter while limiting the mean response time.

Ferdaus [7] had focused on the wastage of resources and energy consumption. He addressed the energy and resource related issues in server farms by focusing at the datacenter level resource management.

Beloglazov and Buyya [8] they introduced a VM merging technique using adaptive threshold. The adaptive threshold was based on statistical analysis with the history of the data collected during the lifetime of VMs. They analyzed that the average interruption in service and the migration time of the virtual machines for Web-based applications was nearly about 10 percent of the aggregate CPU usage. Resources within Cloud Data Centers are normally over-provisioned (inclusion of extra storage capacity) to assure high quality of service and service availability.

Kusic et al [9] used Limited Lookahead Control (LLC) technique to address the problem of power management in virtualized environments as a sequential optimization. The author's objective was to increase the resource provider's profit by minimizing both power consumption and SLA violation.

Khanna et al [10] introduced a heuristic technique that sorts the virtual machines based on the CPU and memory usage in the ascending order to minimize the migration costs. Then hosts list is sorted in the ascending order on remaining capacity to maximize resource utilization. Authors demonstrated that the number of servers running can be decreased using the virtualization technology.

The selection of the Virtual machines and hosts for migration is by developing a mathematical optimization model.

Deshpande and Keahey [11] aim was Traffic-sensitive live migration of virtual Machines. They proposed network aware live migration to alleviate the influence of migration on SLA and application QoS.

Naha et al [12] proposed the load balancing and cloud brokering method for the cloud server farms.

Son and Buyya [13] proposed Software Defined Networking (SDN) which was a powerful feature in Cloud computing that provides a centralized view of topology and bandwidth on every path.

The open-source virtual switch, Open vSwitch (OVS) [14] provides the virtualization switching stack supporting Open Flow and other standard protocols.

CLOUDS-Pi [15] alow-cost testbed environment for SDN-enabled cloud computing, is used as the research platform to test virtual machine block live migration.

Architecture of Virtual Machine Migration

The architecture of VM migration has been demonstrated below as per the diagram.

As per figure 1 we can see that there is a datacenter broker which is responsible for sending instructions and assigning cloudlets (user requests) to their respective physical machines hosting the virtual machines. A host is a physical server that houses virtual machines. A virtual machine is mimicking of the computer system which provides the same functionality as that of a physical machine.

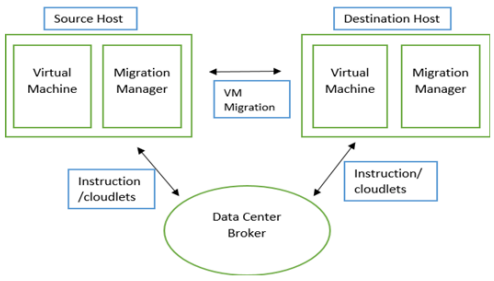


Figure 1: shows the architecture of VM migration.

Published by: The Mattingley Publishing Co., Inc.

TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 5212-5216

The datacenter broker is responsible for managing hosts as well as assigning virtual machines to a particular host. Now how does a datacenter broker decide to which host a virtual machine is allocated to? That's where our virtual machine migration policy comes into action. The migration policy decides which host is suitable for a particular virtual machine. It also decides whether a host is over-utilized or under-utilized. Once the host has been selected for migration a virtual machine must be selected from that host for migration that's when the VM selection policy is used. The VM selection policy selects a VM from a list of VMs running on the host. Based on their CPU, RAM and Bandwidth usage a VM is selected for migration then the migration manager of the host takes care of the migration progress from one host to another.

Proposed Algorithm for VM Migration

In this segment we are proposing a pre-existing VM migration algorithm with a few enhancements. In this proposed scheme in continuation with the previous algorithm, we use a modified version of local regression whichidenti identifies a host to be either under-utilized or over-utilized where in choosing host for further operations. The following algorithms have been mentioned below:

Algorithm 1: Detecting Over Utilized Host

```
Input: A host from datacenter
Result: Boolean (True or False)
1 if host.utilizationHistory< 10
2          if host.currentUtilization> 0.7
3                    return true;
4          else
5                    return false;
6   end
7 else
8           utilHistoryR = new double [length];
9 for i = 0 to length do
10         utilHistoryR [i] = utilHistory [length - i - 1];
11 end for
12 estimates ← null;
13 estimates = getParameterEstimates (utilHistoryR);
14 predictedUtil = estimates [0] + estimates [1] ∗
(length + migrationIntervals);
15 if predictedUtil * safetyParameter>= 1
16       return true;
17   else
18       return false;
19END
```

Algorithm 2: Detecting Under Utilized Host

```
Input: A host from datacenter and Excluded Hosts Set
Result: Underutilized Host
1 minUtilization← 0.35;
2 undUtilHost← NULL;
3 HostList←getHostList ();
4 foreach host in HostList
5          if excludedHosts.contains (host)
6                    continue;
7   utiliz = host.getUtilizationOfCpu ();
8          if utiliz> 0 &&utiliz<= minUtil&&
!areAllVmsMigOutOrAnyVmMigIn (host)
9          minUtilization = utilization
10         underUtilizedHost = host
11 return undUtilHost
```

Brief Description

The Migration Policy considers migrating of VMs from a host if it is considered to be under-utilized or over-utilized. The migration algorithm is also responsible for sending the chosen host to the VM selection algorithm for selection of VM for migration. Our modified algorithm, Modified Local Regression (MLR) detects when the host is over-utilized or under-utilized by predicting the future CPU utilization of a host.

Proposed Solution

As per algorithm 1, it is used to detect an over-utilized host. There are two approaches for this:
1) Adaptive utilization threshold base.
2) Non-threshold base algorithm.
The non-threshold base algorithm determines if a host is over-utilized then decides if migration of one or more VMs from that host is required. There is no static upper threshold but based on past data predicted utilization of host in next time frame is determine. In linear algebra regression means to find a relation between two variables. These two variables are time and percentage CPU utilized by VM for each time interval. The MLR algorithm needs utilization history of a host to predict the future utilization of that host, till we get the past utilization a static threshold will be used to consider the host as overloaded when the utilization crosses the threshold. When the utilization history of a host is available then the future CPU usage is predicted, if the future CPU usage of the physical machine will be more than or close to 100%, then that host will be considered as over-utilized host to avoid any further SLA violations.
Variables used in algorithm 1:
utilHistoryR => utilization History Reserve = the utilization History in reserve order.
utilHistory => utilization History of a host.
getParameterEstimates => gets utilization estimates of the host.
migrationIntervals => it defines the gap between VMs migration.
safetyParameter => it is a tuning parameter used by the allocation policy to estimate host utilization (load). The host overload detection is based on this estimation.
The algorithm 2 as mentioned above is used to detect host which are under-utilized. When a host is found to be under-utilized all the running VMs of that host will

May-June 2020
ISSN: 0193-4120 Page No. 5212-5216

be migrated to suitable hosts and the host will beset to hibernation mode. A host which has CPU utilization of 35% or less, that host is considered as under-utilized. If a host with utilization is less than 35%, the host's utilization will be set as the new lowest limit. The lowest limit for utilization will be updated as new host with least utilization are encountered.
The algorithm 1 and 2 are the enhancements that have been made to the existing Local Regression policy.

3. Experimental Setup And Simulation Results

CloudSim 3.3 toolkit [19] has been used for simulation in this paper. Cloudsim is a modern day simulation tool which allows us to create cloud data center which provide on demand virtualization resources and application management. For our simulation we created a data center with 800 physical servers and over 1050 active VMs.

Workload Characterization

Workload used in our simulation contains real system data which are the main sources of running our simulation and have been taken from Planet Lab.
Planet Lab contains real system data collected from various systems like HP and IBM servers. These are used as benchmarks to run the simulation. The workload data use in this simulation have been collected between March 2011 and April 2011. The workload used for simulation, CPU usage is below 50% in terms of workload data and VM assignments are random during the simulation run.

Experimental Setup

Each physical machine belonging to a data center contains a dual core processor and as per the workload system model performance of each core is set to 1860 MIPS for HP ProLiant ML110 G4 server and 2660 MIPS for HP ProLiant ML110 G5 servers. Each physical machine hasa capacity to work with 1 GBPS of network bandwidth. The VMs contain a single core processor as the workload data supports only single core VMs.

Simulation Scenario

In the simulation scenario the data center is connected to the internet. User requests are sent to the data center over the internet. The user requests are generated as per the workload data which are then sent to the cloud data center. The cloud data center is also responsible for processing the user requests and sending it to the respective physical machine on which the user's virtual machine is running.

4. Simulation Results And Discussion

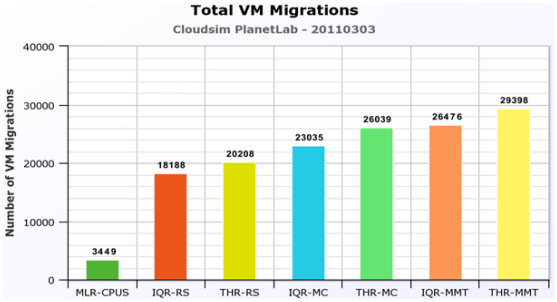
Our proposed algorithms MLR+CPUS have been simulated with the workload data. The simulation environment was created and run using the CloudSim toolkit [19] with real life workload data which has been borrowed from Planet Lab's physical machines which are located all around the world. The results obtained are presented in figures 2 and 3.

Virtual Machine Migration

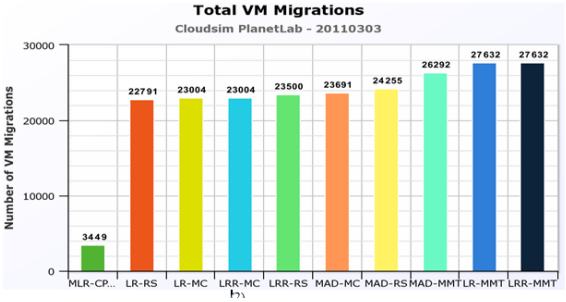
As per figures 2 and 3 we can see that the proposed MLR+CPUS algorithms have 3449 number of VM migrations during one simulation. As per results MLR+CPUs has the least number of VM migrations. Comparing our proposed MLR+CPUS with algorithms like LR-MMT, LRR-MMT and THR-MMT which have more than 27000 VM migrations and are based on Local Regression comparatively MLR+CPUS has 88% lesser number of VM migrations. MLR+CPUS's VM migrations are very less compared to any other algorithm making it a very efficient algorithm.

5. Conclusion

As per the results presented in this paper, we can state that our proposed VM migration (MLR) combined with VM selection algorithm (CPUS)together, are more efficient compared to the previous works done. As per the results, our proposed solution completes the simulation with the least power consumption, with a SLA violation rate of only 0.04% and number of VM migrations being only 3449 making it the most efficient solution under every criteria that has been considered. Dynamic VM consolidation has been considered for the simulation scenario. Finally, an analysis of the simulation results have been obtained and presented.



(2)



(3)

Figures 2 and 3: show the number of VM migrations.

Published by: The Mattingley Publishing Co., Inc.

May-June 2020
ISSN: 0193-4120 Page No. 5212-5216

References

[1] Cardosa M, Korupolu M. R & Singh A. (2009). "In Integrated Network Management, 2009". IM'09. IFIP/IEEE International Symposium on (pp. 327-334). IEEE

[2] Keller G, Tighe M, Lutfiyya H & Bauer M. (2012). "In Network and service management (cnsm),

[3] 2012 8th international conference and 2012 workshop on systems virtualization management (svm)" (pp. 406-413). IEEE

[4] Xiao Z, Song W, Chen Q (2013) "Dynamic resource allocation using Xiao Z, Song W, Chen Q (2013) "Dynamic resource allocation using virtual machines for cloud computing environment "IEEE Trans Parallel Distributed Syst 24(6):1107-1117

[5] Srikantaiah S, Kansal A, Zhao F "Energy aware consolidation for cloud Computing" Cluster Computing 12 (2009) 1–15.

[6] Speitkamp B, Bichler M (2010) "A mathematical programming approach for server consolidation problems in virtualized data centers." IEEE Trans Serv Computing :266–278

[7] Gandhi A, Harchol-Balter M, Das R, Lefurgy C (2009)"Optimal power allocation in server farms, in: Proceedings of the 11th International Joint Conference on Measurement and Modeling of Computer Systems, ACM" New York, NY, USA, 2009, pp. 157–168.

[8] Ferdaus M. H. (2016) "Multi-objective virtual machine Management in Cloud Data Centers (Doctoral dissertation, Monash University)"

[9] Beloglazov A, Buyya R (2012) "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. Concurrency Computing :1397–1420"

[10] Kusic D, Kephart J O, Hanson J E, Kandasamy N, Jiang G "Power and performance management of virtualized computing environments via lookahead control, Cluster Computing (2009) 1–15."

[11] Khanna G, Beaty K, Kar G, &Kochut A. (2006). "In Network Operations and Management Symposium, 2006." NOMS 2006. 10th IEEE/IFIP (pp. 373-381). IEEE

[12] Deshpande U, Keahey K, "Traffic-sensitive live migration of virtual machines, Future Gener." Comput. Syst. 72 (2017) 118–128.

[13] Naha R K, Othman M and Akhter N." Diverse approaches to cloud brokering: innovations and issues." International Journal of Communication Networks and Distributed Systems 19.1 (2017): 99-120.

[14] Son J, Buyya R "A taxonomy of software defined networking enabled cloud computing, ACM Comput. Surv. 51 (3) (2018) 59:1

[15] OpenvSwitch, https://www.openvswitch.org/, 2016

[16] Toosi A N, Son J, Buyya R "Clouds-pi: A low-cost raspberry-pi based micro data center for software-defined cloud computing" IEEE Cloud Comput. 5 (5) (2018) 81–91.

[17] Son J, Buyya R "A taxonomy of software defined networking enabled cloud computing, ACM Comput. Surv. 51 (3) (2018) 59:1

[18] Toosi A N, Son J, Buyya R "Clouds-pi: A low-cost raspberry- pi based micro data center for software-defined cloud computing" IEEE CloudComput. 5 (5) (2018) 81–91.

[19] Energy-aware virtual machine selection method for cloud data center resource allocation Nasrin Akhter1_, Mohamed Othman12_ (Member, IEEE), Ranesh Kumar Naha3.

[20] Modeling and Simulation of Scalable Cloud Computing Environments and the CloudSim Toolkit: Challenges and Opportunities Rajkumar Buyya1, Rajiv Ranjan2 and Rodrigo N. Calheiros1,33 Pontifical Catholic University of Rio Grande do Sul Porto Alegre, Brazil {raj, rodrigoc} @csse.unimelb.edu.au, rajiv@unsw.edu.au