SERVER CONSOLIDATION USING MULTI-OBJECTIVE OPTIMIZATION IN CLOUD COMPUTING

[1]Rajeshvari Panchal, [2]Bela Shrimali, [3]Dr.Hiren B. Patel
[1]PG Scholar, [2]Lecturer, [3]Professor&Head
[11]Department of computer engineering,
[1] LDRP-Institute of technology and Research, Gandhinagar, India.

Abstract: Cloud computing is emerging as a new paradigm of large-scale distributed computing. It is a framework for enabling convenient, On-demand network access to a shared pool of computing resources. Server consolidation has become increasingly important for improving efficiencies of resource usage and power consumption in datacenters. Consolidation is most effective method for energy saving in cloud environment with dynamic workloads. In this paper we proposes host selection policy to optimize energy, SLA, and number of migration in cloud datacenters. We using multicriteria weighted decision making technique for select virtual machines from overloaded PMs.

***Keywords*- Cloud computing; Consolidation; data center; energy consumption; Migration.**

I. Introduction

Cloud computing has become a popular computing paradigm that provides access to computing resources and application services as a pay-as-you-go business model. There are three common cloud computing service model known as Infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). In IaaS, the cloud service provider facilitates services to the users with virtual machines and storage to improve their business capabilities. Cloud provides a managed pool of resources which includes storage processing power and software services.

Virtualization technology which is the platform of cloud computing facilities the process of resource management in cloud environment. Virtualization is an important feature of cloud computing that allows providing multiple virtual machine on single physical machine as well as migration of virtual machines. Consolidation of VMs on the least possible PMs and switching ideal PMs off is the most novel method to save energy. Server consolidation using virtualization is an effective approach to achieve better energy efficiency of cloud data center [1].

Considering various goals that sometimes are contradicted with each other makes the resource management problem in cloud data centers a recently challenging issue. The two most important requirement that have to be considered for resource allocation in cloud environments are energy consumption and service level agreement (SLA) fulfilment. The SLA is an agreement that specifies the quality of service between service provider and service consumer [2].

The basic consolidation problem in cloud data centers is divided four sub-problems are (1) determination of overloaded PMs, (2) determination of underloaded PMs, (3) Selection of VMs that should be migrated from overloaded PMs, and (4) placement of migrating VMs on PMs [5]. Moreover, this paper considers the important criteria including SLA violation and Resource utilization in proposed algorithms.

This paper is organized as follows. Section II presents related work. Section III presents system model and proposed policy. Section IV concludes the paper.

II. Related Work

The authors in [1] have proposed enhanced optimization (EO) policy as a novel resource management procedure in cloud data centers. The main idea behind EO policy is solving the resource allocation problem for the VMs that are selected to be migrated from either overloaded or underloaded PMs. Besides, they have introduced a solution based on the Technique for Order of Preference by Similarity to ideal solution (TOPSIS) for optimizing different targets in clouds data centers at the same time including energy consumption, LA violation, and number of VM migrations. Based on this idea, they have proposed TOPSIS power and SLA Aware Allocation (TPSA) and TOPSIS –available capacity –number of VMs-Migration Delay (TACND) policies as novel multi-criteria algorithms for resource allocation and determination of underloaded PMs in cloud data centers, respectively.

The authors in [2] have focused the problem of multi-criteria resource allocation in cloud data centers considering important objectives including energy consumption, SLA violation, and number of VMs migrations. This paper proposed Multi-objective power and SLA (MOPS) policy based AHP algorithm as a novel multi criteria resource allocation solution that simultaneously consider five criteria including the power increase, Available capacity, number of VMs, Resource Correlation ,Migration delay.

Antonescu [3] et al. present a framework for SLA management based on multi-objective optimization. The framework features a forecasting model for determining the best virtual machine-to-host allocation given the need to minimize SLA violations, energy consumption and resource wasting. They describes the genetic group allocation algorithm, the multicriteria evaluator and forecaster.

The authors in [5] have conducted competitive analysis and proved competitive ratios of optimal online deterministic algorithms for the single VM migration and dynamic VM consolidation problems. They have divided the problem of dynamic VM consolidation into four parts: (1) determining when a host is considered as being overloaded; (2) determining when a host is considered as being underloaded; (3) selection of VMs that should be migrated from an overloaded host; and (4) finding new placement of the VMs selected for migration from the overloaded and underloaded hosts. They have proposed novel adaptive heuristics for all parts. They have used Power Aware Best Fit Decreasing (PABFD) algorithm to solve resource allocation problem.

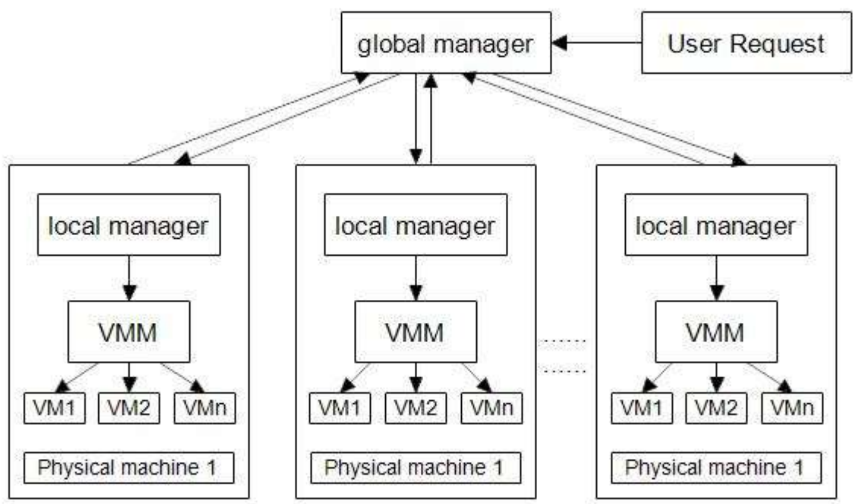
III. PROPOSED SYSTEM MODEL

3.1 Problem Statement

To propose multi-criteria decision making method for both determination of underloaded hosts and placement of the migrating VMs by considering criteria for power increase, Available capacity, Number of virtual Machines, Resource Correlation, Migration delay, resource utilization and Resource utilization. To propose such method that optimize energy consumption, reduce SLA Violation.

3.2 System Architecture

As shown in figure the system architecture consist following steps:



First, user submit new request for provisioning for virtual machines.

The local manager send information about resource to the global manager and monitoring PMs. the local manager which are implemented by VMM are connected to global manager through network interfaces.

The global manager is resource manager of specific data center which allocate virtual machine to available hosts in the data centers.

The global manager handles placement of VMs to PMs as well as resource distribution among VMs in the data center.

VMM performs actual resizing and migration of VMs.

3.3 Performance, Energy and Utilization model

This energy model is basis on CPU utilization. This approximation comes from the idea that CPU is the major power consumer in a data center. However, is that by introducing multi-core CPUs with modern power management techniques, as well as utilization of virtualization technique, CPU is not the only major power consumer in data centers any more. This fact combined with the difficulty of modeling power consumption in modern data centers, makes building precise analytical models a complex research problem. Hence, instead of using a complex analytical model for power consumption of a server. Table 1 shows the power consumption of the servers used in this study which provided in [5].

Energy consumption is modeled as the summation of power consumed during period of time according to equation (1).Which is used in the literature such as [6].

$$E(t) = \int p(t)\, dt \quad (1)$$

Quality of service requirement are commonly formalized in the form of SLAs, which can be determined in terms of such characteristics as minimum throughput or maximum response time [4]. SLA violation metric used in this equation which is multiplication of two metrics: the SLA violation time per active host (SLATAH) and performance degradation due to migration as defined in below equation (2) and (3).

$$SLAV = SLATAH * PDM \quad (2)$$

$$SLATAH=\frac{1}{N}\sum_{i=1}^{N}\frac{Tsi}{Tai}\ ,\quad PDM=\frac{1}{M}\sum_{j=1}^{M}\frac{Cdj}{Crj} \quad (3)$$

Where,

- N=Total No. of Nodes.
- Ts_i=Total time, node *i* experiencing full utilization.
- Ta_i=Total time, node *i* in active state.
- M=Total no.of VMs.
- Cd_j=Estimation of performance degradation of VM j due to migration.
- Cr_j=CPU capacity requested by VM j during its lifetime.

Host Utilization U_H is defined as [7].

$$U_H = \frac{\sum_{i=1}^{n}(U_{vmi} \times C_{vmi})}{C_H} \quad (4)$$

Where,

- n=Total no.of VM on a host.
- C_H=Host capacity in MIPS.Which is defined eq. (5)
- $U_{vmi=}$VM utilization.
- $C_{vmi=}$Total VM capacity in MIPS.

C_H= Number of core× Individual core capacity (5)

3.4 Proposed Power and SLA weighted placement Policy

The general steps of the consolidation technique.

1. Identification of overloaded PM
2. Selection of VM Migrate from overloaded PM based on VM Selection Policy
3. Creation of Migration list
4. Identification of Underloaded host and add all VMs of underloaded PMs to the migration list.
5. Find new placement for all the VMs in the using TOPSIS power and SLA placement policy

Our technique works on find new placement for all VMs in the using power and SLA weighted TOPSIS policy

PSWT is a multi-objective resource allocation method to identify solutions from a finite set of alternatives based upon simultaneous distance minimization from an ideal point and distance maximization from a nadir point and improves the TPSA policy proposed in [1] in three aspects. First, it considers available maximum capacity of a PM based on its maximum frequency instead of its current frequency level. Second, PSWT introduces weighted TOPSIS method as a functional tool for Cloud administrators to effectively specify their desired resource management configurations by applying different weights for the criteria defined in Table 1.

This method let the administrator to set a trade-off for fulfilling different objectives including energy consumption, SLA violation, and number of VM migrations. PSWT policy takes advantage of TOPSIS as a multi-criteria algorithm that considers seven criteria depicted in Table 1 in its decision process. This policy computes the scores of all the PMs that are candidate for hosting a VM using the method that is described in this section and selects the PM with the highest score. Criteria considered in PSFT policy can have either benefit or cost type. The more the value of criteria with the benefit type, and the lower the value of criteria with the cost type, the closer is the answer to the optimum point.

PSWT computes the score of PMs so that the following conditions simultaneously exist in the answer: (1) the selected PM has the least power increase(2)the selected PM has the Most available capacity (3) the selected PM has the least number of VMs, (4) VMs on the selected PM have the least resource correlation with the VM to be allocated, (5) the migration delay of the VM to be allocated to the selected PM is the least, and (6) the selected PM has the most resource utilization (7)the selected PM has least SLA Violation.

PSWT policy defines two ideal points including ideal positive point (*PM*□) and the ideal negative□ point (*PM*-) by composition of the best and worst values of the criteria shown in Table 1. PSWT chooses a□□PM for hosting a VM that has the shortest distance from *PM*+ and the farthest distance from *PM*-.

Table1: considered criteria in PSWT policy.

No.	Notation	Parameter	Description
1	PI	Power increase	Power increase of allocating a VM on a PM
2	AV	Available capacity	Available resource capacity of a PM
3	NV	Number of VMs	Number of VMs on PM
4	RC	Resource correlation	Resource correlation of a VMs with the VMs on a PM
5	MD	Migration delay	The delay incurred due to migration of VMs to PM
6	RU	Resource utilization	Utilize the resource of a PM
7	SLAV	SLA violation	QOS agreed between the service provider and the service consumer

All the information assigned to the PMs in the time slot t from a decision matrix PMC (physical machine configuration) as shown in figure

PI_{PM1}	AC_{PM1}	RU_{PM1}	NV_{PM1}	RC_{PM1}	MD_{PM1}	$SLAV_{PM1}$
...	...	…	...	...	...	…
PI_{PMi}	AC_{PMi}	RU_{PMi}	NV_{PMi}	RC_{PMi}	MD_{PMi}	$SLAV_{PM1}$
...	...	…	...	...	...	….
PI_{PMN}	AC_{PMN}	RU_{PMN}	NV_{PMN}	RC_{PMN}	MD_{PMN}	$SLAV_{PM1}$

Where PM1, PM2….PMn are available PMs that are candidates of selection by TPSA; PI, AC, NV, RC, MD are the Five criteria.
Step1: First we normalize the decision matrix PMC (Physical Machine Configuration) to have dimensionless decision matrix PMC. The decision matrix is made dimensionless by dividing each entry by maximum value of each columm according to below matrix.

$\frac{PIPM1}{PImax}$	$\frac{ACPM1}{ACmax}$	$\frac{RUPM1}{RUmax}$	$\frac{NVPM1}{NVmax}$	$\frac{RCPM1}{RCmax}$	$\frac{MDPM1}{MDmax}$	$\frac{SLAVPM1}{SLAVmax}$
…	…	….	…	…	…	…
$\frac{PIPMi}{PImax}$	$\frac{ACPMi}{ACmax}$	$\frac{RUPMi}{RUmax}$	$\frac{NVPMi}{NVmax}$	$\frac{RCPMi}{RCmax}$	$\frac{MDPMi}{MDmax}$	$\frac{SLAVPMi}{SLAVmax}$
…	…		…	…	…	…
$\frac{PIPMN}{PImax}$	$\frac{ACPMN}{ACmax}$	$\frac{RUPMN}{RUmax}$	$\frac{NVPMN}{NVmax}$	$\frac{RCPMN}{RCmax}$	$\frac{MDPMN}{MDmax}$	$\frac{SLAVPMN}{SLAVmax}$

Step2: In the next step, PM^+ and PM^- are determined. In general, the criteria can be classified into two types: benefit and cost. The benefit criteria means that a higher value is better, while for the cost criteria is the opposite. Larger values for a benefit type attribute leads to less distance from PM+ and more distance from PM-, while opposite condition is cost type variable.AC is benefit type attribute and all other is cost type attribute. Then determining PM+ and PM-.we want to place a VM on a PM that the PM has least power increase, the highest available capacity, least number of VMs, least resource correlation, and least migration delay.

PM+ = {PI-, AC+, RU+, NV-, RC-, MD-, SLAV-}
PM- = {PI+, AC-, RU+, NV+, RC+, MD+, SLAV-}

Where criteria+ and criteria- are the maximum and minimum value in each columm of decision matrix.
Step3: In this Step, the relative distance for each criteria of a PM from PM+ and PM- are calculated using below equation

$$\text{Score} = \frac{\sqrt{(PM^{j}_{criterian} - PM^{-}_{criterian})^2}}{\sqrt{(PM^{j}_{criterian} - PM^{-}_{criterian})^2}\sqrt{(PM^{j}_{criterian} - PM^{+}_{criterian})^2}}$$

Where Score (PM^j) the score of a specific criterian of j^{th} PM.

Step4: To identify the optimize host we are using weighted method of multi-objective optimization.it is the method of optimizing technique in which weights are assigned to score of each PMs

$$\text{Score } (PM^{j}) = \sum_{criterian=1}^{\#criterian} Weight_{criterian} * Score_{Criterian}^{PMj}$$

Step5: Rank PMs according to their Score and select the one with highest score. The PM with the highest score has the maximum distance from PM- and the Minimum distance from PM+.

4. Experimental Evaluation

4.1 Example Scenario:

The explanation of the example is carried out from the matrix as shown in table2.there are seven criteria and six PMs .we have considered two criteria in addition i.e. resource utilization and SLA violation as shown in table 2.seven criteria are PI, AC, NV, RC, MD, RU, SLAV in table 2.in order to select the best PM we go through the following steps:

Table2: Value of all criteria

Criterion	Power Increase	Available Capacity	Number of VMs	Resource Correlation	Migration Delay	Resource Utilization	SLA Violation
Physical Machines	PI(Watt)	AC (MIPS)	NV (Number)	RC (%)	MD (mS)	RU (%)	SLAV (%)
PM_1	20	200	10	20	2.2	40	0.00125
PM_2	30	300	8	25	0.75	28	0.00119
PM_3	10	320	12	80	1.71	43.75	0.015873
PM_4	20	800	5	50	2.88	54	0.009259
PM_5	15	300	7	90	3.67	56.333	0.016789
PM_6	10	500	9	70	2.44	52	0.010417

Step2: All the criterian are divided by maximum value of each column.

Table3: All the criteria are divided by the maximum value of each columm

Criterion	Power Increase	Available Capacity	Number of VMs	Resource Correlation	Migration Delay	Resource utilization	SLA violation
Physical Machines	PI(Watt)	AC (MIPS)	NV (Number)	RC (%)	MD (mS)	RU (%)	SLAV (%)
PM_1	0.67	0.25	0.83	**0.22**	0.60	0.71	0.07
PM_2	1.00	0.38	0.67	0.28	0.20	0.50	0.07
PM_3	0.33	**0.40**	1.00	0.89	**0.47**	**0.78**	**0.95**
PM_4	**0.67**	1.00	**0.42**	0.56	0.78	0.96	0.55
PM_5	0.50	0.38	0.58	**1.00**	**1.00**	**1.00**	**1.00**
PM_6	**0.33**	**0.63**	0.75	0.78	0.66	0.92	0.62

Step3: PM+ are determined whose power increase least ,the highest capacity, least number of VMs, least resource correlation, most resource utilization, least SLA violation .and for PM- it is vice versa. We place VM on such PM+.

PM^{+}= {PI^{-}, AC^{+}, NV^{-}, RC^{-}, MD^{-} , RU^{+}, $SLAV^{-}$} = {0.33, 1.00, 0.42, 0.22, 0.20, 0.50, 0.07}
PM^{-} = {PI^{+}, AC^{-} , NV^{+}, RC^{+}, MD^{+}, RU^{-}, $SLAV^{+}$} = {1.00, 0.25, 1.00, 1.00, 1.00, 1.00, 1.00}

Step4: The relative distance for each criterian of a PM

Table3: find the score of each criterian

Criterion	Power Increase	Available Capacity	Number of VMs	Resource Correlation	Migration Delay	Resource Utilization	SLA Violation

Physical Machines	PI(Watt)	AC (MIPS)	NV (Number)	RC (%)	MD (mS)	RU (%)	SLAV (%)
PM1	0.50	0.00	0.71	0.00	0.50	0.58	0.00
PM2	0.00	0.17	0.43	0.07	0.00	1.00	0.00
PM3	1.00	0.20	1.00	0.86	0.33	0.44	0.94
PM4	0.50	1.00	0.00	0.43	0.73	0.08	0.52
PM5	0.75	0.17	0.29	1.00	1.00	0.00	1.00
PM6	1.00	0.50	0.57	0.71	0.58	0.15	0.59

Step5: Modified PMC with assigned to score of PMs and all the VM assigned highest score of PM.

Case1: Existing Method with same weight assigned (weight=0.2)

Table5: existing method with same weight assigned

Physical machine	PM1	PM2	PM3	PM4	PM5	PM6
Score	0.34	0.13	0.67	0.53	0.64	0.67

Table6: Existing method with different weight assigned.

Physical machine	PM1	PM2	PM3	PM4	PM5	PM6
Score	0.41	0.23	1.01	0.88	1.03	1.00

Case2: Proposed method with Different weight assigned

Table7: Proposed method with same weight assigned.

Physical machine	PM1	PM2	PM3	PM4	PM5	PM6
Score	0.46	0.33	0.95	0.65	0.84	0.82

Table8: Proposed method with different weight assigned.

Physical machine	PM1	PM2	PM3	PM4	PM5	PM6
Score	0.70	0.73	1.79	1.23	1.63	1.43

4.2 Performance Metrics Calculation

The main targets for comparing the efficiency of the algorithms are energy consumption by physical nodes and SLA violations, however, these targets are typically negatively correlated as energy can usually be decreased by the cost of the increased level of SLA violations [5]. Therefore, we use the Multi parameter Energy and SLA Violation (MESV) metric defined in Eq. (6) to assess the simultaneous optimization of energy and SLA violation. Moreover, since the objective of the resource management system is to minimize energy consumption, SLA violation, and number of VMs' migrations, we use the Multi parameter Energy, SLA violation, and Migrations' count (MESM) metric which is defined in Eq. (7).

$$MESV = Energy \times MPASV \quad (6)$$

$$MESM = MESV \times Migrations\ Count \quad (7)$$

4.3 Results and Discussion

In this section we compare TPSA policy and our proposed policy. In existing method, PM5 is selected which has the highest score. Existing system considers five criteria and same weight is considered. In proposed policy we have considered seven criteria and weight used is different. Our technique chooses the best PM (in our case PM3 is considered as it has the highest score) as compared to the existing method because it has least power increase, more available capacity, least resource correlation, least migration delay.

Table 9: Comparison between existing work and proposed technique

Criteria	In existing method the select PM5	In proposed method the select PM3
Power increase	15	10
Available capacity	300	320
Number of VMs	7	12
Resource correlation	90	80
Migration delay	3.67	1.71
Resource utilization	-	43.75
SLA Violation	-	0.015873

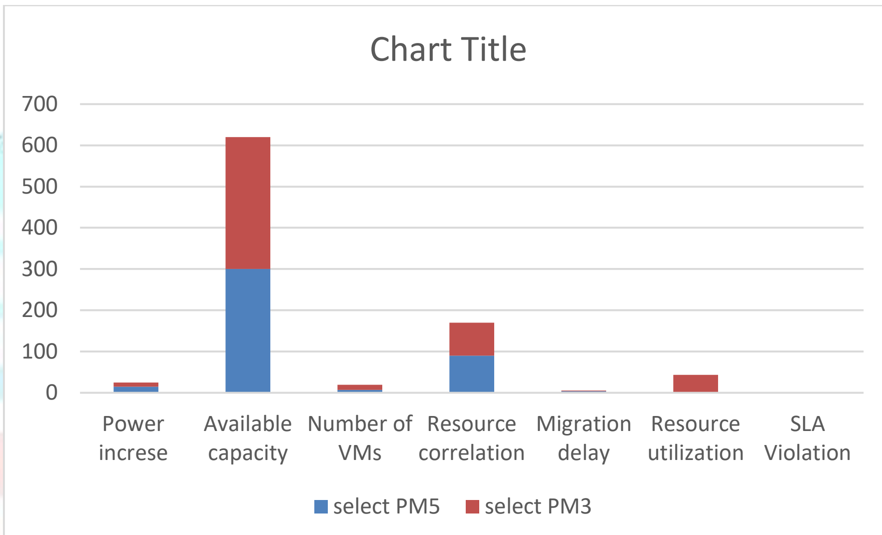


Figure2: comparison of existing method and proposed method

IV Conclusion

Server consolidation is a key feature in current virtualized data centers. It focuses on minimizing the amount of resources required to handle the data center workload and efficiently handle the different parameters like proper resource utilization, energy consumption and performance of the data center. For our work we have considered Different parameter used to consolidation technique such as power increase, available capacity, Number of VMs, Resource correlation, Migration delay, Resource utilization, SLA violation. it shows experimental evaluate result depicts that out technique PSWT is performing well 20% in power increase, 20% available capacity and improvement 20% in migration.in future work, we may use heuristic approach to using weight.

References:

[1] Arianyan, Ehsan, Hassan Taheri, and Saeed Sharifian. "Novel energy and SLA efficient resource management heuristics for consolidation of virtual machines in cloud data centers." *Computers & Electrical Engineering* 47 (2015): 222-240.

[2] Arianyan, Ehsan. "Multi objective consolidation of virtual machines for green computing in Cloud data centers." *Telecommunications (IST), 2016 8th International Symposium on*. IEEE, 2016.

[3] Antonescu, Alexandru-Florian, Philip Robinson, and Torsten Braun. "Dynamic SLA management with forecasting using multi-objective optimization." *Integrated Network Management (IM 2013), 2013 IFIP/IEEE International Symposium on*. IEEE, 2013.

[4] Horri, Abbas, Mohammad Sadegh Mozafari, and Gholamhossein Dastghaibyfard. "Novel resource allocation algorithms to performance and energy efficiency in cloud computing." *The Journal of Supercomputing* 69.3 (2014): 1445-1461.

[5] Beloglazov, Anton, and Raj Kumar Buyya. "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers." *Concurrency and Computation: Practice and Experience* 24.13 (2012): 1397-1420.

[6] Beloglazov, Anton, Jemal Abawajy, and Raj Kumar Buyya. "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing." *Future generation computer systems* 28.5 (2012): 755-768.

[7] Shrimali, Bela, and Hiren Patel. "Multi-objective optimization oriented policy for performance and energy efficient resource allocation in Cloud environment." *Journal of King Saud University-Computer and Information Sciences* (2017).