International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-5, March 2019

A Novel Approach of Virtual Machine Consolidation for Energy Efficiency and Reducing Sla Violation in Data Centers

Pardeep Singh, Jyotsna Sengupta, P.K. Suri

Abstract*: *Infrastructure as a Service (IaaS) virtually divides the hardware resources of datacenter and produces several Virtual Machine (VM) objects on one or more physical machines (PM). These VMs are further allocated to the different clients as per their requirements. So, virtualization has a significant role to enhance the utilization of PM. And, this is managed by optimally constructing and destructing VMs over PMs. In this work,VM consolidation algorithm is the center of research that includes the procedures of initial VM allocation, detecting overloaded hosts, selecting appropriate VMs for migration, then placing the migrated VMs overs PMs and finally detecting the idle or under-loaded hosts. Main issues discussed are dynamically placement of VMs and finding the idle or under-loaded hosts. VM Placement problem is solved by introducing a novel Utility Aware Best Fit Decreasing (UABFD) algorithm that increases the utility of the servers. By increasing the utility, servers can finish the assigned load in a shorter makespan, which leads to less energy consumption. Another problem of detecting underloaded hostsis also modified by applying some heuristics. Both the Proposed algorithms are then replaced with the existing algorithms inVM consolidation scheme, to designa Modified VM Consolidation (MVMC) scheme. Cloudsim is used for evaluating various performance parameters of cloud environment with the proposed scheme.Resultsare compared and analyzed with the existing VM consolidation scheme. That has proved thatMVMC has significant improvement for various parameters over the existing scheme.

Index Terms*: *Energy Consumption, VM Selection, VM Consolidation, VM Placement, SLA Violation, Cloud Computing, Data Center Management

I. INTRODUCTION

A Cloud environment which can fulfil user's expectations and achieve various objectives, requires a cost effective resource management. This is the reason that many immense IT companies like Amazon, Google, IBM and Microsoft are researching progressively for optimal management of their Cloud centers [1]. Moreover, performance of cloud system always remains a crucial factor; as if in any case service provider be unable to provide the services according to the SLA, then, clients can even leave the services permanently. Another important concern of excess power consumption also cannot be ignored [2]. Not only, it escalates the running cost of data centers, but also responsible for large amount of CO2 emission in the environment. Approximately 1,00,00,000 MW power was consumed by data centers in 2011, that produced 40,568,000 tons of CO2 emissions [5]. According to an estimate, internet consumes approximately 10% power of total world energy. And, after every 5 years, the price of power for operating data centers get doubled [3]. Between 2005 and 2010, the volume of energy used by data centers got raised by 56%. According to Karthikeyan&Chitra [4], data centers consumed almost 1.1% to 1.5% of the total world energy. Key point is that, only 20-30% energy is actually consumed by data center for performing server operations and remaining 70-80% power utilization is unused due to over provisioning of idle servers and other hardware resources [6], [7].

So, relating to the IaaS model, in which cloud user's requirements for computing or infrastructure are primarily served by allocating VMs, these parameters can be handled by managing virtualization [8]. The virtualization technique creates a number of VM on one or more PM also called as Host or Server in data centers [9]. As shown in Fig. 1, Cloud architecture includes numerous data centers that have a layer of hosts and these hosts are virtually divided into various VMs. When a client need to perform a task, VM Allocator on the basis of VM Allocation policy, creates the VM on a suitable host and assign the task to that VM. Resources required by the VM are considered and accordingly suitability of PM is defined on the basis of various factors, which are responsible for the energy consumption, service quality and some other important cost effecting properties of the Cloud.

Assignment of Tasks to the available VMs is done through Task Scheduler defined by the Broker. On next level, scheduling of VMs at a particular host is done by VM Scheduler. This scheduling is necessary for sharing and optimal utilization of resources among VMs. Most of the services often need a slight amount of the total existing resources. So, in this scenario a large portion of space and resources remain idle and not justified with the workload. This results to the wastage of resources and consumption of extra energy, referred as server sprawl problem [10]. To avoid this, minimum hosts should be packed with maximum number of virtual machines. And idle host machines can be turn down to sleep mode, which leads to less power consumption. This approach to prevent the server sprawl is called Server Consolidation.

Manuscript published on 30 March 2019.
***Correspondence Author(s)
Pardeep Singh, Department of Computer Science, Punjabi University, Patiala, India.
Jyotsna Sengupta, Department of Computer Science, Punjabi University, Patiala, India.
P.K. Suri, Department of Computer Science and Applications, Kurukshetra University, Kurukshetra, India.

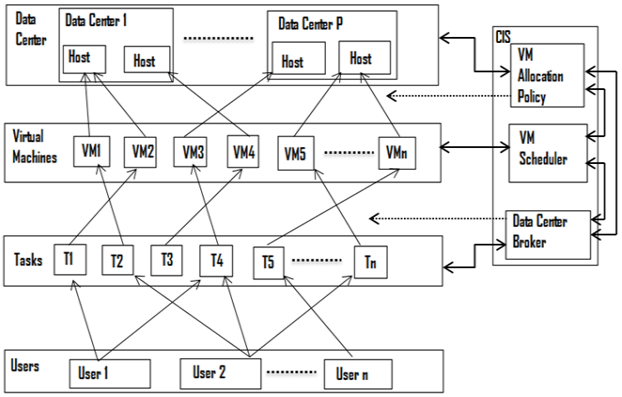

Fig. 1.Layers in Cloud Architecture

Virtualization has provided the feature to control the hardware resources, by dividing the physical resources virtually into different independent machines called Virtual Machines (VMs). A host or server or PM, can be virtually divided into several VMs at a time, with in a limit defined by resource capacity of that PM. In the process of Virtualization, Virtual Machine Monitor (VMM) acts as a virtual layer, and has the responsibility to divide the hardware layer of PMs into VMs. This process of dividing the physical resources into virtual resources, have to consider mainly the total capacity of all PMs and resource requirements of each and every VMs. The process of virtual resource description becomes more complex when PMs and/or VMs are heterogeneous.

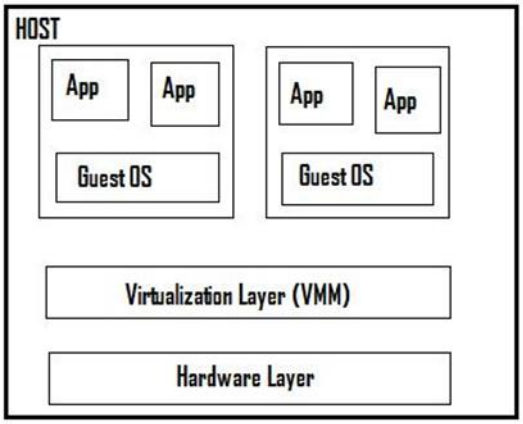

Fig. 2. Virtual Machine Abstraction

Live Migrations of VMs is another important aspect needs to be handled for optimal resource management [11]. Virtualization process just creates the VMs, but these VMs are dynamically managed by a VM consolidation strategy. That facilitates the virtual machines to be switched among various hosts without any disturbance in their running tasks. VMs are consolidated among the hosts, based upon the current load, carried by each host. In this load balancing VMs from heavier loaded hosts are migrated towards lighter loaded hosts. If any host gets idle, it should be shut down and become inactive for saving the energy. Various consolidation strategies have been defined by researchers and are further in research, for obtaining different objectives. An effective strategy can achieve more than one objective without affecting the others. However, there are some issues occurred while implementing the consolidation process. Some main issues are:

1. As the total workload to be process at a time, depends upon the amount of available tasks, provided by multiple users at that instant. Sharing of this unstable workload becomes somewhat difficult to handle. Mainly when some VMs are consolidated on a host, it might be getting overloaded. A host can execute the number of VMs, according to the total capacity it have. Over burden can leads to delay in processing of VMs. This would increase the response time of VMs and hence leads to SLA violation. So, in this phenomenon, QoS would not be according to the expectation of the user, which is termed as Service License Agreement (SLA). SLA violation more than defined amount, could results to a penalty to the service provider. Therefore, a story policy is required to handle the issue of overloaded hosts which can control the load over the host during VM Consolidation Process.
2. Live VM migration can affect the system performance as discussed in [11]. The complete process of migrating one VM from one PM to another requires some computing resources. So, if a huge number of migrations would took place for the consolidation, it would consume a big amount of hardware and system resources, again this extra burden of resource consumption will violate the SLA. Hence, some methodology is needed to reduce the VM migrations, while processing the VM consolidation.

A dynamic VM consolidation mechanism is proposed in this work, for the improvement of various performance metrics, which are described in later section. The main contributions of this research work are:

- Develop a VM placement algorithm based upon the maximum utilization of hosts by VMs, affecting the energy efficiency.
- Develop an under-loaded host detection algorithm which considers number of VMs on active hosts and utilization of those hosts, for detecting a suitable under-loaded host.
- Modify the existing VM consolidation mechanism by introducing newly developed algorithms in place previous existing algorithms.
- Compare and analyse the results of modified mechanism with the existing one based upon the various performance measurements.

Proposal is structured into different sections: Introduction section introduced the domain and prospects from the research. Section II describes the related work done by various researchers in the same field. Section III defines the problem identified during the research with the description of sub-problems. Further section IV refers to the preliminary and explains Power Aware Best Fit Decreasing (PABFD) and Underload Host Detection (UHD) algorithms [12]. Next, section V explores the proposed algorithms step by step. Sixth section provides the information related to experimental setup, performance metrics and data input. This section also analyzes and compares the results obtained from proposed technique with existing. Last section concludes the research and guidance about future work.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

II. RELATED WORK

Many researchers have paid their attention to optimize the energy usage, as well as improve the QoS in data centers. Because, more the energy consumption is, more it affects the environment and running cost. And, better QoS helps in retaining a better relationship with clients. Most of the researches have focused on solving this problem by applying some modifications in server consolidation scheme, out of which some latest has been reviewed here. Beloglazov and Buyya [13] emphasized on reducing the power consumed by CPU using live VM migration and dynamic load balancing techniques in cloud data centers. In this concept, Overloaded host detection problem is solved by determining the dynamic upper threshold value. This adapted threshold value is calculated based upon historical utilization data of resources. Local Regression (LR), Robust Local Regression (LRR), Interquartile Range (IQR) and Median Absolute Deviation (MAD), are four variants designed in the work for finding overloaded hosts. In addition, the researchers stated two policies for VM Selection problem, Minimum Migration Time (MMT) that chooses the VMs with less RAM and Maximum Correlation policy (MC) that selects those set of VMs that have maximum correlation of the CPU utilization. In this mechanism, power consumption from all server components is taken into account; however, only CPU is the considered factor, which is a weak point. So other factors could also be considered for improvement in power consumption. Zhou et al. [14] defined a three threshold energy saving algorithm (TESA), which classified hosts into four different classes according to the load handled : heavy-load hosts, moderate-load hosts, light-load hosts and little-load hosts. VMs from heavy-load hosts or little-load hosts are migrated to light-load hosts. However, the VMs on moderate load hosts and light load hosts remained same. Further, Zhou et.al. [15] modified TESA by making the thresholds adaptive and proposed a novel virtual machine deployment algorithm named Adaptive Threshold Energy Aware (ATEA). Then, two adaptive three-threshold algorithms; KAI and KAM are used to determine the thresholds. Three VM selection policies are also proposed: MMS, LCU and MPCM for energy optimization and SLA violation. Chowdhury et.al. [16] have considered VM placement as bin packing problem and three variants of PABFD [12] i.e. modified worst fit decreasing, second worst fit decreasing and first fit decreasing with decreasing host are generated. A clustering technique i.e. modified k-means algorithm is applied to create clusters of VM based upon CPU utilization and currently allocated RAM. High density cluster i.e. cluster with more numbers of VMs, will allocate the VMs firstly than next dense cluster and so on, till all the VMs are not allocated. The adopting clustering technique developed in this work, performed best with modified first fit decreasing with decreasing host algorithm (FFHDVP_C).

Han et. al. [17] defined a new algorithm for under loaded host detection and VM placement. For under load host detection they considered power efficient value (PE) of host. PE is the ratio between power consumption and number of VMs running on a particular host. All the hosts with the CPU utilization less than lower threshold would be considered in candidate host set. And, the host with the maximum PE value is selected as the under loaded host. Authors also defined a remaining utilization aware (RUA) VM placement algorithm that considered the remaining available CPU resources of PMs for placing the VMs. Then, a new integrated VM consolidation algorithm consisting of these two techniques is applied on five deferent types of planetlabwork load data using cloudsim simulator. The results are compared with PABFD [12] and UMC [18], which shows that the proposed algorithm has less impact for energy consumption. However, SLA violation, SLATAH, number of VM migrations and ESV metrics reduced significantly. Khoshkholghi et al. [19] researched on all the four sub problems of VM consolidation defined by Beloglazov&Buyya [12] and generate a new consolidation scheme comprised of all proposed techniques. In this, considering two utilization thresholds, which are defined by Iterative weighted linear regression technique, an over loaded host detection algorithm is developed. Also, a new VM selection algorithm based upon three different policies i.e. maximum power reduction policy, time and power tradeoff policy & violated Mips–VMs policy, is generated. Than a two phase VM placement algorithm is proposed. In first phase, VMs selected from over loaded hosts are placed and in second phase all VMs selected from under loaded hosts are allocated host. Further a Multiple Resources Under-loaded Host Detection algorithm (MRUHD) is introduced. This algorithm defines a adaptive lower threshold value. If CPU, RAM and bandwidth utilizations are lower than this threshold, the host is considered as under-loaded. The proposed consolidation scheme is evaluated using CloudSim toolkit. The results depicts that new proposed scheme outperformed in all the benchmarks and can reduced SLA and energy consumption up to 87% and 28 % respectively. Kuo et.al. [20] stated the resource based first fit algorithm for assigning VMs to host. In this procedure resource requirements of requesting VM are analyzed before assignments and available resources of each host are updated after the termination of VM assigned to it. Than the defined work is compared with resource based worst fit algorithm (RWFA) and resource based best fit algorithm (RBFA). The Performance evaluation shows that RFFA scheme performed well than RWFA & RBFA for decreasing energy consumption.

III. PROBLEM STATEMENT

With the analysis of the related work, some research gaps has been identified, which are helpful in designing the proposal of this research work. All the reviewed papers and related research have modified the existing techniques or developed some new policies for improvements in one or more performance metrics defined later. Researchers have developed many strategies for reducing the energy consumption, and for this they have considered the required or available resources, or amount of energy consumed by applying different allocation policies. For reducing the energy main concepts used are to allocate the VMs in such a way that it can reduce the number of active PMs and shutting down the idle servers.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

However, reduction in active hosts would results into aggressive consolidation and impose excess load of VM migrations. Opposite to it if we can retain the active hosts, it would increase the rate of energy consumption, but, overall tasks would be finished earlier with less VM migrations. This gap in research leads us to concentrate over increasing the utilization of servers and allocate them to preferably those VMs which can utilize the server for long time.

According to the Beloglazov&Buyya [12], during the VM consolidation process for optimal placement of VMs over hosts, four sub problems needed to be researched: first is to find out the overloaded hosts which have been allocated more number of VMs than their serving capacity. This results into idle VMs and hence promotes the SLA violation; second, selecting the adequate number of VMs from the overloaded hosts for migration, that should select appropriate VMs depending upon various parameters such as CPU utilization, Size of VMs, VM correlation etc.; third, detecting the under-loaded hosts , in which a threshold value is defined for some particular parameters and hosts with less than this threshold value is considered as under-loaded. Purpose for detecting under-loaded host is to find suitable host that can easily be shut down without affecting other parameters, as all the VMs from under-loaded hosts would migrate to other active hosts; lastly, to place all the VMs selected for the migration on suitable physical hosts in an optimal way. Placement here means how the available active physical hosts which are neither overloaded nor under-loaded could allocate their available resources to migrated VMs. This mapping between hosts and VMs is not stable, because, new VMs keep on generated to satisfy dynamic user demands and old VMs would get on terminated progressively after the completion of tasks assigned to them. This mapping of VMs and PMs have impact over all the performance measurements such as energy consumed by hosts, how much hosts can further be get overloaded or under-loaded after a specific interval , how much amount of VM migrations would be occurred and last but not least how much violation of SLA be happened.

It has been analyzed that there is a tradeoff between SLA violation and energy consumption [15]. Algorithms reducing the energy consumption might be impact on SLA violation and similarly to reduce the SLA violation more energy could be consumed [10]. So, a consolidation scheme that could maintain the balance between both parameters needs to be defined.

IV. PRELIMINARY

Out of four sub problems defined in previous section, two are researched and modified in this paper i.e. VM placement and under-loaded host detection. First proposal is for VM placement over hosts. In the research [12], power aware best fit decreasing technique is applied that is defined in Algorithm 1. This is the default algorithm for VM placement in cloudsim. In this algorithm, VMs to be migrated are selected one by one from a list, and their suitability is checked against each and every available host. Power consumption is the factor considered mainly in this method for finding the suitable host for a VM. As shown in step 8 of algorithm 1, power is estimated for each VM for a particular host. Then, based upon the condition applied in step no. 9, VM will be allocated to the specific host only if it has the minimum power consumption on that host.

Algorithm 1: Power Aware Best Fit Decreasing (PABFD)

```
1   Input : ListOfHosts, ListOfVMs  Output: Mapping  of Hosts with VMs
2   ListOfVMs.sortAccordingToDecreasingUtilization()
3   for each VM from ListOfVMs do
4       MinPower ← MAXIMUM
5       HostAllocated ← null
6       for each Host from ListOfHosts do
7           if ( Available Resources of Host are sufficient for VM)
8               Power ← CaluclatePowerEstimation (Host, VM)
9           if (Power < MinPower)
10              HostAllocated ← Host
11              MinPower ← Power
12  if HostAllocated ≠ null then
13    HostToVMMap.add(VM,HostAllocated)
14  return HostToVMMap
```

Fig. 3. Power Aware Best Fit Decreasing (PABFD): Algorithm 1

Another sub problem of detecting the under-loaded host defined in [26] is shown in algorithm 2. In this the host from the hostList, with the zero CPU Utilization and have no VM migrating in or out from the host is considered as the under-loaded host. Upper threshold value in this algorithm is set fixed as 1.

Algorithm 2: Underutilized Host Detection (UHD)

```
1       Input : ListOfHosts, ListOfExcludedHosts  Output: Underutilized Host
2       UnderUtilizedHost ← null
3       MinUtilization ← 1
4       For each Host from ListOfHosts do
5           If (Host not belongs to ListOfExcludedHosts
6                 Utilization ← GetCpuUtilization(Host)
7                 If ( 0 < Utilization < MinUtilization )
8                     If ( no VM is migrating in or out on Host )
9                         MinUtilization ← Utilization
10                        UnderUtilizedHost ← Host
11      return UnderUtilizedHost
```

Fig. 4.Underutilized Host Detection (UHD): Algorithm 2

V. PROPOSED ALGORITHMS

Applying the modifications in the default algorithms, a new VM placement is proposed in algorithm 3. In this modified version, shown in step no. 8, instead of estimating the power for a VM on particular host, the utilization valueis considered for the same searching. Then, in step no. 9, utilization value is checked, if it is less than zero or more than 100% than it is invalid for that host otherwise in step no. 10, if it is a valid utility value, then that VM from VMlist will be allocated to the particular host, conditionally, its utilization value is maximum.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Hence, in this proposal VMs will be allocated in the descending order of their utilization value. This phenomenon would lead to execute the heavy workload first and hence active hosts could be retained more and tasks would be finished fast with high energy consumption rate in less time.

```
Algorithm 3: Utility Aware Best Fit Decreasing (UABFD)
1   Input : ListOfHosts, ListOfVMs  Output: Mapping  of Hosts with VMs
2   ListOfVMs.sortAccordingToDecreasingUtilization()
3   for each VM from ListOfVMs do
4       MaxUtility ← MINIMUM
5       HostAllocated ← null
6       for each Host from ListOfHosts do
7           if ( Available Resources of Host are sufficient for VM)
8               Utility ← CalculateUtilityEstimation(Host, VM)
9           if (Utility ≥ 0 and Utility ≤ 1)
10              if (Utility > MaxUtility
11                  HostAllocated ← Host
12                  MaxUtility ← Utility
13  if HostAllocated ≠ null then
14    HostToVMMap.add(VM,HostAllocated)
15  return HostToVMMap
```

Fig. 5. Utility Aware Best Fit Decreasing (UABFD): Algorithm 3

Than next Algorithm 4, is proposed for under-loaded host detection sub-problem of VM consolidation. In this, an extra metric value defining numbers of VMs on a particular host, is additionally used for selecting the under-loaded hosts. Based upon this host with least CPU utilization or host with less number of VM and have no VM migrating in or out from the host, will be considered.

```
Algorithm 4: Modified Underutilized Host Detection (MUUHD)
1   Input : hostList, excludedHostList  Output: underutilized host
2   underUtilizedHost ← null
3   minUtilization ← 1
4   minMetric ← MAX
5   hostList.sortIncreasingUtilization()
6   for each host in hostList do
7       if host not belongs to excludedHostList
8           utilization ← getCpuUtilization(host)
9           metric ← getSizeOfVmList(host)
10          if utilization >0 and utilization < minUtilization or metric <minMetric
11              if no VM is migrating in or out on host
12                  minUtilization ← utilization
13                  minMetric ← metric
14                  underUtilizedHost ← host
15  return underUtilizedHost
```

Fig. 6. Modified Underutilized Host Detection (MUHD): Algorithm 4

This will reduce the numbers of VM migrations to be performed from underutilized hosts and hence improve the SLA. Applying these two modified algorithms in the existing VM Consolidation (VMC) problem, a new Modified VM Consolidation (MVMC) is generated, which is described in Algorithm 5. In this, at step no. 9 and 13, proposed Algorithm 3 (UABFD) and at step no. 10, algorithm 4 (MUUHD) are called. So, whenever there is a need of new VM or existing VM finish the task assigned to it and get destroyed , VM consolidation happens in the data center.

```
Algorithm 5: Modified VM Consolidation (MVMC)
1   Input : hostList, vmList  Output: Optimized  Allocation of VMs to Hosts
2   excludedHostList ← null
3   overUtilizedHosts ← null
4   vmsToMigrate ← null
5   overUtilizedHosts ← getOverUtilizedHost(hostList) // Call algorithm for finding
    overutilized hosts
6   excludedHostList ← excludedHostList + overUtilizedHosts
7   vmsToMigrate ←getVmsToMigrateFromHosts(overUtilizedHosts) // call algorithm for VM
    selection
8   HostsForMap ← hostList - excludedHostList
9   UtilityAwareBestFitDecreasingAlgorithm(HostsForMap, vmsToMigrate)
10  underUtilizedHosts ← ModifiedUnderUtilizedHostDetectionAlgorithm(
    hostList,excludedHostList)
11  excludedHostList ← excludedHostList + underUtilizedHosts
12  vmsToMigrate ← all vms from underUtilizedHosts which are not in Migration
13  UtilityAwareBestFitDecreasingAlgorithm(underUtilizedHosts, vmsToMigrate)
14  switchoffHosts(underUtilizedHosts)
```

Fig. 7. Modified VM Consolidation (MVMC): Algorithm 5

VI. EXPERIMENT AND PERFORMANCE ANALYSIS

A. Performance Metrics

Various performance metrics can be used for evaluating the effectiveness of the proposed VM Consolidation mechanism and comparing it with the baseline mechanism. These performance metrics are energy consumption, number of VM Migrations, Performance Degradation due to Migration (PDM), SLA violation time per active host (SLATAH) and overall SLA Violation (OSLAV) [12].

- *Energy Consumption*: Physical Machines can be turn on or shut down, depending upon the server consolidation mechanism. Therefore, the total amount of energy consumed by all Physical Machines during running time is defined as Energy Consumption. Lower the value of energy consumption mean less expenditure. So, an algorithm can be considered better if it helps to reduce this metric.
- *Performance degradation due to Migration (PDM):* This parameter can be expressed as:

$$PDM = \sum_{i=1}^{V} \frac{P_{d_i}}{P_{c_i}}$$

Where parameter V defines the number of VMs in data center, P_{d_i} defines the estimated performance degradation due to i^{th} VM Migration and P_{c_i} means total cpu capacity required by i^{th} VM during its lifetime.

- *SLA violation time per active host (SLATAH):* It defines the percentage of SLA Violation time, during which CPU of any active PM experienced the 100% utilization, as defined below:

$$SLATAH = \frac{1}{H} \sum_{j=1}^{H} \frac{T_{C_i}}{T_{A_i}}$$

Where H represents the total number of PMs in data center, T_{C_i} corresponds to the total time during which CPU of i^{th} host stay on 100% utilization mode that leads to SLA Violation and T_{A_i} describe the total time of i^{th} host for which it has remained in active state.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Importance of SLATAH can be seen as the fact that VM cannot be assigned to a host with 100% CPU utilization, which put a VM into waiting for same amount of time, which violates the SLA.

• *Overall SLA Violations (OSLAV)*: Due to overutilization of hosts and migrations of VMs, SLA is occurred, because over-utilized hosts cannot be assigned more VMS and migration of VMs consume the extra CPU cycles for maintain the status of migrations. So the overall SLA violation is resultant of all factors violating the SLA. The overall SLA violations are represented as a percentage and calculated as follows:

$$\text{OSLAV} = \frac{\text{totalRequestedMips} - \text{totalAllocatedMips}}{\text{totalRequestedMips}}$$

Where totalRequestedMips are MIPS requested by all the VMs in data center and totalAllocatedMips is the actually allocated total MIPS to the VMs based on the resource demand.

• *Number of VM Migrations*: This metric shows how often the VMs are migrated among PMs. Too many migrations degrade the performance and increase the SLA violations, and too few migrations might lead to an inappropriate assignment, and hence a balance is required to consider this trade-off.

Table 1. Power Consumption by Two Servers at different load levels in watts

Server	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
HP G4	86	89.4	92.6	96	99.5	102	106	108	112	114	117
HP G5	93.7	97	101	105	110	116	121	125	129	133	135

B. Experimental Setup

To check the effectiveness of algorithms, a simulation model of Cloud Environment is required, which can depict a IAAS model. We have selected CloudSim tool as the simulation platform [26]. By using CloudSim, developers can focus on specific systems design issues that they want to investigate, without getting concerned about details related to cloud-based infrastructures and services [21]. The default VM Consolidation scheme (VMC) defined in CloudSim and modified VM consolidation scheme (MVMC) are applied and compared. Total sixteen experiments has been performed that are combination of four overloaded host detection algorithms i.e. 1) Median Absolute Deviation (MAD), 2) Inter Quartile range (IQR), 3) Local Robust Regression (LRR) and 4) Local Regression (LR) with four basic VM selection techniques i.e. 1) Minimum Migration Time (MMT), 2) Maximum Correlation (MC), 3) Random Selection (RS) and 4) Minimum Utilization (MU). Each experiment name is defined, in such a way so that it can represent the overload detection policy and VM selection policy used in it, e.g. MAD_MMT depicts that MAD overload detection policy is used with MMT VM selection policy. In our experiments 800 heterogeneous hosts are configured, out of which half are HP ProLiant ML110G4 servers (Intel Xenon 3040,dual-core 1860 Mhz,4 GB, 1Gbps) ,and rest are HP ProLiant ML110 G5 servers (Intel Xenon, 3075, dual-core 2660, 4GB, 1 Gbps). We have use the real data of energy consumption defined by SPECpower benchmark [22]-[23]. The energy consumed by these two types of servers on different load levels is mentioned in Table 1 [12]. Reason for selecting this configuration is to measure the effectiveness of proposed Server consolidation algorithms, servers with less resource capacity can be overloaded easily with light workload. Four types of virtual machines corresponds to Amazon EC2 instance types are created as Extra Large instance (2500 MIPS, 3.75 GB) , High CPU Medium instance (2500 MIPS, 0.85 GB), Small instance (1000 MIPS, 1.7 GB) , and Micro instance (500 MIPS, 613 MB).

C. Workload Data

Since simulation can be better applied to real systems data, real workload traces taken from CoMon system are implied in the simulation. The CoMon project [24] developed a monitoring infrastructure for PlanetLab [25]. These data are measurement about CPU utilization from thousands of virtual machines that running on various physical servers around the world. We experiments on the data produced by CPU utilization of more than thousand VMs and these VMs are allotted at more than 500 servers. The characteristics of data are shown in Table 2 [12].

Table 2. Workload Data Characteristics (Cpu Utilization)

Date	Number of VMs	Mean	Standard Deviation	Quartile 1	Median	Quartile 3
03/03/2011	1052	12.31 %	17.09%	2%	6%	15%

D. Experiment Results

Following are the tables, showing the experiment results for five main performance metrics considered in the work. Tables 3 explain the outcomes for both existing and modified scheme, in application with MAD overload detection policy and four VM selection policies. Similarly table 4 describes the results for experiments using IQR and these selection policies. And, then table 5 and 6 also defines the data values obtained from experiments based on LRR and LR policy respectively.

E. Result Analysis

Energy Consumption: The bar graph in fig.8 illustrates the amount of energy consumed in kilowatt hours for all the sixteen experiments performed. It can be analyzed from fig.8 that energy consumption by applying MVMC is less than VMC in all the cases.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

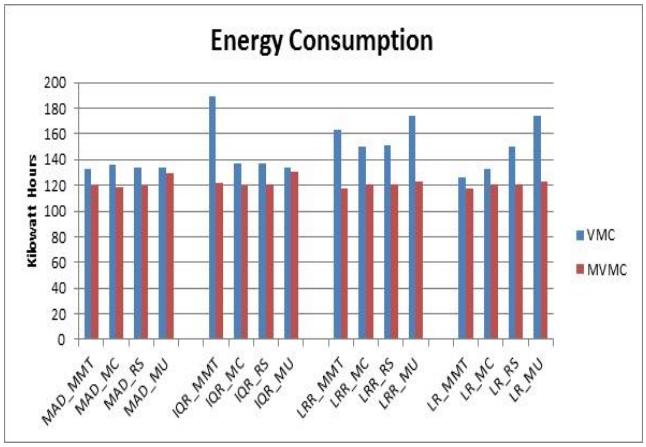

Fig. 8. Comparison for Energy Consumption in MVMC and VMC

No. of VM Migrations: Fig. 9 illustrates the amount of VM migrations occurred while the whole VM consolidation process for all the cases. Again, the proposed scheme MVMC has performed better than the default scheme and reduced the number of VM migrations in situations except in IQR_MU. Reason behind less or no improvements in IQR_MU is selection process. Because, VMsare selected based upon maximum utilization and same concept is applied for the proposed scheme during VM placement.

And, with the modified scheme, LRR_MMT is consuming minimum energy than other experiments. IQR_MMT shows highest decrease in energy consumption by applying proposed approach than default.

Table 3. Performance Metrics with MAD

	Energy Consumption		No. of VM Migrations		PDM		SLATAH		OSLAV	
	VMC	MVMC	VMC	MVMC	VMC	MV MC	VMC	MVMC	VMC	MVMC
MAD_MMT	132.42	119.88	28105	21744	0.1	0.08	4.01	4.12	0.15	0.13
MAD_MC	135.5	118.47	24737	15519	0.13	0.1	4.8	5.42	0.19	0.17
MAD_RS	175.81	119.6	24026	15807	0.11	0.11	7.09	5.71	0.13	0.17
MAD_MU	133.6	129.41	48244	42862	0.13	0.13	5.33	7.68	0.22	0.26

Table 4. Performance Metrics with IQR

	Energy Consumption		No. of VM Migrations		PDM		SLATAH		OSLAV	
	VMC	MVMC	VMC	MVMC	VMC	MVMC	VMC	MVMC	VMC	MVMC
IQR_MMT	188.86	121.94	26476	22540	0.06	0.07	4.96	3.88	0.07	0.12
IQR _MC	137.23	119.75	24646	15537	0.13	0.1	4.75	5.1	0.18	0.14
IQR _RS	136.78	120.67	25312	15618	0.13	0.1	4.27	5.21	0.18	0.15
IQR _MU	134.28	130.65	26162	43903	0.12	0.12	4.58	6.02	0.18	0.21

Table 5. Performance Metrics with LRR

	Energy Consumption		No. of VM Migrations		PDM		SLATAH		OSLAV	
	VMC	MVMC	VMC	MVMC	VMC	MVMC	VMC	MVMC	VMC	MVMC
LRR_MMT	163.15	117.96	27632	10077	0.08	0.02	5.84	4.28	0.14	0.12
LRR_MC	150.33	120.95	23004	7876	0.1	0.04	6.97	4.15	0.16	0.11
LRR_RS	151.15	120.22	23028	7541	0.1	0.05	7.03	4.47	0.16	0.13
LRR_MU	174.24	123.02	29555	15673	0.07	0.03	8.18	7.55	0.17	0.27

Table 6. Performance Metrics with LR

	Energy Consumption		No. of VM Migrations		PDM		SLATAH		OSLAV	
	VMC	MVMC	VMC	MVMC	VMC	MVMC	VMC	MVMC	VMC	MVMC
LR_MMT	126.66	117.96	11624	10077	0.05	0.02	4	4.28	0.15	0.12
LR _MC	132.27	120.95	11587	7876	0.05	0.04	3.98	4.15	0.15	0.11
LR _RS	150.51	120.7	23825	7491	0.1	0.04	7.15	4.36	0.17	0.12
LR _MU	174.24	123.02	29555	15673	0.07	0.03	8.18	7.55	0.17	0.27

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

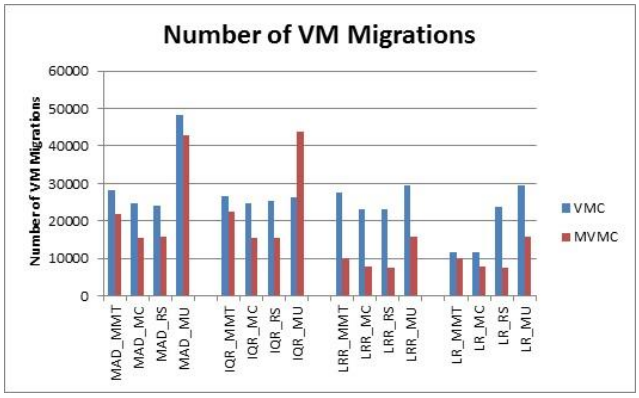

Fig. 9. Comparison for No. of VM Migrations in MVMC and VMC

Performance degradation due to Migration (PDM): As the no. of VM migrations is reduced with MVMC, PDM also decreased with this proposed technique. It is lucid from fig. 10 that withmodified approach, PDM factor is decreased in each experiment result, except IQR_MMT and it shows best improvementfor experiments with LRRand LR overload detection policies. Hence, reduction in performance degradation reflect the better quality service with MVMC.

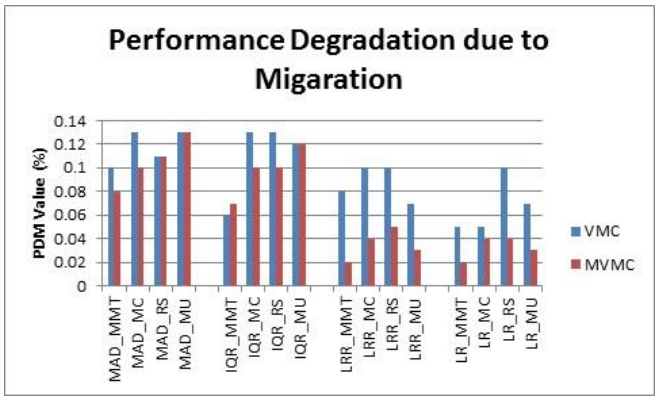

Fig. 10. Comparison for PDM in MVMC and VMC

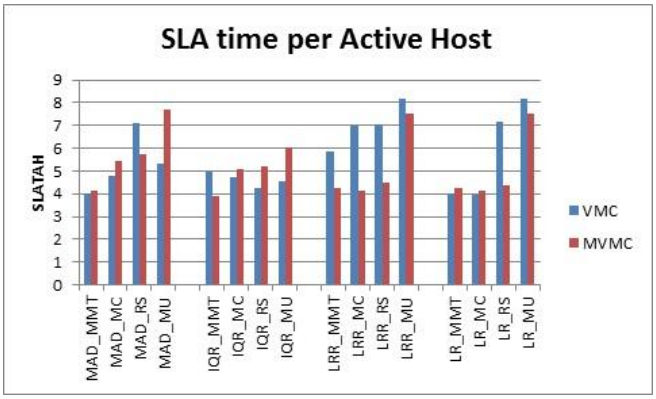

Fig. 11. Comparison for SLATAH in MVMC and VMC

*SLA Time per active host :*Fig. 11, defines the analysis of SLA time per active host applying proposed and default algorithm. And it can be seen that, this parameter has not been improved for all the cases. In MAD policy and IQR policy SLATAH is increased, whereas in LR policy SLATAH is approximately similar in both MVMC and VMC schemes. Only LRR policy is performing well with MVMC and decreasing the SLATAH. Reason for bad SLATAH in three policies is that VMs with heavy load are handled primarily by the host then VMs with lighter loads.So, lighter VMs have to wait for active hosts. Similar to SLATAH, another quality factor i.e. Overall SLA violation (OSLAV), graphed in fig. 12, depicting that OSLAV is also reduced for most of the combinations, mainly in the LRR policy and not showing improvement for MU selection policy in any combination.

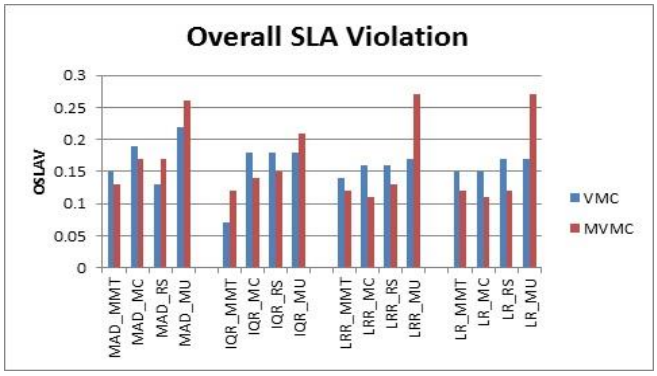

Fig. 12. Comparison for OSLAV in MVMC and VMC

VII. CONCLUSION AND FUTURE WORK

Based upon the modified solutions for VM Placement and Under-loaded host detection problems, proposed algorithm MVMC could make remarkable improvements over the existing algorithm. It managed to get lower power consumption and less number of VM Migration for all the experiments. Also, it results to less amount of PDM in LRR and LR policy. In addition, SLATAH and Overall SLA violation parametersare minimized in LRR policy. So, the proposed heuristic proves that energy consumption can be reduced by increasing the utility of hosts. Because increased utility leads to better energy consumption rate during a short period and hence tasks are completed earlier. As active hosts are retained for more time period, it reduces the overhead of migrating the VMs also.Comparing to the existing approach, this approach has improved the energy consumption with less number of host shutdowns. Results analysis concludes that MVMC performed best with Local Robust Regression policy with combination ofall the four VM selection policies for improving energy consumption and service quality of service providers. As a future work, other heuristics can be applied in VM allocation policy or VM selection policy. Also, some new policies can be introduced to better handle the overloaded or underloaded hosts problem.

REFERENCES

1. S. Lohr, "Google and I.B.M. Join in 'Cloud Computing' Research", Nytimes.com, 2018. [Online]. Available: http://www.nytimes.com/2007/10/08/technology/08cloud.html. [Accessed: 11- Jan- 2019].
2. B. Sosinsky, Cloud computing bible. Hoboken, N.J.: Wiley, 2013.
3. R. Buyya, S. Selvi and C. Vecchiola, Mastering cloud computing. Waltham, MA: Morgan Kaufmann, 2013.
4. R. Karthikeyan and P. Chitra, "Novel heuristics energy efficiency approachfor cloud data center," inAdvanced Communication Control and Computing Technologies (ICACCCT), 2012 IEEE International Conference on.IEEE, 2012, pp. 202–207

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

5. University of Michigan," Computer Center PowerNap Plan Could Save 75 Percent Of Data Center Energy", ScienceDaily, 2018. [Online]. Available: https://www.sciencedaily.com/releases/2009/03/090305164353.htm. [Accessed: 11- Jan- 2019].
6. C. Cassar, "Nationwide, electricity generation from coal falls while natural gas rises. Today in Energy", Eia.gov, 2015. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=23252. [Accessed: 8- Jan- 2019].
7. E. Rubin, A. Rao and C. Chen, "Comparative assessments of fossil fuel power plants with CO 2 capture and storage.", in Greenhouse gas control technologies 7, 2005, pp. 285-293.
8. H. Kim, W. Kim, and Y. Kim, “Experimental study to improve resource utilization and performance of cloud systems based on grid middleware,” Journal of Communication and Computer, vol. 7, no. 12, pp. 32-43, 2010.
9. R. Bianchini and R. Rajamony, “Power and energy management for server systems,” Computer, vol. 37, no. 11, pp. 68-76, 2004.
10. Z. Usmani and S. Singh, “A survey of virtual machine placement techniques in a cloud data center," Procedia Computer Science, vol. 78, pp. 491-498, 2016.
11. C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Warfield, “Live migration of virtual machines,” in Proceedings of the 2nd Conference on Symposium on Networked Systems Design & Implementation-Volume 2. USENIX Association, 2005, pp. 273-286.
12. A. Beloglazov and R. 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, vol. 24, no. 13, pp. 1397-1420, 2012.
13. A. Beloglazov and R. Buyya, “Adaptive threshold-based approach for energy-efficient consolidation of virtual machines in cloud data centers." in MGC@ Middleware, 2010,p. 4.
14. Z. Zhou, Z.-g. Hu, T. Song, and J.-y. Yu, “A novel virtual machine deployment algorithm with energy efficiency in cloud computing,” Journal of Central South University, vol. 22, no. 3, pp. 974-983, 2015.
15. Z. Zhou, Z. Hu, and K. Li, “Virtual machine placement algorithm for both energy-awareness and sla violation reduction in cloud data centers,” Scientific Programming, vol. 2016, p. 15, 2016.
16. M. R. Chowdhury, M. R. Mahmud, and R. M. Rahman, “Implementation and performance analysis of various vm placement strategies in cloudsim,” Journal of Cloud Computing, vol. 4, no. 1, p. 20, 2015.
17. G. Han, W. Que, G. Jia, and L. Shu, “An efficient virtual machine consolidation scheme for multimedia cloud computing,” Sensors, vol. 16, no. 2, p. 246, 2016.
18. A. Verma, G. Dasgupta, T. K. Nayak, P. De, and R. Kothari, “Server workload analysis for power minimization using consolidation," in Proceedings of the 2009 conference on USENIX Annual technical conference. USENIX Association, 2009, pp. 28-28.
19. M. Khoshkholghi, M. Derahman, A. Abdullah, S. Subramaniam and M. Othman, "Energy Efficient Algorithms for Dynamic Virtual Machine Consolidation in Cloud Data Centers", IEEE Access, vol. 5, pp. 10709-10722, 2017.
20. C.F. Kuo, T.H. Yeh, Y.F. Lu, and B.R. Chang, “Efficient allocation algorithm for virtual machines in cloud computing systems,” in Proceedings of the ASE Big Data & Social Informatics 2015. ACM, 2015, p. 48.
21. M. Soni, "The CloudSim Framework: Modelling and Simulating the Cloud", Open Source For You, 2018. [Online]. Available: http://opensourceforu.com/2014/03/cloudsim-framework-modelling-simulating-cloud-environment/. [Accessed: 14- Jan- 2019].
22. Standard Performance Evaluation Corporation, "SPECpower_ssj2008", Spec.org, 2018. [Online]. Available: https://www.spec.org/power_ssj2008/results/res2011q1/power_ssj2008-20110124-00338.html. [Accessed: 14- Jan- 2019].
23. Standard Performance Evaluation Corporation, "SPECpower_ssj2008", Spec.org, 2018. [Online]. Available: https://www.spec.org/power_ssj2008/results/res2011q1/power_ssj2008-20110124-00339.html. [Accessed: 14- Jan- 2019].
24. K. Park and V. Pai, " CoMon: a mostly-scalable monitoring system for PlanetLab ", ACM SIGOPS Operating Systems Review, vol. 40, no. 1, p. 65, 2006.
25. B. Chun, D. Culler, T. Roscoe, A. Bavier, L. Peterson, M. Wawrzoniak, and M. Bowman, “Planetlab: an overlay testbed for broad-coverage services,”ACM SIGCOMM Computer Communication Review, vol. 33, no. 3, pp.3–12, 2003.
26. R. N. Calheiros, R. Ranjan, A. Beloglazovss , De Rose, C. A., &Buyya, R. (2011). CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Software: Practice and experience, 41(1), 23-50.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)