International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 22 Issue 1 – MAY 2016.

EXPLORING STOCHASTIC OPTIMIZATION APPROACH FOR RESOURCE RENTAL PLANNING IN CLOUD COMPUTING

M.Mohanraj[#1], Dr.M.Kannan[*2]

Head Of The Department, Mahendra Engineering College

Abstract- **In cloud computing, the modern cloud data centers are hosting a variety of advanced applications and the IT infrastructure over the recent years because of the demand for computational power infrastructure which are widely used by some of the applications increasing rapidly. Due to the enormous amount of electrical energy consumed by the huge cloud data centers, the operating cost and the emission of carbon dioxide (Co2) produces the high value as a result. In order to reduce the energy consumption and to increase the physical resource utilization in data centers, the most effective way used is a dynamic consolidation of virtual machines (VMs). The main purpose of this paper is to provide a novel method which is used in dynamic virtual machine consolidation. This proposed novel method has outperformed the existing policies in terms of energy consumption, SLA violation and VM migration time by surveying the determination of underloaded hosts, determination of overloaded hosts, selection of VM and placement of the migrating VMs.**

Keywords-- cloud computing, consolidation, energy consumption, SLA violation

I. INTRODUCTION

Cloud computing is a rapidly growing pace in Information and Communication Technology (ICT) industry and delivers three services: 1) Platform as a Service (PaaS), 2) Software as a Service (SaaS) and 3) Infrastructure as a Service (IaaS) under pay-as-you-go model (PAYG). The proliferation of cloud computing, various cloud service providers such as Amazon, Google, IBM and Microsoft have initiated to inculcate increasing numbers of energy greedy data centers for satisfying the resources demanded by customers (e.g. storage and computational resources) [3]. The continuous increase in customers' demands in cloud data centers leads to the high energy consumption of huge data centers which raise a great concern for both governments and service providers to utilize energy more completely. High energy consumption increases the operating costs and the total cost of acquisition (TCA), and also it has an environmental impact in terms of carbon dioxide (CO2) emissions [5]. The hardware infrastructure including servers (Hosts or Physical machine), storage, and network devices in cloud data centers uses the major portion of energy consumption.

At present, virtualization is a technique which is widely used in most cloud data centers. Virtualization allows a creation of multiple instances from a single physical instance of a resource or an application and share among multiple customers among organizations. It achieves by referring a logical name to a physical storage in the data center and providing a pointer to that physical resource when expected. User's resource requests are packed as virtual machines (VMs) and then placed in different hosts based on specific criteria, such as meeting the Service Level Agreement (SLA) requirements between cloud providers and cloud customers, bettering the resources utilization, reducing the number of VM migrations and so on. Each VM in physical machine needs a certain amount of resources like CPU, memory, storage and bandwidth, to support application performance. Virtualization helps to improve resource utilization, scalability, reducing the active users and reduce energy consumption. Moreover, virtualization also helps cloud providers to orderly deploy resources on-demand, which provides an efficient solution to the low energy utilization and flexible resource management. However, worthless VM migrations open extra management cost, e.g., virtual machine reconfiguration, online VM migration, and creation and destruction of VMs, which causes extra energy consumption. Therefore, we attempt to reduce the number of VM migrations to reduce energy consumption.

One method used to reduce energy consumption is a dynamic consolidation of VMs. Here the VMs in cloud data centers are periodically reallocated which minimizes the number of active hosts using live migration. Live migration transfers a VM between hosts without suspension and with a short downtime. Nevertheless, application performance should also be considered when placing these VMs. That is to say, if we keep all VMs on a single server, the server's performance will be degraded due to its limited physical resources. In that case, the condition for migration of VM is that if the resource utilization of the PM exceeds a certain value, VMs on the PM cannot meet the SLA between providers and users. Therefore, we set an upper threshold of CPU utilization to avoid overloaded hosts and maintain the SLA agreement.

Another method to reduce energy wastage is to turn off PMs with low utilization rate. The average utilization of the whole data center in Google [10] is only 30%, which encourages us to set a low threshold. If a host's resource utilization is lower than the threshold, then all the VMs on that PM are migrated and now the unused host is turned off, resulting in fewer active hosts of which each one is highly utilized. The process of VM dynamic consolidation involves CPU utilization threshold setup, the VMs selection, and the VM placement.

Dynamic consolidation of virtual machines is an effective technique which turns off idle or underutilized servers to reduce the power utilization in the data center. However, achieving the desired level of Quality of services (QoS) between user and data center is critical. Therefore dynamic consolidation of virtual machines can redeem energy at the same time maintaining an acceptable QoS. Because VM placement is an NP-hard problem and the workload is unstable and unpredictable, it makes dynamic VM consolidation, even more complicated. So, VM dynamic consolidation is split into four subproblems (1) Determination of overloaded host (2) Determination of under loaded host (3) VM selection and (4) VM placement which reduces the energy and improves utilization of resources without compromising SLA requirement.

The rest of the paper is organized as follows. We discuss the target system model in Section II. In this section firstly introduces power and energy model, and SLA violation metrics for the data center. Section III presents VM consolidation for data centers especially heterogeneous physical nodes. Finally, we conclude in Section IV.

II. RELATED WORKS

The authors in [7] have proposed a building system and guideline for vitality effective distributed computing went for the advancement of vitality productive provisioning of cloud assets, while meeting QoS prerequisites characterized by SLA. The VM allotment issue is partitioned into two sections: the first part is the confirmation of new demands for VM provisioning and setting the VMs on PMs, while the second part is the improvement of the current VM allotments. The primary part is demonstrated as a receptacle pressing issue and fathomed it by MBFD calculation in which sort all VMs in diminishing request of their present CPU usages, and allots each VM to a PM that gives minimal expansion of force utilization because of this portion. Also, the improvement of the current VM assignments is completed in two stages: 1) select VMs that should be moved, 2) the picked VMs are put on the PMs utilizing the MBFD calculation.

The authors in [10] have led focused investigation and demonstrated aggressive proportions of ideal online deterministic calculations for the single VM movement and element VM combination issues. They have isolated the issue of element VM union into four sections interestingly including: (1) deciding when a host is considered as being over-burden; (2) deciding when a host is considered as being underloaded; (3) choice of VMs that ought to be relocated from an over-burden host; and (4) finding another position of the VMs chose for movement from the over-burden and underloaded has. They have proposed novel versatile heuristics for all parts. They have utilized PABFD calculation to tackle asset distribution issue.

The authors in [11] have proposed various VM solidification calculations for cloud server farm vitality decrease considering auxiliary components, for example, racks and system topology of the server farm hidden the cloud. All the more accurately, the cooling and system structure of the server farm which facilitating the PMs are considered while merging the VMs. Thusly, less racks and switches are utilized, without trading off the administration level understandings, so sit out of gear directing and chilling hardware can be turned to decrease the vitality utilization.

The authors in [12] have proposed productive solidification calculations which can lessen vitality utilization and in the meantime the SLA infringement at times. A proficient SLA-mindful asset allotment calculation was presented that considers the exchange off between vitality utilization and execution. Their proposed asset portion calculation considers both PM usage and connection between's the assets of a VM with the VMs present on the PM. Besides, a novel calculation for determination of underloaded PMs was proposed during the time spent asset administration in cloud server farms considering PM CPU usage and number of VMs on the PM.

The main drawback of all these works is that they consider either vitality utilization or SLA infringement as their primary target and build up their answers in light of that. Be that as it may, this paper considers all objectives including vitality utilization, SLA infringement, and number of VM relocations in the meantime utilizing novel multi-criteria calculations which prompts outstanding enhancements in yield results.

III. TARGET SYSTEM MODEL

The target system model consists of cloud data centers with heterogeneous resources which serve different applications for various users and runs multiple heterogeneous VMs on data center nodes. As a result, each PM has dynamic mixed workload. VMs and PMs are characterized with parameters including CPU computation power (Millions Instructions Per Second-MIPS), Disk capacity, Network bandwidth, and RAM. The target system model [1] is depicted in Fig. 1. This model has two important parts: the central manager and the agents. In a cloud data center, the central manager acts as a resource scheduler which allocates virtual machines to the available hosts in the data center based on specific criteria. Also, it manages VMs by resizing according to their resource needs and makes decisions about when and which VMs should be migrated from PMs. Next important part is the agents which are incorporated on hypervisors. The agents and the central manager are connected through network interfaces. Agents have the responsibility for monitoring PMs besides transferring accumulated information to the central manager. Hypervisor performs actual resizing and migration of VMs besides the shift in power modes of the PMs. Here, to provide FT (fault tolerance) and HA (High Availability) capabilities, the central manager runs on any of the VM instead of a PM.

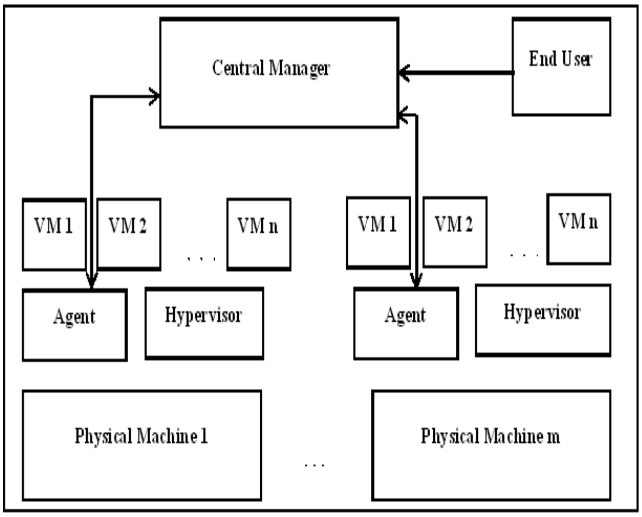

Figure 1. System model

A. Power and energy models

In cloud data centers, server's power utilization and CPU utilization has a linear relationship [7, 8]. Because of the proliferation of multi-core CPUs with utilization technique, CPU is not the only power consumer in data centers [2].

Based on the system that performs work, power and energy are defined. Power is defined as the rate at which the system performs the work, although energy is defined as the total amount of work performed over a period of time by the system. The measurement of power and energy are watts (W) and watt-hour (Wh), respectively. The technique of switching the idle server to sleep mode justify the reduction of the total power consumption.

For this work, power model defined in (1).

$$P(u) = P_{idle} + (P_{busy} - P_{idle})\, u, \qquad (1)$$

Where, P is the estimated power consumption of the system, P_{busy} is the server's power consumption when it is fully utilized, and u is the current CPU utilization, P_{idle} is the power consumption by an idle server.

Due to the variability in workload, the CPU utilization may change over time. So, the CPU utilization is defined as the function of time and is represented as u(t). Therefore, the total energy consumption by a physical node in a data center can be defined in (2).

$$E(t) = \int_t P(t)dt \qquad (2)$$

B.SLA violation metrics:

In cloud data centers, QoS requirements are commonly formalized in the form of SLAs. SLAs determined in terms of characteristics such as maximum response time or minimum throughput delivered by the system [2]. As these characteristics can vary for different applications, workload independent metric can be used to evaluate the SLA delivered to any VM deployed in an IaaS such as OTF (Overload Time Fraction) metric defined in [6]. In this study, we use the SLA Violation (SLAV) metric introduced in [2] as defined in Eq. (3) which is composed of multiplication of two metrics: the SLA violation time per active host (SLATAH) and performance degradation due to migration (PDM) as defined in Eq. (4).

$$SLAV = SLATAH \times PDM \qquad (3)$$

$$SLATAH = \frac{1}{N}\sum_{i=1}^{N}\frac{T_{s_i}}{T_{a_i}}, PDM = \frac{1}{M}\sum_{j=1}^{M}\frac{C_{d_j}}{C_{r_j}} \qquad (4)$$

where T_{Si} is the total time during which the host i has experienced the utilization of 100%; T_{ai} is the total time during which the host i has been in the active state; N is the number of PMs; C_{dj} is the estimate of the performance degradation of the VMj caused by migrations which are estimated as 10% of the average CPU utilization in MIPS during all migrations of the VMj; C_{rj} is the total CPU capacity requested by the VMj during its lifetime; and M is the number of VMs.

IV. PROPOSED WORK

In cloud data centers, an effective way to improve energy efficiency is dynamic VM consolidation as a dynamic control procedure. The main aspect of this procedure is to optimize resource utilization and energy-performance trade-off inside cloud data center.

A.Determination of Under loaded Host:

The TOPSIS Available Capacity, Number of VMs, and Migration Delay (TACND) policy is a multi-criteria decision-making method that takes advantage of Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) and considers three criteria depicted in Table 1. TACND policy estimates the scores for all the hosts in the system that is a candidate for underloaded hosts and selects a host with the highest score as underloaded.

TACND policy selects the host as underloaded if the conditions exist: (1) the selected host has the least available capacity, (2) the selected host has the least number of virtual machines and (3) the selected host has the least migration delay of all the VMs.

TABLE 1: Considered criteria in TACND policy

No	Notation	Parameter	Description	Benefit/Cost
1	AC	Available capacity	Available resource capacity of a host	Cost
2	NV	Number of VMs	Number of VMs on a host	Cost
3	MD	Migration delay	The delay incurred due the migration of all VMs on host	Cost

B.VM Placement:

TOPSIS Power and SLA-aware Allocation policy for resource allocation is a multi-criteria algorithm that takes the advantages of TOPSIS method by considering five criteria depicted in Table 2 for its decision process [4]. This policy computes the scores for all the hosts that are a candidate for hosting a VM and selects the host with the highest score as the destination host. In TPSA policy, the criteria considered can have either benefit or cost type. The benefits type has more value for criteria and the cost type has lowered value for criteria, and the closer is the answer to the optimum point.

TABLE 2: Considered criteria in TPSA policy

No	Notation	Parameter	Description	Benefit/Cost
1	PI	Power increase	Power increase of allocating a VMs on a host	Cost
2	AC	Available capacity	Available resource capacity of a host	Benefit
3	NV	Number of VMs	Number of VMs on a host	Cost
4	RC	Resource correlation	Resource correlation of a VM with the VMs on a host	Cost
5	MD	Migration delay	The delay incurred due the migration of all VMs on host	Cost

TPSA computes the score of hosts so that the following conditions exist in the answer: (1) the selected host has the least power increase, (2) the selected host has the most available resource, (3) the selected host has the least number of VMs, (4) VMs on the selected host have the least resource correlation with the VM to be allocated, and (5) the selected host has the least migration delay of the VM.

By selecting host with least number of VMs, higher the probability that the VM has a lower number of competent for the shared resources which leads to the reduction in SLA violations. Moreover, the host with the highest available capacity ensures the higher probability of allocation of the resources for the requested VMs and also consequently reduces the SLATAH metric. Based on the idea given in [9], is that the higher the resource correlation among the applications which use the same resources on an oversubscribed server, then higher the probability of the server being overloaded. According to this idea, the host is selected such that the allocated VM has the least resource correlation with the VMs on that host. Also, considering the migration delay of the VM to be allocated on the selected host, this lowers the SLA violation during the migration process. Also, due to smart decisions based on multiple criteria and omission of migrations with longer delays, it reduces the number of VM migrations. In TPSA method, the chosen destination host has the shortest distance from the ideal positive point (PM+) and the farthest distance from the ideal negative point (PM-). PM+ and PM- are formed as a composite of best and worst values of considered criteria for all hosts. Distance from each of these poles is measured in the Euclidean distance.

V. SIMULATION RESULT

Since the target system is generic cloud computing environment, it is vital to analyze it on a large-scale virtualized data center infrastructure. The simulation uses CloudSim toolkit which provides the desired environment. The infrastructure setup has real configurations of cloud computing comprising a data center with 800 installed heterogeneous hosts and five types of VMs (Amazon EC2 VM types).

A.PERFORMANCE METRIC

In order to assess the simultaneous minimization of energy, SLA violation, and number of VMs' migrations, we use a new metric which is denoted as Energy-SLAV-Migration (ESM) in (5)

$$ESM = Energy * SLAV * MigrationsCount \quad (5)$$

A.ENERGY CONSUMPTIONS

The Fig2.shows the energy comparison between LR/MMT and EO policy. The proposed EO policy has better performance over energy consumption.

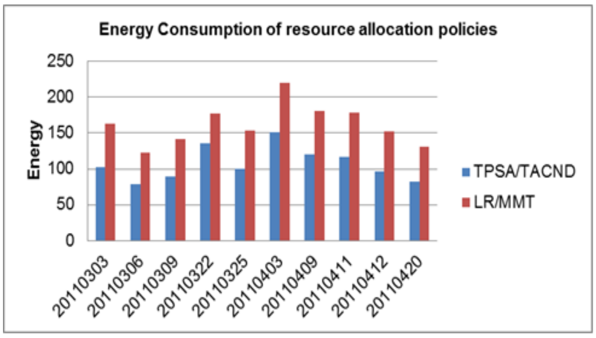

Figure 2. Energy Consumptions

B.NUMBER OF VM MIGRATIONS

The Fig3 shows the number of VM migrations comparison between LR/MMT policy and EO policy. The proposed EO policy has reduced number of VM migrations compared to LR/MMT policy.

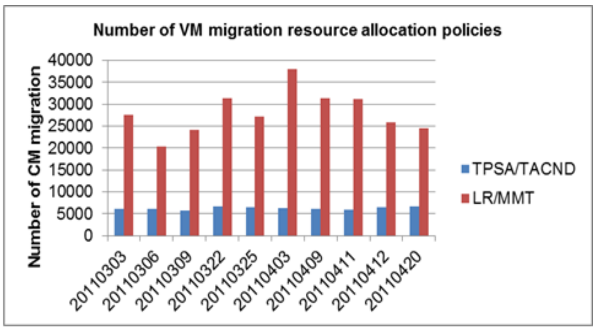

Figure3. Number of VM Migrations

C.SLA VIOLATION

The Fig 4 shows the SLA violation comparison between LR/MMT and EO policy. The EO policy has significant improvement when compared to LR/MMT policy.

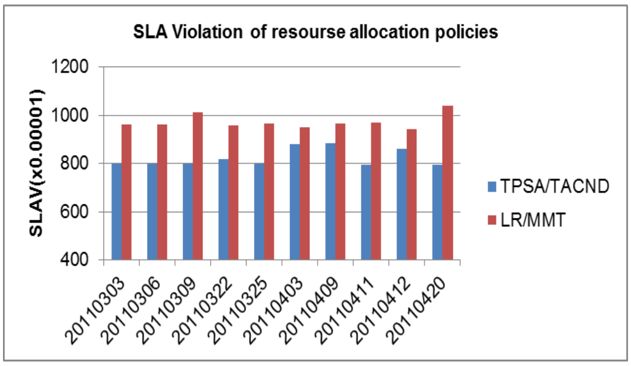

Figure4. SLA Violations

C. ENERGY-SLAV-MIGRATION

The Fig 5.shows the ESM metric comparison between LR/MMT and TPSA/TACND policies. TPSA/TACND policy provides better performance.

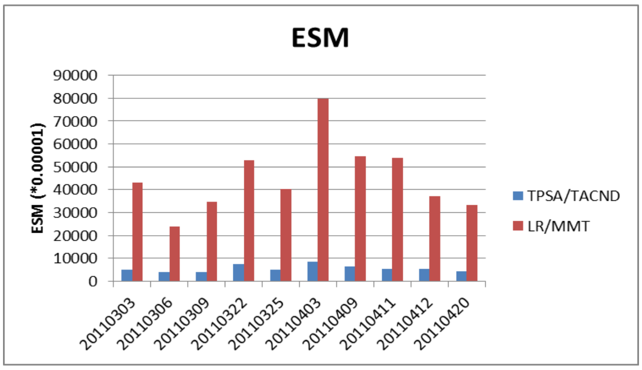

Figure5. Energy-SLAV-Migration

VI. CONCLUSION

Development of huge cloud data centers all around the world leads to the enormous energy consumption and a steady increase in carbon emissions. It is necessary to reduce the energy consumption without SLA violation and performance degradation in virtualized data centers. The energy consumption and SLA violation can be reduced by performing the energy-efficient resource management strategies like dynamic VM consolidation which switch off the idle hosts into sleep mode. A new approach for dynamic VM consolidation was proposed which provides an efficient resource management procedure across data centers for reducing the energy consumption, SLA Violation and number of VM migration. This policy gathers all the VMs to be migrated from either over-utilized or under-utilized PMs in the VM migration lists and allocating the resource at once using TPSA policy which is a multi-criteria algorithm. More precisely, the proposed approach provides the maximum user satisfaction with reducing the energy consumption, SLA violation, and number of VM migrations in cloud data centers.

REFERENCES

[1] EhsanArianyan, Hassan Taheri, Saeed Sharifian. "Novel energy and SLA efficient resource management heuristics for consolidation of virtual machines in cloud data centers".ComputElectrEng 2015.

[2] Beloglazov A, Buyya R. "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in Cloud data centers". ConcurrComput: PractExper 2012;24:1397–420.

[3] Gao Y, Guan H, Qi Z, Song T, Huan F, Liu L. "Service level agreement based energy efficient resource management in cloud data centers". ComputElectrEng 2013.

[4] Chen C-T. "Extensions of the TOPSIS for group decision-making under fuzzy environment". Fuzzy Sets Syst 2000;114:1–9.

[5]Beloglazov A, Buyya R, Lee YC, Zomaya A. "A taxonomy and survey of energy efficient data centers and cloud computing systems". AdvComput 2011;82:47–111.

[6]Beloglazov A, Buyya R. "Managing overloaded hosts for dynamic consolidation of virtual machines in cloud data centers under quality of service constraints". ParallDistribSyst IEEE Trans 2013;24:1366–79.

[7] Beloglazov A, Abawajy J, Buyya R. "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing". Future GenerComputSyst 2012;28:755–68.

[8] Kusic D, Kephart JO, Hanson JE, Kandasamy N, Jiang G. "Power and performance management of virtualized computing environments via look ahead control". ClustComput 2009;12:1–15.

[9] Verma A, Dasgupta G, Nayak TK, De P, Kothari R. "Server workload analysis for power minimization using consolidation". In: Proceedings of the 2009 conference on USENIX annual technical conference, USENIX Association, 2009. p. 28–8.

[10] Barroso L A, Hölzle U. "The datacenter as a computer: an introduction to the design of warehouse-scale machines". Synthesis lectures on computer architecture, 2009, 4(1): 1–108

[11] Esfandiarpoor S, Pahlavan A, Goudarzi M, (2014), 'Structure-aware online virtual machine consolidation for datacenter energy improvement in cloud computing' In ComputElectr Eng.

[12] Horri A, Mozafari MS, Dastghaibyfard G., (2014), 'Novel resource allocation algorithms to performance and energy efficiency in cloud computing' In J Supercomput, Vol. 69, pp. 1445-61.