International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 5, Issue 7, July 2016)

Energy Efficient Dynamic Resource Scheduling for Cloud Data Center

Rakesh Kumar Vishwarkarma[1], Dr. Syed Imran Ali[2], Anidra Katiyar[3]

[1]Research Scholar, [2]Head, CSE, [3]Asst.Professor, Sagar Institute of Science, Technology & Research

***Abstract*—Massive-scale cloud data centers host many applications and comprise of millions of servers thus consuming more power than ever. In order to efficiently manage the power usage of these data centers, Green computing offers schemes like load balancing across physical machines, live migration of virtual machines and Sever Consolidation which aims at minimizing the number of Active Physical Machines (APM). Server consolidation is a result of Virtual Machine (VM) scheduling which involves— *VM selection, VM placement* and *VM placement re-optimization.* In this paper, we present a VM placement optimization technique used in green cloud, particularly based on the classical problem of Bin Packing. Bin packing is inspired by the NP-Hard knapsack problem and reduces the total number of Active Physical Machines (APM). Further these placements are optimized using Rank based VM scheduling algorithm. The proposed approach subsequently reduces the energy consumption and provides improved server consolidation.**

***Keywords*— Cloud Computing, Data Center, Load Balancing, Server Consolidation, Virtual Machine Scheduling.**

I. Introduction

Cloud Computing has come up as a new paradigm for large scale distributed computing. Based on a pay-as-you-go model, it relieves an organization or individual from actually buying a real cloud service. Cloud users can easily access the resources offered by cloud providers, based on their needs. The three existing cloud service models being: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). For instance, Amazon Webb Services is a well-known Iaas offering an open source Elastic Compute Cloud (EC2) for users to perform their computations. In the same way, Salesforce and Google's AppEngine are public clouds for SaaS and PaaS respectively.

As per research, the massive-scale data centers are responsible for causing 2 percent of global carbon emission and also, 1.3 percent of world's electricity consumption is by data centers, which is being estimated to grow to 8 percent by the year 2020. Generally, the servers only operate at 10-50 percent of their total capacity, which clearly indicates the considerable amount of energy loss. So the major concern of the research community is the efficient power management of data center resources in an effective and feasible manner.

In order to achieve such kind of power management efficiently and economically, Green Computing comes to the rescue.

Green computing successfully meets the business requirements while delivering energy-efficient, cost-effective, flexible, stable and secure solutions. Companies like Amazon with its EC2 (Elastic Compute Cloud) are switching towards greening their data centers, i.e. trying to use least numbers of actively running servers. This calls for appropriate solutions like Virtualization, Virtual Machine Migration, Load Balancing, Server Consolidation and Dynamic Resource Management. In particular, cloud data center resources (e.g. CPU, memory, network bandwidth and storage) need to be allocated with equal focus on reduction of energy usage as on satisfaction of Quality of Service (QoS) requirements specified by users via Service Level Agreements (SLAs).

A. Server Consolidation

Virtualization enables sharing of computer hardware by partitioning the computational resources. In a datacenter server, often many services only need a small fraction of the total available resources. This can lead to a scenario in which several virtualized servers operate and consume a lot more space and resources than expected. This problem is referred to as 'server sprawl'. To prevent such wastage of resources and to increase the energy efficiency, those services can be virtualized and run in a virtual machine. In consideration of this concept, multiple virtual machines can run on a single physical server. This approach can be followed in response to server sprawl and is termed as Server Consolidation. The reduction in the number of servers has a noticeable benefit for data centers by improving system availability, reducing infrastructure complexity and of course saving energy and money.

Consolidation can save energy by reducing the power consumption and thus resulting in decrease of overall operational costs for data center administrators. All this is achieved via Live Migration, which is the process of transferring the memory, storage and network connectivity of a running application or virtual machine from its current host to destination host with no impact on the VM availability to users.

If a certain server is under-utilized, the aim of server consolidation is to identify that server and migrate all of its virtual machines to other active physical machines which have their utilizations above a certain threshold which is already set. Thus the under-utilized server is freed up & so it can be switched to sleep/idle mode to save power. With the help of live migration of VMs Server Consolidation aims at achieving—least possible number of Active physical machines, packing these Active PMs with VMs as tightly as possible to increase energy efficiency and switching the non active PMs. Although it is hard to come up with a green computing approach which is energy-efficient, dynamic, high in performance and to strike a full balance between allocated resource with the minimum migration overhead. Some or the other trade off ought to exist. But one can design resource allocation approach which can support green computing using dynamic allocation of data center resources. This can be done using Bin Packing algorithm, which can also avoid overloading at servers.

B. Bin Packing

The classical problem of bin packing is similar to the NP-Hard Knapsack problem which consists of a number of weighed objects which need to be filled into a sack of given capacity. To model this problem as a resource allocation algorithm, we consider each item as a Virtual Machine (VM) to be tightly packed in minimum number of bins, each considered as a Physical Machine (PM). The bin packing problem is NP hard. The quality of a polynomial time approximation algorithm, A is measured by its approximation ratio, R (A) to optimal algorithm, OPT:

$$R(A) = \lim_{n\to\infty} sup_{OPT(L)=n} \frac{A(L)}{OPT(L)} \quad (1)$$

Where A (L) is the number of bins used under the algorithm A, OPT (L) is the number of bins used under the optimal algorithm OPT and L is the list of input sequence.

II. Related Work

In this section we throw some light on the existing VM scheduling techniques which aim at improving server consolidation:

W. Song et al. [1] formulated a dynamic resource allocation algorithm based on Bin packing which optimizes the number of actively running services using virtualization. They designed a slight variation of the Relaxed Online Bin Packing algorithm [23] and named it as VISBP (Variable Item Size Bin Packing).

They implemented it using extensive trace-driven simulation and also compare it with three well known server consolidation algorithms: the Black Box & Gray box algorithm [12], the Vector Dot algorithm [1] and the Offline-Bin Packing algorithm [7]. The core of VISBP is its ability to handle the change in size of an item (VM) at runtime. This “change” operation supports an on demand, dynamic resource allocation. VISBP excels in load balancing and hot-spot detection & normal but it violates service level agreements to an extent and also there is need to improve the VM to PM ratio.

C. Ghribi et al. [2] investigated two exact energy-efficient Virtual Machine scheduling algorithms to achieve an optimal VM management comprising optimal placement and migration of VMs. *Y. Zhang et al.* [4], addressed the problem that the existing Bin packing heuristics, whether single dimensional or multi-dimensional, do not dig much into the resource requirement heterogeneity of VMs. So they proposed several algorithms which are heterogeneity aware..

J. Dong et al. [5] applied a few constraints such as network link capacity and Physical Machine (PM) size on scheduling of Virtual Machines (VM) via a two-staged VM scheduling algorithm. *M. Li et al.* [6], observed that Bin packing problems do not consider overhead of resource consumption and so they interpreted overhead as the difference between the resource consumption of a server and all its encompassed VMs. Although it was observed that overhead estimation should depend on the application type and marked as a future enhancement. *S. Srikantaiah et al.* [11] tried to optimally reduce energy consumption problem as a multi-dimensional Bin Packing problem, mainly focusing on the disk and CPU as important resources. This study proves fruitful by throwing light on the key aspects of energy-performance relationship and uncovers many research issues but memory is not considered as a primitive resource. *A. Beloglazov et al.* [13] propose a novel dynamic virtual machine consolidation technique which assures Quality of Service provisioning by meeting Service Level Agreements.

One can easily observe that bin packing based algorithms can substantially decrease the required amount of active servers (APMs) and so reduces the required energy usage. Many techniques are based on stochastic bin packing and make use of historical data for optimization of VM allocation. The VM resizing techniques are highly efficient in terms of cost reduction. Although these techniques may seem fine from outside, there exist some or the other kind of trade- offs when deeply analyzed. The strict server consolidation techniques impose an extra overhead of migration.

Owing to the workload-variability in applications, there is a need to constantly optimize the VM allocation algorithm.

III. Proposed Methodology

A. System Overview

The architecture of a virtualized data center on which our approach can be applied is explained herein. A Hypervisor or a Virtual Machine Monitor (VMM) is run by each server or PM. This VMM like Xen, supports virtualization by creating multiple instances of independent runtime environments on a single computer. Therefore, each PM (host or server) has multiple VMs running on it. All VMs are loaded by different cloudlets or tasks given by user. These cloudlets may be any application such as remote desktop, mail server, web server, Map/Reduce, etc. All this information about the cloud data center, its characteristics and components, is sent to a centralized VM Scheduler, which implements green computing by adjusting the VM to PM mapping. The VM Scheduler keeps a track of the available and allocated VMs, the history of VM resource demand, history of load on PM and the currently assigned VM-PM mappings.

Virtual Machine allocation problem works mainly in 2 steps:

(1) VM Placement: When new requests arrive, VMs are provisioned and placed onto available hosts (PMs).

(2) Placement Optimization: Re-adjustment of current placements based on performance parameters.

(1) VM Placement: The first sub-problem is solved using MBFD algorithm where CPU utilization is the prominent resource. MBFD sorts the CPU utilizations of all VMs in decreasing order and places a VM on such a host whose power consumption is increased the least due to this placement. MBFD algorithm's pseudo-code is as under:

Algorithm 1: Modified Best Fit Decreasing (MBFD)

```
Input: hostList, vmList Output: allocation of VMs
vmList.sortDecreasingUtilization()
foreach vm in vmList do
  minpower ← MAX
  allocatedHost ← NULL
  foreach host in hostList do
    if host has enough resource for vm then
      power ← estimatePower(host, vm)
      if power < minpower then
        allocatedHost ← host
        minpower ← power
  if allocatedHost ≠ NULL then
    allocated vm to allocatedHost
return allocation
```

(2) Placement Optimization:

Once the initial VM placement is done, we step towards optimizing it. Our algorithm deals with the optimization of VM resources through overload avoidance. This subproblem, may also be called as VM selection problem, involves two major selections:

(i) Selecting *which* VM(s) should be migrated

(ii) Selecting *where (on which PM)* this VM(s) should be migrated.

The second sub-problem is solved Rank based VM scheduling algorithm (RBVM). The proposed algorithm states: "Select a VM whose CPU time is maximum and assign it to the PM which has highest rank among all the resources". .The proposed RBVM algorithm not only considers the maximization of resource utilization of IaaS cloud environment but will also, in a way, minimizes the carbon footprint scheduling them on the virtual machine selected based on their rank. The rank of virtual machine is calculated by considering the characteristics of virtual machines.

It is important to note that the existing algorithm uses MMT algorithm [13] for placement optimization.

IV. Implementation And Performance Analysis

The target system being an IaaS cloud computing environment it calls for employment of a large-scale infrastructure of a virtualized data center which is quite crucial and not easy to manage. Thus we opt for simulation. This work makes use of GridSim as the simulator, since it supports most of the characteristics of a good simulation environment.

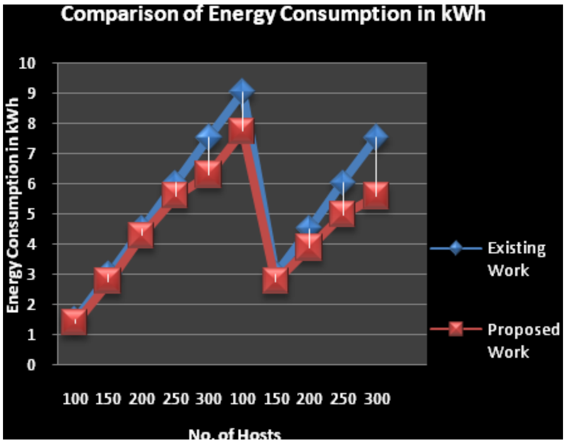
Different Simulation parameters considered for the simulation are defined in Table-I.

Table I.
Simulation Parameters

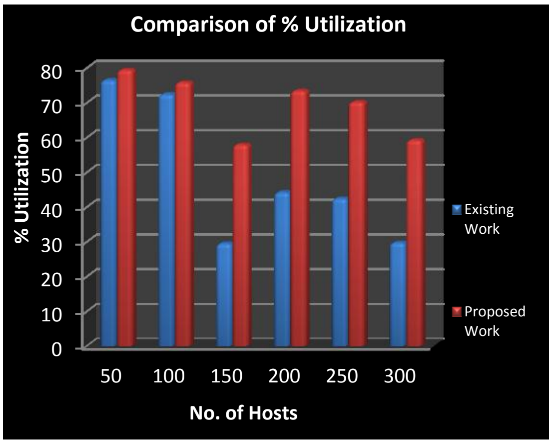
Virtual Machine Parameters	
Parameter Name	***Value***
Number of VMs	50 to 300
RAM	Varying size from 256 to 2048 Minimum 256 MB to Maximum 2048 MB, Varying by 256 MB per Virtual Machine
MIPS	Varying size from 1000 to 2000 Minimum 1000 MIPS to Maximum 2000 MIPS, Varying by 100 MIPS per Virtual Machine
Bandwidth	1000
Number of PMs	50 to 100
Physical Memory	10000 MB

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 5, Issue 7, July 2016)

The results and the performance of the proposed VM Scheduling Algorithm for server consolidation is depicted with the help of graphs 1 and 2.



Graph 1. Comparison of energy consumed by existing and proposed algorithms



Graph 2. Comparison of resource utilization by existing and proposed algorithms

It can be clearly seen that energy consumption is further minimized from 1.508 KWh to 1.389 KWh and the percentage of overall resource utilization has scaled up by 3% on an average.

V. Conclusion

The data center servers and their resources need to be managed using energy-efficient virtual machine provisioning and management techniques. The paper delved with the currently existing virtual machine scheduling techniques mainly focusing on the use of the classical Bin Packing heuristic. Owing to the workload-variability in applications, there is a need to constantly optimize the VM allocation algorithm. The novel technique of selecting migration candidates on the basis of ranks proves to be more energy efficient and helps to utilize the available cloud more judiciously.

As a future work the algorithm will be tuned and implemented with different sets of parameters using different VM placement approach. The algorithm can also be scaled up to accommodate more number of machines running on a single server. The same algorithms can be implemented in the real cloud scenario so that it can be critically analyzed and more parameters can be taken into consideration.

References

[1] Song, Weijia, et al. "Adaptive resource provisioning for the cloud using online bin packing." Computers, IEEE Transactions on 63.11 (2014): 2647-2660.

[2] Ghribi, Chaima, Makhlouf Hadji, and Djamal Zeghlache. "Energy efficient vm scheduling for cloud data centers: Exact allocation and migration algorithms."Cluster, Cloud and Grid Computing (CCGrid), 2013 13th IEEE/ACM International Symposium on. IEEE, 2013.

[3] Alahmadi, Ahmed, et al. "Enhanced First-Fit Decreasing Algorithm for Energy-Aware Job Scheduling in Cloud." Computational Science and Computational Intelligence (CSCI), 2014 International Conference on. Vol. 2. IEEE, 2014.

[4] Zhang, Yan, and Nayeem Ansari. "Heterogeneity aware dominant resource assistant heuristics for virtual machine consolidation." Global Communications Conference (GLOBECOM), 2013 IEEE. IEEE, 2013.

[5] Dong, Jiankang, Hongbo Wang, and Shiduan Cheng. "Energy-performance tradeoffs in IaaS cloud with virtual machine scheduling." Communications, China 12.2 (2015): 155-166.

[6] Li, Mingfu, et al. "Unveiling the resource consumption overhead of virtual machine consolidation in data centers." Global Communications Conference (GLOBECOM), 2012 IEEE. IEEE, 2012.

[7] Bobroff, Norman, Andrzej Kochut, and Kirk Beaty. "Dynamic placement of virtual machines for managing sla violations." Integrated Network Management, 2007. IM'07. 10th IFIP/IEEE International Symposium on. IEEE, 2007.

[8] Wang, Meng, Xiaoqiao Meng, and Li Zhang. "Consolidating virtual machines with dynamic bandwidth demand in data centers." INFOCOM, 2011 Proceedings IEEE. IEEE, 2011.

[9] Chen,Ming,et al. "Effective VM sizing in virtualized data centers." Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on. IEEE, 2011.

[10] Srikantaiah, Shekhar, Aman Kansal, and Feng Zhao. "Energy aware consolidation for cloud computing." Proceedings of the 2008 conference on Power aware computing and systems. Vol. 10. 2008.

[11] Wood, Timothy, et al. "Sandpiper: Black-box and gray-box resource management for virtual machines." Computer Networks 53.17 (2009): 2923-2938.

[12] Beloglazov, Anton, and Rajkumar Buyya. "Adaptive threshold-based approach for energy-efficient consolidation of virtual machines in cloud data centers."Proceedings of the 8th International Workshop on Middleware for Grids, Clouds and e-Science. Vol. 4. ACM, 2010.

[13] Li, Kangkang, Huanyang Zheng, and Jie Wu. "Migration-based virtual machine placement in cloud systems." Cloud Networking (CloudNet), 2013 IEEE 2nd International Conference on. IEEE, 2013.

[14] Bein, Doina, Wolfgang Bein, and Swathi Venigella. "Cloud storage and online bin packing." Intelligent Distributed Computing V. Springer Berlin Heidelberg, 2012. 63-68.

[15] Singh, Aameek, Madhukar Korupolu, and Dushmanta Mohapatra. "Server-storage virtualization: integration and load balancing in data centers."Proceedings of the 2008 ACM/IEEE conference on Supercomputing. IEEE Press, 2008.

[16] Ho, Yufan, Pangfeng Liu, and Jan-Jan Wu. "Server consolidation algorithms with bounded migration cost and performance guarantees in cloud computing."Utility and Cloud Computing (UCC), 2011 Fourth IEEE International Conference on. IEEE, 2011.

[17] Lin, Ching-Chi, Pangfeng Liu, and Jan-Jan Wu. "Energy-efficient virtual machine provision algorithms for cloud systems." Utility and Cloud Computing (UCC), 2011 Fourth IEEE International Conference on. IEEE, 2011.

[18] Beloglazov, Anton, and Rajkumar 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.

[19] A. Basu, H. B. Manasa. "Energy Aware Resource Allocation in Cloud Data Centers ." International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-2, Issue-5, June 2013

[20] Buyya, Rajkumar, Anton Beloglazov, and Jemal Abawajy. "Energy-efficient management of data center resources for cloud computing: a vision, architectural elements, and open challenges." arXiv preprint arXiv:1006.0308(2010).

[21] Mishra, Mayank, and Anirudha Sahoo. "On theory of vm placement: Anomalies in existing methodologies and their mitigation using a novel vector based approach." Cloud Computing (CLOUD), 2011 IEEE International Conference on. IEEE, 2011.

[22] Gambosi, Giorgio, Alberto Postiglione, and Maurizio Talamo. "Algorithms for the relaxed online bin-packing model." SIAM journal on computing 30.5 (2000): 1532-1551.

[23] Zoha Usmani and Shailendra Singh / Procedia Computer Science 78 (2016) 491 – 498, Paper Title, A Survey of Virtual Machine Placement Techniques in a Cloud Data Center in Conference "International Conference on Information Security & Privacy (ICISP2015), 11-12 December 2015, Nagpur, INDIA".