A Survey of Virtual Machine Placement Techniques and VM Selection Policies in Cloud Datacenter

S.Poovizhi
PG Scholar: dept. of Computer Science and Engineering
Velalar College of Engineering and Technology
Erode, India
e-mail: poovizhi.s@gmail.com

N.Jayanthi
Asso. Professor: dept. of Computer Science and Engineering
Velalar College of Engineering and Technology
Erode, India

Abstract— The large scale virtualized data centers have been established due to the requirement of rapid growth in computational power driven by cloud computing model . The high energy consumption of such data centers is becoming more and more serious problem .In order to reduce the energy consumption, server consolidation techniques are used .But aggressive consolidation of VMs can lead to performance degradation. Hence another problem arise that is, the Service Level Agreement(SLA) violation. The optimized consolidation is achieved through optimized VM placement and VM selection policies . A comparative study of virtual machine placement and VM selection policies are presented in this paper for improving the energy efficiency.

Keywords*- Server consolidation, SLA violation, over loaded host, virtual machine.*

*****

I. INTRODUCTION

Instead of raising total cost of ownership by purchasing and maintaining IT infrastructure, the cloud computing model allows customers to provision resources on a pay-as-you-go basis. This leads to the establishment of thousands of computing nodes which leads to the high energy consumption. The process of hardware and software partitioning ,time sharing and partial or complete machine simulation is called Virtualization, in which each simulation can act as a complete system is provided for a promising approach through which hardware resources on one or more machines can be used in multiple execution environments. Hypervisor or Virtual Machine Monitor (VMM) are the software used to virtualize physical servers .The virtualized physical servers are called the virtual machines(VM's).The Virtual Machine (VMs), improves the utilization of resources. Live migration is a technique for moving a running virtual machine between different physical machines for load balancing .But SLA violation can occur due to this load balancing. Service level agreement (SLA) is a type of contract that must be signed between service provider and service customer before actual delivery of services. It defines functional and non-functional characteristic of a service including quality of services requirements and penalties in case of SLA violation [39]. Some of the SLA parameter are Response time, Storage, Network bandwidth and memory which are used as metrics to determine the provisioning of resource requirement[39] .In data centre due to poor hardware utilization ,an increased amount of space and resources are occupied by virtual servers in spite of their workload ,which results in the occurrence of server sprawl .Thus green cloud computing is required to reduce the number of active servers which is achieved through server consolidation. In server consolidation technique ,the number of physical servers used are reduced by migrating multiple virtual machines into active servers and turning off other servers to sleep mode. This in turn reduce the overall system complexity and provides energy efficiency and reduced SLA violation. Some of the other problems due to VM consolidation are, The energy-performance profiling technique is required for the calculation of the server energy consumption , which leads to overhead problems . And the energy efficiency of the cloud data centre depends on the energy consumption of each of the cloud resources which should be minimum and must match the specifications of service level agreements.

II. SERVER CONSOLIDATION

The sever consolidation requires four major steps namely host over-load Detection, host Under-load Detection, VM Selection and Migration ,VM placement.

A. Host over-load detection

In this step, the scheduling techniques will produce a threshold value based on the history of resource utilization of the cloud data centre . And it is set as the overloaded threshold limit and it is compared with each of the host's utilization and if it is above the threshold limit it is set as overloaded host. And the VM's in the overloaded host are migrated for server consolidation.

B. Host under-load detection

In this step, the host's utilization which are less than the threshold limit are identified and are set as under loaded host .

ISSN: 2321-8169
77 – 82

And all the VM's in these host are migrated to other hosts and are put to sleep mode for low power consumption.

C. Vm selection and migration

In this step, the VM's to be migrated from overloaded host are selected and VM List is created for migration .There are several VM selection policies available with varying cloud data centre resources.

D. Vm placement

In this step, the vm's in the VM list are migrated to other suitable physical machines based on VM to PM mapping , which should reduce the energy consumption by data centre.

III. VIRTUAL MACHINE PLACEMENT ALGORITHMS

Virtual machine placement is the process of allocating the physical machines for virtual machines from VM list of overloaded host . The study on algorithms for allocation of virtual machines (VMs) to physical machines (PMs) in infrastructure clouds has been done recently. Initial placement, consolidation, or tradeoffs between honoring service-level agreements and constraining provider operating costs are some of the problems which are covered in those algorithms .Of these power saving and delivering QOS are the two major goals of the VM placement techniques in VM consolidation. In this paper we are concentrating on power-based VM placement algorithms which uses Dynamic VM placement algorithms.

A. Classification of vm placement algorithms

Classification of VM placement algorithms are Constraint programming, Bin packing problem, stochastic integer programming, Genetic algorithm, Adaptive algorithms.

1) Constraint programming

Constraint programming is a logical programming which is widely used in production planning, scheduling ,timetabling and in variety of combinatorial problems .Here we are using it to find the optimal placement of virtual machines. The constraint programming considers real-world-problem as a constraints satisfaction problem and a general purpose constraint solver calculates solution for it. The basic idea of constraint programming is applied on the definition of virtual machine placement problem, the CSP is modeled .

Zhang et al. 30.used constraint programming technique for virtual machine placement in physical servers for reducing the total cost on resource usage which also does not violates the Quality of service requirements. The performance measures and workload types are considered and fulfilled by the author.

DuPont et al.29. proposed a energy-aware resource allocation model for cloud datacenters which considers SLA violation reduction in VM placement. A VM Repacking Scheduling Problem is introduced which is a flexible and extensible framework which improves energy efficiency and allows to derive SLA constraints by users.

J.Dong et al5. Introduced a two-staged VM scheduling algorithm which includes network link capacity and physical machine size as constraints and modeled the problem. They combined Best Fit heuristic of Bin Packingas constraints and modeled the problem. They combined Best Fit heuristic of Bin Packing with min-cut hierarchical clustering algorithm to place VM's. Here the network congestion is reduced by MLU(Maximum Link Utilization) and also the number of active P;s used is also reduced.

2) Bin packing

Bin packing in a combinatorial NP-hard problem. In this problem in a finite number of bins of capacity 1, objects of different sizes should be packed in such a way that the number of bins used should be minimized. Here we are considering each VM's as objects and Physical Machines as the bins to fill in the objects. And the resources should be allocated in an optimized manner so as to reduce the number of physical machines used.

W.Song et al[1]. proposed a work using relaxed on-line bin packing algorithm VISBP(Variable Item Size Bin Packing).Here they used trace-driven simulation in order to compare VISBP with Black-box, Gray-box and Vector Dot algorithms.VISBP algorithm used only CPU and Memory. And the algorithm achieves good green computing effect and stability compared to other algorithms. Also it excels in hot spots mitigation and load balance. Due to the support of 'change' operation the algorithm supports dynamic resource allocation. Since here the VM to PM ratio is not optimized SLA violation is not reduced optimally.

Y.Zhang et al[4], proposed a heuristic algorithm for VM consolidation which includes the heterogeneity aware resource allocation. They have used algorithms like Dominant-Residual Resource aware FFD.The performance evaluation shows that its performance is same as dimension-aware heuristics but with the same cost as single dimension heuristics.

C.Lin.et.al[18], introduced Dynamic Round-Robin(DRR) algorithm and Hybrid algorithm which is the combination of DRR and First Fit which helps in energy-effective virtual machine provisioning and consolidation. This proposed approach includes two rules for consolidating VM's.First is to restricting the physical machines from allocating new VM. And the second is to shut down the physical machines by speeding up the consolidation process. This work reduced energy consumption productively compared with Round Robin algorithm.

ISSN: 2321-8169
77 – 82

3) Stochastic integer programming

Stochastic programming is a framework in the field of mathematical optimization which includes problems with uncertainty. The idea on stochastic integer programming is to use probability distributions. Here since the VM's future requirement is not known , this technique can be used to predict the suitable VM-PM mapping.

N.Bobroff et al[7], introduced a dynamic server migration and consolidation algorithm. The proposed algorithm reduces SLA violation by reducing the amount of physical capacity required .The management algorithm (MFR) dynamically remaps VMs to PMs required, so as to reduce the SLA violation .The combination of Bin packing heuristics and Time series forecasting techniques are used to reduce the number of physical machines used.

SpeitKamp et al[31]. Proposed a optimized server consolidation problem using LP-relaxation –based heuristic and historical workload analysis .They introduced a capacity planning approach to minimize the number of server used and their operational cost .This approach combines data preprocessing approach and an optimization model .

M.Chen et al[9] proposed a VM sizing approach known as Effective sizing .In this approach the stochastic optimization problem is simplified by associating VM’s dynamic load to a fixed demand using statistical multiplexing principle. Here the VM resource is calculated as a aggregation of intrinsic demand and correlation aware demand.

4) Genetic algorithm

Genetic algorithm begins with a initial set of population .Each solution can be represented as a tree,with VM’s as child node, physical nodes as the parent node and global resource manager as the root node.

Mi et al. [34] introduced adaptive self-reconfiguration of VM reallocation on heterogeneous PMs using Genetic Algorithm Based Approach(GABA).To catch up with the changing workloads, request forecasting module is used. GABA results in conservation of power and deals with multi-objective optimization.

Ferdaus et al. 20 model the problem of VM placement as an NP-Hard Multi-Dimensional Vector Packing Problem(mDVPP) focusing on balancing the cloud resource utilization, making use of the ACO (Ant Colony Optimization) meta heuristic. This is an effective approach where computation time is also remarkably lesser.

Gao et al. 23 minimized the power consumption and resource wastage of VM placement problem using a modification of Ant Colony System (ACS) algorithm.The residual resources were effectively balanced along different dimensions on the servers. This combinatorial problem is modeled as a multi-objective algorithm named VMPACS.

5) Adaptive algorithms

Anton Beloglazov and Rajkumar Buyya1.1 proposed a adaptive utilization thresholds, which ensures a high level of meeting the Service Level Agreements.They used a modified Best Fit algorithm .The technique outperforms in terms of SLA-violation.

Zhou Zhou1.2,proposed a adaptive three-threshold energy - aware algorithm in order to reduce the high energy consumption and SLA-violation .Here the data centre is divided into four classes and Energy aware Best Fit Decreasing algorithm is used. This paper shows that dynamic thresholds are more efficient than fixed threshold.

IV. VM SELECTION POLICY

On detecting the overloaded host the first step is to choose VMs which have to be migrated, and the second step is to place the chosen VMs to hosts. The VM's to migrate are selected based on the VM selection policies listed below.

TABLE I VM SELECTION POLICY

Selection policy	Technique	Parameters used
MIMT[37]	Policy selects a vm that requires minimum time to complete a migration relatively to other vm's allocated to the host.	CPU utilization
MAMT[37]	Policy chooses the maximum number of VM's which must be migrated from the host inorder to reduce the CPU utilization for the host with heavily loaded.	CPU utilization

HPGT[37]	Highest potential growth policy migrates VM from that has the lowest usage of the CPU relative to the total CPU capacity of VM for a host with heavy load.	CPU utilization
RC[19]	Random choice policy selects a vm to be migrated according to the uniformly distributed random variable.	CPU,RAM,Bandwidth
MC[19]	maximum correlation selects the vm to be selected that has the highest correlation of the cpu utilization with other vms.	CPU
MMS[38]	MMS policy selects a VM with the minimum memory size to migrate compared with the other VMs allocated to the host.	Memory
LCU[38]	LCU policy chooses a VM with the lowest CPU utilization to migrate compared with the other VMs allocated to the host	CPU
MPCM[38]	Minimum Product of Both CPU Utilization and Memory Size selects a VM with the minimum product of both CPU utilization and memory size to migrate compared with the other VMs allocated to the host.	CPU and Memory

TABLE II VM ALLOCATION TECHNIQUES

Algorithm	**Technique**	**Parameters Used**	**Advantages**	**Disadvantages**
Best Fit heuristic of Bin Packing with min-cut hierarchical clustering algorithm.[5]	Constraint programming	Network-link capacity and Physical machine size.	Number of active PM's used is reduced.	More Migration cost
MFR(Measure Forecast Remap)[7]	Stochastic Integer Programming	CPU	Meets SLA requirements and reduced number of PM's	Needs more extension to multiple resources.
Effective sizing algorithm[9]	Stochastic Integer Programming	CPU , Memory	Reduced number of PM's used.	Needs more extension to multiple resources.
VISBP[1]	Bin packing	CPU,Memory and network	Achieves good green computing effect,load balancing,dynamic resource allocation and stability.	SLA violation
Enhanced FFD[3]	Bin packing	CPU	More energy-efficient,High system through-put.	SLA violation
GABA[34]	Genetic algorithm	CPU	Improved CPU utilization,reduced number of PM's	Over head of large searching space.
VMPACS[23]	Genetic algorithm	CPU, network and storage	Energy efficient, minimum resource usage.	High time consumption
BGM-BLA[25]	Genetic algorithm	CPU, memory, storage	Reduced energy consumption	High consumption time.
Modified	Adaptive	CPU	Outperforms in	Needs to

Best-Fit algorithm[37]	threshold algorithm		SLA violation	reduce the number of VM migration.
Energy-aware best fit decreasing algorithm[38]	Adaptive threshold algorithm	CPU	Reduced energy consumption	Needs to reduce the number of VM migration.

V. Conclusion

Server Consolidation in data centres is an important area of research in the recent years. In the survey [40] consolidations mainly detailing the VM placement algorithms are mentioned.In this paper the main focus is on listing the different VM selection policies in addition to the VM Placement techniques. The objective of these techniques can either be minimization of power consumption or providing reduced SLA violation both being in conflict. Ranking these algorithms or stating the best one out of the lot is not a proper suggestion because every other technique has some specific target, migration technique, prominent resources and influential parameters. Although these techniques may seem fine from outside, there exist some or the other kind of trade-offs when deeply surveyed. Owing to the workload variability and continuously changing demands of applications, there is a need to constantly optimize these VM placement algorithms and VM selection policies.

References

[1] Song, W., Xiao, Z., Chen, Q. and Luo, H., 2014. Adaptive resource provisioning for the cloud using online bin packing. Computers, IEEE Transactions on, 63(11), pp.2647-2660.

[2] Ghribi, C., Hadji, M. and Zeghlache, D., 2013, May. Energy efficient vm scheduling for cloud data centers: Exact allocation and migration algorithms. In Cluster, Cloud and Grid Computing (CCGrid), 2013 13th IEEE/ACM International Symposium on (pp. 671-678). IEEE.

[3] Alahmadi, A., Alnowiser, A., Zhu, M.M., Che, D. and Ghodous, P., 2014, March. Enhanced First-Fit Decreasing Algorithm for Energy-Aware Job Scheduling in Cloud. In Computational Science and Computational Intelligence (CSCI), 2014 International Conference on (Vol. 2, pp. 69-74). IEEE.

[4] Zhang, Y. and Ansari, N., 2013, December. Heterogeneity aware dominant resource assistant heuristics for virtual machine consolidation. In Global Communications Conference (GLOBECOM), 2013 IEEE (pp. 1297-1302). IEEE.

[5] Dong, J., Wang, H. and Cheng, S., 2015. Energy-performance tradeoffs in IaaS cloud with virtual machine scheduling. Communications,China, 12(2), pp.155-166.

[6] Gambosi, G., Postiglione, A. and Talamo, M., 2000. Algorithms for the relaxed online bin-packing model. SIAM journal on computing, 30(5), pp.1532-1551. 7. Bobroff, N., Kochut, A. and Beaty, K., 2007, May. Dynamic placement of virtual machines for managing sla violations. In Integrated Network Management, 2007. IM'

[7] 10th IFIP/IEEE International Symposium on (pp. 119-128). IEEE. 498 Zoha Usmani and Shailendra Singh / Procedia Computer Science 78 (2016) 491 – 498

[8] Wang, M., Meng, X. and Zhang, L., 2011, April. Consolidating virtual machines with dynamic bandwidth demand in data centers In INFOCOM, 2011 Proceedings IEEE (pp. 71-75). IEEE.

[9] Chen, M., Zhang, H., Su, Y.Y., Wang, X., Jiang, G. and Yoshihira, K., 2011, May. Effective VM sizing in virtualized data centers. In Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on (pp. 594-601). IEEE.

[10] Rieck, B., 2010. Basic Analysis of Bin-Packing Heuristics. Publicado por Interdisciplinary Center for Scientific Computing. Heildelberg University.

[11] Srikantaiah, S., Kansal, A. and Zhao, F., 2008, December. Energy aware consolidation for cloud computing. In Proceedings of the 2008 conference on Power aware computing and systems (Vol. 10).

[12] Wood, T., Shenoy, P., Venkataramani, A. and Yousif, M., 2009. Sandpiper: Black-box and gray-box resource management for virtual machines.Computer Networks, 53(17), pp.2923-2938.

[13] Beloglazov, A. and Buyya, R., 2010, November. Adaptive threshold-based approach for energy-efficient consolidation of virtual machines in cloud data centers. In Proceedings of the 8th International Workshop on Middleware for Grids, Clouds and e-Science (Vol. 4). ACM.

[14] Li, K., Zheng, H. and Wu, J., 2013, November. Migration-based virtual machine placement in cloud systems. In Cloud Networking (CloudNet), 2013 IEEE 2nd International Conference on (pp. 83-90). IEEE.

[15] Paul, I., Yalamanchili, S. and John, L.K., 2012, December. Performance impact of virtual machine placement in a datacenter. In Performance Computing and Communications Conference (IPCCC), 2012 IEEE 31st International (pp. 424-431). IEEE.

[16] Singh, A., Korupolu, M. and Mohapatra, D., 2008, November. Server-storage virtualization: integration and load balancing in data centers. In Proceedings of the 2008 ACM/IEEE conference on Supercomputing (p. 53). IEEE Press.

[17] Singh, A., Korupolu, M. and Mohapatra, D., 2008, November. Server-storage virtualization: integration and load balancing in data centers. In Proceedings of the 2008 ACM/IEEE conference on Supercomputing (p. 53). IEEE Press.

[18] Lin, C.C., Liu, P. and Wu, J.J., 2011, December. Energy-efficient virtual machine provision algorithms for cloud systems. In Utility and Cloud Computing (UCC), 2011 Fourth IEEE International Conference on (pp. 81-88). IEEE.

[19] Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers

[20] Ferdaus, M.H., Murshed, M., Calheiros, R.N. and Buyya, R., 2014. Virtual machine consolidation in cloud data centers using ACO metaheuristic. InEuro-Par 2014 Parallel Processing (pp. 306-317). Springer International Publishing.

[21] Buyya, R., Beloglazov, A. and Abawajy, J., 2010. Energy-efficient management of data center resources for cloud

computing: a vision, architectural elements, and open challenges. arXiv preprint arXiv:1006.0308.

[22] Mishra, M. and Sahoo, A., 2011, July. On theory of vm placement: Anomalies in existing methodologies and their mitigation using a novel vector based approach. In Cloud Computing (CLOUD), 2011 IEEE International Conference on (pp. 275-282). IEEE.

[23] Gao, Y., Guan, H., Qi, Z., Hou, Y. and Liu, L., 2013. A multi-objective ant colony system algorithm for virtual machine placement in cloud computing.Journal of Computer and System Sciences, 79(8), pp.1230-1242.

[24] Camati, R.S., Calsavara, A. and Lima Jr, L., 2014. Solving the virtual machine placement problem as a multiple multidimensional knapsack problem. ICN 2014, p.264.

[25] Tao, F., Li, C., Liao, T. and Laili, Y., 2015. BGM-BLA: a new algorithm for dynamic migration of virtual machines in cloud computing.

[26] Beloglazov, A., Buyya, R., Lee, Y.C. and Zomaya, A., 2011. A taxonomy and survey of energy-efficient data centers and cloud computing systems.Advances in computers, 82(2), pp.47-111.

[27] Beloglazov, A. and Buyya, R., 2013. Managing overloaded hosts for dynamic consolidation of virtual machines in cloud data centers under quality of service constraints. Parallel and Distributed Systems, IEEE Transactions on, 24(7), pp.1366-1379.

[28] Beloglazov, A. and Buyya, R., 2010, May. Energy efficient resource management in virtualized cloud data centers. In Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (pp. 826-831). IEEE Computer Society.

[29] Dupont, C., Giuliani, G., Hermenier, F., Schulze, T. and Somov, A., 2012, May. An energy aware framework for virtual machine placement in cloud federated data centres. In Future Energy Systems: Where Energy, Computing and Communication Meet (e-Energy), 2012 Third International Conference on (pp. 1-10). IEEE.

[30] Zhang, L., Zhuang, Y. and Zhu, W., 2013. Constraint Programming based Virtual Cloud Resources Allocation Model. International Journal of Hybrid Information Technology, 6(6), pp.333-344.

[31] Speitkamp, B. and Bichler, M., 2010. A mathematical programming approach for server consolidation problems in virtualized data centers. Services Computing, IEEE Transactions on, 3(4), pp.266-278.

[32] Bichler, M., Setzer, T. and Speitkamp, B., 2006. Capacity planning for virtualized servers. In Workshop on Information Technologies and Systems (WITS), Milwaukee, Wisconsin, USA.

[33] Chowdhury, M.R., Mahmud, M.R. and Rahman, R.M., 2015, June. Study and performance analysis of various VM placement strategies. In Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2015 16th IEEE/ACI International Conference on (pp. 1-6). IEEE.

[34] Mi, H., Wang, H., Yin, G., Zhou, Y., Shi, D. and Yuan, L., 2010, July. Online self-reconfiguration with performance guarantee for energy efficient large-scale cloud computing data centers. In Services Computing (SCC), 2010 IEEE International Conference on (pp. 514-521). IEEE.

[35] Tang, C., Steinder, M., Spreitzer, M. and Pacifici, G., 2007, May. A scalable application placement controller for enterprise data centers. In Proceedings of the 16th international conference on World Wide Web (pp. 331-340). ACM.

[36] Wang, X., Chen, X., Yuen, C., Wu, W. and Wang, W., 2014, December. To migrate or to wait: Delay-cost tradeoff for cloud data centers. In Global Communications Conference (GLOBECOM), 2014 IEEE (pp. 2314-2319). IEEE.

[37] ZHOU Zhou, HU Zhi-gang, SONG Tie, YU Jun-yang 2015,march, A novel virtual machine deployment algorithm with energy efficiency in cloud computing. Article in Journal of Central South University, DOI: 10.1007/s11771-015-2608-5.

[38] Zhou Zhou,1 Zhigang Hu,1 and Keqin Li2, Virtual Machine Placement Algorithm for Both Energy-Awareness and SLA Violation Reduction in Cloud Data Centers,,in Hindawi Publishing Corporation, Volume 2016, Article ID 5612039.

[39] Praveen Shukla , R. K. Pateriya ,may,2015,Review on Various Energy-Aware and SLA Violation Reduction Techniques, International Journal of Computer Applications (0975 – 8887) Volume 117

[40] Zoha Usmani , Shailendra Singh ,Feb 2016, A Survey of Virtual Machine Placement Techniques in a Cloud Data Centre ,volume 78, 1st International Conference on Information Security & Privacy.