iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** XI **Month of publication:** November 2019

DOI: http://doi.org/10.22214/ijraset.2019.11117

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XI, Nov 2019- Available at www.ijraset.com

Virtual Machine Placement Optimization in Cloud Environment: A Survey

Swasthi Shetty C S
BMS College of Engineering

Abstract: Energy conservation in data centers has been an active research area in cloud computing in recent times. Effective energy conservation can be achieved using server consolidation, which aims at utilizing server resources efficiently and minimizing the number of active Physical Machines (APMs) running in a data center. Effective placement of virtual machines is necessary to optimize server consolidation. Virtual machine placement techniques provide a suitable mapping of hosts to VMs to reduce energy consumption and minimize SLA violation in data centers. This paper presents a comprehensive survey of different Virtual Machine placement techniques utilized in cloud computing, revealing the advantages and limitations of the algorithms.
Keywords: VM Placement, VM consolidation, Data Centers, Virtual Machines, SLA Violation

I. INTRODUCTION

Cloud computing provides access to on demand computing resources to the users in a pay-as-you-use pricing model. Three different models are being offered by cloud service providers such as IaaS (Infrastructure as a service), PaaS (Platform as a service) and SaaS (Software as a service). One of the significant challenges for cloud service providers is to reduce energy consumption in data centers. The cloud service providers spend a significant amount in setting up data centers in the beginning. They have to incur data center management costs later to maintain data centers. This includes power costs, software and hardware maintenance costs etc. According to a recent study [1], 13% of the overall data center management cost is incurred by power consumption. So, it is essential to optimize power consumption in data centers to reduce the operational cost for cloud service providers. To prevent wastage of resources in data centers, Virtual Machines (VM) are packed on to the fewest possible physical machines and idle physical machines are later shut down, thereby reducing energy consumption. This process of consolidating VMs on to the servers is called server consolidation. It comprises of 4 steps: 1) Host underload detection, where hosts with utilization under a certain threshold are selected, all the VMs on the host are migrated to other servers, and underloaded hosts are shutdown. 2) Host overload detection, where hosts with utilization greater than a certain threshold are detected and some of the VMs are migrated to other hosts. 3) VM selection, where appropriate VMs are selected for migration from over utilized hosts. 4) VM placement, where VMs selected for migration in the 3rd step is mapped to different Physical Machines (PM). In this paper, we focus on Virtual Machine Placement algorithms. Virtual machine placement (VMP) is the process of mapping Virtual machines to Physical machines in order to reduce energy consumption and minimize SLA violation in data centers. VMP has been an active research area in cloud computing throughout the last decade. Many VMP algorithms have been proposed to maximize utilization and to reduce power consumption, in turn reducing operational costs in data centers. VMP algorithms can be traffic-aware, load-aware, application-aware, power-aware or a combination of these. To achieve better performance, VMs are migrated to other hosts when servers become over utilized or underutilized. So, when the resource demands of a Virtual machine cannot be fulfilled by the physical machine on which the VM is hosted, VMs are migrated to another PM for the fulfillment of the demands. . VMs are migrated from over utilized hosts to prevent Service level Agreement violation. In the case of underutilized hosts, all the VMs hosted on the PM are migrated and the host is shut down. The remainder of this paper is organized as follows. Section II describes the classification of VM placement algorithms. Section III presents a detailed discussion of different approaches used in VM placement algorithms and Section IV presents concluding remarks and future research directions.

II. VM PLACEMENT CLASSIFICATION

A. Power and Quality of Service

1) *Power based:* The objective of power based virtual machine placement algorithm is to map virtual machines to physical machines in a manner to reduce energy consumption in data centers. Virtual machines are aggressively packed in physical machines and underutilized physical machines are shut down to reduce power consumption [2].
2) *QoS based*: The objective of this approach is to meet the quality of service guaranteed by cloud service providers. Service Level Agreement (SLA) is signed between the user and cloud service provider when users opt for cloud services. Service provider will have to pay the penalty when they fail to deliver quality of service. QoS based approaches are used to minimize SLA violation, in turn ensure quality of service to the customers [3].

B. Dynamicity

1) *Static VM Placement*: VMs are mapped to PMs during the initial placement and reallocation is not considered for a long period of time. The VM-PM mapping is fixed throughout the lifetime of a VM
2) *Dynamic VM Placement*: This technique is used to modify VM placement on regular time intervals depending on load, energy consumption etc. Dynamic VM allocation requires more resources compared to static allocation as dynamic solutions are implemented for shorter timescales and could also affect the performance of applications running on virtual machines. Dynamic VM placement techniques can be further classified into two types [4]:
   a) *Reactive:* VM reallocation is implemented after the systems reach a certain state. After the initial placement of VMs on PMs, changes are done only when hosts become underloaded or overloaded.
   b) *Proactive:* After the initial placement, reallocation of VMs are done at regular intervals before the system reaches a certain state. It aims at preventing overloading or underloading of servers and reallocation is carried out before the servers are overloaded. This approach is better than reactive placement as this prevents SLA violation.

C. Optimization Strategy

1) *Heuristic*: Heuristic algorithms are problem dependent solutions that can provide a satisfactory solution to a specific problem in limited time and cost. Greedy algorithm is one such example for heuristic algorithms that has been implemented in many VM placement strategies. Heuristic algorithms run fast and hence can be applied in various online VM placement techniques.
2) *Metaheuristic*: Metaheuristic algorithms are problem independent solutions, that can be applied to a wide range of problems. These algorithms need more time to execute and find solutions in comparison to heuristic algorithms. ACO (Ant Colony Optimization), genetic algorithms, honeybee foraging algorithms are examples of metaheuristic algorithms that have been used in VM placement techniques to deal with load balancing. These algorithms have been found to provide better load balancing compared to heuristic algorithms.
3) *Hybrid:* In hybrid algorithms, both heuristic and metaheuristic approaches are combined to come up with an optimal solution. In some cases, heuristic algorithms are used for initial placement of VMs and metaheuristic algorithms are used to optimize VM placement during VM migration. Alternatively, metaheuristic algorithms can be used to provide solutions to a set of problems and heuristic algorithms can be used to find an optimized solution based on these solutions. Hybrid algorithms increase implementation complexity but at the same time reduce time and cost to arrive at the solution.

III. VM PLACEMENT ALGORITHMS

A. Bin Packing-based VM Placement

VM placement can be seen as a bin packing problem where Physical machines are considered as bins with minimum resource capacity and Virtual machines are objects that need to be packed into these bins. Requirements of virtual machines are to be met by the PMs. Aggressive packing of virtual machines on to physical machines will allow idle physical machines to be shut down, in turn reducing power consumption in data centers. Bin packing algorithms can be classified as follows:

1) *First Fit*: This greedy based approach allocates Virtual machine to the first PM that has enough resources required by VM. As and when a new VM arrives, the algorithms start to find a PM that has enough capacity to host the VM. A new PM is switched on, in case it fails to find a PM that has enough capacity [3,4,5].
2) *Next Fit Method:* In this technique, an initial VM is placed on the first PM that has adequate resources. For the second VM, the search starts from the Physical machine that hosted the first VM. For the next VM, search starts from the physical machine that hosted the previous VM.
3) *Random Fit:* A physical machine is chosen in a random manner for Virtual machine allocation [6]. If the physical machine chosen does not have enough resources, another physical machine is chosen randomly for the Virtual Machine to be placed
4) *Best Fit Decreasing:* This technique sorts all the PMs in decreasing order of their utilization and selects PM with the highest utilization for VM allocation [7]. If PM with the highest utilization has no adequate resources to host a VM, PM with the second highest utilization is taken into consideration. This technique ensures aggressive packing of VMs on to the PMs and allows PMs with lower utilization to be shut down after migrating VMs to the higher utilized servers. This algorithm has been proven to be better than first fit, next fit and random fit algorithms.

Fikru et al. proposed a VM placement algorithm based on bin packing. They introduced two algorithms - modified best fit decreasing algorithm to minimize the number of servers and Medium-Fit (MF) algorithm to reduce SLA violation. MF rule calculates desired resource utilization level(L_d) and selects a host with utilization level that has a minimum distance from L_d VM allocation. MF algorithm performs lesser number of VM migrations and lowers SLA violation compared to other VM placement algorithms [8]. Lianpeng et al. proposed a RobustSLR Power-Aware Best fit decreasing (RPABFD) algorithm for VM placement where future host load is predicted using Robust Simple Linear Regression (RobustSLR) algorithm. Overloaded and underloaded hosts found using RobustSLR algorithm is excluded from the list for VM placement. According to the author, this algorithm has reduced SLA violation rates by 99.16% and energy consumption by 25.43% [9]. Hui et al. proposed Space Aware Best fit decreasing algorithm for VM placement, where the host with minimum available MIPS (Millions of Instructions Per Second) is selected for placement. This algorithm has outperformed Power-Aware Best fit decreasing algorithm both on assuring SLA and energy conservation [10].

B. Constraint Programming

VM placement problem can be solved using constraint programming technique where the solution must satisfy the constraints on relations between variables. The goal of this technique is to maximize global utility function that is finalized based on SLA and operational cost. The virtual machine placement can be seen as a complex combinatorial problem that can be solved using constraint programming techniques in two stages. The first stage is to arrive at local decisions associated with each application hosted on the cloud. The second part is to feed local decisions of all applications to a global decision module that maps VMs to PMs based on the CPU load [11]. Dong et al. proposed a two stage VM scheduling scheme. In the first stage, the authors proposed a static VM placement scheme minimizing active servers and network elements to reduce energy consumption. They combined best fit algorithm with minimum cut hierarchical clustering. In the second stage, they used local search algorithms during VM migration to optimize VM allocation and to minimize Maximum link Utilization (MLU) in VMs and at the same time, to reduce link congestion [12]. Yu et al. proposed a virtual machine placement algorithm by formulating the VM placement problem as a Constraint satisfaction problem. The objective of this algorithm was to cut operational costs and increase resource utilization with lesser number of servers. They used Choco software library as a constraint solver [13].

C. Stochastic Integer Programming (SIP)

This mathematical optimization technique is used when future demands and cost of resources are unknown and expected distributions can be calculated or are known. This technique is used in VM placement optimization as the future demands of VM utilization are unknown. This technique aims to minimize the operational cost when future demands are unknown [7].

Sivadon Et al. proposed an Optimal VM placement (OVMP) algorithm to minimize the cost plan for renting resources from cloud providers. OVMP makes a decision based on the optimal solution of SIP in a multi cloud environment where future demands of resources are unknown. The proposed algorithm is carried out in 2 stages. First, the number of VMs in the reservation phase are defined. In the second stage, actual number of VMs required and actual rent are defined by the providers. This algorithm reduces the cost of resource provisioning in cloud data centers [14]. Bobroff et al. proposed a VM consolidation and migration technique that operates in 3 steps. The first step is to measure historical data of VM and PM utilization. Second step is to predict future demands and then mapping VMs to PMs based on the predicted data. Hence it is also called as Measure-Forecast-Remap algorithm. This algorithm reduces the capacity demands of the server and thereby reducing the operational cost [15].

D. Genetic Algorithm

This algorithm performs a natural selection of solutions among the possible solutions. Genetic algorithm can be considered as an extension of bin packing approach with additional constraints. It takes more computing time and resources compared to bin packing algorithms. Genetic algorithms focus on multiple objectives like minimizing power consumption, cost, resource wastage etc. They solve problems of conflicting objectives, that bin packing algorithms cannot. Tang et al. proposed a hybrid genetic algorithm to come up with the energy efficient virtual machine placement technique. Authors have extended the Genetic algorithm approach to incorporate infeasible solution repairing procedure and local optimization procedure. This technique turns infeasible solution to a feasible one by resolving all the constraint violations and later uses local optimization procedure to improve the solution. It takes into account energy consumption and communication network in data centers [16]. Ferdaus et al. proposed a server consolidation mechanism using ACO metaheuristic approach. This technique minimized resource wastage and power consumption in data centers. Computation time taken is higher compared to other methods [17].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XI, Nov 2019- Available at www.ijraset.com

All the above stated VM placement algorithms are summarized in the table below. Table 1 also lists the strengths and weaknesses of all the algorithms.

TABLE I
summary of vm placement techniques

Name of Algorithms	Based upon	Considered Resource	Strengths	Weaknesses
Medium-Fit [8]	Bin packing	CPU	Lower SLA violation, Reduced number of migrations	Requires peak power of servers which takes extra memory
Robust SLR Power Aware BFD [9]	Bin packing	CPU	Reduced energy consumption and SLA violation	High computation time
Min-Cut hierarchical clustering [12]	Constraint Programming	CPU & network bandwidth	Reduced energy consumption and network traffic	More VM migration cost
CP-based VM Placement [13]	Constraint Programming	CPU, memory & network bandwidth	Improved resource utilization and reduction in data center costs	Slightly slow execution
OVMP (Optimal Virtual Machine Placement) [14]	Stochastic integer programming	CPU, storage & network bandwidth	Reduction in resource provisioning cost	Slightly slow execution
MFR (measure forecast-remap) [15]	Stochastic integer programming	CPU	Meet SLA targets, reduce the number of APMs	Need for extension to multiple resources
HGA [16]	Genetic algorithm	CPU and memory	Improvement in performance and efficiency	Increase in computation workload
VMPACS [17]	Genetic algorithm	Storage, CPU &network	Near optimal solution, Energy efficient, Minimum resource management	High computation time

IV. CONCLUSION

In this paper, I have presented a survey of different types of Virtual machine Placement algorithms. Many VM placement techniques are investigated and advantages and limitations are discussed. The objective of some of the algorithms is to reduce power consumption, whereas others have addressed SLA violation issues. Few techniques have tried to achieve a balance between reducing power consumption and delivering Quality of Service to customers. Although these techniques look accurate and suitable for VM placement, there is a need to constantly optimize these techniques because of the unpredictable demands of applications running on VMs. As a future enhancement, joint optimization of virtual machine selection and virtual machine placement algorithms can be considered to achieve a trade-off between energy consumption and SLA violation. Joint optimization algorithms can ensure that selected virtual machines are placed on suitable hosts depending on the criteria used for the selection of virtual machines.

REFERENCES

[1] Beloglazov, A. and Buyya, R. (2012), Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in Cloud data centers. Concurrency Computat.: Pract. Exper., 24: 1397-1420. doi:10.1002/cpe.1867

[2] 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.

[3] 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.

[4] M. Masdari et al. /Journal of Network & Computer Applications 66 (2016) 106–127.

[5] Gupta RK, Pateriya R. Survey on virtual machine placement techniques in cloud computing environment. Int J Cloud Comput: Serv Archit 2014.

[6] Choudhary, A., Rana, S., & Matahai, K. (2016). A Critical Analysis of Energy Efficient Virtual Machine Placement Techniques and its Optimization in a Cloud Computing Environment. Procedia Computer Science, 78, 132-138. doi:10.1016/j.procs.2016.02.022

[7] Kaur et al., International Journal of Advanced Research in Computer Science and Software Engineering 6(5), May- 2016, pp. 521-526.

[8] Moges, F., Abebe, S. Energy-aware VM placement algorithms for the OpenStack Neat consolidation framework. *J Cloud Comp* **8,** 2 (2019) doi:10.1186/s13677-019-0126-y

[9] L. Li, J. Dong, D. Zuo and J. Wu, "SLA-Aware and Energy-Efficient VM Consolidation in Cloud Data Centers Using Robust Linear Regression Prediction Model," in IEEE Access, vol. 7, pp. 9490-9500, 2019. doi: 10.1109/ACCESS.2019.2891567

[10] H. Wang and H. Tianfield, "Energy-Aware Dynamic Virtual Machine Consolidation for Cloud Datacenters," in IEEE Access, vol. 6, pp. 15259-15273, 2018. doi: 10.1109/ACCESS.2018.2813541

[11] Zoha Usmani and Shailendra Singh / Procedia Computer Science 78 (2016) 491 – 498.

[12] J. Dong, H. Wang and S. Cheng, "Energy-performance tradeoffs in IaaS cloud with virtual machine scheduling," in China Communications, vol. 12, no. 2, pp. 155-166, Feb. 2015. doi: 10.1109/CC.2015.7084410

[13] Yu Y., Gao Y. (2012) Constraint Programming-Based Virtual Machines Placement Algorithm in Datacenter. In: Shi Z., Leake D., Vadera S. (eds) Intelligent Information Processing VI. IIP 2012. IFIP Advances in Information and Communication Technology, vol 385. Springer, Berlin, Heidelberg

[14] S. Chaisiri, Bu-Sung Lee and D. Niyato, "Optimal virtual machine placement across multiple cloud providers," 2009 IEEE Asia-Pacific Services Computing Conference (APSCC), Singapore, 2009, pp. 103-110. doi: 10.1109/APSCC.2009.5394134

[15] N. Bobroff, A. Kochut and K. Beaty, "Dynamic Placement of Virtual Machines for Managing SLA Violations," 2007 10th IFIP/IEEE International Symposium on Integrated Network Management, Munich, 2007, pp. 119-128. doi: 10.1109/INM.2007.374776

[16] Tang, M. & Pan, S. Neural Process Lett (2015) 41: 211. https://doi.org/10.1007/s11063-014-9339-8

[17] Ferdaus M.H., Murshed M., Calheiros R.N., Buyya R. (2014) Virtual Machine Consolidation in Cloud Data Centers Using ACO Metaheuristic. In: Silva F., Dutra I., Santos Costa V. (eds) Euro-Par 2014 Parallel Processing. Euro-Par 2014. Lecture Notes in Computer Science, vol 8632. Springer, Cham