

A Load Balancing Aware Virtual Machine Live Migration Algorithm

Chengjiang Liu¹

¹Southwest Branch of State Grid Corporation of China, No. 266, Shuxiu Road, Chengdu, China
13908176712@163.com

Keywords: Virtual machine, live migration, load balancing, cloud data center, green computing.

Abstract. Green computing has become a research hotspot in service-oriented cloud center. Resources in a data center are always virtualized into a resource pool for uniform management and scheduling. Thus load balance in the cloud data center is of great importance to reach some kind of service levels and has become an extremely important part of the cloud data center. In this paper, we focus on virtual machine migration algorithm with adaptive load balancing. We model virtual machine resource application as a mathematical optimization problem. Resources importance, such as memory, CPU and network bandwidth and network topology are modeled as weight in a linear combination of these kinds of resources. We solve this problem with dynamic programming and get an optimized virtual machine migration plan. We validate the proposed method in a simulated data center with CloudSim and results show that with our method, data center load can be reduced effectively.

Introduction

As cloud service become popular and cost effective throughout the world, virtual machine (VM) technology has out staged as the key of many cloud services providers. To maximize data center utilization, cloud data center usually utilizes supervisor to manage virtualized resources. Virtual migration is then used to maintain a balanced system load. VM migration can be classified into two categories: static and dynamic (online) migration. In static VM migration, the objective VM would be closed and status of it would be migrated to another destination physical machine, so cloud service between the migration period is suspended. Another class of VM migration is dynamic migration which would keep some statuses of the VM on the objective physical machine and copy some status to the destination physical machine. Online migration can guarantee a kind of level of services but need resources to maintain running status of VM, so the process is much more complicated than the static one. Today, many data centers management software mainly uses dynamic VM migration mechanism to provide better services, to do load balancing, and to minimize power usage, etc. As high level of cloud service and lower energy consumption is always hard to obtain at the same time, so data centers always sacrifice one for another.

In a data center, VMs are running on different racks of physical machines to provide fast services. As services requirements increase, load of some machine would become very high while other physical machine's load become stably low leading to resources wastes.

In this paper, we investigate live VM migration method in a load balancing manner. When applying VM migration, there are many candidate physical machines that fulfill resources requirement of the VM. To better manage these resources, only one physical machine would be scheduled to corresponding VM. Previous works have presented some heuristic virtual machine placement algorithms in order to find optimal solutions to minimize system overhead and energy consumption. Considering both energy and load balance in VM management is a hot research and industry direction. We have proposed a load balancing aware live VM migration method which utilizes the VM require resource statistics and run-time resources statistics of both virtual machines and physical machines. We model these factors into a linear formula and find solution of it over runtime feature space of both VM and physical machines.

Organization of paper is as follows. First, we introduce the related work on VM placement algorithms. Next, we present the main design and solution of the proposed method. To validate

effectiveness of our method, we present experimental results and analysis on the popular cloud data simulation platform-- CloudSim. Finally, we summarize work of this paper.

Related works

Virtual machine migration methods in cloud data center have been studied in many aspects for years. Ahmad Shakeel et al. [1] classify virtual methods into two categories: energy aware approach and service level based approach. These two kinds of methods can be further classified as dynamic and static migration method. Wood[2] et al. propose software test based method to monitor and manage memory, CPU and network usage in the data center. Their approach monitors statistics of VMs and operating systems. But their approach does not take energy and data center topology into account. Beloglazov et al. [3] present heuristic VMs consolidation algorithms. They use statistical method to calculate the CPU utilization bound. If a physical machine is bound to be over-utilized, one VM on this physical machine is chosen and will be migrated to another physical machine. Their statistical based method can be combined to form other kinds of VM consolidation strategies.

Dias et al. [4] propose VM migration method to reduce data center traffic cost by grouping highly communication intensive VMs into clusters, and partitions these physical machine with straightforward bin-packing algorithm. [5][6] choose physical machine with minimal migration cost to consolidate virtual machines. Traffic cost for migration one single VM can be minimized, however, author does not tell whether traffic cost of the whole system is reduced or not.

Srikantaiah et al. [7] have studied VM dynamic consolidation in data centers to reduce the energy consumption. Verma et al.[8] use energy management capabilities of virtualization to maintain energy and migration cost-aware applications. [9] has introduced system EnaCloud to considerate VM placement dynamically in order to save energy. Jeyarani et al. [10] have proposed artificial based method to provide efficient virtual machine services in cloud data center.

Problem statement

Generally, VM placement problem is the problem of migrating n VMs onto m physical machines(hosts). Solution of this problem can be represented by a vector of resources. We make the assumption that there are available physical machines in resource pool of one cloud data center. Each VM is associated with a list of required resources. We assume that VMs are independent of each other with equal priority. The problem can be defined as: to find a VM location selections such that the total energy cost of the migrated VMs to destination physical machines is minimized and system load is balanced.

We define a sextuple $S=\{PM, VM, EC, RCM, NS, LS\}$ to represent the problem. PM is a set of physical machine available for online migration. VM is a virtual machines set that need to be monitored during migration time. EC is the energy cost by the physical machine in the resource pool. RCM represents the remaining CPU and memory resources of each PM at migrating start time . NS is the static and runtime network bandwidth statistics. LS is a PM location vector for current migration VMs. Obviously, this problem is a multiple optimization modal which means there are more than one location selection that meets the resource constraints of the VM requests. We aim to find LS that fulfills the resources constraints while minimizing energy consumption and load of physical machines.

$$\min \begin{cases} \sum_{i=1}^n (a_{EC} EC_{t,i} - (1-a_{EC}) EC_{t-1,i}) \\ \sum_{i=1}^n (a_{RCM} RCM_{t,i} - (1-a_{RCM}) RCM_{t-1,i}) \end{cases} \quad \text{s.t. PM in the resource pool} \quad (1)$$

The problem is described in Equation(1), where a_{EC} and a_{RCM} are weights at different time of migration. We use this parameter to balance energy and CPU-memory usage before and after virtual machine migration. Due to space limitation, some details are omitted.

Algorithm Description

In this section, we present the Load Aware Virtual Machine Placement (LAVPM) algorithm which combines the energy cost and runtime and static VM traffic into optimization. LAVPM is illustrated in Algorithm 1.

Algorithm.1. Load Aware Virtual Machine Placement (LAVPM)

1. Sort PMs in the resource pool by CPU, memory, and network status;
 2. For a VM request :
 3. If the resource in the resource pool is not sufficient to allocate,
 4. Compute Equation (1), and compute EC and NS of the VM;
 5. Record the PMs IDs whose resource is of least load;
 6. Place the VM on the first PM;
 7. The next VM request;
 8. return PM list of selection.
-

Evaluation

The simulation experiments is carried out on the CloudSim-3.0[11]. We simulated 4 racks, and each rack can deploy some VMs.

We used 2 load balancing policies for comparison:

- (1) Traffic-aware. This algorithm place VMs in a traffic manner across different racks/cabinets.
- (2) Energy-ware. This method selects servers in the rack/cabinets with the greatest energy reduction.

We use energy and number of migration to benchmark these methods. The energy is defined as the total energy to migrate a VM request list and the number of migration means during a time slot, the migration VMs on different physical machines. Simulated results are shown in Figure 1.

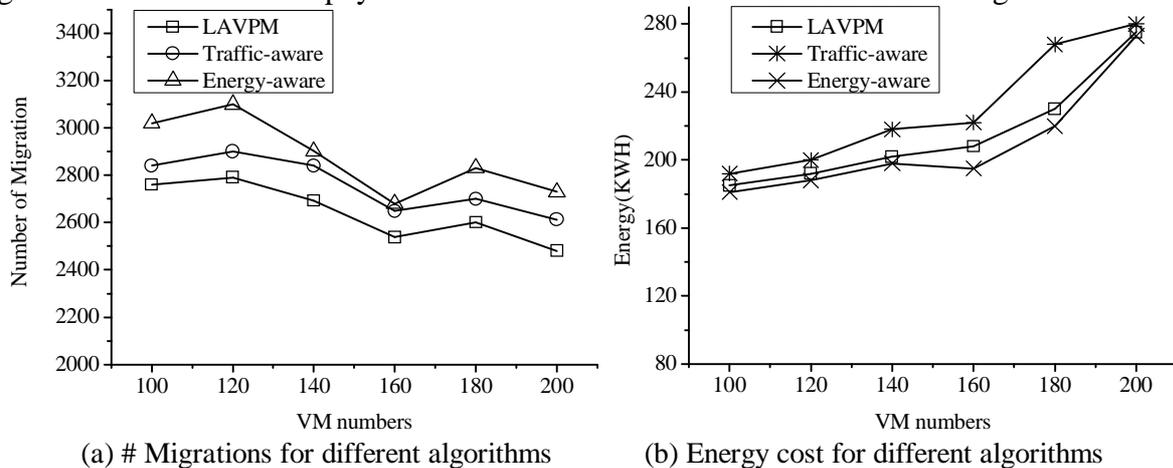


Figure 1. Benchmark for different virtual machine placement algorithms

From Figure 1 (a) we can see, LAVPM use fewer number of virtual machine migration than the other two methods because LAVPM takes both static and runtime statistics to get a better location list. Figure 1 (b) shows that LAVPM spends nearly the same energy as energy-aware method when VM number varies, so our method is a nearly optimal energy aware method.

Conclusions

This paper proposes a load balancing aware VM placement algorithm in loud data centers to minimize network cost and energy consumption. We model the VM placement problem into a

multi-model optimization problem and give a algorithm to find an optimal list of physical machine based on the model. Experimental results show that our method needs fewer migration number and requires nearly the same energy compared with energy aware placement methods. In the future, we are to implement these algorithm into a distributed cloud data center to improve system resource usage and provide better service while save energy cost.

References

- [1] S. Ahmad, B. Ahmad, M. S. Sheikh, and M. K. Rashid: International Journal of Advanced Science and Technology, Vol. 44(2012), p.69-80,.
- [2] Augustin S., Antonio N., Rene C., Emilio L., and Nina T: *How to identify and estimate the largest traffic matrix elements in a dynamic environment*. In Proceedings of the joint international conference on Measurement and modeling of computer systems, SIGMETRICS '04/Performance '04, pages 73-84, New York, NY, USA, (2004).
- [3] Anton B., Rajkumar B., Young Choon, and Albert Y. Z.: *A taxonomy and survey of energy-efficient data centers and cloud computing systems*. CoRR, abs/1007.0066,(2010).
- [4] D.S. Dias and L.H.M.K. Costa: *Online traffic-aware virtual machine placement in data center networks*. In Global Information Infrastructure and Networking Symposium (GIIS), 2012, pages 1-8, (2012).
- [5] Jun C., Weidong L., and Jiaying S: *Network performance-aware virtual machine migration in data centers*. CLOUD COMPUTING 2012, The Third International Conference on Cloud Computing, GRIDs, and Virtualization, pages 65-78, (2012).
- [6] Anjana S. and Umesh B: *Virtual machine placement in computing cloud*. Technical report, Indian Institute of Technology Bombay,(2010).
- [7] S. Srikantaiah, A. Kansal, and F. Zhao: Cluster Computing, Vol. 12(2009), p.1–15.
- [8] A. Verma, P. Ahuja, and A. Neogi: *PMapper: power and migration cost aware application placement in virtualized systems (PMapper '08)*. In Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware, pp. 243–264, Springer, New York, NY, USA, (2008).
- [9] B. Li, J. Li, J. Huai, T. Wo, Q. Li, and L. Zhong: *EnaCloud: an energy-saving application live placement approach for cloud computing environments*. In IEEE International Conference on Cloud Computing (CLOUD '09), pp. 17–24, Bangalore, India, (2009).
- [10] R. Jeyarani, N. Nagaveni, and R. V. Ram: International Journal of Intelligent Information Technologies, vol. 7, no. 2(2011), p. 25–44.
- [11] Garg S. and Buyya R.: *Networkcloudsim: Modelling parallel applications in cloud simulations*. In Proc. 4th Int. Conf. on Utility and Cloud Computing (UCC' 11). IEEE: pp. 105–113(2011).