ARTICLE IN PRESS

Future Generation Computer Systems ■ (■■■■) ■■■–■■■

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Energy-aware and multi-resource overload probability constraint-based virtual machine dynamic consolidation method

Zhihua Li [a,*], Chengyu Yan [a], Lei Yu [b], Xinrong Yu [a]

[a] Department of Computer Science and Technology, School of Internet of Things Engineering, Jiangnan University, Jiangsu Wuxi 214122, PR China
[b] School of Computer Science, College of Computing, Georgia Institute of Technology, Atlanta, GA, 30332, USA

ARTICLE INFO

Article history:
Received 8 December 2016
Received in revised form 29 August 2017
Accepted 30 September 2017
Available online xxxx

Keywords:
Virtual machine consolidation
Multi-resource constraints
Optimization model
Artificial bee colony-like search
Energy aware

ABSTRACT

In a data center, virtual machine consolidation has been proposed to improve the resource utilization and energy efficiency. An effective and efficient virtual machine consolidation method should achieve an appropriate balance among multiple goals, including guaranteeing service quality, reducing energy consumption and maximizing resource utilization. This problem is a multi-objective optimization problem with multiple resource constraints. To solve this problem, we propose an energy-aware dynamic virtual machine consolidation (EC-VMC) method that migrates virtual machines while satisfying constraints on the probabilities of multiple types of resources being overloaded. In our method, a series of algorithms for selecting and placing virtual machines to be migrated are utilized, with constraints on the probabilities of various resources in a physical machine being overloaded. Our algorithms integrate and cooperate similarly to artificial bee colony foraging behavior to perform an optimized search for the mapping relation between virtual machines and physical machines for consolidation. Extensive simulation is conducted to compare our EC-VMC method with previous virtual machine consolidation methods. The simulation results demonstrate that the EC-VMC method effectively overcomes the deficiencies of some existing heuristic algorithms and is highly effective in reducing VM migrations and energy consumption of data centers and in improving QoS.

1. Introduction

High energy consumption is a major challenge for resource management in data centers. As a data center continues to expand, the problem of high energy consumption becomes more prominent [1]. A survey by International Business Machines (IBM) [2] suggests that the average CPU utilization of physical machines in a data center is only 15%∼20%, and a physical machine in an idle state typically consumes 70% of its peak energy consumption [2,3]. Idle physical machines (PMs) with underutilized resources in a data center indicate low energy efficiency and tremendous energy waste.

Virtualization technology enables a cloud-computing service provider to create multiple VMs in a single physical machine (PM) and perform load balancing via virtual machine (VM) migration. As a major technology for improving energy efficiency and resource utilization in a data center, virtual machine (VM) consolidation has been extensively investigated [4–10]. A data center can periodically consolidate VMs and turn off some underutilized PMs based on VM and PM resource utilization to reduce energy consumption and improve resource utilization. However, because of the stochastic variation of workload in data centers, overly radical VM consolidation will negatively influence resource reservations in PMs, which leads to Service Level Agreement (SLA) violations. Therefore, minimizing energy consumption and maximizing resource utilization while guaranteeing quality of service (QoS) is a major challenge for VM consolidation in data centers.

The main strategy of VM consolidation is to define the static overload threshold or upper bound of CPU utilization to identify the overload status of the physical machine and the lower bound of CPU utilization to identify the underutilized PMs. VM migration or VM consolidation is triggered based on these thresholds to achieve the goal of reducing energy consumption and improving QoS. Such methods [3,11] are simple but lack flexibility to adapt to the dynamic workload in a data center. A dynamic overload threshold [12] is proposed for VM migration that considers the workload variations on source and destination hosts after VM migrations, but it does not consider load rebalancing of data centers if the workloads on these hosts have changed. Some existing methods [7–10] only focus on VM migration but ignore the impact of VM migration on resource consumption and service quality. For instance, as a VM in live migration suspends its service, prolonged VM migration likely affects the service quality [13].

* Corresponding author.
E-mail address: zhli@jiangnan.edu.cn (Z. Li).

https://doi.org/10.1016/j.future.2017.09.075
0167-739X/

Please cite this article in press as: Z. Li, et al., Energy-aware and multi-resource overload probability constraint-based virtual machine dynamic consolidation method, Future Generation Computer Systems (2017), https://doi.org/10.1016/j.future.2017.09.075.

Usually, VM consolidation methods can be either single resource based [3–11] or multiple resource based [14,15]. Because the CPU resource is a key factor that influences energy consumption of data centers, numerous studies have investigated how to improve the energy efficiency via the allocation of CPU resources, such as the methods proposed in [3–11]. However, these methods only consider the CPU resource. Factors such as memory, bandwidth and disk utilization should also be considered for VM consolidation, since they are also key factors that affect QoS. However, at the same time, multi-resource based VM consolidation can be more complex. For instance, previous work [14] proposed a multi-resource VM consolidation method that involves more complex assumptions and mathematical calculations due to multi-resource consideration.

To address the limitations of existing VM consolidation methods, we consider the VM consolidation problem with the following constraints: (1) VM dynamic consolidation should trigger as few VM migrations as possible to minimize the negative impact on QoS; (2) in a data center, the VM-generated workload is complex and dynamic, and the VM consolidation strategy should minimize the probability of the physical machines being overloaded; (3) VM dynamic consolidation should turn off as many underloaded PMs as possible to reduce energy consumption of data centers; and (4) resources such as memory and network are the major factors that affect QoS, so VM consolidation should allow a comprehensive treatment of such multi-resources. Based on these considerations, we propose an energy-aware dynamic optimization approach, called EC-VMC, for VM consolidation with constraints over overload probabilities for multiple types of resources. It consists of multiple algorithms corresponding to different phases of VM dynamic consolidation. Our algorithms simulate artificial bee colony foraging behavior to find the mapping relation between PMs and VMs. By using the searching mechanism and optimization strategy of the artificial bee colony (ABC) algorithm [16], an approximate feasible solution is obtained via the iterative searches. Our major contributions, in detail, are as follows.

(1) First, through deeper research, we note that VM consolidation is a multi-objective optimization problem with multiple resource constraints;

(2) Then, by assuming the mapping relation between PMs and VMs as the food source, the proposed sub-algorithms integrate and cooperate to simulate the artificial bee colony foraging behaviors; by using both the searching mechanism and optimization strategy of the artificial bee colony ABC algorithm, the optimum mapping relation with multi-resource constraints between PMs and VMs is obtained;

(3) Next, the issue of "where from and where to" for live VM migration is globally optimized and examined. Results on achieving appropriate balance among guaranteeing service quality, reducing energy consumption and maximizing resource utilization are promising;

(4) At last, the proposed EC-VMC method overcomes the limitation of falling into local optima that some existing algorithms have, such as the well-known heuristic BFD (Best Fit Decreasing) algorithm. Validation and experimental comparison are conducted using the CloudSim platform. The experimental results indicate that the proposed EC-VMC method has a distinct advantage in terms of reducing energy consumption, VM migrations and improving QoS.

The rest of the paper is organized as follows. Section 2 presents the related works. Section 3 identifies and describes our optimization objective with multi-resource constraints. Section 4 presents the overview and the detailed design of our VM dynamic consolidation. Section 5 presents the performance evaluation of our scheme compared with other VM consolidation methods. Finally, Section 6 concludes this paper and discusses our future work.

2. Related works

A VM consolidation scheme should identify the VMs that should be migrated and the PMs that can be turned off. It subsequently solves the issue of to where the VMs should be migrated, i.e., the issue of identifying the source and destination hosts for the VMs in live migration. In summary, "where from where to" is the core issue of VM migration or VM consolidation. Many works [3,12,14,15,17–31] examine this issue from different perspectives, such as VM placement [19–22], host overload detection [3,12,22–27], and VM migration selection [22,28,29]. In this section, we mainly discuss single-resource VM consolidation and multi-resource VM consolidation from the aspect of resource allocation.

2.1. Single-resource VM consolidation

It is relatively simple to study host overload detection, VM migration and VM consolidation via the assignment of CPU resource. In previous studies [3,11], the upper and lower boundaries of CPU utilization were defined to classify PMs into three categories: underloaded, normal and overloaded. Some VMs were migrated from the overloaded physical machine to achieve the goals of performing load balancing, improving QoS and degrading the risk of overloading the CPU resource. The migrated VMs were redeployed in the normal PMs. All VMs in the underloaded physical machine were migrated, and the hosts were turned off to reduce energy consumption. However, the workload is random, and the host overload detection method, which is based on a fixed threshold, is incapable of adjusting the reserved idle resources according to the uncertain workload, which hinders the use of the VM consolidation method to properly allocate resources and causes undesired situations, such as poor service performance and high energy consumption.

Beloglazov A. et al. [22] made improvements based on the developments in [3] and proposed an adaptive heuristic VM dynamic consolidation method that analyzed the historical workload variation pattern of PMs and adaptively adjusted the overload threshold. Beloglazov A. et al. [22] proposed three host overload detection algorithms: the Median Absolute Deviation (MAD) algorithm, Interquartile Range (IQR) algorithm and Local Regression (LR) algorithm. The MAD and IQR algorithms measure the workload stability by calculating the median absolute deviation and interquartile range of recent CPU utilization, respectively. The overload threshold is decreased to reserve more resources for PMs with unsteady workloads. In this manner, the resource demands in the next phase are guaranteed and service quality is improved. However, the MAD and IQR algorithms disregard the recent workload variation trend. The result is that a physical machine with an unsteady load always needs to reserve a large number of resources, which actually decreases resource utilization and increases energy consumption. The LR algorithm forecasts CPU utilization via a local regression method and leverages the forecast value to proactively prevent physical machine overloading. However, for a highly fluctuating workload, obtaining an accurate prediction is difficult for the LR algorithm. To transfer workload by migrating some VMs from a physical machine with overload risk, Beloglazov A. et al. [22] proposed three VM migration selection algorithms: Minimum Migration Time (MMT), Maximum Correlation (MC) and Random Selection (RS). To realize VM placement, Beloglazov A. et al. [22] proposed the power aware best fit decreasing (PABFD) algorithm. First, the PABFD algorithm arranges the VMs in descending order based on the resource demand. Second, the migration of each VM is evaluated, and a physical machine with a minimum increase in energy consumption after VM placement is selected as the migration destination. The PABFD algorithm allocates the migrated VMs to the hosts with high energy efficiency; however,

Please cite this article in press as: Z. Li, et al., Energy-aware and multi-resource overload probability constraint-based virtual machine dynamic consolidation method, Future Generation Computer Systems (2017), https://doi.org/10.1016/j.future.2017.09.075.