Future Generation Computer Systems 50 (2015) 62–74

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Energy efficient scheduling of virtual machines in cloud with deadline constraint

Youwei Ding, Xiaolin Qin *, Liang Liu, Taochun Wang

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China

HIGHLIGHTS

- We develop a new VM scheduler to reduce energy cost for the cloud service providers.
- We deduce that there exists an optimal frequency for a PM to process certain VMs.
- We define the optimal performance–power ratio to weight the heterogeneous PMs in the cloud.
- The deadline constraint is satisfied by the definition of required resource of each VM.
- We achieve over 20% reduction of energy and 8% increase of processing capacity in best cases.

ARTICLE INFO

Article history:
Received 14 May 2014
Received in revised form
18 November 2014
Accepted 1 February 2015
Available online 11 February 2015

Keywords:
Energy efficiency
Virtual machine scheduling
Dynamic voltage and frequency scaling
Cloud computing

ABSTRACT

Cloud computing is a scale-based computing model, and requires more physical machines and consumes an extremely large amount of electricity, which will reduce the profit of the service providers and harm the environment. Virtualization is widely used in cloud computing nowadays. However, existing energy efficient scheduling methods of virtual machines (VMs) in cloud cannot work well if the physical machines (PMs) are heterogeneous and their total power is considered, and typically do not use the energy saving technologies of hardware, such as dynamic voltage and frequency scaling (DVFS).

This paper proposes an energy efficient scheduling algorithm, EEVS, of VMs in cloud considering the deadline constraint, and EEVS can support DVFS well. A novel conclusion is conducted that there exists optimal frequency for a PM to process certain VM, based on which the notion of optimal performance–power ratio is defined to weight the homogeneous PMs. The PM with higher optimal performance–power ratio will be assigned to VMs first to save energy. The process of EEVS is divided into some equivalent schedule periods, in each of which VMs are allocated to proper PMs and each active core operates on the optimal frequency. After each period, the cloud should be reconfigured to consolidate the computation resources to further reduce the energy consumption. The deadline constraint should be satisfied during the scheduling. The simulation results show that our proposed scheduling algorithm achieves over 20% reduction of energy and 8% increase of processing capacity in the best cases.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Cloud computing has been developed to store, manage and analyze the massive data. The initial aims of cloud computing are high performance, scalability, capacity, cost of infrastructure and so on, not including energy. With the growth of the number and the size of data centers, energy consumption becomes a challenge for both companies and governments. It is shown that the cost of energy consumed by a server during its lifetime will exceed the cost of server itself [1]. A report of US Environmental Protection Agency (EPA) indicated that IT infrastructures in USA consumed about 61 billion kWh for a cost of 4.5 billion dollars in 2006 [2]. This electricity consumption is about 1.5% of the total US electricity consumption, and is more than double of that consumed by IT in 2000. It was also noticed in [3] that servers consume 0.5% of the world's electricity produced, and it will quadruple by 2020 if the trend continues. However, the utilization of a typical data center is around 20%–30% [4], which means a large amount of energy will be wasted.

Virtualization is an important technology typically adopted in cloud to consolidate the resources and support the pay-as-you-go service paradigm. It has been reported that virtual machines could be used for scientific applications with tolerable performance

* Corresponding author.
E-mail addresses: dingyouwei@nuaa.edu.cn (Y. Ding), qinxcs@nuaa.edu.cn (X. Qin), liangliu@nuaa.edu.cn (L. Liu), wangtc@nuaa.edu.cn (T. Wang).

http://dx.doi.org/10.1016/j.future.2015.02.001
0167-739X/© 2015 Elsevier B.V. All rights reserved.

punishment, and could provide desirable, on-demand computing environments for any users. Adapting virtualizations in cloud computing, the platform often provides various virtual machine templates, the jobs will be allocated with preconfigured virtual machines once they arrive at the cloud, and then the virtual machine is started at proper physical machines, finally it will be shut down if the job is finished.

Virtual machine scheduling is one of the most important and efficient technologies of reducing energy consumption in cloud. Beloglazov et al. [5] reported a survey of energy efficient data centers and cloud, they classified the research into hardware, operating system, virtualization and data center levels, and introduced the techniques used to save energy consumption for each level.

The main idea of scheduling VMs energy efficiently is placing them on only part of the physical machines and transforming the other ones into low power state (sleep or off). Since the service providers own all the details of the physical machines in the cloud and the resource requirements of VMs from the users, they can place VMs to proper physical machines to minimize the energy and thus maximize their profit. Existing scheduling methods mainly focus on minimizing the number of physical machines used to run the VMs. However, less attention is paid to energy saving technologies in the hardware level though the information of hardware in the cloud is known to the provider.

Dynamic voltage and frequency scaling (DVFS) is an efficient technology to reduce processor energy consumption. Manufacturers of processor have developed their patents on DVFS to make their products operate on several frequencies with different supply voltages. DVFS has been mainly used to achieve energy efficiency in embedded, multicore and multiprocessor systems. However, DVFS technology is rarely adopted in the virtualized cloud systems to save energy. Since DVFS technology is adopted for the processors, it is most efficient for computation-intensive VMs, but does not suit for I/O-intensive or network-intensive VMs. We only focus on the computation-intensive VMs in this paper.

Existing VMs scheduling methods using DVFS to reduce total energy are mostly developed in homogeneous clusters, and only the power of processors are measured. For example, a power-aware algorithm of VMs scheduling [6] was proposed to allocate the virtual machines in a DVFS-enabled cluster. They use as low as possible frequency for the processor to run the VMs, and all physical machines are homogeneous. The criterion of selecting physical machines for each VM is the power of the processors, which does not fit the practical cloud, because the processors only consume about 25% of the total energy of the server [7].

Example 1. Given two computation-intensive VMs vm_1 and vm_2, and two servers n_1 and n_2 with the same processors but different other components. Suppose the power of the processor is P_{cpu} on fixed frequency f, and the power of the other components of n_1 and n_2 are $P_{s1} = P_{cpu}$ and $P_{s2} = 0.5P_{cpu}$ respectively. The servers operate on frequency f unless they are powered down, and the execution time of vm_1 and vm_2 are t_1 and t_2. Obviously, there are four solutions for this case: (a) vm_1 for n_1 and vm_2 for n_2, (b) vm_1 for n_2 and vm_2 for n_1, (c) vm_1 and vm_2 for n_1, and n_2 is powered down, (d) vm_1 and vm_2 for n_2, and n_1 is powered down. Therefore the energy of n_1 and n_2 for processing the two VMs are $E_a = (2t_1 + 1.5t_2)P_{cpu}$, $E_b = (1.5t_1 + 2t_2)P_{cpu}$, $E_c = 2(t_1 + t_2)P_{cpu}$, $E_d = 1.5(t_1 + t_2)P_{cpu}$. It can be seen that solution d is the optimal one in terms of energy consumption. However, on the condition that the deadline is less than $t_1 + t_2$, solution a is the optimal one if $t_1 < t_2$, otherwise solution b is the optimal one.

We can see that heterogeneities and the total power of the physical machines are the main challenges of energy efficient scheduling in cloud computing, while the adoption of DVFS technology is another challenge for the scheduling of virtual machines.

This paper focuses on dynamic scheduling of virtual machines to achieve energy efficiency and satisfy deadline constraints in the cloud with heterogeneous physical machines. The main contributions of this paper are as follows. We conduct that there exists optimal frequency for a physical machine to process certain virtual machines, and each PM should operate on at least the optimal frequency. Then the notion of optimal performance–power ratio is defined to weight the heterogeneities of the physical machines, VMs will be allocated prior to the PMs with higher optimal performance–power ratio. The scheduling is divided into some equivalent periods, and the cloud will be reconfigured after each period to consolidate the computation resource to further reduce the energy consumption. Finally, the deadline constraint is maintained by the definition of required resource, VM can be completed on time as long as it is allocated successfully to a PM.

The rest of this paper is organized as follows: Section 2 reviews the existing research on energy efficient cloud computing and scheduling of virtual machines. Section 3 defines the problem and describes the power and virtual machine models used in this paper. In Section 4, we present an energy efficient algorithm to schedule virtual machines in cloud computing using DVFS technology. Section 5 shows the simulation results for the proposed algorithm, and Section 6 concludes the paper and points out future work.

2. Related works

Energy efficient scheduling of tasks in cloud is studied widely. Jacob Leverich et al. [8] proposed a strategy of selecting part of the physical machines in a Hadoop cluster to execute the tasks while powering down other ones to reduce the power consumption. However, Willis Lang et al. [9] indicated that using all physical machines to run the workloads and then powering down them simultaneously can save more energy. Both the two methods cannot work well if workloads are data-intensive, because powering down some physical machines will result in data unavailability while using all ones will cause frequent data migrations. A replication scheme named Chained Declustering was used in [10] to ensure data available when powering down partial physical machines in a cluster, and it also guaranteed load balance between the active machines. Since the replication scheme is the basis of powering down physical machines and load balance, it is not suitable for the deployed clusters. Considering the energy consumption during both execution time and idle periods, powering down fractional machines is an accepted method to make clusters energy efficient.

Virtualization is widely used in cloud computing to fully utilize the resources and improve the performance. Various VM scheduling methods [11–13] have been proposed to dynamically allocate and consolidate the VMs in cloud computing environment. The allocation algorithms can be mainly divided into two types, allocating VMs onto PMs and assigning PMs to VMs. The consolidation is typically achieved by VM migrations. Energy consumption was not considered in traditional VM scheduling in cloud computing.

Energy efficient VMs scheduling in data centers mainly focuses on fully utilizing each physical machine to reduce energy consumption. They take a physical machine as a whole and use the formula $P = P_{idle} + (P_{\max} - P_{idle})^* u$ to compute the power of each node, where u is the utilization of a physical node, P_{idle} and $P_{\max}$ are the idle and peak power of the node. It is an experimental formula, and has been tested in many data centers for a period of time. It is the simplest to estimate the total energy consumption of the data center using this formula, but may be not the optimal solution for the computation-intensive VMs.

It is reported that we can save energy via appropriate VM scheduling [14]. Energy efficient virtual machine scheduling is often viewed as an allocation or mapping problem, which is an optimization problem. It was abstracted to multi-dimensional space

Download English Version:

https://daneshyari.com/en/article/425623

Download Persian Version:

https://daneshyari.com/article/425623

Daneshyari.com