
Li Mao, Department of Computer Science, Guang Dong Police Officer College, Guangzhou, China & School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
De Yu Qi, School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
Wei Wei Lin, School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
Bo Liu, School of Computer, South China Normal University, Guangzhou, China
Ye Da Li, Guangdong Engineering Polytechnic, Guangzhou, China

ABSTRACT

With the rapid growth of energy consumption in global data centers and IT systems, energy optimization has become an important issue to be solved in cloud data center. By introducing heterogeneous energy constraints of heterogeneous physical servers in cloud computing, an energy-efficient resource scheduling model for heterogeneous physical servers based on constraint satisfaction problems is presented. The method of model solving based on resource equivalence optimization is proposed, in which the resources in the same class are pruning treatment when allocating resource so as to reduce the solution space of the resource allocation model and speed up the model solution. Experimental results show that, compared with DynamicPower and MinPM, the proposed algorithm (EqPower) not only improves the performance of resource allocation, but also reduces energy consumption of cloud data center.

KEYWORDS

Cloud Computing, Constraint Satisfaction Problem, Energy Consumption, Resource Equivalence

1. INTRODUCTION

In recent years, with the rapid increasing of power consumption in global data centers and IT systems, high energy-consumption gradually becomes a prominent issue. Especially, with the arrival of the cloud computing era (Chen & Zheng, 2009; Foster, Zhao, Raicu, & Lu, 2008; W.-W. Lin & Qi, 2012), more and more computing and storage resources pile up in the cloud centers, which brings great challenges to the management of the energy consumption (Foster, Zhao, et al., 2008; Zhao-Hui & Qin-Ming, 2012). According to statistics (Hooper, 2008), about 6000 data centers consumed approximately 61 billion kilowatt-hours of electricity costing $4.5 billion in the United States in 2006, which is more than the total energy consumption of all televisions in the United States. The data from the United States energy department shows that the energy consumption in data centers accounts for 1.5% of all the energy consumption in USA, and the demand for electricity in date centers is still growing at an annual rate of 12%. By 2011, the data centers consume 100 billion kilowatt-hours of electricity costing about $7.4 billion per year. In addition, the market research to all the electricity cost in the whole world made by IDC, the International Data Corporation shows that the global enterprises would
cost about $40 billion a year on the energy consumption. High energy consumption in data center not only wastes too much energy and makes the running systems not stable, but also brings negative effects to the environment of nature. The federal agency of USA has pointed out that high energy consumption makes serious influences to the air quality, national security, climate changes and the reliability of power grid, and etc. Therefore, the energy consumption optimization has become an important issue to be solved urgently in current cloud data centers.

Based on the important virtualization technologies, cloud computing has become the delivery of the resources and supported the pay-as-you-go model. In a virtualized cloud data center, thousands of physical machines (PMs) are composited of a pool of computing resources to host the various virtual machines (VMs) that provide desirable environments for any users. VM migration is an efficient tool for resource provisioning, by dynamically rearranging the previous placement (Kangkang Li, Zheng, & Wu, 2013). Taking the advantage of VM migration, VMs save costs by reducing the need for PMs, and lower the power and cooling demands for more efficiently using hardware.

The energy-efficient scheduling VMs aims at sharing resources (CPU etc.) fairly among multiple VMs, assigning VMs to PMs in accordance with the algorithm, which minimizes the number of PMs used. In practice, the service providers of cloud data centers own all the details of the PMs in the cloud and different hardware requirements of VMs from the users, they can place VMs to proper physical machines to minimize the energy and thus maximize their profit (Ding, Qin, Liu, & Wang, 2015). It is important that an energy-efficient VM scheduling algorithm should consider the heterogeneous characteristics of PMs, such as the number of cores and the size of memory, because the servers in the cloud data center are always heterogeneous.

Many researchers applied virtualization technology to reduce power consumption of cloud data center. This technology allows one to consolidate several servers to one physical node as Virtual Machines (VMs) reducing the amount of the hardware in use. (Beloglazov & Buyya, 2010) Based on a model for estimating the energy consumption among virtual machine, (Kim, Cho, & Seo, 2014) proposed a virtual machine scheduling algorithm that can provide computing resources according to the energy budget of each virtual machine. (Ding, Qin, et al., 2015) proposed an energy efficient scheduling algorithm considering the deadline constraint: in each of equivalent schedule periods, VMs were allocated to proper PMs and each active core operates on the optimal frequency. A virtual machine placement and migration approach (Piao & Yan, 2010) was presented to minimizing the data transfer time consumption by optimizing the data transfer between the virtual machine and data. An energy-aware task consolidation (ETC) technique (Hsu, Slagter, Chen, & Chung, 2014) was presented to minimize energy consumption by restricting CPU use below a specified peak threshold. Three power-saving policies (Chiang, Ouyang, & Hsu, 2015) were implemented in cloud systems to mitigate server idle power. The paper also proposed an efficient green control (EGC) algorithm for solving constrained optimization problems and making costs/performance tradeoffs in systems with different power-saving policies. A strategy (Ouyang, Chiang, Hsu, & Yi, 2014) is presented to implement working vacation (WV) to lower and eliminate unnecessary power consumed by idle servers. Two green systems are also proposed. One implements a single WV and the other implements multiple WVs in an operational cycle. Control procedures in both green systems are mapped into Petri net-based models which contribute to designing a multiple decision process and describing system behaviors.

Constraint satisfaction problem (Haralick & Elliott, 1980) is a fundamental problem in the area of artificial intelligence, which is widely used in practice, especially used in the scheduling of allocating resources. An autonomic resource manager was used in (Van, Tran, & Menaud, 2009) to control the virtualized environment which decouples the provisioning of resources from the dynamic placement of virtual machines. Constraint Satisfaction Problems (CSP) are responsible for two phase problem: determining the VM allocation vectors for each application, and placing these VMs on PMs in order to minimize the number of active PMs. Similarly to (Van, Tran, et al., 2009), the manager aims to optimize a global utility function was presented in (Nguyen Van, Dang Tran, & Menaud, 2009), with a Constraint Programming approach to formulate and solve the optimization problem.
Agent-Based Infrastructure for Dynamic Composition of Grid Services
[www.igi-global.com/chapter/agent-based-infrastructure-dynamic-composition/40833?camid=4v1a](www.igi-global.com/chapter/agent-based-infrastructure-dynamic-composition/40833?camid=4v1a)

Communication Analysis and Performance Prediction of Parallel Applications on Large-Scale Machines
[www.igi-global.com/chapter/communication-analysis-and-performance-prediction-of-parallel-applications-on-large-scale-machines/159041?camid=4v1a](www.igi-global.com/chapter/communication-analysis-and-performance-prediction-of-parallel-applications-on-large-scale-machines/159041?camid=4v1a)