Research on OLES Algorithms Based on Energy Consumption in Cloud Computing

Guanghui Wei

Chongqing College of Electronic Engineering, Chongqing, 401331 China

qdbiq@163.com

Keywords: OLES Algorithms, Energy Consumption, Cloud Computing

Abstract: The energy consumption of cloud computing is studied. Because cloud computing system needs to consume a lot of energy in the process of running, and the more energy consumption, the more serious the environmental pollution. For most of cloud computing energy-saving algorithms are NP-Hard problems, but most of the previous studies did not take into account the life cycle of virtual machines, and the relationship between the number of virtual machines and the number of physical machines on energy consumption. In this paper, OLES algorithm is proposed to reduce cloud computing energy consumption. Energy consumption, operation cost reduction and environmental pollution reduction are of great significance to cloud computing operators, the whole country and even the whole human race.

1. Introduction

D.F. Parkhill boldly predicted in his classic book The Challenge of the Computer Utility that computing power would be available to the public like water and electricity. Indeed, today this prediction has basically come true. In the IT industry, cloud computing is the fifth public resource after water, electricity, gas and oil. Cloud computing, as a business computing and information service mode, distributes computing tasks on a large number of physical computer servers or virtual machines, enabling different users'various computing and storage needs to acquire computing and storage capabilities on these physical machines or virtual machines, thus providing information services for users. With only one device that can access the Internet, we can easily enjoy high-speed and high-quality services. As a result, cloud computing has also been called a "poor supercomputer".

With the continuous development of China's scientific and technological strength, energy consumption per unit GDP is also getting lower and lower. But even in 2016, China's energy consumption per unit GDP is still three times that of the United States and six times that of Japan. It can be seen that there is still much room for improvement in energy consumption per unit GDP in China.

2. Existing research results

Guo Bing and others described some examples and current situation of cloud computing, redefined and discussed green computing[1]. Liu Peng and others introduced the research background and challenges of cloud computing, including energy consumption. In the existing literature, the energy saving mechanism, energy consumption modeling and evaluation methods in green network are studied. Especially, the current research status and progress are analyzed in detail in four aspects: energy consumption measurement, energy consumption modeling, management mechanism and management methods for energy consumption management of virtualized cloud platform. Ten further studies are put forward. To solve the problem, several optimal scheduling algorithms are proposed based on M/M/1 queuing model, and the scheduling management of basic resources in cloud computing is introduced in detail. Lee et al considered integrating real-time task requests to maximize utilization and energy saving[2]. Two real-time heuristic scheduling

  DOI: 10.25236/eeems.2018.007

algorithms always choose to assign tasks to maximize current resource utilization to save energy.

Khandekar et al. further designed an offline scheduling algorithm with approximation of 5, which can be used to schedule tasks with fixed processing time constraints under arbitrary request capacity[3]. For some special cases, an online scheduling algorithm with competition ratio of physical machine capacity G is designed, and for some special cases, the competition ratio is smaller than that of physical machine capacity G. Scheduling algorithm. In the existing literature, the optimization and approximate optimization modeling method considering energy saving and performance in server cluster environment is discussed[4]. It is suggested that the product of minimizing energy consumption and response time be used as a metric to measure different scheduling algorithms[5]. Tian et al. put forward the framework of managing and implementing the platform in virtual computing environment as a service, which mainly realizes the basic functions of user management, resource management and access management[6]. By virtualization and 7 *24 hours of remote online service, the sharing and utilization of resources can be improved, and at the same time, it is convenient for more users to use[7]. Tian et al. introduced the physical machine quantity allocation analysis model considering different random queuing models with satisfactory quality of service; introduced several off-line scheduling algorithms in the existing literature and compared their energy-saving effects; used the method of minimizing the total running time of all physical servers in parallel task scheduling An off-line energy-saving scheduling algorithm with approximation of 3 is proposed, and the related results are applied to the energy-saving scheduling of cloud data centers. For the dynamic scheduling algorithm, an on-line dynamic dichotomy scheduling algorithm with better energy-saving effect than the current general one is proposed.

3. Energy Consumption Model

Users submit requests for the use of virtual machines to the scheduling system. After receiving the virtual machine requests submitted by users, the scheduling system will coordinate the physical servers according to the requirements of the task requests and the current utilization of resources of the physical servers in the cloud computing center, and then deploy the corresponding virtual machines to the physical servers. It is for user's use.

The energy-saving algorithm mainly decides which physical server to allocate virtual machine requests to, and at the same time combines appropriate migration to minimize the energy consumption of the entire data center. Requests submitted by users include information about the start time of the request, the duration of the request, and the virtual machine specifications of the request.

The total energy consumption of cloud computing center is composed of physical server energy consumption, refrigeration system energy consumption, lighting energy consumption and network equipment energy consumption. Physical server energy consumption and refrigeration system energy consumption are the main energy consumption, while other energy consumption is the secondary energy consumption. Total energy consumption can be expressed by the following formula:

$$P_{total}=P_{pm}+P_{AC}+P_{odditional} \tag{1}$$

Among them, P_{total} is the total energy consumption of cloud computing center, P_{pm} is the total energy consumption of Cloud Computing Center server, P_{AC} is the energy consumption of refrigeration system, Podditional refers to other secondary energy consumption.

4. Energy Saving Algorithm

The results of energy-saving modeling and scheduling algorithm studied in this paper are based on the following preconditions:

1) The system start time S0 = 0, and the time interval of task request I can be time slot format [start time, end time, processing capacity]= [si, ei, di], where start time Si and end time EI are non-negative integers.

2) All tasks are relatively independent, and there are no other preemptive modes except the specified start and end times, where priority constraints are not taken into account.

3) When each task request is allocated to a physical machine, it is allocated to a single physical machine.

4) Each computer does not consider damage, that is to say, it is guaranteed to be continuously available.

Definition 1 gives a time interval I=[s, c], and the length of I is [C, S]. Extended to Paired Discontinuous Interval Joint $I = \bigcup_{i=1}^{k} Ii$, The length of I is $|i| = \sum_{i=1}^{k} |I_i|$.

Define 2 pairs of any interval set I, whose span is $sp(I) = |\bigcup I|$ and the length of I is $len(I) = \sum_{i=1}^{k} |I_i|$.

Definition 3 represents the cost of the optimal solution for any task scheduling instance J and the degree of parallelism g (> 1), which minimizes the total processing time of the computer.

For a given batch of virtual machine scheduling tasks, minimizing the total energy consumption of the data center is equivalent to minimizing the total running time of all the physical machines in the data center under the worst case and physical machine isomorphism, while meeting the system capacity and performance constraints.

OLES algorithm analysis. In the MI problem, the start and end times of jobs are fixed, and the general case of parallelism g > 1 is considered. When allocating virtual machines, we need to consider fixed start time, fixed end time, and the limitation of parallelism. The flow chart of MFFDE algorithm is as follows. The jobs are arranged in non-ascending order with the same processing time. If the processing time is equal, the start time is earlier. If both of them are the same, they can be arranged arbitrarily. The computational complexity of OLES algorithm is O (nmax (m, log2n), where n is the number of jobs and M is the number of physical machines used. It takes O (nlog2n) time to arrange jobs in non-ascending order with processing time. Next, a physical machine will be found to assign the jobs. Here, O (m) steps are needed, and a total of N jobs need O (m) steps. Therefore, the complexity of the algorithm is O (nmax (m, log2n), usually n > M.

OLES algorithm flow:

1) Input: (J, g), J is the task set, G is the maximum parallelism of the physical machine.

2) Output: Scheduled jobs, total busy processing time of all physical machines, total number of physical machines used

3) Arrange all jobs in non-ascending order with the same processing time (the jobs with the same processing time and the jobs with earlier start time are in the first place; if both are the same, they are arranged randomly);

4) for i=l to n do

Find the first physical machine I that can satisfy the parallelism.

Assign job J to physical machine I and update the load of physical machine

5) Calculate the workload and the busy processing time of all physical machines

When all long jobs are assigned to physical machine M1, if OLES (J1) = span (J1) has upper bound OPT (J), then the optimal OPT (J) is mainly determined by OLES (J1). In this case, the allocation of other physical machines has little effect on OPT (J), and $\frac{3}{g} w(J_m)$ is also small (negligible compared with span (J1), otherwise OLES (J1) = span (J1) cannot reach the upper bound OPT (J). $\sum_{i=1}^{m} OLES(J_i)$ is mainly determined by span (J1), which is close to or equal to OPT (J).

(2) If OLES (J1) = span (J1) is smaller than OPT (J), for example, when OPT (J) is not determined by OLES (J1), we consider the worst case, because it is the upper bound. In the worst case, $span(I_i + 1) \leq \frac{3w(I_i)}{g}$, so we can easily get $OLES(J_i) < \frac{3}{g} w(J_m)$.

5. Summary

Off-line energy saving scheduling is an ideal off-line algorithm, which can minimize the energy consumption of cloud data centers.

References

[1] Bruns chwiler T, Smith B, Ruetsche E, et al. Toward zero-emission data centers through direct reuse of thermal energy. IBM J Res Dev, 2009, 53(3): 11. 13.

[2] Eduardo P, Ricardo B, Enrique V, et al. Load Balancing and Unbalancing for Power and Performance in cluster-based Systems. Proceedings of the Workshop on Compilers and Operating Systems for Low Power (COLP), 2001.

[3] M. Flammini, G. Monaco, L Moscardelli, H. Shachnai, M. Shalom, T. Tamir and S. Zaks Minimizing Total Busy Time in Parallel Scheduling with Application to Optical NetworksTheoretical Computer Science, 2010, 411(40-42): 3553-3562.

[4] A. Gandhia, V. Guptaa, M. harchol-baltera, M. A. Kozuchb. Optimality Analysis ofenergy-performance trade-off for Server Farm Management Performance Evaluation (2010)1-23.

[5] Lee YC, Zomaya AY. Energy Efficient Utilization Of Resource In Cloud ComputingSystems. J. Super Comput. Publised online March 19, 2010.

[6] Pakbaznia E, Ghasemazar M, Pedram M temperature-aware dynamic resource provisioning in a power-optimized datacenter. In Proceedings of the conference ondesign, automation and test in Europe DATE10), Dresden, Germany, pp 124-129.

[7] Shalom, M, Voloshin, A, w.h. Wong, P. w.h., Yung, f.c. C, Zaks, S. Online Optimization of Busy Time on Parallel Machines. THEORY AND APPLICATIONS OF MODELS OFCOMPUTATION, Lecture Notes in Computer Science, 2012, Volume 7287/2012, 448-460.