Advances in Computer Science Research (ACSR), volume 90
3rd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2019)

Research on Energy Saving Algorithm of Datacenter in Cloud Computing System

Xiaozhao Liu, Jian Wu*

Suzhou Institute of Trade & Commerce, School of Mechanical & Electric Technology, Suzhou, Jiangsu; 215009 China

* 181573466@qq.com

Abstract. Energy consumption of datacenter in cloud computing system not only increases the operating costs of cloud service providers, but also has gradually evolved into an environmental issue. The definition and characteristics of cloud computing are described briefly, and the high energy consumption problem of cloud datacenter is focused. Based on the classification of the energy saving algorithm, the three kind of energy saving algorithms are researched in a big way, including the energy saving algorithm based-DVFS, energy saving algorithm based-virtualization and the energy saving algorithm based-turn off/on of hosts, and we make a comparative analysis of the advantages, disadvantages and applicable scene of these algorithms. Finally, we sum up the further research problems for energy consumption management in cloud computing datacenter.

Keywords: cloud computing system; DVFS; energy saving algorithm; high energy consumption.

1. Introduction

Cloud Computing [1,2], as the latest computing model, provides dynamic and flexible architecture and QoS services for users' pay-as-you-go behavior. It can be defined as an Internet-based computing method. In this way, the relevant IT capabilities based on the Internet can be provided to users in a service way. Users need not know the details of providing cloud services, nor need to have corresponding expertise and equipment operation capabilities. They can obtain the required services through the Internet. The whole operation mode is like the power grid [3]. Cloud computing has the characteristics of super-large scale, virtualization, reliable data storage, high sharing, scalability and low cost. At present, many large IT vendors have launched their own cloud computing platforms. The cloud computing infrastructure such as Google App Engine, Amazon EC2, IBM Blue Cloud and Microsoft Azure are widely used. Cloud computing research focuses on massive data storage, resource management, virtualization, task scheduling, data center energy-saving algorithms and cloud security. This paper will focus on the energy consumption problem and energy-saving algorithm of cloud system data center.

Energy consumption has become a key issue restricting the development of cloud computing. EPA reports that 1.5% of total energy pumps in the United States were used in data centers in 2006, and this trend will become increasingly serious. The April 2007 report noted that the ICT sector contributed 2 per cent of global CO2 emissions, on a par with the aviation industry. At the same time, the research shows that the decrease of CO2 emission by 15%-30% can ensure that the global temperature rise will be less than 2oC before 2020. In 2010, Google emitted 1.46 million metric tons of CO2, and 100 searches used the equivalent of 60-watt bulbs for 28 minutes. It can be seen that solving the problem of high energy consumption in cloud system data center has become an environmental problem.

2. Classification of Energy Saving Algorithms in Data Center

At present, the energy-saving mechanism of cloud computing has different classification according to different classification standards. According to the different ways of power management, it can be divided into two categories: dynamic power management (DPM) technology and static power management (SPM) technology. According to the different stages of energy consumption reduction, it can be divided into three categories: Resource Hibernation, DVFS and Virtualization. The former mainly reduces idle energy consumption, while the latter two mainly execute energy consumption.

The main premise of DPM is that the load conditions faced by cloud systems are changing over time, which allows dynamic adjustment of power status according to current performance requirements. There are mainly DVFS and virtualization strategies using DPM technology. In contrast, SPM mainly uses efficient hardware devices, such as CPU, hard disk storage, network equipment, UPS and energy supply equipment, etc. The structural changes of equipment can usually reduce energy consumption.

3. Energy-saving Algorithm based on DVFS

DVFS (Dynamic Voltage and Frequency Scaling), Dynamic Voltage and Frequency Adjustment is one of the effective energy-saving ways to control CPU power consumption on modern processors. Its main idea is that when CPU is not fully utilized, CPU performance can be reduced actively by reducing the supply voltage and clock frequency of CPU, which can bring about a reduction of dynamic power consumption of cubic magnitude without affecting performance. For example, considering that a system has only one task to execute, and the load requires 10 cycles to complete, and the deadline of the task is 100 seconds, in this case, the processor may slow down to 1/10 cycles/s, which meets the deadline and saves energy consumption. This dynamic frequency reduction process is more effective than running tasks at maximum speed and idling the CPU for the rest of the time. DVFS has been widely used in CPUs including mobile terminal devices and servers.

Literature [3] proposes three energy-aware virtual machine delivery strategies based on DVFS in real-time cloud service environment. The cloud service model is defined as: $\{T_i(r_i,c_i,d_i,p_i,f_i)|i=1,2,\ldots,n\}$ n is the number of tasks, r_i representing the start time of task i, c_i the execution time of the worst situation, d_i the relative deadline, p_i the cycle and f_i the completion time. T_i Subtasks start in time r_i, and c_i the worst execution time is required. In order to achieve the application goal, T_i should be completed in time r_i+d_i. p_i The cycle is given T_i so that the sub-task of execution time r_i+kp_i can be opened again and completed on time. The virtual machine model is defined as $\{V_i(u_i,m_i,d_i)|i=1,2,\ldots,n\}$, u_i representing the CPU utilization of application requests, m_i the virtual machine benchmark MIPS, and d_i the deadline. After users submit services to virtual machines, resource providers use DVFS to provide virtual machines to reduce energy consumption. The following are three energy-saving strategies based on DVFS:

1)Lowest-DVFS: Adjust the CPU rate to the lowest and meet the virtual machine deadline. Each virtual machine executes tasks with the requested MIPS. This strategy consumes the least energy when the task arrival rate is low and all requests can be received.

2)δ-Adanced-DVFS: In order to overcome the problem of low service reception rate of Lowest-DVFS strategy, this strategy improves the MIPS of current virtual machine requests by δ%. δ% Definition in advance according to system load: $s=\min\left\{1,(1+\frac{\delta}{100})\times\frac{1}{Q_k}\sum_{V_i\in T_k}\frac{w_i}{d_i-t}\right\}$

3)Adaptive-DVFS: When the arrival rate of requests and their service time are predicted in advance, the optimal scaling model can be obtained through analysis.

Literature experiments show that the performance of δ-Adanced-DVFS is the best in terms of revenue per unit power consumption. Because CPU performance can be adjusted automatically according to system load, the performance of Adaptive-DVFS is limited by the specific queue model.

Literature [1] Based on DVFS technology, a heuristic scheduling algorithm is proposed to reduce the energy consumption of parallel tasks in cluster environment. This algorithm aims at tasks on non-critical paths in parallel task graphs, and reduces the voltage of CPU scheduled by non-critical tasks without affecting the completion time of the whole task. Document [1] also makes a similar study, proposes an energy-aware task scheduling algorithm in cluster environment, and proposes to reduce power consumption by controlling appropriate voltage levels.

Reference [2] For embedded multiprocessor systems, proposed heuristic task scheduling algorithms EGMS and EGMSIV based on energy consumption perception. The algorithm also considers the dynamic adjustment of task scheduling sequence and voltage, and proposes using energy consumption gradient as the evaluation index of task scheduling. EGMSIV realizes the dynamic adjustment of voltage when the same task is executed on the basis of EGMS.

Literature [4] studies the energy minimization of periodic preemptive hardware real-time tasks in multi-processor platforms with DVS capabilities. This paper uses a piecewise scheduling mechanism to assign a static priority to each task. Once the task is assigned to the processor, the processor rate allocation mechanism will be activated to reduce energy consumption and ensure flexibility.

DVFS is designed to reduce execution energy consumption. Execution energy consumption is defined as the energy consumption generated by the operation of instruction and data-driven computer hardware when tasks are running on a computer. During the execution of the same task, its execution power will change with the change of operation stage and execution characteristics . Because the power of dynamic power of CMOS circuit is proportional to the voltage and frequency, reducing the voltage or frequency of CPU can reduce the executive power of CPU. The disadvantage of this method is that when the voltage or frequency of the CPU is reduced, the performance of the CPU will also be reduced.

4. Energy-saving Algorithm based on Virtualization

Another way to improve resource utilization and reduce power consumption is virtualization. Virtualization technology is one of the key technologies in cloud computing. It allows multiple virtual machines to be created on a single host, thus reducing the amount of hardware resources used and improving resource utilization. Virtualization can not only isolate the performance of applications sharing the same computing nodes, but also realize the migration of virtual machines from one node to another using dynamic or offline migration technology. Real-time virtual machine redistribution can achieve dynamic load consolidation, so that virtual machines can be merged into a smaller number of physical nodes, thus switching idleness. Set nodes as energy-saving mode.

Document [4] combines energy consumption management technology with virtualization technology, and develops an energy consumption optimization management method, Virtual Power, for large-scale data centers. This method supports the energy consumption control method of virtual machines running independently, and can reasonably coordinate the energy consumption control requests between different virtualization platforms and between different virtual machines on the same virtualization platform to achieve energy consumption. Overall optimization.

Literature [3] uses constraint satisfaction problem to model virtual machine deployment, and the constraint is the user's service level agreement, so as to maximize energy saving. The idea is to maximize the number of idle physical machines and save energy by closing the idle physical host.

Virtual machine deployment problem is an example of multi-knapsack problem. Using constraint satisfaction problem can easily introduce additional application constraints, such as collocation constraints and similar constraints. This strategy considers the deployment of virtual machines from the aspect of constraint satisfaction, and achieves the goal of minimizing energy consumption on the premise of satisfying the user service level agreement. However, the strategy also seeks the utilization of resources itself, and when the problem domain is too large, the strategy may not guarantee the optimal solution.

Document [1], by formalizing the virtual resource allocation problem in cloud computing environment as a path construction problem, proposes a high efficiency allocation strategy EEVRAS. The strategy uses ant colony system with limited elite strategy to generate optimized resource allocation scheme, which reduces the number of servers and the energy consumption of the system.

The above strategies are all aimed at reducing the number of servers used. Literature [2] proposes a method of combining dynamic migration of virtual machines with closing idle computing nodes to improve the utilization of physical resources and balance the demand between power and performance. This paper proposes a dynamic virtual machine redistribution mechanism, which first ranks the utilization of resources and migrates virtual machines on physical hosts with less utilization of resources to the principle of minimum power increase. The idle physical host is turned off to save energy. The migration criterion given in the literature is to set a fixed occupancy rate of host resources.

The goal of virtual machine migration is to keep the occupancy rate between the upper and lower limits of occupancy rate. The basic idea is that when the upper limit is violated, HPG migrates a set of virtual machines with the lowest CPU occupancy relative to the CPU capacity defined in the parameters of the virtual machine, which is defined as:

$$R = \begin{cases} \left\{ S \middle| S \in P(V_j),\ u_j - \sum_{v \in S} u_a(v) < T_u, \sum_{v \in S} \frac{u_a(v)}{u_r(v)} \to \min \right\} & , \ if\ u_j > T_u; \\ V_j, \quad if \quad u_j > T_l; \\ \varnothing, \quad otherwise \end{cases}$$

$u_r(v)$ is the CPU capability of the initial request of the virtual machine v, it is defined as a virtual machine parameter.

Although the proposed threshold can control the CPU occupancy of the host, the optimal threshold cannot be obtained through analysis. The approximate optimal threshold can be obtained through many experimental analyses, and the cost of virtual machine migration is not discussed in this paper.

Literature [2] developed a hierarchical energy consumption control system, including host-level and user-level subsystems. The former allocates the hardware resources reasonably according to the user's request, facilitating the energy consumption of each virtual machine to not exceed the prescribed upper limit, and then reallocates the virtual hardware resources in the virtual machine layer, so that the energy consumption generated by each user's task does not exceed the prescribed upper limit.

Virtualization technology improves the utilization of physical resources and reduces the energy consumption of execution, but its drawback is that it increases the complexity of system management. With the introduction of virtualization technology, resource management will involve two aspects: Virtual Machine Provision (VM Provision) and Virtual Machine Placement (VM Placement). The former is mainly responsible for allocating resources to application tasks in the form of virtual machines. This stage is generally related to the SLA of hosted applications, and the delivery mechanism is driven by performance objectives, and the choice of performance objectives is general. For the average task response time, the job volume per unit time is described; the latter mainly completes the mapping between virtual machines and physical hosts, and the deployment of virtual machines is driven by the strategy of data centers related to resource management costs.

5. Energy Saving Algorithms based on Host Open/Close

Energy-saving strategies based on host shutdown/turn-on technology can be classified into three categories: random strategy, timeout strategy and predictive strategy. The stochastic strategy regards the closing/opening time of the server as a stochastic optimization model, and uses the stochastic decision model to design the control algorithm. The timeout strategy presets a series of timeout thresholds. If the duration of idle time exceeds the threshold, it will turn the server world into shutdown mode. At the same time, the threshold can be fixed or adjusted adaptively with the system load. The predictive strategy predicts the idle time at the initial stage, and once the predicted value is large enough, it switches directly to the closed mode. The goal of the three strategies is to minimize the idle energy consumption, but the disadvantage is that when the computer starts for a long time, it will lead to a certain degree of performance degradation. The main studies of such strategies are as follows:

Document [4] proposes that the optimal energy-saving strategy based on closing/opening technology is timeout strategy because of the self-similarity of business requests in computer systems. It also proposes robust and effective estimation algorithm for shape parameters of Pareto distribution and non-stationary business based on window size adaptive technology when idle time length service

Pareto distribution is based on truncated mean method with small samples. The DPM control algorithm is obtained.

Literature [3] introduces load sensing mechanism and proposes a dynamic energy-saving algorithm in virtualized computing platform. It uses dynamic provisioning mechanism to close unnecessary host subsystems to achieve energy-saving. Compared with stochastic strategy and predefined timeout strategy, this method can improve energy consumption efficiency on the premise of guaranteeing quality of service.

Energy-saving technology based on host shutdown/turn-on can be combined with virtual machine migration method in virtualization. When load information can be predicted, this method can greatly save idle host energy consumption.

6. Other Energy Saving Algorithm

Cooling system accounts for about 40% of the total energy consumption of cloud data center. The high-speed operation of computing resources leads to the increase of equipment temperature. Excessive temperature not only reduces the reliability of data system, but also reduces the life cycle of equipment. Therefore, it is necessary to cool down the cooling equipment of cloud data center and effectively reduce the cooling energy, which is important for the stable operation and power saving of cloud data center. Significance. Literature [2] proposed that the cooling system of data center should be taken into account. Variable speed fans and temperature sensors should be installed in the server, and the speed of fans should be adjusted according to the temperature of the server, so as to maintain thermal safety and save electricity. Document [4] proposes that the instruction data stream of data center includes temperature sensor data, server instructions and air conditioning unit data in data center space. This paper analyses these data streams and presents a simple and flexible model. According to the given load distribution and cooling system configuration, the thermal distribution of data center can be predicted, and then the thermal load management system can be configured statically or manually. Literature [1] proposes a data center level power and hot spot management solution. PTM (power and thermal management) engine determines the number and location of active servers, and adjusts the cooling temperature to improve the energy efficiency of the data center. Literature [2] presents a fine way to control fans in data centers. The fans on each rack adjust the speed of fans according to their own thermal system, hardware usage and other information. Document [4] Considering space size, rack and fan placement and air flow direction, a multi-level design idea of data center cooling equipment is proposed, and the air-to-air flow and heat exchange are modeled and simulated to provide theoretical support for data center layout. In addition, after the completion of the data center, dynamic cooling strategy can be used to reduce energy consumption. For example, for dormant servers, some refrigeration facilities can be properly closed or the direction of air conditioning can be changed to save costs. Document Aiming at the problem of unbalanced heat distribution in cloud data center, it is the first time to use wireless multimedia sensor networks (WMSNs) to monitor local hot spots in real time, and to reduce the heat load of hot spots by means of task migration. The goal of balancing heat distribution and improving refrigeration efficiency is achieved by taking "hot spot discovery-hot spot location-feature extraction-hot spot elimination" as the main line.

In addition, Dynamic Component Deactivation (DCD) is also an important energy-saving means used in hardware.

DPM uses real-time resource utilization and application load to optimize energy consumption. The disadvantage of DPM is that if the load is always in peak state, power consumption cannot be reduced.

7. Concluding Remarks

This paper mainly studies the energy-saving algorithms of cloud computing system data center, and classifies and compares the typical algorithms currently used. Summarizing the work, we can see that each algorithm has its own advantages and disadvantages based on different application environments. Based on the existing research, there are still some further research focuses in the energy consumption management of data center: 1) how to decide the running status of physical hosts in a given real cloud computing system according to task type, arrival rate and distribution, and how to optimize the energy efficiency of the system by combining DVFS and host shutdown/turn-on technology; 2) cloud is service-oriented, which must satisfy certain quality of QoS service. Requirements, how to define a QoS/energy efficiency model to measure the energy consumption optimization objectives of cloud systems, and clarify the master-slave relationship between them, are also issues that need further attention in the future.

Acknowledgements

Foundation Project: The Natural Science Foundation of the Jiangsu Higher Education Institutions of China(18KJB510042).

References

[1]. von Laszewski, G., W. Lizhe, A.J. Younge, et al. Power-aware scheduling of virtual machines in DVFS-enabled clusters. IEEE International Conference on Cluster Computing and Workshops, 2009. CLUSTER '09. 2009: p. 1-10.

[2]. Goh, L.K., B. Veeravalli, S. Viswanathan. Design of Fast and Efficient Energy-Aware Gradient-Based Scheduling Algorithms Heterogeneous Embedded Multiprocessor Systems. IEEE Transactions on Parallel and Distributed Systems,2009, 20(1):1-12.

[3]. AlEnawy, T. A., H. Aydin. Energy-aware task allocation for rate monotonic scheduling. in 11th IEEE Real Time and Embedded Technology and Applications Symposium, 2005: p. 213-223.

[4]. Blume, H., J. v. Livonius, L. Rotenberg, et al. OpenMP-based parallelization on an MPCore multiprocessor platform - A performance and power analysis. Journal of System Architecture, 2008, 54(11):1019-1029.