Energy and Power Reduction of Cloud Data Centers

Shoufei Gan, Guolong Chen, Xianwei Li

School of Information Engineering, Suzhou University, Suzhou 234000, Anhui, China

Abstract
As a new computing paradigm, cloud computing has received much attention in the past few years, and the demand for cloud resources is skyrocketing. For example, Liu et al. proposed a high-priority and high response scheduling algorithm. Yu et al. studied Intelligent Transportation Control System Based on Cloud Computing. In order to meet the pressing demand for cloud resources, more and more data centers will be built in the future. There are typically tens of thousands of servers in cloud data centers, for instance, it has been reported that Google has more than 900,000 servers. However, large data centers are major consumers of energy and main contributors to carbon emissions. Moreover, recent studies show that energy costs have become a major concern for cloud providers and data centers are far from energy efficiency. Therefore, how to reduce energy and power consumption of these servers is a big challenge for the owners of cloud data centers. In this paper, we make a survey of energy reduction techniques of cloud data centers. Especially, we survey energy reduction methods from different levels of cloud data centers.
Key words: ENERGY COST, CLOUD DATA CENTERS, DATA CENTER NETWORKS, DVFS

1. Introduction

As a new computing paradigm, cloud computing has received much attention in the past few years, and the demand for cloud resources is skyrocketing. In order to meet the pressing demand for cloud resources, more and more data centers will be built in the future. There are typically tens of thousands of servers in cloud data centers, for instance, it has been reported that Google has more than 900,000 servers, in order to meet the need for cloud resources. However, large data centers are major consumers of energy and main contributors to carbon emissions and there are far from energy efficiency according to the studies of H.H. Kramer et al.[1] and J.A. GallegoArrubla et al.[2]. Moreover, recent studies show that energy costs have become a major concern for cloud providers. Vasan et al.[3] pointed out that the average utilization of CPUs in data center servers is only about 12.5%, and a large portion of energy can be saved if the CPU utilization is 86%.

The number of data centers is increasing rapidly throughout the world, which is motivated by the proliferating demand for cloud computing services. As a consequence, the expenditures spent on cloud data centers are skyrocketing in recent years. Figure 1 shows the cost of data centers from 1994 to 2010 and Figure 2 shows the power consumption in different parts of data centers from 1994 until 2014, L. Parolini et al.[4].

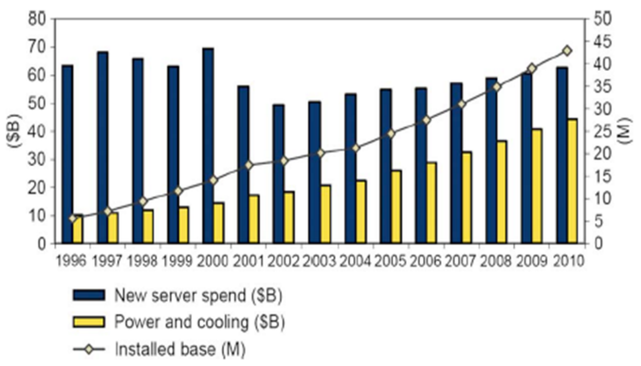

Figure 1. Expenditure cost in data centers

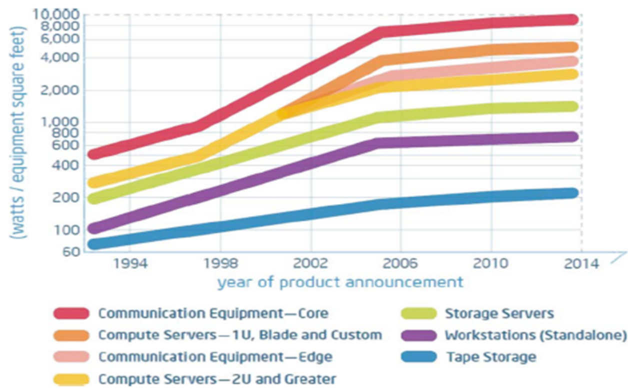

Figure 2. Comparison of positioning errors of different anchor nodes

For cloud providers, they mainly face two problems. The first problem is the concern of large power consumption and the electricity costs, e.g., Google spends more than 38 million dollars per year, A. Qureshi et at. [5]. In this paper, making a survey of energy and power reduction techniques in cloud data centers, especially, we analyze from different levels of data centers.

2. Data center networks

Data center networks play an important role of interconnecting a large number of servers with networking devices, e.g., switches, and high-speed communication links. Traditionally speaking, data center network architectures can be classified into different types, for instance, two-tiered, three-tiered, and three-tiered high speed, D. Kliazovich et al. [6]. Figure 3 shows the three-tiered designed data center networks structure, M. Al-Fares et al. [7]. Cloud providers deploy these network architectures according to the requirements of the applications that data centers can host.

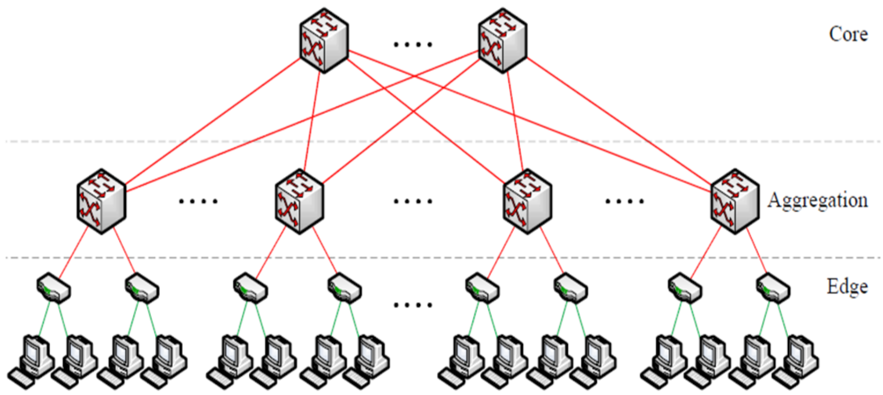

Figure 3. Three-tire data center networks structure

With the development of data centers, the energy consumption in data center networks is becoming a concern that should not be neglected. It has been reported that 10–20% of the total power was caused by the consumption of the network elements in the data center, W. Fang et al.[8]. While most of the current works have overlooked the energy consumption of data center networks, there are some works emerging to study this issue. Fang et al, [8] proposed to reduce power cost by optimal virtual machine (VM) placement and optimizing traffic flow routing, and Figure 4 shows the adopted optimization methods. Their study showed that the proposed method can save more energy than the baseline method elastic tree structure. M. Al-Fares et al. [7] proposed a flat-tree topology of data center networks, which is shown in Figure 5.

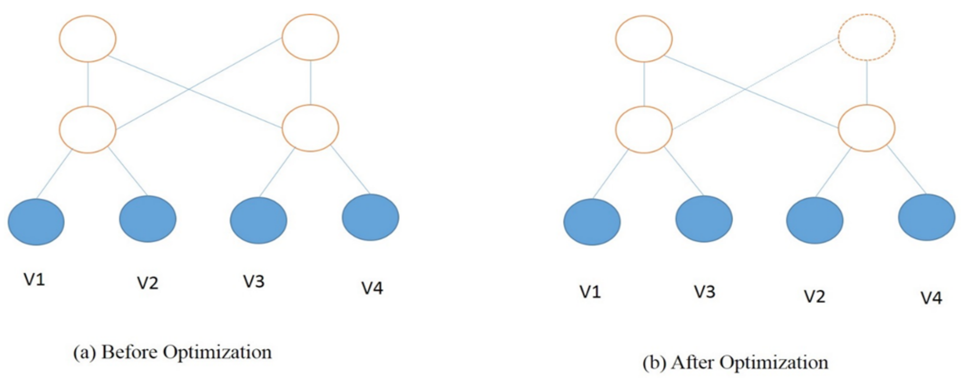

Figure 4. Optimization illustration

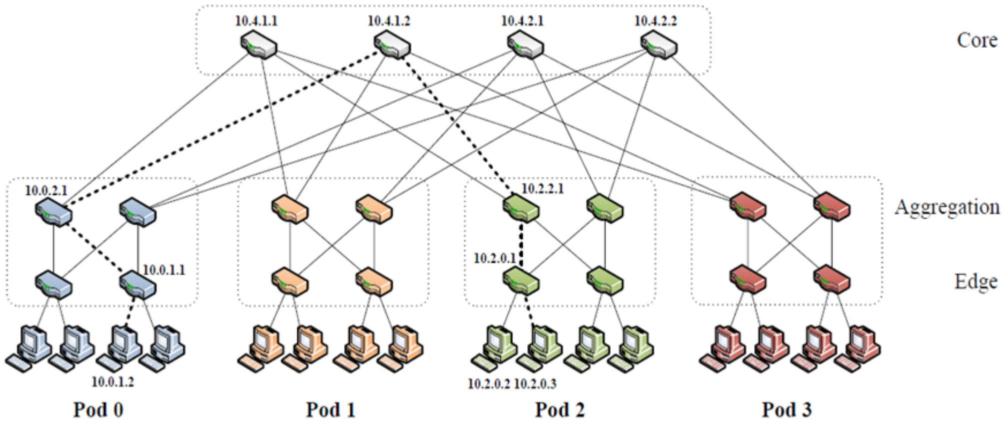

Figure 5. Simple flat-tree topology of data center networks

A. Greenberg et al. [9] proposed a scalable and flexible data center network, which is called VL2. And this new network structure is a simple design and can benefit the cloud service programmer and is efficient. C. Guo et al. [10] presented DCell, a novel network structure shown in Figure 6, which can provide higher network capacity than the traditional data center networks structures.

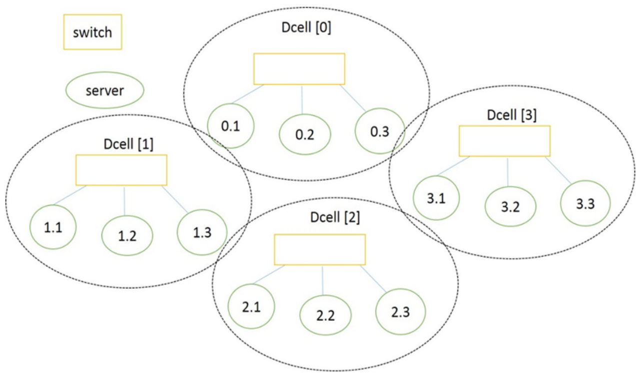

Figure 6. A Dcell data center networks structure

3. Methods for reducing power of CPUs

Power consumption caused by CPU accounts for a large part of total power in cloud data centers, especially for computing sensitive applications. Improving CPU consumption is an effective method to reduce the operating costs of cloud providers. Vasan et al. [3] pointed out that the average utilization of CPUs in data center servers is only about 12.5%, and a large portion of energy can be saved if the CPU utilization is 86%. R. Urgaonkaret at. [11] Studied optimal resource allocation and power management in virtualized cloud data centers with computing sensitive heterogeneous applications by leveraging the technique of Lyapunov Optimization in contrast to the previous works using prediction based approaches. They analyzed how CPU consumption varies with CPU frequency, which is shown in Fig.7. Their analysis indicated that at each utilization level, the power-frequency relationship is well-approximated by a quadratic model, which can be expressed as

$$P(f)=P_{min}+\alpha(f-f_{min}) \qquad (1)$$

Figure 7 showed that the minimum power Pmin should be substantial enough in order to maintain the server to be in active state.

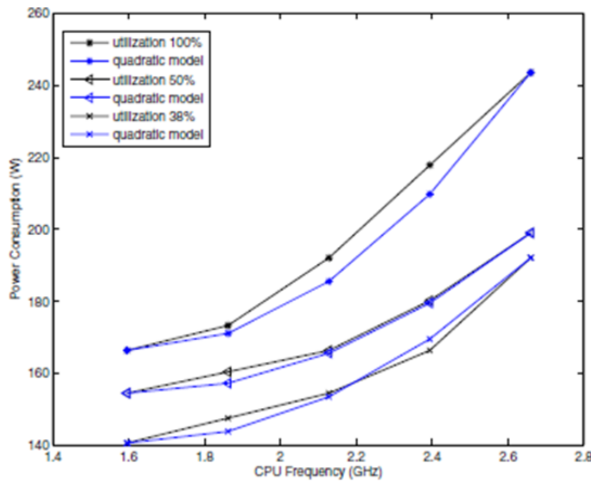

Figure 7. Power vs CPU frequency for a Dell PowerEdge R610 server

Another model used to denote the power consumption of CPU is

$$P(s)=\alpha s^{v}+(1-\alpha) \qquad (2)$$

Where s can be considered as CPU utilization, when s=0, the server is in idle state and when s=1, the servers are running at the maximum speed. v is empirically estimated and a typical value is v=2. The above model is widely used in the literature [12]. The power consumption with respect to s is shown in Figure 8 with v=2.

A commonly used mechanism that is proposed to reduce power consumption is dynamic voltage and frequency scaling (DVFS) [13]. DVFS can reduce processor voltage by dynamically scaling the server speed when the jobs or service request are light. How to model the power consumption with service rate s is an open topic [8], and dynamic power consumption mainly depends on CPU utilization and clock rate, which can be approximately modelled as

$$P_d = aCV^2 f \quad (3)$$

Where a is the switching activity, C is the physical capacitance, V is the supply voltage, and f is the clock rate. When the servers are idle, the power consumption in this case is static power consumption. Therefore, the power consumption of a server is characterized as

$$P = P_d + P_s \quad (4)$$

A typical power consumption function with regard to the service rate a server is denoted by

$$P = \rho k \mu^{\alpha} + P_s \quad (5)$$

Where k is some constant, and the value of α is about 3. Figure 9 shows how power varies with regard to service rate, with $P_s = 2$, k=0.2.

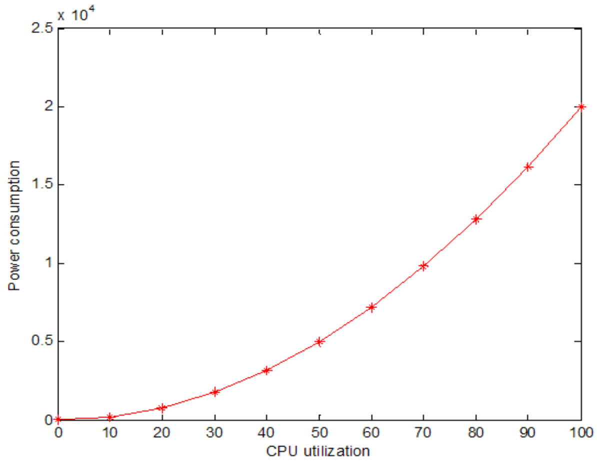

Figure 8. Power consumption varies with CPU utilization

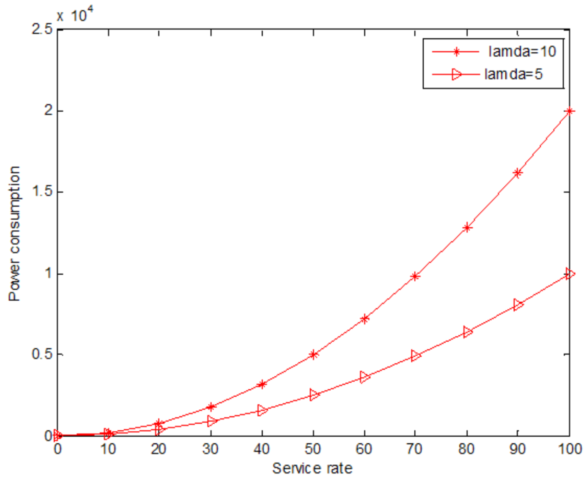

Figure 9. Power consumption varies with service rate

The main objective of the work [8] is managing performance which is the metric of response time and the power consumption, and the goal of the work [12] is to balance the throughput which can be equivalently recognized as the measure of revenue as SaaS cloud service providers charge according the number of service requests and the power consumption in the SaaS clouds.

4. Reducing power consumption cooling systems

It has been claimed by T. Mukherjee et.al [14] that a large part of power was consumed by cooling system in data centers due to the growing power density in servers, which is over 30% of total power consumption. A model used to describe the power consumption in cool systems of a data center is

$$P^c = \frac{P}{CoP(L^{\sup})} \quad (6)$$

Where P is the power consumption of servers in this data center, $CoP(L^{\sup})$ is the coefficiency of the performance of the cooling system, and $L^{\sup}$ is the temperature of supplied cold air. $CoP(L^{\sup})$ can be given by

$$CoP(L^{\sup}) = 0.0068(L^{\sup})^2 + 0.0008L^{\sup} + 0.458 \quad (7)$$

The relationship between $CoP(L^{\sup})$ and $CoP(L^{\sup})$ is shown in Figure 10.

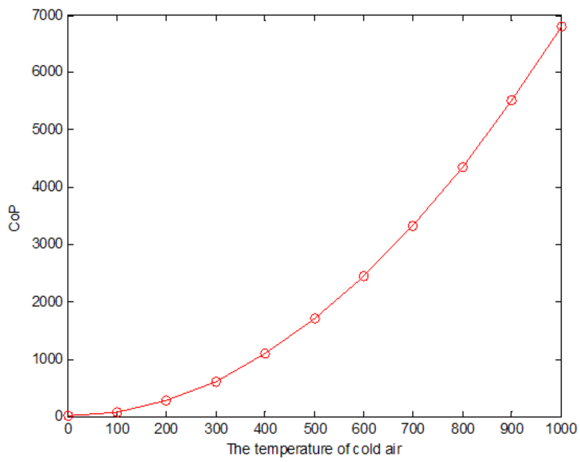

Figure 10. The CoP of cooling systems vs. temperature of cold air

The above two equations are given under the assumption that the impact of the outside temperature is neglected and is supposed as a fixed value.

M. Polverini et al. [15] studied thermal-aware scheduling of batch jobs in geographically distributed data centers by exploiting the variations of prices in different times and locations. Compared with the baseline method, their proposed online method based on the framework of Lyapov optimization not only can reduce the power consumptions of the cooling systems in cloud data centres, but also can meet with the maximum server inlet temperature constraint.

H. Xu et al. [16] studied thermal-aware scheduling of batch jobs and interactive jobs in geographically distributed data centers. They also develop a distributed algorithm. Different from M. Polverini et al. [15], their method was based on an m-block alternating direction method of multipliers (ADMM) algorithm, and their simulation results showed that temperature aware workload management saves 15%–20% cooling energy. Morkun et al. analyzed how temperature affected energy costs in distributed systems by using heat pumps technology [17].

5. Conclusions

With the development of cloud computing, the increasing number of data centers lead to the soaring increase of power consumption and carbon emission, how to reduce them becomes a main concern for cloud providers and the need for green data centers. This paper makes a survey of power and energy reduction methods in cloud data centers. We analyzed some methods that are commonly used to reduce power and energy consumption.

Acknowledgements

This work is partly supported by the Major project of nature science of Department of Education of Anhui Province (KJ2014ZD31).

References

1. H.H. Kramer, V. Petrucci, A. Subramanian, E. Uchoa (2012) A column generation approach for power-aware optimization of virtualized heterogeneous server clusters. *Comput. Indus. Eng*, 63(3), p.p.652-662.
2. J.A. Gallego Arrubla, Y.M. Ko, R.J. Polansky, E. Pérez, L. Ntaimo, N. Gautam (2013) Integrating virtualization, speed scaling, and powering on/off servers in data centers for energy efficiency. *IIE Trans*, 45 (10), p.p.1114-1136.
3. A. Vasan, A. Sivasubramaniam, V. Shimpi, T. Sivabalan, R. Subbiah (2010) Worth their watts? —an empirical study of data center servers. *Proc. of 2010 IEEE 16th International Symposium on High-Performance Computer Architecture*, p.p. 1-10.
4. L.Parolini, B.Sinopoli, B. H.Krogh, Z.Wang (2012) A cyber–physical systems approach to data center modeling and control for energy efficiency. *Proceedings of the IEEE,* p.p.254-268.
5. A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs (2009) Cutting the electric bill for internet-scale systems. *Proc. ACM SIGCOMM*, p.p. 123–134.
6. D.Kliazovich, P.Bouvry, Y.Audzevich, S. Khan (2010) Greencloud: A packet-level simulator of energy-aware cloud computing data centers. *Proc. of 2010 IEEE Conference on Global Telecommunications*, p.p. 1–5.
7. M. Al-Fares, A.Loukissas, A. Vahdat (2008) A scalable, commodity data center network architecture. *ACM SIGCOMM Computer Communication Review*, 38(4), p.p.63-74.
8. W. Fang, X. Liang, S. Li, L.Chiaraviglio, and N.Xiong (2013) VMPlanner: Optimizing virtual machine placement and traffic flow routing to reduce network power costs in cloud data centers. *Computer Networks*, 57(1), p.p.179-196.
9. A. Greenberg, J.R. Hamilton, N. Jain, S. Kandula, C. Kim, P. Lahiri, D.A.Maltz, P. Patel, S. Sengupta (2009) Vl2: a scalable and flexible data center network. *Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication*, p.p.51–62.
10. C.Guo, H. Wu, K. Tan, L. Shi, Y. Zhang, and S. Lu (2008) Dcell: a scalable and fault-tolerant network structure for data centers. *ACM SIGCOMM Computer Communication Review,* p.p.75-86.
11. R.Urgaonkar, U. C. Kozat, K. Igarashi, and M. J. Neely (2010) Dynamic resource allocation and power management in virtualized data centers. *Proc. of 2010 IEEE Network Operations and Management Symposium (NOMS)*, p.p.479-486.
12. F.Liu, Z.Zhou, H.Jin, B. Li, B. Li, H.Jiang (2014) On arbitrating the power performance trade off in SaaS clouds. Parallel and Distributed Systems. *IEEE Transactions on Cloud Computing*, 25(10), p.p.2648-2658.
13. E.Le Sueur, G.Heiser (2010) Dynamic voltage and frequency scaling: the laws of diminishing returns. *Proceedings of the 2010 International Conference on Power Aware Computing and Systems*, p.p.1–8.
14. T. Mukherjee, Q. Tang, C. Ziesman, S. K S Gupta, P. Cayton (2007) Software Architecture for Dynamic Thermal Management in Data centers. *Proc. of 2nd International Conference on Communication Systems Software and Middleware,* p.p.7-12.
15. M. Polverini, A. Cianfrani, S. Ren, A. V. Vasilakos. (2014) Thermal-aware scheduling of batch jobs in geographically distributed data centers. *IEEE Transactions on Cloud Computing,* 2(1), p.p.71-84.
16. H. Xu, C. Feng, and B. Li (2015) Temperature aware workload management in geo-distributed data centers. *IEEE Transactions on Parallel and Distributed Systems*, 26(6), p.p. 1743 - 1753.

17. V. Morkun, O. Savytskyi, S. Ruban (2015) The use of heat pumps technology in automated distributed system for utilization of low-temperature energy of mine water and ventilation air, *Metallurgical and Mining Industry*, 7(6), p.p.118-121.

Multidimensional Data Classification Based on BP Neural Network Algorithm

Rong Deng

Department of Information Engineering, Chongqing Vocational Institute of Engineering, Chongqing, 402260, China

Abstract
This paper selects the BP neural network algorithm for data classification, and carries out a detailed derivation of its work principle. For the existing slow network convergence and prone to falling into local minimum and other defects of this algorithm, the paper adopts the method of the combination of using the variable learning rate and added with the momentum factor to improve the traditional BP algorithm. In the specific network training experiment, it is found that the improved algorithm has improved the speed of network convergence to certain extent. And the final data classification results show that, BP neural network has a relatively high success rate when applied for the multi-dimensional data classification in the Internet of Things.
Key words: INTERNET OF THINGS, DATA CLASSIFICATION, BP NEURAL NETWORK, CLOUD COMPUTING, MULTI-DIMENSIONAL DATA

1. Introduction

In the process of data mining, data classification is a very important method of data analysis [1]. In the world that we live in, all human activities are derived from the accumulated wisdom and knowledge of mankind himself. Human makes the right judgments and decisions based on the combination of the knowledge which has been gained through experience and the perception to the external things. In fact, if we only study the data itself, it will not make any sense, because it's just the raw materials that human has obtained through a variety of available tools and instruments. While the internal logic of things hidden behind these data is what we want to explore, and these existing experiences will be the foundation to provide direct assistance for us to make decisions [2]. Therefore, we need to make analysis on the raw data so as to find the internal relations between them; hence the things that just make sense mathematically will be turned into useful information. However, they cannot be directly identified by human, and need to summarize and classify a lot of multi-dimensional data before it can be converted to the knowledge that human can make use of. Thus, when faced with huge amount of original data, classification and extraction of the data can provide us the knowledge and wisdom