2015 International Conference on Computational Science and Computational Intelligence

A Greedy Approach to Cost-Aware Virtual Machine Allocation for 100% Green Data Centers

Hai Wang
Sobey School of Business
Saint Mary's University
Halifax, Canada
hwang@smu.ca

Haikun Wei
School of Automation
Southeast University
Nanjing, China
hkwei@seu.edu.cn

***Abstract* — This paper shows the cost-aware virtual machine allocation problem for 100% green data centers is an integer programming problem which is NP-hard. The paper presents a greedy algorithm for cost-aware virtual machine allocation which attempts to minimize the total cost of biofuel while maintaining the quality-of-service requirements of the tasks.**

Keywords—green data center; virtual machine; cost-aware.

I. Introduction

As electricity use is the largest portion of the operational cost of data centers [13], energy consumption and energy efficiency are two key issues for designing and building data centers. With the increasing availability of green energy sources such as solar, wind and biofuel, there are increasing number of data centers powered by green energy sources [10].

A *100% green data center* employs a variety of green energy sources, such as solar panels, wind turbines and biofuel generators, to provide 100% power supply to the data center. 100% green data centers are environmental friendly and sustainable. A number of projects have already been initiated to build 100% green data centers [10].

Figure 1 illustrates the typical power supply system of 100% green data centers. In Figure 1, the data center employs a *grid tie system* (GTS) for integrating the power from different sources, and an *uninterruptible power supply system* (UPS) with rechargeable batteries or supercapacitors for ensuring continuous, uninterrupted and stable power supply to the data center. The main purpose of the UPS system is to provide short-term backup power supply [8]. As the power generated from the solar panels and wind turbines depends on many natural factors, it is the biofuel generators that provide the main backup power supply for long periods of time.

Different tasks being processed at a data center often have different *quality-of-service* (QoS) requirements. *Service-level agreements* (SLA) have been used by data centers to specify the QoS requirements of tasks [5-7, 13]. An SLA is a contract between the data center and the customer for the expected QoS requirement of a particular task.

For modern data centers, the virtualization technology enables multiple virtual machines running on the same physical server. Each individual virtual machine allocates necessary computational resources to process a task while ensuring the QoS requirements specified in the SLA. Using the virtualization technology, data centers can minimize the number of physical servers and thus increase the energy efficiency [1, 9, 11].

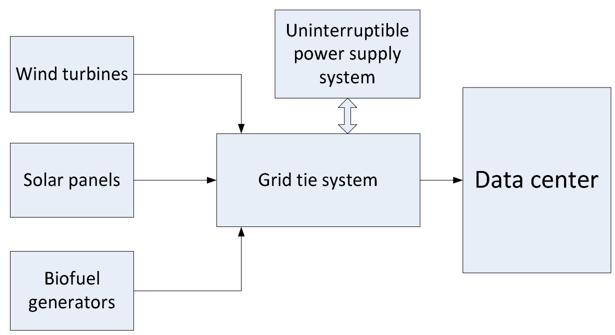

Figure 1. An example of the typical power supply system of 100% green data centers

The virtual machine allocation problem is one of the key challenges for data centers [5]. In this paper, we propose a greedy algorithm for cost-aware virtual machine allocation for 100% green data centers. It assumes that the investment cost of the data center is fixed, and the equipment depreciation rate is fixed for all equipments. The proposed greedy algorithm attempts to minimize the total cost of biofuel while maintaining the QoS requirements of the tasks. For 100% green data centers, the total cost of biofuel is a largest portion of the operational cost. The algorithm takes into account the biofuel price which may fluctuate over the time.

The remainder of this paper is organized as follows. Section 2 formulates the virtual machine allocation problem for 100% green data centers. Section 3 presents the proposed greedy algorithm. Section 4 describes the related work. Section 5 presents the conclusions.

II. Problem Formulation

We define the following power demand and supply for a 100% green data center:

978-1-4673-9795-7/15 $31.00 © 2015 IEEE
DOI 10.1109/CSCI.2015.89

- $D(t)$ = The total power demand of the data center at time t.
- $S(t)$ = The total power supply from the solar panels at time t.
- $W(t)$ = The total power supply from the wind turbines at time t.
- $B(t)$ = The total power supply from the biofuel generators at time t.

The power demand, $D(t)$, corresponds to the overall power consumption of the data center. The power supplies from the solar panels and wind turbines, $S(t)$ and $W(t)$, depend on many natural factors and are hard to predict. When the investment cost of the 100% green data center is fixed and the equipment depreciation rate is fixed for all equipments, power generated from the solar panels and wind turbines can be treated as a free commodity.

When the power demand exceeds the total power generated from the solar panels and wind turbines, biofuel generators will then be used to deliver additional power to the data center. Let $P(t)$ denote the price of biofuel at time t. If we neglect the effects of the UPS system which provides only a short-term backup to ensure uninterrupted power supply, a simplified relationship between the power demand and power supplies can be expressed as

$$B(t) = \begin{cases} D(t) - S(t) - W(t) & \text{if } D(t) > S(t) + W(t), \\ 0 & \text{Otherwise.} \end{cases}$$

The total cost of biofuel, T, during a time period U can be expressed as

$$T = \int_{\forall t \in U} B(t) \cdot P(t) \cdot dt \qquad (1)$$

Suppose that there are N identical physical servers $\mathbf{S} = \{S_1, S_2, \ldots, S_N\}$ at the data center, and the maximum computational processing capacity of each server is identical and is C. Let $\mathbf{V} = \{V_1, V_2, \ldots, V_M\}$ denote a set of M possible virtual machines to be used. When a task request arrives at the data center, the task will be immediately assigned to a virtual machine to be processed. Once the task is completed, the corresponding virtual machine will be available for processing another task. We classify all tasks into X classes, where all tasks in the same class have the same SLA. Suppose that a virtual machine requires computational processing capacity Q_x for each task in class x, $1 \leq x \leq X$, in order to ensure the QoS requirements specified in the SLA.

For a virtual machine $v \in V$ and a physical server $s \in S$, we define the follow variable corresponding to the allocation of virtual machines:

$$H_{s,v,x} = \begin{cases} Q_x & \text{if } v \text{ is running a class } x \text{ task on } s, \\ 0 & \text{otherwise.} \end{cases}$$

and

$$\sum_{\forall v \in V} \sum_{x=1}^{X} H_{s,v,x} \leq C \qquad \text{for } \forall s \in S.$$

The utilization of a physical server s can then be calculated as

$$U_s = \frac{\sum_{\forall v \in V} \sum_{x=1}^{X} H_{s,v,x}}{C} \qquad (2)$$

Let D_{idle} denote the power consumption of a physical server when it is idle (i.e., the utilization is 0%), and let D_{peak} denote the power consumption of a physical server when it is 100% busy (i.e., the utilization is 100%). Previous work [2, 3, 7] reported that there is a nearly linear relationship between the utilization and average power consumption of servers of data centers such that

$$D(t) = \sum_{\forall s \in S} \left[D_{idle} + \left(D_{peak} - D_{idle}\right) \cdot U_s\right] \qquad (3)$$

Eq. (3) indicates that a single server with 100% utilization consumes less power than two servers each with 50% utilization even if the former completes the same amount of tasks as the latter.

For 100% green data centers, the virtual machine allocation problem is to assign virtual machines to physical servers such that the total cost of biofuel, T, is minimized. By Eq. (1), (2) and (3), it can be shown minimizing T is equivalent to

$$\text{minimizing} \int_{\forall t \in U} P(t) \cdot \left\{ \sum_{\forall s \in S} \left[D_{idle} + \frac{D_{peak} - D_{idle}}{C} \sum_{\forall v \in V} \sum_{x=1}^{X} H_{s,v,x} \right] \right\} dt \qquad (4)$$

$$\text{subject to} \sum_{\forall v \in V} \sum_{x=1}^{X} H_{s,v,x} \leq C \qquad \text{for } \forall s \in S.$$

Hence, the virtual machine allocation problem which minimizes the total cost of biofuel T is to find the optimal values of $H_{s,v,x}$ in Eq. (4).

III. The Cost-Aware Virtual Machine Allocation

Even if the biofuel price $P(t)$ remains as a constant all the time and all tasks are known in advance, the optimization problem associated with Eq. (4) is an *integer programming* problem which is known to be *NP-hard*. In real-life, the price of biofuel may fluctuate over the time, and tasks will arrive at the data center in an unpredictable fashion.

As there is no fast algorithm to solve this optimization problem, we propose a greedy algorithm for cost-aware virtual machine allocation.

Once a task arrives at the data center, the greedy algorithm will immediately allocate the corresponding virtual machine to a particular physical server that minimizes Eq. (4). Intuitively, when the biofuel price remains as a constant, the algorithm will allocate the virtual machine to the busiest physical server that can accommodate the task. As implied by Eq. (3), increasing the utilization of the busiest physical server generally results in only a small increase of the overall power demand. The algorithm also takes into account the biofuel price in order to minimize the total cost of biofuel.

The algorithm is summarized as follows.

Algorithm Input:

(1) The constants: C, D_{peak} and D_{idle}.
(2) E = The newly arrived task.
(3) Q = The corresponding Q_x of the new task E.
(4) G = A list of tasks that are currently being processed at the data center.
(5) All existing values of $H_{s,v,x}$ for the tasks in G.
(6) All existing values of $[P(t) \cdot H_{s,v,x}]$, where $P(t)$ is the biofuel price at the arrival time of the task corresponding to the existing $H_{s,v,x}$.
(7) P = The biofuel price $P(t)$ at the time of the new task E's arrival, which may be different from the values of previous $P(t)$.

Algorithm Description:

1. Examine all virtual machines to see if any tasks in G has been completed.
2. Deleted the completed tasks from G, and update the corresponding values of $H_{s,v,x}$ and $[P(t) \cdot H_{s,v,x}]$. Make the virtual machines of the completed tasks available again.
3. Assign an available virtual machine to E, and add E to G.
4. For each physical server s, compute
$$\delta_s = P \cdot D_{idle} + \frac{D_{peak} - D_{idle}}{C} \cdot \left\{ P \cdot Q + \sum_{\forall\, v \in V} \sum_{x=1}^{X} [P(t) \cdot H_{s,v,x}] \right\}$$
5. Find the physical server β such that $\delta_\beta = \min_{\forall s} \delta_s$ subject to the constraint $Q + \sum_{\forall\, v \in V} \sum_{x=1}^{X} H_{s,v,x} \le C$
6. If the physical server β is currently in idle, wake this server
7. Allocate the virtual machine of the new task E to the physical server β
8. Update the values of $H_{s,v,x}$ and $[P(t) \cdot H_{s,v,x}]$.

IV. RELATED WORK

The virtual machine allocation problem has been extensively studied for data centers. One group of related work focused on the virtual machine allocation for a network of data centers distributed at different geographical locations [1, 3, 9, 10, 15]. The main goal of these studies is to minimize the network traffic and/or system latency.

Another group of related work focused on the virtual machine allocation of data centers to minimize the operational cost or to maximize the profit [4, 6, 7, 11, 14]. Some research explored the feature of a warning period before a sudden power change, which is called as energy buffering [11]. Others focused on the optimization methods to maximize the profit [6, 7], or to minimize the overall power cost [4], or to minimize the overall power consumption [14].

All previous related work dealt with either traditional data centers or green data centers with electrical grid as the backup power supply. This paper is the first of the kind that deals with 100% green data centers.

V. CONCLUSIONS

This paper has shown that the cost-aware virtual machine allocation problem for 100% green data centers is an integer programming problem which is NP-hard. The proposed greedy algorithm minimizes the total cost of biofuel while maintaining the quality-of-service requirements of the tasks.

REFERENCES

[1] K. Bilal, S. U. R. Malik, O. Khalid, A. Hameed, E. Alvarez, V. Wijaysekara, R. Irfan, S. Shrestha, D. Dwivedy, M. Ali, U. S. Khan, A. Abbas, N. Jalil, S. U. Khan, "A taxonomy and survey on Green Data Center Networks", *Future Generation Computer Systems*, 36, pp. 189-208, 2014.

[2] M. Blackburn, "Five ways to save server power", White Paper #7, The Green Grid, http://www.thegreengrid.org/ (Accessed August 30, 2015).

[3] A. P. M. De La Fuente Vigliotti, and D. M. Batista, "A green network-aware VMs placement mechanism", *Proceedings of the 2014 Global Communications Conference*, pp.2530-2535, 2014.

[4] W. Deng, F. Liu, H. Jin, C. Wu, and X. Liu. "MultiGreen: cost-minimizing multi-source datacenter power supply with online control", *Proceedings of the 4th international conference on Future energy systems*, pp. 149-160, 2013.

[5] F. Kong, and X. Liu. "A survey on green-energy-aware power management for datacenters", *ACM Computing Survey*, 47(2), Article 30, 38 pages, 2014. DOI=10.1145/2642708

[6] M. Ghamkhari, and H. Mohsenian-Rad, "Profit maximization and power management of green data centers supporting multiple SLAs", *Proceedings of the 2013 International Conference on Computing, Networking and Communications*, pp. 465-469, 2013.

[7] I. Goiri, J. L. Berral, J. O. Fito, F. Julia, R. Nou, J. Guitart, R. Gavalda, J. Torres, "Energy-efficient and multifaceted resource management for profit-driven virtualized data centers", *Future Generation Computer Systems*, 28(5), pp. 718-731, 2012.

[8] I. Goiri, M. E. Haque, K. Le, R. Beauchea, T. D. Nguyen, J. Guitart, J. Torres, R. Bianchini, "Matching renewable energy supply and demand in green datacenters", *Ad Hoc Networks*, 25(B), pp. 520-534, 2015.

[9] A. Hammadi, and L. Mhamdi, "A survey on architectures and energy efficiency in Data Center Networks", *Computer Communications*, 40, pp. 1-21, 2014.

[10] W. Hu, A. Hicks, L. Zhang, E. M. Dow, V. Soni, H. Jiang, R. Bull, and J. N. Matthews, “A quantitative study of virtual machine live migration”, *Proceedings of the 2013 ACM Cloud and Autonomic Computing Conference*, Article 11, 10 pages, 2013. DOI=10.1145/2494621.2494622

[11] R. Singh, D. Irwin, P. Shenoy, and K. K. Ramakrishnan, “Yank: Enabling green data centers to pull the plug”, *Proceedings of the 10^{th} USENIX Symposium on Networked Systems Design and Implementation*, pp. 143-155, 2013.

[12] L. Tomas, and J. Tordsson, “An autonomic approach to risk-aware data center overbooking”, *IEEE Transactions on Cloud Computing*, 2(3), pp. 292-305, 2014.

[13] L. Wang, and S. U. Khan, “Review of performance metrics for green data centers: a taxonomy study”, *Journal of Supercomputing*, 63(3), pp. 639-656, 2013.

[14] Y. Wang, and X. Wang, “Power optimization with performance assurance for multi-tier applications in virtualized data centers”, *Proceedings of the 39^{th} International Conference on Parallel Processing Workshops*, pp. 512-519, 2010.

[15] Z. Xiao, W. Song, and Q. Chen, “Dynamic resource allocation using virtual machines for cloud computing environment”, *IEEE Transactions on Parallel and Distributed Systems*, 24(6), pp.1107-1117, 2013.