3L-2

Decentralized Management for Power Optimization in Cloud Data Centers Using VM Migration

Islam A.K.A.Md. Shariful, Koji Hasebe, Kazuhiko Kato
Dept. of Computer Science, School of Systems and Information Engineering
University of Tsukuba

1 Introduction

In recent years, cloud computing has become very popular. Cloud data centers are also being used for many purposes like storage, data analysis etc. For maintaining energy supply to huge resources of data centers, power optimization in data centers is getting strong attention lately. We propose a power optimization approach in our research using VM migration. In the previous works, VM migration based power optimization and network traffic performance in data centers has been studied separately when both has significance for efficiency. In our research, we consider both and propose a model for power optimization with simulation result and evaluate the performance. As our future work, we consider decentralized management for this approach.

2 Related Work

Beloglazov et al. [1] used VM migration for power optimization. They consolidates VMs into minimum number of hosts and switches the idle hosts into sleep mode. They have trade off with network performance related to VM dependency. Shrivastava et al. [2] has taken VM dependency into their account to ensure efficient network resource utilization and proposed an efficient mechanism for application-aware VM migration. Though power optimization was not studied by them, their approach can help network traffic-aware power optimization technique. In our research, we consider the both approaches of these two works and propose a network traffic-aware VM migration based power optimization technique.

3 System Design

3.1 Architectural Framework

In this study, we consider a large-scale cloud data center comprising with physical computing infrastructure of n homogeneous physical hosts, m heterogeneous VMs and a tree network architecture constructed with q switches having heterogeneous capacity of traffics. Let $P = \{p_1, p_2, ..., p_m\}$ be the set of total number of physical machines and $V = \{v_1, v_2, ..., v_n\}$ be the set of total number of VMs in the datacenter. Let $S = \{s_1, s_2, ..., s_q\}$ be the set of switches that is used for network formation.

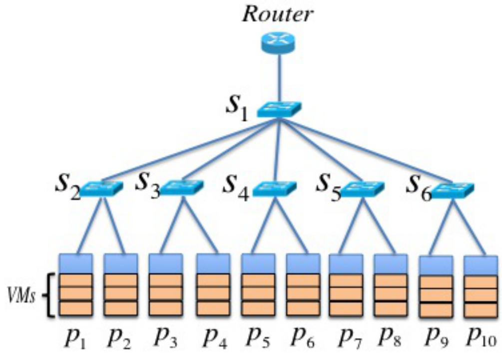

Figure 1: Network Graph, NG for $m = 10, n = 30, q = 6$

Let $NG = (E, N)$ be the network graph for our framework where E is the set of edges represent VM dependencies and $N = P \cup S$ is the set of physical machines and switches (Figure: 1). Let $\mathcal{G} = (V, E, w)$ be the weighted cost graph where $w : E \to R$ is a function that assigns each edge a weight of traffic cost between two VMs that is a real number.

3.2 Power Consumption Model

Several recent studies have shown that CPU of a physical host consumes the main part of energy used by a physical machine and approximately 70% of the power is used in its idle state. So, switching idle host into sleep mode can be an honest way to optimize power. Therefore, in this work we use the power model defined below.

$$A_p(u) = A_{min} + \alpha \cdot u$$

Here, $u = \frac{Sum\ of\ accesses\ per\ second\ of\ all\ VMs\ in\ p}{Access\ capacity\ persecond\ of\ p}$ the CPU utilization of p; $A_p(u)$ is the consumed power of physical machine p in active state; A_{min} is the minimum power consumed by a physical machine in idle state; α is the increase rate of power for CPU utilization. Now, let $A_t = \sum_{i=1}^{m} A_p(u_{p_i})$ be total power consumed by the datacenter. Here, u_{p_i} is the CPU utilization of the physical machine p_i. Now the optimization framework can be defined as,

$$minimize\ A_t = \sum_{i=1}^{m} A_p(u_{p_i})$$
$$subjects\ to$$
$$T_l < u_{p_i} < T_u,\ \forall s \in S\ tr(s) \leq Cap(s)$$

Here, $Cap(s)$ is the traffic capacity for switch s; $tr(s)$ is the total traffics of switch s caused by underlying VMs; T_u and T_l are the upper and lower utilization thresholds of a physical machine respectively.

4 Power-aware VM Allocation

For efficient allocation of VMs, we need to follow two steps: one is VM selection and another one is VM placement. These two are described below.

VM Selection: VMs should be selected from a physical machine only when its CPU utilization goes under lower threshold, T_l or higher the upper utilization threshold, T_u. We follow the minimization of migrations algorithm showed in previous work [1]. This policy ensures minimum number of migrations.

VM Placement: For VM placement, we follow the bin packing problem algorithm partially which is used in previous work [1] also. . The capable physical machine is selected first and VMs are placed to that maximally. Along with that, it confirms that no switch goes having the traffics higher its capacity.

5 Simulation Results

For simulation, we considers 100 homogeneous physical hosts having 300 VMs of different capacity when 10% of VMs are dependent on other VMs. At first, every host posses 3 VMs each. The maximum capacity for a host is set to 3000 MIPS when its utilization is measured by requested MIPS by holding VMs.

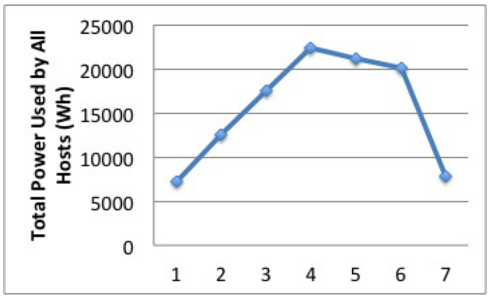

Figure 2: Change of the Total Power of Active Physical Hosts

Upper and lower utilization threshold are set to 0.7 and 0.3 respectively. Total number of switches are 6 and their traffic capacity varies from 25MB/min to 40MB/min. Total number of taken reading is 7 times in an interval of one hour. Our simulation shows that power can be optimized with the change of accesses to lower state. Figure 2 shows the change of total power of all hosts with respect to total access numbers of different time of day. It represents, in low access period, load can be consolidated to minimum hosts and in high access period, load can be run by maximum hosts.

6 Conclusion and Future Work

In our research, we consider a framework for power optimization in data center taking the matter of VM dependency which is important for network performance. We have made a study on how the power used in data center changes with respect to change of access to VMs. We showed that it can be possible to optimize power by consolidating VMs to minimum number of physical hosts when they are in less use. This time we could not study on network performance. In our future work, we do study on the trade off between network performance and the QoS. This time we manage the model with centralized management. We consider decentralized management as our future work.

References

[1] A. Beloglazov, J. Abawajy, and R. Buyya. Energy-Aware Resource Allocation Heuristics for Efficient Management of Data Centers for Cloud Computing. In *Future Generation Computer Systems*, 2011.

[2] V. Shrivastava, P. Zerfos, K. won Lee, H. Jamjoom, Y.-H. Liu, and S. Banerjee. Application-aware Virtual Machine Migration in Data Centers. In *In Proc. of IEEE INFOCOM Mini-conference*, 2011.