Journal of Engineering Sciences
Vol 10, Issue 12, Dec /2019
ISSN NO:0377-9254

CLOUD DATA CENTER BASED ENERGY EFFICIENT SCHEDULING OF SERVERS WITH MULTI-SLEEP MODES

[1]PALIVELA SIVA JYOTHI, [2]B.HARI BABU
M.Tech Students, Professor (M.Tech, Ph.D)
DEPT OF CSE
KITS Engineering College, Ramchandrapuram

ABSTRACT:

In a cloud data center, servers are always over-provisioned in an active state to meet the peak demand of requests, wasting a large amount of energy as a result. One of the options to reduce the power consumption of data centers is to reduce the number of idle servers, or to switch idle servers into low-power sleep states. However, the servers cannot process the requests immediately when transiting to an active state. There are delays and extra power consumption during the transition. In this paper, we consider using state of- the-art servers with multi-sleep modes. The sleep modes with smaller transition delays usually consume more power when sleeping. Given the arrival of incoming requests, our goal is to minimize the energy consumption of a cloud data center by the scheduling of servers with multi-sleep modes. We formulate this problem as an integer linear programming (ILP) problem during the whole period of time with millions of decision variables. To solve this problem, we divide it into sub-problems with smaller periods while ensuring the feasibility and transition continuity for each sub-problem through a Backtrack-and-Update technique. We also consider using DVFS to adjust the frequency of active servers, so that the requests can be processed with the least power. Our simulations are based on traces from real world. Experiments show that our method can significantly reduce the power consumption for a cloud data center.

I. INTRODUCTION

IN recent years, cloud data centers are expanding rapidly to meet the ever increasing demand of computing capacity. It is the powerful servers of the data centers that consume a huge amount of energy. According to a report, data centers consume about 1.3% of the worldwide electricity, which is expected to reach 8% in 2020 [1]. Meanwhile, much of the energy is wasted, because servers are busy only 10%30% of the time on average, with most time in idle state. What's worse, a server can even consume 60% or more of its peak power when in idleness [2]. To handle the possible peak demand of user requests, servers are always overprovisioned, wasting a lot of energy as a result. Therefore, there is an urgent need to enhance energy efficiency for cloud data centers. The existing work has mainly focused on dynamic voltage frequency scaling (DVFS) and dynamic power management (DPM). The former is to adjust the voltage/frequency of CPU power according to the demand of computing capacity, while the latter reduces the total energy by putting servers into sleep states or turning off idle servers. However, a difficult issue is that the servers cannot process the incoming requests immediately when transiting to active state. There are delays and extra power consumption during the transitions, which have been ignored in the existing work. Besides, modern servers are usually designed with several sleep states, and the sleep states with smaller transition delays consume more power when sleeping.

In this paper, we study the issue of minimizing energy consumption of a data center by scheduling servers in multisleep modes and at different frequency levels to reduce the total energy of active servers. That is, given the arrival of user requests, schedule the servers (to active state with different frequencies or to different sleep states), such that the total energy consumption of the data center can be minimized while satisfying the QoS requirement. The scheduling algorithm will determine:

1) how many of the active servers should be switched into which sleep state in each timeslot;
2) how many of the sleeping servers in sleep states should be woken up in each timeslot;
3) What frequency levels should the active servers be set to in each timeslot.

The scheduling period of our problem consists of T small timeslots.We solve the problem in two steps. In the first step, we aim to minimize the total number of active servers to meet the QoS requirement by assuming that all servers run at the highest frequency. The problem is formulated as a constraint optimization problem with millions of decision variables due to the large number of timeslots. It is not feasible to solve the problem of such a large size using existing methods. We group multiple timeslots into a segment with equal length, and formulate the scheduling in each segment independently as an integer linear programming (ILP) subproblem. By using Cplex to solve each sub-problem, the optimal solution can be obtained for each segment. However, the scheduling of the current segment doesn't consider the arrival of the requests in the next segment. It may lead to the situation that some servers are put into sleep at the end of this segment, but cannot be woken up immediately to cope with request burst at the beginning of the next segment.

We propose a Backtrack-and-Update technique to solve this issue. In the second step, we make scaling of the frequency levels of the active servers, so that the requests can be processed with the least necessary power. In each timeslot, this problem can also be formulated into an independent ILP problem of a small size that the optimal solution can be obtained. Our simulations are based on traces from real world. Experiments show that our method can significantly reduce the total energy consumption for a cloud data center.

II.PRELIMINARY INVESTIGATION

The first and foremost strategy for development of a project starts from the thought of designing a mail enabled platform for a small firm in which it is easy and convenient of sending and receiving messages, there is a search engine ,address book and also including some entertaining games. When it is approved by the organization and our project guide the first activity, ie. preliminary investigation begins. The activity has three parts:

- **Request Clarification**
- **Feasibility Study**
- **Request Approval**

III.EXISTING SYSTEM

In the existing system, DVFS mechanism scales the CPU chipset power through adjusting the voltage and frequency of CPU. That is, the processing capacity varies with different power levels. Gandhi et al. in [17] combine DFS (Dynamic Frequency Scaling and DVFS to optimize the power allocation in server farm to minimize the response time within a fixed peak power budget. Gerards et al. in [18] try to minimize energy cost through global DVFS on multi-core processors platform while considering the precedence constraint in task scheduling.

Elnozahy et al. in [19] employ a DVS (Dynamic Voltage Scaling) and node On/Off method to reduce the aggregate power consumption of cluster during periods of reduced workload. They also use both DVS and requests batching mechanisms to reduce processor energy over a wide range of workload intensities [20]. Rossi et al. in [21] build power models to estimate the energy consumption of user applications under different DVFS policies.

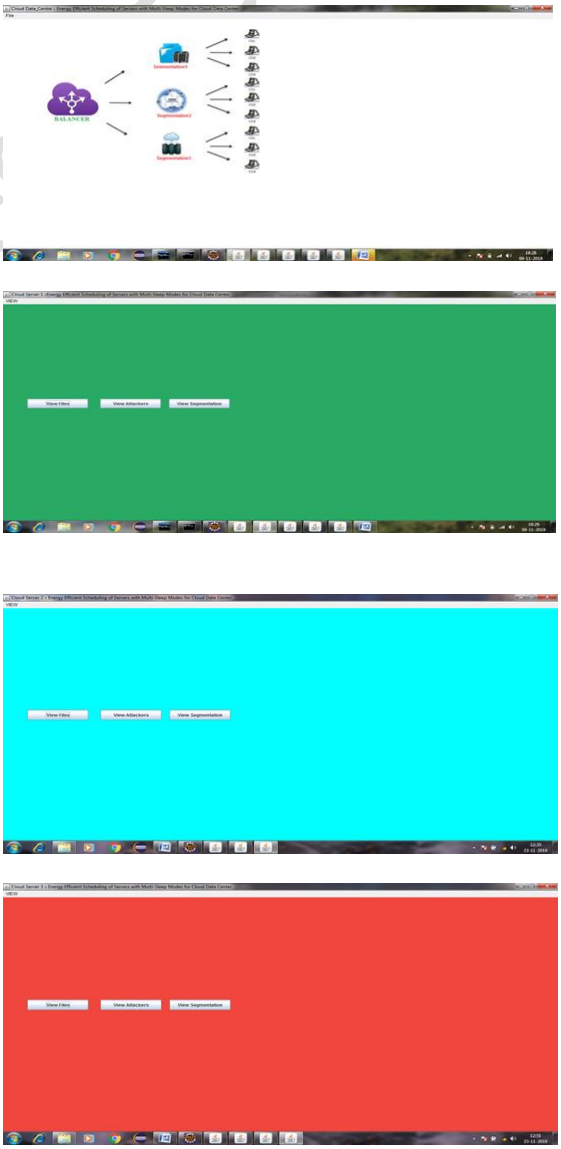
Florence et al. in [22] first study the flow pattern of tasks of cloud, and then try to tune the incoming VM tasks with required frequency using DVFS. Lin et al. in [23] use DVFS to reduce the power consumption in task scheduling in mobile cloud computing environment, but with no consideration of On/Off servers. Chen et al. in [24] combine the three approaches of request dispatching, service management and DVFS to improve energy efficiency for large scale computing platform.

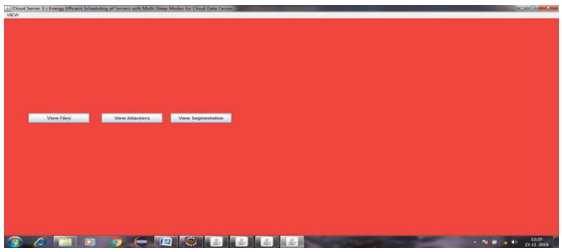
IV.PROPOSED SYSTEM

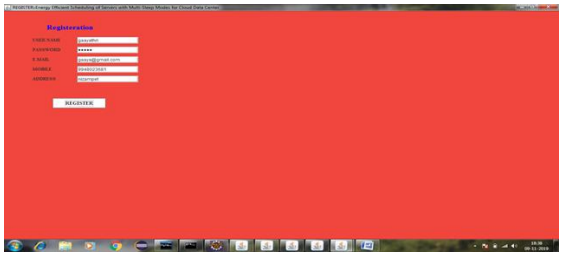
In the proposed system, the system studies the issue of minimizing energy consumption of a data center by scheduling servers in multi sleep modes and at different frequency levels to reduce the total energy of active servers. That is, given the arrival of user requests, schedule the servers (to active state with different frequencies or to different sleep states), such that the total energy consumption of the data center can be minimized while satisfying the QoS requirement.

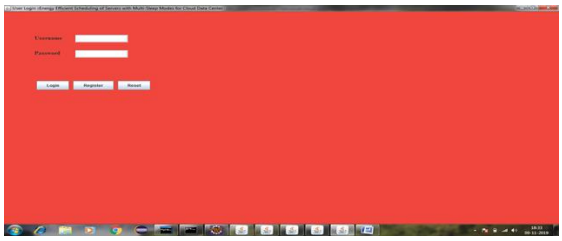
The scheduling algorithm will determine: 1) how many of the active servers should be switched into which sleep state in each timeslot;
2) How many of the sleeping servers in sleep states should be woken up in each timeslot; 3) what frequency levels should the active servers be set to in each timeslot.

V.SAMPLE OUTPUT SCREENS









VI.CONCLUSION

In this paper, we studied the problem of scheduling of servers with multi-sleep modes for cloud data centers. The servers can make transitions between one active state and different sleep states, which involves different sleep power and transition delays for the sleep modes. We proposed Backtrack-and-Update method to make schedule of the servers, deciding how many servers in each state should be switched to which states in each timeslot, so that the total power consumption can be minimized while satisfying the QoS requirement. The problem is too large to be solved by existing methods, so we divide the whole problem and then conquer them one by one while considering the ongoing transitions during the breakpoints. We also consider using DVFS to further reduce the energy caused by the over provisioned computing capacity. Experiments show that our scheduling using multi-sleep modes can significantly reduce the total energy with QoS of less than 10ms. Against the over-provisioned strategy of AlwaysOn, our method can reduce more than 28% of the total energy on average.

REFERENCES

[1] P. X. Gao, A. R. Curtis, B.Wong, and S. Keshav, "It's not easy being green," ACM SIGCOMM Computer Communication Review, vol. 42, no. 4, pp. 211–222, Aug. 2012.

[2] A. Gandhi, M. Harchol-Balter, and M. A. Kozuch, "Are sleep states effective in data centers," in Proc. IEEE IGCC, Jun. 2012, pp. 1–10.

[3] L. Rao, X. Liu, L. Xie, and W. Liu, "Minimizing electricity cost: Optimization of distributed internet data centers in a multielectricity-market environment," in Proc. IEEE INFOCOM, 2010, pp. 1–9.

[4] Y. Zhang, Y. Wang, and X. Wang, "Greenware: Greening cloudscale data centers to maximize the use of renewable energy," in USENIX International Conference on Distributed Systems Platforms and Open Distributed Processing, Dec. 2011, pp. 143–164.

[5] S. Wang, Z. Qian, J. Yuan, and I. You, "A DVFS based energyefficient tasks scheduling in a data center," IEEE Access, vol. 5, pp. 13 090–13 102, 2017.

[6] C. Gu, H. Huang, and X. Jia, "Green scheduling for cloud data centers using ESDs to store renewable energy," in Proc. IEEE ICC, Apr. 2016, pp. 1–6.

[7] IBM, "CPLEX Users Manual," https://www.ibm.com/support/knowledgecenter/SSSA5P 12.7.0/ilog.odms.studio.help/pdf/usrcplex.pdf, 2017, [Online; accessed 25-December-2017].

[8] M. C. P. T. L. T. C. Hewlett-Packard Corporation, Intel Corporation, "Energy Management of ACPI," https://www.intel.com/content/dam/www/public/us/en/documents/articles/acpi-config-power-interface-spec.pdf, 2017, [Online; accessed 25-December-2017].

[9] D. G. Feitelson, D. Tsafrir, and D. Krakov, "Experience with using the parallel workloads archive," Journal of Parallel and Distributed Computing, vol. 74, no. 10, pp. 2967–2982, Oct. 2014.

[10] "Logs of real parallel workloads from production systems," http://www.cs.huji.ac.il/labs/parallel.

[11] D. D. Gutierrez, "The Intelligent Use of Big Data on an Industrial Scale," https://insidebigdata.com/2017/02/16/the-exponential-growth-of-data/, 2017, [Online; accessed 25-December-2017].