Green Latency-aware Data Deployment in Data Centers: Balancing Latency, Energy in Networks and Servers*

Yuqi Fan, Hongli Ding, Donghui Hu
School of Computer and Information
Hefei University of Technology

ABSTRACT

Two concerns exist in service provisioning by data centers. One is that users require to experience low latency while accessing data from the data centers. The other is to reduce the power consumed by network transport and servers in the data centers. In this paper, we tackle the problem of green data deployment in the data centers, taking into account the three factors of latency, energy consumption of the data centers and the network transport.

Categories and Subject Descriptors

C.2 [Internetworking]: Network Architecture and Design

Keywords

Energy-efficient, Latency-aware, Data deployment

1. INTRODUCTION

We use a network model similar to the one in [1]. The data accessed by the users must be deployed in some server(s) which may be in different data centers, and each data chunk is deployed only once. The objective is as follows.

Minimize:

$$\begin{aligned} & \lambda_1 \sum_{u_i, dc_j, s_m, d_k} rep(dc_j, s_m, d_k) p(u_i | d_k) l(u_i, dc_j) \\ & + \lambda_2 \sum_{dc_j, s_m} rep(dc_j, s_m) e_S(dc_j, s_m) \\ & + \lambda_3 \sum_{u_i, dc_j, s_m, d_k} s(d_k) rep(dc_j, s_m, d_k) p(u_i | d_k) e_I(u_i, dc_j) \end{aligned} \quad (1)$$

λ_1 , λ_2 , and λ_3 are the weights of the sub-objectives of the latency, the energy consumption of the data centers and the network transport, respectively. $rep(dc_j, s_m, d_k)$ indicates whether data d_k is deployed in server s_m in data center dc_j . $p(u_i | d_k)$ is the probability that a given request is asking for data d_k and it comes from user group u_i . $l(u_i, dc_j)$ is the latency between user group u_i and

*The work reported in this paper was supported in part by Anhui Provincial Natural Science Foundation (1208085MF101) and National Natural Science Foundation of China (61272540)

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

DCC'14, August 18, 2014, Chicago, Illinois, USA.

ACM 978-1-4503-2992-7/14/08.

<http://dx.doi.org/10.1145/2627566.2627584>.

data center dc_j . $rep(dc_j, s_m)$ is the indicator whether server s_m in data center dc_j has been deployed some data. $e_S(dc_j, s_m)$ is the energy consumption of server s_m in data center dc_j . $s(d_k)$ is the size of data d_k . $e_I(u_i, dc_j)$ is the energy required to transport one bit from a data center to a user through the Internet.

We propose an algorithm GLDD (Green Latency-aware Data Deployment) to allow operators to tune the three factors during data deployment. When processing each data chunk d_k , GLDD searches the servers in all the data centers with the least cost to deploy data d_k . Each server in each data center is checked to obtain the cost to accommodate data d_k on the server if the server has enough capacity. The cost of deploying data d_k on server s_m in data center dc_j integrates the three factors above.

We evaluate the performance of the algorithm GLDD by comparing GLDD with the algorithm FORTE proposed in [2]. We investigate four cases in the simulation. *Case 1* and *case 2* only consider the sub-objectives of the latency and the data center power consumption in the objective function, respectively. Energy in transport is ignored in *case 3*. *Case 4* tries to strike a three-way tradeoff among the three sub-objectives.

Figure 1: Performance comparison of GLDD and FORTE.

The performance versus different number of data and users is depicted in Fig. 1. GLDD achieves better results than FORTE in *case 3* and *case 4*, since the reduce of the latency and the network power consumption cannot compensate the increase of the data center power consumption with FORTE. However, FORTE outperforms GLDD in *case 1* and *case 2*. GLDD cares about only one of the two factors of the latency and the data center energy consumption, and each data chunk has only one copy, which potentially increases the latency or the power consumed by transport.

2. REFERENCES

- [1] J. Baliga, R. W. A. Ayre, K. Hinton, and R. S. Tucker. Green cloud computing: Balancing energy in processing, storage, and transport. *Proceedings of the IEEE*, 99:149–167, January 2011.
- [2] P. X. Gao, A. R. Curtis, B. Wong, and S. Keshav. It's not easy being green. In *SIGCOMM*, 2012.