Optimal Routing and Resource Allocation for Multimedia Cloud Computing

Wenqiang Gong
Communication University of China
School of Information Engineering
Beijing,100024,China
gwq7000@126.com

Jinyao Yan
Communication University of China
Beijing,100024,China
jyan@cuc.edu.cn

Zheng Chen
Communication University of China
Computer and Network Center
Beijing,100024,China
zhchen@cuc.edu.cn

***Abstract*—Routing and resource allocation are two major research directions for cloud computing, especially for improving the response time in multimedia cloud computing. In this paper, we propose network model for the transmission time between data centers and design routing algorithms for multimedia cloud computing. We further propose resource allocation with the goal to minimize the resource cost. We show some first simulation results of our proposed algorithms.**

I. Introduction

As one of the most popular IT technology, cloud computing technology has been sweeping various industries in an irresistible trend. The media cloud services are hosted by multiple geographically distributed data centers. The network routing strategy and resources allocation in the media cloud, has been studied by many scholars in related fields, which are respectively impact on the resource transmission time and the response times of cloud service. The shorter response time means the better Quality of Experience (QoE).

Routing and resource allocation are two major research directions for cloud computing. However, there is much research work on the routing topology inside data center while little on routing algorithms among data centers and little on the resource allocation scheme. In our previous work [2], we construct a simple network model between data centers and propose a heuristic algorithm to solve the resource allocation scheme problem. Nan et al. [1] propose the optimal VMs allocation schemes for single-site cloud and multi-site cloud. Zhan F B. [3] and Goldberg A V. [4] show us some ways of computing shortest paths over a network.

To address aforementioned challenges, we analyze the network factors of transmission time and study the influence of network factors on the resource allocation in clouds. Specifically, we study on the optimal routing strategy and the resource allocation strategy for VMs in the media cloud.

The rest of the paper is organized as follows. In Section II we analyze factors of transmission time, propose routing algorithms, and form resource allocation problem. Experimental evaluations and result analysis are performed in Section III. Finally, we conclude the paper in Section V.

This work is supported by Beijing Higher Education Young Elite Teacher Project.

II. Routing and Resource Allocation

A. The network model of multimedia cloud

In the PaaS cloud network architecture, data centers provide VMs to process and storage resource for applications and transmit resource file through fiber-optic network. Clients connect to data centers through access-network to request multimedia applications services which are performed at multiple data centers. The whole cost time of resource transmission T between two data centers is consist of four parts of time: the propagation delay T_{prop}, the transmission delay T_{tran}, the route processing delay T_{proc}, the route queuing delay T_{queu}. We derive the formula below [2][4]:

$$\begin{aligned} T &= T_{tran} + T_{prop} + T_{proc} + T_{queue} \\ &\approx \frac{Filesize}{B} + \frac{Distance}{C} \end{aligned} \tag{1}$$

Through the analysis of the above formula, we find T_{proc} and T_{queue} can be ignored when compared with the values of T_{tran} and T_{prop}. Through further analysis, we we can get that there are three main factors impact on the transmission time between cloud networks, there are the size of file to be transmit $Filesize$, the bandwidth of the link B, the physical distance between data centers $Distance$. C here is the light speed, which is a constant.

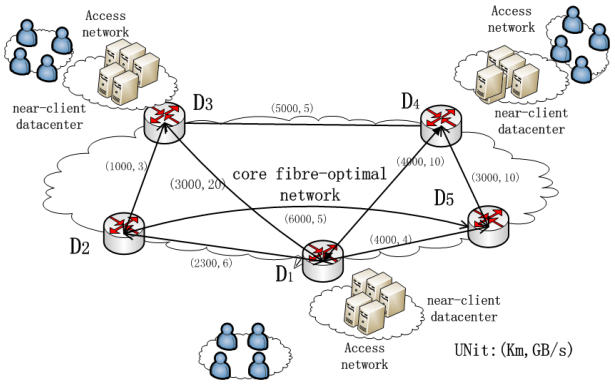

Fig. 1. The network environment of multimedia cloud data centers

QSHINE 2014, August 18-20, RHODES, Greece
Copyright © 2014 ICST
DOI 10.4108/icst.qshine.2014.256415

B. Optimal Routing based on Dijkstra Algorithm

In the muti-site cloud, a centralized controller is typically used for resource planning and routing between data centers. Dijkstra Algorithm is a routing algorithm in graph theory [6], which is mainly used to compute the best path in consideration of different metrics in real network environment. In the basic routing Dijkstra Algorithm, hop count is generally used as a reference index. However, we set some network parameters as metric to improve the transmission performance for the multimedia cloud.

Based on Eq.(1), we propose a bandwidth-based routing algorithm based on Dijkstra algorithm using the bandwidth of network link B as metric, and a distance-based routing algorithm based on Dijkstra algorithm using distance between data centers $Distance$ as metric. In the bandwidth-based routing algorithm, we take the smallest value of bandwidths of the links in the path as the bandwidth metric of the path [7]. For example, datacenter D_1 connect D_5 via D_4 is the shortest path, the bandwidth metric of path B_{1-5} is minimize $\{B_{1-4}, B_{4-5}\}$. Both algorithms are to reduce the time of file transmission in data center networks and further reduce the cost in the resource allocation scheme finally.

C. The Resource Allocation Scheme

Now we present our resource allocation model and then form into an optimization problem. We define I as set of VMs, J as set of applications. The goal is to minimize the cost. Let K be the number of VMs. r and d are the reservation and on-demand price plans P for I types VMs and J types application respectively. We consider the optimal resource allocation problem as minimization problem in Eq.(2).

$$Minimize\ C(t) = \sum_{i \epsilon I} \sum_{j \epsilon J} P_{ij}^{r} K_{ij}^{r(t)} + P_{ij}^{d} K_{ij}^{d(t)} \qquad (2)$$

The optimization problem in (2) is proved to be an NP-Hard problem [1]. Besides, the number of VMs is decided by the processing time that the application allows. With shorter time at the same condition, the more VMs that needed. As shown in Eq.(3), the application processing time T_j^{Pro} must be less than the value of application response time T_j^{res} minus the data transmission time T_j^{tra} in data centers network and T_j^{Acc} in user access network. T_j^{tra} is computed by Eq.(3) and T_j^{Acc} can be treated as a constant in our network model.

$$T_j^{Acc} + T_j^{Pro} + T_j^{Tra} \leq T_j^{Res}, \forall j \epsilon J \qquad (3)$$

In previous work [2], we proposed a heuristic algorithm to solve the problem (2). This time we will further study optimal allocation algorithms based on our proposed optimal routing algorithms.

III. SIMULATION

In our initial simulations, we assume five data centers $D = \{D_1, \cdots, D_5\}$ with distance and bandwidth shown in Figure 1.

We first set the file size as a fixed value, and compute the transmission time of ten paths between the five data centers, namely $\{D_1-D_2, D_1-D_3, \cdots, D_2-D_3, \cdots, D_4-D_5\}$. Paths are computed by two proposed routing algorithms. Then we change the value of file size for different multimedia applications. Figure 2 shows the simulation results of transmission time for two proposed algorithms .

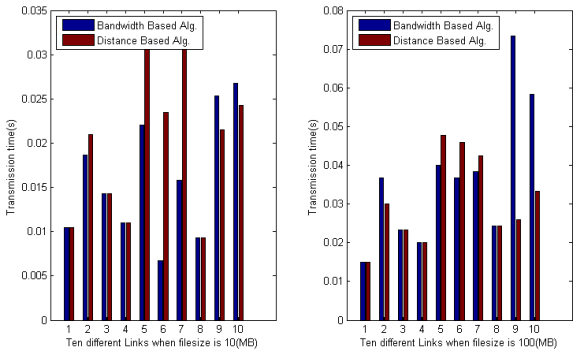

Fig. 2. Transmission time of file sizes of 10MB and 100 MB

We find the transmission time change greatly when bandwidth and the distance are set as reference factor. While increasing the size of files to be transferred ten times, the transmission time of each path increase and the growth rate of different paths changes differently for two proposed algorithms.

IV. CONCLUSION AND FUTURE WORK

In this paper, we study optimal routing and resource allocation for multimedia cloud computing. We propose two routing algorithms to reduce the transmission time, and propose resource allocation schemes to minimize the resource cost. In our future work, we will extend our simulations in more realistic environments and jointly study optimal routing algorithms and resource allocation schemes.

REFERENCES

[1] Nan X, He Y, Guan L, *Optimal allocation of virtual machines for cloud-based multimedia applications*[C]. 3rd ed. Multimedia Signal Processing (MMSP). 2012 IEEE 14th International Workshop on. IEEE, 2012.

[2] Gong W, Jinyao Y, Chen Z, *An Optimal VM Resource Allocation for Near-Client-Datacenter for Multimedia Cloud*[C]. International Conference on Ubiquitous and Future Networks(ICUFN).2014.In publishing.

[3] Zhan F B, *Three fastest shortest path algorithms on real road networks: Data structures and procedures*[J]. Journal of geographic information and decision analysis, 1997, 1(1): 69-82.

[4] Goldberg A V, Harrelson C, *Computing the shortest path: A search meets graph theory*[C]. Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics, 2005: 156-165.

[5] Ramaswamy R, Weng N, Wolf T, *Characterizing network processing delay*[C]. Global Telecommunications Conference2004. GLOBECOM'04. IEEE, 2004, 3: 1629-1634.

[6] Mooney P, Winstanley A, *An evolutionary algorithm for multicriteria path optimization problems*[J]. International Journal of Geographical Information Science, 2006, 20(4): 401-423.

[7] Sahni S, Rao N, Ranka S, *Bandwidth scheduling and path computation algorithms for connection-oriented networks*[C]. Networking, 2007. ICN'07. Sixth International Conference on. IEEE, 2007: 47-47.