A Bandwidth Allocation Scheme to Improve Fairness in Data Center Networks

Yusuke Ito
University of Kitakyushu,
Kitakyushu, Japan
wash000i@net.is.env.kitakyu-u.ac.jp

Hiroyuki Koga
University of Kitakyushu,
Kitakyushu, Japan
h.koga@kitakyu-u.ac.jp

Katsuyoshi Iida
Tokyo Institute of Technology,
Tokyo, Japan
iida@gsic.titech.ac.jp

ABSTRACT

The development of various Internet technologies has enabled users to connect to the Internet through a wide variety of access networks and communication terminals. This has brought a great demand for cloud computing which allows users to enjoy various Internet services provided by data centers (DCs) anytime, anywhere. In cloud computing, though, users more distant from DCs receive a poorer quality of service in terms of transmission rates because of the TCP RTT-unfairness problem. Therefore, we propose a bandwidth allocation scheme based on collectable information in DC networks to enable fair communication service to each user. We have confirmed the effectiveness of this approach through simulation evaluations.

Categories and Subject Descriptors

C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design

Keywords

Bandwidth Allocation, TCP, Data Center

1. INTRODUCTION

Users can now connect to the Internet through a wide variety of access networks and communication terminals such as smart phones or PCs. This has created a great demand for cloud computing which allows users to enjoy various Internet services provided by data centers (DCs) [1] anytime, anywhere. Major services available through cloud computing include file sharing or transaction services which commonly use TCP as a reliable data transmission protocol.

In cloud computing, a user's quality of service in terms of transmission rate depends on the distances from DCs. This problem is caused by how TCP controls congestion. TCP cannot exactly identify the conditions of other flows, so it estimates an available bandwidth based on packet losses. Since congestion control is applied according to round-trip time (RTT), the transmission rate for a user with a short RTT is effectively increased compared with that of a user having a long RTT. Fairness in terms of throughput among flows with different RTTs needs to be improved.

To improve fairness among flows, CHOKeW [2] has been proposed as one of the active queue management (AQM) technologies applied on routers. This approach adjusts the packet dropping rate per flow based on service classes and network conditions. It realizes fair communication among flows by preferentially discarding packets which belong to a high rate flow. However, the processing cost on routers increases enormously as the number of flows increases.

In this study, we propose a bandwidth allocation scheme based on collectable information in DC networks to provide fair communication service to each user. This scheme collects flow information including the bandwidth of each link, the number of competing flows, and the RTT of each flow from routers and servers in DC networks and then fairly allocates transmission rates among flows based on the collected information. We show the effectiveness of this approach through simulation evaluations.

2. PROPOSED SCHEME

The proposed scheme fairly allocates transmission rates among flows based on collectable information in DC networks as shown in Fig. 1. Flow information such as the bandwidth of each link, the number of competing flows, and the RTT of each flow is collected by a Unified Congestion Control Architecture (UCCA) [3], which uniformly manages the congestion information in the networks. Routers in DC networks periodically inform servers of the number of existing flows, which can be measured by extending the OpenFlow architecture. The servers identify bottleneck links based on the information about the number of flows received from the routers and the bandwidth of each link, and calculate the transmission rate which should be allocated to each flow by dividing the bandwidth of the bottleneck link by the number of competing flows.

For down direction communication from a server to a client, the server can simply allocate the calculated transmission rates to each flow. On the other hand, for up direction communication from a client to a server, the server needs to inform the client of the allocated transmission rate. Our scheme uses TCP's advertised window size to notify the client of the transmission rate. The server calculates the

CoNEXT Student Workshop'14, December 2, 2014, Sydney, Australia.

http://dx.doi.org/10.1145/2680821.2680834.

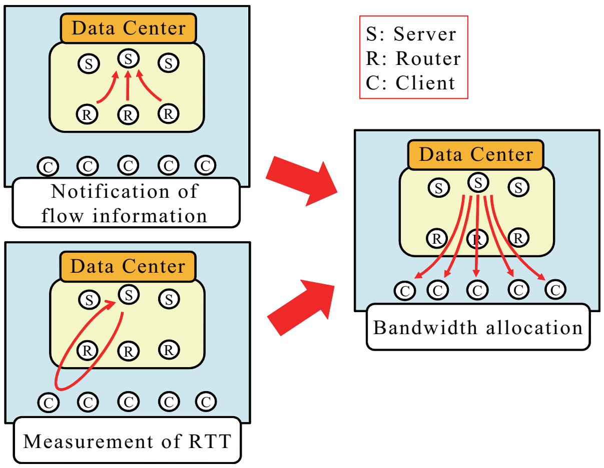


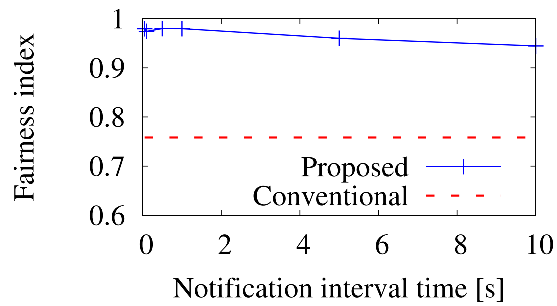
Figure 1: Proposed Scheme

bandwidth delay product of each flow as the advertised window size based on the bandwidth which should be allocated and the RTT of each flow. That is, the advertised window size of a flow having a long RTT is larger than that of a short RTT flow. The RTT of each flow is measured by the server's TCP. If the server does not send any data to the client, it periodically sends probe packets to measure RTT. This bandwidth allocation should enable fair communication among flows which have different RTTs. However, the number of competing flows and the RTT of each flow vary dynamically in an actual environment. Therefore, the accuracy of the bandwidth allocation is likely to depend on the interval time needed to collect the flow information.

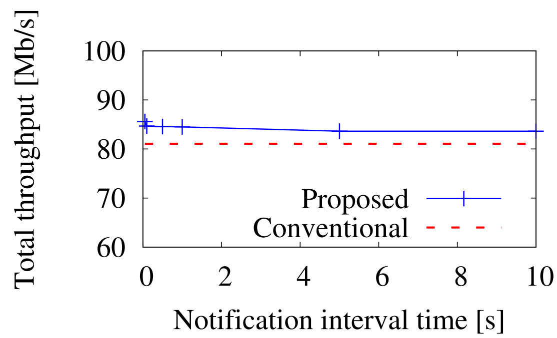
3. PERFORMANCE EVALUATION

To investigate the efficiency of our scheme, we evaluated it through simulation using Network Simulator ns-3 after its implementation. In the simulation, we used a tree topology to focus on the basic characteristics of the scheme. A server connected two edge routers through a core router. Each edge router connected five clients separately. Thus, there were two client groups and these had access links with different delay times: 2 and 7 ms. Other links had a delay time of 1 ms. The bandwidth of all links was 100 Mb/s. One client of each group started to send 100 MB of data to the server every 10 seconds, so two flows were joined every 10 seconds. The TCP algorithm was NewReno. The interval time to notify the flow information varied from 0.05 to 10 s. We evaluated Jain's fairness index and total throughput for the proposed scheme and conventional TCP.

Figures 2(a) and 2(b) show the fairness index and total throughput of the proposed and conventional schemes when the notification interval time varied from 0.05 to 10 s. The proposed scheme provided higher fairness and throughput than the conventional scheme regardless of the notification interval time. This was because the proposed scheme allocates an adequate advertised window size to each flow according to RTT. In addition, the proposed scheme suppresses packet drops on routers through bandwidth allocation, thus improving throughput.



(a)Fairness index



(b)Total throughput

Figure 2: Effect of notification interval time

4. CONCLUSION

Our proposed bandwidth allocation scheme uses collectable information in DC networks to provide fair communication service to each user. Simulation evaluations show that this scheme enables higher fairness and throughput than can be achieved with conventional TCP. In our future work, we will consider the most effective way to fairly allocate link bandwidth in DC networks to each user even when access networks are bottlenecks.

5. ACKNOWLEDGMENTS

This work was supported in part by JSPS KAKENHI Grant Number 25280028.

6. REFERENCES

[1] J. Zhang, F. Ren, and C. Lin, "Survey on transport control in data center networks," *IEEE Network Magazine*, vol. 27, no. 4, pp. 22–26, July/August 2013.

[2] S. Wen, Y. Fang, and H. Sun, "Differentiated bandwidth allocation with TCP protection in core routers," *IEEE Transactions on Parallel and Distributed Systems*, vol. 20, no. 1, pp. 34–47, January 2009.

[3] K. Kusuhata, I. Kaneko, K. Iida, H. Koga, and M. Shimamura, "A unified congestion control architecture design to improve heterogeneous wireless network efficiency and accommodate traffic by various rich applications," *Proc. IFIP HET-NETs2013*, p. 5, November 2013.