A QoS-Based Scheduling Mechanism for Overlay Aggregate Traffics

Yunbo Wu[1,2], Zhishu Li[1], Zhihua Chen[3],Yunhai Wu[4], Li Wang[1],Tun Lu[1]

[1]School of Computer Science, Sichuan University,24 southern section1,1st Ringroad, ChengDu, China
(ybwutm,tm_001,tm_002,ths_01)@tom.com
[2]Ningbo Fashion Institute, Ningbo, China
[3]Kunming Meteorological Administration, west of XiHua Park, Kunming, China
khzitm@tom
[4]Yunnan Design Institute of Water Conservancy and Hydroelectric Power, Kunming, China
juan01_tom@tom.com

Abstract. This paper presents a control mechanism in SON(Service Overlay Network) architecture to provide delay guarantees of different aggregate traffics over IP network. In addition to the bandwidth provision problem of aggregate-traffic in service overlay network, queuing delay in the service gateway is another critical QoS parameter, and diverse multimedia flows have different delay requirements. To assure the delay requirements of aggregate flows in service overlay network, we present a simple but effective adaptable control mechanism, which consists of two-queues with exhausitive service in cyclic order and characterizes the abilities to ensure delay requirements of various aggregate flows according to their burstiness. Moreover, Associated performance is analyzed by the Markov chain and probability generation function, and simulations validate this analysis.

1 Introduction

With the explosive of growth of the Internet and multimedia applications, the need for QoS guarantees is becoming more and more ubiquitous in the networks of future. Moreover, Everything over IP has been regarded as a preferred solution of next-generation Internet, then IPQoS emerges paramount importance in multimedia delivery. In order to achieve IP QoS, several new significant architectures such as IntServ, DiffServ, and MPLS have been proposed by IETF in last few years. However, these and other proposals for Internet QoS, have two key requirements: first, they require all routers along a path to implement QoS mechanisms for scheduling and buffer management, and second, they require the right incentives for Internet Service Providers (ISPs) to enable these functions. Unfortunately, these demands have often turned out to be difficult to meet. Internet QoS remains to be an open issue till now.

On the other hand, the Internet has evolved to become a commercial infrastructure of service delivery. Many value-added and content delivery services are developed via overlay networks rather than IP-layer, such as content delivery overlays [1], p2p file sharing overlays [2]. Besides, Ref. [3,4] discussed service overlay networks and service composition, while Ref. [5] studied service grid. SON architecture provides more effective and efficient solution in QoS guarantee. Relying on the bilateral SLAs

the SON can deliver end-to-end QoS sensitive services to its users via appropriate provisioning and service-specific resource management. The underlying network domains can aggregate traffic based on the SONs they belong to and perform traffic and QoS control accordingly based on the corresponding SLAs.

In this paper, we'll deploy a control mechanism in SON architecture to provide delay guarantees of different aggregate traffic in service overlay networks. Considering that different types of multimedia traffics characterize various burstiness, our control model consists of two asymmetric queues with exhaustive service in a cyclic order. The model function independently in the service gateway of SON and has the ability of providing lower waiting delay to the flows with higher burstiness without any priority pre-assigning. Furthermore, we adopt an ingenious but efficient method to deal with this asymmetric polling system, and obtain exact average waiting time formula. Simultaneously, associated simulation verifications are presented.

The rest of the paper is organized as follows: in section 2 we describe our control model and make some analysis of this model, in section 3 we address our concluding remarks.

2 Model Evaluation

Our model consists of two queues with exhaustive service in a cyclic order. Considering that different types of multimedia traffics characterize various burstiness, we may classify multimedia traffics in to two queues according to burstiness, viz. one queue with larger burstiness and another with smaller ones. When server visits a queue in turn, not only all the packets already waiting in the queue will be served, but also comprise packets arriving during the service intervals. After completing the service of one queue, the server will move to the succeeding queue and incur a switching-over time, then start to serve the queue in the same way. There are three independent random processes relating to each queue separately, i.e., the packets arriving, servicing upon a queue, and switching-over from one queue to another. Considering difference between two traffics, parameters associated with these random processes are generally not the same. Namely, this queuing model is asymmetric queuing system. In terms of QoS requirements many metrics are concerned, such as delay, delay jitter and packet loss. Our work will focus on the performance of delay of multimedia resources, which incurred during networking node. In our analysis, we condition the discrete system and infinite capacity in each queue. Then Related joint generating functions are defined as follows:

$$G_1(z_1, z_2) = R_2\big(A_1(z_1)A_2(z_2)\big)\, G_2\big(z_1\ ,\ B_2\big(A_1(z_1)F_2(A_1(z_1))\big)\big) \tag{1}$$

$$G_2(z_1, z_2) = R_1\big(A_1(z_1)A_2(z_2)\big)\, G_1\big(B_1\big(A_2(z_2)F_1(A_2(z_2))\big),\ z_2\big) \tag{2}$$

From equations (1)-(2), mean packet waiting time in queue Qi can be obtained as following:

$$\overline{w_i} = \frac{g_i(i,i)}{2\lambda_i g_i(i)} + \frac{\left(2\rho_i^2 + \rho_i - 1\right)A_i''(1)}{2\lambda_i^2\left(1-\rho_i^2\right)} + \frac{\lambda_i B_i''(1)}{2\left(1-\rho_i\right)}, \quad i = 1, 2. \tag{3}$$

Moreover, we'll present some numerical results of packet average waiting time in each queue according to Equations (3) (Refer to Fig.1 and Fig.2). Where w_{1s}, w_{2s} represent the simulation value of w_1, w_2, respectively.
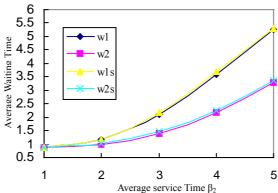

Fig. 1. Average waiting time versue average service time
($\gamma_1 = \gamma_2 = 1, \lambda_1 = \lambda_2 = 0.1, \beta_1 = 1$)

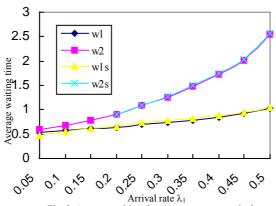
Fig. 2. Average waiting time versus average arrival rate
($\gamma_1 = \gamma_2 = 1, \beta_1 = \beta_2 = 1, \lambda_2 = 0.01$)

Fig.1 shows that the queue with larger value of β has lower mean waiting time than the other one. The reason is that service related to the queue has been intensified. While Fig.2 depicts that the average waiting time related to having higher arrival rate is always smaller than the other related to having lower arrival rate. Moreover, in both cases the waiting time increases as λ_1 raising,   just for the reason that the load of system increases also. In addition, simulation results show good agreement with the analytical results.

3   Conclusions

From the discussions above-mentioned, we can draw the following conclusions: our scheduling model can provide lower waiting delay to the resource with higher burstiness without any priority pre-assigning. Meanwhile changing the value of β can influence the waiting delay, which implies that the length of information packet can affect waiting delay. These will benefit multimedia applications delivery over IP networks.

QoS provision for multimedia application delivery in Internet is a crucial and complex issue. Traditional IP QoS mechanisms, such as the ones offered by Interserv and Diffserv, have encountered more difficulties in implementation and deployment in current Internet. Instead, the method that provides value-added with QoS guarantee based on overlay networking is getting more attractive and to be a promising solution in Internet multimedia applications. Our control machinery can be embedded independently in the service gateway of SONs as a supplement, it can provide delay guarantee to various traffics with different burstiness. This combination architecture may implement all-around end-to-end QoS guarantees. To extend our research, we will further our control model to satisfy QoS demands of multi-types multimedia traffics, and continue our research on the end-to-end QoS issue of various overlay networks.

References

1. S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker. CAN: A scalable content-addressable network. In Proceedings of ACM Sigcomm, August 2001.
2. Gnutella. http://gnutella.wego.com/.
3. L. Subramanian, I. Stoica, H. Balakrishnan, and R. H. Katz. OverQoS: Offering QoS using Overlays. Proc. of First Workshop on Hop Topics in Networks (HotNets-I), Princeton, New Jersey, October 2002.
4. X. Gu, K. Nahrstedt, R. N. Chang, and C. Ward: "QoS-assured service composition in managed service overlay networks," in the 23rd IEEE International Conference on Distributed Computing Systems (ICDCS2003), May 2003.
5. Gill Waters, John Crawford, Sei Guan Lim: Optimising multicast structures for grid computing. Computer Communications 27 (2004) 1389–1400