International Journal of Grid and Distributed Computing
Vol.7, No.2 (2014), pp.69-78
http://dx.doi.org/10.14257/ ijgdc.2014.7.2.07

Admission Control for Aggregate Flow in Cloud Computing

Yanbing Liu[1], Yunlong He[2] and Jun Huang[2]

School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
liuyb@cqupt.edu.cn, iminor_he@163.com, xiaoniuadmin@gmail.com

Abstract

With the development of cloud computing, the number and types of cloud services experience an explosive increase. Towards improving the Quality of Service provided in cloud computing, we model cloud computing's admission control in this paper. Based on the observation that cloud service requests are typically concurrent, we propose admission control method for aggregate flow by computing the its effective bandwidth using network calculus. Acceptance decision is consequently made by computation with the effective bandwidth. Experiments are conducted to investigate properties of effective bandwidth, and to evaluate the performance of the admission control method under different settings.

Keywords: *Admission control; cloud services; aggregate flow; network calculus*

1. Introduction

In recent years, cloud computing has gradually become a popular research subject [1, 2]. With the interest in cloud computing continuously growing, the number and types of cloud services experience an explosive increase. As cloud services must be done via the Internet, the traditional Internet with limited resources is challenged unprecedentedly. Meanwhile, users demand higher Quality of Service (QoS), which makes the traditional "best effort" service offered by the network unfeasible anymore. Call Admission Control (CAC), as a means of preventive traffic control, is one of the most effective methods that can be used to avoid network congestions and to guarantee the QoS for cloud services [3]. In order to provide a guaranteed QoS for cloud services, admission control algorithm needs the bound functions in the worst-case scenario to make acceptance decisions. At present time, one of the most effective ways to calculate the bound functions for worst cases is the network calculus.

The rest of this paper is organized as follows. Section 2 overviews the related work in admission control for cloud computing in the literature. Section 3 elaborates the foundations of network calculus. Section 4 outlines our admission control model in cloud computing. Section 5 proposes the effective bandwidth model for aggregate flow and defines the decision-making strategy. Section 6 presents the performance analysis of admission control models with different delay constraints. Section 7, finally, concludes the paper.

2. Related Work

The theory of Network Calculus (NC) was first developed by Chang [4] and Cruz [5, 6] and then extended by others (*e.g.*, see [7-10]). NC is a novel and effective tool of quantitative mathematics for network performance analysis, and computes deterministic bounds on delays,

ISSN: 2005-4262 IJGDC

backlogs and other QoS parameters using arrival curves and service curves [11, 12]. The most notable advantage of NC, when compared with other traditional statistical theory, is that it provides deterministic boundary analyses [13]. NC aims at analyzing critical behaviors and usually focuses on worst-case performances so that it can provide strict QoS guarantees for services [14, 15]. Evidently, NC has been widely used as a tool to analyze network performance, thereby obtaining tight bound functions to facilitate the admission control of flows [16, 17].

At present, most of the studies (*e.g.*, see [3, 18]) in admission controls deal only with per flow. As the cloud computing develops rapidly, there are thousands of users simultaneously requesting cloud service every day, thus per-flow-based admission control approach fundamentally restricts the network's ability to handle concurrent requests. In [19], Le Boudec et al. defined the notion of Effective Bandwidth (EB) of flow which is based on delay constraint, and proposed that the notion can be used to implement the admission control. However, it is not clear that how EB can be used in the context of admission control for aggregate flow in cloud computing [20].

In this paper, we present a network model and a system model in terms of the admission control in cloud computing, and propose to improve the network's ability of handling the concurrent cloud service requests by devising admission control methods that target on aggregate flow and utilize EB in making acceptance decisions. Relevant experiments are conducted to investigate the properties of EB for aggregate flow, and to evaluate the performances of EB-based Admission control (EBAC) with different parameters.

3. Network Calculus Theory

We now give the definition of some of the notions in NC that will be needed in the rest of the paper to acquire EB for aggregate flow in cloud computing. The detailed descriptions of these concepts can be found in [19].

DEFINITION 1 (Wide-sense Increase Function). Given a function f defines for $\forall s,t \in (R \bigcup \{+\infty\})$ we say that f is wide-sense increase function if and only if $s \le t : f(s) \le f(t)$.

DEFINITION 2 (Arrival Curve). Given a wide-sense increasing function α, we say that a flow R has α as an arrival curve, or R is α-smooth, if and only if R meets one of the following two equivalent conditions for $\forall t \ge 0, s \le t$:

$$R(t) - R(s) \le \alpha(t-s) \tag{1}$$

$$R(t) \le (R \otimes \alpha)(t) \tag{2}$$

where $\otimes$ is the min-plus convolution and is given as follows:

$$(R \otimes \alpha)(t) = \begin{cases} \inf_{0 \le s \le t}\{R(s) + \alpha(t-s)\}, & t \ge 0 \\ 0, & t < 0 \end{cases} \tag{3}$$

DEFINITION 3 (Service Curve). Consider a system S and a flow through S with input and output function R and R^*. We say that S offers to the flow a service curve β if and only if β is wide sense increasing, $\beta(0) = 0$ and $R^* \ge R \otimes \beta$.

DEFINITION 4 (Effective Bandwidth). For a flow with an arrival curve α, the effective bandwidth $e_D(\alpha)$ of the flow is defined to be the bit rate required to serve the flow in a work conserving manner, with a delay constraint D. That is,

$$e_D(\alpha) = \sup_{s \ge 0} \left\{ \frac{\alpha(s)}{(s+D)} \right\} \tag{4}$$

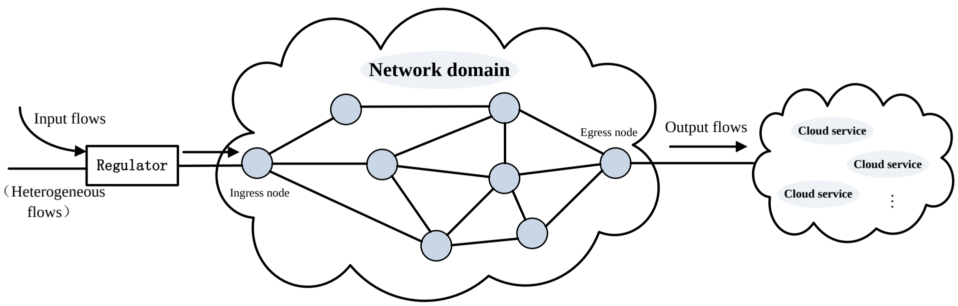
Regarding the effective bandwidths of aggregate flow, we may assume $D_i = D$ for any i (where i represents different types of flows), then we have:

$$e_D(\sum_i \alpha_i) \le \sum_i e_D(\alpha_i) \tag{5}$$

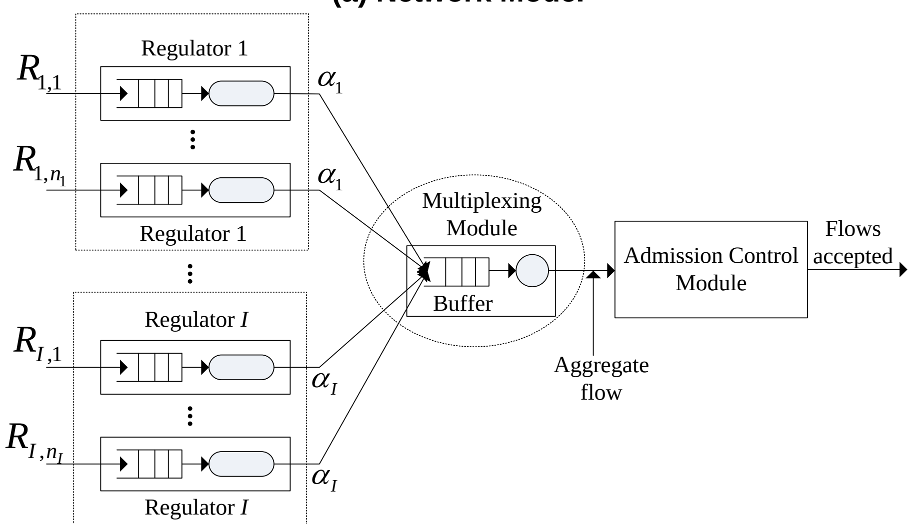
If this assumption causes trouble, then we may provide a guaranteed delay for every cloud service flow by letting $D = \min\{D_i\}$.

4. The Framework

The framework of admission control in cloud computing is shown in Figure 1 which consists of a network model and a system model.



(a) Network Model



(b) System Model

Figure 1. Admission Control Model in Cloud Computing

Figure 1(a). is the network topology by which users request the cloud services. The process of users' requesting and accepting cloud services must be done via the traditional Internet. Figure 1(b) is the system model for admission control. The heterogeneous flows of user requests for cloud services are first shaped by the regulator; then the shaped flows are multiplexed through the FIFO multiplexing module; and finally the output from the multiplexing module, which is the aggregate flow, goes through the phase of admission control producing the accepted flows. The multiplexing module and admission control module are deployed at the position of the ingress node in Figure 1(a), so that those flows that are accepted will enter the network domain.

As shown in Figure 1(b), the heterogeneous flows share a common buffer when they are multiplexed, which indicates that this mechanism requires less bandwidth resources than that where each flow is allocated a fixed size buffer. This is exactly the situation described by the inequalities in (5). In other words, the admission control with respect to aggregate flow can admit more cloud services than that of per flow when the bandwidth of the ingress node is a constant.

Heterogeneous flows' bursts are smoothed through the regulator before being multiplexed. We consider the case where the output flows of regulators are constrained by the traffic specification $T\text{-}SPEC(p,M,r,b)$. $T\text{-}SPEC(p,M,r,b)$ is the shaping curve of the regulator, and is the arrival curve of the output flows of regulators as well, where parameters *p, M, r, b* are peak rate, maximum packet size, sustainable rate (average rate), and burst tolerance of a flow, respectively. A flow which has $T\text{-}SPEC(p,M,r,b)$ as its arrival curve is the so called Variable Bit Rate (VBR) flow. The specific constraint function is as follows:

$$\alpha(t)=\begin{cases}\min(pt+M,rt+b), & t\geq 0\\ 0, & t<0\end{cases} \tag{6}$$

5. The Admission Control Method

We now give the admission control details in this section.

5.1. Effective Bandwidth for per Flow

Equation (4) defined in Section 3 are the EB of a single flow in networks. Figures 2 below depict the EB of the VBR flow which satisfies the condition of *i*=3 in Table 1 in Section 6.

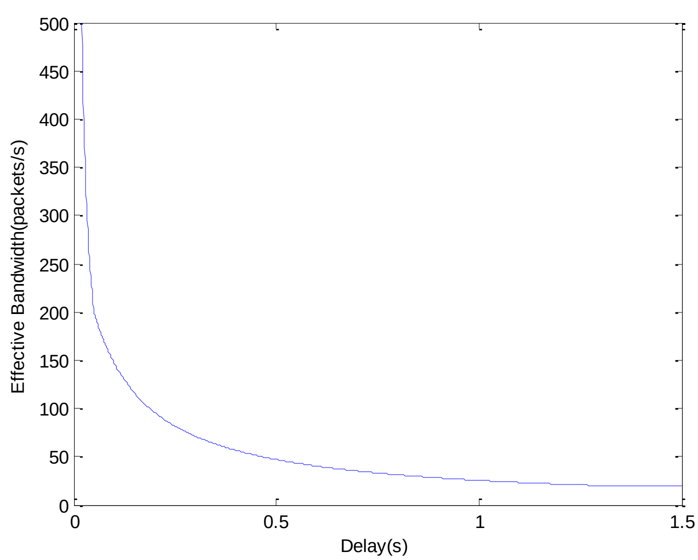


Figure 2. EB for VBR Flow

It can be seen clearly from the figure that EB keep decreasing as the delay increase, until a minimum value is hit. In other words, the cloud service rate at the ingress node may keep dropping but cannot be lower than the average rate of flow.

5.2. Effective Bandwidth for Aggregate Flow

We now turn our attention to EB for the aggregate flow. Suppose there are *I* types of cloud services, and the multiplexing module multiplexes n_i flows of type *i*, where n_i is the number of cloud services of type *i* (*i*=1,2,…,*I*). Every flow has $T\text{-}SPEC(p_i, M_i, r, b_i)$ as an arrival curve. The multiplexing process is as shown in Figure 1(b), and a fixed, but arbitrary, delay constraint *D* is set for each flow. The Effective Bandwidth for aggregate flow is given by the following formula:

$$e_D(\sum_{i=1}^{I} n_i \alpha_i)$$

where $\alpha_i = T\text{-}SPEC(p_i, M_i, r, b_i)$ and $\sum_{i=1}^{I} n_i \alpha_i$ is the aggregated arrival curve of the *I* types of flows.

The issue of how to compute the EB for aggregate flow is addressed below. In the following calculations, $\Gamma_i = (b_i - M_i)/(p_i - r_i)$, and it is assumed that $\Gamma_1 \le \Gamma_2 \le \cdots \le \Gamma_I$.

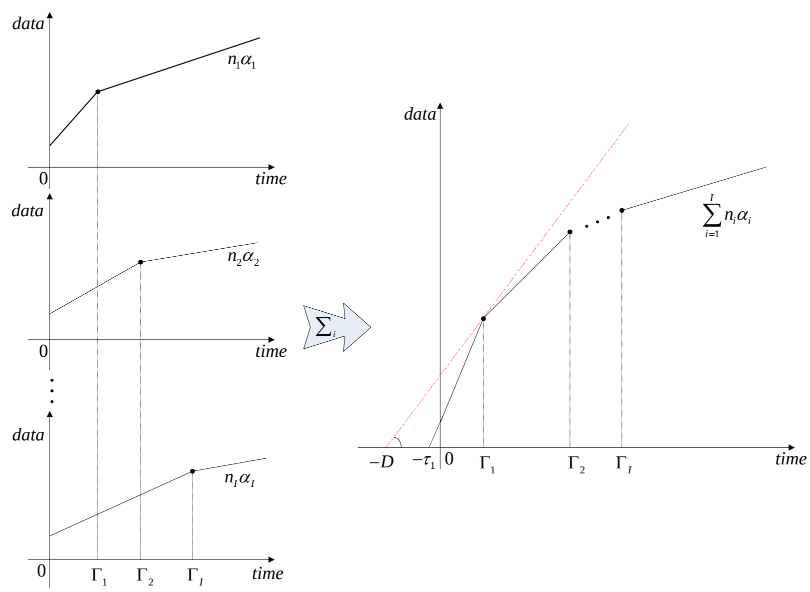


Figure 3. Calculation of Effective Bandwidth for Aggregate Flow

Based on Definition 4, EB for the aggregate flow after *I* types of cloud service flows are multiplexed can get easily from the mathematical induction, as shown in Figure 3, as follows: the slope of the tangent line to the aggregated arrival curve at *time* = –*D* is the desired EB. Specifically,

$$e_D(\sum_{i=1}^{I} n_i\alpha_i) = \max\{e_1, e_2, e_3, \cdots, e_{I+1}, e_{I+2}\}$$
$$= \begin{cases} e_1, & 0 < D \le \tau_1 \\ e_2, & \tau_1 < D \le \tau_2 \\ e_3, & \tau_2 < D \le \tau_3 \\ & \vdots \\ e_{I+1}, & \tau_I < D \le \tau_{I+1} \\ e_{I+2}, & D > \tau_{I+1} \end{cases} \tag{7}$$

where,

$$e_1 = \sum_{i=1}^{I} n_i M_i \Big/ D,$$
$$e_2 = [(\sum_{i=1}^{I} n_i p_i)\Gamma_1 + \sum_{i=1}^{I} n_i M_i] \Big/ (\Gamma_1 + D),$$
$$e_3 = [(n_1 r_1 + \sum_{i=2}^{I} n_i p_i)\Gamma_2 + n_1 b_1 + \sum_{i=2}^{I} n_i M_i] \Big/ (\Gamma_2 + D),$$
$$e_{I+1} = [(\sum_{i=1}^{I-1} n_i r_i + n_I p_I)\Gamma_I + \sum_{i=1}^{I-1} n_i b_i + n_I M_I] \Big/ (\Gamma_I + D),$$
$$e_{I+2} = \sum_{i=1}^{I} n_i r_i$$
$$\tau_1 = \sum_{i=1}^{I} n_i M_i \Big/ \sum_{i=1}^{I} n_i p_i,$$
$$\tau_2 = (n_1 b_1 + \sum_{i=2}^{I} n_i M_i) \Big/ (n_1 r_1 + \sum_{i=2}^{I} n_i p_i),$$
$$\tau_3 = (\sum_{i=1}^{2} n_i b_i + \sum_{i=3}^{I} n_i M_i) \Big/ (\sum_{i=1}^{2} n_i r_i + \sum_{i=3}^{I} n_i p_i),$$
$$\tau_I = (\sum_{i=1}^{I-1} n_i b_i + n_I M_I) \Big/ (\sum_{i=1}^{I-1} n_i r_i + n_I p_I),$$
$$\tau_{I+1} = \sum_{i=1}^{I} n_i b_i \Big/ \sum_{i=1}^{I} n_i r_i$$

5.3. Admission Control Strategy for Aggregate Flow

Given that there are I types of cloud service flows to be multiplexed, and the delay is D for every flow; if the output rate of the ingress node is a constant C, then the acceptance/rejection of the aggregate flow is determined by the following conditions:

$$e_D(\sum_{i=1}^{I} n_i\alpha_i) \le C \tag{8}$$

Inequality (8) represents the EBAC method. A flow is accepted if and only if the relevant condition adopted in the admission control is satisfied. The set of maximum values satisfying the admission control condition denotes as $(n_1, n_2, ..., n_I)$. It is used to analyze the performance of the admission control and indicate how many cloud services the network can accept. And where $n_i \in N$.

6. Numerical Results and Analysis

We in this section analyze the characteristics of the EB for aggregate flow, and evaluate the performance of EBAC for cloud services. Specific *T-SPEC* parameters are shown in Table 1, and the selection of these parameters can be supported by [19].

Table 1. Parameters of Three Types of Cloud Services Flow

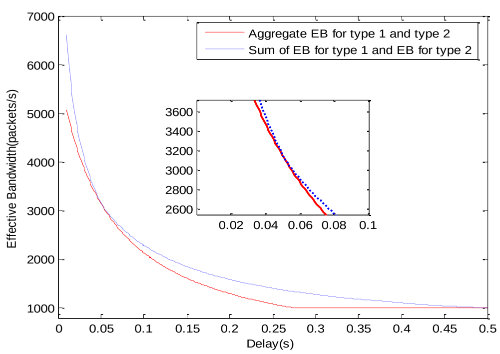
i	p_i (packets/s)	M_i (packets)	r_i (packets/s)	b_i (packets)	Γ_i (ms)
1	20000	1	500	26	1.3
2	5000	1	500	251	55.5
3	200	10	20	26	88.9

6.1. Properties of EB for Aggregate Flow

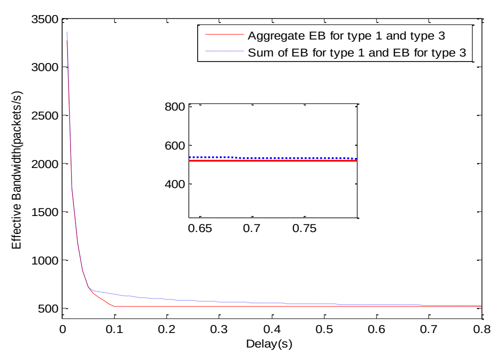
Figure 4 shows the characteristics of EB. As shown in Figures 4, the EB for an aggregate flow is always less than or equal to the sum of their respective EBs of each individual flow no matter I=2 or I=3. This also shows the correctness of Inequalities (5). Combined with the EB calculation discussed in the previous section, it can be seen that

$$e_D(\sum_i \alpha_i) = \sum_i r_i = \sum_i e_D(\alpha_i)$$

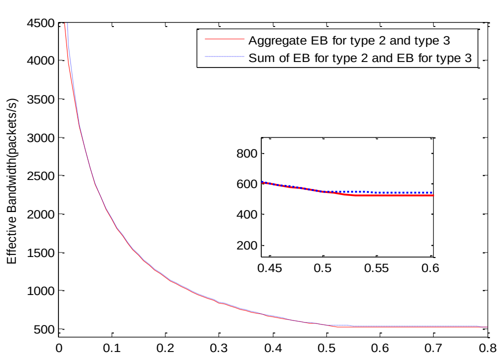
when the delay constraint D is sufficiently large. EB for the aggregate flow is equal to the sum of the EBs of all individual flows.



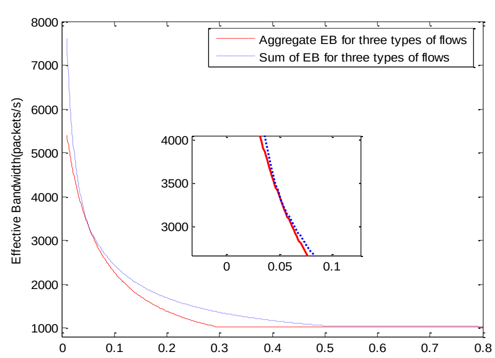
(a) *D* vs. EB for *i*=1 and *i*=2



(b) *D* vs. EB for *i*=1 and *i*=3



(c) *D* vs. EB for *i*=2 and *i*=3



(d) *D* vs. EB for *i*=1, *i*=2 and *i*=3

Figure 4. Characteristic of Effective Bandwidth

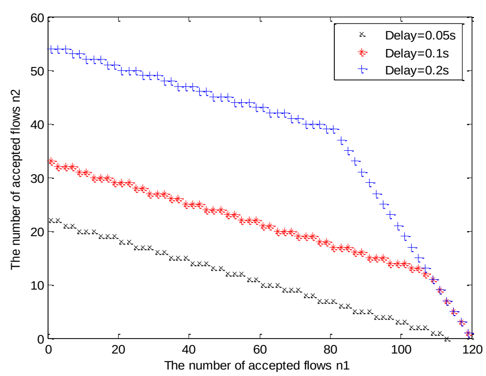
EB increases as the delay constraint *D* decrease. When *D* approaches 0, EB approaches infinity.

6.2. Performance Analysis of Admission Control Method

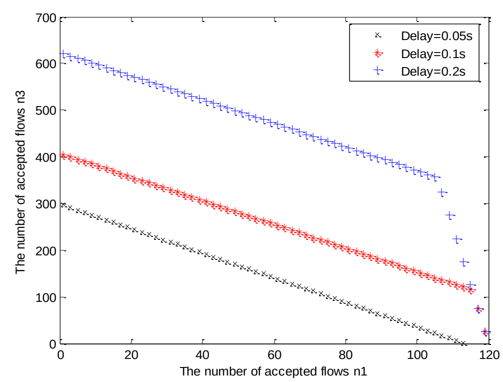
Figure 5 depicts the acceptance performance for the admission control method EBAC when the output rate of the ingress node is *60000 packets/s.*

It shows that a longer delay constraint can give rise to larger acceptance ability for flows, the bigger delay constraint, as presented previously, causes a smaller EB. So the node can accept more cloud services when the EB is smaller. By combing the computations in (7) with the admission control criteria in (8), we can derive the fact that the number of accepted flows $n_i(i=1,2,3)$ are linearly related under a certain delay. Specifically, (7) show that the values of EB for aggregate flow depend not only on the delay constraints, but on the number of accepted flows $n_i(i=1,2,3)$ as well. Changes on values of EB directly affect the truth values of (8), and therefore cause curves or surfaces to bend.

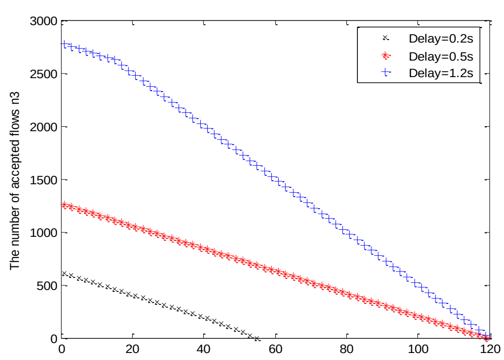
Figures 5(a), 5(b), and 5(c) demonstrate the acceptance performance for the case of *I*=2. Figures 5(d) depicts the acceptance performance for the case of *I*=3. Particularly, as shown in Figure 5(a), since the EB for flows of type 1 is less than or equal to that of type 2, the maximum number of acceptable flows of type 1 when being accepted alone (*i.e.*, n_1 when n_2=0) is more than that of type 2 (*i.e.*, n_2 when n_1=0), with respect to the delay constraint *D* specified in the figure. Figure 5(b), 5(c) and 5(d) conveys the analogous result to that of Figure 5(a).



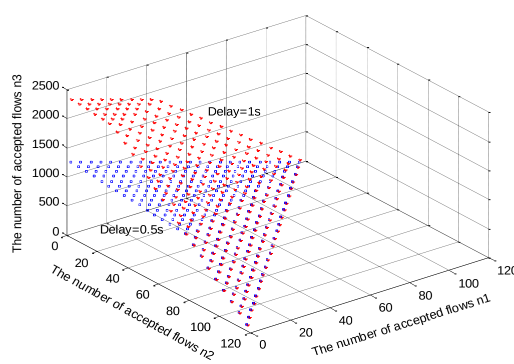
(a) n_1 vs. n_2 for *I*=2



(b) n_1 vs. n_3 for *I*=2



(c) n_2 vs. n_3 for *I*=2



(d) n_1, n_2 vs. n_3 for *I*=3

Figure 5. The Performance of Admission Control Methods for *C=60000packets/s*

It may appear that the maximum number n_1 is less than or equal to n_2 for other delay constraint D. Note that these figures only depict experiment result with some particular parameter values; in the actual experiment the selection of parameter values is arbitrary.

7. Conclusions

As cloud computing becomes increasingly popular, the issue of quality of service for cloud computing also becomes critical. Toward a guarantee for quality of service in cloud computing, a flow admission control model was proposed in this paper. Considering that cloud services are typically concurrent and for the purpose of improving these concurrent requests, we investigated the computations of Effective Bandwidth for aggregate flow by means of Network Calculus, and proposed admission control method EBAC that are based on the computation. Numerical experiments are conducted to investigate properties of effective bandwidth, and to evaluate the performance of the admission control method EBAC under different settings.

Acknowledgements

This work is supported by the program New Century Excellent Talents in University and the following grants: National Science Foundation of China (Grant No. 61272400, 61309031), Chongqing Innovative Team Fund for College Development Project (Grant No. KJTD201310), Natural Science Foundation of Chongqing (Grant No. cstc2013jcyjA40026), The Research Project of Chongqing Education Committee (Grant No. KJ130523), and CQUPT Research Fund for Young Scholars (Grant No. A2012-79).

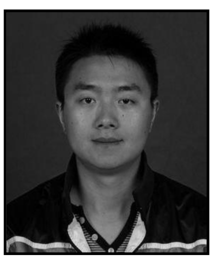
References

[1] M. Li, Y. Liu and Y. Xiao, “Services selection strategy based on user behavior in cloud computing”, Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), vol. 25, no. 5, **(2013)**, pp. 639-643.
[2] S. Shi and Y. Liu, “Cloud computing task scheduling research based on dynamic programming”, Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), vol. 24, no. 6, **(2012)**, pp. 687-692.
[3] Y. Chen, H. Wang, G. Fen and Y. Gao, “Multiscale admission control algorithm for statistical QoS guarantee”, Journal on Communications, vol. 32, no. 11, **(2011)**, pp. 140-146.
[4] C.-S. Chang, “Performance guarantees in communication networks”, Springer-Verlag, **(2000)**.
[5] R. Cruz, “A calculus for network delay. part I. network elements in isolation”, IEEE Transactions on Information Theory, vol. 37, no. 1, **(1991)**, pp. 114-131.
[6] R. Cruz, “A calculus for network delay, part II: Network analysis”, IEEE Transactions on Information Theory, vol. 37, no. 1, **(1991)**, pp. 132-144.
[7] A. Parekh and R. Gallager, “A generalized processor sharing approach to flow control in integrated services networks: The single node case”, IEEE/ACM Transactions on Networking (TON), vol. 1, no. 3, **(1993)**, pp. 344-357.
[8] R. Agrawal, R. Cruz, C. Okino and R. Rajan, “Performance bondsfor flow control protocols”, IEEE/ACM Trans. Netw., vol. 7, no. 3, **(1999)**, pp. 310-323.
[9] Y. Jiang and Y. Liu, “Stochastic network calculus”, Springer Verlag, London, **(2008)**.
[10] J.-Y. Le Boudec, “Application of network calculus to guaranteed service networks”, IEEE Transactions on Information Theory, vol. 44, no. 3, **(1998)**, pp. 1087-1096.
[11] F. Ciucu and J. Schmitt, “Perspectives on network calculus-no free lunch, but still good value”, Proceedings of the ACM SIGCOMM, **(2012)**, pp. 311-322.
[12] A. Bouillard and E. Thierry, “An algorithmic toolbox for Network Calculus”, Discrete Event Dynamic Systems, vol. 18, no. 1, **(2008)**, pp. 3-49.
[13] L. Zhang, D. Huang and Z. CHEN, “Models of bounds on end-to-end delay of long-range dependence traffic based on fractal leaky buckets”, Journal on Communications, vol. 29, no. 7, **(2008)**, pp. 32-38.

[14] L. Bisti, L. Lenzini and E. Mingozzi, "Numerical analysis of worst-case end-to-end delay bounds in FIFO tandem networks", Real-Time Systems, vol. 48, no. 5, **(2012)**, pp. 527-569.
[15] A. Bouillard, L. Jouhet and E. Thierry, "Tight performance bounds in the worst-case analysis of feed-forward networks", Proceedings of the IEEE INFOCOM, **(2010)**, pp. 1-9.
[16] T. Shan and O. Yang, "An admission control scheme for the real-time VBR traffic in the ATM network: deterministic bandwidth allocation", Proceedings of the IEEE International Conference on Communication (ICC), **(1998)**, pp. 1505-1509.
[17] J. Schmitt, F. Zdarsky and M. Fidler, "Delay bounds under arbitrary multiplexing: when Network Calculus leaves you in the lurch", Proceedings of the IEEE INFOCOM, **(2008)**, pp. 1669-1677.
[18] A. Ashraf, F. Jokhio, T. Deneke, S. Lafond, I. Porres and J. Lilius, "Stream-based admission control and scheduling for video transcoding in cloud computing", Cluster, Cloud and Grid Computing (CCGrid), **(2013)**, pp. 482-489.
[19] J.-Y. Le Boudec and P. Thiran, "Network calculus: a theory of deterministic queuing systems for the internet", Springer, **(2001)**.
[20] W. Ji-yi, F. Jian-qing, P. Ling-di and X. Qi, "Study on the P2P Cloud Storage System", Acta Electronica Sinica, vol. 39, no. 5, **(2011)**, pp. 1100-1107.

Authors

Yanbing Liu, he works as a Professor and Ph. D. supervisor of Chongqing University of Posts and Telecommunications. He received an M.S. degree in computer application from Beijing University of Posts and Telecommunications. He received the Ph.D. degree at the University of Electronic Science and Technology of China. His current research interests include cloud computing and network security.



Yunlong He, he received his B.S. degree in Electronic Information Science and Technology from the Xinzhou Teachers University, China in 2011. Currently he is studying for his master degree in Communication and Information System at the Chongqing University of Posts and Telecommunications. His research interest is admission control in cloud computing.

Jun Huang, he is an Associate Professor of Chongqing University of Posts and Telecommunications. He received an M.S. degree in Computer Science and Technology from Chongqing University of Posts and Telecommunications and the Ph.D. degree at the Beijing University of Posts and Telecommunications. His current research interests include network routing optimization and the QoS of the Future Network.

