Network-Adaptive QoS Control for Relative Service Differentiation-Aware Video Streaming

Gooyoun Hwang[1], Jitae Shin[2], and JongWon Kim[1]

[1] Networked Media Lab., Department of Information and Communications,
Gwangju Institute of Science and Technology (GIST), Gwangju, 500-712, Korea
{gyhwang,jongwon}@netmedia.gist.ac.kr
[2] School of Information and Communication Engineering,
Sungkyunkwan Univ., Suwon, 440-746, Korea
jtshin@ece.skku.ac.kr

Abstract. Emerging networked multimedia applications in the Internet require special supports from the underlying network such as real-time treatment, guaranteed service, and different levels of network service quality. However, the current Internet having the same-service-for-all paradigm is not suitable for these multimedia communications. In this paper, firstly we introduce a scalable and adaptive quality of service (QoS) mapping framework over the differentiated services (DiffServ, or DS) network. The framework is composed of the functionalities of proactive and reactive QoS mapping controls to provide reliable and consistent end-to-end service guarantee. On this framework, we propose a network-adaptive QoS control to have a feedback of instantaneous network fluctuation. The main idea of our proposal is to employ explicit congestion notification (ECN) mechanism in conjunction with the proactive QoS mapping control at the ingress of a DiffServ domain. It is possible not only that the status of network classes is notified to the end-host applications but also that a reactive QoS adjustment is triggered at the ingress side. Our simulation results illustrate how to enhance the QoS performance of streaming video in the under-provisioned network.

1 Introduction

Delivering networked multimedia services over the Internet, such as IP telephony, video-conferencing, and online TV broadcast, demand very stringent level of quality of service (QoS) guarantee. However, the current IP Internet is based on a simple service commitment, so called best-effort (BE) service, thus lacking the ability to support the QoS requirement from the multimedia applications. The differentiated services (DiffServ, or DS) model specified by the Internet Engineering Task Force (IETF) has been proposed as a more scalable and manageable architecture for network QoS provisioning [1]. In this model, resources are allocated differently not for individual packet flows but for aggregated traffic flows based on a set of bits (i.e., DSCP) and two other service models, premium service (PS) and assured service (AS), are currently defined to provide different

level of forwarding assurance for IP packets at each node [1, 3, 4]. The PS is designed to provide a low-loss, low-latency, and assured bandwidth service (i.e., absolute service differentiation) [4]. It requires an admission control to prevent resource starvation of other service classes and thus trades off the flexibility for more guarantees. In contrast, the assured service provides a relative service differentiation among DS classes, in the sense that high-priority classes receives a better (or at least not worse) service than low-priority ones [3]. Since most of recent multimedia applications, which include streaming video as a special example, have become more and more resilient to occasional packet loss and delay fluctuation, the assured service of DiffServ networks seems a more attractive choice due to its simplicity and flexibility.

However, since the assured service only provides qualitative differentiation between a number of classes, service guarantee is limited. Recently, some research works have tried to strengthen the service guarantees that can be provided within the context of the relative service differentiation without sacrificing its scalability and simplicity. Probably the best-known approach is the proportional differentiated services (PDS) model [6, 8], which quantifies the difference in the service by making the ratios of delays or loss rates of different classes roughly constant. This model can allow the network operator to control the quality spacing between a number of classes independent of the class load variations. Thus, we can focus on the PDS model for relative service differentiation-aware video streaming.

Within the DiffServ architecture, various types of video delivery scenarios have been proposed to integrate applications and networks for enhanced media streaming [5, 7, 10]. One possible scenario is to adaptively combine application requirements into appropriate network resource parameters such as end-to-end delay, jitter, packet loss, and bandwidth. Accordingly, an adaptive QoS mapping[3] mechanism depending on the network situation is required to support the end-to-end QoS guarantee dynamically to end systems and improve the network utilization. From this point of view, we presents a scalable and adaptive QoS mapping-control (shortly, control) framework, which consists of proactive QoS control and reactive QoS control in network class/flow-based granularity, over the DiffServ domain for service differentiation-aware video streaming. The QoS control framework is strongly coupled with a pricing or policy infrastructure of the corresponding domain to make higher classes more costly than lower ones. Given a certain cost constraint, the users/applications are supposed to find an optimal way based on the contracted QoS level.

Meanwhile, unstable network service situations caused by instantaneous class load fluctuations still occur even though the resource provisioning policy of underlying network is strictly managed. The unstable situations obstruct video applications to achieve their desired service requirements persistently. Since the end-to-end QoS perceived by applications mainly depend on the current net-

---

[3] The issue of *QoS mapping* occurs when we map prioritized and classified groups of some applications(or users or flows) based on their importance into different DS levels or network classes.

work load condition, video applications should response to this fluctuation in a proper way. Furthermore, the adjustment of proactive QoS control is needed when a video receiver is not satisfied with the received quality compared to its expectation. For these reasons, a network-adaptive QoS control is proposed to enhance relative service differentiation-aware video streaming in this paper. With the help of network feedback, the end-host video application can recognize the status of network classes and can now react in advance. Basically, the provided feedback information is incorporating latest status of underlying routers. By leveraging the end-to-end feedback (i.e., in tie with the required congestion control), the feedback is relayed to the sender in a similar way as the explicit congestion notification (ECN) [2]. The network feedback-based QoS control triggers the QoS mapping adjustment at the ingress of DiffServ networks. NS 2-based network simulation is performed to demonstrate the enhanced performance of the proposed QoS mapping control framework (i.e., more efficient and adaptive to network variation).

The remainder of this paper is organized as follows. Section 2 presents a scalable and adaptive QoS mapping framework as our future direction. In section 3, we propose the network-adaptive control to enhance relative differentiation-aware video streaming. Various sets of performance evaluation through computer simulations are presented in section 4. Finally, section 5 concludes this paper.

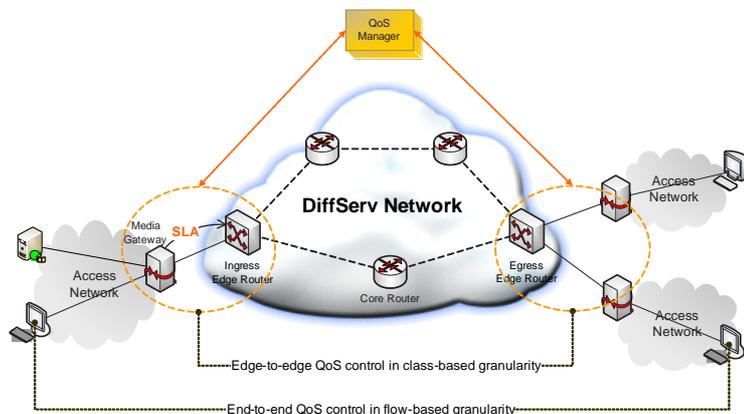## 2 Scalable and Adaptive QoS Mapping Framework over the DiffServ Network

Fig. 1. Overview of scalable and adaptive QoS mapping framework.

To provide reliable and effective end-to-end video streaming, this section presents a scalable and adaptive QoS mapping framework over the DiffServ

networks. As illustrated in Fig. 1, the proposed QoS mapping framework can be divided into three QoS controls : (1) proactive QoS control for aggregated flows at the ingress point of the network, (2) reactive edge-to-edge QoS control between the borders of the network (e.g., edge router and/or media gateway), and (3) reactive end-to-end QoS control. The former two controls are based on network-class QoS, while the last reactive control is based on flow QoS as a complementary role and fine-tuned for end-to-end QoS provisioning. A key assumption in this architecture is that if a user (or application) has absolute QoS requirements, he/she has to dynamically choose a class in which the observed QoS is acceptable. If the user is also interested in minimizing the cost of the session, he would choose the least expensive or minimum class that is acceptable.

For the end-to-end video streaming, sources within the access network send videos to the corresponding clients. The access network subscribes to DS services, and traffic is delivered to the clients through the DS domain. The access network has service level agreements (SLAs) [1] with the DS domain. In this framework, the video applications at the source assign a relative priority-based index (RPI) to each packet in terms of loss probability and delay as studied in [7] so that each packet can reflect its influence to the end-to-end video quality. The source then furnishes each packet with this prioritization information for a special edge node called media gateway (MG)[4] as shown in Fig. 2. Thus, the video streams from the sources are merged at the MG. In order to prevent the sources from violating their SLAs and protect resources from a selfish source, the MG exercises the traffic shaping on a per-flow basis through the token buckets (TBs) assigned for individual flows, as seen in Fig. 2. The packets violating this agreement are assigned with the lowest DS class (i.e., best-effort class).
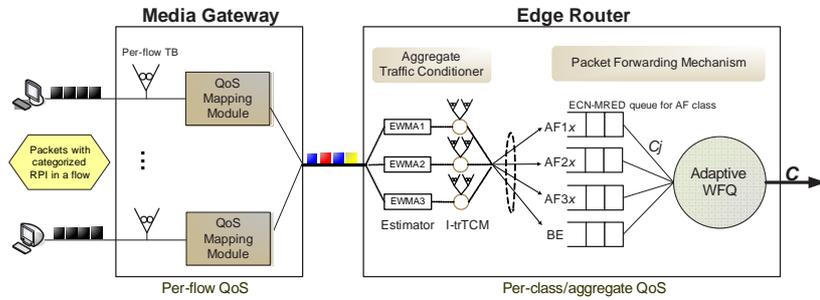

Fig. 2. Overall structures of media gateway and edge router.

The main function of the MG is to make a cost-efficient coordination between the prioritized packets (or flows) and the DS service classes, which we call *QoS*

---

[4] The MG can be existed as an independent entity or combined with the edge router (ER) if necessary. The deployment issue of the MG is not fully investigated in this work yet.

mapping. For the optimal QoS mapping of the relative prioritized packet on to the DS level, we just refer [7] and mention briefly due to page limitation. That is, for packets conforming to the TB, the MG assigns to each packet a DiffServ codepoint (DSCP) on the basis of the packet's RPI. Then, the MG forwards the packet streams to the edge router (ER) at the ingress of the DS network. The ER is composed of an aggregate traffic conditioner (ATC) and a packet forwarding mechanism [9]. The ATC is employed to observe the traffic conditioning agreement (TCA) with the DS domain and the packet forward mechanism provides proportionally relative QoS spacing between network service classes by using a joint buffer management (e.g., multiple random early detection (RED)) and a dynamic scheduler (e.g., adaptive WFQ). In the QoS mapping framework, the proactive QoS control is realized as discussed.

However, the network QoS of a DS domain may be time-varying due to traffic load fluctuation. Thus, the egress MG monitors the QoS performance of each class and notifies the ingress MG about the observed QoS level through a feedback channel or a QoS manager. Based on this feedback, the ingress MG decides whether to stay in the same class or switch to the higher or lower class in order to satisfy the specified QoS performance of a class in the SLA. In this paper, we put aside the reactive edge-to-edge QoS control out of scope (but under our further investigation). Finally, a proper network-aware feedback control (i.e., reactive flow-based QoS control), which can give the applications a guidance of network status, enables the fine-tuned refinement on top of coarse proactive QoS control. For this, a network-adaptive QoS control will be proposed in next section.

3 Proposed Network-Adaptive QoS Control

The intermediate routers in the DS domain have a better understanding of dynamic network fluctuation. Thus, in our opinion, a proper combination of congestion signalling from network and source reaction will be an effective solution to the instantaneous network fluctuation in the Internet. A feedback mechanism encompassing both network and end-systems can contribute to enhance the performance of video applications. Explicit congestion notification (ECN) for IP routers [2] has been proposed as a mean of giving more direct feedbacks of congestion information back to end-systems. By marking packets, the ECN does not waste the network bandwidth used to forward packets and detects incipient congestion in the network. Many researches have accomplished active queue management, e.g., RED or RED variants, with and without ECN support even in TCP networks. It is also possible for media applications using UDP protocol to react to ECN marks. This motivates us to design a network feedback-based QoS control with the extension of ECN. The major idea of our feedback control is employing ECN mechanism in conjunction with the proactive QoS control at the ingress of a DiffServ domain. It is possible that not only the congestion status of network class is notified to the end-host video applications but also the reactive QoS mapping control is triggered in a faster manner.

Fig. 3 shows the outline of proposed network-adaptive QoS control on top of the scalable and adaptive QoS mapping framework. Routers in the DS domain are assumed to be ECN-enabled and equipped with multiple RED (MRED) queues. The CE (congestion experienced) bit on incoming packets is set when a DS class queue enters into an unstable state.
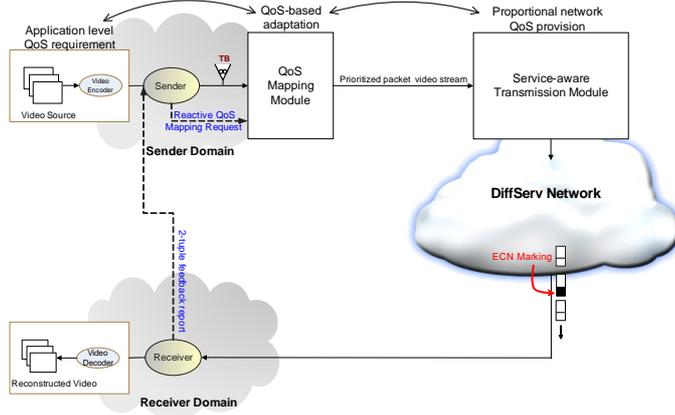

Fig. 3. Overview of network-adaptive QoS control.

On the other hand, the ECN-aware receiver monitors the CE bit on each arrived i-th packet of a flow and calculates the received average cost of a flow, $\overline{C}_{recv}^f$, using the following equation.

$$\overline{C}_{recv}^f = \frac{\sum_{i=1}^{N^f} C_q(i)}{N^f}, \tag{1}$$

where N^f is total received packet numbers of flow f and $C_q(i)$ is unit cost of DS level q, which i-th packet is mapped to, predefined by the SLA. This observed value is considered as a barometer to interpret the degree of service degradation which the receiver experiences. When a ECN-marked packet is arrived, the receiver immediately sends a report of its DS level q and the average cost $\overline{C}_{recv}^f$ to the MG through the corresponding sender. Hence, the MG examines the feedback information and triggers the reactive QoS mapping control. The MG considers the DS level q of the report as a congested class and compares the received average cost $\overline{C}_{recv}^f$ with the requested average payment $\overline{C}_{send}^f$[5]. Then, on the basis of the feedback information, the categorized packet k belonging to the congested class is reassigned to non-congested classes appropriately. The overall algorithm of this feedback control is summarized in the box.

---

[5] $\overline{C}_{send}^f$ is computed by the equation (1) when video packets of a flow are initially sent.

ECN marking at the routers

1. When a DS class queue of ECN-MRED router experiences congestion, the router sets the CE bit to indicate the onset of congestion to the end nodes.

Receiving end-system

1. Upon the receipt of a ECN-marked packet, the receiver checks its DS class $\tilde{q}$ and computes $\overline{C}_{recv}^f$ at that time.

2. Send a 2-tuple $\{\tilde{q}, \overline{C}_{recv}^f\}$ feedback information to the MG.

Reaction of the MG for reactive QoS control

1. Upon the receipt of a feedback report, the MG regards $\tilde{q}$ as a congested class and compares $\overline{C}_{recv}^f$ with $\overline{C}_{send}^f$.

2. The MG adjusts the mapping of source category k to DS level q based on the above comparison. That is, k involved in the expected congestion class is re-mapped to higher non-congested classes, for $\overline{C}_{recv}^f \leq \overline{C}_{send}^f$. Otherwise, k is reallocated to lower non-congested ones.

   – `DS level:` Let q be a DS level, where $1 \leq q \leq Q$ with the increasing order of network service quality and Q is the total number of DS levels.

   – `RPI partitioning :` Let $R_q^k(i)$ be a partition i among the RPI k category and be assigned into DS level q. Each k category has equal number of packets initially and is sorted in an increasing order, that is, $1 \leq k \leq K$, where K is the category with highest RPI values. Generally, the packets within the same k could be assigned into different q levels.

   – `Proactive mapping:` Each packet, whose RPI is k ($k \in R_q$), is mapped to q in order to meet the requested $\overline{C}_{send}^f$.

   – `Reactive mapping:` When a congestion feedback, i.e., ECN, from a class $\tilde{q}$ is received, $R_q^k(i)$ is distributed into $S_q(t)$ subset, where $S_q(t)$ is the number of non-congested DS levels at time t and $1 \leq j \leq S_q(t)$, which are higher levels than the level q for $\overline{C}_{recv}^f \leq \overline{C}_{send}^f$. Then, the packets belonged to $R_q^k(i)$ are re-mapped to DS level j.

3. This reactive control is operated in time of $[t_{ACK}, t_{ACK} + \Delta]$, where t_{ACK} and Δ denote the times of receiving feedback packet and a certain time interval, respectively.

   – `Receiving a congestion feedback:` The MG sets Δ to 0.1 (sec) and performs the reactive mapping during Δ.

   – `Monitoring congestion:` After the expiration of Δ, the MG returns back to the normal state and observes a congestion feedback during $2RTT$.

   – `Increasing` Δ: If a congestion feedback is received again within $2RTT$, the MG increases $\Delta = 2\Delta$ and goes to the reactive control state. Otherwise, $\Delta = 0.1$.

4 Simulation Results

In this section, we will demonstrate the performance evaluation to compare the two types of QoS mapping control, i.e., the proactive class-based QoS control (PQ-only) and our network feedback-based QoS control (PQ-NBF, that is PQ-

only plus reactive flow-based QoS control). The overall simulation setup is illustrated in Fig. 4. The standard-based H.263+ encoding/decoding and the NS-2 network simulator are used to evaluate the end-to-end video performance.

While H.263+ encoder encodes video, mean-square-error (MSE) value of each group of block (GOB) is calculated and stored as a data file. Since each GOB is packetized into a separate packet in the simulation, priority will be assigned to each packet according to the relative loss importance of payload. Error patterns generated from results by the NS-2 simulations are used to decide whether the packet is lost or not. Then, at the receiver side, encoded bitstream is decoded with the error pattern file. Consequently, peak signal to noise ratio (PSNR) between original and reconstructed video is calculated to quantify how much video quality is degraded by the packet loss during transmission. The PSNR is computed as

$$PSNR_{n_{th}} = 10 \cdot log_{10} \frac{255^2}{\sum_{i \in frame} MSE_{(n,i)}} \tag{2}$$

where

$$MSE_{(n,i)} = \frac{1}{N} \sum_{n \in video} |\hat{R}_n^i(x,y) - R_n(x,y)|^2 \tag{3}$$

$MSE_{(n,i)}$ is the mean square error of n_{th} frame when i_{th} packet is lost. Using (x,y) as a 2-D coordinate of pixel in a frame, $\hat{R}_n^i(x,y)$ and $R_n(x,y)$ are the reconstructed n_{th} frames when i_{th} packet is lost and kept, respectively.
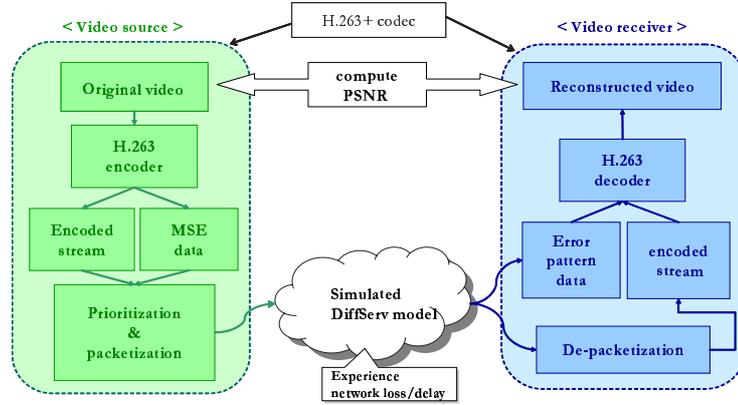

Fig. 4. Overall simulation diagram for H.263+ streaming video over a simulated DiffServ network.

Fig. 5(a) shows the simplified simulation topology model, which is used to generate underlying network dynamics. One test video source and several background traffic sources are connected to the DS domain through the MG and communicate with one of different destination nodes. The link between the ER

and the core router(CR) is the bottleneck link with a capacity of 3Mbps, where the video flow competes with other background UDP flows. In order to support the relative service differentiation, we assume that the DS domain provides three assured forwarding (AF) classes and that each class queue has three drop precedences [3]. That is, the DS level order from high to low is $\{AF_{11}, AF_{12}, AF_{13}; AF_{21}, AF_{22}, AF_{23}; AF_{31}, AF_{32}, AF_{33}; BE\}$. On the other hand, a video source (S_0) in Fig. 5(a) transmits a video stream from the video trace (*Foreman* sequence) with RPI. In the simulations, we use a reference QoS mapping for the video flow as follows: $k=0\sim1 \rightarrow BE$, $k=2\sim5\rightarrow AF_{32}$, $k=6\sim9\rightarrow AF_{31}$, $k=10\sim13\rightarrow AF_{22}$, $k=14\sim16\rightarrow AF_{21}$, $k=17\sim18\rightarrow AF_{12}$, and $k=19\rightarrow AF_{11}$, respectively. In order to get a fair comparison, we do not differently use the reference QoS mapping between categorized packets and DS levels.

As described in Section 2, the ER is responsible for realizing the specialized traffic managements. An extended version of two-rate three-color-maker (trTCM), so called inter-connected trTCM, is adopted to monitor and mark all incoming traffics based on the TCA of SLA. The inter-connected trTCM measures the aggregated ingress rate for each class j. If the aggregated rate reaches the maximum assigned rate C_j for each class queue as presented in Fig. 5(b), the incoming packets into this queue are handed over randomly. Please refer [9] for the detailed description of the inter-connected trTCM. Queueing with MRED and WFQ, also shown in Fig. 5(b), is applied for the drop precedence of queue management and link sharing of scheduler. For our simulations, the routers in the DS domain use the MRED with the values of [50, 70, 0.01] for AF_{x1}, [30, 50, 0.02] for AF_{x2}, and [10, 30, 0.1] for AF_{x3}. Note that the notation [x, y, z] represents minimum threshold, maximum threshold, and maximum drop probability, respectively, of the RED queue.

(a) Simulation topology   (b) Inter-connected trTCM
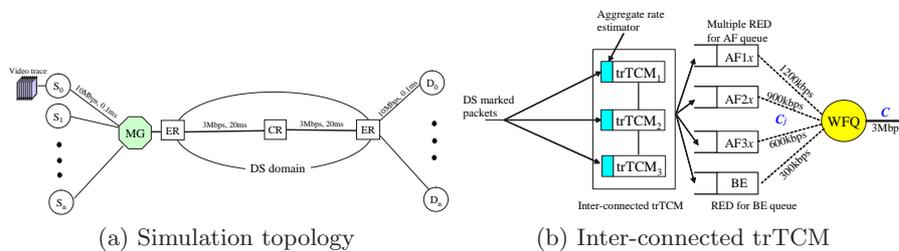
**Fig. 5.** Simulation topology model and inter-connected trTCM.

The main objective of the experiments is to investigate the effectiveness of our network feedback-based control. With this network-application feedback control mechanism, it is expected that the video applications can manage their flows more effectively, thereby achieving an enhanced QoS performance. In order to verify this expectation, each QoS control is simulated in different runs under the same network load conditions. To make a severe congestion period, the sending rates of AF_{2x} sources and AF_{3x} sources are adjusted in runtime from 10 to 20

(sec) alternately so that it corresponds to total provision level of 110%. For the sake of simplicity, the range of unit cost per packet in PQ-NBF is 9 to 0 and is associated with the same order of DS level.

Table 1. The end-to-end performance results of the QoS mapping controls.

QoS mapping controls	End-to-end QoS Parameters			
	Achieved throughput (kbps)	Average loss rate (%)	Average delay (msec)	Average PSNR (dB)
PQ-only	391.446	7.672	73.608	29.797
PQ-NBF	406.518	1.387	61.727	33.378

Table 1 presents the end-to-end performance results of two QoS controls, i.e., PQ-only and PQ-NBF, in terms of achieved throughput, average loss rate, and average delay. It represents that the proposed feedback control effectively responses to the network congestion and has competitive advantage for video streaming. Based on the performance results of the achieved throughput and average packet loss rate, the PQ-NBF can achieve better resource utilization than the PQ-only. Furthermore, the delay and jitter performance can be improved significantly under the PQ-NBF, as shown in Fig. 6. This result is quite desirable because video clients feel much annoyed when the video packet does not arrive on time. Recall that the congestion period in the simulation starts at run time 10 sec. Additionally, the PQ-only cannot achieve better performance than the PQ-NBF. It means that a suitable feedback mechanism can assist video applications to adapt dynamically to underlying network and to stabilize the end-to-end QoS within an acceptable range.


(a) Delay (b) Jitter
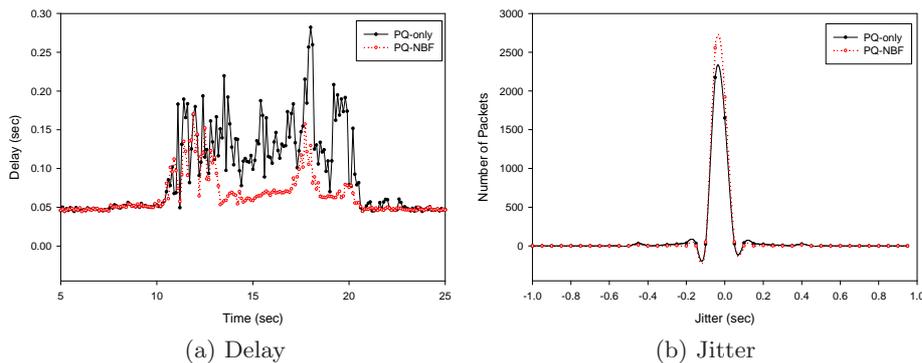
Fig. 6. Performance comparison of the controls in terms of average delay and jitter.

Finally, to examine the perceptual quality of the H.263+ video, we play out the decoded video sequence at the receiver and measure the PSNR as an objective quality metric, which calculates the difference between the original source video sequence and the received video sequence. Note that the average original PSNR for the video trace is about 34.76 dB (i.e., ideal case). Fig. 7(a) and Fig. 7(b) present PSNR performance comparison of different QoS controls and achieved average PNSR at different traffic loads . The objective PSNR quality measure of the PQ-NBF is better than that of the PQ-only over various under-provisioned situations. Also, the corresponding snap shots of a decoded video frame are shown in Fig. 8(a) and Fig. 8(b). The visual quality under the PQ-NBF can outperform in the average or instance sense over time-varying network load conditions.

(a) PSNR performance comparison

(b) Achieved average PSNR
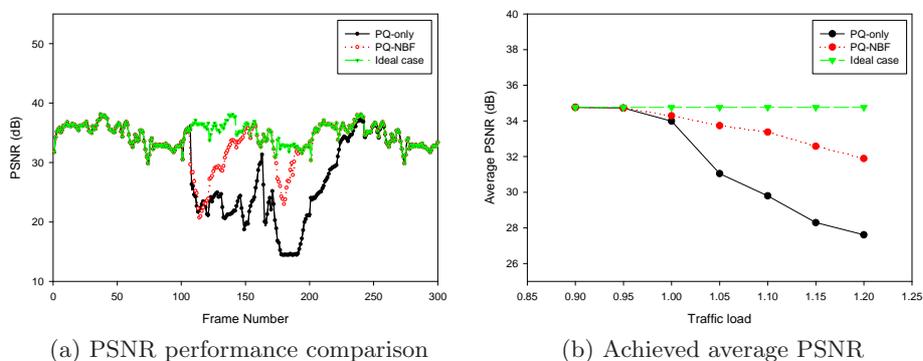
Fig. 7. PSNR performance comparison at 110% network load level and achieved average PSNR at different traffic loads.

(a) PQ-only

(b) PQ-NBF
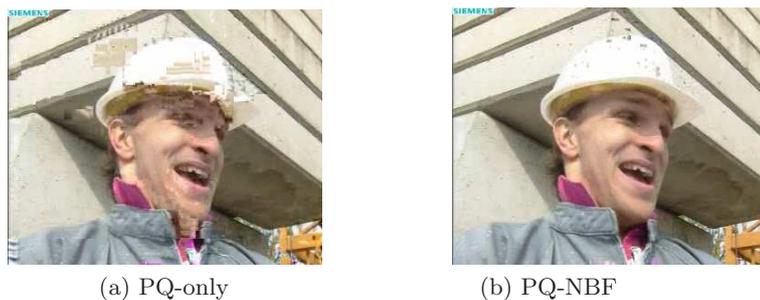
Fig. 8. Video quality comparison of a sample frame of the *Foreman* sequence at 110% network load level.

5   Conclusion

In order to provide reliable and consistent end-to-end service guarantee, we present a scalable and adaptive QoS mapping framework over the DiffServ networks. The framework is composed of the functionalities of proactive QoS control and reactive QoS control. We investigate mainly network-adaptive reactive QoS control to obtain the QoS enhancement of a video streaming by proposing a network feedback mechanism on top of proactive class-based QoS control. The key point of this feedback control is interacting between network and end systems using the ECN mechanism. Through simulation experiments with video trace, we show that the proposed QoS control improves the quality of video streaming significantly in a DiffServ network. Future work would include the reactive edge-to-edge QoS control between network border entities to take care of the end-to-end video streaming over multiple DiffServ domains.

Acknowledgment

References

1. S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services,", IETF RFC 2475, Dec. 1998.
2. K. K. Ramakrishnan and S. Floyd, "A proposal to add explicit congestion notification (ECN) to IP,", IETF RFC 2481, Jan. 1999.
3. J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski, "Assured forwarding PHB group,", IETF RFC 2597, June 1999.
4. K. Nichols, V. Jacobson, and L. Zhang, "A two-bit differentiated servicesarchitecture for the Internet,", IETF RFC 2638, July 1999.
5. Y. T. Hou, D. Wu, B. Li, T. Hamada, I. Ahmad, and H. J. Chao, "A differentiated services architecture for multimedia streaming in next generation Internet," *Computer Networks*, vol. 32, pp. 185-209, Feb. 2000.
6. C. Dovrolis, D. Stiliadis, and Parmesh Ramanathan, "Proportional Differentiated Services, Part II: Loss Rate Differentiation and Packet Dropping," in *Proc. of International Workshop on Quality of Service (IWQoS)*, Pittsburgh PA, June 2000.
7. J. Shin, J. Kim, and C.-C. J. Kuo, "Quality-of-Service mapping mechanism for packet video in differentiated services network," *IEEE Transaction on Multimedia*, vol. 3, no. 2, pp. 219-231, June 2001.
8. C. Dovrolis, D. Stiliadis, and P. Ramanathan, "Proportional differentiated services: delay differentiation and packet scheduling," *IEEE/ACM Transactions on Networking*, vol.10, no. 1, pp. 12-26, Feb. 2002.
9. J. Shin, "An analysis of aggregated traffic marking for multi-service networks," *IEICE Transactions on communications*, vol. E86-B, no.2, pp. 682-689, Feb. 2003.
10. S-R. Kang, Y. Zhang, M. Dai, and D. Loguinov, "Multi-layer active queue management and congestion control for scalable video streaming," in *Proc. IEEE ICDCS*, Mar. 2004.