The Knowledge Engineering Review, Vol. 29:4, 484–495. © Cambridge University Press, 2014
doi:10.1017/S0269888914000186

QoS-based active dropping mechanism for NGN video streaming optimization

CHI-YUAN CHEN[1], TIN-YU WU[2], WEI-TSONG LEE[3], HAN-CHIEH CHAO[2] and JEN-CHUN CHIANG[3]

[1]*Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan, ROC;*
e-mail: chiyuan.chen@gmail.com;
[2]*Department of Computer Science and Information Engineering, National Ilan University, Yilan, Taiwan, ROC;*
e-mail: tyw@niu.edu.tw, hcc@niu.edu.tw;
[3]*Department of Electrical Engineering, Tamkang University, Taipei, Taiwan, ROC;*
e-mail: wtlee@mail.tku.edu.tw, 491351192@s91.tku.edu.tw

Abstract

Video streaming over mobile wireless networks is getting popular in recent years. High video quality relies on large bandwidth provisioning, however, it decreases the number of supported users in wireless networks. Thus, effective bandwidth utilization becomes a crucial issue in wireless network as the bandwidth resource in wireless environment is precious and limited. The NGN quality of service mechanisms should be designed to reduce the impact of traffic burstiness on buffer management. For this reason, we propose an active dropping mechanism to deal with the effective bandwidth utilization in this paper. We use scalable video coding extension of H.264/AVC standard to provide different video quality for users of different levels. In the proposed dropping mechanism, when the network loading exceeds the threshold, the dropping mechanism starts to drop data of the enhancement layers for users of low service level. The dropping probability alters according to the change in network loading. With the dropping mechanism, the base station increases the system capability and users are able to obtain better service quality when the system is under heavy loading. We also design several methods to adjust the threshold value dynamically. By using the proposed mechanism, better quality can be provided when the network is in congestion.

1 Introduction

Next-generation network (NGN) is a concept of network integration. Through transmitting and receiving the packets, users are able to access the needed network services by NGN infrastructure. NGN basically has two characteristics: first, the service supply is not related to the transmitting technologies of the sub-layer network; and second, although NGN is integrated by multi-networks, users' mobility is unobstructed and unlimited. In NGN environment, the service supply and the management of the sub-layer network are separated. Nowadays, WiMAX technology provides an easy, time-saving, and low-cost method for the deployment of NGN.

The wireless broadband network technology has been developed and the demand of wireless network keeps increasing. The wireless network is extensively used from the personal wireless network IEEE802.11 system to the latest 3G telecommunication network. Various applications of wireless network system are adopted not only in data and voice transmission but also in video streaming. For the video streaming, we need wide bandwidth network such as IEEE 802.16 metropolitan area network (wireless MAN).

https://doi.org/10.1017/S0269888914000186 Published online by Cambridge University Press

WiMAX is the abbreviation of Worldwide Interoperability for Microwave Access that conforms to the 802.16 standard. In the 802.16e-2005 standard, the system provides users to access network with mobility (e.g. cell phone and PDA). According to users' requirements, WiMAX can provide different service flows. In this way, the flexible and effective bandwidth makes WiMAX a popular wireless broadband network system. WiMAX also defined five quality of service (QoS) service classes and designed to reduce the impact of traffic burstiness on buffer management.

On the other hand, video streaming needs wide bandwidth to transmit video data, but the bandwidth was limited in the wireless environment. For this reason, the extension of H.264/AVC scalable video coding (SVC) was designed to provide different video quality according to the received data. Zhang *et al.* (2007) uses the rate control algorithm to control the transmission rate and to reduce the client buffer size. Chuang *et al.* (2004) uses the adaptive media playout buffer control to mitigate the risks of buffer outage. Bajic *et al.* (2003) focuses on the congestion control and uses the buffer management to control loss of less-important data. Markopoulou and Han (2001) use the priority dropping mechanism to increase the PSNR of video quality. Nevertheless, these researches do not mention how to provide different video quality for users of different levels. When the video server provides users with the best video quality but not all users need it, the network system undoubtedly wastes lots of bandwidth.

In this paper, we have three major contributions. First, we adopt algorithms including earliest deadline first (EDF) and largest weighted delay first (LWDF) with SVC. Hence the priority sequence of the dropping packets is decided when the queue is in congestion and the effects on the video quality can be reduced to the minimum. Second, by using the proposed active dropping mechanism, the NGN system not only provides different video quality for users of different levels, but also reduces the waste of bandwidth and allows more users. Finally, our proposed mechanism is suitable for different NGN architectures and do not affect their security strength. Security is a very important issue for NGNs. The related security mechanisms are strictly defined in specifications (Chen *et al.*, 2008; Chang *et al.*, 2010). For this reason, we proposed our optimization mechanism based on the basic NGN QoS mechanism and did not modify any component or policy.

The rest of the paper is organized as follows. The related works and background is introduced in Section 2. The proposed scheme, active dropping mechanism mode, is elaborated in Section 3. The simulation results are illustrated in Section 4. Finally, the conclusion is given in Section 5.

2 Background and related works

In this section, we will introduce the QoS architecture in the NGNs, WiMAX's suggestion over QoS, and SVC briefly.

2.1 Quality of service of next-generation networks

In the NGN environment, the service supply and the management of the sub-layer network are separated. Users request NGN services simply and do not have to understand what kind of transmitting technology they use. In such a manner, the service management of NGN needs some mechanisms to choose a suitable network and request the chosen network for the needed QoS to achieve the requested end-to-end service quality. Thus, QoS of NGN is established upon the existing QoS mechanisms of the included networks. After receiving users' requested QoS, the core network needs to satisfy the end-to-end QoS requests. The responsibility of satisfying QoS is allotted to each network region of this connection and is asked to be completed. In other words, the requested end-to-end QoS is accomplished by job responsibility. After the centralization and allotment, QoS request will be fulfilled by the chosen network regions. This concept is illustrated in Figure 1.

From NGN standard architecture, we further analyze several functions of QoS, which is shown in Figure 2. Service control function (SCF) is located in the service stratum and the rest are located in the transport stratum. The detail of components are given below.

SCF located in the service stratum, which is responsible for controlling the service and transmitting the needed message of QoS to RACF (resource and admission control function) of the transport stratums.

https://doi.org/10.1017/S0269888914000186 Published online by Cambridge University Press

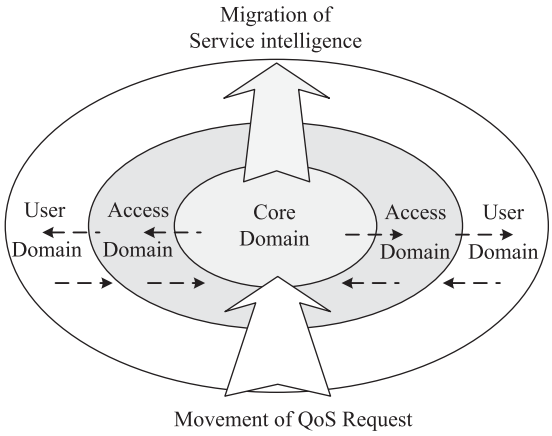


Figure 1 QoS concept of next-generation networks (NGNs)

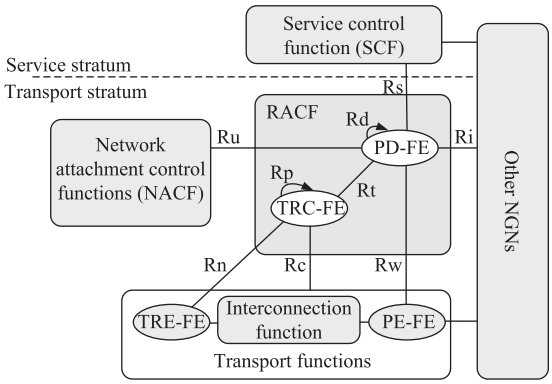


Figure 2 QoS management architecture of next-generation networks (NGNs) (ITU-T, 2006)

Network attachment control function provides users the membership identification of QoS service and makes sure the available maximum bandwidth for users while setting up the connection. RACF provides network resource and admission control. This function includes Policy Decision Functional Entity (PD-FE) and Transport Resource Control Functional Entity (TRC-FE). According to user profile, SLA, network operation policy, service priority information, and resource availability, PD-FE determines whether the network accepts the admission or not. When the connection is allowed to be established, the control message is delivered to ask the network to deploy the resource. Namely, PD-FE manages and controls the network resource by PE-FE (Policy Enforcement Functional Entity). Based on the characters of different network technologies, TRC-FE monitors the utilization of network resource to co-operate with PD-FE for resource admission control. As its management control function alters according to different technologies, TRC-FE is a resource control method related to the transport network technology and topology. On the contrary, PD-FE in the management architecture is not related to the transport network technology and topology. PE-FE is located mostly on the edge of the regional network, which is responsible for controlling the transport elements.

2.2 Overview of IEEE 802.16e network

The IEEE 802.16 MAN (IEEE 802.16-2004, 2004; IEEE 802.16e-2005, 2005) is developed in two structures. The first structure is a fixed broadband wireless access (BWA) system, called IEEE 802.16-2004. In 802.16-2004 system, the modulation technique in PHY layer is orthogonal frequency-division multiplexing (OFDM). Using the OFDM technique, 802.16-2004 system can support broadband access and long transmission distance, but the range restriction of each base station (BS) must be line-of-sight. The 802.16-2004 system is

https://doi.org/10.1017/S0269888914000186 Published online by Cambridge University Press

used to deal with the last mile problem and it is quick and low cost. The second structure is a mobile BWA system, called the IEEE 802.16e-2005. The IEEE 802.16e-2005 standard, extended from the 802.16-2004 standards, provides mobility support in cellular deployments. In this standard, the modulation technique is orthogonal frequency-division multiple access (OFDMA). Because of the OFDMA modulation in PHY layer and medium access control (MAC) layer, the 802.16e system can provide stable service for users. The range restriction in this system is non-line-of-sight (NLOS) and the advantage of NLOS is the signal with no serious fading when the signal is obstructed by buildings. As the 802.16e system allows users to move freely in the cities, the IEEE 802.16e-2005 system is suitable to use in metropolis because better service quality depends on better signal quality.

IEEE 802.16e provides different service flows according to different requirements and characteristics of services (Mai *et al.*, 2007; Chakchai *et al.*, 2009; Cheng *et al.*, 2010; Kim & Lim, 2010). WiMAX defines five QoS service classes: unsolicited grant scheme (UGS), extended real-time polling service (ertPS), real-time polling service (rtPS), non-real-time polling service (nrtPS), and best effort service (BE). The first is UGS, which is designed to support real-time service flows that will generate fixed-size data packets periodically. The service of ertPS is close to UGS in that the BS allocates the maximum sustained rate in active mode. This service is designed to support VoIP with silence suppression. The characteristic of rtPS is that the BS needs to poll each MS regularly to determine and allocate. This service class is for variable bit rate (VBR) real-time traffic (e.g. Moving Picture Experts Group (MPEG) compressed video). The feature of nrtPS is non-real-time VBR traffic with no delay guarantee with service classification for the minimum rate guaranteed. BE supports most of data traffic flow to allow specifying the minimum reserved traffic rate and the maximum sustained traffic rate even for this class.

The UGS, ertPS, and rtPS are real-time traffic. UGS has a static allocation. ertPS is a combination of UGS and rtPS. The UGS will be granted periodically without polling-request procedure. The second is real-time-variable rate service (RT-VR), which is designed to support real-time service flows, which will generate variable-sized data packets on a periodic basis. This service needs more request and latency than UGS, but it can support variable grant sizes. The third is extend real-time-variable rate service (ERT-VR), which is designed to support real-time service flows that generate variable-sized packets on a periodic basis. ERT-VR has the service characteristic of the UGS and RT-VR. The fourth is non-real-time-variable rate service, which is designed to support delay tolerant data streams that consist of variable-sized data packets. The last is BE that is designed to support data streams that have no minimum transmission delay. Each service flow has its unique service parameters that are used to scheduling and the transmission priority. In IEEE 802.16e, the MAC layer protocol is shown in Figure 3 (Huang *et al.*, 2007).

Each service includes one or multiple connections and each connection is identified by a 16-bit connection identifier. The convergence sub-layer classifies the service data units (SDUs) to a proper connection with specific QoS parameters. The common part sub-layer controls most functions such as fragmentation, packing, scheduling and retransmission, etc. By these service flows, IEEE 802.16e-2005 provides suitable service flows for different applications and systems that offer better service quality. For the video streaming service, network systems should provide real-time service and large variable-sized data packets on a periodic basis. The UGS has less latency than RT-VR but the UGS only provides a fixed data rate. Therefore, the RT-VR service flow is the most suitable for video streaming application.

The IEEE 802.16e provides wireless broadband access but the bandwidth resource in wireless networks is limited. Therefore, the network system should need a mechanism to drop the packet, which is less important or delayed from the transmitting queue. With this mechanism, the network system avoids unnecessary transmission and enhances the bandwidth utilization (She *et al.*, 2007).

2.3 Scalable video coding

Video streaming is a popular application in the wireless network systems. In the video processing, video encoder needs more effective compressing rate to reduce the streaming bit rate (Zhou *et al.*, 2009). The lower video bit rate makes more users gain access to the same wireless network system. SVC (Wiegand *et al.*, 2007) can improve the video streaming adaptation to unstable network environment. When the network is in congestion, the network system maintains video streaming quality with the low bit rate. SVC

https://doi.org/10.1017/S0269888914000186 Published online by Cambridge University Press

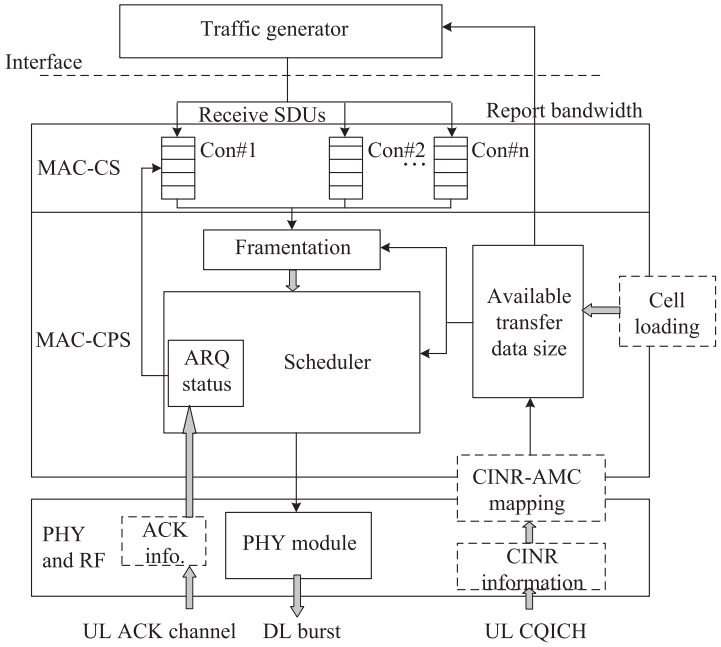


Figure 3 IEEE 802.16e medium access control (MAC) layer (ITU-T, 2006)

is three dimensional: spatial scalable, temporal scalable, and quality scalable; thus, video server encodes videos into several layers that consist of one base layer and one or several enhancement layers. The base layer provides users with basic video quality with low bit rate and enhancement layer is used to refine the video quality of the base layer. Users who receive more enhancement layer data can decode the videos in better quality while the base layer only provides basic video quality with low bit rate. When it makes users uncomfortable to watch videos intermittently, maintaining the video streaming is important. For users, basic video quality with less delay is of course acceptable, but it is even better to have more enhancement layer data to better the video quality.

Non-SVC has only single bit rate streaming but not all users adapt themselves to the bit rate. In the 802.16e network system, users of different positions influence the signal strength and each modulation technology provides different bandwidth size. For this reason, the single bit rate streaming cannot satisfy the requirement of each user. Compared with the non-SVC, SVC provides different bit rate streaming for users in different network environments. SVC uses the layered structure to assign for each layer according to the importance. For the important layer, the unequal erasure protection can be used to protect the important data. Thanks to these advantages, the SVC is more suitable for transport technology in wireless network system than non-SVC system.

The SVC extension of H.264/AVC standard is developed by The Joint Video Team of the ITU-T Video Coding Experts Group and the ISO/IEC MPEG (Reichel *et al.*, 2006; Wiegand *et al.*, 2006). This standard involve temporal scalability (frame rate), spatial scalability (picture resolution), and SNR scalability (quality). Figure 4 shows the H.264/AVC SVC structure (Schierl *et al.*, 2007).

To support the temporal scalability, the concept of hierarchical prediction structure is used. Each set of temporal layers can be decoded independently and the enhancement layer is coded as B pictures and the group of pictures is classified into several layers. To support the spatial scalable coding, the multi-layer structure with inter-layer prediction is adopted and the inter-layer prediction includes the motion prediction, residual prediction, and intra prediction.

The video is decomposed into several layers and each spatial layer can get prediction from the lower spatial layer. Quality scalability is the special case of spatial scalability with the same picture size for base and enhancement layers. Each spatial layer is referred to coarse-grain quality scalable coding (CGS) or medium-grain quality (MGS). When CGS is used for quality scalability, switching between different CGS layers must be at the defined point. On the contrary, when MGS is used, switching different MGS layers

https://doi.org/10.1017/S0269888914000186 Published online by Cambridge University Press

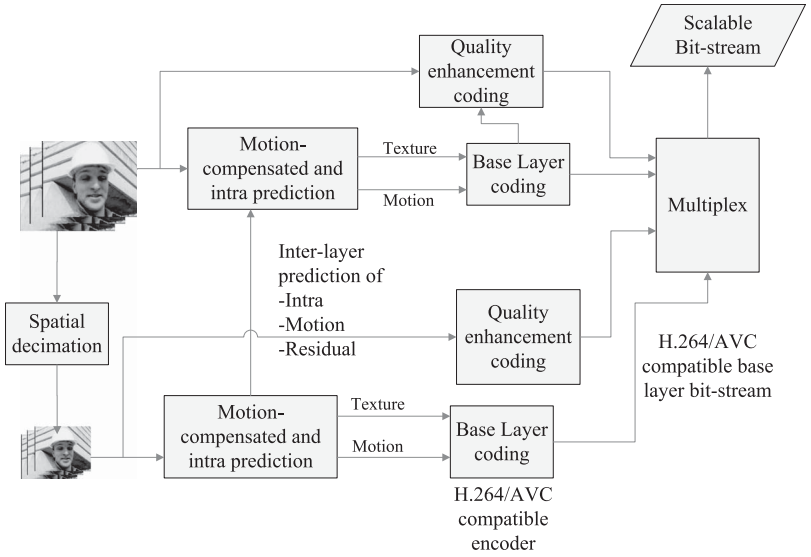


Figure 4 The scalable extension of H.264/AVC structure

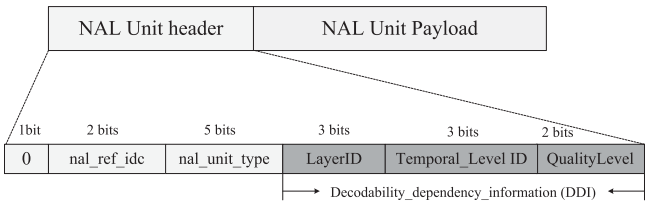


Figure 5 The Network Abstraction Layer (NAL) header of scalable H.264/AVC

can be in any access unit. In the way of the MGS concept, any enhancement layer NAL (Network Abstraction Layer) unit can be discarded from a quality scalable bit stream. For the first quality layer is restricted to CGS, and enhancement layer is consisting of fine-grained scalability (FGS).

The header information of H.264/AVC is depicted in Figure 5. The first byte is used in basic video, which is encoded by H.264AVC standard. In order to support the H.264/AVC scalable video, a byte extension of the NAL header is adopted. In this byte, the first three bits refer to different spatial resolutions. The next three bits stand for a temporal resolution, while the last two bits represent a quality level. The quality level for the base layer is zero and will increase one for each FGS enhancement layer.

3 The proposed scheme

In this section, we analyze QoS of WiMAX service classes in NGNs (Deng *et al.*, 2008). Among the five WiMAX QoS service classes, UGS and ertPS are suitable for use in VoIP and Video Conference, while rtPS is suitable for use in streaming media. Delay-based algorithms define real-time traffic, UGS, ertPS, and rtPS service classes, for which the delay bound is the primary QoS parameter and basically the packets with unacceptable delays are discarded.

In this paper, by applying algorithms including EDF and LWDF to our proposed dropping mechanism, we deal with video streaming over QoS. In IEEE 802.16, there are two strategies in QoS scheduler (as shown in Figure 6): scheduler to serve the connection based on the deadline, and largest delay to avoid missing its deadline. SVC can encode the videos into levels of different quality. According to different classifications, we encode the video and put it in the queue of uplink and down link of BS and subscriber station in WiMAX. Whenever the congestion occurs, the QoS-based active dropping mechanism is adopted to drop packets according to the present network loading and different user levels. If the network

https://doi.org/10.1017/S0269888914000186 Published online by Cambridge University Press

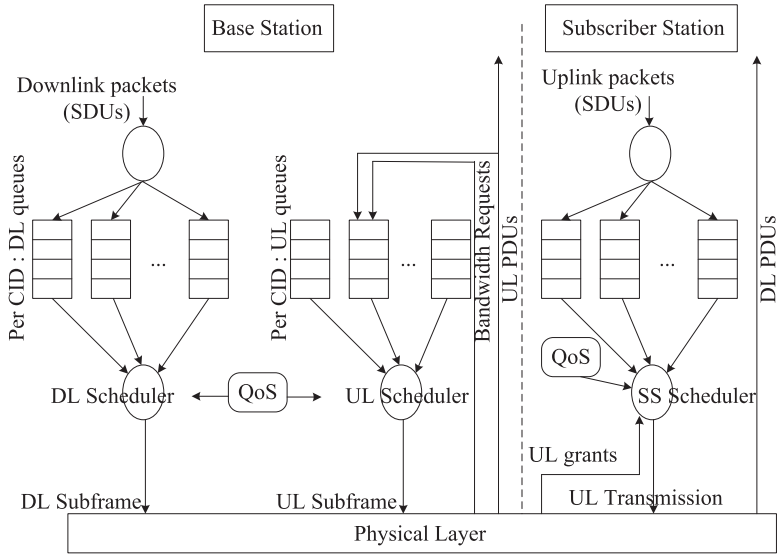


Figure 6 The scheduler structures of base station (BS) and subscriber station (SS) (Cicconetti *et al.*, 2006)

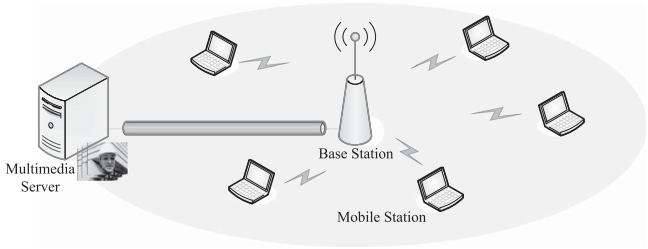


Figure 7 The point to multi-point system

loading exceeds the threshold, the dropping mechanism starts to drop the enhancement layer data for users of low levels and the dropping probability is varying with different network loading.

3.1 The dropping model

Figure 7 depicts the 802.16e point to multi-point system that is simulated in this work. The system consists of a multimedia server, one BS and several mobile stations (MS). The connection between video server and BS is an IP-based network. MS receives the video stream form the video server through the connection between BS and MS. For video transmission, we choose the RT-VR service flow between the BS and the MS.

The video streaming is encoded by scalable extension of H.264/AVC in this work. In the video server, we provide several spatial and temporal resolutions. In the resolutions, the quality layer consists of three layers, including one base layer and two enhancement layers. In Schwarz *et al.* (2007), the scalable bit streams are generated by scalable H.264/MPEG-4 AVC extension with the SVC coding structure. Table 1 shows the average bit rates in each spatial–temporal resolution.

In the simulation, the 802.16e-2005 OFDMA-TDD mode is adopted. The bandwidth is 5 MHz and the FFT size is 2048. With the standard (Reichel *et al.*, 2006), we calculate the OFDMA parameters and the results of the calculated parameters are shown in Table 2.

In the BS, the enhancement layer information will be dropped out from the queue when the network loading goes beyond the threshold. In the following, we set two conditions that represent these two stages of the network loading thresholds.

https://doi.org/10.1017/S0269888914000186 Published online by Cambridge University Press

Table 1 The average bit rates of the scalable video coding bit stream at different spatial–temporal quality layer

Format	Bit rates (kbit/second)				
QCIF 15 Hz	96	112	128	160	192
CIF 7.5 Hz	192	224	256	320	384
CIF 15 Hz	256	320	384	448	512
CIF 30 Hz	384	448	512	640	768

Table 2 The parameters of 802.16e PHY layer

Parameters	Value
Channel bandwidth	5 M(Hz)
Used of sub-channels	60
Size of FFT	2048
Symbol time (T_b)	358.4 μs
Guard time (T_g)	44.8 μs
OFDMA symbol time ($T_s = T_b + T_g$)	403.2 μs

When the network loading exceeds this threshold, the active dropping mechanism will start. The first-stage threshold is set with network loading in p percent and the second threshold is q percent. The active dropping mechanism is defined as:

$$DroppedData = \sum_{i=1}^{3} P_i \cdot quality_i, \tag{1}$$

where *quality_i* is the video stream of quality layer. The base layer is assigned from *quality*_1 to *quality_i* stands for its enhancement layer. In this work, the video server encodes the video to one base layer and two enhancement layers. The P_i is the probability to show whether the data should be dropped or not. Because we guarantee the basic video quality for users, P_1 is equal to 0 before the network loading exceeds 95%. When the system loading exceeds 95%, the network and video quality become unstable. The first-level enhancement data is dropped by the P_2 probability when the network loading goes beyond the first-level threshold. The P_2 is defined as:

$$P_2 = \begin{cases} 0 & \text{if} \quad 0 \le x \le p\%, \\ 0 < \frac{x-p}{q-p} < 1 & \text{if} \quad p\% < x < q\%, \\ 1 & \text{if} \quad x \ge q\%, \end{cases} \tag{2}$$

where the value of x refers to the network loading. The second-level enhancement data is dropped by the P_3 probability. The dropping probability of the second-level enhancement is larger than the probability of the first level because the second-level enhancement data is less important than that of the first level. The P_3 is defined as:

$$P_3 = \begin{cases} 0 < \frac{x}{p} < 1 & \text{if} \quad 0 \le x \le p\%, \\ 0 < \frac{x-p}{q-p} < 1 & \text{if} \quad p\% < x < q\%, \\ 1 & \text{if} \quad x \ge q\%. \end{cases} \tag{3}$$

By using these equations, we can calculate the dropping probability of each layer data by the network loading presently.

https://doi.org/10.1017/S0269888914000186 Published online by Cambridge University Press

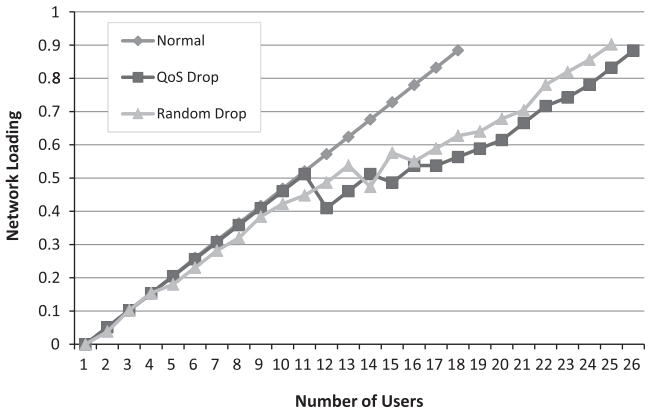


Figure 8 Network loading fluctuation with different mechanisms

3.2 Mechanism optimization

In order to make the bandwidth threshold value suitable for use in different kinds of network loading situations, we propose three kinds of mechanisms to achieve the optimization of the bandwidth threshold value dynamically.

The principle of the threshold adjustment can be illustrated in two situations. The first situation is when the estimation of video service traffic is larger than the estimation of the available bandwidth, the adjustment mechanism decreases the threshold value. On the contrary, when the estimation of video service traffic is smaller than the estimation of the available bandwidth, the adjustment mechanism increases the threshold value. By this adjustment, we get the optimization when the estimation of video service traffic equals to the estimation of the available bandwidth. The threshold value is adjusted according to three kinds of proposed mechanisms: step adjustment (SA), over measure adjustment (OA), and dependence adjustment (DA).

In the SA mechanism, the adjustment value is set as one. The threshold value decreases by 1 when the estimation of video service traffic is larger than the estimation of the available bandwidth and the threshold value increases by 1 when the estimation of video service traffic is smaller than the estimation of the available bandwidth. In the OA mechanism, the adjustment value is defined as:

$$P_n = P_{n-1} + \frac{(available_bandwidth - traffic_estimation)}{2}, \tag{4}$$

where $P_{(n-1)}$ represents the previous threshold value. In the DA mechanism, the adjustment value is defined as:

$$P_n = (available_bandwidth - traffic_estimation) \cdot \alpha + P_{n-1} \cdot (1 - \alpha), \tag{5}$$

where $P_{(n-1)}$ is the previous threshold value and α the dependence factor.

4 Simulation results

In the simulation, we assign three levels of user priority to each user and set p value as 50 and q value as 75. Users of the highest priority can obtain the whole video stream data including one base layer and two enhancement layers in any network loading condition. Users of medium and low priority can receive the whole video data when the network loading is under the first-stage threshold. When the network loading exceeds the first-stage threshold, users of medium priority receive only one base layer and one enhancement layer data, and users of low priority receive only base layer data. The users (MS) accessing the video server increase one per second, and user priority is assigned to each user randomly.

The network loading fluctuation with QoS dropping mechanism is shown in Figure 8. The normal scheme (non-active dropping) can be observed that the network loading increases with the number of users. In this case, the system providing service for users (MS) restrains a fixed number of users because all

https://doi.org/10.1017/S0269888914000186 Published online by Cambridge University Press

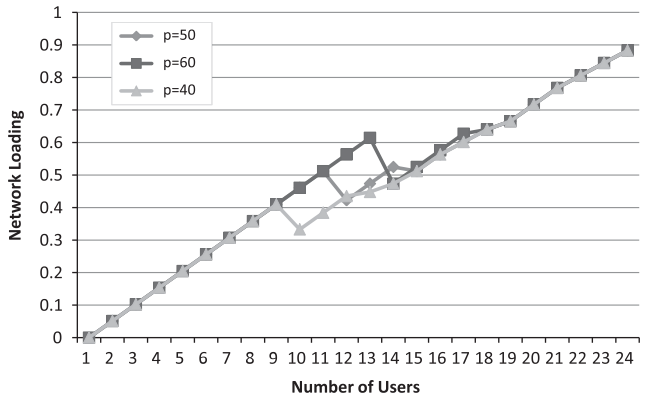


Figure 9 The influence of different threshold values

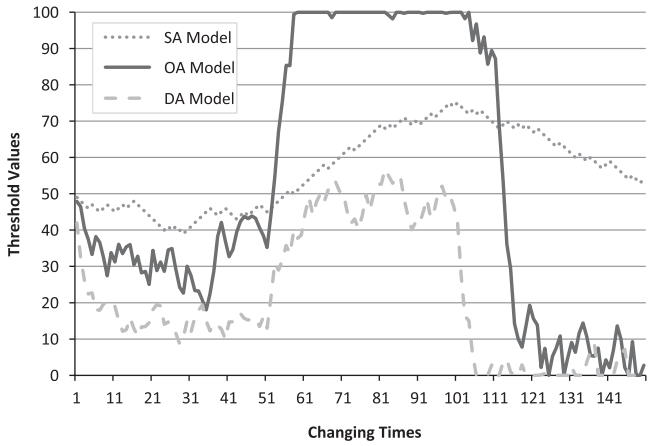


Figure 10 Optimization of threshold values

users receive the whole video data including the base layer and enhancement layers. In the random dropping scheme, the system drops the data of two enhancement layers from the buffer with the P_3 probability value in Equation (1) in any network loading. The data of level one enhancement layer is dropped with the P_2 probability value in Equation (1) when the network loading goes beyond 50%. In the proposed QoS-based active dropping mechanism, all users are able to obtain better video quality when the system loading is low. When the network loading is high, the video quality for high-level users is still guaranteed and users of low levels can at least obtain the basic video quality. With the same bandwidth, more users gain access to the BS, which means that the BS has more effective system capability.

The influence of different threshold values is shown in Figure 9. Different threshold values that we choose effect the start time of the dropping mechanism. Generally speaking, larger threshold value provides better video quality because the dropping mechanism starts with high network loading. In this situation, the BS provides fewer users to access the wireless network, but when the threshold is small, BS can provide more users to access wireless network. In this situation, the video quality will be reduced because the dropping mechanism starts with low network loading.

After optimizing the dropping mechanism, the simulation result is shown in Figure 10. To observe the optimization model, Figure 10 can be divided into three parts. The first part is the region of the changing times from 0 to 50. It can be known that the estimation of video service traffic is close to the estimation of the available bandwidth. The second part is the region of the changing times from 51 to 100. In this part, the estimation of video service traffic is smaller than the estimation of the available bandwidth with high probability. The final is the region of the changing times from 101 to 150. In this part, the estimation of

https://doi.org/10.1017/S0269888914000186 Published online by Cambridge University Press

video service traffic is larger than the estimation of the available bandwidth with high probability. According to our observation, the SA model adjusts smoothly while the OA model adjusts acutely. Smooth adjustment allows users to maintain smooth video quality but SA model cannot adjust immediately when the network loading vary acutely. On the other hand, the OA model can provide immediate adjustment but cannot provide users smooth video quality. The DA model, having the advantage of both SA and OA, provides not only immediate adjustment but also smooth video quality for users.

5 Conclusion

Effective bandwidth utilization is a crucial issue in video streaming over wireless network, as the video quality and bandwidth utilization usually cannot be maintained at the same time. Thus, we focused on the SVC video and video streaming in IEEE 802.16e network. In this paper, an effective QoS-based active dropping mechanism is proposed to deal with the bandwidth utilization. We use user priority to assign different levels to each user, and the priority level influences how much enhancement layer data users can receive. The dropping mechanism runs when the loading of the BS loading exceeds the chosen bandwidth threshold. This mechanism can release bandwidth that was used to transmit the dropped data previously and the BS allows more users. In order to optimize our proposed dropping mechanism, we use three kinds of threshold adjustments to adjust threshold value dynamically. Owing to this optimization, the threshold can be suitable for use in all kinds of situations of network loading. The simulation results show that by the proposed mechanism, we not only effectively improve the bandwidth utilization and the network capability, but also satisfy users of different levels for their requirement of video quality. In the optimization results, the threshold value can be adjusted when the network loading varies with time.

Acknowledgement

This research was partly funded by the Ministry of Science and Technology of R.O.C. under grant 103-2221-E-197-018 and 101-2221-E-197-008-MY3.

References

Bajic, I. V., Tickoo, O., Balan, A., Kalyanaraman, S. & Woods, J. W. 2003. Integrated end-to-end buffer management and congestion control for scalable video communications. In *Proceedings of 2003 International Conference on Image Processing (ICIP 2003)*, **3**, 257–260.

Chakchai, S., Jain, R. & Tamimi, A. K. 2009. Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey. *IEEE Journal on Selected Areas in Communications* **27**(2), 156–171.

Chang, K. D., Chen, C. Y., Chen, J. L. & Chao, H. C. 2010. Challenges to next generation services in IP multimedia subsystem. *Journal of Information Processing Systems* **6**(2), 129–146.

Chen, C. Y., Wu, T. Y., Huang, Y. M. & Chao, H. C. 2008. An efficient end-to-end security mechanism for IP multimedia subsystem. *Computer Communications* **31**(18), 4259–4268.

Cheng, S. T., Hsieh, M. T. & Chen, B. F. 2010. Fairness-based scheduling algorithm for time division duplex mode IEEE 802.16 broadband wireless access systems. *IET Communications* **4**(9), 1065–1072.

Chuang, H. C., Huang, C. Y. & Chiang, T. 2004. On the buffer dynamics of scalable video streaming over wireless network. In *Proceedings of IEEE 60th Vehicular Technology Conference (VTC2004-Fall)* **4**, 2582–2586.

Cicconetti, C., Lenzini, L., Mingozzi, E. & Eklund, C. 2006. Quality of service support in IEEE 802.16 networks. *IEEE Network* **20**(2), 50–55.

Deng, D. J., Chang, L. W., Wu, T. Y. & Hu, C. C. 2008. Guaranteed QoS provision scheduling mechanism for CBR traffic in IEEE 802.16 BWA systems. *Journal of Internet Technology* **9**(4), 403–409.

Huang, C., Juan, H. H., Lin, M. S. & Chang, C. J. 2007. Radio resource management of heterogeneous services in mobile WiMAX systems. *IEEE Wireless Communications* **14**(1), 20–26.

IEEE 802.16-2004 2004. IEEE standard for local and metropolitan area networks part 16: air interface for fixed broadband wireless access systems. In *IEEE*.

IEEE 802.16e-2005 2005. IEEE standard for local and metropolitan area networks part 16: air interface for fixed and mobile broadband wireless access systems, amendment 2: physical and medium access control layers for combined fixed and mobile operation in licensed bands and corrigendum 1. In *IEEE*.

ITU-T 2006. Functional requirements and architecture for resource and admission control in next generation networks. In *ITU-T Recommendation Y.2111*.

https://doi.org/10.1017/S0269888914000186 Published online by Cambridge University Press

Kim, T. & Lim, J. T. 2010. Quality of service supporting downlink scheduling scheme in worldwide interoperability for microwave access wireless access systems. *IET Communications* **4**(1), 32–38.

Mai, Y. T., Yang, C. C. & Lin, Y. H. 2007. Cross-layer QoS framework in the IEEE 802.16 Network. In *Proceedings of The 9th International Conference on Advanced Communication Technology* **3**, 2090–2095.

Markopoulou, A. & Han, S. 2001. Transmitting scalable video over a DiffServ network. `http://www.stanford.edu/class/ee368c/Projects/project06/index.html`.

Reichel, J., Schwarz, H. & Wien, M. 2006. Joint scalable video model JSVM-5. In *Doc. JVTR202*.

Schierl, T., Hellge, C., Mirta, S., Gruneberg, K. & Wiegand, T. 2007. Using H.264/AVC-based scalable video coding (SVC) for real time streaming in wireless IP networks. In *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS 2007)*, 3455–3458.

Schwarz, H., Marpe, D. & Wiegand, T. 2007. Overview of the scalable video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology* **17**(9), 1103–1120.

She, J., Hou, F. & Ho, P. H. 2007. An application-driven MAC-layer buffer management with active dropping for real-time video streaming in 802.16 networks. In *Proceedings of 21st International Conference on Advanced Information Networking and Applications (AINA '07)*, 451–458.

Wiegand, T., Sullivan, G. J., Reichel, J., Schwarz, H. & Wien, M. 2006. Joint draft 5: scalable video coding. In *Doc. JVTR201*.

Wiegand, T., Sullivan, G. J., Reichel, J., Schwarz, H. & Wien, M. 2007. Joint draft 9 of SVC amendment (revision 2). In *Doc. JVTV201*.

Zhang, Y., Huangfu, W., Xu, J., Li, K. & Xu, C. 2007. Integrated rate control and buffer management for scalable video streaming. In *IEEE International Conference on Multimedia & Expo (ICME)*, China.

Zhou, L., Geller, B., Wang, X., Wei, A., Zheng, B. & Chao, H. C. 2009. Multi-user video streaming over multiple heterogeneous wireless networks: a distributed, cross-layer design paradigm. *Journal of Internet Technology* **10**(1), 1–12.

https://doi.org/10.1017/S0269888914000186 Published online by Cambridge University Press