Delay-Constrained Adaptive Early Drop forReal-time Video Delivery over IEEE 802.11 Wireless Networks

Sooyong Lee*, Giwon Yang†, Yohaan Yoon†, Kyunghee Lee‡,Soon Joo Hyun†,Ben Lee§and Myungchul Kim†

* Samsung Electronics, Suwon, Republic of Korea

E-mail: syong82.lee@samsung.com

† Department of Computer Science, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

E-mail: {giwone, straightfor, sjhyun, mck}@kaist.ac.kr

‡Electronics and Telecommunication Research Institute, Daejeon, Republic of Korea

Email: leekhe@etri.re.kr

§School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97330, USA

E-mail: benl@eecs.orst.edu

ABSTRACT

The recent proliferation of mobile devices increases the demand for high-quality real-time video services in wireless network. However, the characteristics of wireless medium such as interferences and ease of congestion make it difficult for the delivered video to be played out in a timely manner.The latency introducedby those wireless characteristics severely degrades the perceptual quality of real-time videodue to intermittent playoutat the client side. In this paper, we propose adelay-constrained adaptive early drop scheme for real-time video delivery over IEEE 802.11 wireless networks. In our scheme, an intermediate router forcibly discards the less important video packetswhich are expected not to be played out in time. The simulation results demonstratethat the proposed scheme can achieve 20% lower end-to-end delay and60% less frame loss ratio with32% better Peak Signal-to-Noise Ratio (PSNR) compared to the existing cross-layer mapping schemes.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Wireless communication

General Terms

Algorithms, Design, Experimentation, Performance

Keywords

Video delivery, IEEE 802.11 Wireless networks, H.264/AVC, Quality of Service.

1. INTRODUCTION

Recently, the exploding growth of mobile devices with high computing power and display resolution has increased the demands for high-quality video services in wireless networks. Even though the physical link rate of IEEE 802.11 wireless networks [1] has continuously increased, supporting such services over wireless multi-hop network is challenging due to the interferences, the ease of network congestion and the error-prone medium. As the number of hops increases, packet delay and packet loss increase due to the carrier-sense and hidden node interferences [2]. In particular, the packet delay severely degrades the video quality because a video frame cannot be played outat the receiver in time after the specified playout time even if it is successfully transmitted. Buffering at the receiver side has drawbacks in terms of the delay and this can be a critical challenge against the real-time video delivery.

Although one of the widely-usedvideo codecs, H.264/Advanced Video Coding (AVC) [3] has the advantages in error-resilient and network-friendly features, video codecs themselves cannot appropriately deal with but just drop the delayed video frames due to the playout time constraint.IEEE 802.11e Extended Distributed Channel Access (EDCA) [1] was proposed to provide QoS differentiation with the Access Category (AC) according to traffic types. However, EDCA cannot guarantee the video quality in a congested network because all the video packets are transmitted with being allocated to one AC for video traffic. As a result, the queue overflows so that the video packets are delayed and lost.

Accordingly, the cross-layer mapping approaches were proposed with assigning each video frame or slice to ACs by its priority [4, 5]. Their approachesgive more transmission opportunity to more important video frames or slices by allocating them to different ACs. However, they did not consider the playout time constraint. To provide interactive services, a video frame should be played within 400msafter it is encodedaccording to ITU-T [6]. Another approach [7] drops a video packet which exceeds time-to-live in an intermediate router, but it requires the cache for the packet retransmission on every node. Besides, the receiver should inform which packets it does not receive to everyupstream node.

In this paper, we propose an adaptive priority-based early drop scheme which assigns a delay budget, the remaining time constraint until playback, to video packets. An intermediate router forcibly discards video packets which are expected not to be played in time constrained by the

ICUIMC(IMCOM)'13, January17–19, 2013, Kota Kinabalu, Malaysia.

delay budget. Furthermore, we differentiate the delay budget based on the priority of video frames to reduce distortion and error propagation especially in a congested network. The performance of our scheme is evaluated by Open Evaluation Frame work for Multimedia Over Networks (OEFMON) [8] to demonstrate our scheme significantly enhances the video quality in terms of end-to-end delay, frame loss ratio and PSNR. The evaluation results demonstrate that our scheme can achieve 20% lower end-to-end delay and 60% less frame loss ratio with 32% better PSNR compared to the existing cross-layer mapping schemes.

This paper is organized as follows. Section 2 introduces the background and related work. Section 3 presents our proposed scheme in detail. The simulation results are described in Section 4. We finalize the paper with the conclusions and future work in Section 5.

2. BACKGROUND AND RELATED WORK

2.1 H.264/AVC

H.264/AVC is a widely used video coding standard proposed by Joint Video Teams [9]. H.264/AVC consists of Video Coding Layer (VCL) and Network Abstraction Layer (NAL). After a video frame is encoded by VCL, it is encapsulated into several NAL Units (NALUs) by NAL. Each NALU consists of NAL header and Raw Byte Sequence Payload (RBSP) which is the encoded video data. NAL header contains the precedence of RBSP, and the type information of RBSP.

H.264/AVC defines three frame types: I-frame, P-frame, and B-frame. The loss of I- or P-frame not only affects the decoding of the following frames but incurs the distortion until the next I-frame is successfully decoded. In particular, I-frame is referenced by P- or B-frame so that the impact of I-frame loss is more severe than that of P-or B- frame loss. On the other hand, the loss of B-frame does not severely affect decoding of any frame because, in general, it is not used as a reference frame for more efficient compression. As a consequence, the order of importance in video frame type is I-, P-, and B-frame.

2.2 Cross-layer mapping schemes

There are few studies which consider ensuring the transmission of the important video frame to improve quality of H.264 video transmission in a single-hop wireless local area network [4, 5]. They commonly use the mapping strategy between the EDCA ACs and types of video frames.

In [4], I-frame is mapped toAC(3) with the highest priority, and P- and B-frames are mapped toAC(2) and AC(1), respectively. With this mapping, I-frame takes advantage of more frequent channel access than P- and B-frame under network congestion. Therefore, the error propagation in the decoded video frames due to reference frame loss can be reduced by granting more transmission opportunity to I-frames. As a result, overall performance can be enhanced compared with the conventional Distributed Coordinate Function (DCF).

In [5], the enhanced cross-layer architecture was proposed to improve data-partitioned H.264/AVC video transmission over IEEE 802.11e network. Data partitioning (DP) is an extension of H.264/AVC standard for error-resilience. Each type of video frame is separated into three partitions, partitions A, B and C by means of multiple macro blocks. Partition A includes the essential information of video data such as header, vector, and quantization information. Partitions B and C are additional video data to improve the quality of decoded frame. If partition A is delivered successfully, H.264 video data can be restored when other partitions are lost, but there is no way to recover the video data if partition A is missing. The approach in [5] maps between the EDCA ACs and types of data partitions with their priorities. This cross-layer approachcan achieve better quality of video transmission under congested networks with the heavy-loaded best-effort traffic.

Those two cross-layer mapping approaches utilize the advanced features of H.264 which is the precedence of video frames or data partitions. However, they cannot properly handle the delayed video packets because an AC with higher priority can be swamped with more important video frames or partitions in a congested network. This can incur queue overflows and significantly affect the user-perceived video quality. Video frames are sequentially generated and encoded from source. Then, video frames should be decoded and played out at receiver sequentially in time. The aforementioned cross-layer mapping methods may increase the overall delay to transmit video packetsdue to asynchronous transmission time of each AC. This delay makes users experience intermittent video playback at the receiver.

3. ADAPTIVE EARLY DROP SCHEME

In order to determine whether a video packet can be played successfully at the destination or not, *delay budget* is assigned to each video packet based on playout time constraint according to the ITU-T recommendation. The initial *delay budget*, B_{init}, is configured at a source to be less

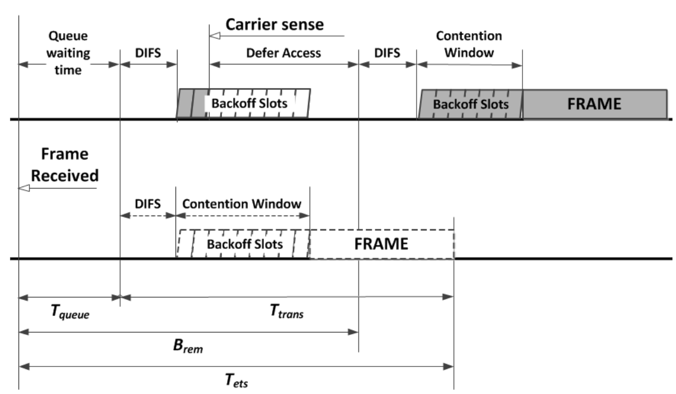


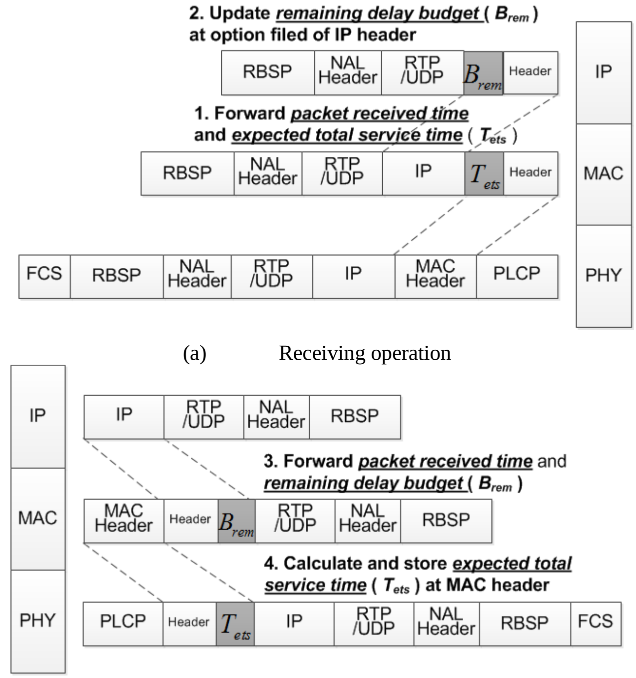
Figure1. Expected total service time in IEEE 802.11 DCF basic access mode

than or equal to the playout time constraint (i.e., 400ms). As the corresponding video packet passes through intermediate routers, the *delay budget* is decremented in hop by hop fashion. This remaining *delay budget* is denoted as B_{rem}. The *expected total service time*, T_{ets}, is the expected transmission time on a single wireless link. For the recognition of B_{rem} and T_{ets} at MAC layer, the option field in IP header is used and just a short 4-byte field is added to MAC header, respectively. The proposed scheme operates according to the following steps:

- Initialization – A video sender initially configures B_{rem} as B_{init} and stores it into IP header of a video packet.
- Reception –Upon receiving a video packet, an intermediate router updates B_{rem} in IP header of the packet.
- Transmission – When a video packet arrives at the end of the output queue, a source node or an intermediate router calculates T_{ets} and compares it to B_{rem}. If T_{ets} is less than B_{rem}, then the packet is transmitted after recording T_{ets} to MAC header. Otherwise, the packet is forcibly dropped.

In a basic access mode of IEEE 802.11 Distributed Coordination Function (DCF), if a node has a data packet to transmit, it waits for Distributed Inter-Frame Space (DIFS). If the channel is idle during DIFS, the backoffprocedure is started as shown in Figure 1. The actual packet transmission starts after the backoff slots are decremented to zero. If the node senses channel busy during DIFS or backoff time, it defers the transmission until the channel is idle. Unless the channel is busy during DIFS, we could estimate a packet transmission time based on the backoff slots and the packet size.

The expected total service time, T_{ets}, can be estimated as the sum of the queue waiting time, T_{queue}, and the expected packet transmission time, T_{trans}. $T_{trans}$is the sum of DIFS, $T_{backoff}$ and T_{data} which are the remaining backoff time and data frame transmission time, respectively. The intermediate routers simply calculate T_{queue} by recording the packet arrival time at the output queue and the time the packet is located at the end of the queue. Upon the packet arrives at the end of the queue, $T_{backoff}$ is configured as the product of the remainingbackoff slots and the slot time. If the packet is retransmitted due to the loss of ACK frame, the intermediate router just adds the time for transmission failure and reconfigures $T_{backoff}$. T_{data} is estimated by the quotient of the size of the packets divided by the link bandwidth. If the network is not congested, T_{ets} is so small that it may not affect the packet delay. However, packet retransmission occurs frequently in a congested network so that T_{ets}can easily exceed the remaining delay budget, B_{rem}.Note that the packet can be retransmitted at MAC layer



(b) Transmitting operation

Figure2. Early drop operation at an intermediate router

as long as T_{ets} does not exceed B_{rem}.

In our scheme, the initial delay budget, B_{init} is configured differently according to a video frame type to give more transmission opportunity to high priority frames. For this differentiation, the B_{init} of P-frame is set by simply subtracting *gap* from the B_{init} of the I-frame. For example, the B_{init} of P-frame is 380ms when the B_{init} of the I-frame is 400ms and *gap* is 20ms. The tunable parameter, *gap*, should be greater than or equal to the product of the average number of NALUs segregated from a single I-frame and the average transmission time per hop in order to ensure I-frame transmission. Therefore, as the *gap* increases, the transmission probability of I-frames increases whereas that of other frames decreases.

The proposed scheme operates at both a video sender and an intermediate router. At a video sender, the type information of the video frame is stored at NAL header and its encoding time is forwarded to lower layers for calculating T_{queue}. After that, IP layer initiates B_{rem} as B_{init} and stores B_{rem} at the option field in IP header. At MAC layer, T_{ets} is calculated and stored at the additional field in MAC header. Note that only four-byte fields are added to both IP and MAC headers for storing B_{rem}andT_{ets}.

When an intermediate router receives a video packet, the receiving time of the video packet and T_{ets} are forwarded to upper layer. Then, IP layer updates the B_{rem} based on forwarded T_{ets} as shown in the sequence of Figure2(a). The updated B_{rem} is calculated simply by subtracting T_{ets} from the received B_{rem}. To transmit the video packet, IP layer

Table 1.Simulation Parameters

Parameters	Values
Radio type	802.11a (Router), 802.11g (AP)
Data link rate	18Mbps
Basic rate	6Mbps
Resolution	4CIF (704*576)
Data rate	800Kbps
Frame per second	30fps
Video length	20s
Playout time	400ms
Video frame type	I, P-frame
I-frame interval	30
Video flows	One way, 8 flows
B_{init}	400ms
gap	20ms

forwards the receiving time of the video packet and updated B_{rem} to MAC layer as depicted in the sequence of Figure2(b). And then, MAC layer calculates T_{ets} and determines either to transmit or to drop the video packet. If T_{ets} is larger than B_{rem}, the video packet is discarded.

4. EVALUATION

The performance of our scheme is evaluated via network simulator, EXATA 2.0.1 [10]. The simulated wireless network is configured as a linear topology of three nodes, A, B, and C connected in sequence. The video senders are connected to the node A via a wired hub whereas the video receivers are connected to the node C with IEEE 802.11g radio with 18Mbps link rate. We use OEFMON [8] to generate and measure the quality of H.264 video traffic. One of the ITU-T reference video, *Soccer* (4CIF), is used as video source. In order to deliver this video source, H.264/AVC codec creates I-frame for every second with a baseline profile defined in the standard and the encoding data rate is set to 800Kbps on average. The size of NALU is set to 1,300 bytes. In order to make network congested, eight video flows are transmitted twice. We compare the performance of our scheme to the existing cross-layer mapping scheme. NALUs created from I- and P-frame are mapped to AC(3)and AC(2)of EDCA, respectively.

Table 2 shows the average frame loss ratio, FLR for short, in a network and a video codec. The FLR in the network, FLR_{net}, of both schemes are the same as 0.11. However, the reasons for the frame loss are different. In the cross-layer mapping scheme, the queue overflow dominates the frame loss. In addition, FLR_{net} affects the frame loss at application layer, FLR_{app}, of the destination. About 68% of frames cannot be played due to excessive queuing delay. On the contrary, in our scheme, the network loss is incurred by compulsory packet drops for saving the network resource. Therefore, all the delivered video frames are played

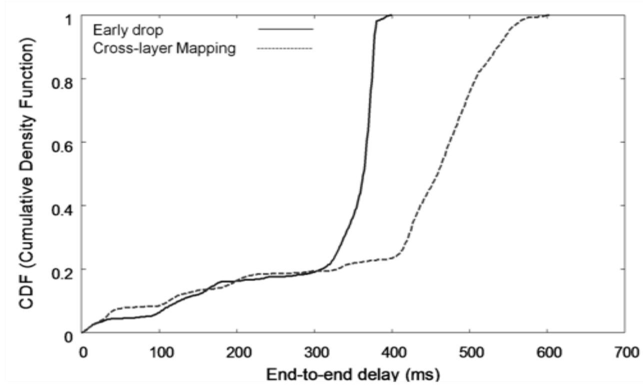


Figure3. End-to-end delay CDF

Table 2. Frame Loss Ratio

Scheme	FLR_{net}	FLR_{app}	$FLR_{I\text{-}frame}$	$FLR_{P\text{-}frame}$
Cross-layer mapping	0.11	0.68	0.00	0.71
Early drop	0.11	0.00	0.025	0.13

Table 3. Average End-to-End Delay and PSNR

Scheme	Average end-to-end delay (ms)			PSNR (dB)
	Total	I-frame	P-frame	
Cross-layer mapping	399.6	35.9	418.4	18.5
Early drop	317.3	330.7	316.7	24.4

because there is no frame loss at application layer.

We also measured the FLRs according to frame types to evaluate the capability of delivering more important video frames. The loss ratio of I-frame, $FLR_{I\text{-}frame}$, with our scheme is 2.5% whereas no I-frame loss occurs with the cross-layer mapping scheme. However, the difference of those values is not significant. On the other hand, the loss ratio of P-frame, $FLR_{P\text{-}frame}$, in the cross-layer mapping scheme is about 5.5 times larger than $FLR_{P\text{-}frame}$ in our scheme. It indicates that P-frames delivered with the cross-layer mapping scheme suffer from severe queue overflow at intermediate routers because of their low transmission priority. Accordingly, our scheme can reduce the entire frame loss ratio by 60% compared to the existing cross-layer mapping scheme.

Figure3shows the cumulative distribution of the end-to-end delay of the transmitted video frames. As can be seen, all the video frames are played within 400ms with our scheme while only 23% of video frames are played within 400ms with the cross-layer mapping scheme. This implies that the delay of video transmission in the network including queuing delay is longer than 400ms under the network congestion.

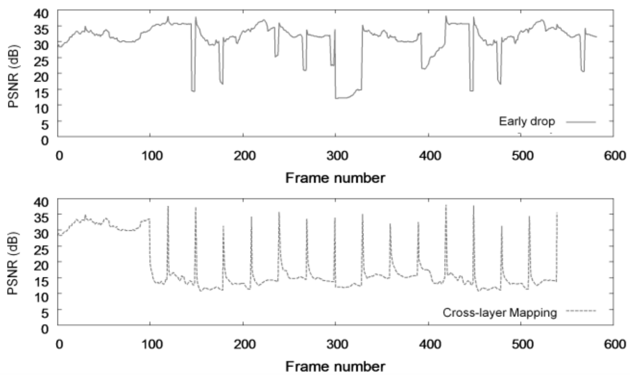


Figure4.PSNR values of video frames delivered at node 13

Figure5.The snapshot from frame #267 and #271 at node 13

Table 3 shows the end-to-end delay according to frame type and PSNR values on average. Although the end-to-end delay of I-frame with our scheme is about 9.2 times longer than that with the cross-layer mapping schemebecause each intermediate router has one output queue for video traffic, the delay variation between I- and P-frame is less than the video frame transmission interval (33ms), which means most video frames can be transmitted sequentially. By contrast, the delay variation is large (382ms) with the cross-layer mapping scheme because the prioritized I-frame transmission causes the starvation of P-frame transmission. Therefore, the transmissions of formerly encoded P-frame are delayed due to the transmissions of latterly encoded I-frame. These delay and disorder severely affects the perceptual video quality which can be measured by PSNR. The average PSNR value of the delivered video frames can be enhanced as 32% with our scheme from 18.5 dB to 24.4 dB.

The PSNR values of each frame are plotted in Figure 4. The PSNR value greater than 30 dB indicates that the corresponding video frame is successfully decoded and played.With the cross-layer mapping scheme, only the PSNR values of I-frames are greater than 30 dB because most P-frames cannot be decoded due to the excess of playout time constraint. On the other hand, the PSNR values of most frames are greater than 30 dB with our scheme. This quality enhancement can be found perceptually in the video frame snapshots as shown in Figure 5. All of five frames from 267 to 271 are successfully decoded and played with our scheme while thesame image is played with the cross-layer mapping scheme before and after the 269th frame which is I-frame.

5. CONCLUSIONS

In this paper, we proposed a delay-constrained adaptive early drop scheme for real-time video delivery over IEEE 802.11 wireless networks. The proposed early drop scheme selectively drops unnecessary video frames expected to be delivered without satisfying the playout time constraint. Therefore, we can save the limited wireless network resources and prevent additional queuing delay and queue overflow under network congestion. Note that our scheme can be easily applied to other types of video codecs. The simulation results show that the proposed scheme can provide enhanced video quality to mobile users. As a future work, we will find the optimized gap and initial delay budget by theoretical analysis considering more complicated wireless environments.

6. ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (No. 2012R1A2A2A01008244).

7. REFERENCES

[1] Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE 802.11 WG, June 2007.

[2] J. Li, C. Blake, D.S.J. De Couto, H.I. Lee, and R. Morris, Capacity of Ad Hoc Wireless Networks, InProc. 7th Annu. In Proc. ACM Mobile Computing and Networking (MOBICOM), pp. 61-69, 2001.

[3] T. Wiegand, G. Sullivan, G. Bjontegaard, and a Luthra, Overview of the H.264/AVC Video Coding Standard, IEEE Trans. Circuits and Systems for Video Technology, vol. 13, pp. 560-576, Jul. 2003.

[4] U. Choudhry and J. Kim, Performance Evaluation of H. 264 Mapping Strategies over IEEE 802.11e WLAN for Robust Video Streaming, Advances in Multimedia Information Processing-PCM 2005, pp. 818–829, 2005.

[5] A. Ksentini, M. Naimi, and A. Gueroui, Toward an Improvement of H.264 Video Transmission over IEEE 802.11e through a Cross-layer Architecture, IEEE Communications Magazine, vol. 44, pp. 107-114, Jan. 2006.

[6] ITU-T G.1010, End-user Multimedia QoS Categories, ITU-T Rec., 2001.

[7] T.P. Van, Video-Cooperative Design for Ad hoc Networks, In Proc. IEEE Wireless Communications and Networking Conference (WCNC), pp. 2849-2854, 2007.

[8] C. Lee, M. Kim, B. Lee, S. Hyun, K. Lee and S. Lee, OEFMON: An Open Evaluation Framework for Multimedia Over Networks,IEEE Communications Magazine, vol. 49, pp. 153-161, 2011.

[9] ITU-T Recommendation for Audiovisual and Multimedia Systems: Infrastructure of Audiovisual Services – Coding Moving Video, ITU-T Rec. H 264, March 2010.

[10] EXata2.0.1, http://www.scalable-networks.com/exata/.

[11] D. Wu, Y.T. Hou, and Y.Q. Zhang, Transporting Real-Time Video overthe Internet: Challenges and Approaches, In Proc. The IEEE, vol. 88, pp.1855–1875, Dec. 2000.