Performance Analysis of H.264/AVC, H.264/SVC, and VP8 over IEEE 802.11 Wireless Networks

Yohaan Yoon*, Myungchul Kim*, Sooyong Lee*, Ben Lee†, Soon Joo Hyun* and Kyunghee Lee‡
*Department of Computer Science, KAIST, Daejeon, Republic of Korea, Email: {straightfor, mck, torshong, sjhyun}@kaist.ac.kr
†School of Electrical Engineering and Computer Science, Oregon State University, Oregon, USA, Email: {benl}@eecs.orst.edu
‡Electronics and Telecommunication Research Institute, Daejeon, Republic of Korea, Email: {leekhe}@etri.re.kr

Abstract— **Although the latest video codecs such as H.264/AVC, H.264/SVC, and VP8 were developed with network-friendly features, provisioning the quality of video delivery over IEEE 802.11 wireless networks is still challenging because of error-prone medium and random access. This paper evaluates the quality of video delivery using those latest video codecs over IEEE 802.11. Our results show that reduction of coded video data, consideration of the queue size of each Access Category, and the mandatory implementation of error recovery features in H.264/SVC can help provide robust multimedia transmission over IEEE 802.11. In addition, the performance of the latest video codecs is compared via extensive simulation scenarios. Our comparison shows that VP8 achieves good video quality with the basic medium access control techniques of IEEE 802.11, whereas H.264/AVC and H.264/SVC benefit significantly from appropriate mapping schemes between encoded video fragments and IEEE 802.11e Access Categories.**

Keywords - Video over IEEE 802.11; IEEE 802.11e; QoS; H.264/AVC; H.264/SVC; VP8

I. INTRODUCTION

As devices such as smartphones, tablet PCs and smart TVs proliferate, the demands for multimedia services such as video conferencing, VoIP, and IPTV over IEEE 802.11 wireless networks, i.e., WLANs, are also increasing. Although the various video codecs such as H.264 Advanced Video Coding (H.264/AVC), H.264 Scalable Video Coding (H.264/SVC), and VP8 were developed with network-friendly features, provisioning the quality of video delivery over error-prone wireless networks with uneven distributed access control is still a challenging issue.

IEEE 802.11e was proposed to support Quality of Service (QoS) by providing Access Categories (ACs) for services differentiation. Several cross-layer approaches have been proposed to map encoded video data to different ACs based on the video fragment priority for robust video transmission [1-3]. However, these prior studies do not provide comprehensive performance comparison of various AC mapping schemes and are limited only to H.264/AVC and H.264/SVC.

Hence, this paper evaluates the quality of video delivery over WLANs using the latest video codecs including H.264/AVC, H.264/SVC and VP8. The quality was evaluated by means of both the network performance metrics, such as packet loss ratio and end-to-end delay, and the video quality metrics, such as Peak Signal to Noise Ratio (PSNR) and Mean Opinion Score (MOS).

Our evaluation leads to the following four important findings that should be considered for research on multimedia transmission over WLANs.

- Reducing the sudden surge in the amount of encoded video data helps reliable multimedia transmission over WLANs. This is because when there is burst of traffic, significant packet losses occur that degrade the quality of received video [4].
- The queue size of each AC should be carefully considered when AC mapping is applied. This is because the queue size requirement of high-priority video fragments is larger than that of low-priority ones, and thus a large number of more important video fragments will cause AC queues to overflow resulting in video quality degradation.
- The error recovery features should be implemented in the base layer for H.264/SVC. Although the base layer of H.264/SVC is assumed not to be lost, there is no guarantee that all packets of the base layer will be received in IEEE 802.11e.
- In general, VP8 provides better user-perceived quality than the other two codecs using Distributed Coordination Function (DCF) and Enhanced Distributed Channel Access (EDCA). However, H.264/AVC and H.264/SVC provide better user-perceived quality with mapping schemes between encoded video fragments and ACs.

The rest of this paper is organized as follows: Section II discusses the related studies. Section III presents the evaluation methodology including the experiment scenarios. Experimental results and analysis are discussed in Section IV. Finally, Section V concludes the paper and discusses future work.

II. RELATED WORK

A. Research on Multimedia over IEEE 802.11

In IEEE 802.11 DCF, there is no QoS support for multimedia services because all the stations equally contend for the wireless medium. The IEEE 802.11e provides basic QoS features with EDCA, which is a new channel access method that differentiates traffic into four ACs: voice, video, best efforts, and background [5].

Prior work in [1-3] studied AC mapping schemes to support QoS with finer granularity based on the importance of video fragments. Methods presented in [1, 2] map video slices generated by Data Partitioning (DP) in H.264/AVC to different ACs, but they compared only one mapping scheme to the conventional transmission in DCF and EDCA. Moreover, these approaches were studied using only H.264/AVC. In [3], the authors proposed the Traffic Prioritization Algorithm (TPA) for the transmission of H.264/AVC and H.264/SVC. In this approach, half of the video traffic is mapped to AC1 while the other half to AC2 based on the order of importance, reserving AC3 for VoIP and AC0 for background traffic. Thus, the previous studies do not consider different AC mapping schemes and have been limited only to H.264/AVC and H.264/SVC. Moreover, they are limited to evaluating video quality without observing the performance of wireless networks.

B. Video Codecs

The encoded video frames and layers have different importance depending on the video codec. They can be transmitted with differently mapping schemes over IEEE 802.11e. Therefore, we discuss the video frame and layer characteristics of the latest codecs.

H.264/AVC [6] is a standard for video compression developed by ITU-T together with the ISO/IEC. H.264/AVC contains a number of new features including error resilience/concealment and Network Abstraction Layer (NAL). The NAL facilitates the mapping H.264/AVC data to packets on IEEE 802.11 and provides header information that enables a simple and effective customization in accordance with the underlying networks. H.264/AVC has three types of frames: I-frame, P-frame, and B-frame. I-frame (intra-frame) is independently coded and does not reference any other frames. P-frame (predicted frame) is decoded using the previous P or I-frame. B-frame (bidirectional predicted frame) depends on both past and future frames for decoding. Therefore, I-frame is more important than the other two frames.

H.264/SVC is an extension of H.264/AVC that encodes video sources into the base layer and several enhancement layers [7]. The layers of H.264/SVC can be customized according to the needs of the applications and the conditions of the networks. Thus, it is possible to transmit a single video stream to multiple heterogeneous clients and networks. Moreover, the base layer can be decoded by any H.264/AVC-compliant decoder for backward compatibility. The base layer has less amount of video data compared to the enhancement layers due to lower quality versions of video. Thus, H.264/SVC assumes that the base layer will be preserved in any situations, and is required to decode the video.

VP8 [8] is an open-source video codec developed by On2 technologies and released by Google. VP8 consists of intra-frames and inter-frames. Intra-frame and inter-frame correspond to I-frame and P-frame of H.264/AVC, respectively. An intra-frame is called a key frame and is independently coded and has no reference to any other frames. Decoding an inter-frame depends on previous frames including recent intra-frame. Therefore, intra-frames have higher importance than inter-frames.

III. EXPERIMANTAL METHODOLOGY

This section presents the simulation study of various video delivery scenarios using the *Open Evaluation Framework Multimedia Over Networks* (OEFMON) [9] to measure both the network performance and the user-perceived quality. The encoders used for H.264/AVC, H.264/SVC, and VP8 are MONOGRAM x264 encoder 1.0.10.0 [10], JSVM 9.8 [12], and WebM VP8 encoder 0.9.12 [8], respectively. Although these software codecs do not completely implement the full features of H.264/AVC, H.264/SVC, and VP8, they are sufficient to evaluate the transmission properties of encoded video data over WLANs.

A. Simulation Environment

The simulated network topology and traffic are illustrated in Fig. 1, which consist of an ad hoc network of six IEEE 802.11a nodes with a data rate at 9 Mbps and Ad hoc On Demand Distance Vector (AODV) routing protocol. The distance between nodes is 90 m, and each node has transmission range of 100 m and a carrier sense range of 200 m. The size of jitter buffer is 150 bytes. Node C transmits the video stream under study to node D for 10 seconds. Meanwhile, nodes C, A, and E transmit background data to nodes D, F, and B, respectively, based on packet size, interval, and start and end times shown in Table I.

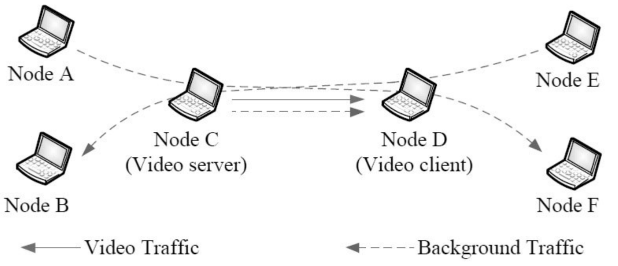


Figure 1. Network topology and traffic

TABLE I. BACKGROUND TRAFFIC

Source	Destination	Packet size	Interval	Start time	End time
C	D	1500 bytes	20ms	8s	16s
A	F	1500 bytes	20ms	9s	15s
E	B	1500 bytes	20ms	11s	13s

A video clip *Soccer* that consists of 300 frames in CIF resolution is used as the video source. The target bitrate of the video stream encoded by each codec is set to 4,000 Kbps according to the default value in commercial software such as FFmpeg [14]. The features of H.264/AVC follow the baseline profile used for video conferencing and mobile applications [6], and I-frame interval is set to 30 frames. H.264/SVC has the same features as H.264/AVC, and consists of three layers including the base layer and two enhancement layers. VP8 is configured as real-time CBR and intra-frame interval is set to 30 frames. The other features in codecs depend on commercial/referenced products.

The encoding parameters of the three video codecs are adjusted so that they all generate the same quality video streams with respect to PSNR and MOS. PSNR is the most commonly used metric for video quality and MOS provides a numerical indication of the perceived video quality after compression and transmission. While PSNR needs original video and received video, MOS requires an extensive formalized test of the system with a large number of users. Therefore, we measure MOS based on PSNR. As shown in Table II, MOS is indicated by a score of five steps, and PSNR can be converted to MOS according to ITU-R BT.500-11 [13]. Fig. 2 depicts PSNR and MOS for encoded video data for the codecs. As can be seen, all three encoded video data have similar user-perceived quality.

TABLE II. CONVERT TABLE FROM PSNR TO MOS

MOS	Quality	Impairment	PSNR (dB)
5	Excellent	Imperceptible	> 37
4	Good	Perceptible but not annoying	31 – 37
3	Fair	Slightly annoying	25 – 31
2	Poor	Annoying	20 – 25
1	Bad	Very Annoying	< 20

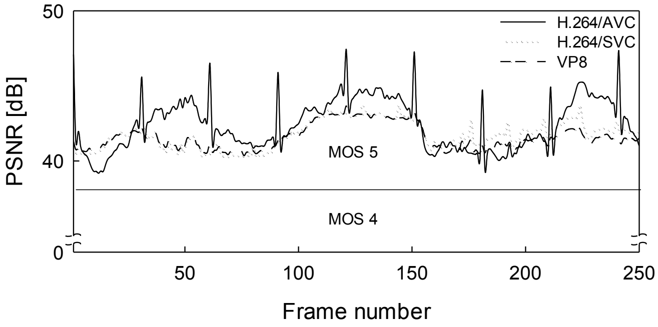


Figure 2. PSNR and MOS of encoded video data using the codecs

B. Simulation Scenarios

Table III shows the seven AC mapping scenarios used for our evaluation of differentiated transmission of the video stream over 802.11e. For DCF, all encoded video frames are mapped to AC0. For EDCA, all video frames are mapped to AC2. In TPA, important video fragments, such as I-frame or the base layer (BL), are mapped to AC2 and the others are mapped to AC1.

TABLE III. MAPPING TABLE IN 802.11E ACS

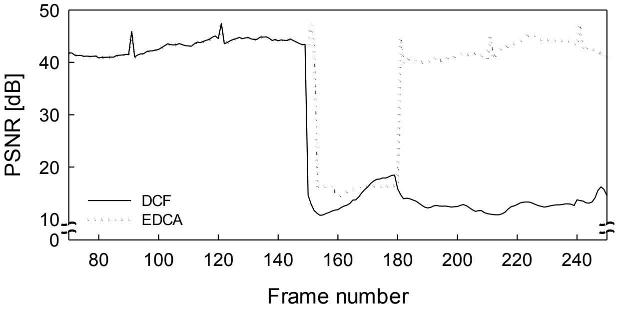
	AC3 (7)	AC2 (5)	AC1 (2)	AC0 (0)
DCF				All video frames
EDCA		All video frames		
TPA [3]		I-frame or BL	P-frame or ELs	
Scheme 1	I-frame or BL	P-frame or ELs		
Scheme 2	All video frames			
Scheme 3	I-frame or BL		P-frame or ELs	
Scheme 4	BL	EL 1	EL 2	

Schemes 1 through 4 utilize AC3, which is originally reserved for VoIP, to map encoded video data. In Scheme 1, I-frame or the base layer is mapped to AC3, and the rest of video fragments are mapped to AC2. In Scheme 2, all encoded video frames are mapped to AC3. For Scheme 3, P-frames or the enhancement layers (ELs) are mapped to the AC1. In Scheme 4, which is configured only for H.264/SVC, the base layer, enhancement layer 1, and enhancement layer 2 are mapped to AC3, AC2, and AC1, respectively. Note that background traffic is mapped to AC0.

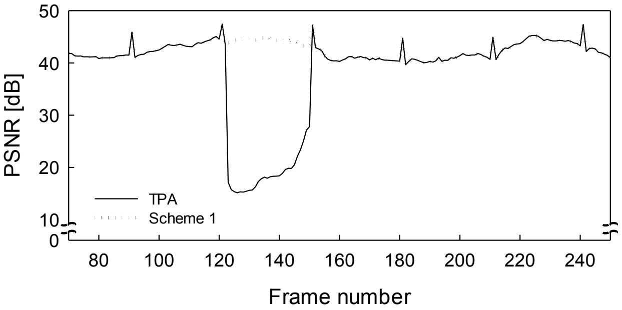
IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental Results

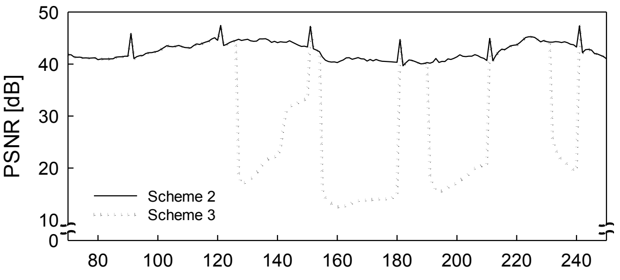
Fig. 3 shows the PSNR results for the decoded video frames between the 70th and the 250th frames for H.264/AVC. Note that PSNR values are calculated between the raw and received frames. With DCF, the 150th frame (an I-frame) is lost, and thus distortion begins from this frame and PSNR drops below 20 dB until the end of the video stream due to frequent frame losses. When EDCA is applied, video distortion occurs for about 1 second because of lots of P-frame drops but PSNR is restored after receiving the 180th frame, which is an I-frame. PSNR is better with TPA than with DCF, but it decreases for 1 second due to the loss of P-frames starting from the 121st frame. Thus, EDCA and TPA show similar average PSNR values of around 39 dB. In Schemes 1 and 2, there are no I-frame and P-frame losses with H.264/AVC. In contrast, PSNR decreases to below 20 dB for Scheme 3 because of frequent P-frame losses.



(a) DCF and EDCA



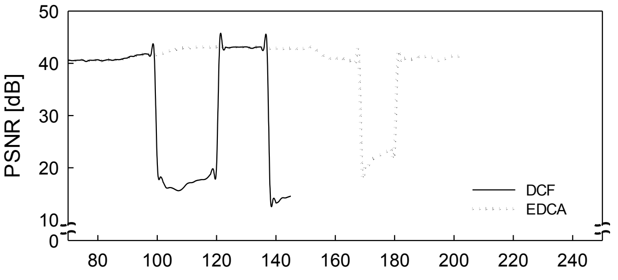
(b) TPA and Scheme 1



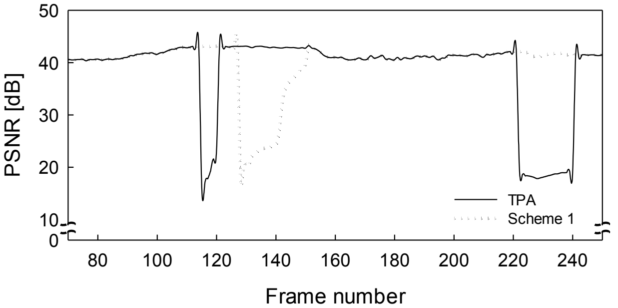
(c) Schemes 2 and 3

Figure 3. PSNR of H.264/AVC over IEEE 802.11 using the mapping schemes

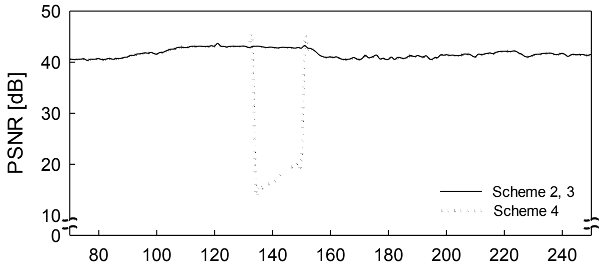
Fig. 4 shows the PSNR results for H.264/SVC. As shown in Fig. 4(a), distortion occurs for H.264/SVC with DCF starting at the 99th frame due to the consecutive P-frame losses. PSNR recovers after receiving the 120th frame (an I-frame). However, the 148th frame cannot be decoded due to the dropped base layer in spite of error resilience/concealment features. With EDCA, the 170th frame is lost and PSNR drops to 20 dB until the 180th frame is received. H.264/SVC with EDCA also cannot decode frames after the 204th frame. With TPA, all the received video frames are decoded, but frequent frame loss occurs between 225th and 273rd frames and PSNR decreases to 20 dB. As shown in Figs. 4(b) and 4(c), PSNR for Scheme 1 is lower than 20 dB between 130th and 150th frames. However, Schemes 2 and 3 show the best video quality without any layer drops. In Scheme 4, PSNR decreases between 134th and 150th frames due to the lost P-frames.



(a) DCF and EDCA



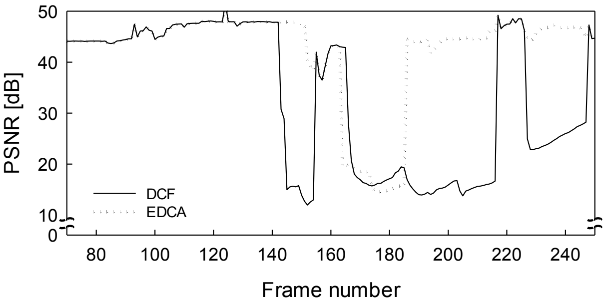
(b) TPA and Scheme 1



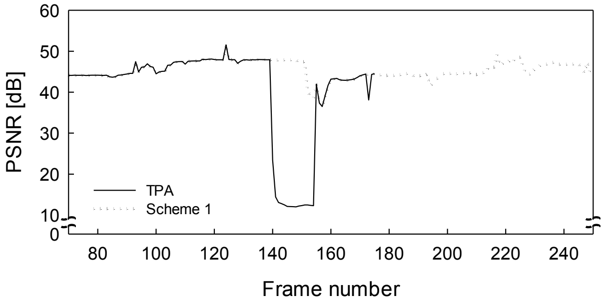
(c) Schemes 2, 3 and 4

Figure 4. PSNR of H.264/SVC over IEEE 802.11 using the mapping schemes

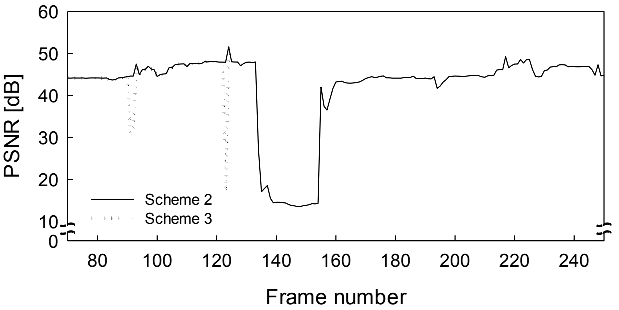
Fig. 5 shows the PSNR results for VP8. In Fig. 5(a), video distortion begins from the 133rd frame, and PSNR drops below 20 dB until the 250th frame because of frequent frame losses when VP8 used with DCF. Compared with DCF, PSNR for EDCA is 5 dB higher on average, but there is distortion for 1 second after the 163rd frame. For VP8 with TPA, video freezes at the 183rd frame even though the other frames are received correctly. Figs. 5(b) and 5(c) show that Scheme 1 achieves the best PSNR without any lost frames. Scheme 2 also shows good PSNR even though some frames are lost. However, video freezes with Scheme 3 because of lots of frame losses. By default, VP8 generates intra-frames every 300 frames. However, the intra-frame interval of VP8 is configured to 30, which is equivalent to the I-frame interval of H.264/AVC and H.264/SVC. To avoid the video freeze problem, the intra-frame interval should be configured to be as long as possible within the default range of 1 to 300 frames, which implies that intra-frame interval should be 300. The video freeze problem does not occur when the intra-frame interval is set to 300.



(a) DCF and EDCA



(b) TPA and Scheme 1



(c) Schemes 2 and 3

Figure 5. PSNR of VP8 over IEEE 802.11 using the mapping schemes

B. *Analysis*

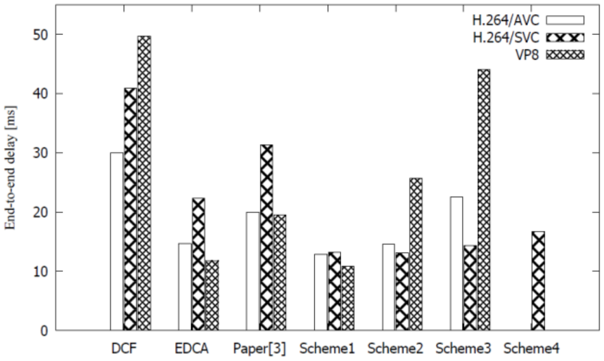
Table IV summarizes the size of encoded video data, the size of I-frame or the base layer data, the total number of packets, and the number of I-frame or the base layer packets for each codec. Note that the size of Maximum Transmission Unit (MTU) is set to 1500 bytes.

TABLE IV. SIZE OF ENCODED VIDEO AND NUMBER OF PACKETS

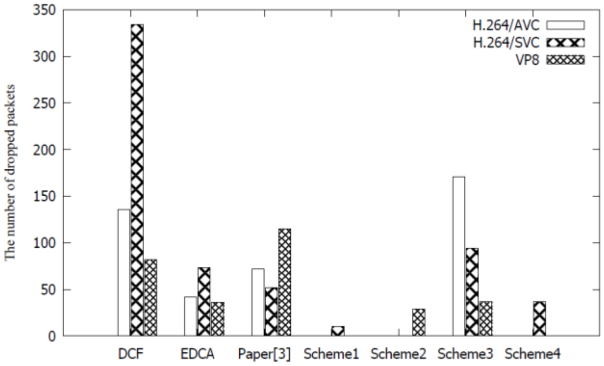
	H.264/AVC	**H.264/SVC**	**VP8**
Total size of encoded video data (Total number of packets)	5,212 KB (3,739)	5,377 KB (4,276)	4,163 KB (3,018)
Total size of I-frame or base layer data (Total number of packets)	434 KB (302)	2,433 KB (1,833)	336 KB (237)

Video data causes heavy traffic, and thus bursts of packet losses may occur. Although H.264/AVC and H.264/SVC generate similar sized encoded video data, the size of the H.264/SVC base layer is roughly six times larger than that of H.264/AVC I-frame. The base layer is often lost which eventually degrades the video quality. The average sizes of I-frames encoded by H.264/AVC and VP8 are 43 Kbytes and 36 Kbytes, respectively. The difference between the two I-frame sizes is one of the key factors for achieving better video quality for VP8 with DCF and EDCA. On the other hand, the ratio of I-frames or the base layers to the entire video data is 8.3%, 45%, and 8% for H.264/AVC, H.264/SVC, and VP8, respectively. Therefore, it is difficult to support QoS with H.264/SVC due to the high proportion of data for the base layer. In addition, the base layer is essential for decoding video with H.264/SVC [7], thus it must be received safely over networks. However, there is no guarantee that the transmitted packets will be successfully received. As a consequence, H.264/SVC is not appropriate for real-time transmission over IEEE 802.11 compared to H.264/AVC and VP8 because the base layer may not be received even with IEEE 802.11e.

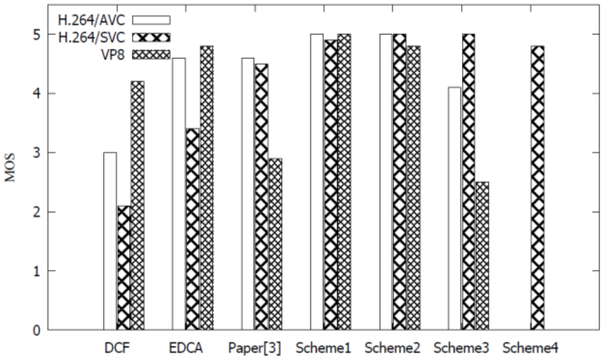
Figs. 6(a)-6(c) illustrate the end-to-end delay, the number of dropped packets, and MOS for the codecs. Fig. 6(a) shows that Scheme 1 has the lowest end-to-end delay due to higher priority mapping of ACs. This is because the highest priority AC has the shortest Arbitration Inter Frame Space (AIFS) allowing packets to quickly access the channel.



(a) End-to-end delay



(b) The number of dropped packets



(c) MOS

Figure 6. Network status for video data over IEEE 802.11

As shown in Fig. 6(b), the number of dropped packets in H.264/SVC with DCF is three times higher than that of H.264/AVC and VP8. In case of Scheme 1, there are ten dropped packets for H.264/SVC while no packets are dropped with H.264/AVC and VP8. As shown in Table IV, the number of packets with H.264/SVC is more than that of H.264/AVC and VP8. While Scheme 1 is applied with H.264/SVC, the quality of video delivery is not guaranteed because the total size of video data in H.264/SVC is more than that of H.264/AVC and VP8. Thus, the exceeding amount of the base layer data incurs AC queue overflow resulting in the video quality degradation. Therefore, the queue size of each AC should be carefully considered when AC mapping is applied. In Scheme 3, the number of dropped packets for H.264/AVC increases to 171, which decreases the average MOS to 4. Therefore, Figs. 6(b) and 6(c) show that burst of packet drops that causes video distortion is an obstacle to guarantee video quality. In addition, reducing the amount of encoded video data will help to support QoS because a large amount of video data generates burst of packet losses in WLANs. As a result, VP8 with DCF and EDCA shows decent user-perceived quality, but H.264/AVC and H.264/SVC can achieve better user-perceived quality utilizing AC mapping schemes in 802.11e as shown in Fig. 6(c).

V. CONCLUSION

In this paper, we evaluated the quality of video delivery using the latest video codecs over WLANs. In order to evaluate the quality of video transmission, we simulated various video delivery scenarios based on the several AC mapping schemes using the OEFMON. Even though the results may be controversial, we get the meaningful results based on conventional/referenced products such as H.264/AVC, H.264/SVC and VP8. Our results showed that reducing the coded video data, considering the queue size of each ACs, and the mandatory implementation of error recovery features for H.264/SVC facilitates a more robust multimedia transmission over WLANs. Moreover, we compared the performance of the latest video codecs on the various mapping schemes.

As a future work, we will carry out experiments with more test sequences, resolutions and target bitrates. Through those various experiments, we plan to develop AC mapping schemes for encoded video delivery over the wireless networks that considers the performance and characteristic of various codecs. In order to obtain realistic results, we will perform our evaluation using a real testbed.

REFERENCES

[1] A. Ksentini, M. Naimi, and A. Gu´eroui, “Toward an improvement of H.264 video transmission over IEEE 802.11e through a cross-layer architecture,” IEEE Communications Magazine, vol. 44, January, 2006.

[2] U. I. Choudhry and J. W. Kim, “Performance Evaluation of H.264 Mapping Strategies over IEEE 802.11e WLAN for Robust Video Streaming,” Lecture Notes in Computer Science, Volume 3768/2005, 818-829, 2005.

[3] A. Fiandrotti, D. Gallucci, E. Masala, and E. Magli, “Traffic prioritization of H.264/SVC video over 802.11e ad hoc wireless networks,” IEEE ICCCN 17th International Conference on Computer Communications and Networks, 2008.

[4] Y. J. Liang, J. G. Apostolopoulos, B. Girod, “Analysis of Packet Loss For Compressed Video: Does Burst-Length Matter?,” IEEE International Conference on Acoustics, Speech and Signal Processing, Hong Kong, April 2003.

[5] IEEE Std 802.11e-2005, Amendment 8: Medium Access Control Quality of Service Enhancements, Nov. 2005.

[6] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H. 264/AVC video coding standard," IEEE Transactions on Circuits and Systems for Video Technology, 2003.

[7] H. Schwarz, D. Marpe and T. Wiegand "Overview of the Scalable H.264/MPEG4-AVC Extension," The IEEE International Conference on Image Processing, 2006.

[8] WebM Project, http://www.webmproject.org/

[9] C. Lee, M. Kim, B. Lee, S. J. Hyun, K. Lee and S. Lee, "OEFMON: An Open Evaluation Framework for Multimedia Over Networks", IEEE Communications Magazine, Vol. 49, No. 9, Sep. 2011, pp.153-161.

[10] *x264*, http://www.videolan.org/

[11] *MONOGRAM*, http://blog.monogram.sk/janos/

[12] *JSVM*, http://ip.hhi.de/imagecom_G1/savce/downloads/SVC-Reference-Software.htm

[13] Methodology for the subjective assessment of the quality of television pictures, ITU-R BT.500-1, 2002.

[14] FFmpeg, http://www.ffmpeg.org/