

# Video Capacity of WLANs with a Multiuser Perceptual Quality Constraint

Jing Hu, Sayantan Choudhury and Jerry D. Gibson

Department of Electrical and Computer Engineering  
University of California, Santa Barbara, California 93106-9560  
Email: {jinghu, sayantan, gibson}@ece.ucsb.edu

**Abstract**—As wireless local area networks (WLANs) become a part of our network infrastructure, it is critical that we understand both the performance provided to the end users and the capacity of these WLANs in terms of the number of supported flows (calls). Since it is clear that video traffic, as well as voice and data, will be carried by these networks, it is particularly important that we investigate these issues for packetized video. In this paper, we investigate the video user capacity of wireless networks subject to a multiuser perceptual quality constraint. As a particular example, we study the transmission of AVC/H.264 coded video streams over an IEEE 802.11a WLAN subject to a constraint on the quality of the delivered video experienced by  $r\%$  (75%, for example) of the users of the WLAN. This work appears to be the first such effort to address this difficult but important problem. Furthermore, the methodology employed is perfectly general and can be used for different networks, video codecs, transmission channels, protocols, and perceptual quality measures.

## I. INTRODUCTION

### A. Motivation

Recently there has been broad interest in using packetized video over wireless networks such as IEEE 802.11 wireless local area networks (WLANs). As a result, there have been numerous studies related to video over WLANs [1]. In particular, research has been performed on cross-layer designs for video over WLANs, including: (a) designing the network matched to the special characteristics of video [2]–[4], (b) compressing and transporting video adaptively with respect to the lower layers in the OSI stack of the network [5]–[12], and (c) solving the cross-layer design problem as an optimization problem with the objective of selecting a joint strategy across multiple OSI layers [13]–[15]. However, the video capacity of a WLAN in terms of the maximum number of video users the WLAN can support, a fundamental limit in video communications over WLANs, has received relatively little attention. The reasons are as follows.

First, videos can be compressed to essentially any desired bit rate, resulting in different reconstructed video quality, and therefore the investigation of video capacity should always be accompanied by a quality constraint. However, video quality

measurement is a very difficult problem on its own. The conventional measures such as the mean squared error (MSE), or equivalently the peak signal-to-noise ratio (PSNR) of the distorted videos, are often criticized for correlating poorly to perceptual video quality. On the other hand, the objective perceptual video quality measures that are based on the lower order processing of human vision systems (HVS) [16], [17] are computationally very intensive [18], [19]. Furthermore, in the situation when multiple users are using the same network, which is inevitably the case when video capacity is under investigation, the assessment of the quality of multiple videos delivered over the network has not been studied.

Second, even if the quality of multiple video users is defined, it is not clear how video capacity should be calculated. Ideally, rate distortion bounds for videos will produce the lowest rate that is required to compress a video subject to a certain distortion constraint. However, due to the difficulty of modeling the correlation among the pixel values in natural video sources, studying their theoretical rate distortion bounds is often considered infeasible [20], [21]. Therefore the lowest rate required for a certain video quality depends on the specific video codec that is employed. This is also true for voice, the capacity of which has been investigated extensively in [22]; but unlike voice where codecs are usually designed for a specified quality at a given compression rate, the video coding standards do not specify a compressed video bit rate and corresponding quality. Instead, the international video coding standards [23]–[26] provide vast flexibility for each application to design its own encoder according to its specific compression requirement. Rate control (RC) [27]–[32] and rate distortion optimization (RDO) [20], [21], [33], [34] techniques have been included to optimize the video codecs to compress videos with minimum distortion at a certain bit rate. Conventionally, RC selects the quantization stepsizes and RDO selects the intra/inter prediction modes, motion vectors and other coding parameters. The RC/RDO problem itself is a subject of intensive study, and unfortunately, the separate treatment of RC and RDO and the interdependency between them have caused a “chicken and egg” dilemma that has prevented a global optimum from being obtained [35].

### B. Contributions

In this paper we address these aforementioned challenges in deriving the quality constrained video capacity of WLANs.

This work was supported by the California Micro Program, Applied Signal Technology, Cisco, Inc., Dolby Labs, Inc., Marvell, Inc. and Qualcomm, Inc., by NSF Grant Nos. CCF-0429884 and CNS-0435527, and by the UC Discovery Grant Program and Nokia, Inc..

Since a video codec has to be specified when studying video capacity of a WLAN, we choose AVC/H.264 because of its performance and dominant popularity for video over WLAN applications [36]. Among the different WLAN technologies, we are particularly interested in IEEE 802.11a WLANs [37] in a realistic frequency selective multipath fading environment. The IEEE 802.11a WLAN operates in the relatively clean 5 GHz frequency band and uses 52-subcarrier orthogonal frequency-division multiplexing (OFDM) with a maximum raw data rate of 54 Mbit/s. It is a widely available WLAN solution.

We start our investigation with a simulation of AVC/H.264 coded video over an 802.11a WLAN with multipath fading. From the simulation we get a collection of videos whose compressed data rate versus delivered quality performance is independent of a specific RC/RDO algorithm or a specific channel coding algorithm, and can be exploited to derive a multiuser video quality indicator and video capacity formulas. These proposed methodologies, however, can be generalized to other video over WLAN applications that are different from the simulation scenario that we investigate.

In this simulation, it is observed that even when the average PSNR over all transmitted frames of a video with packet losses is reasonably high, PSNRs vary significantly across the video frames. Furthermore, the video quality varies dramatically across the different transmissions over the channel. In order to capture the distribution of the distortion across the video frames and channel uses (transmissions, or realizations), we propose a new statistical video quality indicator  $PSNR_{r,f}$  as the PSNR achieved by  $f\%$  of the frames in each one of the  $r\%$  of the transmissions. This quantity has the potential to capture the performance loss due to damaged frames in a particular video sequence ( $f\%$ ), as well as to indicate the probability of a user experiencing a specified quality over the channel ( $r\%$ ). The percentage of transmissions also has the interpretation as what percentage out of many video users who access the same channel, would experience a given video quality. We further investigate the correspondence between  $PSNR_{r,f}$  and perceptual video quality through a subjective experiment which results in a linear equation connecting  $PSNR_{r,f=90\%}$  and  $MOS_r$ , the mean opinion score (MOS) achieved by  $r\%$  of the transmissions. It is shown from this subjective experiment that  $PSNR_{f=90\%}$  correlates much better with the delivered perceptual video quality than the average PSNR across all frames of a video, with no extra computation.

There are more sophisticated perceptual video quality measures, such as those included in ITU recommendations ITU-R BT.1683 and ITU-T J.144 [16], [17], for calculating the MOS of a single video sequence. These sophisticated perceptual video quality measures are based on comprehensive studies of the human vision system (HVS). They study the perceptual impacts of the compression artifacts and are shown to perform better than the average PSNR across the video frames for the compressed videos [38]. For video delivered over WLANs, however, the quality degradation due to compression can be overwhelmed by the quality degradation caused by the possible

packet losses in the wireless channel, even with packet loss concealment. As a result, the MOS calculated from  $PSNR_{r,f}$  should be sufficient to indicate the perceptual quality of a delivered video sequence, without the huge computation required by the more sophisticated video quality measurements.

On the video capacity side, due to the significant difference in the intra-coded and inter-coded frame sizes of a compressed video, we formulate upper and lower bounds for video capacity of an 802.11a WLAN operated under the Distributed Coordination Function (DCF), when there is no buffering at the receiver. We further investigate the video capacity when there is buffering at the receiver and obtain the minimum buffer size for the video capacity to reach its upper bound. These results appear to be the first work on the video capacity of WLANs.

Combining the above contributions, we propose a methodology for video over WLAN communication system design and evaluation, which consists of determining the video capacity in the context of the delivered video quality constraints calculated by  $PSNR_{r,f}/MOS_r$ . Practical issues such as the usage of a specific rate control and rate distortion optimization scheme and a specific channel coding scheme are also discussed.

### C. Organization of this paper

The remainder of this paper is organized as follows. The video over WLAN simulation setup is explained in detail in Section II. Section III presents the results of the simulation in terms of delivered quality and coded video data rate. The results on the distribution of the delivered video quality then motivates the proposal of a new video quality measure  $PSNR_{r,f}$  and its corresponding new multiuser perceptual quality indicator  $MOS_r$  in Section IV. Section V presents the video capacity calculation under 802.11a DCF. Due to the significant difference in the intra-coded and inter-coded frame sizes of a compressed video shown in the coded data rate part of Section III, the timing of intra-coded frames among different video users is studied carefully in the video capacity calculation. We discuss the video capacity jointly with the perceptual quality constraint as well as some application issues in Section VI. Section VII concludes and lists the key contributions of this paper.

## II. VIDEO OVER WLAN SIMULATION SETUP

We simulate the transmission of AVC/H.264 coded video over 802.11a WLANs. AVC/H.264 provides a huge selection of coding schemes and values for their parameters. As illustrated in Fig. 1, the center of this diagram is a simplified AVC/H.264 encoder and the options for the major schemes and parameters are presented in the callout balloons. We choose the Baseline Profile of AVC/H.264 in its reference software [39] version JM10.1 with low delay and low computational complexity. Ninety frames each from a group of videos, representing different types of video content, are coded using combinations of group of picture sizes (GOPs) (10, 15, 30, 45 frames), quantization parameters (QP) (26 for fine quantization and 30 for coarse quantization) and payload sizes (PS) (small-100 bytes and large-1100 bytes). QP dominates the

quantization error and has a major effect on the coded video data rate. GOPS determines the intra-frame refresh frequency and plays an important role when there is packet loss. PS is the parameter that is carried forward from the source to the PHY layer. The remainder of the parameters to be selected in Fig. 1: the intra-mode, block size and inter-frame prediction precision are optimally chosen in the encoder to yield the minimum source bit rate. We do not employ rate control schemes to dynamically choose QPs to compress the video sequences at a constant bit rate. Instead we focus on a constant QP case. This is further justified by the preference of constant video quality over constant bit rate [40], especially when the bandwidth allocated to a video user need not be constant, which is likely to be the case for WLANs when QoS is enforced. In Section VI we discuss how the results of this paper can be used even when a specific rate control or a specific rate distortion optimization algorithm is employed.

The Nafteli Chayat model [41], an important indoor wireless channel model with an exponentially decaying Rayleigh faded path delay profile, is employed. The rms delay spread used is 50 nanoseconds, which is typical for home and office environments. In order to estimate the packet error rate under different channel conditions, we modify a readily available OFDM simulator for the IEEE 802.11a PHY [42]. Non-fading channels are also considered for comparison. Noise is modeled as AWGN for both the fading and non-fading cases. The decoding at the receiver is based on soft decision Viterbi decoding. We also assume perfect synchronization and channel estimation.

We consider one-hop WLANs, in which case we limit our attention to the PHY, MAC and APP layers. In the medium access control (MAC) layer of IEEE 802.11, a cyclic redundancy check (CRC) is computed over the entire packet, and if a single bit error is detected, the packet is discarded. For data, a retransmission would be requested, however, for our particular video applications we do not request a retransmission, but rely on packet loss concealment. Each realization of the multipath delay profile corresponds to a certain loss pattern for that fading realization. Two hundred and fifty packet loss realizations are generated for each combination of the chosen PHY data rate 6 Mbps, different average channel SNRs (3.5 dB for bad channel, 5 dB for average channel, 7 dB for good channel), and two video PSs (small–100 bytes and large–1100 bytes).

Each compressed bit stream is corrupted based on the packet loss patterns generated by the multipath fading channel and then reconstructed in the AVC/H.264 decoder with its nominal packet loss concealment (PLC) scheme. Different PLC schemes will have an impact on the concealed video quality and there exists an exhaustive literature proposing different error concealment techniques. However, only a few simple schemes are commonly used in practical applications [5]. As a baseline, we apply the basic PLC method integrated in AVC/H.264 reference software [39]. This PLC method recovers the missing MBs in an I frame through spatial interpolation and the missing MBs in a P frame by searching

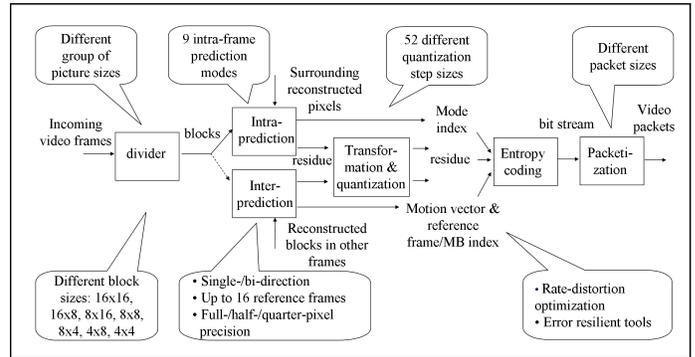


Fig. 1. A simplified diagram of AVC/H.264 encoder with different coding options and parameters

and copying the most likely MBs in the correctly received reference frames. The previous frame is copied when the whole frame is lost. This method is shown to be effective in both PSNR and perceptual quality [43].

In summary, in this AVC/H.264 coded video over IEEE 802.11a WLAN simulation, we investigate six different parameters across the APP, MAC and PHY layers of the OSI stack. They are the characteristics of a video (VIDEO), quantization parameter (QP), group of picture size (GOPS), payload size (PS), PHY data rate, and average channel SNR. They are listed in the center of Fig. 2. Among these six parameters, PS, PHY data rate and the average channel SNR have an impact on the packet loss realizations, which together with the other three parameters, decide the delivered video quality. On the other hand, the four parameters on the top: VIDEO, QP, GOPS and PS decide the coded video data rate, which in turn determines the video capacity of a certain PHY data rate.

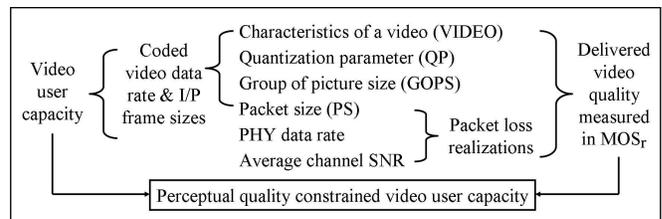


Fig. 2. Parameters investigated in video over WLAN simulation and their impacts on the video capacity vs. delivered quality tradeoff

### III. VIDEO OVER WLAN SIMULATION RESULTS

In this section we take a close look at the video over WLAN simulation results that motivate the proposals of the new video quality indicator and the video capacity bounds in the next two sections.

#### A. Packet loss and video quality

Figure 3 plots the cumulative distribution function (cdf) of the packet error rate (PER) over 250 realizations of each channel for 100 byte and 1100 byte packets in a multipath fading environment at average channel SNRs of 3.5 dB, 5 dB

and 7 dB when the 6 Mbps PHY data rate is used. The cdf of PER for 100 byte packets and a channel SNR of 0.5 dB for an AWGN channel is also plotted (the top curve in this figure, marked with squares '□'). It shows that for the same channel SNR and the same PS, the PER of an individual multipath fading channel realization can range from 0 to 1, since the cdf's of the PER of all fading channels are greater than 0 at PER = 0 and less than 1 for PER < 1. This suggests that the average PER over all realizations of a multipath fading channel is not an appropriate indicator of the channel performance. For example, for the multipath fading channel with PS = 100 bytes and average SNR = 7 dB (the second curve from top, marked with circles 'o' in Fig. 3), the average PER across the realizations is 5.5%, but this average PER across the realizations only represents the PER of a very small portion of the realizations. As we can see from this figure: 70% of the realizations actually have no packet loss; another 20% of the realizations have packet loss less than 2%; and 10% of the realizations have a higher PER than 5.5%. This is less a problem for AWGN channels, since we see that the variation of the PER of the AWGN channel is significantly less, only from 1% to 3% in this figure. Figure 3 also shows that the average PER of an AWGN channel is much lower than that of a multipath fading channel even at a much poorer channel SNR. Furthermore, for multipath fading channels, the 1100 byte packets are more likely to be lost than the 100 byte packets.

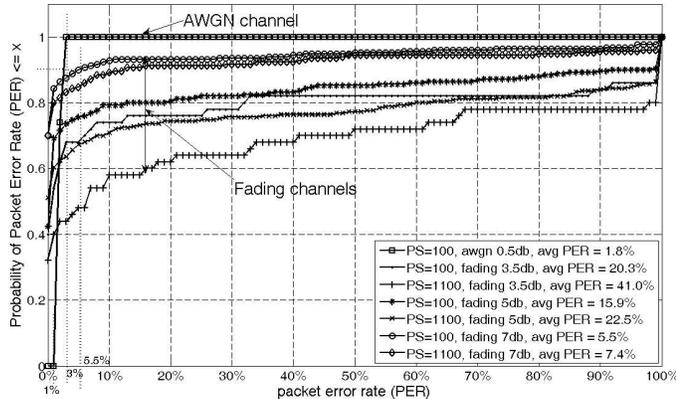


Fig. 3. Cumulative distribution function (cdf) of packet error rate over 100 realizations of each channel in AWGN and multipath fading environments for 100 byte and 1100 byte packets and PHY data rate as 6 Mbps

We also study the performance variation in a single channel realization and obtain a PSNR for each frame and each packet loss pattern, for a combination of the codec parameters. Only the PSNR of the luminance component of the video sequences is considered and the peak signal amplitude picked in this paper is 255. Figure 4 plots the PSNRs of each frame of the video *silent.cif* coded at QP = 26 and 30, GOPS = 15, PS = 100 for 100 realizations of the multipath fading channel of average SNR 7 dB and AWGN channel of SNR 3 dB, respectively, when PHY data rate 6 Mbps is used. The thick lines in each

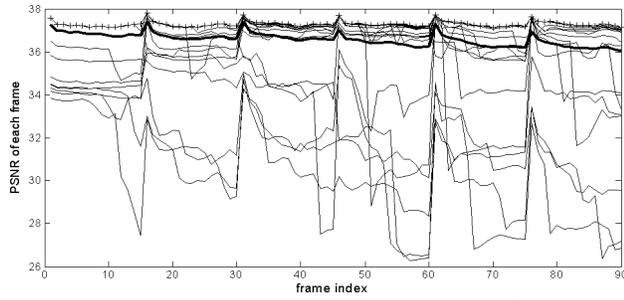
plot represent the average PSNRs across the 100 channel realizations. This average should be slightly different than the PSNR calculated from averaging the MSEs. In practice, however, there is no significant difference between the two definitions [44].

It is clear in Fig. 4 that even for the same video, coded using the same parameters for the same average channel SNR, the quality of the delivered video in terms of PSNR varies significantly across different channel realizations. The plots in Fig. 4 are typical for all of the videos and codec parameters we tested. PSNRs also can vary dramatically from one frame to another in the same processed video sequence. From Fig. 3 we know that for the multipath fading channel, about 70% of the realizations have no packet loss. These realizations overlap and form the lines marked with “+” in Figs. 4(a) and 4(c). For the AWGN channel, all realizations have similar PERs. However, because of the prediction employed in video coding, it is shown in Figures 4(b) and 4(d) that the realizations of similar PER can generate completely different concealed video quality. This suggests that neither the average PER, nor the average PSNR across all the frames and all the realizations, is a suitable indicator of the quality a video user experiences and therefore these quantities should not serve as the basis for developing or evaluating video communications schemes for WLANs.

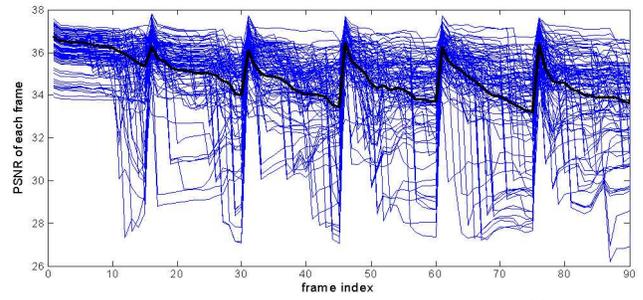
### B. Coded video data rate

In order to study the number of video users that can be supported by any network, we need to study the coded/compressed video data rate. Different from other types of signals such as voice, the coded video data rate depends strongly on the properties of individual videos. Also, the sophisticated video codecs such as the AVC/H.264 standard, are suites of coding options and parameters whose values are to be chosen for specific videos and channel conditions. This flexibility inevitably shapes the coded video data rate, and hence, alters the number of video users a WLAN can support. In the following we study the compressed intra (I) frame and inter (P) frame sizes of a video which are shown later to be critical in the video capacity calculation.

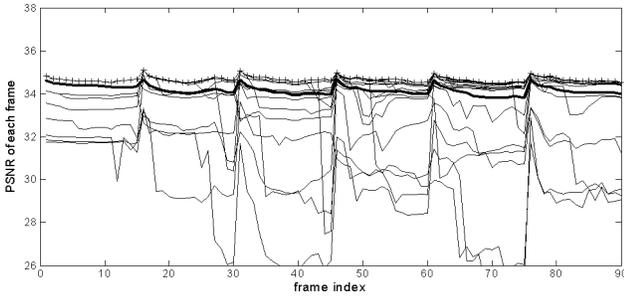
When all the other parameters are fixed, the coded I-frame size of a video mostly depends on the complexity in a scene, and the coded P-frame size depends on the motion across the frames. In Fig. 5 we plot the coded frame size of each of 90 frames of three different videos: *silent.cif*, *paris.cif* and *stefan.cif*. All of these videos are coded using QP = 22, PS = 100, and GOPS = 10. First, we can see that the I frames (the 1<sup>st</sup>, 11<sup>th</sup>, 21<sup>st</sup>, ... frames) consume a much higher bit rate than the P frames (the remaining frames). Second, although the I-frame and P-frame sizes do vary throughout the 90 frames of each video, they stay close to a certain level for each video. Third, among the three videos, *stefan.cif*, which is a sports video of a tennis player in the foreground and the audience in the background, has the busiest scene and the highest motion; *silent.cif* which is a head-and-shoulders video, encountered regularly in videoconferencing, has the simplest scene and the



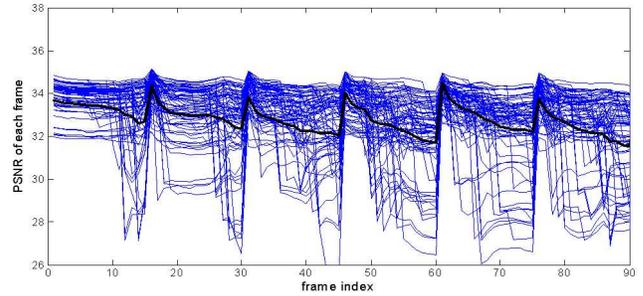
(a)  $QP = 26$ ,  $fading@7dB$ ,  $avgPER = 5.5\%$



(b)  $QP = 26$ ,  $AWGN@3dB$ ,  $avgPER = 1.5\%$



(c)  $QP = 30$ ,  $fading@7dB$ ,  $avgPER = 5.5\%$



(d)  $QP = 30$ ,  $AWGN@3dB$ ,  $avgPER = 1.5\%$

Fig. 4. PSNRs of each frame of the video *silent.cif* coded at  $GOPS = 15$ ,  $PS = 100$  for 100 realizations of multipath fading channel of average SNR 7 dB and an AWGN channel of SNR 3 dB, when PHY data rate 6 Mbps is used. The thick lines in each plot represent the average PSNRs across the 100 channel realizations

lowest motion; *paris.cif* with two people talking with relatively stable background, most likely to be a news streaming video, has intermediate scene complexity and intermediate level of motion. The coded bit rate difference between *silent.cif* and *stefan.cif* shown in this figure is as large as around 1 Mbps, which cannot be overlooked when video capacity is under investigation.

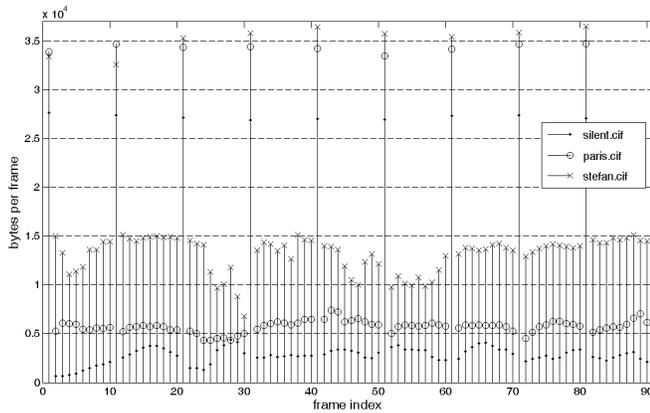


Fig. 5. Coded frame size of each frame of three videos compressed at  $QP = 22$ ,  $PS = 100$ ,  $GOPS = 10$

The  $GOPS$  decides the I-frame refresh rate, and therefore the smaller the  $GOPS$ , the higher the coded video data rate. However we notice that the difference in coded P-frame sizes when different  $GOPS$ s are employed is negligible. For example

in Fig. 6, we plot the coded frame size of each frame of *stefan.cif* compressed at  $QP = 26$ ,  $PS = 100$  and different  $GOPS$ s. We can see that if a frame is inter-coded, its coded frame size is almost unaffected by its distance to the previous intra-coded frame. This happens because quantization is the major lossy compression process in the encoder, and therefore the same frame compressed using the same  $QP$  will have similar reconstructed frames so that the next frames predicted from the similar reconstructed frames will have similar frame sizes as well.

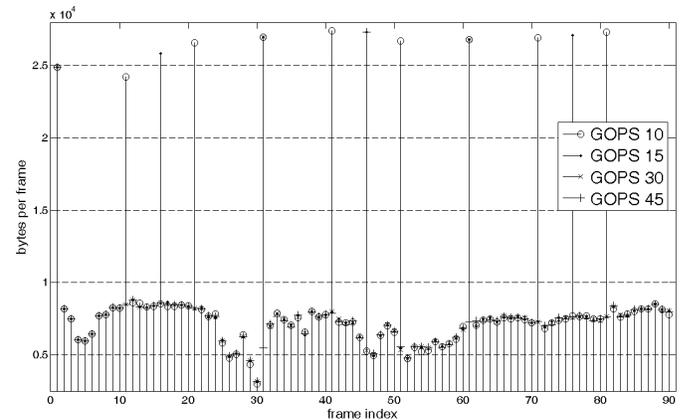


Fig. 6. Coded frame size of each frame of *stefan.cif* compressed at  $QP = 26$ ,  $PS = 100$  and different  $GOPS$ s

#### IV. DEFINITION OF $PSNR_{r,f}$ AND ITS CORRESPONDENCE TO PERCEPTUAL QUALITY OF MULTIPLE USERS $MOS_r$

As shown in Section III-A and in particular in Fig. 4, for video communication over WLANs, the PSNRs of the delivered videos vary significantly across the video frames and across the different realizations of the channel. In order to capture the distribution of the distortion across the video frames and channel uses, in this section we propose a statistical PSNR based video quality measure,  $PSNR_{r,f}$ , which is defined as the PSNR achieved by  $f\%$  of the frames in each one of the  $r\%$  of the realizations. Parameter  $r$  captures the reliability of a channel over many users and can be set as a number between 0% to 100% according to the desired consistency of the user experience.

The proposal of using  $PSNR_f$ , i.e., the lowest PSNR achieved by  $f\%$  (usually set as a majority) of the frames in a single video sequence, to measure the perceptual quality of a single video sequence is based on three observations that are recognized by researchers in video quality assessment [19]: 1) the frames of poor quality in a video sequence dominate human viewers' experience with the video; 2) however, if only a very small portion of the video frames are of poor quality, the quality drop due to these few frames are not perceivable by the human viewers; 3) when the PSNRs are higher than a threshold, increasing PSNR does not correspond to an increase in perceptual quality that is already excellent at the threshold.

To confirm these observations, i.e., to study the correlation between  $PSNR_f$  and the perceptual quality of videos, as well as to find a suitable range for the parameter  $f$ , a subjective experiment is designed and conducted. Stimulus-comparison methods [45] are used in this experiment, where two video sequences of the same content were presented to the subjects side by side and were played simultaneously. The video on the left is considered to be of perfect quality while the video on the right is compressed and then reconstructed with possible packet loss and concealment. Three naive human subjects are involved in this experiment. They are asked to pick a number representing the perceptual quality of the processed video compared to the perfect video from the continuous quality scale shown in Figure 7. Fifty video pairs were tested and 20% of them appear twice in this experiment to test the consistency of the subjects' decisions.

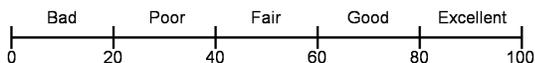


Fig. 7. Perceptual video quality scale in MOS

Figure 8 plots the opinion scores given by the three subjects in circles ('o'), dots ('.') and crosses ('+'). Of the 50 tested videos, 18 are silent.cif, reconstructed from different levels of packet losses. They are arranged from left to right with ascending average of three subjects' opinion scores. The same is done for the 16 videos of paris.cif and the 16 videos of stefan.cif. For each tested video, the PSNRs are calculated for

each frame, from which both average PSNR across all frames and  $PSNR_f$  with  $f$  as any value can be further calculated. Since the PSNRs and the opinion scores are of different scales, in order to compare them, the average PSNR and  $PSNR_f$  with  $f$  ranging from 0.5 to 0.99 are mapped to the opinion scores through linear functions which yield the minimum mean square errors in the fit.

We find that among all the values of  $f$  we investigate,  $PSNR_f$  with  $f = 90\%$  correlates to the opinion scores the best, whose linear mapping is plotted as solid lines in Fig. 8. We also plot the best linear mapping of average PSNRs in dashed lines for comparison. As seen from these curves,  $PSNR_{f=90\%}$  correlates significantly better than average PSNR, to the perceptual quality for all three videos that are given in circles ('o'), dots ('.') and crosses ('+') for each video. The average PSNR underestimates the quality at high quality level and overestimates the quality at low quality level. This is because average PSNR treats all frames equally. At high quality level, however, only a few frames with relatively lower quality bring down the average PSNR but do not affect the perceptual quality. At low quality level, on the other hand, there are frames with extremely bad quality which affect the overall video quality significantly while the average PSNR is still quite high. This subjective experiment shows that  $PSNR_{r,f}$  can serve as an effective video quality measure, and that  $f$  should be set around 90% for medium video frame rates, such as 15 fps used in this paper. In this case the linear mapping from  $PSNR_{r,f=90\%}$  to  $MOS_r$ , the mean opinion score (MOS) achieved by  $r\%$  of the transmissions, is

$$MOS_r = 19 + 3.6(PSNR_{r,f=90\%} - 19). \quad (1)$$

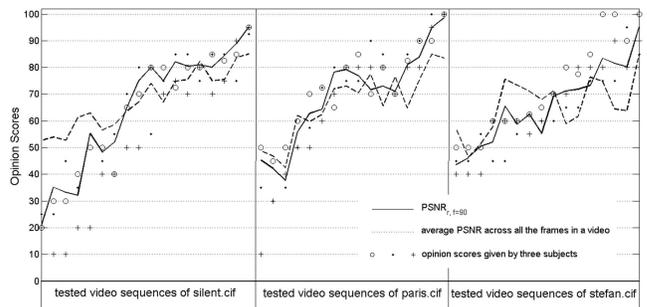


Fig. 8. The opinion scores given by the three subjects and the the best linear mappings of  $PSNR_{f=90\%}$  and average PSNR. Of the 50 tested videos, 18 are silent.cif, 16 are paris.cif and 16 are stefan.cif, reconstructed from different levels of packet losses. They are arranged from left to right with ascending average of three subjects' opinion scores.

From here onwards in this paper, we call the new multiuser perceptual video quality indicator  $PSNR_{r,f}/MOS_r$ .  $PSNR_{r,f}$  focuses on the distribution of the video quality across the video frames and channel uses, while  $MOS_r$  also provides guidance on the perceptual quality across different users. The  $MOS$  in  $MOS_r$  can be calculated from

$PSNR_{f=90\%}$  using Eq. (1). More subjective experiments involving more human subjects need to be conducted to confirm Eq. (1). The definition of  $PSNR_{r,f}/MOS_r$  is motivated by the AVC/H.264 coded video over IEEE 802.11a WLAN simulation, but this indicator is independent of the simulation setup and can be exploited in different video communication systems. In real video communication systems, the receiver can be set up to monitor the decoded video quality at regular intervals. With a fixed frame rate of transmitted video, it is then straightforward to calculate the number of frames under investigation for that period of time. Here we briefly discuss an example of how  $PSNR_{r,f}/MOS_r$  can be used.

In Fig. 9 we plot the  $MOS_r$  of three videos coded by AVC/H.264 using QP = 26, GOPS = 10, PS = 100 bytes and transmitted over an 802.11a WLAN with a PHY data rate of 6 Mbps at average channel SNRs of 5 and 7 dB, respectively. In the 7 dB channel (the three curves on the right) for example, if all users are assumed to be communicating the same type of videos and an 80% consistency in user experience is desired, i.e.,  $r=0.8$ , the videoconferencing users (silent.cif) experience an “excellent” video quality; the news watchers (paris.cif) experience a “good” video quality, but the sports fans only receive “bad” quality videos, corresponding to a MOS of 30 out of 100, which is 40 to 50 points lower than those of the other two groups of users. This information can then be utilized for link adaptation, system performance evaluation, or system design purposes. For example, depending on the type of videos a specific communication system targets, a lower PHY data rate might need to be used, if one is available, in order to achieve a good user experience. We discuss in more detail the multiuser perceptual video quality indicated by  $PSNR_{r,f}/MOS_r$  together with the video capacity bounds in Section VI.

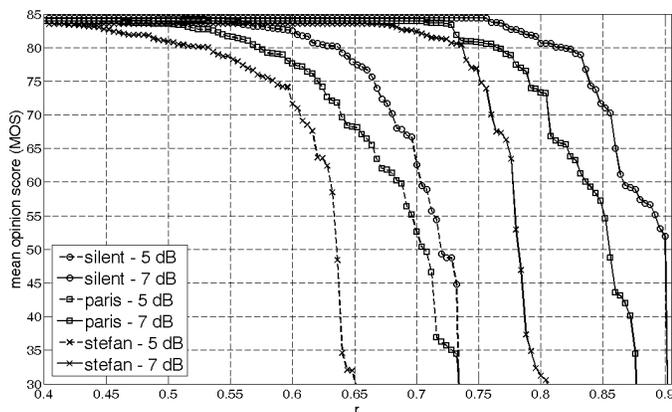


Fig. 9.  $MOS_r$  of three videos coded by AVC/H.264 using QP = 26, GOPS = 10, PS = 100 bytes and transmitted over 802.11a WLAN with a PHY data rate of 6 Mbps at average channel SNRs 5 and 7 dB

## V. VIDEO CAPACITY OF AN IEEE 802.11A WLAN WITH DCF

In this section we investigate video capacity of an 802.11a WLAN with DCF. The rule of thumb in the video capacity

calculation is that all video frames of all users should arrive at the playout buffer before their respective playout deadlines. If we do not consider extra buffering at the playout buffer, after an unavoidable initial delay for the first video frame to be transmitted to the receiver, the playout deadline of each following frame is simply 1 over the frame rate (FR) (in frames per second, or fps) of the video sequence after the time this frame is generated at the transmitter. For example, if the videos are to be displayed at the receiver side at a FR of 20 fps, the current frames of all users need to be transmitted within 0.05 second, or 50 ms. In the following we first calculate the time required to transmit one frame of a single video user. We then divide the transmission deadline by that transmission time to calculate the number of users that can be supported.

IEEE 802.11 defines two different MAC mechanisms: the contention based Distributed Coordination Function (DCF) and the polling based Point Coordination Function (PCF). At present, only the mandatory DCF is implemented in most 802.11 compliant products. DCF achieves automatic medium sharing between compatible stations by implementing Carrier-Sense Multiple Access with Collision Avoidance (CSMA/CA). In Appendix A we discuss this mechanism in more detail and calculate the transmission time  $T_{packet}(PS, DR)$  (in milliseconds) of a packet of a certain payload size, PS, over a certain 802.11a PHY data rate, DR, under DCF.

With  $T_{packet}(PS, DR)$  calculated, we only need to know how many packets are in a video frame in order to calculate the total time required for a video frame to be transmitted successfully. We already investigated the AVC/H.264 coded video frame sizes of different types of videos in Section III-B, where we show that when all the other parameters are fixed, the coded I-frame size and P-frame size stay close to a certain level for each video. We use  $S_I$  and  $S_P$  to denote the average I-frame and P-frame sizes of each video. The values of  $S_I$  and  $S_P$  depend on the video source, the video codec employed, and the schemes and parameters used in video coding. In [46] we study this dependence extensively and propose two heuristic formulas to predict the compressed I frame and P frame sizes from the coding parameters.

In the following video capacity calculation, we assume that all the video users are communicating the same type of videos, i.e., videoconferencing, news streaming, or sports streaming. We also assume that H.264/AVC is the default video codec used by all users and that similar values of the video codec parameters are chosen for all users. Therefore, the values for  $S_I$  and  $S_P$  are considered the same for all video users. Although in reality a WLAN will possibly transmit a mix of different types of video streams, it is still important to know the maximum number of users that can be supported for individual streams. The reasons are: 1) it provides the network operators an idea of how many users can be supported in the network in an identical traffic category; 2) if there is a mix of video users, the capacity number can be approximated by an interpolation of the capacity values for each traffic category present. For instance, if video “Stefan” has a capacity of X users and “News” has a capacity of Y users, the total number

of users that can be supported, if there is a mixture of "Stefan" and "News" users, would probably be something between X and Y depending on the distribution of "Stefan" and "News" users.

Since collisions are a function of traffic mix, traffic loading, retransmission limits, and QoS constraints, we do not consider collisions in our capacity calculation. We do acknowledge that the collisions will possibly reduce both the upper and lower bounds of the video capacity. Also during the course of the capacity calculation, we assume that the frames are always received without errors, however, we study the effect of transmission errors extensively when we investigate the delivered quality of the videos in Sections IV and VI.

In the following we derive the video capacity both with and without extra buffering at the receiver.

#### A. Video capacity when there is no extra buffering at the receiver

As shown in Section III-B, there is a significant difference between  $S_I$  and  $S_P$ , the compressed I and P frame sizes. Since the contention based DCF does not have the capability of coordinating the video users, we need to investigate the following two extreme cases, in terms of the timing of I frame refreshing among the users, to calculate the upper and lower bounds of video capacity. The gap between the upper and lower bounds of video capacity might be reduced if a specific rate control scheme is employed such that the difference between  $S_I$  and  $S_P$  is reduced. This issue will be discussed as part of the application in Section VI.

- 1) Worst case: all users happen to have I-frames to be transmitted at the same time, such that the capacity, denoted by  $C_m$ , is decided by I-frame size, as

$$C_m = \left\lfloor \frac{\frac{10^3}{FR}}{T_{packet}(PS, DR) \times \frac{S_I}{PS}} \right\rfloor, \quad (2)$$

where  $T_{packet}(PS, DR)$  (in milliseconds) refers to the transmission time of a packet of a certain payload size, PS, over a certain 802.11a PHY data rate, DR, under DCF;  $\frac{S_I}{PS}$ , the compressed I-frame size  $S_I$  divided by the payload size PS, is the number of packets contained in a compressed I-frame; the numerator  $\frac{10^3}{FR}$ , is the transmission deadline for the current frames of all video users, as we discuss in the first paragraph of this section.

- 2) Best case: all users happen to be coordinated in I-frame refreshing, i.e., if a WLAN supports  $C_M$  video users, at any time, there are at most  $\lfloor \frac{C_M}{GOPs} \rfloor$  users sending/receiving their I frames. The video capacity in this case, which forms an upper bound for all the different timing of the I-frame refreshing, is

$$C_M = \left\lfloor \frac{\frac{10^3}{FR} - \lfloor \frac{C_M}{GOPs} \rfloor \times \frac{S_I}{PS} \times T_{packet}(PS, DR)}{T_{packet}(PS, DR) \times \frac{S_P}{PS}} \right\rfloor + \left\lfloor \frac{C_M}{GOPs} \right\rfloor \quad (3)$$

where the numerator in the first part of the summation calculates the time left for other users to transmit their

P-frames and the corresponding denominator calculates the total time required to transmit a P-frame.

The above equation can be solved for  $C_M$ . When the timing is such that the video frames of all users are transmitted contiguously end-to-end without collisions,  $C_M$  is calculated based on a group of pictures, as

$$C_M = \left\lfloor \frac{GOPs \times \frac{10^3}{FR}}{T_{packet}(PS, DR) \times \frac{S_I + S_P \times (GOPs - 1)}{PS}} \right\rfloor. \quad (4)$$

The numerator in this formula calculates the total transmission time allowed for a group of pictures, with a group of picture size GOPs, of all users. The denominator in the formula calculates the total transmission time of one video user, which is  $T_{packet}(PS, DR)$ , the transmission time of a packet of a certain payload size, PS, over a certain 802.11a PHY data rate, DR, under DCF, multiplied by  $\frac{S_I + S_P \times (GOPs - 1)}{PS}$ , the number of packets contained in a compressed group of pictures of a single video user.

#### B. Video capacity with a playout buffer of $b$ milliseconds

The playout buffer at the receiver can allow some extra delay of the video frames, 100 ~ 500 ms for videoconferencing and up to a few seconds for video streaming. We use  $b$  in milliseconds or ms to denote the buffer length. The starting time of playing a video is normally delayed for  $b$  ms to ensure smooth video playing once it starts. However, the existence of a playout buffer is not for each frame in a video to have some extra transmission time. For example, when  $FR = 20$  fps and if we were to allow 5 ms extra delay for each video frame, which gives only a 10% increase in their transmission time, within 1 minute, the last frame is already delayed  $5ms/frame \times 20fps \times 60sec = 6 seconds$ , which is obviously too large for any type of video application. The playout buffer is only for the frames that have more bits than the other frames, such as I frames, to take more time to transmit, and then the extra time taken by these frames is compensated by the following less intensive frames, such as P frames. Therefore when the buffer is longer than the time saved by all P frames in a GOP,  $C_b$ , the video capacity with a buffer of length  $b$  (in milliseconds), is simply calculated based on a group of pictures, which is exactly the upper bound of video capacity,  $C_M$ , when there is no buffering at video playout.

For the video capacity to achieve its upper bound, the buffer length  $b$  needs to be

$$b \geq \left( \frac{10^3}{FR} - T_{packet}(PS, DR) \frac{S_P}{PS} C_M \right) (GOPs - 1), \quad (5)$$

i.e.,  $b$  needs to be larger than the time saved by the transmission of each P frame, multiplied by the total number of P frames in a group of pictures. If we substitute for  $C_M$  with the expression in Eq. (4), we get

$$b \geq \frac{10^3 \left( \frac{S_I}{S_P} - 1 \right) (GOPs - 1)}{FR \frac{S_I}{S_P} + (GOPs - 1)}. \quad (6)$$

In Fig. 10 we plot the minimum buffer length  $b$  for the video capacity to achieve its upper bound  $C_M$  for typical values of  $\frac{S_I}{S_P}$  and  $GOPS$  and a FR of 20 fps. Clearly when there is no difference in the compressed I-frame and P-frame, i.e.,  $\frac{S_I}{S_P} = 1$  and  $GOPS = 1$ , the required buffer size is 0. As seen from Fig. 10,  $b$  is always less than 1 second, which is smaller than the buffer length of typical streaming applications. For these two scenarios,  $C_b$  is always equal to  $C_M$ . For video conferencing, if  $\frac{S_I}{S_P}$  and/or  $GOPS$  are set to be relatively large numbers, the requirement for buffer size might not be satisfied. In this case,  $C_b$  fluctuates between  $C_M$  and a lower bound with buffering, denoted by  $C_{m,b}$  as

$$C_{m,b} = \left\lfloor \frac{\frac{10^3}{FR} + b}{T_{packet}(PS, DR) \times \frac{S_I}{S_P}} \right\rfloor. \quad (7)$$

The increase from  $C_m$ , the lower bound of video capacity without buffering, to  $C_{m,b}$ , the lower bound of video capacity with buffering, can be approximated by comparing  $b$  with  $\frac{10^3}{FR}$ .

If we consider the case when there is no extra buffering as a special case of buffer length  $b = 0$ , the results in this section can be summarized as

$$C_{m,b} \leq C_b \leq C_M, \quad (8)$$

i.e., the video capacity with extra buffer (length of  $b$  ms), fluctuates between the lower bound  $C_{m,b}$  and the upper bound  $C_M$ , which are calculated in Eqs. (7) and (4) respectively.  $C_b$  reaches its upper bound  $C_M$  when the buffer length  $b$  satisfies Eq. (6), which is normally the case for video streaming but not always for video conferencing.

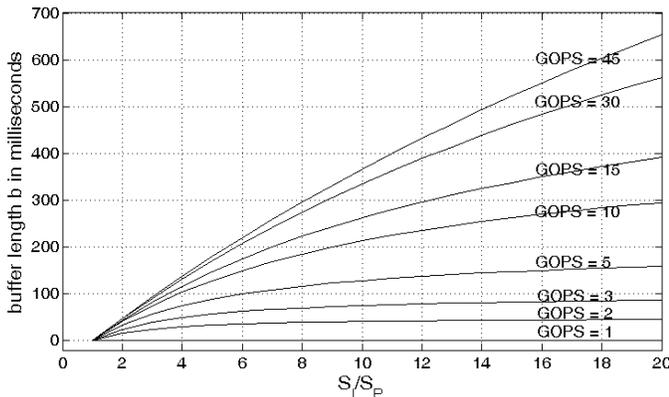


Fig. 10. The minimum buffer length  $b$  for the video capacity to achieve its upper bound  $C_M$ , for typical values of  $\frac{S_I}{S_P}$  and  $GOPS$  and a FR of 20 fps

## VI. PERCEPTUAL QUALITY CONSTRAINED VIDEO CAPACITY AND ITS APPLICATIONS

In Sections IV and V, we propose a new multiuser perceptual quality indicator  $PSNR_{r,f}/MOS_r$  and video capacity bounds, respectively. In this section we exploit these two aforementioned contributions jointly to investigate video capacity with a perceptual quality constraint. We present some of our

video over WLAN simulation results as an example, and we discuss how to use our method in other practical systems.

Table I lists the lower and upper bounds of video capacity  $C_m$  and  $C_M$  for silent.cif and stefan.cif, coded in our simulation using all the combinations of PS, QP and GOPS, at PHY data rate = 6 Mbps and FR = 15 fps. The corresponding  $MOS_{r=75\%}$ , i.e., the MOS achieved by 75% of these video users, are listed in Table II. By looking at Table I and Table II together, we can find out not only the number of video users a WLAN can support under different conditions but also the quality experienced by a majority of the users.

Throughout this paper we have pointed out the impact of the video characteristics on both the delivered quality and the coded video rate, such as in Figs. 5 and 9. In this section we can see the dramatic difference due to the different videos in terms of video capacity bounds as well as  $MOS_{r=75\%}$  in Tables I and II, respectively. For example, when PS = 100, QP = 26 (the underlined numbers in both Tables), from Table I the WLAN can support up to 8 videoconferencing (silent.cif) users, but it can only support at most one sports streaming (stefan.cif) user. Now we further consider the video quality experienced by these users. When the channel SNR = 5 dB, 75% of the 8 videoconferencing (silent.cif) users experience a “fair” quality, a MOS of around 50; however the single stefan.cif user only gets very poor quality, a MOS close to 0. Even though the results of only three videos are shown in this paper, these results demonstrate clearly that when designing or evaluating a video communication system, the characteristics of typical videos to be transmitted need to be decided first.

Now we focus on the video stefan.cif, i.e., a typical sports streaming video, and discuss different ways to employ the video capacity and  $PSNR_{r,f}/MOS_r$ . First, they can be jointly considered to derive perceptual quality constrained video capacity. For example, at PHY DR = 6 Mbps and FR = 15 fps at PS = 1100 and QP = 26 (the underlined numbers in both Tables), from Table I a maximum of 5 sports streaming users can be supported; however from Table II if the channel SNR is on average 5 dB, 75% of the 5 users actually experience poor quality, a MOS of 0 out of 100. If the channel SNR is on average 7 dB, 75% of the 5 users experience a quality between “fair” and “good”, a MOS between 59 to 68. Therefore if we set the quality constraint as “75% of users are guaranteed a fair video quality”, the video capacity of PHY DR = 6 Mbps at channel SNR = 5 dB is not 5 but 0. Second, video capacity and  $PSNR_{r,f}/MOS_r$  can be used to choose the compression and communication parameters. For example, between PS = 100 and PS = 1100, the larger payload size needs to be chosen if higher video capacity is desired, and the smaller payload size is chosen if the quality the users experience needs to be higher. We do not show the results for other payload sizes or other PHY data rates, but it is clear that these parameters can be chosen based on similar criteria.

Some of the numbers in these two tables are a bit surprising. For example, as we increase GOPS from 10 to 45, the video quality in  $MOS_{r=75\%}$  does not necessarily decrease even though the I-frame refreshing becomes less and less frequent.

TABLE I  
ONE WAY VIDEO CAPACITY: PHY DATA RATE=6 MBPS, FR=15 FPS

Video	$PS$ (bytes)	QP=26					QP=30				
		$C_m$	$C_M$				$C_m$	$C_M$			
			GOPS=10	15	30	45		GOPS=10	15	30	45
silent.cif (videoconferencing)	100	$\underline{0}$	$\underline{4}$	$\underline{6}$	$\underline{7}$	$\underline{8}$	1	8	10	12	14
	1100	$\underline{2}$	$\underline{9}$	$\underline{11}$	$\underline{14}$	$\underline{15}$	4	18	21	27	30
stefan.cif (sports)	100	$\underline{0}$	$\underline{1}$	$\underline{1}$	$\underline{1}$	$\underline{1}$	0	2	3	3	3
	1100	$\underline{1}$	$\underline{5}$	$\underline{5}$	$\underline{5}$	$\underline{5}$	2	8	9	10	11

TABLE II  
 $MOS_{r=75\%}$  CALCULATED FROM  $PSNR_{r=75\%, f=90\%}$  FOR SILENT.CIF AND STEFAN.CIF

Channel SNR(dB)	Video	$PS$ (bytes)	QP=26				QP=30			
			GOPS=10	15	30	45	GOPS=10	15	30	45
5	silent.cif	100	$\underline{45}$	$\underline{52}$	$\underline{46}$	$\underline{43}$	55	51	48	45
		1100	$\underline{0}$	$\underline{0}$	$\underline{0}$	$\underline{0}$	0	0	0	0
	stefan.cif	100	$\underline{0}$	$\underline{12}$	$\underline{0}$	$\underline{9}$	13	10	10	13
		1100	$\underline{0}$	$\underline{0}$	$\underline{0}$	$\underline{0}$	0	0	0	0
7	silent.cif	100	84	84	84	84	75	75	75	75
		1100	84	84	84	84	75	75	75	74
	stefan.cif	100	75	82	74	73	71	72	71	70
		1100	$\underline{68}$	$\underline{59}$	$\underline{62}$	$\underline{62}$	60	62	63	51

Also the effect of QP on the delivered video quality with possible packet loss and concealment is quite complicated and dynamic. On one hand, lower QP means finer quantization and yields better video quality when there is no packet loss. On the other hand, lower QP corresponds to more packets for a video sequence, and when there is packet loss, more packets might be lost, and the quality improvement in the received packets because of finer quantization is jeopardized by the quality deterioration caused by the lost and concealed packets. We discuss these observations and the interplay of the parameters in a packet loss environment in [47].

The  $PSNR_{r,f}/MOS_r$  constrained video capacity can also be employed by a video application which might use a specific RC/RDO scheme, and/or a specific channel coding scheme. With RC/RDO, the bit stream should be smoother in terms of the compressed I-frame and P-frame sizes and there may be intra-coded blocks in some P frames. This only makes  $\frac{S_I}{S_P}$  smaller than that in our simulation, which brings the upper and lower bounds of video capacity closer. In some cases, even if a GOPS is fixed initially, there might be a sudden scene change in a video sequence, in which case a new GOPS will need to be incorporated in the capacity bounds and minimum buffer size calculations. If we include channel coding, the packet size and therefore the time required for a video packet to be transmitted successfully needs to be calculated based on the coding scheme and its parameters.

## VII. CONCLUSIONS

In this paper we investigate multiuser perceptual quality constrained video capacity for WLANs. As a particular example, we investigate the delivered quality and coded data rate of AVC/H.264 coded video over IEEE 802.11a WLANs in a frequency selective multipath fading environment.

We show that for the same video coded using the same parameters for the same average channel SNR, the quality of the concealed video varies significantly across different realizations. The PSNRs also vary from one frame to another in the same processed video sequence. Neither the average PER nor the average PSNR across all of the frames and all of the realizations is a suitable indicator of the quality a video user experiences, and therefore they should not serve as the basis for video communications quality assessment. A new multiuser perceptual video quality indicator  $PSNR_{r,f}/MOS_r$  is proposed to capture the distribution of the distortion across the video frames and channel uses (realizations or number of users).  $PSNR_{r,f}$  is defined as the PSNR achieved by  $f\%$  of the frames in each one of  $r\%$  of the transmissions. A subjective experiment sets up a linear equation connecting  $PSNR_{r,f=90\%}$  and  $MOS_r$ , the mean opinion score (MOS) achieved by  $r\%$  of the transmissions.

Due to the significant difference in the intra-coded and inter-coded frame sizes of a compressed video, the upper and lower bounds for video capacity of an 802.11a WLAN operated under the Distributed Coordination Function (DCF) are formulated for the case when there is no buffering at the receiver. The video capacity when there is buffering at the receiver is also studied, and the minimum buffer length for the video capacity to reach its upper bound is obtained.

Combining the multiuser perceptual quality indicator and the video capacity calculation, a methodology for video over WLAN communication system design and evaluation is proposed, which consists of determining the video capacity of a WLAN in the context of the delivered video quality constraints calculated by  $PSNR_{r,f}/MOS_r$ . This work appears to be the first such effort to address this difficult but important problem. Furthermore, the methodology employed is perfectly general and can be used for different networks, video codecs, trans-

mission channels, protocols, and perceptual quality measures.

## APPENDIX A

### TRANSMISSION TIME OF ONE PACKET UNDER 802.11A DCF

IEEE 802.11 defines two different MAC mechanisms: the contention based Distributed Coordination Function (DCF) and the polling based Point Coordination Function (PCF). At present, only the mandatory DCF is implemented in most 802.11 compliant products. DCF achieves automatic medium sharing between compatible stations by implementing Carrier-Sense Multiple Access with Collision Avoidance (CSMA/CA). A station using the DCF has to follow two medium access rules: 1) the station will be allowed to transmit only if its carrier-sense mechanism determines that the medium has been idle for at least a Distributed Inter Frame Space (DIFS) time, and 2) in order to reduce the collision probability among multiple stations accessing the medium, the station will select a random backoff interval after deferral or prior to attempting to transmit another packet after a successful transmission. An acknowledgment (Ack) packet will be sent by the receiver upon successful reception of a data packet. It is only after receiving an Ack packet correctly that the transmitter assumes successful delivery of the corresponding data packet. The Short Inter Frame Space (SIFS), which is smaller than DIFS, is the time interval between reception of a data packet and transmission of its Ack packet. The timing of successful packet transmissions is shown in Fig. 11. For a detailed explanation on the 802.11a DCF the readers are directed to [48].

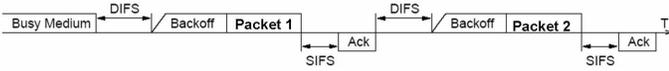


Fig. 11. Timing of successful frame transmissions under the DCF

Let  $T_{packet}(PS, DR)$  be the total time (in milliseconds) required for a packet of payload size of PS to be transmitted successfully over a certain PHY data rate of DR under 802.11a DCF. According to Fig. 11,  $T_{packet}(PS, DR)$  is composed of the DIFS time, the backoff time, the time to send the video packet itself, the SIFS time, and the time to transmit the Ack frame, as

$$T_{packet}(PS, DR) = t_{DIFS} + t_{Backoff} + t_{packet}(PS, DR) + t_{SIFS} + t_{Ack}(DR). \quad (9)$$

The following two formulas of calculating  $t_{packet}(PS, DR)$  and  $t_{Ack}(DR)$  are standard and are presented here without further explanation. Interested readers can refer to the 802.11 standards [37] or other research papers such as [48],

$$t_{packet}(PS, DR) = t_{PLCP\_Preamble} + t_{PLCP\_SIG} + \left\lceil \frac{28 + (16+6)/8 + 40 + PS}{BpS(DR)} \right\rceil \cdot t_{Symbol}. \quad (10)$$

$$t_{Ack}(DR) = t_{PLCP\_Preamble} + t_{PLCP\_SIG} + \left\lceil \frac{14 + (16+6)/8}{BpS(DR)} \right\rceil \cdot t_{Symbol}. \quad (11)$$

For the PSs of 100 bytes and 1100 bytes, and the DR of 6 Mbps that we implement in our simulation and that is used as examples in Section VI,

$$\begin{aligned} t_{packet}(100bytes, 6Mbps) &= 248\mu s, \\ t_{packet}(1100bytes, 6Mbps) &= 1584\mu s, \\ t_{Ack}(6Mbps) &= 44\mu s. \end{aligned} \quad (12)$$

The backoff time  $t_{Backoff}$  is set as  $aCWmin/2 \times t_{SlotTime} = 15/2 \times 9 = 67.5\mu s$ .

Summarizing the above calculations, the time needed to successfully transmit 100 byte and 1100 byte video packets over a PHY DR of 6 Mbps is

$$\begin{aligned} T_{packet}(100bytes, 6Mbps) &= 0.4095ms, \\ T_{packet}(1100bytes, 6Mbps) &= 1.7455ms. \end{aligned} \quad (13)$$

## REFERENCES

- [1] M. van der Schaar and P. A. Chou, editors, *Multimedia over IP and Wireless Networks: Compression, Networking, and Systems*. Elsevier / Academic Press, publishers, Mar. 2007.
- [2] K. Nahrstedt, W. Yuan, S. Shah, Y. Xue, and K. Chen, "Quality of service support in multimedia wireless environments," *Multimedia over IP Wireless Networks: Compression, Networking, and Systems*, M. van der Schaar and P. A. Chou eds. Elsevier / Academic Press, publishers, Mar. 2007.
- [3] IEEE Std. 802.11e/D10.0, "Draft supplement to part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications: Medium Access Control (MAC) enhancements for Quality of Service (QoS)," Nov. 2004.
- [4] Y. Wu et al., "Network planning in wireless ad hoc networks: a cross-layer approach," *IEEE Journal on Selected Areas in Communications, special issue on wireless ad hoc networks*, vol. 23, no. 1, pp. 136–150, Jan. 2005.
- [5] T. Stockhammer and W. Zia, "Error-resilient coding and decoding strategies for video communication," *Multimedia over IP Wireless Networks*, M. van der Schaar and P. A. Chou eds. Elsevier / Academic Press, publishers, Mar. 2007.
- [6] P. A. Chou and Z. Miao, "Rate-distortion optimized streaming of packetized media," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 390–404, Apr. 2006.
- [7] G. Cote, S. Shirani, and F. Kossentini, "Optimal mode selection and synchronization for robust video communications over error-prone networks," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 952–965, Jun. 2000.
- [8] C. Hsu, A. Ortega, and M. Khansari, "Rate control for robust video transmission over burst-error wireless channels," *IEEE Journal on Selected Areas in Communications, Special Issue on Multimedia Network Radios*, vol. 17, no. 5, pp. 756–773, May 1999.
- [9] Q. Li and M. van der Schaar, "Providing adaptive QoS to layered video over wireless local area networks through real-time retry limit adaptation," *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 278–290, Apr. 2004.
- [10] A. Ortega and H. Wang, "Mechanisms for adapting compressed multimedia to varying bandwidth conditions," *Multimedia over IP Wireless Networks: Compression, Networking, and Systems*, M. van der Schaar and P. A. Chou eds. Elsevier / Academic Press, publishers, Mar. 2007.
- [11] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966–976, 2000.
- [12] X. Zhu, E. Setton, and B. Girod, "Congestion-distortion optimized video transmission over Ad Hoc networks," *EURASIP'05*, 2005.
- [13] M. van der Schaar, S. Krishnamachari, S. Choi, and X. Xu, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 10, pp. 1752–1763, Dec. 2003.
- [14] E. Setton, T. Yoo, X. Zhu, A. Goldsmith, and B. Girod, "Cross-layer design of ad hoc networks for real-time video streaming," *IEEE Wireless Communications Magazine*, vol. 12, no. 4, pp. 59–65, Aug. 2005.

- [15] M. van der Schaar, "Cross-layer wireless multimedia," *Multimedia over IP Wireless Networks: Compression, Networking, and Systems*, M. van der Schaar and P. A. Chou, eds. Elsevier / Academic Press, publishers, Mar. 2007.
- [16] ITU Recommendations, "Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference," *ITU-R BT.1683*, Jun. 2004.
- [17] —, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," *ITU-T J.144*, Mar. 2004.
- [18] T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation," *Handbook of Image & Video Processing* (A. Bivok eds.), Academic Press, 2000.
- [19] Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," *The Handbook of Video Databases: Design and Applications* (B. Furht and O. Marqure, eds.), CRC Press, pp. 1041–1078, Sep. 2003.
- [20] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, p. 2350, Nov. 1998.
- [21] Z. He and S. K. Mitra, "From rate-distortion analysis to resource-distortion analysis," *IEEE Circuits and Systems Magazine*, vol. 5, no. 3, pp. 6–18, Third quarter 2005.
- [22] N. Shetty, S. Choudhury, and J. D. Gibson, "Voice capacity under quality constraints for IEEE 802.11a based WLANs," *ACM International Wireless Communication and Mobile Computing Conference (IWCMC), Vancouver*, Jul. 2006.
- [23] ISO/IEC 13818-1:2000, "Information technology – generic coding of moving pictures and associated audio information: Systems," 2000.
- [24] ISO/IEC 14496-1:2001, "Information technology – coding of audiovisual objects – part 1: Systems," 2001.
- [25] ITU Recommendations, "Video coding for low bit rate communication," *ITU-T rec. H.263*, Jan. 2005.
- [26] ITU-T and ISO/IEC JTC 1, "Advanced video coding for generic audiovisual services," 2003.
- [27] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 246–251, Feb. 1997.
- [28] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 6, pp. 878–894, Sep. 2000.
- [29] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 172–185, Feb. 1999.
- [30] S. Ma, W. Gao, and Y. Lu, "Rate control on JVT standard," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, JVT-D030*, Jul. 2002.
- [31] Z. G. Li, F. Pan K. P. Lim, X. Lin and S. Rahardj, "Adaptive rate control for h.264," *IEEE International Conference on Image Processing*, pp. 745–748, Oct. 2004.
- [32] Y. Wu et al., "Optimum bit allocation and rate control for H.264/AVC," *Joint Video Team of ISO/IEC MPEG & ITU-T VCEG Document*, vol. JVT-O016, Apr. 2005.
- [33] G. J. Sullivan and T. Wiegand, "rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 74–90, Nov. 1998.
- [34] Y. K. Tu, J.-F. Yang and M.-T. Sun, "Rate-distortion modeling for efficient H.264/AVC encoding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 5, pp. 530–543, May 2007.
- [35] D.-K. Kwon, M.-Y. Shen and C.-C. J. Kuo, "Rate control for H.264 video with enhanced rate and distortion models," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 5, pp. 517–529, May 2007.
- [36] 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*, vol. 13, pp. 560–576, Jul. 2003.
- [37] IEEE Std. 802.11-1999, Part 11, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications," *ISO/IEC 8802-11:1999(E)*, 1999.
- [38] V. Q. E. Group, "The Quest for Objective Methods: Phase II, Final Report," <http://www.its.bldrdoc.gov/vqeg/>, Aug. 2003.
- [39] "H.264/AVC software coordination - reference software JM10.1," <http://iphome.hhi.de/suehring/tml/>, 2006.
- [40] C.-W. Kong, J. Lee, M. Hamdi, and V. Li, "Turbo-slice-and-patch: an algorithm for metropolitan scale VBR video streaming," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 3, pp. 338–353, Mar. 2006.
- [41] N. Chayat, "Tentative criteria for comparison of modulation methods," *IEEE P802.11-97/96*, Sep. 1997.
- [42] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*. Artech House, Jan. 2000.
- [43] F. Chiaraluce, L. Ciccarelli, E. Gambi, and S. Spinsante, "Performance evaluation of error concealment techniques in h.264 video coding," *Picture Coding Symposium*, no. 1, pp. 163–166, 2004.
- [44] K. Stuhlmüller, N. Farber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, Jun. 2000.
- [45] "Methodology for the subjective assessment of the quality of television pictures," *ITU-R Recommendation BT.500*, 2002.
- [46] J. Hu, S. Choudhury, and J. D. Gibson, "Perceptual quality constrained video user capacity of 802.11a WLANs," *International Symposium on Multimedia over Wireless*, Aug. 2007.
- [47] —, "H.264 video over 802.11a WLANs with multipath fading: Parameter interplay and delivered quality," *International Conference on Multimedia & Expo*, Jul. 2007.
- [48] D. Qiao, S. Choi, and K. G. Shin, "Goodput analysis and link adaptation for IEEE 802.11a wireless LANs," *IEEE Trans. on Mobile Computing (TMC)*, vol. 1, no. 4, Oct.-Dec. 2002.