REAL-TIME TRANSMISSION OF H.264 VIDEO OVER 802.11-BASED WIRELESS AD HOC NETWORKS

E. Masala[1], *C. F. Chiasserini*[2], *M. Meo*[2], *J. C. De Martin*[3]

[1]Dipartimento di Automatica e Informatica / [2]Dipartimento di Elettronica / [3]IEIIT-CNR
Politecnico di Torino
Corso Duca degli Abruzzi, 24 — I-10129 Torino, Italy
E-mail: [masala|michela|chiasserini|demartin]@polito.it

ABSTRACT

The paper aims at evaluating a number of Quality of Service (QoS) indices of a real-time video transmission over an 802.11 ad hoc wireless network. Video is coded according to the state-of-the-art ITU-T H.264 encoder and its transmission is simulated by means of the *ns* network simulator. Objective quality measurements are presented. Moreover, the impact of different parameters, both at the encoded and at the MAC level, and of background interfering traffic, is studied, showing the various trade-offs involved.

1. INTRODUCTION

The great success of the IEEE 802.11 technology for wireless local area networks (WLANs) [1] is creating new opportunities for the deployment of advanced multimedia services. Important applications such as telephony, videoconferencing and audiovisual streaming are on path to move to Wireless Local Area Networks (WLANs), creating a complex, yet highly attractive scenario, where users will be able, at least in principle, to seamlessly switch from, typically expensive, wide-area coverage to cheaper, higher-bandwidth local and micro-local networks.

Real-time multimedia transmission over WLANs, however, poses several challenges. Radio bandwidth is limited, and propagation conditions over the radio channel may significantly vary in time, often leading to quite large error rates. Besides, in the case of 802.11-based networks, sources must contend for the radio channel, thus experiencing access delays that may significantly degrade perceptual quality.

The purpose of this paper is to study the transfer of video sequences over wireless ad hoc networks using the 802.11 technology. The state-of-the-art ITU-T H.264 [2] video encoder is configured to optimally match the ad hoc network scenario as well as to adapt to varying channel conditions. Standard video test sequences are packetized according to the H.264 Network Adaptation Layer (NAL) for transmission using the RTP/UDP/IP protocol stack. Several network conditions are simulated using the *ns* [3] network simulator. In particular, we consider the presence of interfering data traffic carried by TCP connections. The quality perceived by the video user at the receiver is objectively evaluated, using the PSNR as a distortion measure.

This work was supported in part by CERCOM, the Center for Wireless Multimedia Communications, Torino, Italy, http://www.cercom.polito.it.

Design trade-offs involving different layers are analyzed to optimize the overall performance of the system. Error resilience tools provided by the H.264 standard are also configured and adapted to the characteristics of the 802.11 wireless medium.

Multimedia data transmission over WLANs has recently been the focus of several works, but either coding and error resilience aspects are not explicitly covered [4][5], or the effects of interfering traffic are not considered [6].

In this paper, we aim at globally optimizing the parameters involved in a real-time video transmission, ranging from video encoding and packetization to the 802.11 MAC interface parameters. Accurate and objective results, obtained through network simulations and video quality evaluation, are provided, showing the system performance under various network conditions.

2. THE 802.11 AD HOC NETWORK SCENARIO

We consider an ad hoc network composed of stationary wireless stations, using the IEEE 802.11b technology. We assume that the 802.11 stations can transmit at 11 Mbps, and employ the basic DCF scheme to access the channel [1]. The ad hoc network includes one station generating video traffic, and up to 8 stations which generate data traffic and exhibit a greedy behavior. All traffic sources are associated with the same destination and use the same relay station to deliver their traffic to the destination. Also, all source nodes are in the radio proximity of each other, while only the relay station can directly communicate with the destination node. The overall network scenario is shown in Figure 1.

At the video traffic source, standard video test sequences are packetized according to the IP Network Adaptation Layer (NAL) specification of the H.264 standard and transmitted using the RTP/UDP/IP protocol stack. Data traffic, instead, is transmitted by using the TCP/IP protocol suite. At the MAC and physical layers, all stations employ the 802.11 functions.

The 802.11 radio channel is modeled as a Gilbert channel. Two states, *good* and *bad*, represent the state of the channel during an 802.11 slot time: a MAC Protocol Data Unit (MPDU) is received correctly if the channel is in state good for the whole duration of the MPDU transmission; it is received in error otherwise. We denote the transition probability from state *good* to state *bad* by p and the transition probability from state *bad* to state *good* by q; a set of p and q values, representing a range of channel conditions, has been obtained by using the trace-based channel estimation in [7]. The average error probability, denoted by ϵ, and the average

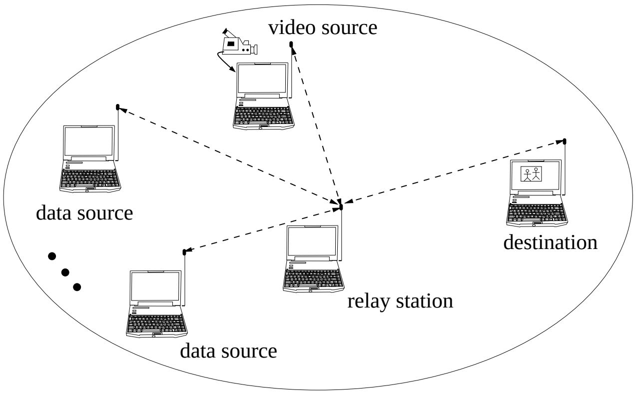

Fig. 1. The 802.11 ad hoc network scenario.

length of a burst of errors are derived as $p/(p+q)$ and $1/q$, respectively.

3. THE H.264 VIDEO CODING STANDARD

We focus on the transmission of video data compressed according to the new ITU-T H.264 standard. The compression scheme follows the ISO MPEG and ITU-T H.26x coding standards, with some new features to achieve a higher compression efficiency. Some of them are briefly outlined in the following; refer to [2] and [8] for more details. The base coding unit for transform coding is a 4x4 sample block. Thus macroblocks are composed of 16 luminance blocks and 4 blocks for each chrominance component. The transform coding is a separable integer transform with essentially the same properties of the DCT. Regarding motion compensation, prediction using multiple reference frames is possible.

Consecutive macroblocks are grouped into a *slice*. The *slice* is important because it has the property to be independently decodable. This is useful to subdivide the coded stream into independently decodable packets, so that the loss of a packet does not affect the decoding of others.

One of the most interesting characteristics of the H.264 standard is the decoupling of the coding aspects from the bitstream adaptation needed to transmit it over a particular channel. The part of the standard that deals with the coding aspects is called Video Coding Layer (VCL), while the other is the Network Adaptation Layer (NAL) [9]. One of the developed NAL is aimed to the problem of transporting data over an IP network using the Real-Time Transport Protocol (RTP) [10], which is well suited for real time multimedia transmissions.

A complete separation between the VCL and the NAL is difficult to obtain because some dependencies exist. The packetization process is an example: error resilience, in fact, is improved if the VCL is instructed to create slices of about the same size of the packets and the NAL told to put only one slice per packet. The error resilience characteristics of the transmission will benefit because all the data contained in a certain packet can be decoded independently from the others. Note that in H.264 the subdivision of a frame into slices has not to be the same for each frame of the sequence; thus the decoder can flexibly decide how to make the slices. However they cannot be too short because a decrease of the compression ratio would occur for two reasons, i.e. the slice headers would reduce the available bandwidth and the context-based entropy coding would become less efficient.

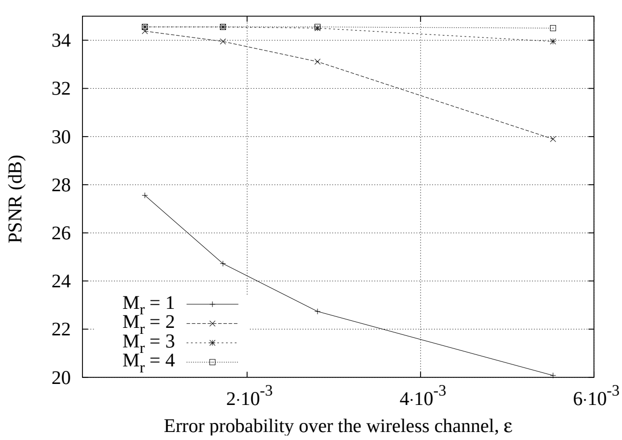

Fig. 2. PSNR values as a function of the channel error probability, for varying maximum number of transmission M_r.

Moreover, in wireless channels, the size of the packet influences its error probability; longer packets are more likely to contain transmission errors but reduce coding efficiency. In this paper the trade-off involved in the packet creation process will be investigated, studying the performances of the video transmission as a function of the packet size.

4. RESULTS

Accurate simulations have been carried out using the *ns* [3] network simulator, to study the the effects of transmission errors in different network conditions. The well known *Foreman* video sequence has been coded using the H.264 test model software [11], enabling most of the new characteristics of the H.264 standard, in particular multiple reference frames and Lagrangian optimized motion search for macroblocks down to 4$\times$4 pixels size. The sequence size is CIF at 15 fps, and is encoded using a fixed quantization parameter, set to achieve a bit rate of about 256 kbit/s. The sequence length is 149 frames, and one B frame is introduced after each P frame. The transmitted sequence is obtained concatenating the base video sequence 80 times, reaching a length of 794.6 s at 15 fps. To improve error resilience, an I frame is interposed at the beginning of each repetition of the sequence (i.e., every 148 frames.)

We supposed to transmit the video sequence using the IP/UDP/RTP protocol stack. At the receiver, a playout buffer mechanism has been implemented to compensate the delay jitter of the packets. The playout buffer size has been set to 1 s. At the MAC layer, the duration of the time slot and of the DIFS time interval has been set to 20 μs and 50 μs, respectively. The average length of a burst of errors over the radio channel is equal to 5 time slots.

The first set of simulations analyzes the video transmission quality when no other traffic is present in the network, while the second is focused on the study of the effect of an increasing amount of TCP traffic on the video transmission.

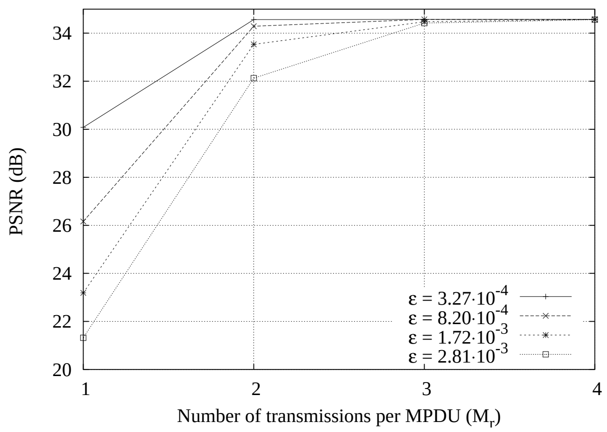

Fig. 3. PSNR values as a function of the maximum number of transmission per MPDU at the MAC level. The results are plotted for various channel conditions.

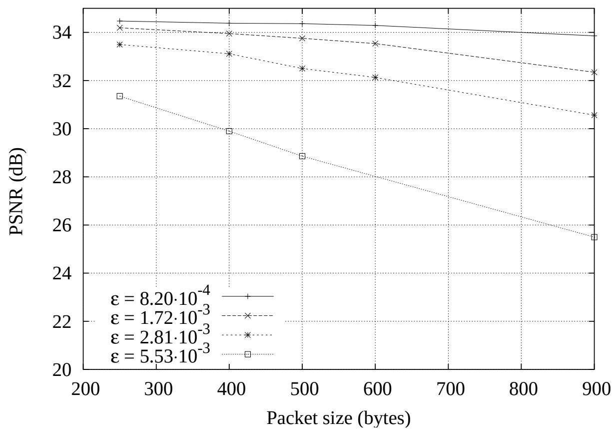

Fig. 4. PSNR values as a function of mean packet size, for various channel conditions; M_r is equal to 2.

4.1. Video quality measurements

The following results show the behavior of the transmission system when no background traffic is present in the network.

Figure 2 shows the peak SNR (PSNR) values for the transmission of video as a function of the error probability over the wireless channel. The curves differ for the maximum number of allowed transmissions per MPDU, denoted by M_r. If no retransmissions are present ($M_r = 1$), the quality rapidly decreases, showing that at least one retransmission at the MAC level is needed to obtain an acceptable video quality when the error probability tends to be high.

Figure 3 shows the impact of the maximum number of transmissions per MPDU on the perceptual video quality, measured by the PSNR, for four different channels. Video quality very near to the maximum can be achieved by setting the maximum number of transmission attempts per MPDU, M_r, to 3.

Figure 4 shows the effect of the mean size of the packets on video quality, for various channel conditions, with M_r equal to

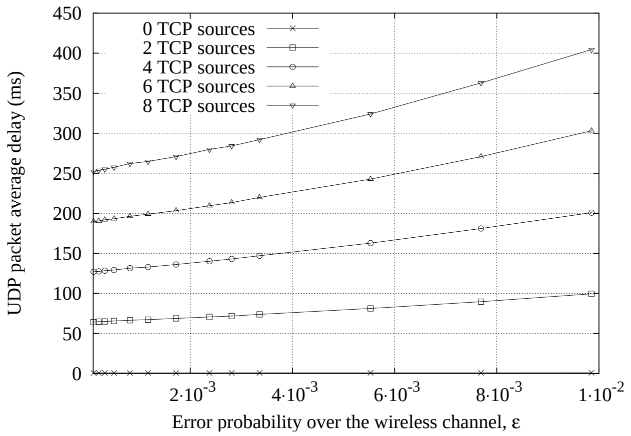

Fig. 5. Average delay of UDP packets as a function of the error probability over the 802.11 wireless channel. The results are plotted for a varying number of TCP traffic sources and by setting the maximum number of transmission attempts for video traffic to be equal to 2.

2. When the channel error probability is low ($\epsilon = 8.20 \cdot 10^{-4}$), the mean packet size has a limited influence on the video quality, thus larger video packets can be used, minimizing the MAC header overhead without incurring in an excessive quality degradation. For high error probabilities ($\epsilon = 5.53 \cdot 10^{-3}$), it is better to create smaller video packets during the encoding process so that the channel errors affect a more limited part of the video sequence.

4.2. Network performance

In this section we evaluate the impact that the following aspects of the system have on the quality of the video service: i) the interfering TCP traffic, ii) the error probability over the radio channel, iii) the setting at the MAC layer of the maximum number of allowed transmissions per MPDU, denoted by M_r.

Figure 5 shows the average delay perceived by UDP packets versus the error probability over the wireless channel when various numbers of interfering TCP sources are considered. The maximum number of transmission per MPDU, M_r, is set to 2 in the MAC layer of the video traffic source. As the error probability over the wireless channel increases, the number of transmissions performed per MPDU increases so that longer times are needed to deliver the MPDUs and, thus, longer average delays are experienced at the UDP level. A similar behavior can be observed by letting the number of interfering TCP sources increase. The effect of a large number of interfering sources is twofold. On the one hand, it translates into a large collision probability which delays the access to the radio channel. On the other hand, when the channel is shared by a large number of sources, the channel capacity perceived by individual sources is smaller.

Let us focus on the case where no interfering TCP sources are considered. Depending on the radio channel conditions, the service time of video MPDUs is either equal to one or to two MPDU transmission times (remember that M_r is equal to 2). The service time results to be small enough that there is no queue at the MAC layer and the delay perceived by UDP packets is extremely small.

These results suggest some criteria for the choice of the set of

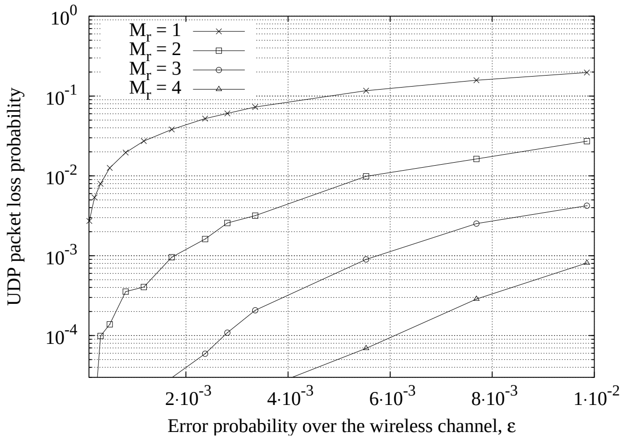

Fig. 6. Loss probability of UDP packets as a function of the error probability over the 802.11 wireless channel, when no TCP traffic sources are considered. The results are plotted for different values of the maximum number of transmission attempts.

services which can be provided by the system. Consider, for example, the case of interactive video services whose QoS constraint consists in the average delay being kept as small as 150 ms. In this case, the number of interfering TCP sources should be limited. Up to 3 interfering sources are acceptable, while 4 sources can be admitted only if the average error probability over the radio channel is small, say smaller than 0.003. More than 4 sources cannot be admitted even in the presence of very good channel conditions.

Figure 6 shows the impact of the radio channel conditions on the loss probability of UDP packets of the video stream when different values of the maximum number of transmissions per MPDU are considered. The curves show that some retransmissions are needed in order to keep the UDP packet loss probability to reasonable values. However, values of M_r as small as 3 are already enough to guarantee UDP loss probability smaller than 1%, even under bad channel conditions.

5. CONCLUSIONS

The behavior of H.264-coded video transmission over a wireless 802.11 ad hoc network scenario has been analyzed. The influence of some of the parameters involved in the transmission system has been studied by means of network simulations. Various network conditions have been tested, with different levels of background interfering traffic. Results give a clear indication on how to select the system parameters. In particular, we have observed that a video packet size as small as 300 bytes should be used when channel conditions are not favorable. A maximum number of transmission attempts at the MAC layer equal to 3 enables us to obtain a high PSNR, for most channel conditions. Moreover, in the presence of interactive video services, the number of TCP sources that can be admitted in the network should be limited in order to meet the QoS requirements.

Further work is needed to investigate the trade-offs existing between perceptual quality and utilization of radio resources, as well as the system performance when user mobility and more complex multihop network scenarios are considered.

6. REFERENCES

[1] "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *ISO/IEC 8802-11, ANSI/IEEE Std 802.11*, 1999.

[2] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "Joint Final Committee Draft (JFCD) of Joint Video Specification (ITU-T Rec. H.264 — ISO/IEC 14496-10 AVC)," *Doc. JVT-D157*, July 2002.

[3] UCB/LBNL/VINT, "Network Simulator – ns – version 2," *URL: http://www.isi.edu/nsnam/ns*, 1997.

[4] T. Suzuki and S. Tasaka, "Performance Evaluation of Integrated Video and Data Transmission with the IEEE 802.11 Standard MAC Protocol," in *Proceedings of GLOBECOMM*, 1999, vol. 1b, pp. 580–586.

[5] M. Freytes, C.E. Rodriguez, and C.A. Marques, "Real-Time H.263+ Video Transmission on 802.11 Wireless LANs," in *Proc. Int. Conf. on Information Technology: Coding and Computing*, Las Vegas, NV, April 2001.

[6] A. Majumdar, D. G. Sachs, I. V. Kozintsev, K. Ramchandran, and M. M. Yeung, "Multicast and Unicast Real-Time Video Streaming Over Wireless LANs," *IEEE Journal on Selected Areas in Communications*, vol. 12, no. 6, pp. 524–534, June 2002.

[7] G. T. Nguyen, R. H. Katz, B. Noble, and M. Satyanarayanan, "A trace-based approach for modeling wireless channel behavior," in *Winter Simulation Conference*, 1996, pp. 597–604.

[8] T. Stockhammer, T. Oelbaum, and T. Wiegand, "H.26L/JVT Coded Video Transmisson in 3G Wireless Environments," in *Int. Conf. on Third Generation Wireless and Beyond*, San Francisco, CA, May 2002.

[9] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "An Overview on the H.26L NAL Concept," *Doc. JVT-B028, URL: ftp://standard.pictel.com/video-site*, January 2002.

[10] R. Frederick H. Schulzrinne, S. Casner and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," *RFC 1889*, January 1996.

[11] ITU-T VCEG, "Test Model Long Term (TML) 9.7," *URL: ftp://standard.pictel.com/video-site*, February 2002.