Intelligence Packet Scheduling for optimized video transmission over wireless networks

Ilias Politis
University of Patras
26500 Rio Patras
Greece
ipolitis@ece.upatras.gr

Michail Tsagkaropoulos
University of Patras
26500 Rio Patras
Greece
mtsagkaro@ece.upatras.gr

Thomas Pliakas
Dept. of Information & Communication Systems Engineering
University of Aegean
Karovasi, Samos, Greece
tpliakas@aegean.gr

Tasos Dagiuklas
Technological Institute of Messolonghi
Greece
ntan@teimes.gr

Stavros Kotsopoulos
University of Patras
26500 Rio Patras
Greece
kotsop@ece.upatras.gr

ABSTRACT

This paper presents an intelligent packet scheduling scheme based on an analytical distortion model for Multi-Referenced H.264 video over 802.11 wireless networks. This model considers that video distortion in H.264 video codecs is based on multi-reference video frames, as opposed to previous video standards. The model has been verified with actual video distortion measurements and it has been compared with a simple additive model that omits the correlation between the frames. The packet dropping that is required in the case of the available transmission rate is limited, takes into consideration the impact of the loss of the frame on the resulting video distortion that depends not only on the distortion of the particular frame but on the distortion effect that propagates to the correlated frames. Furthermore, simulation results show that the proposed model accurately captures the distortion effect for reference based H.264 coding and the improves the reconstructed video quality when is implemented in rate adaptation via packet scheduling.

Keywords

Multi-reference H.264, Packet Scheduling, Rate Adaptation, 802.11

1. INTRODUCTION

As video transmission over lossy packet networks such as the Internet and emerging wireless networks are becoming ever more popular, it is important to develop error resilient source coding and transmission techniques [18][8] .

The video quality of a decoded video transmitted over a wireless network is often associated with the average pixel by pixel distortion of the vide frames. This distortion results from both the compression scheme of the source encoder and the channel losses. The latter is commonly referred to as channel distortion and depends on the communication channel loss characteristics, the intra-update period and the error concealment method applied during decoding at the receiver. Modelling channel distortion accurately is the key for rate-distortion optimization and end-to-end quality of video communications.

The next wireless LAN (WLAN) generation could provide multimedia services to mobile and fixed users through wireless access, with the development of the high-speed physical (PHY) layers IEEE 802.11g (54 Mb/s)[3]. However, wireless channel characteristics such as shadowing, multipath, fading, and interferences still limit the available bandwidth for the deployed applications. Consequently, video compression techniques and transmission techniques are a critical part of multimedia applications over WLAN. The IEEE 802.11 networks are most commonly used due to its low cost and easy deployment. The 802.11 standard [3] provide two access mechanisms, in order the mobile terminals to gain access to shared wireless medium; (1) the *Point Coordination Function* (PCF) and (2) the *Distribution Coordination Function* (DCF). The PCF originally aims at supporting real-time traffic, but is rarely implemented in current commercial products due to its implementation complexity and uncertainty on the efficiency. On the other hand, (DCF) is a contention-based channel access protocol. A shared wireless medium is randomly accessed by contentions among stations in a service area. Accordingly, the access delay increases significantly with contending stations. In addition, at the presence of channel errors, retransmissions of the MAC *Automatic Repeat Request* (ARQ) become another reason of much longer delay. The mechanisms of the contention-based access and the MAC ARQ produce highly random and long delay. These lead to only support of best-effort services, and real-time video communication is even more challenging.

There is a large number of research works that has been reported and regards modelling the impact of packet loss on video distortion. Such models can be fall in two cate-

MOBIMEDIA 2007, August 27-29, Nafpaktos, Greece

DOI 10.4108/ICST.MOBIMEDIA2007.1924

gories. In the first category, the models consider that distortion is proportional to the number of losses within a video sequences [15],[10]. These studies also suggest that the average distortion of multiple losses can be derived as a superposition of the uncorrelated error signals. However, these models are accurate for low residual error rates and when such errors are sufficiently apart of each other and there are no burst errors. As an effect, the impact of multiple losses is considered as the superposition of multiple losses. The models in the second category consider the correlation between error signals, giving rise to more complex loss patterns, burst of losses and losses separated with small lags, than just isolated losses [4]. Evidently, burst losses lead to larger distortion than individual single losses. In this case, the burst length affects the video quality in a distinct way and it has been determined analytically for different packet losses including burst errors and errors with lag,[12] and [7].

However, all the models mentioned above have not considered the inherent feature of H.264/AVC encoder that can select between a number of previously encoded frames highly correlated with the current frame, as reference for motion-compensated prediction of each inter-macroblock or macroblock partition. The use of multiple reference pictures allows H.264/AVC to achieve significantly better compression than any previous standard and on the same time, it affects the error propagation in the case of an error frame. It is evident from all the above studies that accurate distortion models are very important especially when decisions like rate-distortion optimization and packet scheduling are based on these models.

Due to the increase on multimedia transmission over the wireless channel today, multimedia streams share the same channel more frequently. Therefore, cases where users are allocated insufficient transmission bandwidth are very common. In such cases, the sender must reduce its video transmission rate by selecting which packets to drop prior to the transmission. Selecting the proper video packets to drop can become very trivial as it has a very significant effect on the reconstructed video. Several solution have been proposed for adapting the video characteristics to the transmission channel constraints. Video transcoding techniques [9], are applied in order to re-encode the video stream with lower bit rates but increase the complexity of the encoder dramatically. Scalable video coding [14] provides an inherent prioritization of the encoded video packets, thus it allows the sender to quickly select which packets to drop, however scalable coding has not been yet widely accepted. An extended work has been made on the video rate-distortion optimized transmission over wireless networks. [**?**] proposed a cross-layer ARQ for H.264 over 802.11 networks which gives priority to important packets during retransmission while, [11] describes an adaptive quality of service strategy for 802.11 networks that is applied to a single stream and without considering R-D optimization. Finally, [16] studies a R-D optimised bandwidth adaptation of multiple video streams that is performed by a network node that drops packets from all incoming streams.

The scope of this paper is to propose an intelligent packet scheduling mechanism for H.264/AVC video streams, based on a distortion model that accounts for the video coding characteristics and the use of multiple reference frames for motion-compensated prediction. In the case of limited transmission bandwidth the scheduler selects which combination of video frames will be dropped prior to transmission. Hence, the scheme is able to limit the increase in video distortion due to channel errors.

The rest of the paper is organized as follows. In Section 2, a novel distortion prediction model that accounts for the effect of multi-reference frames incorporated in the H.264 is presented. In Section 3, the proposed video coding and packet scheduling framework for providing optimized video streaming over 802.11 wireless lans is described. Furthermore, in Sections 4 and 5 we demonstrate how video streaming applications can benefit from the use of the proposed model. Finally, Section 6 draws the conclusions and discusses directions for further work and improvements.

2. NOVEL VIDEO DISTORTION MODEL

The following analysis considers a video sequence that begins with an *I*-frame and is followed by P-frames with an intra frame period N in order to increase error resilience. It is assumed that this intra frame period equals to the total error recovery period, in case of packet loss. As an effect, losses that occur outside this period are uncorrelated. Let k be the index of video picture. Then the total increase in Mean Square Error (MSE) distortion that will affect the video if picture k is lost is given by: $D(k) = \sum_{i=1}^{L} \Delta d_i$, where L is the number of video pictures in the sequence and Δd_i is the increase in MSE distortion relative to picture i, given that picture k is lost. It is assumed that previous frame concealment is used and there are no prior losses. It has been proved [12] that the MSE of subsequent frames will have a non-zero value, however due to the intra update and the spatial filtering, its amplitude decreases gradually until it becomes zero at a point far enough from frame k.

Based on this definition, the total video distortion due multiple lost frames M will be the sum of all individual MSEs over all the frames L affected by these losses. This is known as the additive distortion model as presented in [15] and [6].

$$D_{total} = \sum_{i=1}^{M} D(k) = \sum_{j=1}^{M} \sum_{i=1}^{L} \Delta d_i \tag{1}$$

We define a list of reference frames with size $MREF$ that is used during the encoding and decoding processes for motion-compensated prediction. Moreover, without loss of generality, each frame is coded into a single packet, although this condition can be extended to support different packetization schemes. Finally, in our analysis (similar to previous studies [17], [15] and [12]), a simple error concealment mechanism is used that in the case of a frame loss it replaces it with its previous at the decoder.

The error power introduced in a single frame k is denoted by $\sigma_s^2(k)$ and the total video distortion due to error frame k and its error power propagation to the following frames is denoted by $D_s(k)$. For more general loss patterns, $\sigma^2(k)$ and D are the MSE and the sum of the MSE values over all frames in the intra frame period, respectively. The proposed model includes analytical models for a single frame loss, a burst of losses with variable burst length and frame losses separated by a lag.

As it has already been determined in [15] and [12], the distortion metric consist of two factors: a geometric attenuation factor (due to spatial filter) and a linear attenuation factor (due to intra update). We have introduced a third parameter$MREF$ that accounts for the impact of the number of reference frames on the distortion propagation. Hence, the error power propagation at frame $k+l$, due to a single frame loss at k is:

$$\sigma^2(k+l) = \sigma^2(k) \cdot \Lambda_l \tag{2}$$

and the error power propagation effect is:

$$\Lambda_l = \left(1 - \frac{l-1}{N}\right) \cdot \left(r^{l-1} + \frac{r^{N-l+1}}{\Phi(N,r) + \left(\frac{MREF}{N}+1\right) \cdot \left(\frac{(N-l+1)}{N}\right)}\right) \tag{3}$$

Where, $\Phi(N,r)$ depends on the scene content of the particular video sequence and the coding parameters. This value has been estimated through curve fitting for different isolated errors. The value of this parameter for the video sequence *Foreman* is $\Phi(N,r) = \left(3 - \frac{2\cdot(N-l+1)}{N}\right) \times r$. Additionally, $r < 1$ is the spatial filtering factor.

In order to predict the total distortion caused by combinations of different error patterns that may include single errors, burst of errors and errors separated by a small lag, it is required to provide a more generic formula. In such a formula, we will take into account both the type of errors and the different error frame (Multi-Reference) dependencies that are present in H.264. Hence, in the general case of multiple combinations of erroneous frames the distortion D_n, where n is the error pattern size and $n \geq 1$, is modelled by the following recursive formula.

$$D_n = D_{n-1} - \sum_{\substack{k=F_{n-1} \\ i=0}}^{\substack{k=F_{n-1}+i \\ i=N-1+k}} \Lambda_{(i)} \cdot \sigma^2(F_{n-1}) + \sigma^2(F_{n-1}) +$$

$$+ \begin{cases} \sum\limits_{\substack{k=F_n \\ i=0}}^{\substack{k=F_n+i \\ i=N-1+k}} \Lambda_{(i)} \cdot \sigma_s^2(F_n), & uncorrelated \\ \sum\limits_{\substack{k=F_n \\ i=0}}^{\substack{k=F_n+i \\ i=N-1+k}} \Lambda_{(i)} \cdot \sigma^2(F_n), & burst \\ \sum\limits_{\substack{k=F_{n-1} \\ i=0}}^{\substack{k=F_n-1 \\ i=F_n-1-F_{n-1}}} \Lambda_{(i)} \cdot \sigma^2(F_n) + \sum\limits_{\substack{k=F_n \\ i=0}}^{\substack{k=F_n+i \\ i=N-1+k}} \Lambda_{(i)} \cdot \sigma^2(F_n), & lag \end{cases} \tag{4}$$

In the above recursive distortion model the frame number of the n^{th} erroneous frame is denoted by F_n. This recursive formula calculates the total distortion for the first error frame and depending on whether the next error frame is correlated or not with the previous error frame, it combines the above formulas and estimates the total distortion of the resulted error pattern.

3. PROPOSED SYSTEM MODEL

3.1 H.264 Overview

The Moving Picture Experts Group and the Video Coding Experts Group (MPEG and VCEG) have developed a new standard that promises to outperform the earlier MPEG-4 and H.263 standards, providing better compression of video images. The new standard "Advanced Video Coding" (AVC) and is published jointly as Part 10 of MPEG-4 and ITU-T Recommendation H.264 [2].

Some of the important terminology adopted in the H.264 standard is as follows:

- A field or a frame (of progressive or interlaced video) is encoded to produce a coded picture. A coded frame has a frame number, which is not necessarily related to decoding order and each coded field of a progressive or interlaced frame has an associated picture order count, which defines the decoding order of fields.

- Previously coded pictures (reference pictures) may be used for inter prediction of further coded pictures. Reference pictures are organised into one or two lists, referred to as list 0 and list 1.

- A coded picture consists of a number of macroblocks. Within each video picture, macroblocks are arranged in slices, where a slice is a set of macroblocks in raster scan order. An I slice may contain only I macroblock types, a P slice may contain P and I macroblock types and a B slice may contain B and I macroblock types.

H.264/AVC defines a set of three Profiles, each supporting a particular set of coding functions and each specifying what is required of an encoder or decoder that complies with the Profile. The Baseline Profile supports intra and inter-coding (using I-slices and P-slices) and entropy coding with context-adaptive variable-length codes (CAVLC). The Main Profile includes support for interlaced video, inter-coding using B-slices, inter coding using weighted prediction and entropy coding using context-based arithmetic coding (CABAC). The Extended Profile does not support interlaced video or CABAC but adds modes to enable efficient switching between coded bitstreams (SP- and SI-slices) and improved error resilience (Data Partitioning).

Potential applications of the Baseline Profile include videotelephony, video-conferencing and wireless communications; potential applications of the Main Profile include television broadcasting and video storage; and the Extended Profile may be particularly useful for streaming media applications. However, each Profile has sufficient flexibility to support a wide range of applications and so these examples of applications should not be considered definitive.

3.2 Intelligent Scheduling

There may be cases when the transmission bandwidth required from the sender exceeds the capacity limit of the shared wireless channel; hence the sender needs to decide which packets to drop or omit, while the quality degradation of the received video is minimized [6]. Without loss of the generality, the channel is considered lossless and the sender node decides which frames will be optimally dropped according to the distortion model proposed in Section II. The sender adapts its transmission rate to the current channel capacity and calculates the number of packets that needs to drop. As it has been shown, dropping a packet (considering that one packet is one frame) imposes a distortion that affects not only the current frame but all the correlated frames (multi-reference H.264). The intelligence of the packet dropping is based on the consideration of the correlation among the frames that populate the reference list. The distortion prediction model presented above enables the sender to drop the combination of packets that will cause the minimum overall distortion to the video stream.

The impact of dropping packets based on the proposed distortion model is compared against the dropping packets using the additive model in terms of PSNR. This means that the distortion model(s) decide(s) which packet must be dropped according to their importance at the video quality meter. The video sequences used in this experiment follow the coding setup described in Section III. Three standard sequences have been used namely Foreman, News and Salesman in QCIF format and 300 frames at 30fps. A transmission window of 36 frames/packets is considered for both the proposed and the additive model. We have investigated the case where in each transmission window a number of packets are dropped. Specifically, five error patterns have been considered in this study, i.e. 1 to 5 packets are lost at each transmission window. If we increase the number of lost packets the perceived picture quality deteriorates at unacceptable levels. The selected packet drop probabilities correspond to five error frame patterns or combinations according to the proposed and the additive distortion model. For reasons of simplicity this paper presents the results for the worst case scenario according to which the restrained transmission bandwidth requires 40 video packets in total, to be dropped prior to transmission, resulting to a 13% loss probability.

3.3 802.11 Mac Protocol

The DCF [5] access mode is based on carrier sense multiple access with collision avoidance (CSMA/CA) in principle. Each station in a single BSS (basic service set) contends for shared medium to transmit its data packet. If the medium is busy, the station defers its transmission and initiates a backoff timer. The backoff timer is randomly selected between 0 and *Contention Window* (CW). When the station detects that the medium has been free for a duration of DCF interframe spaces (DIFS), it begins to decrement the backoff counter as long as the channel is idle. As the backoff timer expires and the medium is still free, the station begins to transmit. In case of a collision, indicated by the lack of an acknowledgment, the size of the CW is doubled following the (5) until it reaches the CW_{max} value. Furthermore, after each successful transmission, the CW is initialized with CW_{min},

$$CW = (CW_{min} x 2^i) - 1 \qquad (5)$$

where i is the number of transmission attempts. Unlike wired networks, it is relatively difficult to detect collisions in wireless environments due to significant differences between the transmitted and received power levels. Hence, an ACK (acknowledgement) packet should be immediately sent by the receiver after a predefined period (SIFS) upon successful reception of a data packet. Following a successful packet transmission, a station is required to wait for DIFS interval. A random back-off process is then started to prevent collisions between stations. Such a station can initiate its transmission only if the medium remains idle for additional random time called back-off interval. Meanwhile, if an ACK packet is not received within a time-out interval, the sender assumes that the packet loss has been occurred and a back-off procedure is also initiated prior to retransmitting the data packet.

4. SIMULATION SETUP

This section evaluates the performance of the proposed system model through a set of simulations. A NS-2 based simulation environment with the appropriate extensions [1] for simulating 802.11 WLANs is adopted.

Two YUV *Quarter Common Intermediate Format* (QCIF) (176x144) raw video sequences consisting of 300 frames are used as video sources. The video sequences are encoded at a frame-rate of 30 frames/sec using H.264/AVC video compression standard. A GOP length is 36 frames and its structure is IPPPPP... has been used. The video frame are then encapsulated into RTP packets using a simple packetization scheme [13] (by one-frame-one-packet policy). The size of each RTP packet is maximally bounded to 1024 bytes. The generated video packets are delivered through the 802.11 MAC at the form of UDP/IP protocol stack.

The 802.11b is employed for the physical layer, which provides four different physical rates. In our simulation, the physical rates are fixed to 11 Mbps for data and 2Mbps for control packets. The 802.11 MAC parameters are set to the default values of the standard specification (CWmin=15, CWmax=1023, and retry limit=7). Additionally, the streaming node station generates background traffic (500 kb/s) using constant bit rate (CBR) traffic over User Datagram Protocol (UDP). This allows us to increase the virtual collisions at the server's MAC layer. Furthermore, we include five wireless stations where each station generates 500 kb/s of data using CBR traffic in order to overload the wireless network.

5. RESULTS AND DISCUSSION

In this section we examine via simulation experiments the performance of the proposed distortion model and we validate its accuracy by comparing it to real measurements. The model has been tested over several error patterns with single, burst and errors with lag. Furthermore, we examine the performance of the proposed intelligent packet scheduling and we compare it with the simplified additive distortion model.

Fig. 1 illustrates the distortion effect for the loss of a single frame in two cases, where 5 and 10 frames are used as reference frames respectively for the Foreman video sequence. The comparison between the proposed model and the measured distortion clearly has indicated that the proposed model accurately captures the distortion effect as long as there are reference frames stored in the decoder for motion vector prediction.

Fig 2. illustrates the MSE for different burst error lengths. It clearly shows that the distortion due to varying burst error length is not equivalent to the sum of isolated losses, which is also consistent with [15]. Apparently, the proposed model is very accurate in the calculation of the distortion, allowing only a deviation from the actual data of 0.4 dB.

Figs. 3, 4 and 5 illustrate the perceived video quality in terms of PSNR due to the intelligent packet dropping for both the proposed distortion model and the simplified ad-

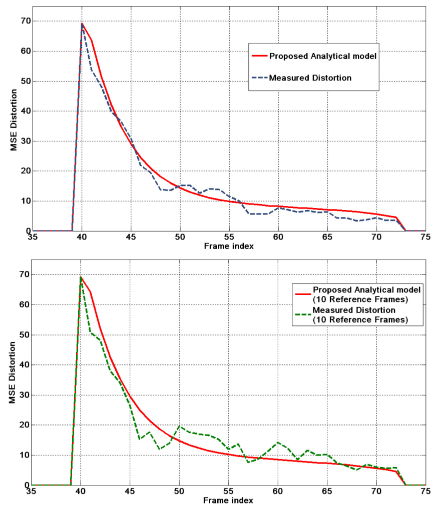

Figure 1: Single frame loss and error propagation. (a) reference frame list size MREF= 5 (b) reference frame list size MREF= 10.

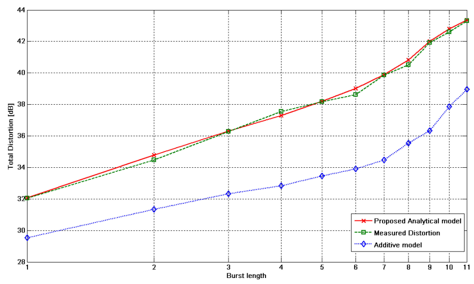

Figure 2: Total Distortion vs burst error length.

ditive model. It can be noticed that there is a significant improvement in the perceived PSNR from the proposed distortion model. This is explained by the fact that the proposed distortion model predicts and identifies not only the most important frames from their size but also their effect to neighboring frames due to its correlation with them. Therefore, in the decision upon which frame to omit prior to its transmission, the proposed distortion model considers the distortion propagation effect in conjunction with the reference frame list size.

Fig. 6 and 7 show the overall PSNR performances of the intelligent packet scheduling and the scheduling scheme based on the additive model for the two sequences *Foreman* and *Salesman* as a function of the available data rate of the shared wireless channel. It can be seen that the proposed scheduling outperforms the additive model over the whole range of values considered. This is due to the fact that the proposed scheme selects the combination of packets to drop that have the minimum distortion effect on the reconstructed video, as opposed to the additive model that just considers the uncorrelated distortion effects of every single packet and calculates the sum of the single distortions. In the case of *Foreman* sequence it is clear that the performance of the proposed scheme increases as the available rate increases, while in the case of *Salesman* sequence the proposed scheduling algorithm and the additive model

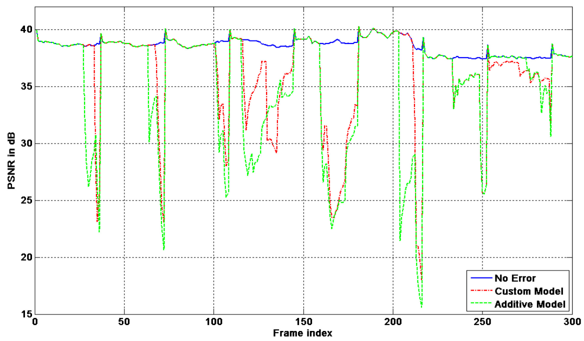

Figure 3: PSNR comparison between Proposed and Additive models for video sequence Foreman

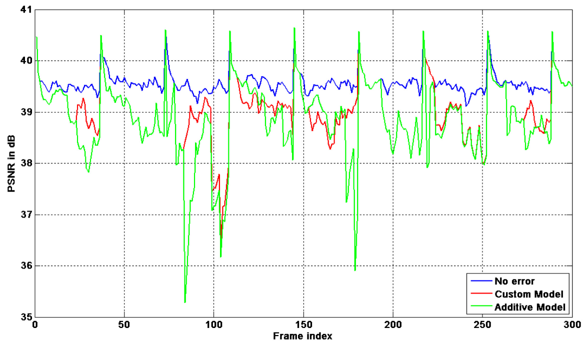

Figure 4: PSNR comparison between Proposed and Additive models for video sequence News

converge as the available transmission rate increases. This is due to the fact that the latter sequence includes static scenes and more highly correlated video frames, hence both the proposed distortion model and the additive models have the same impact on high rates. This is not the case for *Foreman* sequence, which is characterized by highly uncorrelated frames and arbitrary motion. In this case the proposed algorithm has a profound advantage over the additive model by 0.5 to 1 dB in PSNR.

6. CONCLUSIONS

This paper has presented an intelligent packet scheduling algorithm using an analytical distortion model for multi-referenced H.264/AVC video over error-prone channels. Unlike previous research studies, the proposed model considers the use of reference frames for motion-compensated prediction and it holds its validity for different reference list sizes. The proposed distortion model has been extended for complex error patterns and takes into consideration the effect of reference frames in the error power propagation to neighboring frames. Extended simulations proved that the proposed distortion model is very accurate (0.4 dB) as compared with the real measurements and captures the distortion effect more accurately than the additive model.

Additionally, the paper presented an intelligent packet dropping scheme based on the above model. The packet dropping that is required in the case of the available transmission rate is limited, takes into consideration the impact of the loss of the frame on the resulting video distortion that depends not only on the distortion of the particular frame

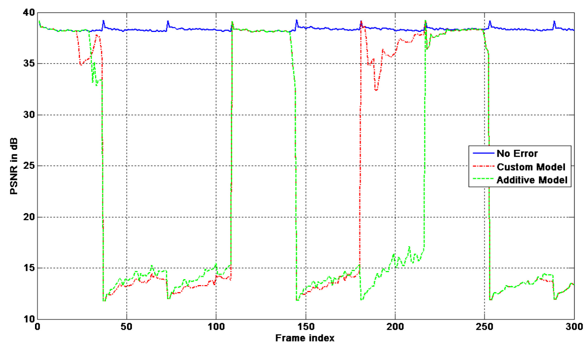

Figure 5: PSNR comparison between Proposed and Additive models for video sequence Salesman

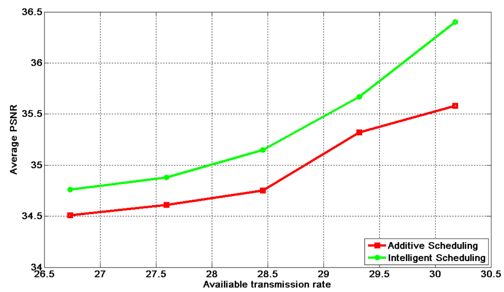

Figure 6: Average PSNR in dB versus available data rate in Kbps for video sequence Foreman

but on the distortion effect that propagates to the correlated frames. Simulations for different error patterns proved that the proposed packet dropping scheme significantly improves the received video quality, in terms of PSNR, when compared with the simplified additive distortion model.

Acknowledgment

This work was supported by the project PENED-No.03$E\Delta$636, which is funded in 75% by the European Social Fund and in 25% by the Greek State-General Secretariat for Research and Technology.

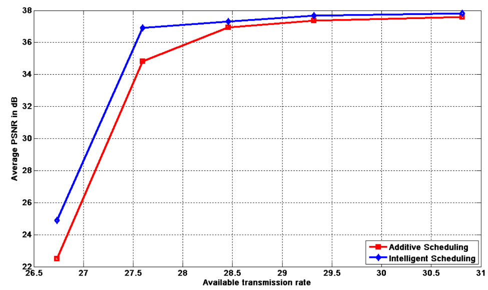

Figure 7: Average PSNR in dB versus available data rate in Kbps for video sequence Salesman

7. REFERENCES

[1] Network simulator v.2 (ns-2) dcf and edca extensions. *http://www.tkn.tu-berlin.de/research/802.11e_ns2.*

[2] I. 14496-10 and I.-T. Rec. H.264 standard, advanced video coding. 2003.

[3] I. 802.11g 2003. Part11 wireless lan medium access control (mac) and physical layer (phy) specification band. *Supp. IEEE 802.11*, 2003.

[4] J. Apostolopoulos, W. Tan, and G. Wornell. Modeling path diversity for multiple video communications. *ICASSP02 Conference Proceedings*, May 2002.

[5] G. Bianchi. Performance analysis of the ieee 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–547, March 2000.

[6] J. Chakareski and P. Frossad. Rate-distortion optimized distributed packet scheduling of multiple video streams over shared communication resources. *IEEE Transaction on Multimedia*, 8(2):207–219, April 2006.

[7] F. Fitzek and M. Reisslein. Mpeg-4 and h.263 video traces for network performance evaluation. *IEEE Network*, 15(6):40–54, December 2001.

[8] M. Ghanbari. *Standard codecs: Image compression to advanced video coding.* IEE Press, 2003.

[9] C. L. J. Xin and M. T. Sun. Digital video transcoding. In *Proc. IEEE*, volume 93, pages 84–97, Jan 2005.

[10] I. Kim and H. Kim. A new resource allocation scheme based on a a psnr criterio for wireless video transmission to stationary receivers over gaussian channels. *IEEE Transanctions on Wireless Communications*, 1(3):393–401, July 2002.

[11] 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 Tran. on Multimedia*, volume 6, pages 278–290, Apr 2004.

[12] Y. Liang, J. Apostolopoulos, and B. Girod. Analysis of packet loss for compressed video: Does burst-length matter. *Proc. Int'l Conf. Acoustics, Speech, and Signal Processing, Hong Kong, China*, April 2003.

[13] H. Schulzrinne. Rtp: A transport protocol real-time applications. *RFC 1889*, 1996.

[14] M. V. S.R.McCanne and V. Jacobson. Low complexity video coding for receiver-driven layered multicast. *IEEE J. Sel. Areas Commun.*, volume 15, pages 983–1001, Aug 1997.

[15] K. Stuhlmuller, N. Farber, M. Link, and B. Girod. Analysis of video transmission over lossy channels. *IEEE Journal on Selected Areas in Communications*, 18(6):1012–1032, June 2000.

[16] W. K. W. Tu and E. Steinbach. Rate-distortion optimized video frame dropping on active network nodes. *Proc. Int. Packet Video Workshop*, Dec 2004.

[17] T. Wiegand, 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*, 13(7):550–576, July 2003.

[18] F. Wu, G. Shen, K. Tan, F. Yang, and S. Li. Next generation mobile multimedia communications: Media codec and media transport perspectives. *China Communications Magazine*, pages 30–44, October 2006.