

Research Article

Distortion Optimized Packet Scheduling and Prioritization of Multiple Video Streams over 802.11e Networks

Ilias Politis,¹ Michail Tsagkaropoulos,² Thomas Pliakas,³ and Tasos Dagiuklas⁴

¹ Department of Electrical & Computer Engineering, University of Patras, Greece

² Mobility Applications Laboratory - Networks, Ericsson Deutschland GmbH - Eurolab, Germany

³ Department of Information & Communication Systems Engineering, University of Aegean, Greece

⁴ Department of Communication Systems and Networks, Technological Educational Institute of Messolonghi, Greece

Received 11 June 2007; Accepted 20 August 2007

Recommended by Stavros Kotsopoulos

This paper presents a generic framework solution for minimizing video distortion of all multiple video streams transmitted over 802.11e wireless networks, including intelligent packet scheduling and channel access differentiation mechanisms. A distortion prediction model designed to capture the multireferenced frame coding characteristic of H.264/AVC encoded videos is used to predetermine the distortion importance of each video packet in all streams. Two intelligent scheduling algorithms are proposed: the “even-loss distribution,” where each video sender is experiencing the same loss and the “greedy-loss distribution” packet scheduling, where selected packets are dropped over all streams, ensuring that the most significant video stream in terms of picture context and quality characteristics will experience minimum losses. The proposed model has been verified with actual distortion measurements and has been found more accurate than the “additive distortion” model that omits the correlation among lost frames. The paper includes analytical and simulation results from the comparison of both schemes and from their comparison to the simplified additive model, for different video sequences and channel conditions.

Copyright © 2007 Ilias Politis et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

The ever growing demand for multimedia services send over wireless access networks leads to cases where multiple users are competing over scarce channel resources. Typical example of such cases is the concurrent media streaming over bottleneck network nodes or the wireless lan-shared channel capacity. Therefore, it is important to optimize the entire transmission system in order to achieve high performance and increase quality of service [1].

In this context, next wireless LAN (WLAN) generation could provide multimedia services to mobiles and fixed users through wireless access, with the development of the high-speed physical (PHY) layers IEEE 802.11g (54Mbps) [2]. 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 their low cost and easy deployment. The IEEE 802.11 provides a best effort service, which indicates that every data

packet-handed over to the 802.11 interfaces receives similar treatment as other packets in terms of delivery guarantees, that is, available bandwidth, latency, jitter, and so forth. It provides two access mechanisms, in order for the mobile terminals to gain access to shared wireless medium [2]; (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 [3]. In order to deliver real-time video traffic, which is sensitive to packet latency and effective bandwidth characteristic of the underlying network, QoS, and differential service [4–6], the 802.11e [7] seems to be an appropriate solution. It introduces the hybrid coordination function (HCF) that concurrently uses a contention-based mechanism and a pooling-based mechanism, enhanced DCF (EDCF), and HCF controlled channel access (HCCA), respectively. Like DCF, EDCA is very likely to be the dominant channel access mechanism in WLANs because it features a distributed and easily deployed mechanism.

As far as the application plane is concerned, video streaming over emerging wireless networks is becoming ever more popular, hence it is important to develop error resilient source coding and transmission techniques [8, 9]. In detail, the video quality of a decoded video transmitted over a wireless network is often associated with the average pixel-by-pixel distortion of the video frames. This distortion results from both the compression scheme of the source encoder and the transmission 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. Modeling channel distortion is accurately the key for rate-distortion optimization and end-to-end quality of video communications. There is a large number of research works that has been reported and regards modeling the impact of packet loss on video distortion. Such models can be fall in two categories. In the first category, the models consider that distortion is proportional to the number of losses within video sequences [10, 11]. 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. 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 [12–14]. 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 intermacroblock 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. Accurate distortion models are very important especially when decisions like rate-distortion optimization and packet scheduling are based on these models.

In the case of concurrent multiple media streams, it is important to develop effective algorithms that will minimize the average distortion of each user. Therefore, there are often cases where users are allocated insufficient transmission bandwidth. In such occasions, the sender has to be able to 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 crucial, as it has a very significant effect on the reconstructed video. Several solutions have been proposed for adapting the video characteristics to the transmission channel constraints. Video transcoding techniques [15] are applied in order to reencode the video stream with lower bit rates but increase the complexity of

the encoder dramatically. Scalable video coding [16] 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 [17] and has proposed a cross-layer ARQ for H.264 over 802.11 networks which gives priority to important packets during retransmission while [18] 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, [19] studies an R-D optimized bandwidth adaptation of multiple video streams that is performed by a network node that drops packets from all incoming streams.

This paper proposes a novel framework based on two scheduling algorithms named “even-loss distribution” and “greedy-loss distribution” for effective packet scheduling and optimized differentiated channel access that minimizes the overall distortion of the video streams. The framework is based on a new distortion prediction model that takes into consideration the high correlation among reference frames of H.264/AVC encoded video streams, the size of the reference list used and the type of error patterns (single errors, burst of errors, errors separated with lag) that are imposed by the transmission channel. The proposed distortion model has been evaluated against real measurements and is compared with the simplified “additive”-based distortion model that omits the correlation among the frames. This model is power efficient as it requires simplified calculations that can be performed prior to the video transmission. Each of the video senders is informed by the system about the current available capacity of the communication channel and allocates its transmission rate accordingly. The tradeoff between transmission rate and video distortion is different for the two proposed scheduling algorithms. The transmitted packets are then processed through the proposed 802.11e MAC mechanism that is able to differentiate the video streams according to their distortion imposed by the scheduling algorithm. The video sequence that has experienced the greater losses is treated as high-priority traffic and is allocated to the appropriate access category. Analytical results indicate the advantages and disadvantages of the proposed scheduling schemes. It is evident that video streams with different type of context and coding characteristics are affected in different way by the “even-loss distribution” and the “greedy-loss distribution” scheduling algorithms.

The rest of the paper is organized as follows. Section 2 provides an overview of the additive distortion model and the mathematical background of the proposed model. In Section 3, the system model is presented. The setup of the 802.11e system and the configuration of the H.264/AVC encoder are discussed. In Section 4, the “even-loss distribution” and the “greedy-loss distribution” scheduling schemes are presented. The simulation setup and an extended discussion upon the results are provided in Section 5 and the paper concludes in Section 6. Appendices at the end of the paper discuss the mathematical model evaluation against actual measurements and compare the behavior of the proposed model and that of the “additive” distortion model, for several

different error patterns (single losses, burst of losses, losses separated with lag).

2. ANALYTICAL DISTORTION MODEL

2.1. Overview of additive model

The following analysis considers a video sequence that begins with an I -frame and is followed by P -frames with an intraframe period N in order to increase error resilience. It is assumed that this intraframe period equals 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 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 [13] that the MSE of subsequent frames will have a nonzero value, however due to the intraupdate 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 [1, 10]:

$$D_{\text{total}} = \sum_{j=1}^L D(j) = \sum_{j=1}^L \sum_{i=1}^L \Delta d_i(j). \quad (1)$$

2.2. Proposed model

We define a list of previously encoded reference frames with size M_{REF} 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 [10, 13, 20]), 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 $D_s(k)$. Correspondingly, $\sigma^2(k)$ and D are the MSE and the sum of the MSE values over all frames in the intraframe period, of more general loss patterns, 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 [10, 13], the distortion metric consists 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 M_{REF} 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

$$\sigma^2(k + l) = \sigma^2(k) \cdot \Lambda_l, \quad (2)$$

and the error power propagation effect is

$$\Lambda(l) = \left(1 - \frac{l-1}{N}\right) \times \left(r^{l-1} + \frac{r^{N-l+1}}{\Phi(N, r) + ((M_{\text{REF}}/N) + 1)((N-l+1)/N)}\right), \quad (3)$$

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*, that was used as reference sequence, is $\Phi(N, r) = (3 - (2 \cdot (N - l + 1)/N)) \times r$. Additionally, $r < 1$ is the spatial filtering factor. Considering a single error at frame k and a period long enough for error recovery, the total video distortion is

$$D_s(k) = \sigma_s^2(k) \cdot A = \sigma_s^2(k) \cdot \sum_{i=0}^{N-1-k} \Lambda_i, \quad (4)$$

where $\sigma_s^2(k)$ is the error power introduced at k and A represents the propagation effect of the initial distortion at k over all the following frames i until the end of the intraframe period N . In order to extend the distortion model for a single loss to a burst of frame losses with a burst length B , we need to consider the correlation between these frames. Assume a correlation coefficient among two error frames $e_s(k-1)$ and $e_s(k)$ to be

$$\rho_{k-1,k} = \frac{(e_s^T(k-1) \cdot e_s(k))/M}{\sigma_s(k-1) \cdot \sigma_s(k)}, \quad (5)$$

where $\sigma_s(k-1)$ and $\sigma_s(k)$ are the MSEs of frames $k-1$ and k accordingly [13]. $M = M_1 \times M_2$ is the 2D array of pixels in each frame k . Then the distortion due to a burst of error frames is

$$\begin{aligned} D(k-B+1, \dots, k) &= \sum_{i=k-B+1}^{k-1} \sigma^2(i) + A(B) \cdot \sigma^2(k), \\ A(B) &= A_0 + c \cdot (B-2), \quad B \geq 2, \\ \sigma^2(k) &= \sum_{i=k-B+1}^k \sigma_s^2(i) + 2 \cdot \sum_{i=k-B+1}^k \sum_{j=i+1}^k \rho_{i,j} \cdot \sigma_s(i) \cdot \sigma_s(j). \end{aligned} \quad (6)$$

In (6), $\sigma^2(i)$ is the sum of the distortion over all i frames, B is the burst length and $B \geq 2$, and $\rho_{i,j}$ is the correlation coefficient of frames i and j . $A(B)$ represents the variation on the shape of the introduced error power relative to burst length B and A_0 is the ratio for $B = 2$. In [13] it was shown that as the burst length B varies, the shape of the initial error signals PSD also varies, which leads to a variation in the

spatial filtering factor r in (3). The process of error power reduction by loop filtering can be modeled with a linear system, and r is the proportion of the power of the introduced error passing through the system. In [10], the loop filter is approximated by a Gaussian lowpass filter. Hence, as B increases, r increases as the PSD of the error is more concentrated in the lower band. The variation of r is relatively small and can be approximated as a linear function of B , that is, $A(B) = A_0 + c \cdot (B - 2)$, where A_0 is the ratio for $B = 2$ and c is the slope of the increase.

Extending the above analysis in order to predict the video distortion in the event of two losses separated by short lag l smaller than the intraframe period N , the distortion can be expressed by the following equation:

$$D(k-1, k) = \sigma_s^2(k-1) + \sum_{i=0}^{l-1} \sigma^2(i) + A \cdot \sigma^2(k), \quad l \in (1, N]. \quad (7)$$

The above distortion prediction model represented by equations (4), (6), and (7) is based on specific error patterns (isolated errors, burst errors, and errors with lag). In order to predict the distortion caused by more complex error patterns that may include different combinations of the above, it is required to provide a more generic formula. This formula will take into account both the type of error and the different error frame (multireference) dependencies that are present in H.264/AVC. 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 modeled by recursive formula (8). In (8), D_1, D_2, \dots, D_n are the total video distortions due to 1, 2n error frames. 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:

$$D_1 = \sum_{\substack{k=F_1+i \\ i=N-1+k}}^{k=F_1+i} \Lambda_{(i)} \cdot \sigma^2(F_1) = \sum_{k=F_1, i=0}^{k=F_1+i, i=N-1+k} \Lambda_{(i)} \cdot \sigma_s^2(F_1),$$

$$D_2 = \sigma_s^2(F_1) + \begin{cases} \sum_{\substack{k=F_2+i \\ i=N-1+k}}^{k=F_2+i} \Lambda_{(i)} \cdot \sigma_s^2(F_2), & \text{uncorrelated,} \\ \sum_{\substack{k=F_2+i \\ i=N-1+k}}^{k=F_2+i} \Lambda_{(i)} \cdot \sigma^2(F_2), & \text{burst,} \\ \sum_{\substack{k=F_2-1 \\ i=F_2-1-F_1}}^{k=F_2-1} \Lambda_{(i)} \cdot \sigma^2(F_2) + \sum_{\substack{k=F_2+i \\ i=N-1+k}}^{k=F_2+i} \Lambda_{(i)} \cdot \sigma^2(F_2), & \text{lag,} \end{cases}$$

$$\vdots$$

$$D_n = D_{n-1} - \sum_{\substack{k=F_{n-1}+i \\ i=N-1+k}}^{k=F_{n-1}+i} \Lambda_{(i)} \cdot \sigma^2(F_{n-1}) + \sigma^2(F_{n-1}) + \begin{cases} \sum_{\substack{k=F_n+i \\ i=N-1+k}}^{k=F_n+i} \Lambda_{(i)} \cdot \sigma_s^2(F_n), & \text{uncorrelated,} \\ \sum_{\substack{k=F_n+i \\ i=N-1+k}}^{k=F_n+i} \Lambda_{(i)} \cdot \sigma^2(F_n), & \text{burst,} \\ \sum_{\substack{k=F_n-1 \\ i=F_n-1-F_{n-1}}}^{k=F_n-1} \Lambda_{(i)} \cdot \sigma^2(F_n) + \sum_{\substack{k=F_n+i \\ i=N-1+k}}^{k=F_n+i} \Lambda_{(i)} \cdot \sigma^2(F_2), & \text{lag.} \end{cases} \quad (8)$$

The proposed model presented in this section has been evaluated against real measurements. Extensive simulations have been performed for different combinations of error patterns. These include single isolated frame losses, burst of frame losses with burst length $B \geq 2$, and error frames separated by a small lag. Additionally, the proposed model was compared against the additive distortion model introduced in [10]. The simulations showed that our model can accurately capture the distortion effect for reference-based H.264 coding. Appendices at the end of the paper present the simulation results for the *Foreman* testing sequence.

3. SYSTEM MODEL

3.1. IEEE 802.11e EDCF

The IEEE 802.11e, recently established working group of IEEE 802.11e, provides applications with differential services and QoS support by priority-based contention service, that is, the enhanced DCF (EDCF). A concept of traffic category is introduced. Each traffic category is associated with the predetermined contention parameters, arbitration IFS (AIFS), CW_{\min} and CW_{\max} , and the backoff persistence factor (PF) [21]. The lower AIFS/ CW_{\min} / CW_{\max} results in the higher probability of winning the channel contention. In EDCF, the contention window is expanded by the PF after collision. In the original DCF, the contention window is always doubled after collision ($PF = 2$), while in EDCF, PF may be of a different value. For each station, multiple traffic categories with different contention parameters can exist in parallel, thus leading to internal contention in each station. The collisions among internal contention are avoided by letting the highest priority traffic category win the contention window.

The values of AIFS[AC], CW_{\min} [AC], and CW_{\max} [AC], which are referred to as the EDCF parameters, are announced by the AP via beacon frames. The AP can adapt these parameters dynamically depending on network conditions. Basically, the smaller AIFS[AC] and CW_{\min} [AC], the shorter the channel access delay for the corresponding priority, and hence the more capacity share for a given traffic condition. However, the probability of collisions increases when operating with smaller CW_{\min} [AC]. These parameters can

be used in order to differentiate the channel access among different priority traffic.

3.2. H.264/AVC standard 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) is published jointly as Part 10 of MPEG-4 and ITU-T Recommendation H.264 [22]. Some of the important terminology adopted in the H.264 standard are as follows.

- (1) 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.
- (2) Previously coded pictures (reference pictures) may be used for interprediction of further coded pictures. Reference pictures are organized into one or two lists, referred to as list 0 and list 1.
- (3) A coded picture consists of a number of macroblocks. Within each 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 intercoding (using I -slices and P -slices) and entropy coding with context-adaptive variable-length codes (CAVLC). The main profile includes support for interlaced video, intercoding using B -slices, intercoding 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, videoconferencing, 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.

The innovative feature on the development of H.264 encoder is the ability to use one or two of a number of previously encoded pictures (or frames) as a reference for motion-compensated prediction of each intercoded macroblock or macroblock partition. This enables the encoder to search for the best match (the most correlated picture) for the current macroblock partition from a wider set of pictures than just the previously encoded picture. The encoder and de-

coder each maintains one or two lists of reference pictures, containing pictures that have previously been encoded and decoded (occurring before and/or after the current picture in display order). Intercoded macroblocks and macroblock partitions in P slices are predicted from pictures in a single list, *list 0*. Intercoded macroblocks and macroblock partitions in a B slice may be predicted from two lists, *list 0* and *list 1*. The configuration of the H.264/AVC encoder used for the following simulations includes a reference *list 0* that contains $M_{\text{REF}} = 5$, number of previously encoded pictures for motion compensated prediction, although in the appendices there are simulation results for a reference list of size $M_{\text{REF}} = 10$ that prove that the proposed recursive model of (8) is valid for any *list 0* as long as $M_{\text{REF}} \geq 2$.

4. PROPOSED SCHEDULING ALGORITHMS

4.1. Even-loss distribution

In this scenario, each video sender has precalculated the distortion effect of each packet and the total video distortion of all the combination of dropped packets. The error combinations may include individual packet losses, burst of dropped packets with burst length $B \geq 2$, or dropped packets separated by a short lag and any combination of the above. These calculations are repeated in each transmission window. In the following analysis, each video sender can transmit packets during a sliding window 36 frames long. The packets in each transmission window, hence the resulted losses, are considered correlated. Each video sender is informed by the shared network about the available transmission rate. According to the precalculated video distortion, each sender may select the optimum pattern of packets to drop in order to optimize its transmission rate. The “even-loss distribution” algorithm ensures that each video stream will be evenly affected by the channel restrictions. Therefore, each video sender will be required to drop on average an equal number of packets. Although which packets will select to drop depends on the distortion prediction model.

Each of the video streams is characterized by the average distortion due to the packet scheduling. The 802.11e network has the ability to differentiate among different priority traffic depending on several QoS criteria. According to the “even-loss distribution” scenario, the 802.11e model that is applied has been configured to differentiate among the multiple video streams based on their average distortion. As it was said before, the number of dropped packets may be equal on the average, among the video streams, however the resulting distortion is different for each video stream. Therefore, the channel access differentiation that is applied in this scenario indicates that the video stream with the highest distortion will be considered as high-priority traffic. The trade-off between the high distortion due the “even-loss distribution” scheduling and the low-loss probability due to prioritization ensures that the resulting average distortion of all video streams will be kept minimum.

TABLE 1: 802.11e MAC parameters.

Access category	AIFS	CW_{\min}	CW_{\max}	Queue length	Max retry limit
AC3	50	7	15	50	8
AC2	50	15	31	50	8
AC1	50	31	1023	50	4
AC0	70	31	1023	50	4

4.2. Greedy-loss distribution

On the other hand, “greedy-loss distribution” scheduling ensures that the most significant video stream will be less affected by packet scheduling. The algorithm considers packet scheduling jointly for all video streams. Thus, the limited channel capacity available for video transmission will cause a larger number of packets to be dropped by these video senders that their video streams are least significant in terms of context, coding, and QoS characteristics. In our analysis, a video sequence that can be classified as significant for the “greedy-loss distribution” algorithm is a video sequence that the frames are highly correlated with each other, and a loss of one will result to a large distortion effect over all correlated frames. This phenomenon is even more distinct in the case of multireference H.264/AVC coding. A lost reference frame will result in a distortion effect that will not fade out quickly and will affect all the forthcoming frames.

The “greedy-loss distribution” algorithm will cause uneven losses among the multiple video streams. However, this will be balanced by the priority categorization of the 802.11e configuration. Similar to the previous algorithm, the prioritization scheme applied by the system will differentiate the video streams in terms of average distortion due to packet dropping. The tradeoff among the video distortion and the priority categorization has different effects on the video streams. The following section presents the simulation results and includes a discussion over the interesting findings.

5. SIMULATION SETUP AND RESULTS

This section evaluates the performance of the proposed framework through a set of simulations. An NS-2-based simulation environment with the appropriate extensions for simulating 802.11e WLANs is adopted [23]. Three YUV Quarter Common Intermediate Format (QCIF) 4:2:0 color video sequences consisting of 300 and coded at 30 frames per second are used as video sources (*Foreman*, *News*, and *Salesman*). Each group of pictures (GOP) is structured as *IPPPPPPP*. and contains 36 frames, and the maximum UDP packet size is at 1024 bytes (payload only). The H.264/MPEG-4 AVC encoder provided in [24] is used for encoding YUV sequences. We simulate a unicast H.264 video transmission (one video server and one video client) utilizing a *basic service set* (BSS) at 1024 Mbps. The video frames are then encapsulated into RTP packets using a simple packetization scheme [25] (by one-frame one-packet policy). The size of each RTP packet is maximally bounded to 1024 bytes. Additionally, the server station generates background traffic

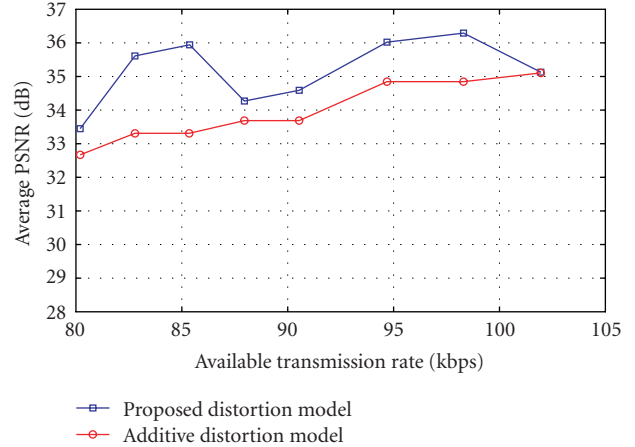


FIGURE 1: Average PSNR (in dB) as a function of available transmission rate (in kbps) for the “even-loss distribution” algorithm. Comparison between the proposed and the simplified “additive” distortion model for the video sequence *Foreman*.

(300 Kbps) using constant bit rate (CBR) over UDP, in order to increase the virtual collisions at the server’s MAC layer. Moreover, we include four wireless stations where each station generates 300 kbps of data using CBR of data, overloading the simulated 802.11e wireless network. The simulated 802.11e draft defines four access categories; AC3 corresponds to the highest access priority, and AC0 to the lowest. Based on this traffic specification, it is possible to differentiate the H.264 frames at the MAC layer according to distortion.

The generated video packets are delivered through the 802.11e at the form of UDP/IP protocol stack. In our simulation, Table 1 depicts the MAC parameters for the simulations.

5.1. Even-loss distribution scenario

In this scenario, we examine the performance of the proposed distortion model (PDM) and the “additive” model of (1) when they are applied in the “even-loss distribution” scheduling algorithm. The available transmission bandwidth is insufficient for the transmission of all three video streams hence the scheduling algorithm is applied to distribute the losses evenly among the streams. The video streams will adapt their transmission rate by selecting to drop the optimum combination of packets. The number of packets to drop is indicated by the “even-loss distribution” algorithm, however which packets will be selected depends on each video source and the distortion model that is used to

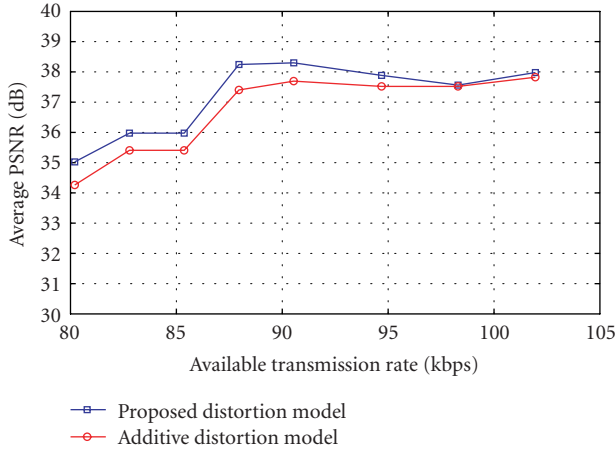


FIGURE 2: Average PSNR (in dB) as a function of available transmission rate (in kbps) for the “even-loss distribution” algorithm. Comparison between the proposed and the simplified “additive” distortion model for the video sequence *News*.

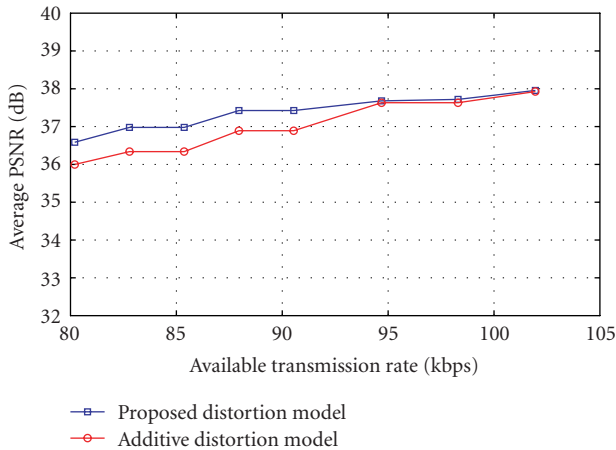


FIGURE 3: Average PSNR (in dB) as a function of available transmission rate (in kbps) for the “even-loss distribution” algorithm. Comparison between the proposed and the simplified “additive” distortion model for the video sequence *Salesman*.

pre-calculate packet distortion. In this case, there are two distortion models that are compared, the PDM presented in Section 2 and the “additive” model that omits the correlation among the video pictures. Figures 1, 2, and 3 show the overall Y-PSNR (dB) performances of PDM and “additive” overall three sequences as a function of the available data rate (kbps) on the shared 802.11e channel. It can be seen that PDM outperforms “additive” over the whole range of values considered for the available transmission rate. This is due to the fact that PDM exploits the knowledge about the correlation of referenced frames and can therefore predict accurately the optimum combination of dropping packets that will result in minimum distortion on the reconstructed video. Thus, by applying the PDM, users will drop packets from their trans-

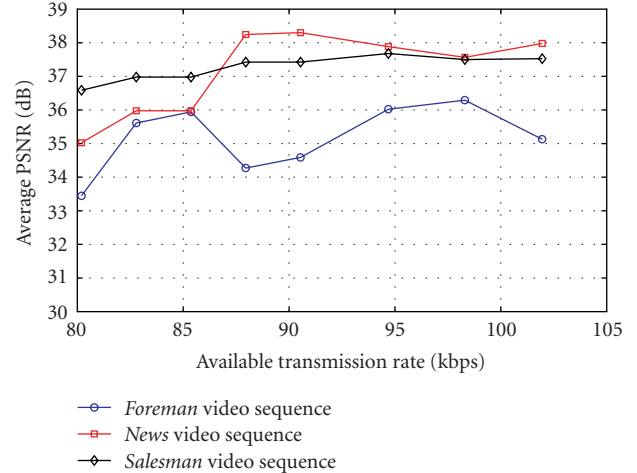


FIGURE 4: Average PSNR (in dB) as a function of available transmission rate (in kbps) for the “even-loss distribution” algorithm. Comparison between the proposed distortion model for all testing video sequences.

mission windows that will have the least impact on the overall quality of the reconstructed videos. As can be seen from the figures, the performance gain of PDM over “additive” is 1 dB on average for the *Foreman* sequence and reaches 2.5 dB when the transmission rate is 85 Kbps. The same applies for the *News* sequence of Figure 2, where on average the distortion due to packet dropping is improved with PDM by 0.5 dB compared with the “additive” model. Figure 3 shows that 0.5 dB on the average is the PDM gain against “additive” for the *Salesman* sequence as well. Additionally, in Figures 2 and 3, it is evident that as the available transmission increases the two distortion models are converging since the dropped packets are less and the resulting video quality degradation is minimum. This phenomenon is not clear in the case of *Foreman* since the “even-loss distribution” algorithm allocates the number of packets to drop evenly to all video streams. This affects *Foreman* sequence significantly as the context of this particular video contains fast moving pictures and motion scenes with low correlation among video frames. It is evident in Figure 1 that there is an occasion during the simulations that the average PSNR of *Foreman* drops roughly by 2 dB between 85 kbps and 90 kbps. This means that the algorithm chooses to drop packets from this video stream and at the same time increase the quality of the others. This is shown clearly in Figure 4 where all PDM-based streams are plotted together. This figure illustrates the tradeoff between the distortion increase due to packet dropping imposed by the “even-loss distribution” algorithm and the prioritization mechanism of 802.11e that tends to minimize further quality degradation and in some cases even increases the video quality. The *Salesman* sequence has on average a constant PSNR of 37.5 dB over all available rates. On the other hand, the *Foreman* and *News* sequences are competing with each other. When the PSNR of *Foreman* is high (38 dB at 88 kbps) the corresponding PSNR of the *News* is minimum (34 dB at 88 kbps). This is due to the fact that these two sequences

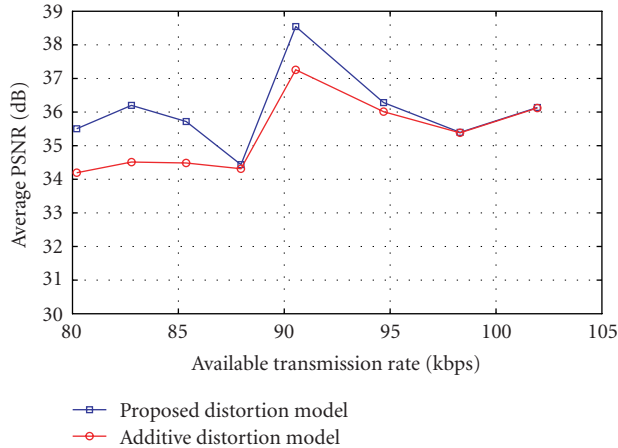


FIGURE 5: Average PSNR (in dB) as a function of available transmission rate (in kbps) for the “greedy-loss distribution” algorithm. Comparison between the proposed and the simplified “additive” distortion model for the video sequence *Foreman*.

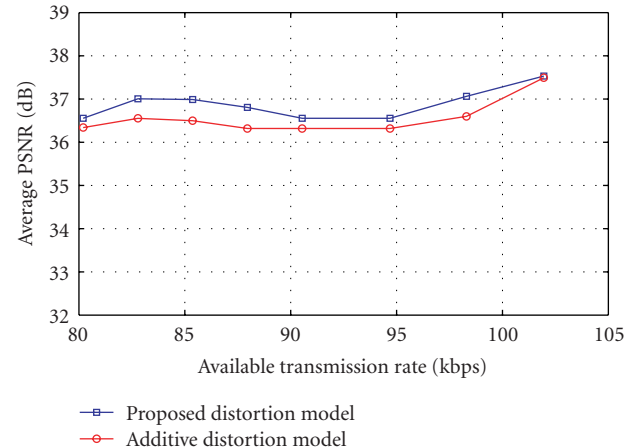


FIGURE 7: Average PSNR (in dB) as a function of available transmission rate (in kbps) for the “greedy-loss distribution” algorithm. Comparison between the proposed and the simplified “additive” distortion model for the video sequence *Salesman*.

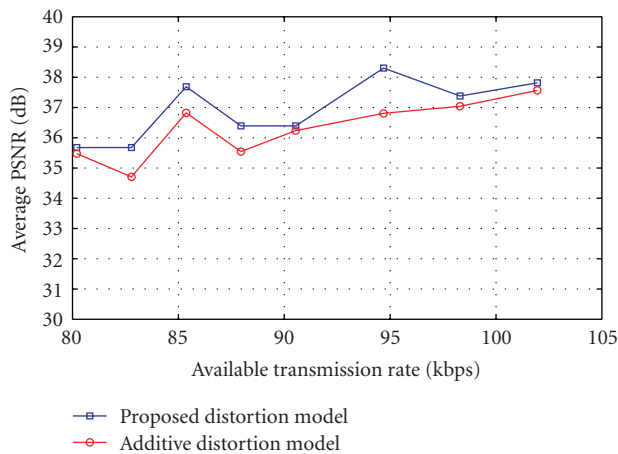


FIGURE 6: Average PSNR (in dB) as a function of available transmission rate (in kbps) for the “greedy-loss distribution” algorithm. Comparison between the proposed and the simplified “additive” distortion model for the video sequence *News*.

have scenes with motion or changing backgrounds; hence the correlation among the reference frames varies accordingly. Therefore, there may be cases where a dropped packet may cause a dramatic effect on the distortion of these two sequences and cases that a dropped packet may not affect the overall distortion significantly. Finally, as it was expected the combination of “even-loss distribution” algorithm and priority categorization provides on the average better quality for *News* and *Salesman* and worse quality for *Foreman* due to its randomly distributed correlations among its frames.

5.2. Greedy-loss distribution scenario

In this scenario, we evaluate the performance of the PDM and the “additive” model when they are applied to the “greedy-loss distribution” scheduling algorithm and the 802.11e channel access differentiation scheme. The difference

with the previous scenario is that in this case the “greedy-loss distribution” algorithm allocates the number of dropped packets unevenly to the three video streams. Since a video sequence, demanding in terms of quality context, is greater affected even by small losses, the logarithm will tend to allocate more transmission bandwidth to this video stream. However, this uneven loss distribution will be balanced by the prioritization scheme that will improve the quality of the most distorted video stream. The simulation results are shown in Figures 5, 6, and 7. In all cases, the PDM outperforms the “additive” model for up to 1.5 dB for *Foreman* at 82 Kbps, 1.5 dB for *News* at 95 Kbps, and 0.5 dB for *Salesman* at 85 Kbps. In addition to that, by comparing Figures 1 and 5, the *Foreman* PSNR for the “even-loss distribution” and the “greedy-loss distribution” algorithms, it is evident that the latter improves the total distortion of the video. However, the same applies less to *News* and does not apply to *Salesman*. These two sequences have been positively affected by the “even-loss distribution” algorithm compared to *Foreman*, but the “greedy-loss distribution” scheduling has improved the distortion effect of *Foreman* sequence as it is the most demanding in terms of context of the three. Figure 8 shows the PDM-based “greedy-loss distribution” algorithm and prioritization plots for all three video streams. The improvement on *Foreman* PSNR compared to Figure 4 is evident.

6. CONCLUSIONS

This paper proposes a novel framework based on two scheduling algorithms named “even-loss distribution” and “greedy-loss distribution” for effective packet scheduling and optimized differentiated channel access that minimizes the overall distortion of the video streams. The framework is based on a new distortion prediction model which takes into consideration the high correlation among reference frames of H.264/AVC. The proposed distortion model has been evaluated against real measurements and is compared with the

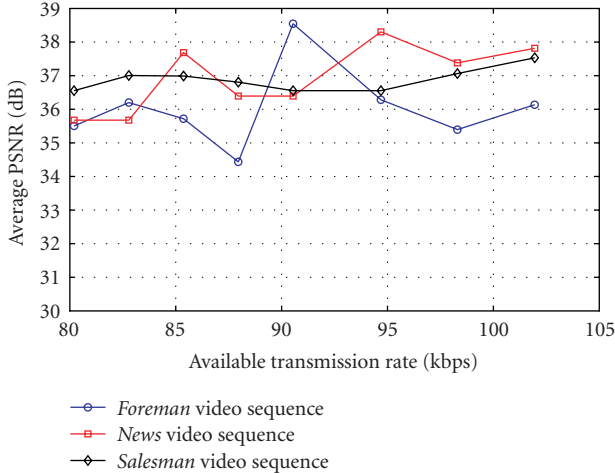


FIGURE 8: Average PSNR (in dB) as a function of available transmission rate (in kbps) for the “greedy-loss distribution” algorithm. Comparison between the proposed distortion model for all testing video sequences.

simplified “additive”-based distortion model that omits the correlation among the frames. The “even-loss distribution” algorithm aims to allocate the packet losses evenly to all video streams. The scheduling algorithm indicated the number of packets that should be dropped by each video source in order to adapt its rate to the available transmission bandwidth. However, which packets will be selected depends on the pre-calculated packet distortion based on the PDM. The imposed distortion due to packet dropping is balanced by a modified 802.11e system that supports video stream categorization to different priority classes based on the average distortion. The simulations show that this scheme can improve the video distortion of the reconstructed video at the receiver end. However, it also shows that demanding video sequences in terms of context are affected more significantly by this approach. The “greedy-loss distribution” algorithm on the other hand, allocates more transmission rate to video streams with complex context and video scenes. Hence, the imposed distortion due to packet dropping is kept minimum, although the prioritization mechanism will treat these video streams as low priority traffic. Simulations show that on average the latter scenario is more effective for all streams, especially effective for demanding sequences with randomly distributed dependencies among the frames of the sequence.

APPENDIX

A. EVALUATION OF DISTORTION MODEL

The proposed distortion model (8) is verified through extensive simulations of a standard video test sequence and is compared with actual measurement and the additive distortion model previously described. Coding and decoding is performed by using JM12 of the H.264/AVC video compression standard [24]. The selected test video sequence is *Foreman* at QCIF resolution with 300 video frames at 30 fps and a constant quantization step that results in an average PSNR of

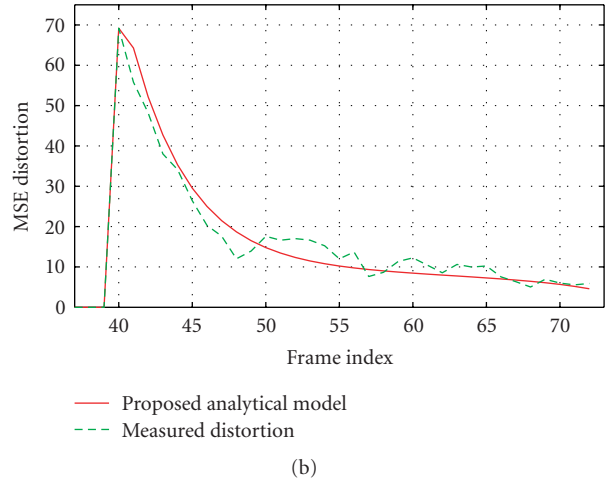
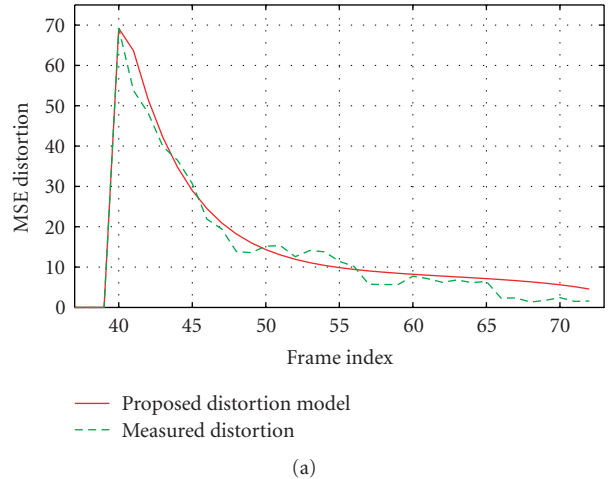


FIGURE 9: Single frame loss and error propagation. (a) Reference frame list size $M_{REF} = 5$. (b) Reference frame list size $M_{REF} = 10$.

36 dB. The size of the reference frames list varies from 5 to 10 frames. According to H.264 standard, the selected reference frames are frames that have been encoded and stored prior to the current frame. Every 4 frames, a slice is intra-updated, corresponding to an intraframe update period of $N = 4 \times 9$.

Figure 9 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. The comparison between the proposed model and the measured distortion clearly 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.

Figure 10(a) illustrates the MSE plotted 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 (1), which is also consistent with [13]. 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. Moreover, Figure 10(b) plots the total distortion due to two frame errors separated by a lag, where the first erroneous frame is Frame 38. It can be shown that the proposed model accounts for the resulted distortion very ac-

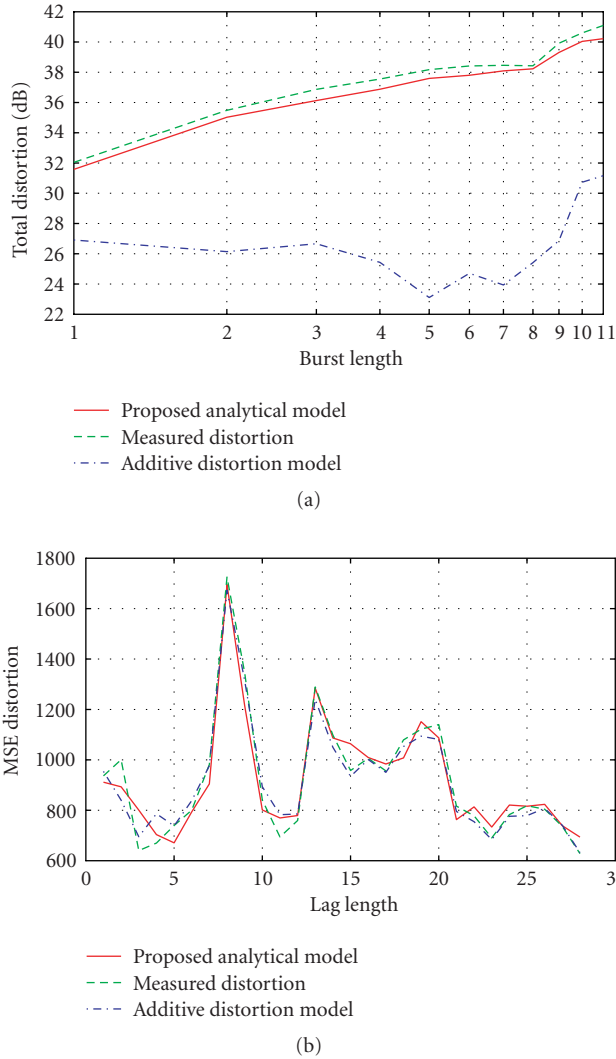


FIGURE 10: (a) Total distortion versus burst error length. (b) Total distortion of two losses with lag (first error frame at 38 and the second 38 + lag.)

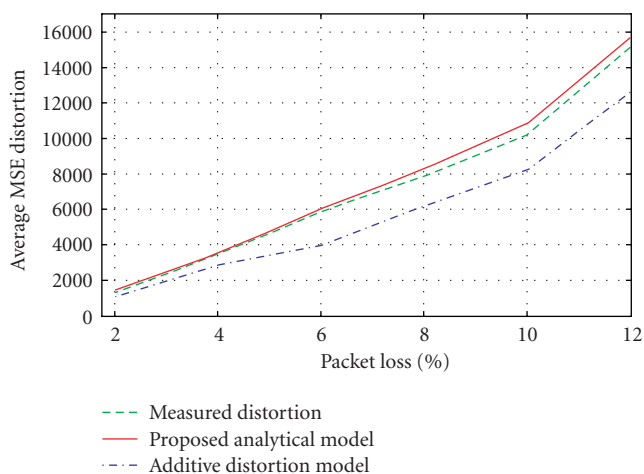


FIGURE 11: Average total distortion versus packet loss.

curately, as it considers the correlation between each pair of error frames as opposed to the additive model.

Finally, in Figure 11 the averaged distortion is plotted for different packet losses. Each packet loss is the result of different error frame patterns. In order for the model to be precise, up to 10 000 random combinations of the 100 video frames have been used for the calculation of the actual video distortion and the proposed recursive approach of (6). It is clear that the model follows the actual distortion very closely even at high packet losses. In particular, at packet loss beyond 9% the distortion model deviates from the actual measurements approximately by 1000 in MSE distortion.

ACKNOWLEDGMENT

This work was supported by the project PENED no. 03636, which is funded in 75% by the European Social Fund and in 25% by the Greek State-General Secretariat for Research and Technology.

REFERENCES

- [1] J. Chakareski and P. Frossard, "Rate-distortion optimized distributed packet scheduling of multiple video streams over shared communication resources," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 207–218, 2006.
- [2] IEEE 802.11g, "Part11: wireless LAN medium access control (MAC) and physical layer (PHY) specification band," Supp. IEEE 802.11, 2003.
- [3] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535–547, 2000.
- [4] J. Deng and R.-S. Chang, "A priority scheme for IEEE 802.11 DCF access method," *IEICE Transactions on Communications*, vol. E82-B, no. 1, pp. 96–102, 1999.
- [5] I. Aad and C. Castelluccia, "Differentiation mechanisms for IEEE 802.11," in *Proceedings of the 20th Annual Joint Conference on the IEEE Computer and Communications Societies (INFOCOM '01)*, vol. 1, pp. 209–218, Anchorage, Alaska, USA, April 2001.
- [6] J.-Y. Yeh and C. Chen, "Support of multimedia services with the IEEE 802.11 MAC protocol," in *IEEE International Conference on Communications (ICC '02)*, vol. 1, pp. 600–604, New York, NY, USA, April-May 2002.
- [7] IEEE 802.11e, "Wireless LAN medium access control (MAC) enhancements for quality of service (QoS)," *802.11e draft 8.0*.
- [8] 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*, vol. 4, no. 5, pp. 30–44, 2006.
- [9] M. Ghanbari, *Standard Codecs: Image Compression to Advanced Video Coding*, IEE Press, London, UK, 2003.
- [10] K. Stuhlmüller, N. Färber, M. Link, and B. Girod, "Analysis of video transmission over lossy channels," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1012–1032, 2000.
- [11] I.-M. Kim and H.-M. Kim, "A new resource allocation scheme based on a PSNR criterion for wireless video transmission to stationary receivers over Gaussian channels," *IEEE Transactions on Wireless Communications*, vol. 1, no. 3, pp. 393–401, 2002.

- [12] J. Apostolopoulos, W.-T. Tan, S. Wee, and G. Wornell, "Modeling path diversity for multiple description video communication," in *Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP '02)*, vol. 3, pp. 2161–2164, Orlando, Fla, USA, May 2002.
- [13] Y. Liang, J. Apostolopoulos, and B. Girod, "Analysis of packet loss for compressed video: does burst-length matter?" in *Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP '03)*, vol. 5, pp. 684–687, Hong Kong, April 2003.
- [14] F. Fitzek and M. Reisslein, "MPEG-4 and H.263 video traces for network performance evaluation," *IEEE Network*, vol. 15, no. 6, pp. 40–54, 2001.
- [15] J. Xin, C.-W. Lin, and M.-T. Sun, "Digital video transcoding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 84–97, 2005.
- [16] S. R. McCanne, M. Vetterli, and V. Jacobson, "Low-complexity video coding for receiver-driven layered multicast," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 6, pp. 983–1001, 1997.
- [17] P. Buccioli, G. Davini, E. Masala, E. Filippi, and J. C. De Martin, "Cross-layer perceptual ARQ for H.264 video streaming over 802.11 wireless networks," in *IEEE Global Telecommunications Conference (GLOBECOM '04)*, vol. 5, pp. 3027–3031, Dallas, Tex, USA, November-December 2004.
- [18] 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, 2004.
- [19] W. Tu, W. Kellerer, and E. Steinbach, "Rate-distortion optimized video frame dropping on active network nodes," in *Proceedings of the International Packet Video Workshop*, Irvine, Calif, USA, Decemeber 2004.
- [20] T. Wiegand, G. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.
- [21] H. L. Truong and G. Vannuccini, "The IEEE 802.11e MAC for quality of service in wireless LANs," in *Proceedings of International Conference on Advances in Infrastructure for Electronic Business, Education, Science, Medicine, and Mobile Technologies on the Internet (SSGRR '03)*, L'Aquila, Italy, January 2003.
- [22] ISO/IEC 14496-10, ITU-T Rec., and H.264 standard, "Advanced video coding," 2003.
- [23] "Network simulator v.2 (ns-2) DCF and EDCA extensions," http://www.tkn.tu-berlin.de/research/802.11e_ns2/.
- [24] "H.264 reference software version," Image Processing Department, Fraunhofer-Institute, <http://ip.hhi.de/>.
- [25] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "Rtp: a transport protocol real-time applications," *RFC 1889*, January 1996.