

RESEARCH

Open Access

An occupancy-based and channel-aware multi-level adaptive scheme for video communications over wireless channels

Husameldin Mukhtar¹, Mohamed Hassan^{2*} and Taha Landolsi²

Abstract

Video streaming over wireless channels is challenged with the time-varying nature of the underlying channels and the stringent requirements of video applications. In particular, video streaming has strict requirements on bandwidth, delay, and loss rate while wireless channels are dynamic and error-prone by nature. In this article, we propose a novel *multilevel* adaptive scheme that is designed to mitigate the challenges facing video streaming over unreliable channels. This is done while preventing potential playback discontinuities and guaranteeing a graceful degradation of the rendered video quality. Scalable video coding, adaptive modulation, and adaptive channel coding are integrated to achieve the objectives of the proposed scheme. If adaptive modulation and channel coding are not enough to guarantee the on-time delivery of decodable video frames, we adopt scalable coding. Simulation results show that the proposed adaptive scheme achieves an improvement of about 2.5 dB in the peak signal-to-noise ratio over a nonadaptive one. In addition, the proposed scheme reduces the number of starvation instances by 50 and 90% in the cases of Stop-and-Wait and Go-Back-N automatic repeat requests, respectively.

Keywords: adaptive modulation, channel coding, error control, source rate control, wire-less channels

1 Introduction

Delivery of multimedia contents over wireless channels is becoming increasingly popular. Recent advances in wireless access networks provide a promising solution for the delivery of multimedia services to end-user premises. In contrast to wired networks, wireless networks not only offer a larger geographical coverage at lower deployment cost, but also support mobility. Nevertheless, wireless channels are dynamic and error-prone by nature while video streaming has strict requirements on bandwidth, end-to-end delay and delay jitter especially for live and interactive video. To make matters worse, compressed video bitstreams are extremely sensitive to losses. This is due to the fact that standard video compression techniques exhibit certain inter-dependencies, whereby correct decoding of a given video frame requires the correct decoding of previous and sometimes future “reference” frames. Hence, *correct and timely*

delivery of reference frames must be guaranteed with a higher probability to limit error propagation that typically results in significant degradation in the decoded video quality.

Different approaches have been proposed in the literature that constitute a solution space for the above challenges. Examples of these approaches are scalable video coding, source rate control, bitstream switching, error control, adaptive modulation, power allocation, transcoding, and adaptive playback [1-7]. The authors in [3] proposed a rate control approach for video streaming over wireless channels. The wireless channel in [3] is characterized by an arguable two-state channel model that provides a coarse approximation of the channel behavior and may not always be acceptable. The source rate and channel code parameters are adaptively computed in a cycle basis subject to a constraint on the probability of starvation at the playback buffer. In [8], the authors employed a wavelet video encoder and proposed a joint packetization and retransmission strategy to minimize the distortion in the decoded video for a

* Correspondence: mshassan@aus.edu

²College of Engineering, American University of Sharjah, Sharjah, UAE
Full list of author information is available at the end of the article

given delay constraint. Average PSNR of the decoded video was used as the performance metric in [8]. The authors in [9] introduced two channel adaptive rate control schemes for slowly and fast varying channels. Both schemes in [9] account for the occupancy of playback buffer in the joint optimization of source rate and channel coding parameters. They assumed Stop-and-Wait automatic repeat request (SW-ARQ) in their proposed video streaming system. While this is an acceptable assumption in wireless environments with small round trip time (RTT), it is typically not a plausible one for wireless networks with large RTT. In [10], the authors presented a system that employs an algorithm to dynamically select the encoding mode of macroblocks as well as the forward error correction (FEC) and the physical layer transmission rate in multirate wireless local area networks (LANs). The algorithm aimed at minimizing the decoded video distortion but ignored the dynamics of the playback buffer to maintain continuous video playback. Moreover, link-layer retransmissions were not considered in [10]. The authors in [11] proposed a rate-distortion optimized packet scheduling and content-aware playout mechanism to maximize the perceived video quality in terms of both picture and playout quality. Non-scalable pre-stored video was assumed in [11]. In [12], the authors proposed a rate control algorithm for streaming on-demand scalable variable bit rate (VBR) video over wireless networks. They used temporal scalability with one base layer (BL) and one enhancement layer (EL) in their simulations and assumed that video packet losses may only occur on missing the playback deadline. A weighted sum of lost BL and EL packets divided by the weighted sum of total BL and EL packets was defined as the performance metric in [12]. The authors in [13] integrated the TCP-friendly rate control (TFRC) algorithm with H.264/AVC source coding and adaptive modulation and channel coding (AMC) for real-time video streaming over wireless multi-hop networks. The performance evaluation in [13] was done in terms of decoded video average PSNR.

While several schemes for video streaming over wireless channels have been introduced in the literature [14-20], the bulk of these schemes aim at the optimization of the performance of the source and/or channel encoders, with little to no considerations of the networking aspects. Many of these studies are concerned with the optimization of the effective throughput of the channel, without considering the impact of source and channel coding on the transport delay and delay jitter. The delay performance of hybrid ARQ schemes has been studied in [21,22] independently of the video content (i.e., without regard to source coding). Most studies on joint source/channel coding address the problem from an information theoretic point of view, and did

not account for network performance and protocol issues, including packetization and retransmissions. In addition, most of the existing work overlooked the impact of playback buffer starvation and overflow at the decoder, both of which are critical to guaranteeing continuous video playback.

In general, we believe that the literature on video streaming is still in a need for comprehensive solutions of the topic, whereby modulation, channel coding, source rate control, ARQ retransmissions, prioritization of video information (and related unequal error protection), power allocation, and error concealment are all performed jointly and adaptively with the objective of maximizing the likelihood of uninterrupted video playback subject to varying channel conditions and frame sizes.

In this study, we propose a multi-level adaptive approach whereby we integrate scalable video coding, adaptive channel coding, and adaptive modulation to achieve efficient video streaming.^a The objective of our multi-level adaptive scheme is to ensure uninterrupted playback with acceptable video quality at the client side. Adaptive modulation is exploited to overcome the performance enhancement limitation in source rate control schemes employing fixed modulation. By integrating scalable video coding with adaptive modulation and channel coding, we significantly increase the probability of successful delivery of video frames within a time constraint that depends on the instantaneous occupancy of the playback buffer. This, in return, reduces the amount of required video scaling, hence, improving the temporal and spatial quality of the reconstructed video. In our analysis and simulations, in addition to SW-ARQ, we consider more practical ARQ schemes such as Go-back-N (GBN) and selective repeat (SR). We also consider two statistical channel models, namely, additive white Gaussian noise (AWGN) and Rayleigh channel models. Moreover, our proposed adaptive scheme takes into account the sensitivity of video frames when implementing source rate control to achieve enhanced video quality.

In the evaluation of the proposed multi-level adaptive scheme, we consider the PSNR as a spatial video quality metric. In addition, we use newly introduced temporal video quality metrics, namely, the skip length (SL) and inter-starvation distance (ISD) [23] which reflect the dynamics of the playback buffer. On the occurrence of any starvation instant, SL indicates how long (in frames) this starvation will last. The rationale behind SL as a metric for temporal quality is the fact that it is better for the human eye to watch a continuously played back video at a lower quality rather than watching a higher quality video sequence that is frequently interrupted. On the other hand, ISD is the distance in frames that separates successive starvation instants. This metric

complements the SL in the sense that if the latter is small but very frequent, then the quality of the played back video would be degraded. Therefore, large ISDs in conjunction with small SLs would result in an interrupted and better quality played back video. Figure 1 illustrates the definitions of these two metrics.

The rest of this article is organized as follows. Section 2 describes our video streaming system and presents the proposed adaptive scheme. Performance evaluation of our scheme is given in Section 3. Finally, conclusions and summary of results are provided in Section 4.

2 Proposed adaptive scheme

Figure 2 describes the proposed video streaming system. In this model, we assume that the receiver continuously monitors the channel state, the playback buffer occupancy, and the quality of the played back video as well as the history of sizes of transmitted video frames. The receiver then feeds back this information to the transmitter/video encoder. Based on this information, the transmitter controls the encoding bitrate of the scalable compressed video and adapts the modulation level and channel coding rate to reduce the likelihood of playback buffer starvation. The video bitstream is transmitted over an unreliable forward channel, whereas we assume that the feedback information is transmitted over a reliable reverse channel. On the transmission of a video frame, the frame candidate for transmission is first segmented into one or more link-layer packets each of which undergoes cyclic redundancy check (CRC) followed by FEC coding. When the FEC decoder at the receiver fails to fully correct transmission errors in any of the packets, we assume that the CRC code will detect these errors and a retransmission request will be triggered. To do so, the deployed hybrid ARQ assumes that the CRC code is first applied to the packet followed by the FEC code. As mentioned earlier, in what follows we consider different ARQ schemes. This includes Stop-and-Wait, Selective Repeat, and Go-back-N.

The wireless channel is represented by a finite-state Markov chain, the states of which are characterized by their bit error rate (BER) denoted by p_i , $i \in \{0, 1, \dots, N\}$. The BER is a function of the ratio of the energy per symbol (E_s) to the noise power spectral density (N_0). Therefore, for a fixed modulation level scheme we have $p_0 > p_1 \dots > p_N$, i.e., state N is the “best” state, and state 0 is the “worst”.

In M-ary modulation schemes, increasing the order of modulation level (i.e., increasing the number of bits per symbol) will increase the error-free channel bitrate by $\log_2 M$ at the expense of the BER performance. For square M-QAM, the analytical expression of the BER, in AWGN channels, is given by [24]

$$p_i^{\text{awgn}} = \frac{2}{\sqrt{M} \log_2 \sqrt{M}} \sum_{k=1}^{\log_2 \sqrt{M}} \sum_{j=0}^{(1-2^{-k})\sqrt{M}-1} \left\{ (-1)^{\lfloor \frac{j2^{k-1}}{\sqrt{M}} \rfloor} \left(2^{k-1} - \left\lfloor \frac{j2^{k-1}}{\sqrt{M}} + \frac{1}{2} \right\rfloor \right) Q \left((2j+1) \sqrt{\frac{6 \log_2 M}{2(M-1)} \frac{E_b}{N_0}} \right) \right\}, \quad (1)$$

where $Q(\cdot)$ is the Q function and $E_b/N_0 = E_s/(N_0 \log_2 M)$ is the per-bit signal-to-noise ratio (SNR). On the other hand, for the BER over Rayleigh fading channels, the expression is given by [24,25]

$$p_i^{\text{Rayleigh}} = \frac{2}{\pi \sqrt{M} \log_2 \sqrt{M}} \sum_{k=1}^{\log_2 \sqrt{M}} \sum_{j=0}^{(1-2^{-k})\sqrt{M}-1} \left\{ (-1)^{\lfloor \frac{j2^{k-1}}{\sqrt{M}} \rfloor} \left(2^{k-1} - \left\lfloor \frac{j2^{k-1}}{\sqrt{M}} + \frac{1}{2} \right\rfloor \right) \int_0^{\pi/2} \prod_{l=1}^L G_{\gamma_l} \left(-\frac{(2j+1)^2 3 / (2(M-1))}{\sin^2 \theta} d\theta \right) \right\}, \quad (2)$$

where L is the number of diversity branches and G_{γ_l} is the moment generating function for each diversity

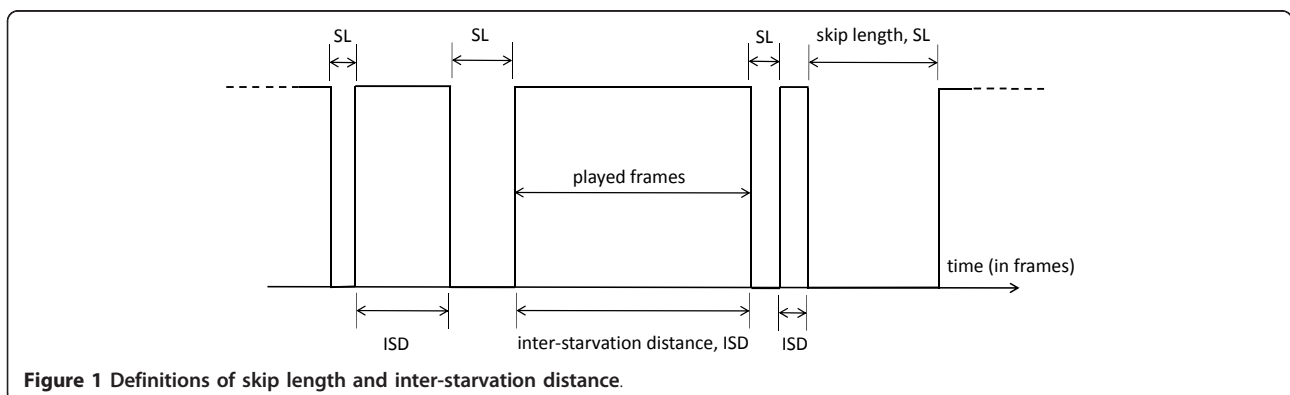


Figure 1 Definitions of skip length and inter-starvation distance.

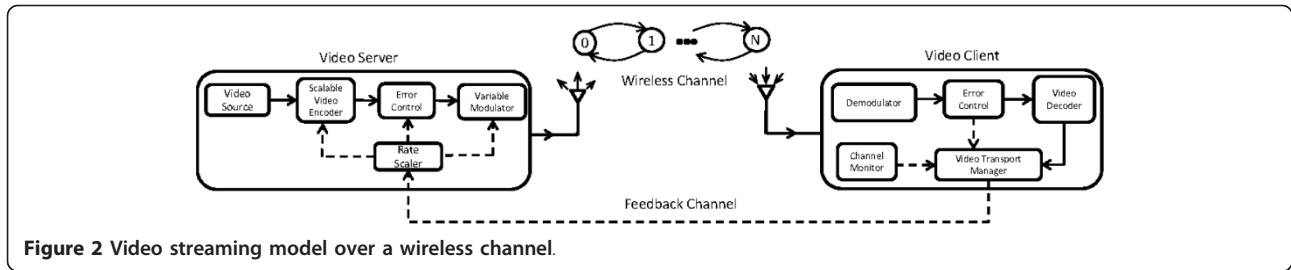


Figure 2 Video streaming model over a wireless channel.

branch defined by $G_{\gamma_l}(s) = 1/(1 - s\bar{\gamma}_l)$. Moreover $\bar{\gamma}_l = (\Omega_l \cdot \log_2 M \cdot E_b/N_0)/L$, where $\Omega_l = E[A_l^2]$ is the power of the fading amplitude A_l . In this study, we assume one diversity branch, i.e., $L = 1$.

2.1 Transmission efficiency (bits/s/Hz)

In this section, we demonstrate the impact of the joint adaptation of the modulation level and channel coding on the achieved spectral efficiency which in turn yields an improved data rate. Let \bar{N}_{r_i} denote the average number of retransmissions needed to successfully transmit a packet in the presence of errors. For SR-ARQ, the number of retransmissions (including the first transmission attempt) is a geometric random variable with mean $\bar{N}_{r_i} = 1/P_{c_i}$ [26] where P_{c_i} is the probability of correctly receiving a packet which is given by

$$P_{c_i} = \sum_{j=0}^{\tau_{\max_i}} \binom{S_p}{j} p_i^j (1 - p_i)^{S_p - j}, \quad (3)$$

where τ_{\max_i} is the number of correctable bits and S_p is the packet size including the FEC bits.

Let C be the error-free channel bitrate for binary phase shift keying and let C_i be the effective channel bit rate when the channel is in state i . When channel coding is implemented an overhead is incurred to the transmitted packets. Therefore, C_i is approximated by

$$C_i = P_{c_i} \frac{k_i}{S_p} C \log_2 M, \quad (4)$$

where $k_i = S_p - h_i$ is the payload size and h_i is the FEC overhead. Let $\varepsilon_i = P_{c_i} k_i / S_p$. Equation 4 is now given by

$$C_i = \varepsilon_i C \log_2 M. \quad (5)$$

Clearly, $0 \leq \varepsilon_i \leq 1$ and reflects the channel condition. For fixed FEC, τ_{\max_i} is usually predefined and has a fixed value. On the other hand, in adaptive FEC, an "optimal" desired value $\tau_{\max_i}^*$ could be determined based on the channel condition and the packet size. In [9], a reasonable approximation for $\tau_{\max_i}^*$ is given by

$$\tau_{\max_i}^* \approx \left\lceil p_i S_p + 3\sqrt{p_i S_p (1 - p_i)} \right\rceil, \quad (6)$$

where $\lceil \cdot \rceil$ is the ceiling function. Therefore, when the channel is in state i , the transmission efficiency η_i for SR-ARQ is

$$\eta_{\text{SR}} = \frac{C_i}{C} = P_{c_i} \frac{k_i}{S_p} \log_2 M. \quad (7)$$

Similarly, based on the analysis in [26], with simple manipulation the transmission efficiency for GBN-ARQ and SW-ARQ protocols is given by

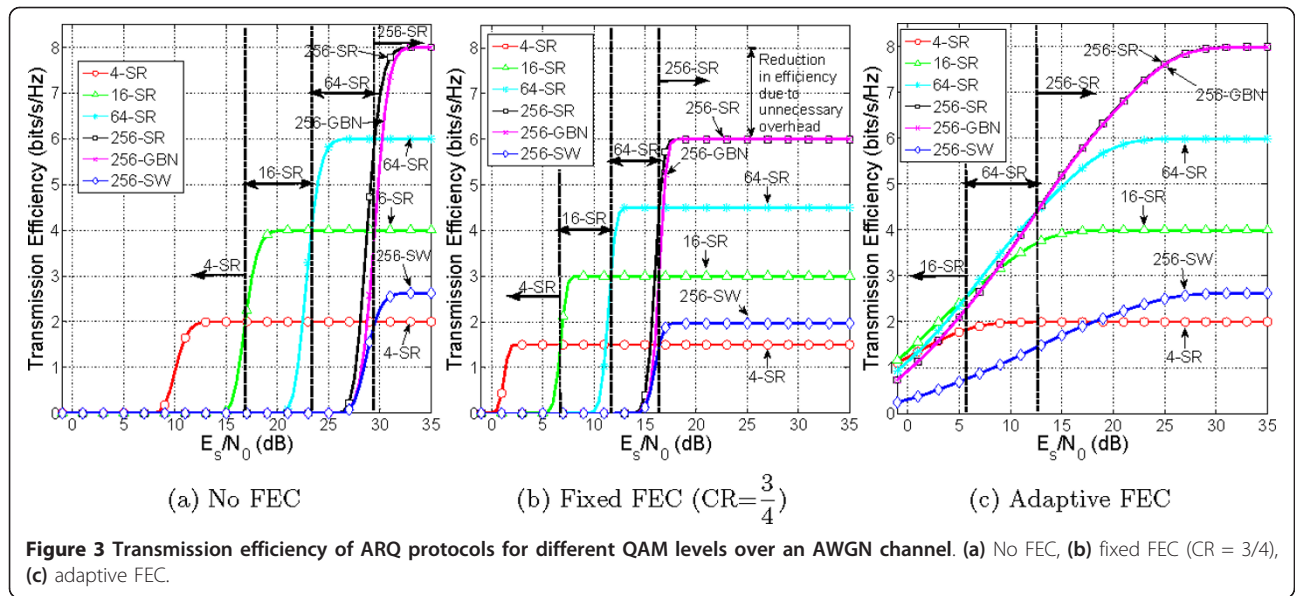
$$\eta_{\text{GBN}} = \left[\frac{P_{c_i}}{P_{c_i} + K(1 - P_{c_i})} \right] \frac{k_i}{S_p} \log_2 M, \quad (8)$$

$$\eta_{\text{SW}} = \frac{P_{c_i} k_i}{K S_p} \log_2 M, \quad (9)$$

where $K - 1$ is the number of packets that can be transmitted during the RTT ($K = \lceil (\text{RTT} \cdot C \cdot \log_2 M) / S_p \rceil + 1$). For the GBN analysis, it was assumed that the window size of the retransmission buffer is selected such that the channel is kept busy all the time. Note that when $K = 1$, Equations 8 and 9 are equal. This is an intuitive result since SW is a special case of GBN.

Figures 3 and 4 compare the transmission efficiency η_i of SR-ARQ for different QAM levels with no FEC, fixed FEC, and adaptive FEC. η_i of GBN-ARQ and SW-ARQ is also shown for 256-QAM. The plots were generated assuming Reed-Solomon FEC, $S_p = 1000$ bits, RTT = 1 ms, and $C = 256$ Kbps. For fixed FEC, a code rate $\text{CR} = k_i / S_p = 3/4$ was assumed whereas for adaptive FEC $\text{CR} = (S_p - 2\tau_{\max_i}^*) / S_p$. In Figure 3, an AWGN channel is assumed whereas in Figure 4 a Rayleigh channel is assumed.

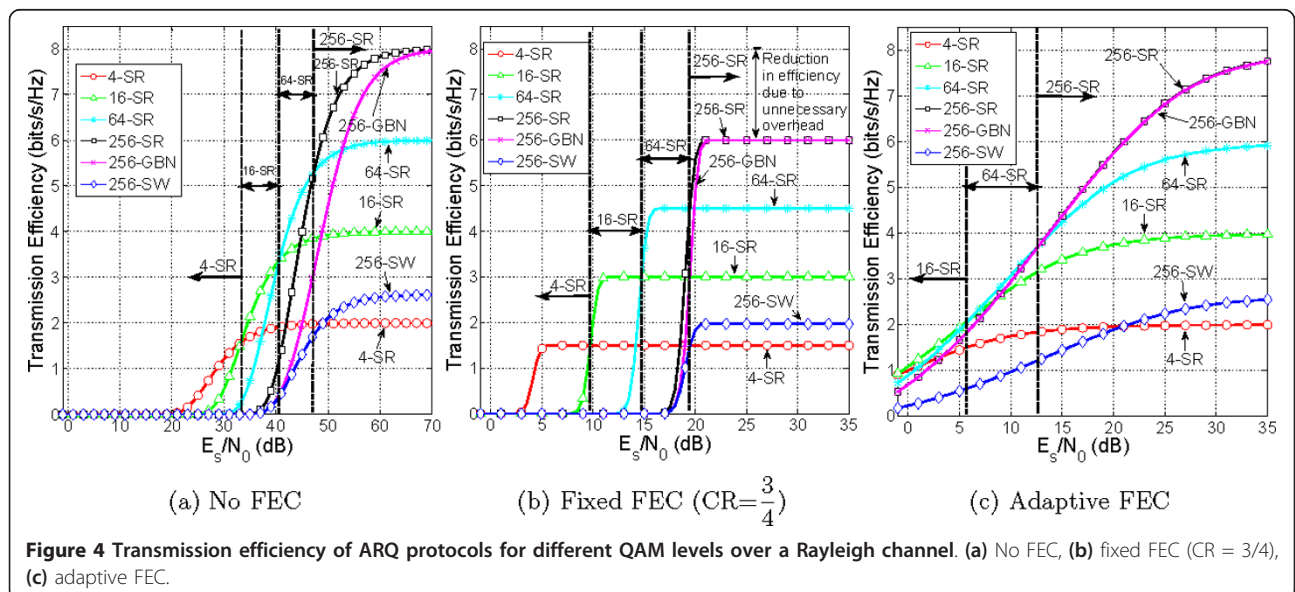
Figure 3a is intuitive and shows that when no FEC is used, 4-QAM is best for low SNR values ($E_s/N_0 < 16.9$ dB). This is a direct conclusion since the BER is minimum for 4-QAM in this E_s/N_0 range. As the SNR increases, the benefit of increasing the modulation level becomes more visible. 16-QAM provides the highest transmission efficiency for $16.9 \text{ dB} < E_s/N_0 < 23.5 \text{ dB}$. 64-QAM efficiency is the highest for $23.5 \text{ dB} < E_s/N_0 < 29 \text{ dB}$. Finally, 256-QAM achieves the highest transmission efficiency for $E_s/N_0 > 29 \text{ dB}$ when compared to the other lower modulation levels.



Moreover, Figure 3b shows that fixed FEC improves the transmission efficiency for low E_s/N_0 values. Notice that the curves are shifted to the left when compared to the case with no FEC. This shift reflects the coding gain which is the difference between the E_s/N_0 values of the uncoded system and the coded system to achieve the same BER performance when FEC is used. However, at high E_s/N_0 values, unnecessary overhead is incurred preventing the modulation scheme from achieving its highest possible transmission efficiency which is equal to $\log_2 M$. Figure 3c shows that adaptive FEC outperforms fixed FEC. With adaptive FEC, the transmission efficiency is improved for even smaller E_s/N_0 values. At the

same time, no unnecessary overhead is added during channel good states (i.e., high E_s/N_0 values) allowing for the realization of the maximum error-free bitrate. Based on these plots a decision can be made to use adaptive FEC with 16-QAM for $E_s/N_0 < 5.5$ dB, 64-QAM for 5.5 dB $< E_s/N_0 < 12.5$ dB, and 256-QAM for $E_s/N_0 > 12.5$ dB to achieve the best bandwidth utilization (when a packet size of 1000 bits is used). It is worth noting that similar computations could be carried out for different packet sizes from which a look up table can be generated to speed up the search process.

Figure 4 shows a significant degradation in the transmission efficiency when the more realistic Rayleigh



channel model is assumed, especially when no FEC or fixed FEC is used. Notice that, for 256-QAM with no FEC, a very high $E_s/N_0 \approx 65$ dB is required to achieve the highest transmission efficiency.

In addition, as shown in Equations 7-9, SR-ARQ performance is not affected by the RTT. However, the performance of SW-ARQ and GBN-ARQ degrades when $RTT \cdot C \cdot \log_2 M$ is relatively large (relative to S_p). For large RTT values, the transmission efficiency of the SW-ARQ becomes unacceptable, whereas the bandwidth efficiency of GBN-ARQ drops rapidly as the channel SNR decreases when fixed FEC (or no FEC) is used.

When adaptive FEC is used, the difference in the performance between SR-ARQ and GBN-ARQ is significantly reduced even for relatively large RTT values. That is because, in adaptive FEC, $P_{c_i} \approx 1$ which makes $\eta_{isr} \approx \eta_{iGBN}$ (see Equations 7 and 8). In other words, when $P_{c_i} \approx 1$, each packet is transmitted once on average making GBN-ARQ less detrimental when compared to a case with higher average number of retransmissions.

2.2 Probability of successful video frame delivery within a time constraint

The proposed multi-level scheme adaptively integrates source rate control, selection of the modulation level, and channel coding to reduce the likelihood of playback buffer starvation while guaranteeing a gracefully degraded quality of the reconstructed video. More specifically, while proper selection of the modulation level (based on the fed back channel SNR) increases the achievable data rate, proper channel coding increases the probability of fast and correct delivery of video frames. This in turn builds up the decoder playback buffer and hence increases the budget time for the transmission of following video frames. This typically results in less scaling (graceful rate control) which leads to better perceptual quality. As will be seen later, the proposed scheme sets a bound on the probability of correct frame transmission within a budget time that is computed using the occupancy of the playback buffer. If this bound on the probability is not met, the multi-level adaptive scheme resorts to scaling the video frames (source rate control). In what follows we show the details of obtaining an expression for the probability of correctly receiving a video frame within a time constraint. Recall that a video frame may consist of multiple packets each of which may require several retransmissions. In what follows we assume a slowly varying channel where the channel state does not change during a frame transmission time.

Let $T_p^{(i)}$ be the time needed to transmit a packet until it is correctly received. $T_p^{(i)}$ is a function of a geometric

random variable which is the number of retransmissions. This time can be approximated by an exponential distribution of mean $\lambda_i^{-1} = E(T_p^{(i)}) = k_i / \eta_i C$. The mean λ_i^{-1} for SR-ARQ, GBN-ARQ, and SW-ARQ is given by [26,27]

$$\lambda_i^{-1} = \begin{cases} \frac{S_p}{C \log_2 M P_{c_i}} & \text{for SR - ARQ,} \\ \frac{S_p}{C \log_2 M} + \left(\frac{S_p}{C \log_2 M} + RTT \right) \frac{1 - P_{c_i}}{P_{c_i}} & \text{for GBN - ARQ,} \\ \left(\frac{S_p}{C \log_2 M} + RTT \right) \frac{1}{P_{c_i}} & \text{for SW - ARQ.} \end{cases} \quad (10)$$

For a given video frame size S_f and a packet size S_p , the required number of packets N_p to contain the video frame is

$$N_p = \left\lceil \frac{S_f}{S_p - h_i} \right\rceil. \quad (11)$$

Hence, the total time $T_f^{(i)}$ needed to successfully deliver a video frame is gamma distributed with parameters λ_i and N_p . Accordingly, the probability of correctly receiving a frame within a time constraint is given by [9]

$$F(T_b, i) = P(T_f^{(i)} \leq T_b) = 1 - e^{-\lambda_i T_b} \sum_{n=0}^{N_p-1} \frac{(\lambda_i T_b)^n}{n!}, \quad (12)$$

where T_b is the budget time defined as follows:

$$T_b = \begin{cases} \frac{0.5}{f_n} & \text{if } B \leq B_{th}, \\ \frac{B - B_{th}}{f_n} & \text{if } B > B_{th}, \end{cases} \quad (13)$$

where f_n is the nominal playback rate, B is the playback buffer occupancy, and B_{th} is a specified buffer occupancy threshold. T_b reflects the urgency of frame arrivals at the playback buffer. For example, when the playback buffer is in an underflow state (i.e., $B \leq B_{th}$), T_b is set to a small value compared to values of T_b when $B > B_{th}$. The smaller the budget time, the more urgently frames should arrive to avoid starvation. B_{th} can be specified differently based on the type (f_{type}) or importance of a video frame. For example, for less important frames such as B frames, B_{th} can be set to a larger value when compared to the value of B_{th} for an I or P frame. This way frame size scaling will be mostly applied to the less important B frames. In addition, more budget time will be allocated for the more important frames and hence reducing the degradation in the video quality due to frame truncation.

In the proposed scheme, the transmitter determines T_b based on the buffer occupancy feedback information. Every time a frame is to be transmitted, the transmitter computes $F(T_b, i)$ for the different modulation levels

and selects the level that achieves the highest $F(T_b, i)$. Nevertheless, if none of the modulation levels can achieve $F(T_b, i) \geq \delta$ where δ is a predefined probability bound, the transmitter reduces the size of the video frame by a scaling increment α such that $S_f^{(new)} = \alpha S_f$. The video frame size is reduced by discarding ELs. Then, the transmitter recomputes $F(T_b, i)$ and repeats the process, if necessary, until $F(T_b, i) \geq \delta$. When compared to other rate control techniques which requires adjustment of encoding parameters, scalable coding is less complex and allows real time adjustment of the video frame size. Our multi-level adaptive video streaming algorithm is outlined in Table 1.

2.2.1 Numerical investigations

We now study the effect of channel coding (τ_{max}), channel condition (E_s/N_0), and frame size on $F(T_b, i)$ for different modulation levels with different ARQ schemes. The modulation levels are 4-QAM, 16-QAM, 64-QAM, and 256-QAM. A Rayleigh fading channel is assumed in the following numerical investigations. Moreover, the following parameters were assumed. $S_f = 9383$ byte which is the average video frame size of the Harry Potter HD sequence when encoded with quantization parameters 28, 28, and 30 for I, P, and B frames, respectively, [28]. $S_p = 2272$ byte which is the maximum transmission unit in IEEE 802.11. $T_b = 167$ ms = $5/30$ ms which corresponds to having five frames available in the playback buffer with a playback rate of 30 fps.

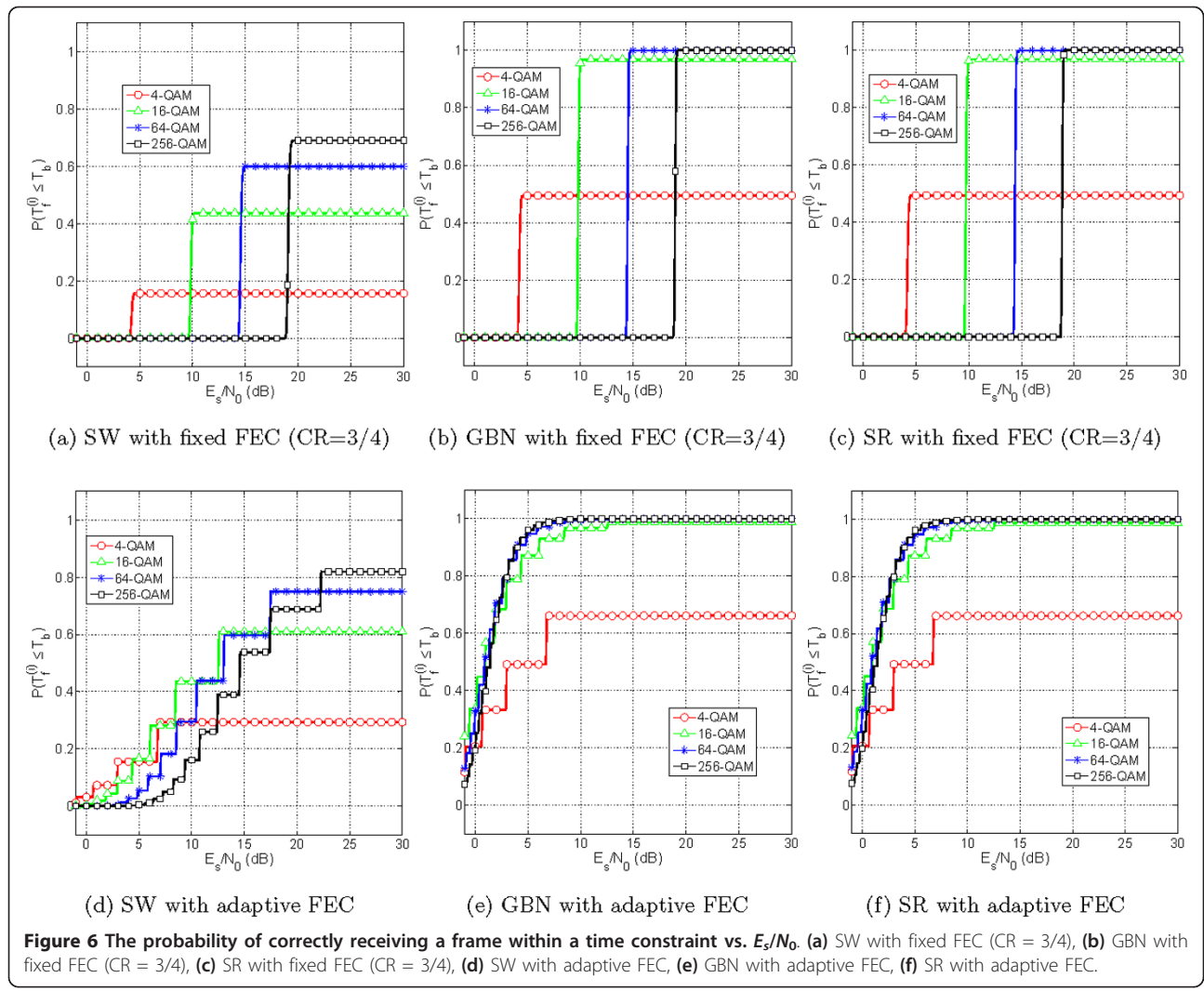
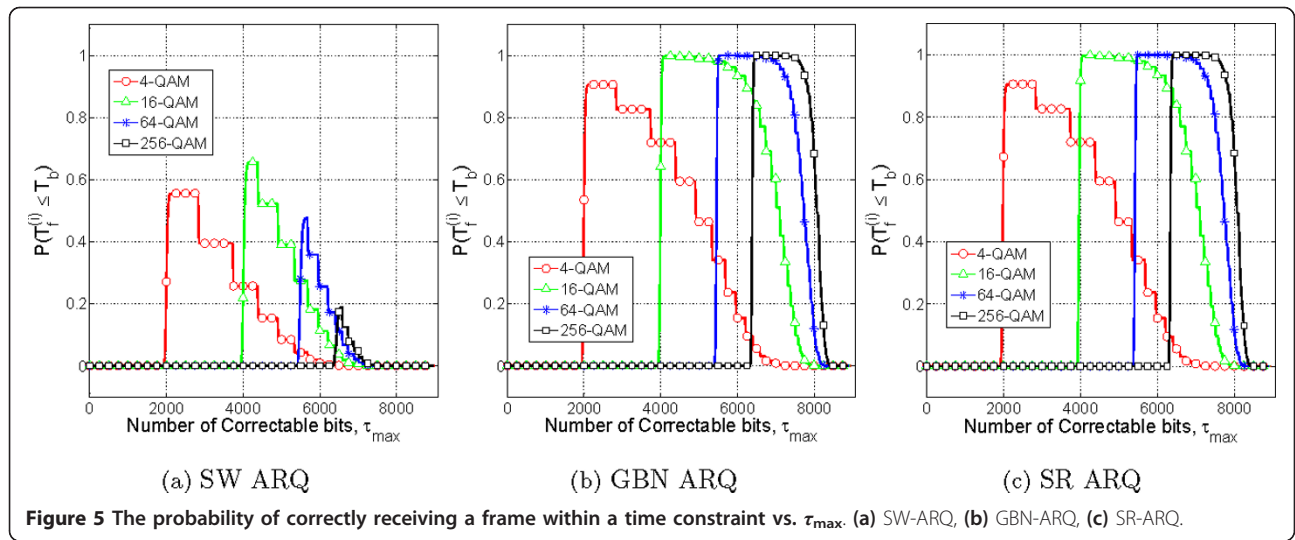
Finally, RTT = 10 ms and $C = 512$ Kbps. These values are used in the rest of our numerical investigations unless stated otherwise.

Figure 5 shows the effect of changing the amount of FEC (τ_{max}) on $F(T_b, i)$ for different levels of QAM for the three considered ARQ schemes. Increasing τ_{max} improves the performance of the different QAM streaming systems by increasing $F(T_b, i)$ up to an optimum point after which the performance starts to degrade. This is due to the fact that increasing the number of FEC bits improves the probability of correctly receiving a packet, but at the same time, the number of required packets per frame increases hindering timely delivery of the video frame. As the modulation level increases the amount of required FEC increases for a low channel SNR which was assumed when generating the plots in Figure 5 ($E_s/N_0 = 5$ dB). As can also be seen from Figure 5, increasing FEC blindly can have a destructive effect on the performance of a transmission system. Moreover, for the same modulation level and the same FEC, GBN, and SR perform better than SW while the difference in performance between SR and GBN is unnoticeable. However, at $\tau_{max} = 2000$ bits, it can be noticed that SR achieves higher $F(T_b, i)$ than the GBN's (notice the line marker at $\tau_{max} = 2000$ bits). The staircase behavior in the plots is attributed to the ceiling function in Equation 11.

Figure 6 shows the impact of varying the modulation level according to the channel conditions on $F(T_b, i)$. In

Table 1 Multi-level adaptive video streaming algorithm

Input: $E_s/N_0, B, S_f, f_{type}, B_{th}$
Output: $M, S_f^{(new)}, h_i$
Initialize: count = 0, $S_f^{(new)} = S_f$
compute T_b using Equation 13
for $j = 1$ to 4 do
$M(j) = 2^{2j}$ {QAM level}
compute $p_i(j)$ using Equation 2
compute $\tau_{max_i}^*(j)$ using Equation 6
compute $N_p(j)$ using Equation 11
compute $F(T_b, i)$ using Equation 12
end for
select QAM level M from \mathbf{M} that achieves maximum of $\mathbf{F}(T_b, i)$
determine required FEC, $h_i = 2\tau_{max_i}^*$ for QAM level M {overhead of Reed Solomon FEC} $F(T_b, i) = \text{maximum of } \mathbf{F}(T_b, i)$
while $F(T_b, i) < \delta$ do
$S_f^{(new)} = \alpha S_f^{(new)}$
count = count + 1
if $\alpha^{\text{count}} > \text{maximum allowed scaling}$ then
Break
Else
compute N_p using Equation 11
compute $F(T_b, i)$ using Equation 12
end if
end while



this figure, variations of the channel condition are represented by changing E_s/N_0 . Fixed FEC and adaptive FEC were considered in this investigation. The plots exhibit a similar trend to the transmission efficiency plots in Figure 4. In Figure 6a-c, fixed FEC is used. It is observed that 256-QAM achieves the highest $F(T_b, i)$ for $E_s/N_0 > 19.5$ dB. However, for lower values of channel SNR, lower modulation levels can provide better performance. Moreover, adaptive FEC significantly improves $F(T_b, i)$ especially for high modulation levels as shown in Figure 6d-f. The plots also support the argument that SR and GBN outperform SW.

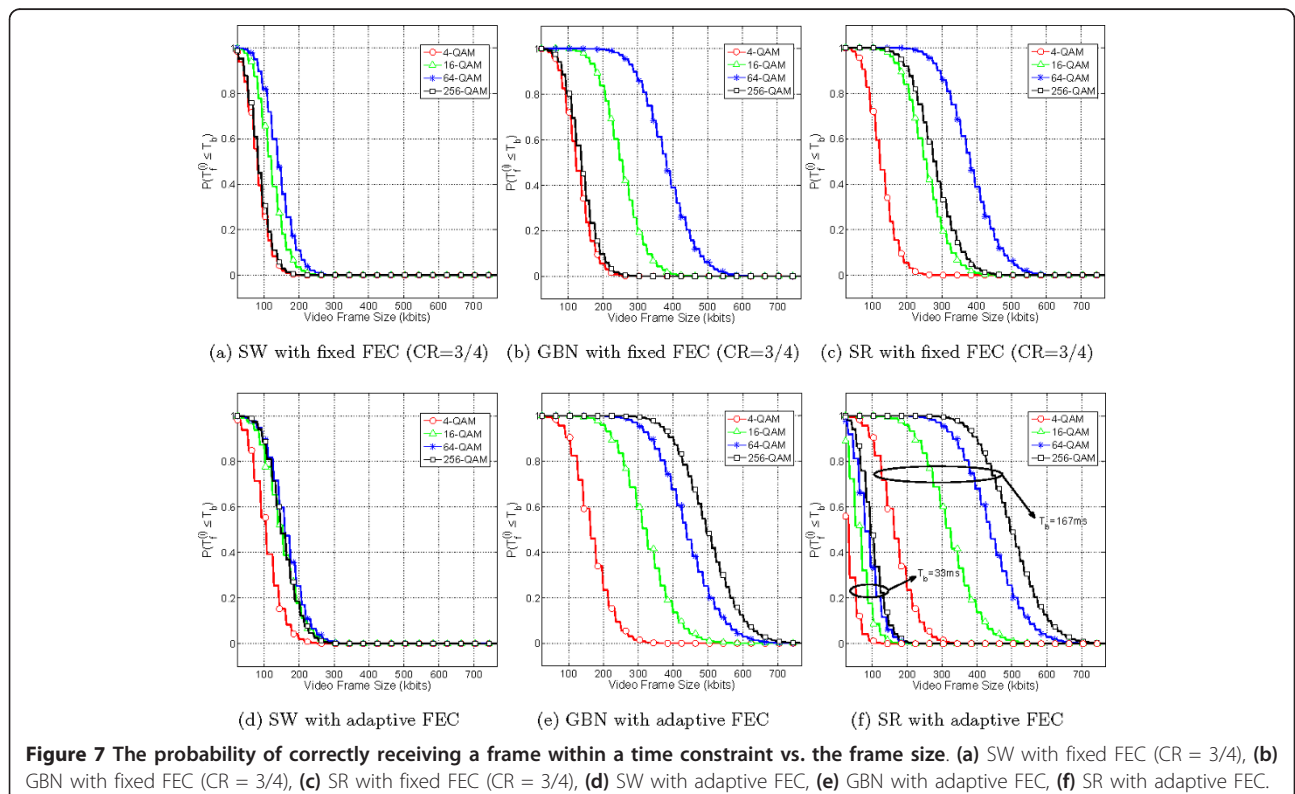
Figure 7 shows the effect of varying the modulation levels on $F(T_b, i)$ for different video frame sizes. The three ARQ schemes with fixed FEC and adaptive FEC were also considered in this investigation. $E_s/N_0 = 19$ dB and $T_b = 167$ ms were assumed when generating the plots. Intuitively, as the frame size is increased, $F(T_b, i)$ is decreased. The performance of the 256-QAM streaming system matches the performance of 4-QAM streaming system when SW and GBN are used with fixed FEC as shown in Figure 7a and 7b. This is attributed to the excessive number of retransmissions in the 256-QAM streaming system for the assumed channel condition. Nevertheless, Figure 7c shows that 256-QAM streaming system is capable of better performance with the efficient SR-ARQ.

Adaptive FEC improves the performance of the video streaming system for a given modulation level and ARQ scheme. Adaptive FEC with GBN or SR considerably enhances the performance of 256-QAM streaming system and allows it to maintain high $F(T_b, i)$ for relatively large frame sizes as shown in Figure 7e and 7f. In other words, adaptive FEC with GBN or SR allows us to transmit larger frame sizes which results in better video quality. Adaptive FEC when combined with adaptive modulation performs better than adaptive modulation alone or adaptive FEC alone.

Moreover, Figure 7f shows the effect of T_b on $F(T_b, i)$. Intuitively, for larger T_b (i.e., larger playback buffer occupancy) the probability of timely delivery of video frames increases and the likelihood of playback buffer starvation decreases.

3 Simulation results

An event-based simulator was used to test our multi-level adaptive algorithm described in Section 2. In our simulations, we considered two video sequences, the “football” sequence and the “Harry Potter” HD sequence. The “football” sequence is a short sequence (260 frames) in YUV format. On the other hand, the “Harry Potter” HD sequence is a long sequence (86384 frames) provided by [28,29].



Every time a frame is to be transmitted, the transmitter computes $F(T_b, i)$. The transmitter scales down, if necessary, the video frame by a scaling increment $\alpha = 0.95$ ($S_f^{(new)} = \alpha S_f$) until a high probability is met ($\delta = 0.9$). In the adaptive QAM scheme, before scaling a frame, the transmitter computes $F(T_b, i)$ of the different modulation levels and selects the level that achieves the highest probability. Nevertheless, if none of the modulation levels could achieve a high probability, scaling is then implemented as necessary.

3.1 Short video sequence

The “football” video sequence with a CIF resolution (352×288) was encoded into 1 BL and 10 quality ELs using the Medium Grain Scalability option in the JSVM H.264/SVC Reference Software [30,31]. This option encodes a video frame and arranges the frame bits in a way that allows discarding parts of the video frame bits (i.e., ELs) while the truncated frame will still be decodable. We used 10 ELs to allow high flexibility for our frame rate control implementation. Moreover, the “football” sequence was encoded with hierarchical B pictures and a group of pictures (GoP) of size 16. A Rayleigh fading channel with an exponentially distributed E_s/N_0 that changes per video frame was assumed. The underlying channel capacity was set to $C = 256$ Kbps. GBN-ARQ and fixed FEC (code rate $CR = 3/4$) were used. The values of B_{th} were set adaptively based on the type of the transmitted video frame where $B_{th} = 3$ for B frames, $B_{th} = 2$ for P frames, and $B_{th} = 1$ for I frames.

The performance of the different fixed QAM streaming systems in addition to the performance of the adaptive QAM streaming system are evaluated in terms of:

- playback buffer occupancy,
- percentage of video frame truncation,
- and decoded video PSNR.

Figure 8a-c describes the video streaming system performance when 4-QAM is used. The preroll threshold is set to 15 frames. During the preroll period scaling is not implemented. We see that the occupancy builds up until there are 15 frames in the buffer. Clearly, this is a very slow start (2.4 s) for only 15 frames. This indicates the poor data rate when low level modulation (4-QAM) is used. When buffer occupancy reaches 15 frames, playback starts and the buffer is drained at 30 fps. When the buffer started to approach starvation at $t = 2.7$ s, scaling was invoked. Nevertheless, the frame arrival rate could not keep up with the playback rate and starvation could not be avoided even though maximum scaling was in effect. Scaling is limited to 50% which is approximately the portion of all ELs in the encoded frames.

Within the period 6.3-7.5 s the buffer occupancy started to increase and scaling was not needed at some instants. During this period the video frame sizes were relatively small which allowed the buffer occupancy to slightly increase.

The scaling affected the quality of the decoded video as shown in Figure 8c. For example, Figure 9 illustrates the visual quality difference between the unscaled and scaled frame number 216. The quality degradation in Figure 9b can be observed in the blurry grass and the writing on the back of player number 82.

The performance of the streaming system when 16-QAM is used is shown in Figure 8d-f. The performance when 64-QAM is used is shown in Figure 8g-i. Figure 8j-l shows the performance when 256-QAM is used while Figure 8m-o shows the performance when adaptive modulation is used. It can be seen that adaptive modulation system outperforms the fixed modulation streaming systems. Adaptive modulation managed to eliminate starvation and reduced the amount of required scaling, hence, enhancing the temporal and spatial quality of the decoded video. Compared to the next best fixed modulation video streaming system, adaptive modulation reduces the average frame scaling from 10.26 to 3.90% and improves the average PSNR by 0.47 dB.

Additional simulations were carried out under the same channel realization but with different random seeds. Figure 10 shows that the adaptive modulation video streaming system outperforms fixed modulation systems in terms of average frame scaling, number of starvation instants, average SL, and average ISD for the different simulation runs.

The performance of the “football” streaming system was evaluated for an average $E_s/N_0 = 18$ dB. Its performance for a different channel realization with higher SNR per symbol (average $E_s/N_0 = 20$ dB) was also simulated (results not shown). 4-QAM performance did not improve due to its data rate limitation. On the other hand, higher modulation level performances improved especially for 256-QAM.

3.2 Long video sequence

The simulations of the “Harry Potter” streaming system were performed with the SW-ARQ and the GBN-ARQ. Each ARQ scheme was combined with fixed FEC and adaptive FEC for comparison. The RTT value was set equal to 10 ms. For the SW-ARQ simulations, $C = 1$ Mbps was assumed, whereas for GBN, $C = 512$ Kbps was assumed. For the SW, we have also simulated the video streaming system with an underlying channel capacity of $C = 512$ Kbps but the communication was infeasible with severe scaling and playback buffer starvation. Thus, we chose a higher channel capacity ($C = 1$ Mbps) for the SW video streaming system in the

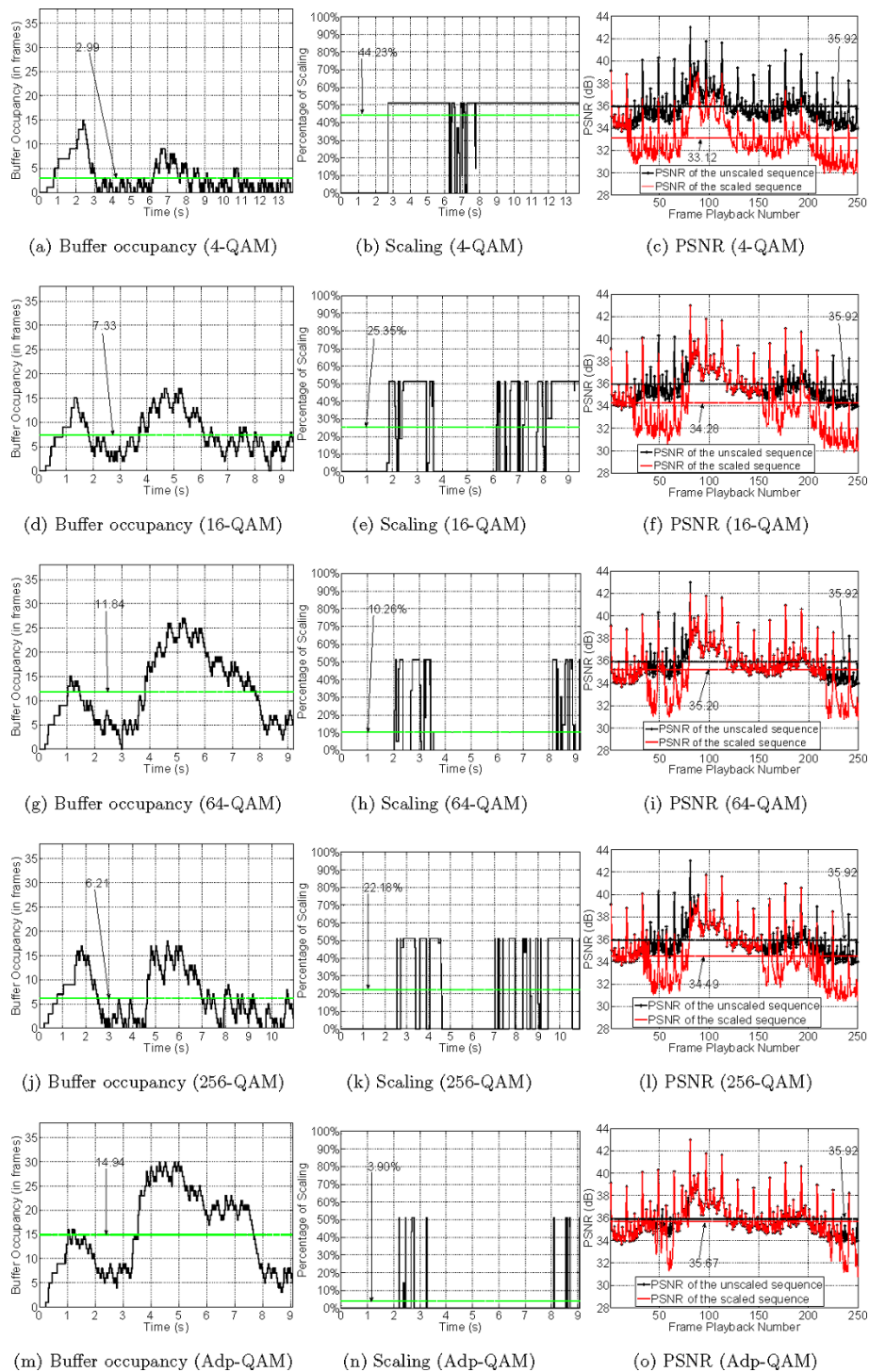


Figure 8 Performance of QAM systems with GBN and fixed FEC for the "football" sequence. (a) Buffer occupancy (4-QAM), (b) scaling (4-QAM), (c) PSNR (4-QAM), (d) buffer occupancy (16-QAM), (e) scaling (16-QAM), (f) PSNR (16-QAM), (g) buffer occupancy (64-QAM), (h) scaling (64-QAM), (i) PSNR (64-QAM), (j) buffer occupancy (256-QAM), (k) scaling (256-QAM), (l) PSNR (256-QAM), (m) buffer occupancy (Adp-QAM), (n) scaling (Adp-QAM), (o) PSNR (Adp-QAM).



Figure 9 Visual quality difference between the (a) unscaled and (b) scaled frame 216 when 4-QAM is used.

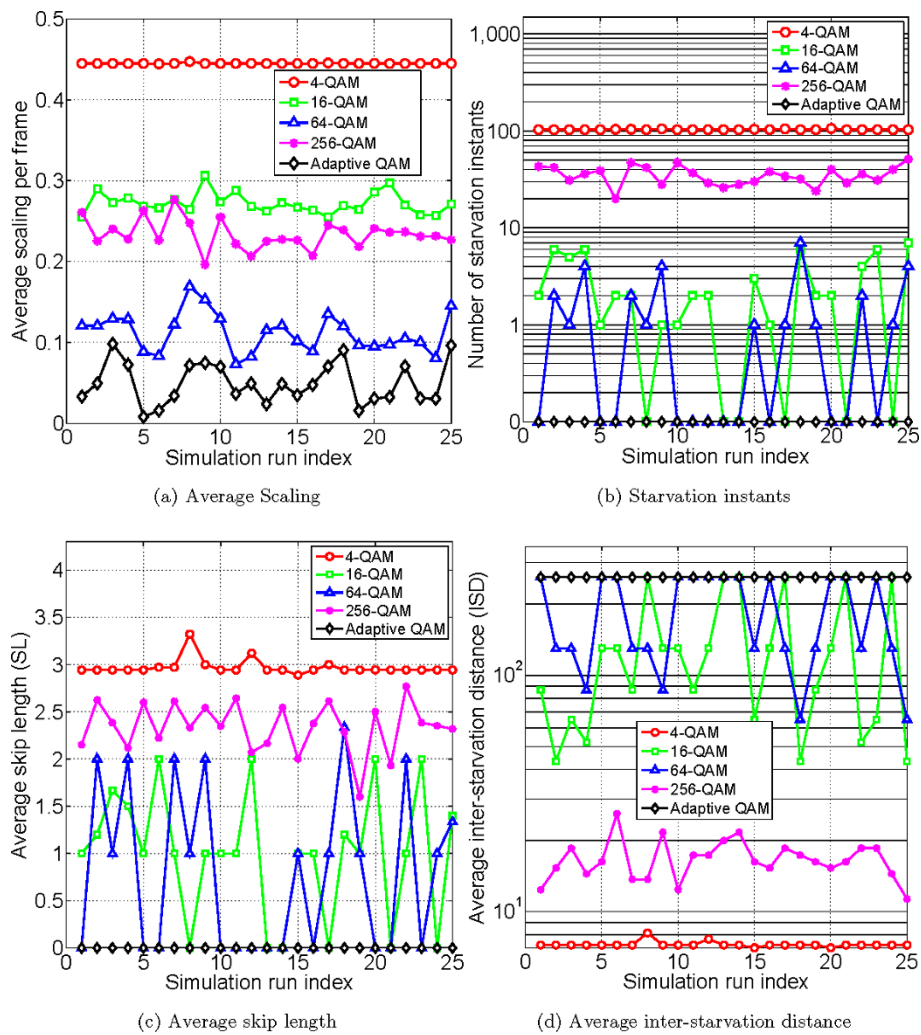


Figure 10 Multiple simulation runs for the QAM systems with GBN and fixed FEC for the "football" sequence. (a) Average scaling, (b) starvation instants, (c) average skip length, (d) average inter-starvation distance.

following simulations. The values of B_{th} were set adaptively based on the type of the transmitted video frame where $B_{th} = 16$ for B frames, $B_{th} = 8$ for P frames, and $B_{th} = 4$ for I frames.

The bar graphs in Figures 11, 12, 13, 14 compare the performances of the different QAM streaming systems for the first 15 min of the “Harry Potter” HD sequence. The comparison is in terms of:

- the average applied scaling,
- percentage of scaled frames,
- and the SL and ISD statistics.

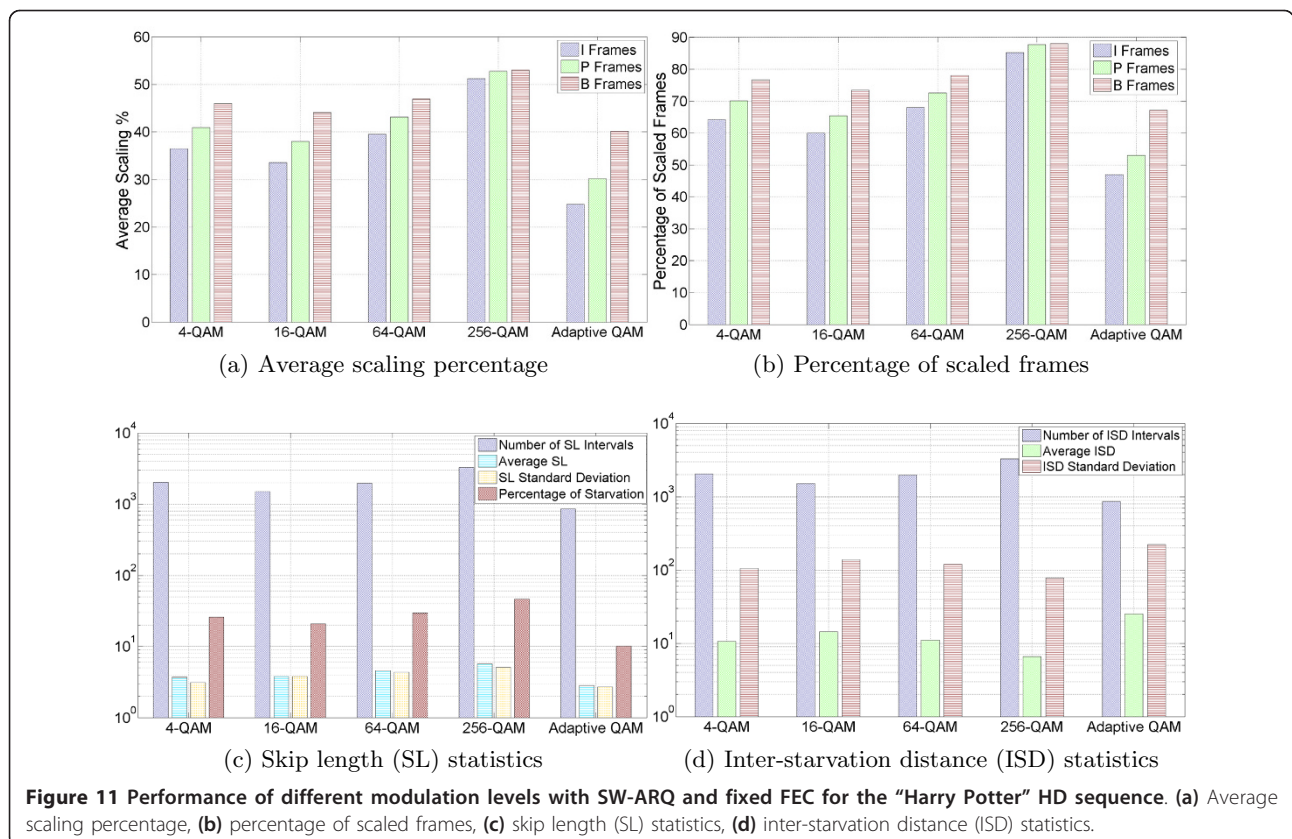
The experiments for the HD sequence were conducted using the video encoding traces from [28]; therefore the PSNR of the decoded video could not be computed using the conventional method which requires the original/reference video. The utilized temporal quality metrics (i.e., the SL and the ISD) provide a useful evaluation of the playback continuity. Large ISDs in conjunction with small SLs would result in an uninterrupted and better quality played back video. Moreover, these temporal metrics can be used with additional bitstream information such as the received video frame sizes (in bits) and the motion vectors statistics to estimate the PSNR quality of the decoded video

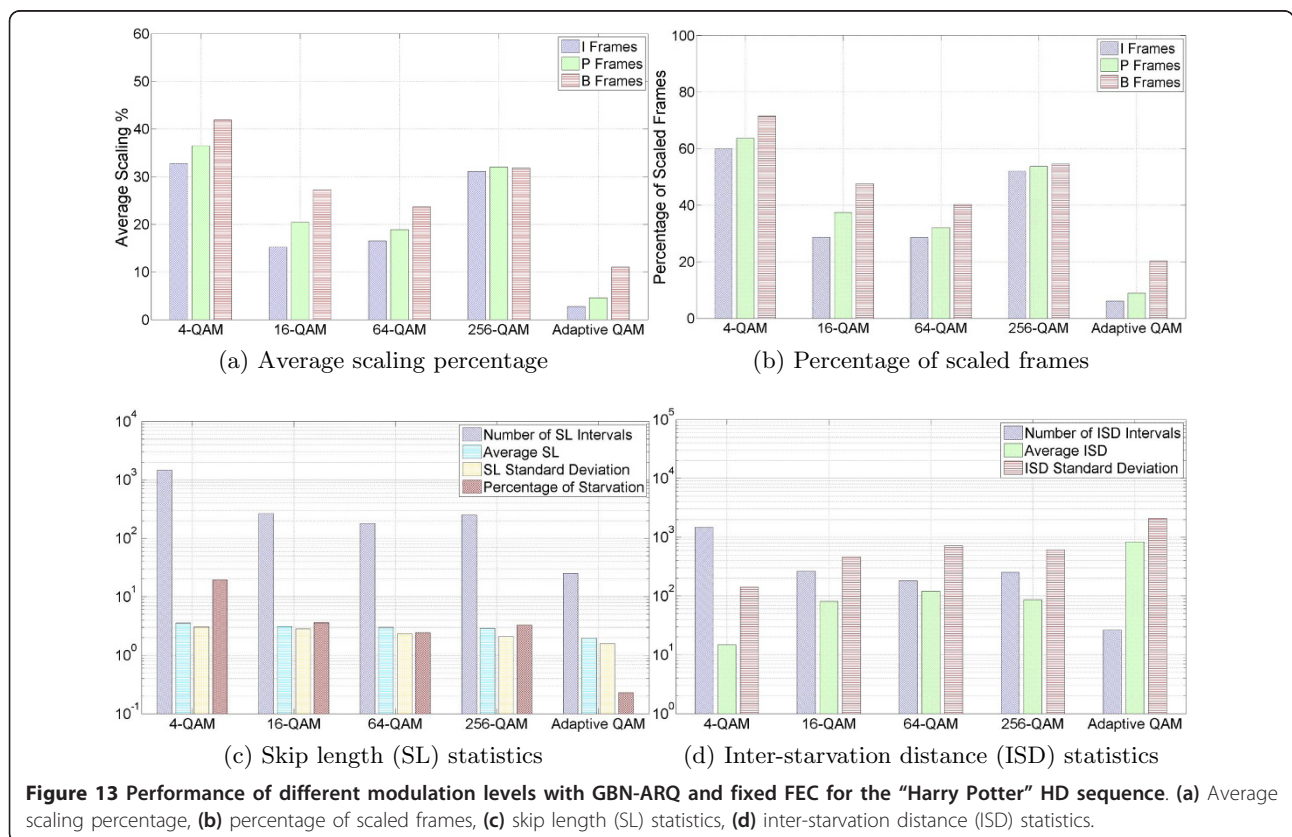
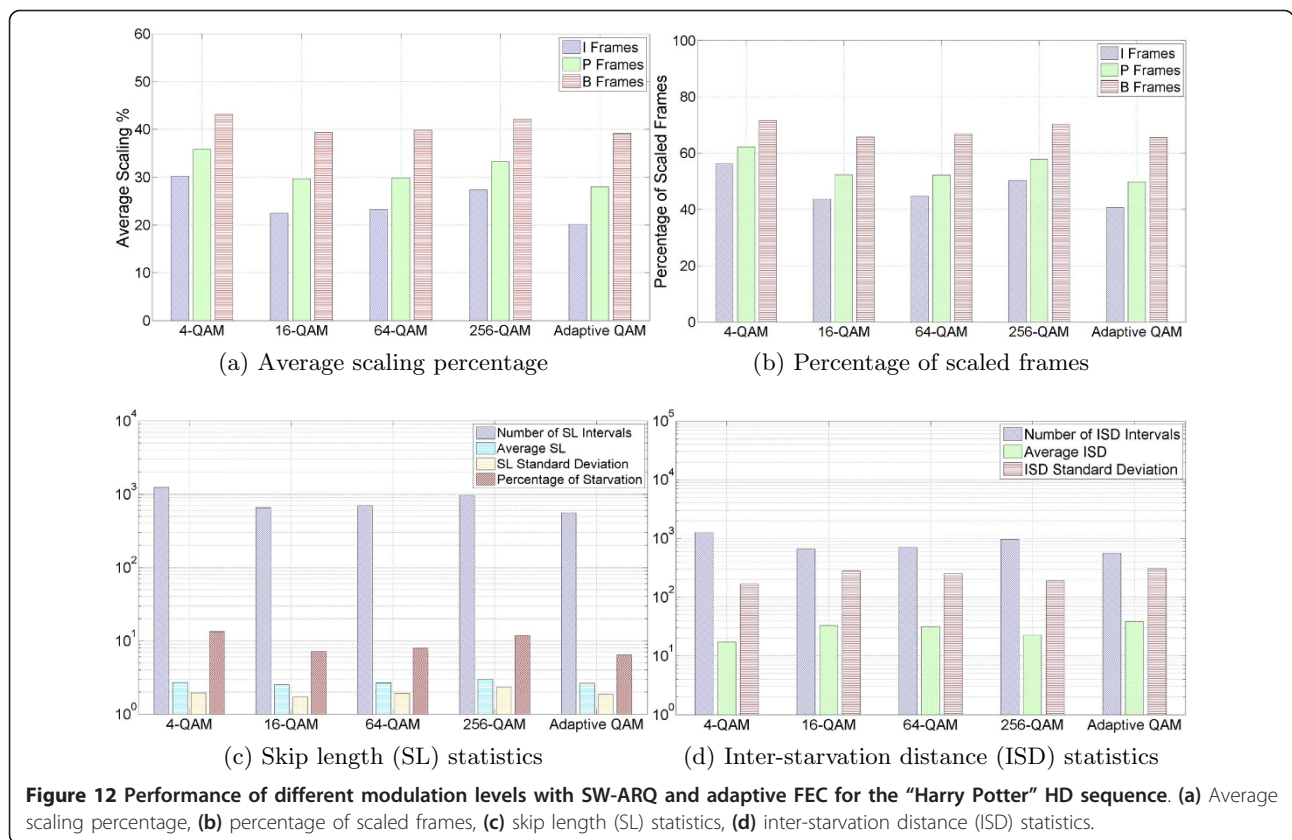
without the need for the reference video. We proposed a classification technique in [23] to predict the PSNR quality based on the SL and ISD.

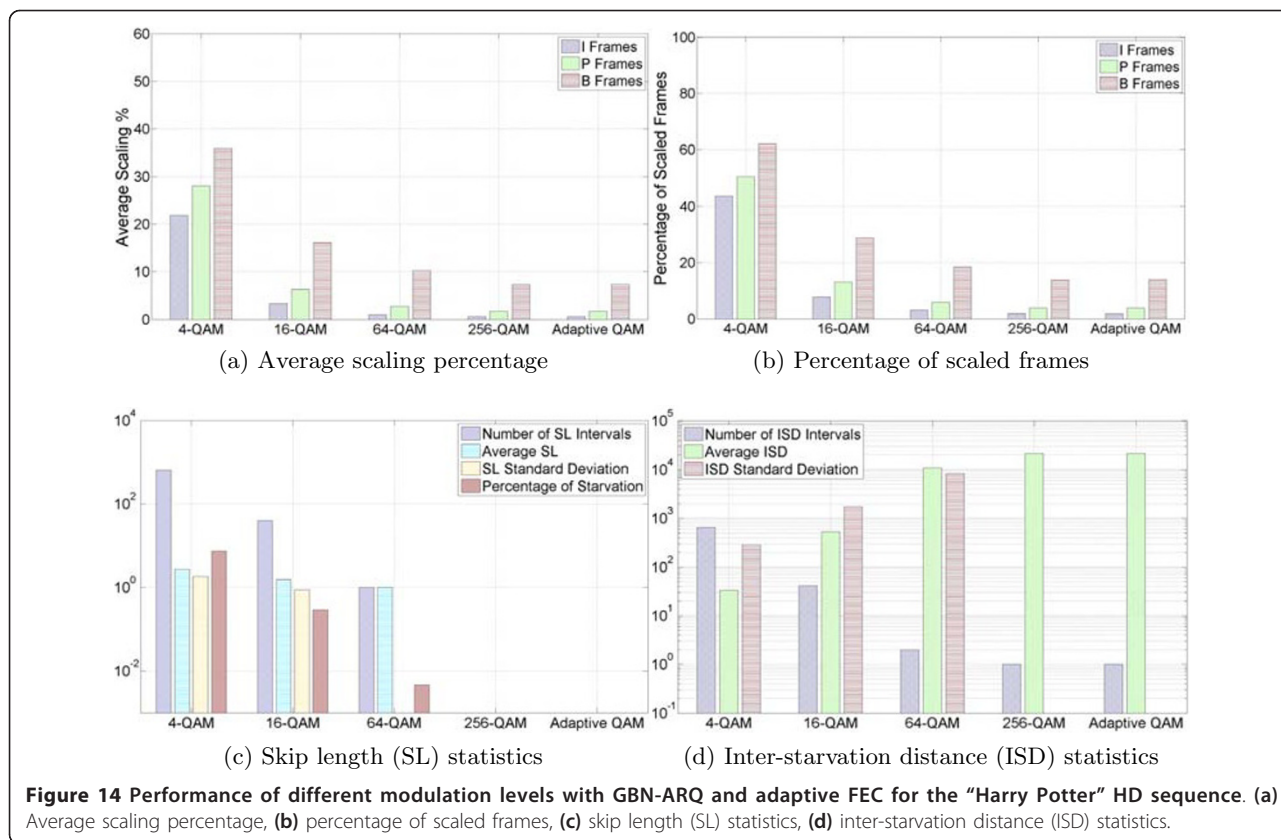
Adaptive FEC provides considerable performance improvement for all fixed modulation streaming systems. This can be observed when comparing Figure 12 with Figure 11 or Figure 14 with Figure 13. Adaptive FEC results in less average frame scaling and less number of scaled frames. It also help achieve larger ISDs and smaller SLs which their combination corresponds to uninterrupted and better quality played back video.

Moreover, adaptive modulation provides significant performance enhancement especially when fixed FEC is employed. For example, adaptive modulation reduces the percentage of playback starvation from 20.85% in the 16-QAM video streaming system (the most efficient fixed modulation level for the assumed channel realization), to 10.07% when SW-ARQ is used and from 2.44 to 0.23% when GBN-ARQ is used. For GBN with adaptive FEC, it can be noticed that the performance of 256-QAM streaming system is the best and matches the performance of adaptive QAM system.

In addition, in Figures 11a, b, 12a, b, 13a, b, 14a, b, it can be noticed that the amount of scaling is inversely proportional to the importance of video frames. Not only the average amount of scaling percentage is







reduced (as the importance of frames increases) but also the number/percentage of scaled frames is reduced. This is due to our selection of B_{th} in Equation 13. Three different values of B_{th} were assigned to each frame type where the smallest B_{th} was assigned to the I frames ($B_{th}^{(I)}$) and the largest was assigned to the B frames ($B_{th}^{(B)}$). This design translates into more budget time allocation and less frame size scaling to important frames, therefore, enhancing the quality of received video.

Table 2 demonstrates the effect of B_{th} on the performance of the adaptive QAM and fixed code rate system for the "Harry Potter" HD sequence. It can be noticed that increasing B_{th} helps in reducing the playback buffer starvation instants which in turn improves the temporal quality of the reconstructed video. However, this causes increased frame truncation which degrades the spatial quality of the reconstructed video. In Table 2, the system with $B_{th}^{(I)} = B_{th}^{(P)} = B_{th}^{(B)} = 16$ achieves the highest temporal playback quality but results in the highest frame truncation. On the other hand, the system with $B_{th}^{(I)} = B_{th}^{(P)} = B_{th}^{(B)} = 0$ results in the lowest frame truncation but achieves the lowest temporal playback quality. Adaptive selection of B_{th} based on the importance of video frames provides a better performance when compared with the fixed B_{th} systems. As can be seen in

Table 2, the proposed adaptive B_{th} assignment ($B_{th}^{(I)} = 4, B_{th}^{(P)} = 8, B_{th}^{(B)} = 16$) provides an equivalent performance in terms of temporal playback quality when compared with the highest fixed B_{th} system ($B_{th}^{(I)} = B_{th}^{(P)} = B_{th}^{(B)} = 16$). However, the adaptive B_{th} system (where $B_{th}^{(I)} = 4, B_{th}^{(P)} = 8, B_{th}^{(B)} = 16$) incurs much less truncation of important frames (I and P frames) which is reflected into a better spatial quality of the played

Table 2 Performance of adaptive QAM (GBN-ARQ and fixed FEC) system with different values of B_{th}

$B_{th}^{(I)}, B_{th}^{(P)}, B_{th}^{(B)}$	0,0,0	4,4,4	8,8,8	16,16,16	4,8,16
Average scaling % per frame					
I frames	6.84	7.54	7.95	9.60	2.67
P frames	5.27	6.17	6.69	7.96	4.07
B frames	3.08	4.32	4.66	5.25	9.41
No. of scaled frames					
I frames	196	206	221	260	79
P frames	448	496	535	648	321
B frame	1049	1364	1443	1621	2802
No. of starvation instants					
	290	210	155	122	127
No. of SL interval					
	139	65	48	37	38
Average SL					
	2.09	3.23	3.23	3.30	3.34
Average ISD					
	154.29	327.27	440.82	568.42	553.85

back video when compared with other systems with fixed B_{th} or even with the case $B_{th} = 0$.

4 Conclusions

A multi-level adaptive video streaming scheme was proposed to overcome the inherent difficulties in wireless channels. Scalable video coding was integrated with adaptive modulation and channel coding. A per-frame rate control technique was implemented based on the channel condition and the decoder buffer occupancy. Unlike other source rate control techniques which requires adjustment of video encoding parameters, the proposed scheme utilizes scalable coding which is less complex and allows real time adjustment of video frame sizes. Video streaming performance was studied for the three main ARQ schemes, Stop-and-Wait, Go-back-N, and Selective Repeat. The analysis and simulation results confirm the advantage of GBN and SR schemes over SW-ARQ in transmission efficiency. It was also shown that the performance of GBN closely matches the performance of SR when adaptive FEC is used. This makes GBN with adaptive FEC a practical and less expensive choice in terms of complexity and buffering requirements when compared to SR. In addition, it was demonstrated that bandwidth utilization is significantly enhanced with adaptive modulation and adaptive channel coding. It was also shown that adaptive modulation and channel coding reduce not only the probability of buffer starvation, but also the amount of required frame size scaling, hence, achieving better temporal and spatial video quality when compared to streaming systems employing fixed modulation. For the football sequence, the adaptive modulation streaming provided a 0.47 to 2.55 dB gain in the average PSNR. In addition, for the Harry Potter HD sequence, adaptive modulation and channel coding achieved larger ISDs and smaller SLs in the playback buffer when compared to the non-adaptive streaming systems. The sensitivity of video frames were also taken into account in our adaptive video streaming scheme to achieve better video quality.

Endnote

^aIntegration of adaptive power allocation and error concealment in addition to prioritization of video information will be considered in a future study.

Author details

¹Khalifa University of Science, Technology and Research, Sharjah, UAE

²College of Engineering, American University of Sharjah, Sharjah, UAE

Competing interests

The authors declare that they have no competing interests.

Received: 4 January 2011 Accepted: 8 December 2011

Published: 8 December 2011

References

1. P Chou, M van der Schaar, *Multimedia over IP and Wireless Networks: Compression, Networking, and Systems* (Academic Press, New York, 2007)
2. C Hsu, A Ortega, M Khansari, Rate control for robust video transmission over burst-error wireless channels. *IEEE J Selected Areas Commun.* **17**(5), 756–773 (1999). doi:10.1109/49.768193
3. M Hassan, L Atzori, M Krunz, Video transport over wireless channels: a cycle-based approach for rate control, in *Proceedings of the 12th Annual ACM International Conference on Multimedia*, ACM, New York, pp. 916–923 (2004)
4. M Hassan, M Krunz, A playback-adaptive approach for video streaming over wireless networks. in *IEEE Global Telecommunications Conference, GLOBECOM'05.* **6** (2005)
5. H Chuang, C Huang, T Chiang, Content-aware adaptive media playout controls for wireless video streaming. *IEEE Trans Multimedia.* **9**(6), 1273–1283 (2007)
6. Y Li, A Markopoulou, N Bambos, J Apostolopoulos, Joint power-playout control for media streaming over wireless links. *IEEE Trans Multimedia.* **8**(4), 830–843 (2006)
7. F Zhai, Y Eisenberg, T Pappas, R Berry, A Katsaggelos, Joint source-channel coding and power allocation for energy efficient wireless video communications. in *Proceedings of the Annual Allerton Conference on Communication Control and Computing*, Citeseer. **41**(3), 1590–1599 (2003)
8. M van der Schaar, D Turaga, Cross-layer packetization and retransmission strategies for delay-sensitive wireless multimedia transmission. *IEEE Trans Multimedia.* **9**(1), 185–197 (2006)
9. M Hassan, M Krunz, Video streaming over wireless packet networks: an occupancy-based rate adaptation perspective. *IEEE Trans Circuits Syst Video Technol.* **17**(8), 1017–1027 (2007)
10. A Argyriou, Error-resilient video encoding and transmission in multirate wireless LANs. *IEEE Trans Multimedia.* **10**(5), 691–700 (2008)
11. Y Li, A Markopoulou, J Apostolopoulos, Content-aware playout and packet scheduling for video streaming over wireless links. *IEEE Trans Multimedia.* **10**(5), 885–895 (2008)
12. G Ji, B Liang, Stochastic rate control for scalable VBR video streaming over wireless networks, in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, IEEE, New York, pp. 1–6 (2010)
13. H Luo, S Ci, D Wu, H Tang, End-to-end optimized TCP-friendly rate control for real-time video streaming over wireless multi-hop networks. *J Vis Commun Image Representation.* **21**(2), 98–106 (2010). doi:10.1016/j.jvcir.2009.06.006
14. N Nejati, H Yousefzadeh, H Jafarkhani, Distortion optimal transmission of multi-layered FGS video over wireless channels. *IEEE J Selected Areas Commun.* **28**(3), 510–519 (2010)
15. E Maani, A Katsaggelos, Unequal error protection for robust streaming of scalable video over packet lossy networks. *IEEE Trans Circuits Syst Video Technol.* **20**(3), 407–416 (2010)
16. B Zhang, M Wien, J Ohm, A novel framework for robust video streaming based on H.264/AVC MGS coding and unequal error protection, in *International Symposium on Intelligent Signal Processing and Communication Systems, 2009. ISPACS 2009*, IEEE, New York, pp. 107–110 (2010)
17. R Lu, J Lin, T Chiueh, Cross-layer optimization for wireless streaming via adaptive MIMO OFDM, in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, New York, pp. 2255–2258 (2010)
18. H Yang, K Rose, Optimizing motion compensated prediction for error resilient video coding. *IEEE Trans Image Process.* **19**(1), 108–118 (2009)
19. R Rasouli, H Rabiee, M Ghanbari, A joint source-channel rate-distortion optimization algorithm for H.264 codec in wireless networks. *Adv Comput Sci Eng*, **1**: 946–950 (2009)
20. J Xu, X Yang, S Zheng, L Song, Group-of-pictures-based unequal error protection for scalable video coding extension of H.264/AVC. *Opt Eng.* **48**, 060502 (2009). doi:10.1117/1.3152774
21. S Kallel, Analysis of a type II hybrid ARQ scheme with code combining. *IEEE Trans Commun.* **38**(8), 1133–1137 (1990). doi:10.1109/26.58745
22. R Deng, Hybrid ARQ schemes employing coded modulation and sequence combining. *IEEE Trans Commun.* **42**(6), 2239–2245 (1994). doi:10.1109/26.293675
23. M Hassan, T Landolsi, H Mukhtar, T Shanableh, Skip length and inter-starvation distance as a combined metric to assess the quality of transmitted video, in *International Workshop on Video Processing and Quality Metrics for Consumer Electronics* (2010)

24. K Cho, D Yoon, On the general BER expression of one- and two-dimensional amplitude modulations. *IEEE Trans Commun.* **50**(7), 1074–1080 (2002). doi:10.1109/TCOMM.2002.800818
25. M Simon, M Alouini, *Digital Communication over Fading Channels* (IEEE, New York, 2005)
26. S Lin, D Costello, *Error Control Coding* (Prentice-Hall, Englewood Cliffs, 1983)
27. A León-García, I Widjaja, *Communication Networks: Fundamental Concepts and Key Architectures* (McGraw-Hill Science/Engineering/Math, New York, 2004)
28. Video Trace Library, <http://trace.eas.asu.edu>
29. P Seeling, M Reisslein, B Kulapala, Network performance evaluation using frame size and quality traces of single-layer and two-layer video: a tutorial. *Commun Surv Tutor.* **6**(3), 58–78 (2009)
30. H Schwarz, D Marpe, T Wiegand, Overview of the scalable video coding extension of the H. 264/AVC standard. *IEEE Trans Circuits Syst Video Technol.* **17**(9), 1103–1120 (2007)
31. H.264/SVC JSVM Reference Software, <http://iphome.hhi.de/suehring/tml/>

doi:10.1186/1687-1499-2011-199

Cite this article as: Mukhtar et al.: An occupancy-based and channel-aware multi-level adaptive scheme for video communications over wireless channels. *EURASIP Journal on Wireless Communications and Networking* 2011 **2011**:199.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
