Media Interactive Streaming in High Speed Downlink Packet Access Wireless Channel

Kostas E. Psannis, Marios G. Hadjinicolaou
School of Engineering and Design, Dept. of Electronic & Computer Engineering, Brunel University, Uxbridge, Middlesex, UB8 3PH, UK

Abstract: This paper presents an efficient method for supporting interactive media streaming in High Speed Downlink Packet Access (HSDPA) wireless channel. The corresponding normal/interactive video streams are obtained by encoding the original uncompressed video file as a sequence of I-P(I)-B(I) frames and I-P(I) frames respectively. Mechanisms for controlling the normal/interactive request are also presented and their effectiveness is assessed through simulations. Wireless media services are supported with acceptable visual quality at the wireless client –end.

Key-Words: Wireless MPEG-Video, Interactive Operations, Broadband Wireless

1 Introduction
Streaming of live or stored video content to group of mobile devices comes under the scope of High Speed Downlink Packet Access (HSDPA) Services [1]-[10]. HSDPA enables operators to provide newer, faster services such as on-demand streaming video/film, high-resolution interactive gaming, multimedia music tracks, “push-to-watch” services and access to large email attachments—all at speeds nearly three times faster than today’s commercial 3G UMTS networks and four times faster than EDGE networks [1]-[6]. High Speed Downlink Packet Access (HSDPA) is a packet based technology for W-CDMA downlink with data transmission rates of 4 to 5 times that of current generation 3G networks (UMTS) and 15 times faster than GPRS. The latest release boosts downlink speeds from the current end-user rate of 384 kbps (up to 2 Mbps according to standards) to a maximum value according to standards of 14.4 Mbps [7]. Real life end-user speeds will be in the range of 2 to 3 Mbps. HSDPA provides a smooth evolutionary path for Universal Mobile Telecommunications System (UMTS) networks to higher data rates and higher capacities, in the same way as Enhanced Data rates for GSM Evolution (EDGE) does in the Global System for Mobile communication (GSM) world[8]. The introduction of shared channels for different users will guarantee that channel resources are used efficiently in the packet domain, and will be less expensive for users than dedicated channels [9]. HSDPA was introduced in the Third Generation Partnership Project (3GPP) release 5 standards. Assuming comparable cell sizes, it is anticipated that by using multi-code transmission it will be possible to achieve peak data rates of about 10 Mbit/s (the maximum theoretical rate is 14.4 Mbit/s). This will result in a six- to seven-fold throughput increase during an average downlink packet session compared with the Downlink Shared Channel (DSCH) standards of 3GPP release 99 [10].

The High Speed Downlink Packet Access service center should be able to accept and retrieve content from external sources and transmit it using error resilient schemes. In recent years several error resilience techniques have been devised [11]-[17]. In [11], an error resilience entropy coding (EREC) have been proposed. In this method the incoming bitstream is re-ordered without adding redundancy such that longer VLC blocks fill up the spaces left by shorter blocks in a number of VLC blocks that form a fixed-length EREC frame. The drawback of this method is that the codes between two synchronization markers are dropped, results any VLC code in the EREC frame be corrupted due to transmission errors. A rate-distortion frame work with analytical models that characterize the error propagation of the corrupted video bitstream subjected to bit errors was proposed [12]. One drawback of this method is that it assumes the actual rate distortion characteristics are known, which makes the optimization difficult to realize practical. In addition the error concealment is not considered. Error concealment has been available since H.261 and MPEG-2 [3]. The easiest and most practical approach is to hold the last frame, that was successful decoded. The best known approach is to use motion vectors that can adjust the image more naturally when holding the previous frame. More complicated error concealment techniques consist of a combination of spatial, spectral and temporal
interpolations with motion vector estimation. In [13] an error resilience transcoder for General Packet Radio Service (GPRS) mobile accesses networks is presented. In this approach the bit allocation between insertion error resilience and the video coding is not optimized. In [14] optimal error resilience insertion is divided into two subproblems: optimal mode selection for macroblocks and optimal resynchronization marker insertion. Moreover in [15] an approach to recursively compute the expected decoder distortion with pixel -level precision to account for spatial and temporal error propagation in a packet loss environment is proposed. Both of these methods [14] and [15], inter frame dependencies are not considered. In MPEG-4 video standard [4], application layer error resilient tools were developed. At the source coder layer, these tools provide synchronization and error recovery functionalities. Efficient tools are Resynchronization Marker and Adaptive Intra frame Refresh (AIR). The marker localizes transmission error by inserting code to mitigate errors. AIR prevents error propagation by frequently performing intra frame coding to motion domains. However AIR is not effective in combating error propagation when I-frames are less frequent. A survey of error resilient techniques for multicast applications for IP-based networks is reported in [15]. It presents algorithms that combine ARQ, FEC and local recovery techniques where the retransmissions are conducted by multicast group members or intermediate nodes in the multicast tree. Moreover video resilience techniques using hierarchical algorithms are proposed where transmission of I- P- and B frames are sent with varying levels of FEC protection. Some of the prior research work on error resilience for broadcast terminals focuses on increasing FEC based on the feedback statistics for the user [16]. A comparison of different error resilience algorithms for wireless video multicasting on wireless local area networks is reported in [17]. However in the literature survey none of the methods applied error resilience techniques at the video coding level to support normal/interactive streaming services. Error resilient (re)-encoding is a technique that enables robust streaming of stored video content over noisy channels. It is particularly useful when content has been produced independent of the transmission network conditions or under dynamically changing network conditions. Developing error resilience technique which provides high quality of experience to the end mobile user is a challenge issue. In this paper we propose a very efficient error resilience technique for HSDPA services. Similar to [18] by encoding separate copies of the video, the normal/interactive video streams are supported with minimum additional resources. The corresponding normal/interactive versions are obtained by encoding every (i.e. uncompressed) frame of the original movie as a sequence of I- P(I)- B(I) and I- P(I) frames using different GOP pattern of MPEG-based video compression. The paper is organized as follows. In Section 2 the preprocessing steps required to support normal/interactive streaming services over HSDPA wireless channels are detailed. Section 3 presents the extensive simulations results. Finally conclusions are discussed in Section 4.

2 Proposed Algorithm

2.1 Normal Playback

The problem addressed is that of transmitted a sequence of frames of stored video using the using the minimum amount of energy subject to video quality and bandwidth constraints impose by the HSDPA wireless channel.

Assume that I- frame is always the start point of a joining multicast session. Since I- frames are decoded independently, switching from leaving to joining multicast session can been done very efficiently. The corresponding video streams are obtained by encoding the original uncompressed video file as a sequence of I-P(I)-B(I) frames using different GOP pattern [18]. P(I) and B(I) frames are coded using motion estimation and each one has a dependency only on the preceding I- frame. This result that corruption of P-frame or B-frame do not affect the next P/B-frame to be decoded. On the other hand, it increases the P(I)-B(I) frame sizes. The problem addressed is that of transmitted a sequence of I- P(I)-B(I) frames using the minimum amount of energy subject to video quality and the delay constraints imposed by the wireless network application [19]-[20].

Intra/Inter Macro_ Block (MB) i is coded using a quantizer q(i) from a set of allowable quatizers Q=(2,4,6,8,10) resulting in video packet of B(i) bits with distortion D(i). This intra/inter packet is transmitted at rate R(i) bits per second chosen from a set of allowable channel rate R=(100Kbps, 200Kbps, 300Kbps, 400Kbps). Our goal is to properly select a quantizer and q(i) and channel rate R(i) for each intra MB in order to minimize the energy required to transmit the intra/inter packet subject to both distortion constraint and delay per intra MB constraint. We consider an optimization model [19] as follows.

We consider a system where source coding decisions are made using the minimum amount of
It should be emphasized that a major limitation in wireless networks is that mobile users must rely on a battery with a limited supply of energy. Effectively utilizing this energy is a key consideration in the design of wireless networks. Our goal is to properly select a quantizer \( q(i) \) in order to minimize the energy required to transmit the sequence of I-B(P)-I frames subject to both distortion and channel constraints. A common approach to control the size of an MPEG frame is to vary the quantization factor on a per-frame basis [21]. To bound the size of predicted frames, an \( P(I)-B(I) \) frames are encoded such that its size fulfills the following constraints

\[
\text{BitBudget} \{ I, B(I), P(I) \} \leq C_{\text{Rate}}
\]

\[
\text{BitBudget} \{ I, B(I), P(I) \} \leq D_{\min}
\]

### 2.2 Interactive Mode

To support interactive functions, the server maintains multiple, different encoded versions of each movie. One version, which is referred to as the normal version is used for normal-speed playback. The other versions are referred to as interactive versions. Each interactive version is used to support Fast/Jump Forward/Backward Slow Down/Reverse and Reverse at a variable speedup. The server switches between the various versions depending on the requested interactive function. Assume that I-frame is always the start point of interactive mode. Since I-frames are decoded independently, switching from normal play to interactive mode and vice versa can be done very efficiently. Note that only one version is transmitted at a given instant time [18].

The corresponding interactive version is obtained by encoding every \( N \)-th (i.e., uncompressed) frame of the original movie as a sequence of I-P(I) frames \( (N_{\text{interactive}} = \text{variable}, M_{\text{interactive}} = 1) \). Effectively this results in repeating the previous I-frame in the decoder, enhancing the visual quality during the interactive mode. This is because it freezes the content of the I-type frames, reducing the visual discontinuities. Moreover P(I) frames are produced between successive I-frames in order to maintain the same frame of normal play and achieve full interactive operations at variable speeds. To improve the marketability of video streaming applications, the client should interact with the content of the presentation deciding the viewing schedule with the full range of interactive functions. The full interactive functions can be supported as follows [18].

Fast Forward/Rewind (FF/FR) is an operation in which the client browses the presentation in the forward/backward direction with normal sequence of pictures. This function is supported by abstracting all the I-type frames of the original (uncompressed) movie in the forward/backward direction and encoding each frame as a sequence of \( I-P(I) \) frames. Jump Forward/Backward (JF/JB) is an operation in which the client jumps to a target time of the presentation in the forward/backward direction without normal sequence of pictures. Therefore the users jump directly to a particular video location. Jump Forward/Backward operation is supported by skipping forward/backward some I-type frames of the original (uncompressed) movie and encoding each of the remaining frames as a sequence of I-P(I) frames.

Rewind operations can be supported by abstracting all the I-type frames of the original (uncompressed) movie in reverse order and generate P(I) frames as many as P- and B-frames in a GoPs of normal playback \( (N = N_{\text{interactive}}) \).

Slow Down/Reverse (SD/SR) is an operation in which the video sequence is presented forward/backward with a lower playback rate. This function can be supported by abstracting all the I-type frames of the original uncompressed movie in the forward/backward order and generate P(I) frames as many as P- and B-frames in a recording ratio (RR) of normal playback \( (N_{\text{interactive}} = \text{RR}) \). We consider a system where source coding decisions are made using the minimum amount of energy

\[
\min E^{q(i)} \{ I, P(I) \} \text{ subject to minimum Distortion (} D_{\min} \text{) at the mobile client and the available Channel Rate (} C_{\text{Rate}} \text{) required by wireless network. Hence}
\]

\[
\min E^{q(i)} \{ I, P(I) \} \leq C_{\text{Rate}}
\]

\[
\min E^{q(i)} \{ I, P(I) \} \leq D_{\min}
\]

### 2.2 Reference Scheme

In order to illustrate the advantage of the proposed algorithm we consider a system where source coding
decisions are made without any constraints for the normal and interactive bitstreams. Thus

\[ E^{\text{normal}}_L(1,B(I),P(I)) \] (Normal stream)

\[ E^{\text{interactive}}_L(1,P(I)) \] (Interactive stream)

3 Experiments

There are two types of criteria that can be used for the evaluation of video quality; subjective and objective. It is difficult to do subjective rating because it is not mathematically repeatable. For this reason we measure the visual quality of the interactive mode using the Peak Signal-to-Noise Ratio (PSNR). We use the PSNR of the Y-component of a decoded frame. The MPEG bitstream used for simulation is the “Mobile” sequence with 180 frames, which was encoded at 2Mbps, with a frame rate of 30fps. The Group of Pictures format was N=15 and M=3.

Figure 1 shows the PSNR plot per frame obtained with the proposed algorithm and the reference scheme during the normal mode.

The number of I-P(I) frames in the interactive mode can be computed as follows

\[ TF^{\text{Interactive}}_{I,P(M)} = I_{\text{number}} \times N_{\text{interactive}} \]

where,

\[ I_{\text{number}} = \frac{TF}{N} \]

is the number of I-frames and 

\[ (N_{\text{interactive}}, M_{\text{interactive}}) \]

are the new re-encoding parameters.

Hence, for TF=180, (N=15,M=3) and 

\[ (N_{\text{interactive}} = 5, M_{\text{interactive}} = 1) \]

\[ TF^{\text{Interactive}}_{I,P(M)} = 60 \].

Figure 2 shows the PSNR plot per frame obtained with the proposed algorithm and the reference scheme during the interactive mode.

4 Conclusions

This paper proposed an efficient method for supporting interactive media streaming in High Speed Downlink Packet Access (HSDPA) wireless channel. The method is based on storing multiple differently encoded versions of the video stream at the server. The corresponding normal/interactive video streams are obtained by encoding the original uncompressed video file as a sequence of I-P(I)-B(I) frames and I-P(I) frames respectively. Clearly the proposed very efficient approach achieves acceptable visual quality for both the normal interactive streams over HSDPA wireless channel.

References:


