H.264/AVC INTERLEAVING FOR 3G WIRELESS VIDEO STREAMING

Thomas Schierl, Markus Kampmann+, and Thomas Wiegand**

*Fraunhofer Institute for Telecommunications
Heinrich Hertz Institute (HHI)
Image Processing Department
Einsteinufer 37, 10587 Berlin, Germany
{schierl, wiegand}@hhi.fraunhofer.de

+Ericsson Research,
Mobile Multimedia Networks
Ericsson GmbH, Eurolab R&D
Ericsson Allee 1, 52134 Herzogenrath, Germany
Markus.Kampmann@ericsson.com

ABSTRACT

We present a streaming system that utilizes interleaved transmission for real-time H.264/AVC video in 3G wireless environments with benefits shown especially in the presence of link outages. In the 3GPP Packet-switched Streaming Service Rel. 6, H.264/AVC and its RTP payload format are specified. The RTP payload format allows interleaved transmission of NAL units of H.264/AVC. Our simulations also include audio into the interleaving framework and are conducted within a testbed that emulates a 3G network including block error rates on the physical layer, a buffer for retransmission on the link layer for different error rates, and link outages. The experimental results demonstrate the superior performance of interleaving for typical link outage settings.

1 INTRODUCTION

Applications that require transmission over 3G mobile networks that have severe delay constraints including real time-streaming [1] have to cope with the problems arising from the physical network layer of these networks, e.g. link outages which arise from handovers between mobile network cells. Usually there is a buffer for a client connection in the mobile network. Such buffers are usually moved into the new cell, the client is visiting. Due to the link outage the connection is interrupted for a while. The application has to cope with this kind of transmission problem. The network buffer can compensate the link outage for some time meaning that not all data that are affected will get lost. Another problem that could arise after a link outage is that the used the network buffer filling level can reach the limit. If the link is up again, the buffer can not compensate other irregularities.

In 3GPP a standard for multimedia streaming over mobile networks called PSS (Packet-switched Streaming Service) [2] has been specified. We present the benefits of using interleaved transmission for H.264/AVC video [3] in 3G wireless environments, especially in the presence of link outages with a network buffer on the link layer. The RTP payload format for H.264/AVC [4] provides a number of features for interleaved transmission, especially slice (NAL unit) interleaving is supported by the format including packets which can contain slices of different pictures (access units). An additional header value allows the reordering to decoding order at the client.

An overview of the 3G streaming system is shown in Fig.1. The rest of the paper is organized as follows. Section 2 gives a detailed description of the used and developed techniques. In section 3 results are presented.

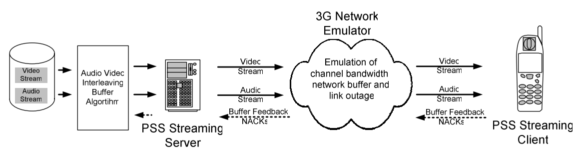

Fig. 1 – 3G Wireless Streaming System

2. AUDIO/VIDEO FRAME INTERLEAVING

The adaptation algorithm of the server is based on feedback and estimation of transmission conditions and client status. As feedback mechanism, RTCP *Receiver Reports* (RR) [5] with the buffer report extension specified in [2] and NACK reports as specified in [6] are used. The *Highest Received Sequence Number* (HRSN) and the *Next Sequence Number* (NSN) to be decoded are exploited for determining the buffer levels of the network and the client.

One important piece of the adaptation algorithm is the *Transmission Rate Control* (TRC). We used a simple approach to avoid overfilling of network as well as client buffers. Such an approach has been proposed in [7]. By recalculating the current buffer levels, network buffer

0-7803-9332-5/05/$20.00 ©2005 IEEE

N_{curr} and client buffer C_{curr}, using the *Highest Transmitted Sequence Number* (HTSN) of the server, the HRSN and the NSN of the client, the server is controlling the transmission of RTP packets. Fig. 2 illustrates the buffer calculation.

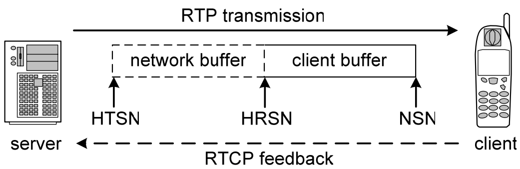

Fig.2 - Reported Sequence Numbers

Two maximum buffer levels are given: The maximum network buffer $N_{\max}$, which has to be an approximated, safe maximum of the used access network technology and the maximum client buffer $C_{\max}$, which is negotiated during the RTSP session establishment via 3GPP extensions [2]. The server is sending an RTP packet *i*, if both of the two following conditions are fulfilled:

$$C_{\max} \geq C_{curr}, C_{curr} = \sum_{i=HTSN}^{NSN} PacketSize_i \qquad (1)$$

$$N_{\max} \geq N_{curr}, N_{curr} = \sum_{i=HTSN}^{HRSN} PacketSize_i \qquad (2)$$

The central part of the adaptation algorithm is the *Audio/Video Frame Interleaving*, which is based on an approach proposed in [8]. The approach extends the well known *Priority Based Scheduling* (PBS) with using feedback for determining the current priority buffer levels at the client and some further changes. We extended this algorithm with a new startup phase in which we are deploying PBS only up to all buffers have reached a certain level. Additionally we extended the scheme to be used with H.264/AVC and its RTP payload format. The client feedback defined in PSS is not only used for the TRC but also for determining the buffer levels at the client. NACKs are additionally used to correct the estimated level values at the server. Fig.3 gives an idea of the algorithm behavior. The priority levels become lower from top to bottom, where the audio frames (‘A’) have the highest priority level and video has the next lower priority levels. A detailed description is given in below.

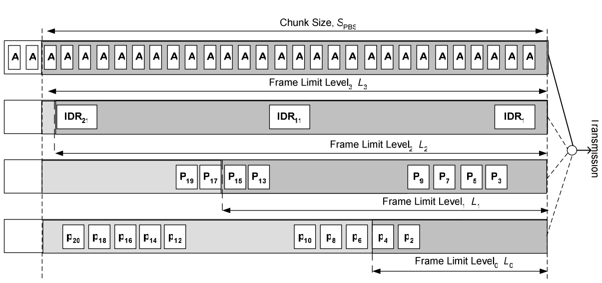

Fig. 3 – A/V Interleaving Algorithm

PBS is fulfilled for a chunk size of S_{PBS} frames. In this chunk the frames are transmitted in priority order, i.e. higher priority frames are transmitted first. If the client reports that a priority level *i* has reached a certain frame limit L_i, the server will fill up the next lower level. If all frame limits are reached, the rest of the chunk will be sent in Earliest Deadline First (EDF) manner, i.e. in the decoding order. At startup, as long as no feedback of the client has been received, straight PBS is deployed. The server uses NACK feedbacks and the reported sequence numbers for calculating L_i for each priority level *i*.

As shown in Fig. 3 the boxes labeled ‘A’ represent audio frames, which have the highest priority. Boxes labeled with ‘IDR’ stand for H.264/AVC IDR pictures and have the next lower priority. Boxes labeled with ‘P’ stand for referenced H.264/AVC P pictures and boxes labeled with ‘p’ stand for non referenced P pictures, which have the lowest priority. The numbers of the video pictures represent the decoding order number (DON) as specified in [4]. In order to achieve a reordering from sending to decoding order at the client, we used the ‘Packetization Mode 2’ of [4]. The H.264/AVC Baseline Profile [3], which is also a part of [2], allows the use of non referenced pictures. The dropping of such pictures does not affect the decoding process of other pictures of the stream. Each lower Reference Layer which is also dropped does not influence the decoding process and thus the quality of higher layer pictures either. This is also known as Temporal Scalability (TS). Thus the algorithm contains an implicit TS behavior.

The A/V Interleaving has two benefits. In case of link outage higher pre buffers for the higher priority levels allow to go on playing with reduced quality, e.g. the full audio stream and additional a ‘slide show’ of IDR pictures. When using EDF, the frequent playout will interrupt earlier. The second benefit is an implicit Bit Stream Thinning algorithm. When the channel does not provide the required transmission rate, the algorithm will transmit only pictures of higher priority levels, since the algorithm only sends the next lower priority level in case feedback has reported that the current sent priority level has reached the desired filling level. Note that in the start up phase it is important to correctly set the level limits in order not to neglect the lower priority levels. And on the other hand it is important to build up a pre-buffer for the highest priority levels, which is large enough to get over a possible link outage. If the values are not correctly selected, it could be that too much time is used for building up a pre buffer, which has not a better performance than a pre buffer built up by plain EDF scheduling. Another important fact is the interaction with the feedback. The control loop of the A/V Interleaving algorithm mainly depends on the feedback interval and delay. If the delay is too large, the algorithm cannot work effectively.

3. EXPERIMENTAL RESULTS

The testbed as shown in Fig. 1 is an implemented real system and was used for obtaining the experimental results. The network emulator allows the emulation of a 3G network including block error rates (BLER) on the physical layer and a buffer for retransmission on the link layer for different error rates. Note that BLER-caused retransmissions can raise the delay and decrease the effective bandwidth available. In addition to that, the 3G network emulator can generate link outages.

For analyzing the system we use network conditions typical for a WCDMA network. The channel rate is fixed to a value of 64 kbit/s with the presence of 5 % Block Error Rate which reduces the available bandwidth for RTP traffic to 56 kbit/s. The network has a buffer for link layer retransmission of 20 kBytes. Especially at the start of the link outage time period this buffer can collect the packets which are still sent out by a server without a transmission rate control. For the following tests we have used a pre buffer of 4 s (27500 byte client buffer startup size, maximum value 50000 byte) at the client and we have emulated a link outage after 20 s of session startup with a duration of 5 s or 8 s. The media contains a bit-stream conforming H.264/AVC Baseline Profile Level 1b with an ‘IDRpPpPp…’ GOP structure and an IDR picture being repeated every 13th picture. Uppercase 'P' stands for P pictures that are reference pictures for motion compensation and lower case 'p' stands for P pictures that are non-reference pictures. The video resolution is QCIF at 12.5 fps and it is encoded at a bit rate of 45 kbit/s. As additional audio stream an AAC LC stream with 16 kHz mono at 10 kbit/s is used. The resulting stream rate of 55 kbit/s optimally fits to the network conditions. We have conducted 3 experiments:

1. No adaptation and 5 s link outage during the session
2. TRC, A/V Interleaving and 5 s link outage during the session
3. TRC, A/V Interleaving and 8 s link outage during the session

The results are shown in Figs. 4, 5, and 6. The right y-axis shows the network buffer level in bytes. The interleaving parameters are shown in Tab.1. The values of S_{PBS} represent number of frames for 4 seconds, which is the pre buffer size.

Interleaving Parameter	Level 3 Audio	Level 2 IDR pics	Level 1 ref. P slices	Level 0 non ref P slices
S_{PBS} (in frames)	16	4	24	24
Frame Limit L (in msec)	5000	5000	1500	1000

Tab. 1 – Interleaving Parameter

For the case of no adaptation, Fig. 4 shows that during the link outage the client buffers are running out and the network buffer is over flowing. It takes some time after the link outage to recover. The network buffer can compensate the link outage to some extend meaning that not all data are lost, but the playout must stop for a while and the over flowed network buffer causes uncontrolled losses in the audio and video stream. Such losses can cause video quality reduction in other pictures as well, if lost pictures are used as reference for motion compensation.

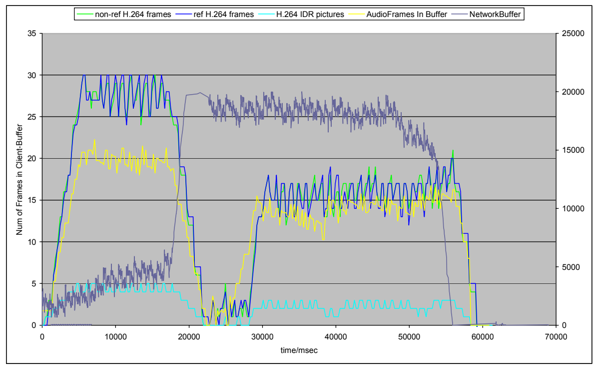

Fig. 4 – Link outage of 5 s without Protection

For the case of TRC and A/V Interleaving, Fig. 5 shows that during the link outage and after that there are still audio frames and video IDR pictures in the buffer, which can be played out. No picture of these priority levels is lost or too late. It looks like that the IDR buffer is empty for a while, but the incoming IDR pictures are consumed immediately. A few seconds after the link outage the next lower video buffer level, the reference (‘P’) pictures, is up again and can be played out, at last the buffer for non-reference (‘p’) pictures is refilled.

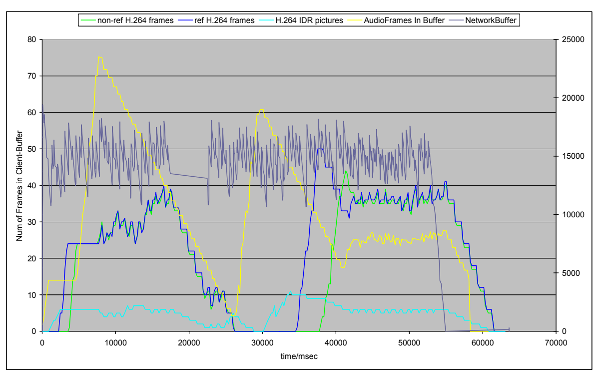

Fig. 5 – Link outage of 5 s with A/V Interleaving and TRC

Fig. 6 shows an example where at least no ‘slide show’ can be played out. Some pictures are too late. Even the audio buffer is running out for a moment. This example should show that it could be complex to determine all parameters for the interleaving algorithm in a way that not more pre buffering time is used than for EDF and the system works for a wide range of link outages. But systems like 3G networks allow a worst case calculation for the maximum link outage length. And all A/V interleaving parameters should be adjusted to that maximum.

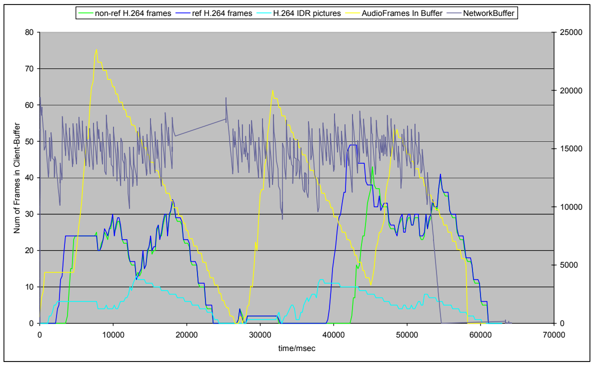

Fig. 6 – Link outage of 8 s with A/V Interleaving and TRC

The results are summarized in Tab. 2.

Loss / Frames	Level 3	Level 2	Level 1	Level 0
link outage 5 s No Protection	56	4	19	18
link outage 5 s A/V Interleaving	0	0	61	67
link outage 8 s A/V Interleaving	0	0	95	102

Tab.2 – Number of lost frames for the 3 experiments for each priority level.

4. CONCLUSION

We have presented a streaming system that utilizes interleaved transmission for real-time H.264/AVC video and AAC LC audio in 3G wireless environments. We have considered audio to be of highest priority. H.264/AVC video is split into 3 priority classes: IDR pictures as highest priority for video followed by reference P pictures and non-reference P pictures. Interleaved transmission is carried out using priority based scheduling with client feedback about the current fill level of each priority class in the client buffer. Experiments are conducted with a testbed that emulates a 3G network including block error rates on the physical layer, a buffer for retransmission on the link layer for different error rates, and link outages. The experimental results demonstrate the superior performance of interleaving. The approach outperforms plain Earliest Deadline First systems. Especially the benefits of using such a system approach for getting over link outages have been shown. But such a system can also be used for media rate adaptation, since it has an implicit Bitstream Thinning behavior.

4. REFERENCES

[1] I. Elsen, F. Hartung, U.Horn, M. Kampmann, L. Peters, “Streaming Technology in 3G Mobile Communication Systems“, *IEEE Computer*, September 2001, pp. 46-52.

[2] 3GPP - TSG-SA4 PSM SWG internal working draft, “Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs (Release 6)”, November 2004.

[3] ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC, “Advanced Video Coding for Generic Audiovisual Services”, 2003

[4] S. Wenger, M. Hannuksela, T. Stockhammer, M. Westerland, D. Singer, “RTP payload Format for JVT Video”, IETF Draft, August 2004

[5] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, “RTP: A Transport Protocol for Real Time Applications”, RFC 3550, July 2003.

[6] J. Ott, S. Wenger, N. Sato, C. Burmeister, J. Rey, “Extended RTP Profile for RTCP-based Feedback (RTP/AVPF)”, IETF Draft, January 2004.

[7] N. Baldo, U. Horn, M. Kampmann, F. Hartung, “RTCP Feedback Based Transmission Rate Control For 3G Wireless Multimedia Streaming”, *PIMRC 2004*, Barcelona, Spain, September 2004.

[8] J. Kritzner, U. Horn, M. Kampmann, J. Sachs, “Priority Based Packet Scheduling with Tunable Reliability for Wireless Streaming”, *HSNMC 2004*, Toulouse, France, July 2004.e.