3GPP Compliant Adaptive Wireless Video Streaming Using H.264/AVC

Thomas Schierl and Thomas Wiegand

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

Markus Kampmann

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

Abstract—**In 3GPP a standard for multimedia streaming over mobile networks called PSS (Packet-switched Streaming Service) has been specified. In order to deal with changing transmission characteristics like varying link bit-rate or handover, adaptive streaming features like client feedback have been included into Release 6 of 3GPP PSS. The adaptive behavior of the streaming server is not specified in PSS. In this paper, an adaptive video streaming system is presented that is compliant with PSS. Using PSS client feedback, transmission characteristics and client buffer status are determined. By means of transmission rate control, packet scheduling, video thinning, and bit-stream switching, the streaming server reacts to the determined transmission characteristics and the client buffer status. As video coding standard, H.264/AVC is used in the adaptive streaming system. Experimental results show that using the presented adaptive streaming system the overall video quality at the client is significantly improved in case of varying link bit-rate or handover.**

3G Wireless Streaming, H.264/AVC Video Rate Adaptation, Audio/Video Frame Interleaving, RTP NACK based Retransmission.

I. Introduction

Streaming services are increasingly becoming important applications for mobile networks. However, provision of sufficient audio and video quality at the mobile user's receiving side is still a challenge due to changing transmission characteristics like varying downlink bit-rates, latency, or handover cases. In this paper, we present a streaming system approach, which is based on Release 6 of the recently standardized 3GPP (3rd Generation Partnership Project) streaming standard PSS [1]. In PSS Release 6, adaptive streaming features like client feedback have been included. The adaptive streaming behavior of the streaming server is not specified in PSS. In order to achieve a better streaming quality in wireless environments, the following features and extensions are proposed:

A. RTCP Feedback including Buffer Reports

B. Transmission Rate Control especially designed for wireless networks

C. Interleaved transmission order of audio and video

D. Media Rate Adaptation with

    a. Bit-stream Switching

    b. Bit-stream Thinning

E. NACK based retransmission

While features A and E are specified within the 3GPP PSS streaming standard [1] as they need protocol support, the extensions B, C, and D are additional features of the streaming server.
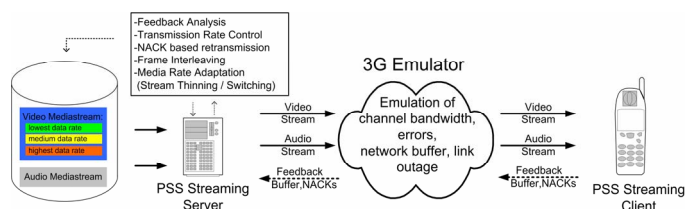

Fig. 1 – Adaptive 3GPP Wireless Streaming Testbed

An overview of the proposed 3G streaming system is shown in Fig. 1. Audio is coded using the MPEG-4 Advanced Audio Coding Low Complexity (AAC LC) Profile. Video is coded based on the H.264/AVC standard [2][3]. The investigation of the behavior of H.264/AVC in a wireless streaming scenario is another important topic of this paper.

The rest of the paper is organized as follows. Section 2 gives a detailed description of the used and developed adaptive streaming techniques. Section 3 describes the implemented system and in Section 4 results are presented.

II. Adaptive Streaming Techniques

The following adaptation techniques are extended or modified in order to work well in a 3G wireless streaming system. All techniques are based on feedback from the client to the server or estimation of either the transmission characteristics or client status. As feedback mechanism, RTCP Receiver Reports (RR) [4] with the buffer report extension specified in [1] and NACK reports as specified in [5] are used. The Highest Received

Sequence Number (HRSN) and the Next Sequence Number to be decoded (NSN) are exploited for determining the buffer levels of the network and the client.

The most important adaptation technique that is deployed in our system is probably *Transmission Rate Control* (TRC). We use a simple approach to avoid overfilling of network as well as client buffers. Such an approach has been proposed in [6]. By recalculating the current buffer levels of the network buffer N_{curr} and the 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.
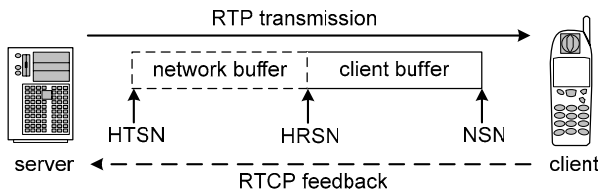

Fig. 2 - Reported Sequence Numbers

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 [1]. The server is sending a RTP packet i, if both of the two following equations are valid:

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

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

The next adaptation technique presented is *Audio/Video Frame Interleaving*, which is based on an approach proposed in [7]. The approach extends the well known Priority Based Scheduling (PBS) with using feedback for determining the current priority buffer levels at the client and using the Earliest Deadline First (EDF) algorithm after having a certain amount of priority pictures in the client buffer. We have extended this approach by using it together with H.264/AVC Picture Interleaving. As first video media type H.264/AVC allows an interleaved transmission within the RTP payload format [8], the so-called 'packetization mode 2'. The RTP payload format from [8] is included in the PSS standard [1]. The algorithm utilizes two parameters per priority level. The first parameter limits the pictures transmitted per priority level and gives the maximum amount of pictures desired in the client buffer (we call this parameter MaxLevel$_i$ for Priority Level i). The second parameter gives the limit of media data in terms of time, which is desired in the client buffer. After this limit has been reached the next lower priority level is filled up. If all levels have reached this TimeLimit$_i$, the levels will be filled up to MaxLevel$_i$ in Earliest Deadline First (EDF) manner, i.e. transmission in decoding order. We use the reported HRSN and NSN in order to recalculate the buffer status and with that the

the time limits. Further we use lost packet sequence numbers reported by NACKs in order to correct these limits. The algorithm is also shown in Fig 3, where (A) stands for audio frames, the rest of the frame types are H.264/AVC pictures: IDR, referenced (P) and non referenced (p) pictures. The algorithm starts with the transmission of the highest level.
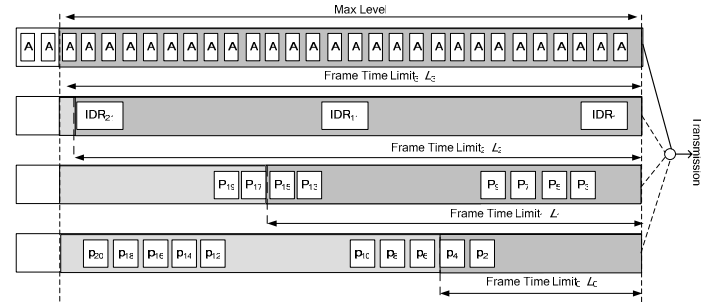

Fig. 3 – Audio/Video Frame Interleaving

The A/V Interleaving has two benefits. In case of link outage higher pre-buffers for the higher priority levels allow to continue playout with these higher levels, e.g. the full audio stream and additional a slide show of IDR/I pictures. When using traditional EDF, the frequent playout will interrupt earlier. The second benefit is an implicit Bit-stream Thinning algorithm, see also next section. If the channel does not provide the required transmission rate, the algorithm will transmit only pictures of high priority levels.

The *H.264/AVC Baseline Profile* [2][3], which is specified in PSS [1], allows the use of pictures that are not referenced for motion compensation of later pictures, so-called non referenced pictures. The dropping of such pictures is not affecting the decoding process of other pictures of the stream. Each lower Reference Layer, which is also dropped, does not influence the decoding process of higher layer pictures either. This is also known as Temporal Scalability; see [9] for more detailed information. Non referenced P Slices are marked as 'p' and referenced P Slices are marked as 'P'. Note that 'P' and 'p' slices cannot be mixed within a picture [2].

In the following, we describe the combination of *Bit-Stream Switching* (BSS) and *Temporal Scalability* (TS). An adaptation approach using both techniques combined has been proposed in [9]. This approach exploits the fact, that switching between H.264/AVC bit-streams with the same encoding parameters except different Quantization Parameters [2] can be fulfilled at IDR pictures with no additional quality loss caused by the switching. It combines the switching approach with the TS approach.
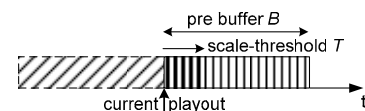

Fig. 4 – Scaling thresholds

The algorithm starts scaling the stream quality down, when the reported buffer level B is below a threshold T. In such a case the system is switching down immediately to a lower data rate track with a rate R_x,

$$\max(R_x) \text{ with } R_x < R \qquad (3)$$

where R is the current estimated receiving rate at the client. Immediate means that switching down is performed at any type of picture. We found that switching down from a higher to a lower QP does not produce any significant artifacts. The rate is calculated dependent on the incoming RTCP RRs [4] and the lost packets reported by NACKs, which could not be retransmitted within the playout timing constraints. An average of the receiving rate can be calculated using the reported HRSN. Here we are assuming that the RRs are sent in a fixed time interval. The exact receiving time at the client can be recalculated using timing information of the RR reporting HRSN, like 'delay since last Sender Report (SR)' d_{HRSN}, which is the time difference between sending the current RR and receiving the last RTCP SR, which has the 'last SR timestamp' t_{SR}. SRs are sent in a fixed time interval. There, the receiving rate at the client is determined according to

$$t_{recv}(HRSN) = t_{SR} + d_{HRSN} \qquad (4)$$

$$R = \frac{\sum\limits_{i=HRSN_k}^{HRSN_l} PktSize_i - \sum\limits_{j=HRSN_k}^{HRSN_l} LostPktSize_j}{t_{recv}(HRSN_l) - t_{recv}(HRSN_k)} \qquad (5)$$

When a track with the requirements of (3) does not exist, the system starts dropping 'H.264/AVC Reference Layers' as proposed in [9]. If the network condition is getting better, i.e., buffer B is again above scaling threshold T, then the server switches to a track with a bit-rate according to (3). The system independently probes for more bit-rate available through the TRC algorithm.

III. EXPERIMENTAL RESULTS

The testbed as shown in Fig. 1 was implemented . 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. Further channel parameters that can be adjusted are PDU size, Transmission Time Interval (TTI) and PDUs per TTI. The client's H.264/AVC decoder is capable of error concealment as proposed in [10]. We have configured the channel with typical parameters in a WCDMA network. We have conducted two experiments as follows:

A. Link Outage Experimental Settings
Link conditions: bit-rate of 64 kbit/s; 5 % BLER, due to BLER 56 kbit/s is available;

Bit-rate profile: 18 s link up with 56 kbit/s → 5 s link outage → 56 kbit/s link is up again until end of simulation.

B. Channel Rate Switching Experimental Settings
Link conditions: bit-rates of 64 and 128 kbit/s, 5 % BLER, due to BLER 56kbit/s and 117 kbit/s are available, respectively;

Bit-rate profile: for 15 s 56 kbit/s → for 20 s 117 kbit/s → until end of simulation 56 kbit/s.

We have used a mobile network buffer of 20 kBytes, which can compensate link outages and data rate peaks. The used H.264/AVC bit-streams are encoded at 45 kbit/s for the link outage experiment and 90/80/56/32 kbit/s for the channel switch experiment. The video resolution is QCIF at 12.5 fps with an IDR interval of 1.04 s (each 13th picture) for the link outage experiment and an IDR interval of 2 s (each 25th picture) for the channel rate switching experiment with a "IDRpPpP.." GOP structure for both experiments. The AAC LC audio stream is coded at about 10 kbit/s for the link outage experiment and coded at about 18 kbit/s for the channel switch experiment. The overall stream length is about 53 s. For both experiments we have used 4 s of pre-buffering period.. The maximum allowed client buffer size is 50000 Bytes.

A. Link Outage Experimental Results
The link outage experiment is conducted with a streaming system without any adaptation and a streaming system doing TRC and A/V Interleaving.
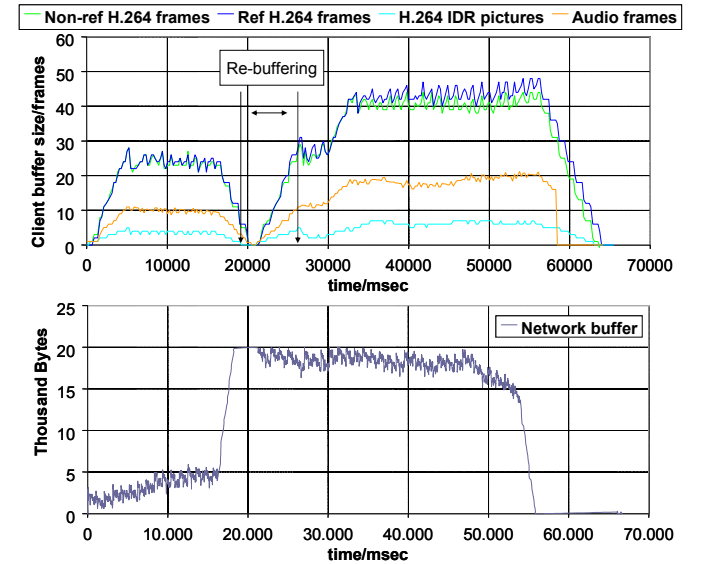

Fig 5 – Link Outage without adaptation

Fig. 5 shows the client buffer fullness and network buffer evolution for the case without adaptation. The colors of the graphs correspond as follows: bytes buffered in the network (grey) and data stored in the client buffer are audio frames (orange), H.264/AVC IDR frames (light blue), referenced P pictures (dark blue) and non-referenced P pictures (green). During the first 20 s the network buffer is slightly growing since stream bit-rate is slightly higher than the channel rate in this time period. The link outage causes a network buffer overflow. Fig. 5 further shows the number of media frames streamed into the client buffer growing simultaneously with decoding order. After the link outage a re buffering is required. It is important to notice that the network buffer overflow caused additional packet losses, while re-buffering and thus an additional quality reduction can be noticed.

Fig. 6 shows the case with TRC and A/V interleaving. Since TRC avoids overflows of the network buffer, we do not have to cope with packet loss here. It can be seen that A/V interleaving provides a certain amount of high priority level pictures in the

client buffer. During the link outage, the non-reference picture buffer (p) and the reference picture (P) buffer are running out. But there are still data for audio play out and a slide show using IDR pictures in the buffer. After the link is up again the algorithm ensures that the buffers are filled up again in prioritized order. First the IDR buffer is brought to a sufficient level, after that the reference picture (P) buffer and last the non reference picture (p) buffer.
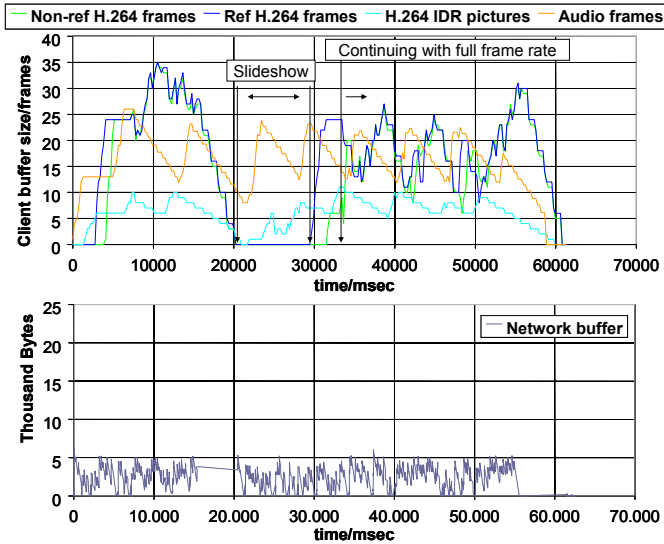

Fig. 6 – Link Outage with TRC + A/V Interleaving

B. Channel Rate Switching Experimental Results

The channel switch experiment is conducted with a streaming system without any adaptation and a streaming system doing TRC, BSS, and TS. The channel switch without using adaptation causes the client buffer running out (Fig. 7) and an overbuffering of the network buffer causing packet loss. The adaptive system's performance is shown in Fig. 8, where it can be seen that the system switches down to the optimal data rate track in case of channel switching. After the channel up switch the system probes for new bit-rate available and goes back to the highest data rate track.
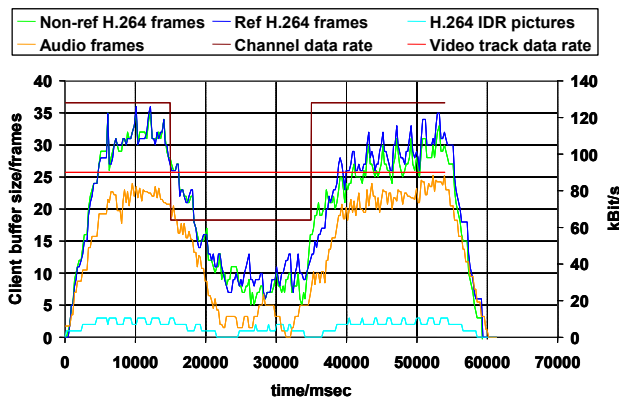

Fig 7 – Channel Switch without adaptation

IV. CONCLUSION

We have presented a streaming system for mobile networks using the lately defined streaming features of the streaming standard PSS Release 6 of 3GPP combined with enhanced adaptation techniques of the streaming server, which are left over by the PSS standard. We have shown the benefits of the system in typical wireless network scenarios such as like link outage and channel rate switching. The advantage of using Transmission Rate Control in combination with media adaptation techniques like Audio/Video Interleaving and Bit-stream Thinning / Switching has been pointed out.
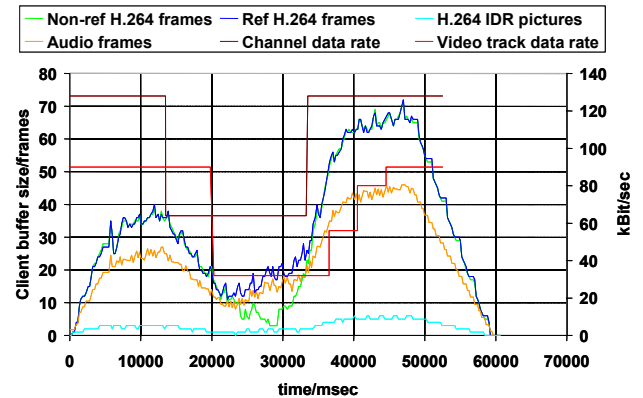

Fig 8 – Channel Switch with TRC + Stream Switching

REFERENCES

[1] 3GPP - TSG-SA4 PSM SWG internal working draft, "Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs (Release 6)", Nov. 2004.

[2] ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC, "Advanced Video Coding for Generic Audiovisual Services", (version 1: 2003, versions 2: 2004) version 3: 2005.

[3] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra: "Overview of the H.264/AVC video coding standard", IEEE Trans. CSVT, vol. 13, no. 7, pp. 560-576, July 2003.

[4] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson: "RTP: A Transport Protocol for Real Time Applications", RFC 3550, Internet Engineering Task Force, July 2003.

[5] J. Ott, S. Wenger, N. Sato, C. Burmeister, J. Rey: "Extended RTP Profile for RTCP-based Feedback", Internet draft, Internet Engineering Task Force, August 2004.

[6] N. Baldo, U. Horn, M. Kampmann, and F. Hartung: "RTCP Feedback Based Transmission Rate Control For 3G Wireless Multimedia Streaming", PIMRC, September 2004.

[7] J. Kritzner, U. Horn, M. Kampmann, and J. Sachs: "Priority Based Packet Scheduling with Tunable Reliability for Wireless Streaming", HSNMC, July 2004

[8] S. Wenger, M. Hannuksela, T. Stockhammer, M. Westerlund, and D. Singer: "RTP payload Format forH.264 Video", RFC 3984, Internet Engineering Task Force, February 2004

[9] T. Schierl and T. Wiegand: "H.264/AVC Rate Adaptation For Internet Streaming," Packet Video Workshop, December 2004

[10] T. Stockhammer, M. M. Hannuksela, and T. Wiegand: "H.264/AVC in Wireless Environments", IEEE Trans. CSVT, vol. 13, no. 7, pp. 657-673, July 2003.