

ADVANCED WIRELESS MULTIUSER VIDEO STREAMING USING THE SCALABLE VIDEO CODING EXTENSIONS OF H.264/MPEG4-AVC

Günther Liebl¹, Thomas Schierl², Thomas Wiegand², and Thomas Stockhammer³

¹ Lehrstuhl für Nachrichtentechnik
TU München
80290 München, Germany
liebl@tum.de

² Fraunhofer – Heinrich Hertz Institute,
Einsteinufer 37,
10587 Berlin, Germany
schierl@hhi.fhg.de

³ Nomor Research
Tannenweg 25
83346 Bergen, Germany
stockhammer@nomor.de

ABSTRACT

We propose a dynamic sharing of radio resources in a wireless multiuser system by combining a flexible video coding scheme, namely the currently developed scalable video coding (SVC) extension of H.264/AVC with appropriate radio link buffer management for multiuser streaming services. The paper introduces the components of the proposed system and the used cross-layer interface. Simulation results verify the benefits of our approach.

1. INTRODUCTION

With increasing popularity of wireless video streaming services, the need for higher capacity radio access networks is obvious. Recent extensions of existing 2G and 3G networks, such as EGPRS, HSDPA, and IEEE 802.1x technologies, allow exploiting multiuser diversity for enhanced system capacity. However, these technologies require new algorithms in the base station, as the incoming packets of different media flows need to be scheduled appropriately. Especially for real-time multimedia transmission, it is important that the delay in the network is not too high for good user experience. One possible approach is the definition of quality-of-service (QoS) attributes for each media flow. The radio access network then attempts to maintain the QoS by appropriate admission control and resource reservation. However, this approach has several drawbacks including that some users might not be able to access the service due to admission control, statistical multiplexing gains cannot be exploited, multiuser diversity is not fully exploited, and the flexibility and bit-rate adaptivity of applications is ignored.

In this work we propose a dynamic sharing of the resources by combining a fully flexible video coding scheme, namely SVC, with appropriate radio link buffer management for multiuser streaming services. In the remainder of this paper we will explain the various system components and present selected simulation results showing the benefits over H.264/MPEG4-AVC compatible solutions as initially introduced in [5].

2. SCALABLE VIDEO CODING AND MEDIA ADAPTATION

2.1 Scalable Video Coding (SVC)

SVC [1] is an extension of H.264/MPEG4-AVC [2]. The basic idea is to extend the hybrid video coding approach in a way that a wide range of spatio-temporal and quality scalability is achieved. Scalability within SVC is a functionality that allows the removal of parts of the bitstream while achieving a reasonable coding efficiency of the decoded video at reduced temporal, SNR, or spatial resolution. An SVC bitstream consists of a base layer and one or several enhancement layers. The base-layer is an H.264/MPEG4-AVC bitstream ensuring backwards compatibility.

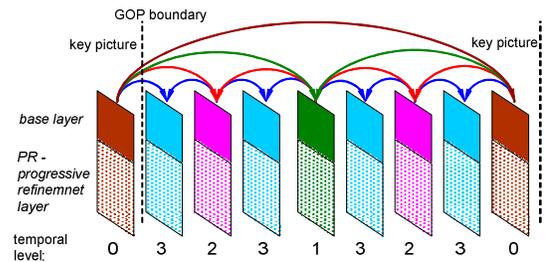


Fig. 1 - Temporal structure of an SVC stream incl. PR

SVC's temporal scaling functionality is often based on a temporal decomposition using hierarchical B pictures. For optimized compression efficiency, each B picture of a higher temporal extension level should be encoded with a higher QP (cascaded QP assignment). For details, see [3]. In Fig. 1 the structure of an SVC stream with a group of pictures (GOP) size of 8 is shown. GOPs can be independently decoded, if the preceding key picture is available. Those key pictures can be used as random access points, if they are intra coded.

Spatial scalability is achieved by different encoder loops with an over-sampled pyramid for each resolution (e.g. QCIF, CIF, and 4CIF), including hybrid video coding with independent motion-compensated prediction (MCP) structures for each layer. In contrast to the encoder, the decoder

is operated with a MCP single loop. Hence, for inter-layer dependency MCP of lower layers is not needed.

SNR scalability is either based on layered coding or progressive refinement (PR) coding. For PR coding, the extension layers contain refinement bits relative to the lower layer. Thus, cutting byte-wise from the end of an PR bitstream is possible. A PR layer only contains refinements for the residual (texture) data, which is also used for prediction in next higher temporal levels. Up to three PR layers can exist in a stream. Typically the quality of each layer is enhanced by a QP delta value of 6. Finally, each layer of a each slice is stored in a separate Network Abstraction Layer (NAL) Unit, which can be transport individually. More details will be discussed later.

2.2 Adaptation and Transport of SVC

The hierarchical B pictures and the PR coding approach are combined so that the resulting priority order of the bitstream is as follows: The priority scale is starting from the lowest temporal layer, which is the most important for decoding, since more pictures are depending on lower temporal levels. Then the respective PR fragments have next lower importance, and so on. Hence, by assuming that the temporal resolution has highest priority we recommend the following simplified *dropping order* in case that a rate adaptation on the fly is necessary:

1. Drop PR fragments of the highest temporal level present.
2. Drop base layer of highest temporal level present.

Although the encoder could apply optimized encoding strategies if the dropping strategies are known in advance, we only apply the above generic priority order for our bitstream rate adaptation. The interface between video coding layer and network is realized by the SVC NAL Unit header [4], which is an extension of the H.264/MPEG4-AVC NAL unit to enable priority signaling across layers.

Bit	0	1	2	3	4	5	6	7	
1. Byte	F		NRI		NAL Unit Type				
2. Byte	Simple Priority ID					D	E		
3. Byte	Temporal Level			Dependency ID		Quality Level			

Fig. 2 - SVC Extended NAL Unit Header

The SVC NAL Unit header as shown in Fig. 2 consists of the H.264/MPEG-4 AVC compliant NAL Unit header (first byte) and (a maximum of) two extension bytes, where

- Byte 1 contains Forbidden Field (F), NAL Unit Reference Indicator (NRI), and NAL Unit Type with syntax and semantics as in H.264/MPEG4-AVC except for using NAL Unit Types 20 and 21 for SVC NAL units;
- Byte 2 signals Simple Priority ID (a mapping of importance specified by the encoder), discardable flag (D), and Extended Bit (E) indicating the presence of Byte 3;
- Byte 3 specifies NAL unit assignment to Temporal, Spatial, and Quality Level.

Typically, PR NAL Units are larger than the transport layer MTU size. To circumvent this problem we apply fragmentation of the NAL Unit by extending the Fragmentation Unit (FU) introduced in [4]: We use the full 3-byte header for each RTP packet and differentiate the importance of the contained FGS refinement data via the Simple Priority ID field (where '0' indicates highest importance, consistent with the definition of the Quality Level field).

2.3 Priority Labeling

Compared to plain H.264/MPEG4-AVC bitstreams, which only offer temporal scalability for coarse bitstream rate adaptation in wireless systems [5], SVC bitstreams offer a wide range of options for rate adaptation. However, whereas for H.264/MPEG4-AVC the NRI value and the NAL unit type are typically sufficient to extract all relevant information inside media aware network elements (MANEs), services can significantly benefit from an extended priority labeling of SVC packets being available within MANEs. Based on the semantics of the individual SVC NAL unit header fields described in section 2.2, we propose the following labeling: Let $b_{i,j}$ denote the j -th bit in the i -th byte of the header. Thus, by combining the 2 bits of the Quality Level and the 6 bits of the Simple Priority ID field, we define the following intermediate priority value

$$\tilde{p}_r \triangleq \sum_{i=6}^7 b_{3,i} \cdot 2^{(13-i)} + \sum_{i=0}^5 b_{2,i} \cdot 2^{(5-i)} \Rightarrow 0 \leq \tilde{p}_r \leq 255.$$

Hence, the higher the importance of the PR fragment included in a packet, the lower the value of \tilde{p}_r . In order to combine this with existing solutions for temporal scalability, we further distinguish base layer fragments (those with $\tilde{p}_r = 0$) via the NRI field. Let

$$p_{NRI} \triangleq \sum_{i=1}^2 b_{1,i} \cdot 2^{(2-i)} \Rightarrow 0 \leq p_{NRI} \leq 3.$$

Then the final priority value for each packet is defined as

$$p_r \triangleq \begin{cases} 255 - \tilde{p}_r + p_{NRI}, & \text{if } \tilde{p}_r = 0, \\ 255 - \tilde{p}_r, & \text{if } \tilde{p}_r > 0. \end{cases}$$

Note that we have "inverted" the intermediate priority value to match the definition of the NRI value (increasing values of p_r indicate higher importance).

3. RADIO LINK BUFFER MANAGEMENT FOR SCALABLE VIDEO STREAMS

3.1 Wireless Multiuser Streaming Environment

In this work we consider a similar wireless multiuser streaming environment as in [5]: In total M users in the coverage area of a base station receive streaming multimedia data from a server. We assume that the core network is over-provisioned such that congestion on the backbone is not an issue. The streaming server forwards the NAL units encapsulated in RTP packets [4] directly into the *radio link buffers* at the base station, where they are kept until transmission to the media clients over the shared wireless link. For each

radio access slot a scheduler decides which users can access the wireless system resources, and a resource allocation unit assigns them appropriately.

If the radio link buffers are not served fast enough because of bad scheduler decisions or too many streams are competing for the common resources, the system is in *overload* and typical congestion problems arise. In [5] it has been shown that for real-time applications it is beneficial to keep the buffer size N_{RL} finite and drop data units already at the radio link buffer to reduce the excess load and avoid late-loss at the media client.

3.2 Drop Strategy

For H.264/MPEG4-AVC the 2-bit NRI header field and the NAL unit type differentiation between single slice and IDR (to determine the GOP structure) have been shown to be sufficient for an efficient drop strategy (called **DDB** in [5]). For SVC, we make use of the extended priority labeling as introduced in section 2.3 and modify the radio link buffer management strategy as follows:

1. If there are still packets which contain PR fragments (i.e. with $p_r \leq 254$) in the buffer: Drop the one with lowest priority which resides longest in the buffer.
2. If there are only packets left which contain base layer fragments (i.e. with $p_r > 254$): Proceed as described for DDB in [5], i.e. first drop all base layer fragments with no further dependencies (i.e. with $p_r = 255$) starting from the beginning of the Head-of-Line (HOL) GOP. If necessary, we propose that dropping shall continue for the remaining fragments from the end of the GOP.

This strategy can be applied to both H.264/MPEG4-AVC and SVC streams, if for H.264/MPEG4-AVC packets the default value $\tilde{p}_r = 0$ is used.

4. SIMULATION AND RESULTS

For the simulation results, both the SVC and the H.264/MPEG4-AVC bitstream have been encoded with JSVM4 [6]. We used a looped Foreman sequence of 300 pictures (10 sec) at CIF resolution encoded at 30 Hz. For both streams we apply a GOP size of 16 pictures and an intra rate of 0.5 Hz. The SVC stream has an H.264/MPEG4-AVC base layer at 160 kbit/s and two PR layers with an overall bit-rate of about 390 kbit/s, which is the same bit-rate as the H.264/MPEG4-AVC anchor. The encoder Y-PSNR is 36.42 dB for SVC and 36.97 dB for H.264/MPEG4-AVC.

The wireless multiuser scenario contains a model of a HSDPA system (including fast fading and shadowing on the mobile radio channel) with identical parameters as in [5], but only $M = 4$ streaming users are connected to the base station. The average channel quality (SNIR) of each user is given in the first row of Tab.1. The scheduling strategy applied at the air interface is maximum throughput. We use pre-buffering of 4 GOPs corresponding to 2.13 sec of the

stream. The size of each radio link buffer is restricted to $N_{RL} = 110$ KBytes.

Tab.1 shows selected simulation results: the overall playable picture rate of the medium (channel) quality user 1 is significantly increased in case of SVC when compared to H.264/MPEG4-AVC. This is due to the fact that the buffer load of this user can be reduced to the most important base layer fragments to achieve continuous playout. Furthermore, priority-based dropping of PR layers results in a smoother variation in the PSNR over the duration of the stream, as can be observed from Figs. 3a and 4a. The performance of the worst quality user 4 is also increased in case of SVC. However, this user is still not able to continuously receive the full base layer, and temporal scalability is required to perceive at least a "slide show" at the media client (see Figs. 3c and 4c). The improvements of both user 1 and 4 are not achieved at the expense of a deterioration of the high (channel) quality users 3 and 5: The latter maintain or even slightly improve the smoothness of their PSNR curve (Figs.3b,d and 4b,d) when changing from H.264/MPEG4-AVC to SVC bit-streams (the slight decrease in PSNR is due to the general scalability loss).

Tab.1-Simulation Parameters/Results

	User 1	User 3	User 4	User 5
Channel SNIR	12.9 dB	17.9 dB	9.7 dB	18.0 dB
H.264 Y-PSNR	30.1 dB	36.7 dB	15.4 dB	37.0 dB
SVC Y-PSNR	34.9 dB	36.1 dB	24.2 dB	36.4 dB
H.264 Pic rate	19.7 Hz	29.5 Hz	0.8 Hz	30.0 Hz
SVC Pic rate	30.0 Hz	29.5 Hz	13.5 Hz	30.0 Hz

5. CONCLUSION

We have shown that SVC allows a straightforward extension of existing radio link buffer management strategies to further improve performance of video streaming over wireless shared channels. Due to the large gains achievable, we propose to add suitable packet priority labeling in MANEs in wireless multiuser systems. Future research targets the simplification of priority computation in MANEs and the usage of these priorities in scheduling processes.

REFERENCES

- [1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable H.264/MPEG4-AVC extension", ICIP, Atlanta, GA, USA, Oct. 2006.
- [2] ITU-T Recommendation H.264 & ISO/IEC 14496-10 AVC, "Advanced Video Coding for Generic Audiovisual Services", v3: 2005.
- [3] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B pictures and MCTF", ICME, Toronto, Canada, Jul. 2006.
- [4] S. Wenger, M. Hannuksela, T. Stockhammer, M. Westerlund, and D. Singer, "RTP payload Format for H.264 Video", RFC 3984, Internet Engineering Task Force, Feb. 2004.
- [5] G. Liebl, H. Jenkac, T. Stockhammer, and C. Buchner, "Radio Link Buffer Management and Scheduling for Wireless Video Streaming", Telecommunication Systems, vol. 30/1-3, Nov. 2005, Springer Science & Business Media B.V., pp. 255-277.
- [6] JSVM 4 Reference Software, JVT-Q203, Nizza, France, Oct. 2005.

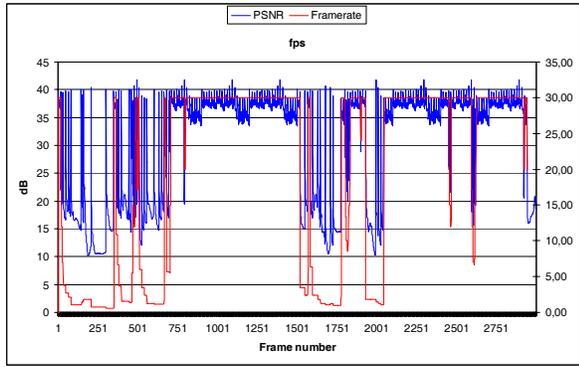


Fig.3a-AVC User 1

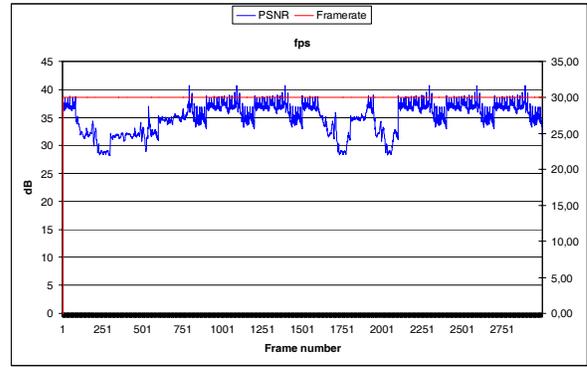


Fig.4a-SVC User 1

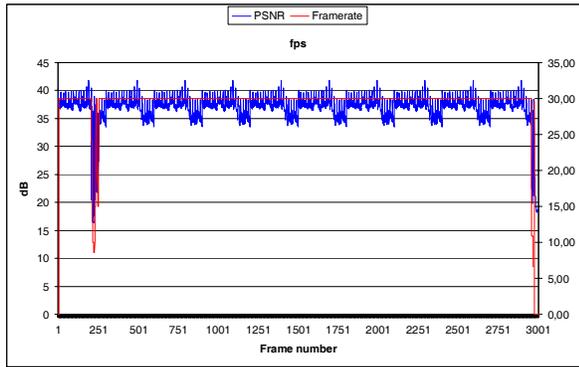


Fig.3b-AVC User 3

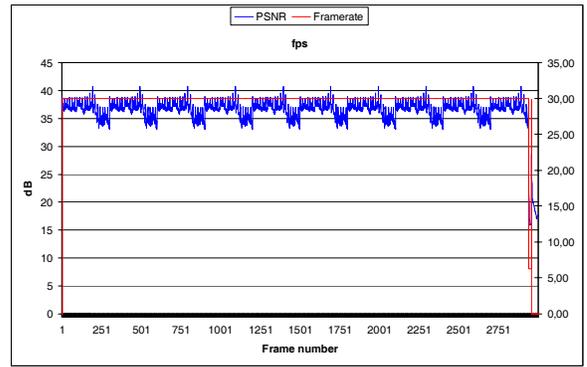


Fig.4b-SVC User 3

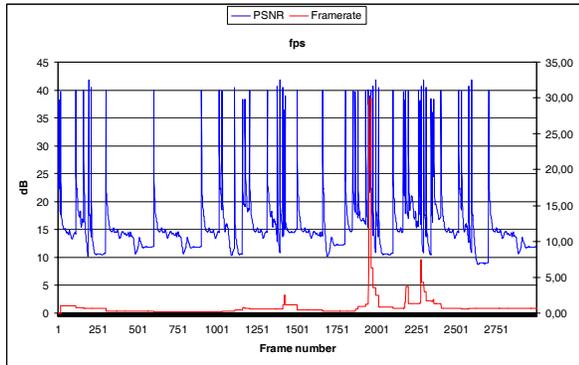


Fig.3c-AVC User 4

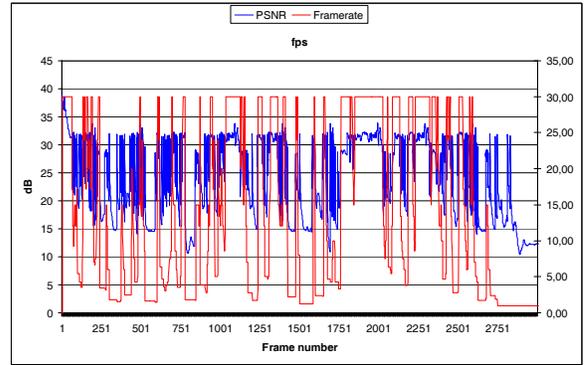


Fig.4c-SVC User 4

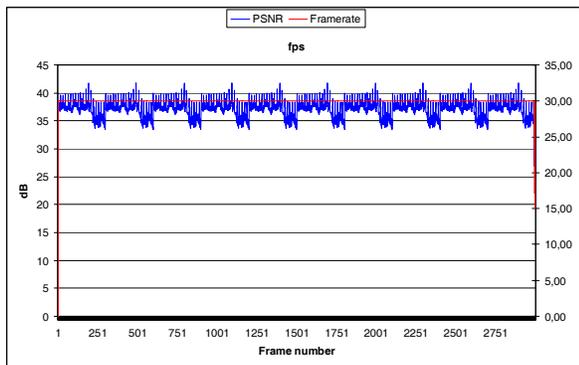


Fig.3d-AVC User 5

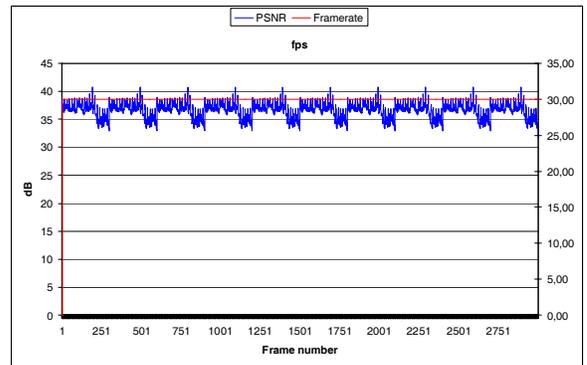


Fig.4d-SVC User 5