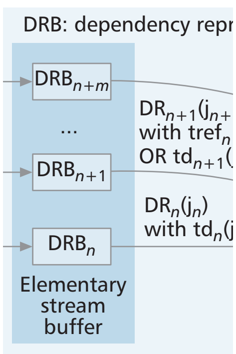
SCALABLE VIDEO CODING OVER RTP AND MPEG-2 TRANSPORT STREAM IN BROADCAST AND IPTV CHANNELS

THOMAS SCHIERL, KARSTEN GRÜNEBERG, AND THOMAS WIEGAND, FRAUNHOFER HHI



The authors give a detailed overview of the recently finished SVC standards on transport over IP/RTP and MPEG-2 Transport stream. Both standards are important for IPTV and video on demand.

ABSTRACT

The ITU-T and ISO/IEC standard for scalable video coding was recently finalized. SVC allows for scalability of the video bitstream in the temporal, spatial, or fidelity domain, or any combination of those. Video scalability may be used for different purposes, such as saving bandwidth when the same media content is required to be sent simultaneously on a broadcast medium at different resolutions to support heterogeneous devices, when unequal error protection shall be used for coverage extension in wireless broadcasting, as well as for rate shaping in IPTV environments. Furthermore, it may also be useful in layered multicast transmission over the Internet or peer-to-peer networks, or in any transmission scenario where prioritized transmission for network flows is meaningful. In order to make usage of SVC in the aforementioned use cases, standards for defining the transport format and procedure are required. Therefore, we give a detailed overview of the recently finished SVC standards on transport over IP/RTP and the MPEG-2 transport stream. Both standards are important for IPTV and video on demand, where the first is important for SVC transport over mobile broadcast/multicast channels, and the latter is also important for SVC transport over traditional digital broadcast channels.

INTRODUCTION

The most recent International Telecommunication Union — Telecommunication Standardization Sector (ITU-T) and International Standards Organization/International Electrotechnical Commission (ISO/IEC) standards for scalable video coding (SVC) [1] specify a video bitstream that is scalable in the temporal, spatial, or fidelity domain, or any combination of those, allowing bit rate reduction while still maintaining reasonable video quality. In order to make use of SVC in media delivery scenarios, standards specifying the transport of SVC over IP/Real-Time Transport Protocol (RTP) [2] and MPEG-2 transport stream [3] were recently finalized. Both standards are important for delivery of SVC over digital broadcast, wireless multicast, and IPTV channels.

Today's digital broadcast channels as specified by the Digital Video Broadcasting Project (DVB) or the Advanced Television Systems Committee (ATSC) rely on MPEG-2 systems [4] for encapsulation and signaling for media delivery. This is a remainder from the early MPEG-2 success in digital video broadcasting. Over the years new standards such as H.263, MPEG-4 visual Part 2, and H.264/MPEG-4 Advanced Video Coding (AVC) have been developed. Even with the new video coding standards, there was no requirement to update the TV delivery chain based on MPEG-2 Transport Stream (TS). Therefore, today's DVB IPTV services are mostly based on MPEG-2 TS delivered over IP, using RTP over the User Datagram Protocol (UDP) or UDP directly. This is the reason to keep both digital broadcast as well as IPTV compliant.

Looking at emerging wireless and mobile broadcast/multicast channels such as the DVB standard for satellite-to-handheld (SH), the DVB standard for next-generation handheld (NH), or 3GPP's multimedia broadcast multicast services (MBMS), the situation is different. These service types target mobile receiver devices, which typically have native IP interfaces and therefore preferably support native media transport over IP; that is, those standards rely on RTP and the Session Description Protocol (SDP) for packetization and signaling.

Recently, second tier standardization committees have already adopted the SVC transport standards as the DVB standards for media delivery over IP, and MPEG-2 TS and the ATSC standard for mobile/handheld are close to the final stage of adoption.

The work of the authors is supported by the European Commission under the contract number FP7-ICT-214063, project SEA.

 1536-1284/09/$25.00 © 2009 IEEE

Using any of the transport formats includes signaling as well as encapsulation and transport mechanisms. Since MPEG-2 TS and RTP comprise different signaling and encapsulation techniques, we highlight both transport methods and also mention the possible carriage of MPEG-2 TS over IP.

The remainder of the article is structured as follows. The next section gives a summary of the SVC standard and highlights important parts of the high level syntax of SVC important for encapsulation operations for SVC bitstreams. The following two sections give a detailed overview of transporting SVC over RTP and over MPEG-2 TS, respectively.

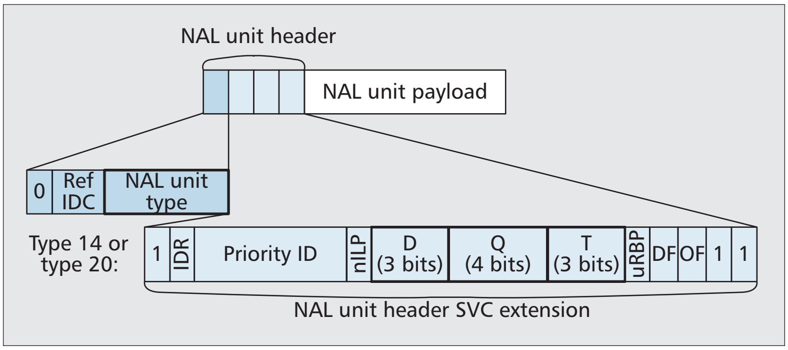


Figure 1. *SVC NAL Unit (top), AVC compatible header section (left, darker blue), and SVC header extension (right, lighter blue).*

SVC Overview

Coding

Video scalability, also known as layered video coding, has always been a desirable feature of a media bitstream. In contrast to earlier approaches included in video coding standards like H.262/MPEG-2 Video, H.263, and MPEG-4 Visual (Part 2), SVC [1] has achieved reasonable coding efficiency, which makes it applicable for a wide range of services. For a more detailed description of the SVC codec see [5], and for SVC use cases see [6].

The SVC design, which is an extension of the H.264/AVC video coding standard adding new profiles, can be classified as a layered video codec. In SVC the hybrid video coding approach of motion-compensated transform coding is extended in such a way that a wide range of spatial, temporal, and fidelity scalability is achieved. An SVC bitstream consists of a base layer and one or several enhancement layers. The removal of enhancement layers still leads to reasonable quality of the decoded video at reduced temporal or signal-to-noise ratio (SNR) fidelity, and/or spatial resolution. The base layer is a bitstream conforming to existing H.264/AVC profiles, ensuring backward compatibility with existing receivers.

High-Level Syntax

In this section we highlight the signaling provided in the SVC bitstream to allow high-level readability in general. Therefore, SVC as well as AVC comprise a so-called network abstraction layer (NAL) which acts as a packet interface to the system and transport layers.

SVC and AVC bitstreams consist of a sequence of NAL unit packets that can be identified by the NAL unit header, as shown in Fig. 1.

A NAL unit may carry different types of payloads:

- Video coding lLayer (VCL) information (e.g., entropy coded intra, residual, or motion data)
- Non-VCL information such as:
  - –Information about decoding process settings as parameter sets
  - –Additional supplemental enhancement information (SEI) messages not required by the decoding process

The payload content of a NAL unit is identified by the NAL unit type field in the 1-byte AVC NAL unit header section. For NAL units that contain SVC VCL data in the scalable extensions of AVC (NAL unit type = 20), three additional header bytes are used, as shown at the bottom of Fig. 1.

This NAL unit header SVC extension contains, among other information, three fields that specify the level with regard to the three scalability dimensions, here labeled D for dependency ID (spatial or coarse-grain quality scalability [CGS] level), Q for quality ID (medium-grain quality scalability [MGS] level), and T for temporal level ID (temporal scalability level). D, Q, and T identify for each SVC VCL NAL unit the specific level of scalability where D and Q are zero. For the AVC base layer, a special prefix NAL unit (NAL unit type = 14) carrying the 3-byte SVC extension header is inserted before each VCL NAL unit of the AVC base layer in order to indicate the temporal level T to which it belongs. For more details on SVC high-level syntax, we refer to [13].

SVC over RTP

This section gives an overview of the transport mechanisms for SVC over IP. Therefore, we discuss RTP and then go into detail about the RTP payload format for SVC. We then discuss the out-of-band signaling for SVC using SDP.

Media over IP: Real-Time Control Protocol

RTP is an application layer protocol that provides the means to transport real-time media data over IP using a variety of transport layer protocols such as UDP, Transmission Control Protocol (TCP), and Datagram Congestion Control Protocol (DCCP). RTP provides encapsulation for real-time media, which requires immediate transport as well as immediate consumption of the data at the receiver. This makes RTP suitable for different scenarios such as live broadcasting and media streaming on demand, as well as conversational services.

Pure RTP provides media encapsulation, media synchronization, and quality of service signaling and basic coordination for multi-party communication. Today's RTP has been fundamentally extended, providing error protection by forward error correction and retransmission, advanced signaling about the received media quality to control the media codec via codec control messages, as well as security. In the following we highlight the RTP details important

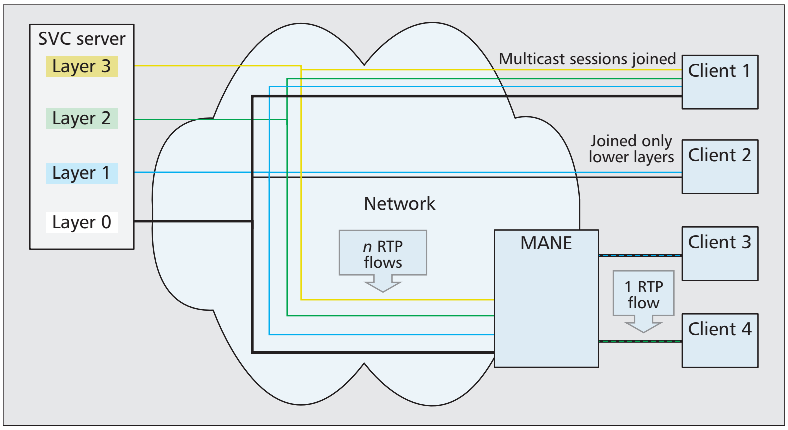


Figure 2. *SVC layered multicast + media-aware network elements (MANEs)*

for TV-like delivery of real-time media over broadcast/multicast as well as IPTV channels.

Besides media transport, the RTP protocol suite provides a second packet type, Real-Time Control Protocol (RTCP) packets. For IPTV, RTCP has two important functionalities: the provisioning of inter-RTP flow media synchronization and receiver feedback on connection statistics.

RTP defines different topologies. Besides typical point-to-point streaming scenarios, these topologies also include point-to-multipoint as well as multipoint-to-multipoint.

Due to the focus of this article, we chose the two examples shown in Fig. 2, where the upper part shows a layered multicast scenario (single source multicast), that is, the use of IP multicast by a single server where clients join sessions of a layered codec depending on their downlink capacity. Layered multicast as introduced by Steven McCanne can also be applied on transmission over a wireless broadcast channel, where different channel protection may be applied in different RTP sessions (containing exactly one RTP flow) through layers.

Assume a further extension of the layered multicast scenario, as shown in the bottom part of Fig. 2, where layered multicast is only applied in a dedicated network (e.g., the IPTV provider's network). At the last hop, which may be a digital subscriber line (DSL) access multiplexer, a media-aware network element (MANE) is present which combines layer of different RTP sessions (RTP flows) in order to build a single session containing the layers matching the link capacity constraints as well as the device capability.

In RTP terminology, a MANE could act in two different modes, first as an RTP mixer, which terminates incoming RTP sessions and establishes new RTP sessions to the client terminal while modifying the RTP payload, or second as an RTP translator, which keeps the end-to-end connection and signaling intact while modifying RTP packets as well as RTP payload content on the fly. The MANE shown in Fig. 2 acts as an RTP mixer.

MEDIA PACKETIZATION RTP (PAYLOAD)

Design of the RTP payload format for SVC video [2] is inherently based on the RTP payload format for AVC video [7]. All packet types, packetization modes, and signaling parameters of [7] are inherited.

The AVC RTP payload format [7] supports encapsulating a single NAL unit, more than one NAL unit, or a fragment of a NAL unit into one RTP packet. A single NAL unit as specified in H.264/AVC can be included in the RTP packet *as is*, and the NAL unit header serves as the payload header. Four types of aggregation NAL units are specified using values in the NAL unit type space that are undefined in SVC [1]. The two single-time aggregation packet types, STAP-A and STAP-B, allow encapsulating multiple NAL units that belong to the same picture (identified by identical RTP timestamps) into one RTP packet. The two multiple-time aggregation packet types, MTAP16 and MTAP24, respectively, can be used to aggregate NAL units from different pictures into a single RTP packet. The AVC RTP payload format [7] also supports two types of fragmentation units, FU-A and FU-B, which enable fragmentation of one NAL unit into multiple RTP packets. In Fig. 3 the use of the single NAL unit packet and the FU as well as STAP are shown. An SVC bitstream with its access units (coded media frames) in decoding order is shown in the upper part. Each square of a different color represents a NAL unit of a particular layer, where white squares identify the base layer NAL units. For details of the packet description see [2, 8].

The AVC RTP payload format [7] further defines three different packetization modes which differ in the supported packet types as well as the method of packet transmission. The single NAL unit mode only supports the use of single NAL unit packets, the non-interleaved packetization mode additionally allows the presence of FU-A and STAP-A packets, and the interleaved packetization mode only allows the use of STAP-B and MTAP packets, and fragmentation units starting with an FU-B packet.

These three packetization modes are also supported by the SVC payload format. In the SVC payload format it is further allowed to separate data of an SVC bitstream onto different RTP flows (RTP streams) in order to allow differentiation in transport for partial content encryption or unequal treatment by a packet loss recovery mechanism such as retransmission or FEC. We highlight the approach of SVC multisession transmission (MST) later on in more detail.

The SVC payload format provides three new packet types:

Payload content specific information (PACSI): A table of content of the NAL units contained in the payload. This optional packet also carries error resilience information, and it can additionally contain a number indicating the decoding order of the NAL units.

Non-interleaved multi-time aggregation packet (NI-MTAP): This packet allows for grouping of non-interleaved, consecutive NAL units from different frames. This could reduce the packetization overhead, especially in hierarchical group of packets (GOP) coding structures for low-bitrate streaming, where the NAL unit size for pictures of the highest temporal level typically is significantly below the network minimum transmission unit (MTU) size for Ethernet.

Empty packet, which is used to produce *dummy* data of the same timestamp in order to align different RTP flows based on timing information.

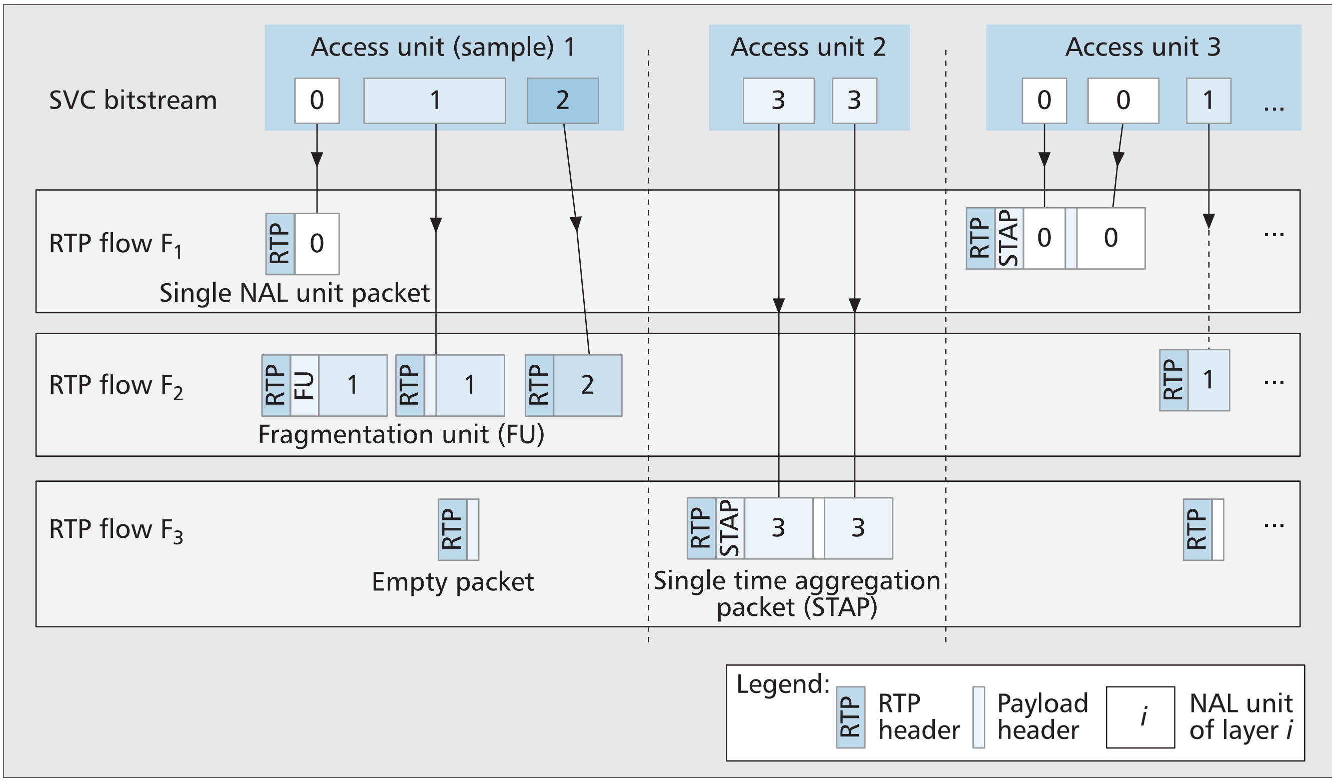


Figure 3. *NI-T multisession transmission of four SVC layers in three RTP flows (example using temporal scalability).*

An important functionality for the transport of SVC is the differentiated transport of NAL unit packets. Therefore, the system layer needs to find out about the importance of packets. Two different approaches can be used here:

Single-session transmission (SST): All NAL units of all layers are transported within one RTP flow; thus, for their differentiation a network element such as a MANE and the receiver need to rely on payload header parsing to indicate the importance of the NAL units.

Multisession transmission (MST): NAL units belonging to one or more layers (identified by a combination of *D*, *Q*, and *T* in the SVC NAL unit header) are transported in a separate RTP flow, which is sent on a separate transport address, as discussed earlier, as session multiplexing.

MST is supported in potentially four different modes:

Non-interleaved timestamp-based (NI-T) mode allows for non-interleaved transmission within each session (a single NAL unit as well as non-interleaved packetization modes can be used) and relies on media timestamps to recover the decoding order if SVC layers are received from different RTP flows.

Non-interleaved cross session decoding order (NI-C) mode allows for non-interleaved transmission within each session (asingle NAL unit as well as non-interleaved packetization modes can be used) and relies on a number provided in the PACSI packet to recover decoding order if SVC layers are received from different RTP flows.

Interleaved cross session decoding order (I-C) mode allows for interleaved transmission within each session (interleaved packetization mode is used) and relies on a number in the payload header to recover decoding order if SVC layers are received from different RTP flows.

NI-TC mode offers both means of decoding order recover (i.e., timestamp-based [NI-T] and number-based [NI-C]).

NI-T mode basically relies on media timestamps to recover the decoding order if NAL units of an SVC stream are received from, say, different transport addresses. One problem may arise because the RTP timestamps represent the presentation time and do not necessarily appear in increasing order within an RTP flow. Therefore, an algorithm is specified in [2, 9] to recover the decoding order starting in the highest RTP flow (i.e., the flow carrying the least important layer in the layered coding hierarchy) to find out the order of RTP timestamp appearance in that RTP flow. The algorithm then proceeds to lower RTP sessions carrying data of the same SVC bitstream. Given the RTP timestamp order from the highest RTP flow, it places the NAL units from lowest to highest RTP flow into the access unit following the RTP timestamp order in the highest RTP flow. This algorithm relies on the precondition that data of a particular timestamp present in one RTP flow is also contained in all higher RTP flows. Therefore, if temporal scalability is used, it may be required to satisfy this rule at the transport layer; in such cases an empty NAL unit is included in higher layers. In [14] a detailed comparison of using NI-T and NI-C mode in error-prone environments is given.

In the example shown in Fig. 3, an empty packet is included in the higher RTP flow (e.g., RTP flow F_3) whenever an access unit exists in a lower layer but not in the higher layer. Figure 3 shows the relation of NAL units and access unit

SDP is required for RTP transport of SVC as well as other purposes, to identify the used packetization modes, the used MST modes, and dependencies between RTP sessions.

present in an SVC bitstream and their mapping to an NI-T multisession transmission. In order to find out dependencies of RTP flows (i.e., the order of RTP flows in terms of *lower* and *higher* RTP flows), a special signaling mechanism is defined as discussed in the next section. Since the NI-T mode depends on timestamps, the RTP timestamps have to be translated to media presentation times; that is, they have to be mapped to wallclock (provided in RTCP) timestamps, in order to allow the aforementioned algorithm the finding of matching timestamps in the different RTP flows. Currently, there is an optional mechanism under specification in [9] to rapidly provide timestamp matching between RTP flows in RTP header extensions.

SIGNALING — SDP

SDP is required for RTP transport of SVC as well as other purposes, to identify the used packetization modes (e.g., non-interleaved mode), the used MST modes (e.g., NI-T mode), and dependencies between RTP sessions. For the latter, a new signaling framework based on SDP grouping has been defined, where each session is identified by a value called *mid*. The SDP dependency grouping [10] in general indicates the potential relation of RTP sessions and additionally provides information with respect to the dependent RTP sessions in the `a=depend` parameters.

Such RTP sessions are typically set up based on a media description starting with `m=` in an SDP message. For each media a single RTP session following a particular RTP payload type can be set up by the client. This is typically negotiated using the Real Time Streaming Protocol (RTSP) for point-to-point streaming, or indicated using Session Announcement Protocol (SAP) for multicast or using other means (e.g., an electronic program guide). RTP payload types are detailed in SDP using a combination of a line starting with `a=rtpmap` and a line starting with `a=fmtp`, where `rtpmap` indicates the payload format to be used (e.g., H264 for AVC or H264-SVC for SVC). The `fmtp` indicates further details about the contained media such as the used profile and level, the parameter sets and transport information, or the mode used for decoding order recovery in multisession transmission (`mst-mode`) and packetization (`packetization-mode`). SDP signaling for SVC also allows for indication of operation points (i.e., the layers) contained in the layers of an RTP session using the parameter `sprop-operation-point-info` in the `fmtp`.

Table 1 shows an SDP example where an SVC bitstream is offered by a server for multisession transmission with up to two potential RTP sessions. The attribute specified in line 2 declares an SDP group having decoding dependencies `DDP`, which contains the media sessions indicated by `mid:` L1 and L2 assigned in lines 8 and 14, respectively, to the media description blocks (shaded with different colors). Each media description is also associated with one or more payload type (PT) numbers in lines 3 and 9, respectively. The `a=depend` attribute (line 15) indicates dependencies per payload type that exist for all layers except the base layer. In Table 1 payload types are marked red, while the media session identifiers are marked blue. For the highest layer of the SVC bitstream indicated in the media description starting at line 9, it is required to also receive either payload type 97 or 98 of the media description starting at line 3 (`mid` = 1).

In order to support SSRC multiplexing (a multiplexing technique based on the RTP header's SSRC field instead of using the network address or port as for RTP session multiplexing) with SVC, there is still a need for a dependency group specification as discussed above to be used with SSRC grouping as defined in [11].

SVC OVER MPEG-2 TRANSPORT STREAM

In this section we discuss the transport mechanism for SVC based on MPEG-2 systems [3]. We give a brief introduction to media transport over the transport stream defined by MPEG-2 systems [4]. Furthermore, we describe the details of using SVC in MPEG-2 transport streams. We then discuss the possibility of using MPEG-2 system streams over IP.

MPEG-2 TRANSPORT STREAM

To encapsulate media elementary stream data (bitstream data) into MPEG-2 TS, MPEG-2 systems defined the packetized elementary stream (PES), which provides packetization of coded media frames and access units. Therefore, an access unit is encapsulated using PES packets. The PES packet header provides framing, size information, stream type identification, as well as media frame related information such as decoding and presentation timestamps. The timing information carried in the PES is relative to a program clock reference (PCR) which is carried in the TS for each program. An access unit may be transported in one or more PES packets, where for SVC one packet per access unit is used.

The PES packets are framed into TS packets, which provide the level for multiplexing other transport streams containing data of other media elementary streams of the same program as well as media data of other programs. Besides PES data, program-specific information (PSI) tables that map the program information to a packet identifier (PID) in the TS packet headers are provided. The TS provides a fixed framing size of typically a 184-byte payload and a 4-byte header per TS packet, which carries, for example, the PID field.

As already mentioned, the transport stream provides the PES data of different elementary streams of different programs, each identified by its own PID as well as table of contents information. For the latter case, a program association table (PAT) is used, which is by default associated to PID = 0 and identifies the different programs (TV channels). Each program has its own program map table (PMT) in a separate PID, which is indicated in the PAT. In the PMT detailed information about each program (i.e., the PIDs of the different PESs) is indicated. The PSI tables such as PMT and PAT are typically provided at fixed intervals in the TS multiplex in order to allow stream access and demultiplexing. Figure 4 shows the relation of the different tables. In the example we assume, in the media data section of the illustrated program, the presence of more than one PID carrying different

SDP message
…
a=group:DDP L1 L2
m=video 20000 RTP/AVP 97 98
a=rtpmap:97 H264/90000
a=fmtp:97 profile-level-id=4d400a ; packetization-mode=1; mst-mode=NI-T; sprop-parameter-sets=…;
a=rtpmap:98 H264/90000
a=fmtp:98 profile-level-id=4d400a ; packetization-mode=2; mst-mode=I-C; sprop-parameter-sets=…;
a=mid:L1
m=video 20002 RTP/AVP 99 100
a=rtpmap:99 H264-SVC/90000
a=fmtp:99 profile-level-id=53000c; packetization-mode=1; mst-mode=NI-T; sprop-parameter-sets=…;
a=rtpmap:100 H264-SVC/90000
a=fmtp:100 profile-level-id=53000c; packetization-mode=2; mst-mode=I-C; sprop-parameter-sets=…;
a=mid:L2
a=depend:99 lay L1:97; 100 lay L1:98

Table 1. *SDP messages describing a simple SVC multi session transmission with two RTP flows.*

PESs of different media (e.g., different layers of an SVC bitstream). More details about the packetization of SVC into MPEG-2 PES and TS packets are given in the next section.

SVC in MPEG-2 Transport Stream

The key to transporting SVC over MPEG-2 TS is the distribution of SVC layers to different PESs, which are indicated by different PIDs in the TS. Different from the RTP payload format for SVC video, the Amendment for the transport of SVC over MPEG-2 system streams [3] defines the transport of SVC data in a more restricted way than the RTP payload format.

The AVC base layer is by definition transported in its own PID and has the stream type assignment of 0x1B, which identifies in the PES header AVC bitstreams conforming to existing AVC profiles as defined in Annex A of [1]. SVC enhancements are transported in a separate PES with a separate PID assigned. Such SVC PESs are indicated by the stream type 0x1F, which indicates AVC bitstreams conforming to SVC profiles defined in Annex G of [1]. An SVC PES carries a video sub-bitstream [3], which is identified by a common value of D for SVC NAL unit header extensions. This rule ensure that data of a particular access unit that is contained in a lower video sub-bitstream (e.g., the AVC base layer) is also contained in all higher SVC video sub-bitstreams. This is important for the access unit reassembling process, which is similar to the one used in the RTP payload format for SVC video in NI-T mode discussed earlier.

Figure 5 illustrates the splitting of an SVC bitstream into different elementary streams, each carried in its own PES. A dependency representation (DR) represents the part of an SVC access unit associated to a particular D-value. The example SVC bitstream includes four different D-values (i.e., an access unit may contain up to DRs). The example stream additionally shows different frame rates in the different video sub-bitstreams; there are different numbers of DRs present in the access units depending on the position in the bitstream (access units AU_n and AU_{n+3} in Fig. 5).

In order to differentiate between different values of Q in the SVC NAL unit header extension, an application has to parse into the PES stream and identify such NAL units, where a quality enhancement to the AVC base layer is contained in the same video sub-bitstream.

For detailed information about the SVC video sub-bitstreams, three different descriptors are available and are contained in the PMT of the program element containing the SVC bitstream:

- Hierarchy descriptor: Indicates the dependencies between PES containing video sub-bitstreams (similar to the dependency grouping discussed earlier) and the type of scalability present in a video sub-bitstream as temporal, spatial, and quality scalability.
- AVC video descriptor: Details about the bit stream such as the used profile and level.
- SVC extension descriptor: Details about the contained operation point in terms of bit-rate, resolution, and ranges of D, Q, and T present in the PES.

At the receiver, the DRs contained in the different PESs need to be reassembled to access unit decoding order. Therefore, [3] defines an access unit and bitstream reassembling process based on the decoding timestamps (DTSs) provided in the PES headers. A similar process as previously presented for the NI-T mode is used, except that MPEG-2 TSs can rely on the presence of DTSs in the TS, where RTP has to rely on presentation timestamps (similar to the PTS). Reference [3] mandates the presence of DTSs and PTSs in all PES packet headers for video sub-bitstreams of an SVC bit stream.

When temporal scalability is used, DTS values of DRs do not necessarily match between layers with different temporal resolutions. In the example

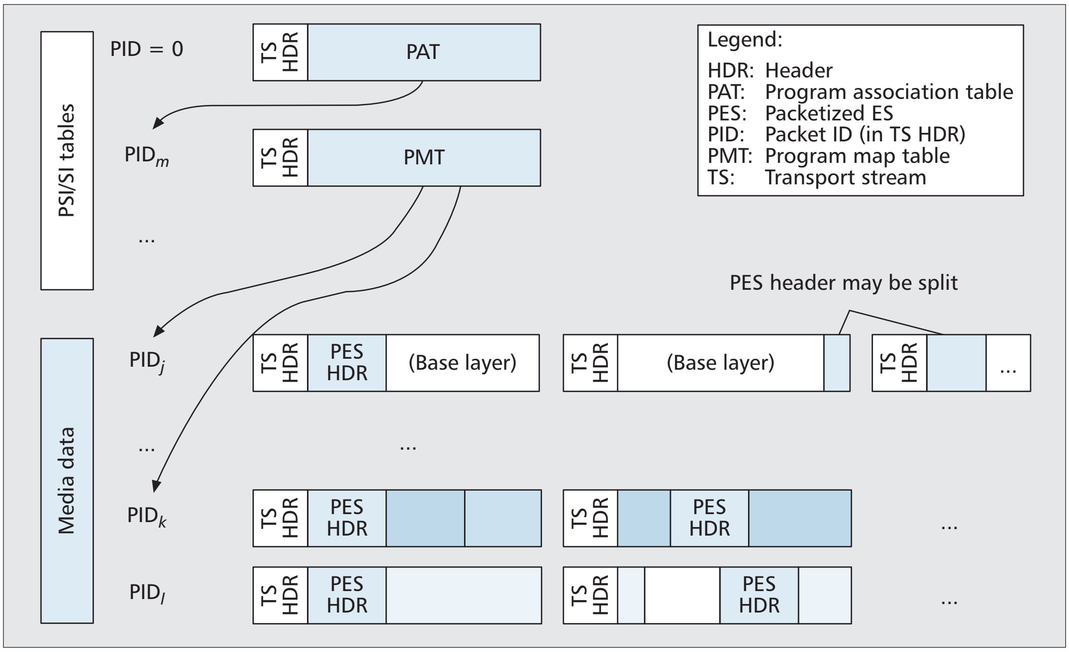


Figure 4. *Meta data structures indicating the location of elementary streams in MPEG-2 TS.*

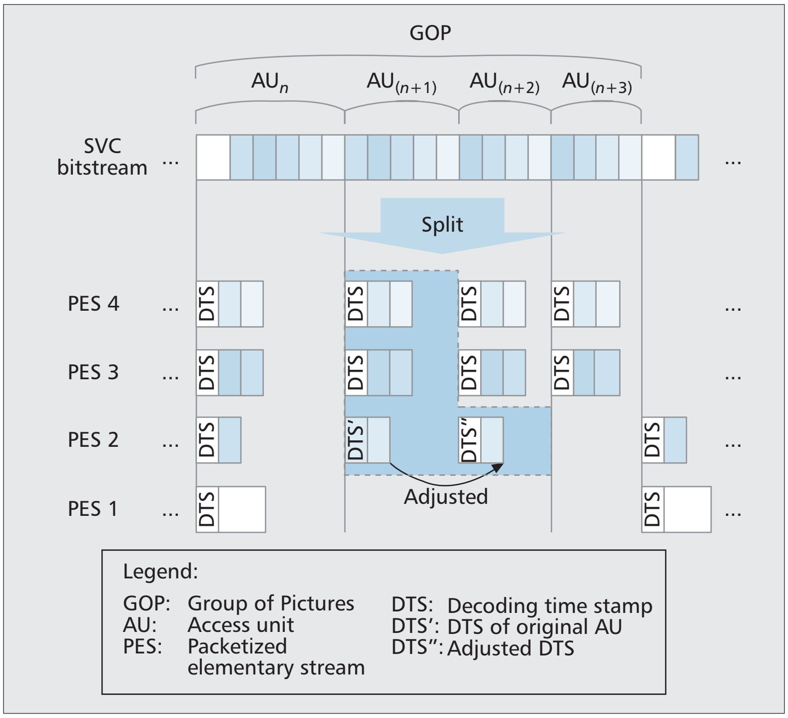


Figure 5. *Splitting an SVC bitstream into elementary streams of MPEG-2 TS.*

shown in Fig. 5 PES 1, carrying the base layer, has the lowest temporal resolution; PES 2, carrying the first SVC enhancement, has twice the temporal resolution of PES 1; and PES 3 again doubles the temporal resolution of PES 2. In order to allow continuous decoding, the DTSs have equidistant intervals, such that the time for decoding an access unit for a lower resolution is larger than that for a higher temporal resolution. In order to allow access unit reassembly in the temporal scalability case, [3] defines, besides the DTS value carried in the PES header, a timestamp reference TREF, which indicates (e.g., for AU$_{(n+3)}$ in PES 3) the DTS″ of the related DR in PES 2.

The timing and buffering model of MPEG-2 systems [4], the system target decoder (STD), defines buffer sizes, data transfer between buffers, as well as the used bit rates for such operations at the receiver. Figure 6 shows the buffer model for SVC. On the left of the figure, the incoming TS is demultiplexed based on the PIDs. According to the PMT of the received program containing the SVC bitstream, the TS packets are forwarded to transport stream buffer TB. Then the data is transferred to the multiplexing buffer (MB), while removing the TS packet header. In the MB the PES packet header data is removed, and the resulting data is byte-wise transferred into the DR buffer (DRB) using a leaky bucket method. Once the DR is completed in each DRB, the data with matching DTS (td) respectively matching TREF (tref) is removed from the DRBs, reassembled in the order of ES (as indicated in the hierarchy descriptor), and transferred to the SVC decoder at time td$_{n+m}$ of the highest received ES with PID$_{n+m}$, (Fig. 6).

MPEG-2 Transport Stream over IP

Since MPEG-2 transport stream provides a completely self-contained transport stream description, it seems obvious that such a stream is also suitable to be contained in IP streams. Therefore, [12] provides a specification to transport an MPEG-2 transport stream over RTP as well as over UDP. Currently there are no means defined to allow multisession transmission of different video subbitstreams over different transport addresses.

Summary

In this article we give a compact overview of the SVC transport standards for digital TV and IPTV channels. We highlight the feature of mul-

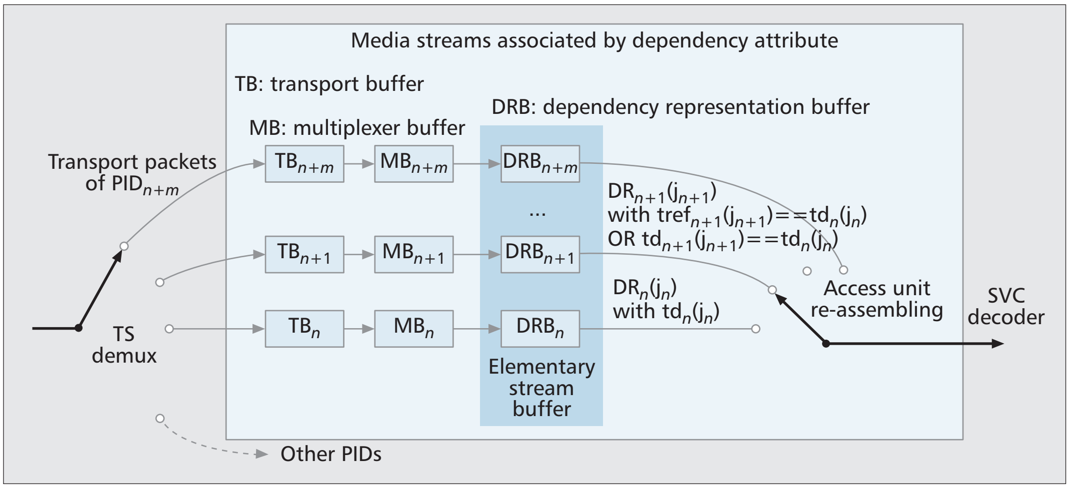


Figure 6. *Decoder buffer allocation and reassembly of access units for SVC layers.*

Since MPEG-2 Transport Stream provides a completely self-contained transport stream description, it seems obvious that such a stream is also suitable to be contained in IP streams.

tisession transmission in RTP and transmission over multiple MPEG-2 transport streams.

References

[1] ITU-T Rec. H.264, "Advanced Video Coding for Generic Audiovisual Services," Nov. 2007; ISO/IEC 14496-10:2008, "Information Technology — Coding of Audio-Visual Objects (MPEG-4) — Part 10: Advanced Video Coding," 2008.

[2] S. Wenger *et al.*, Eds., "RTP Payload Format for SVC Video," IETF AVT, Sept. 2009; http://tools.ietf.org/html/draft-ietf-avt-rtp-svc

[3] T. Schierl, B. Berthelot, and T. Wiegand, Eds., "ISO/IEC 13818-1:2007/AMßD3 — Transport of SVC Video over ITU-T Rec. H.222.0," ISO/IEC 13818-1, to be published.

[4] ITU-T Rec. H.222.0, "Information Technology — Generic Coding of Moving Pictures and Associated Audio Information: Systems," May 2006; ISO/IEC 13818-1:2007, "Information Technology — Generic Coding of Moving Pictures and Associated Audio Information (MPEG-2) — Part 1: Systems," 2007.

[5] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans. Circuits Sys. for Video Tech.*, vol. 17, no. 9, Sept. 2007, pp. 1103–20.

[6] T. Schierl, T. Stockhammer, and T. Wiegand, "Mobile Video Transmission using ," *IEEE Trans. Circuits Sys. for Video Tech.*, vol. 17, no. 9, Sept. 2007, pp. 1204–17.

[7] S. Wenger *et al.*, Eds., "RTP Payload Format for H.264 Video," IETF RFC 3984, Feb. 2005.

[8] S. Wenger, Y.-K. Wang, and T. Schierl, "Transport and Signaling of SVC in IP Networks," *IEEE Trans. Circuits Sys. for Video Tech.*, vol. 17, no. 9, Sept. 2007, pp. 1164–73.

[9] C. Perkins and T. Schierl, Eds., "Rapid Synchronisation of RTP Flows," IETF AVT, July 2009; http://tools.ietf.org/html/draft-ietf-avt-rapid-rtp-sync-05

[10] T. Schierl and S. Wenger, Eds., "Signaling Media Decoding Dependency in Session Description Protocol (SDP)," IETF MMUSIC RFC 5583, July 2009.

[11] J. Lennox, J. Ott, and T. Schierl, Eds., "Source-Specific Media Attributes in the Session Description Protocol," IETF MMUSIC RFC 5576, June 2009.

[12] D. Hoffman *et al*. (Eds.), "RTP Payload Format for MPEG1/MPEG2 Video," IETF RFC 2250, Jan. 1998.

[13] Y.-K. Wang *et al.*, "System and Transport Interface of SVC," *IEEE Trans. Circuits Sys. for Video Tech.*, vol. 17, no. 9, Sept. 2007, pp. 1149–63.

[14] T. Schierl *et al.*, "Decoding Order Recovery for Multi Flow Transmission of Scalable Video Coding (SVC) over Mobile IP Channels," *Proc. ICST Mobimedia 2009*, London, U.K., Sept. 2009.

Biographies

THOMAS SCHIERL (Thomas.Schierl@hhi.fraunhofer.de) received his Dipl.-Ing. degree in computer engineering from Berlin University of Technology, Germany, in December 2003. He has been with Fraunhofer Institute for Telecommunications — HHI since 2004. As project manager he is responsible for HHI's multimedia networking activities, and leads various scientific and industry research projects. He has conducted research work on reliable real-time transmission of H.264/MPEG-4 AVC and SVC in mobile point-to-point, point-to-multipoint, and broadcast environments used by 3GPP PSS/MBMS and DVB-H/SH. He has submitted different inputs on real-time streaming to standardization organizations including 3GPP, JVT/MPEG, ISMA, and IETF. Besides the IETF RTP payload format for SVC and the IETF generic SDP signaling for layered codecs, he is co-authoring various IETF inputs. Furthermore, he is one of the editors of the Amendments for SVC and MVC over MPEG-2 TS. In 2007 he visited the group of Prof. Bernd Girod at Stanford University for research activities on application layer multicast using SVC over wired and wireless channels.

KARSTEN GRÜNEBERG (Karsten.Grueneberg@hhi.fraunhofer.de) received his Diplom-Ingenieur degree in electrical engineering from the University of Hannover in 1987. He joined the research group on HDTV in the Image Processing and Terminal Devices department of Heinrich-Hertz-Institute, also in 1987. He conducted different research activities on signal processing for HDTV transmission and storage devices. Further work concerned the implementation of a real-time system for stereoscopic videoconferencing with view interpolation. Recently, he worked on architectures and implementations for networked media distribution and storage in the context of several European research projects (MediaNet, ASTRALS, SEA). He is also actively participating in and contributing to standardization committees such as MPEG and DVB.

THOMAS WIEGAND (Thomas.Wiegand@hhi.fraunhofer.de) is a professor of electrical engineering at the Technical University of Berlin chairing the Image Communication group, and jointly heads the Image Processing department of the Fraunhofer Institute for Telecommunications — Heinrich Hertz Institute, Berlin, Germany. Since 1995 he has been an active participant in standardization for multimedia with successful submissions to ITU-T VCEG, ISO/IEC MPEG, 3GPP, DVB, and IETF. In October 2000 he was appointed Associate Rapporteur of ITU-T VCEG. In December 2001 he was appointed Associate Rapporteur/Co-Chair of the JVT. In February 2002 he was appointed Editor of the H.264/AVC video coding standard and its extensions (FRExt and SVC). In January 2005 he was appointed Co-Chair of MPEG Video. In 1998 he received the SPIE VCIP Best Student Paper Award. In 2004 he received the Fraunhofer Award for outstanding scientific achievements in solving application related problems and the ITG Award of the German Society for Information Technology. Since January 2006 he has been an Associate Editor of *IEEE Transactions on Circuits and Systems for Video Technology*. In 2008 he received the Primetime Emmy Engineering Award that was awarded to the JVT standards committee by the Academy of Television Arts & Sciences for development of the High Profile of H.264/MPEG-4 AVC, for which he served as Associate Rapporteur/Co-Chair, Editor, and technical contributor.