Network Working Group
Internet Draft
Document: draft-ietf-avt-rtp-svc-00.txt
Expires: June 2007

S. Wenger
Y.-K. Wang
T. Schierl

December 2006

RTP Payload Format for SVC Video

Status of this Memo

Abstract

This memo describes an RTP Payload format for the scalable extension of the ITU-T Recommendation H.264 video codec which is technically identical to ISO/IEC International Standard 14496-10 video codec.

The RTP payload format allows for packetization of one or more Network Abstraction Layer Units (NAL units), produced by the video encoder, in each RTP payload. The payload format has wide applicability, as it supports applications from simple low bit-rate conversational, through Internet video streaming with interleaved transmission, to high bit-rate video-on-demand.

Table of Content

1. Introduction

1.1. SVC -- the scalable extension of H.264/AVC

This memo specifies an RTP [RFC3550] payload format for a forthcoming new mode of the H.264/AVC video codec, known as Scalable Video Coding (SVC). Formally, SVC will take the form of an Amendment to ISO/IEC 14496 Part 10 [MPEG4-10], and likely as one or more new Annexes of ITU-T Rec. H.264 [H.264]. It is planned to keep the technical alignment between the two mentioned specifications, as well as backward compatibility with previous versions of H.264/AVC.

The current working draft of SVC is available for public review [SVC]. In this memo, SVC is used as an acronym for the mentioned scalable extension of H.264/AVC. In that, SVC is a superset of H.264/AVC.

SVC covers the whole application ranges of H.264/AVC. This range is considerable, starting with low bit-rate Internet streaming applications to HDTV broadcast and Digital Cinema with nearly lossless coding and requiring dozens or hundreds of MBit/s.

This memo tries to follow a backward compatible enhancement philosophy similar to what the video coding standardization committees implement, by keeping as close an alignment to the H.264/AVC payload RFC [RFC3984] as possible. It documents the enhancements relevant from an RTP transport viewpoint, defines signaling support for SVC, and deprecates the single NAL unit packetization mode of RFC 3984.

2. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14, RFC 2119 [RFC2119].

This specification uses the notion of setting and clearing a bit when bit fields are handled. Setting a bit is the same as assigning that bit the value of 1 (On). Clearing a bit is the same as assigning that bit the value of 0 (Off).

3. The SVC Codec

3.1. Overview

SVC provides scalable video bitstreams. In SVC, a scalable video bitstream contains a base layer conforming to the existing profiles of H.264 as defined in [H.264], and one or more enhancement layers. An enhancement layer may enhance the temporal resolution (i.e. the frame rate), the spatial resolution, or the quality of the video content represented by the lower layer or part thereof.

Each RTP packet stream can carry NAL units belonging to one or more layers. The NAL unit headers include information of the association of a given NAL unit to a layer. Therefore, extracting individual layers from an RTP packet stream containing more than one layer is a lightweight operation, involving only fixed length bit fields in the header as documented in this memo and in [SVC].

Multiple RTP packet streams, regardless whether they carry a single or multiple layers as discussed above, can be used to transport the whole scalable bitstream, or operation points thereof. When multiple RTP packet streams are in use, they are session multiplexed, i.e. form their own RTP session and therefore have their own SSRC, PT, and Sequence numbering space, among all other properties of a session as spelled out in section xxx of [RFC3550].

The concept of video coding layer (VCL) and network abstraction layer (NAL) is inherited from H.264. The VCL contains the signal processing functionality of the codec; mechanisms such as transform, quantization, motion-compensated prediction, loop filtering and inter-layer prediction. A coded picture in H.264 consists of one or more slices. In SVC, a particular layer consists of all the coded slices required for decoding up to that layer. Within one access unit, a coded picture representing a particular layer consists of all the coded slices required for decoding up to the particular layer at the time instance corresponding to the access unit. The Network Abstraction Layer (NAL) encapsulates each slice generated by the VCL into one or more Network Abstraction Layer Units (NAL

units). Please consult RFC 3984 for a more in-depth discussion of the NAL unit concept. SVC specifies the decoding order of the NAL units.

``Layer'' in the terms ``Video Coding Layer'' and ``Network Abstraction Layer'' refers to a conceptual distinction, and is closely related to syntax layers (block, macroblock, slice, ... layers). ``Layer'' here describes a syntax level of the bitstream in contrast to a part of the layered bitstream, which may be discarded. It should not be confused with base and enhancement layers.

The concept of temporal scalability is not newly introduced by SVC, as H.264 already supports it. In [H.264], sub-sequences have been introduced in order to allow optional use of temporal layers. SVC extends this approach by advertising the temporal layer information within the NAL unit header, or suffix NAL units, as discussed in section 3.3 of this memo and in [SVC]. By our definition, the base layer may be scalable in the temporal dimension.

The concept of scaling the visual content quality in the granularity of complete enhancement layers, i.e. through omitting the transport and decoding of entire enhancement layers, is denoted as coarse-grained scalability (CGS). This is what is commonly understood as scalability in the IETF community. According to SVC, a CGS layer may be a spatial or quality (SNR) enhancement layer.

In some cases, the bit rate of a given enhancement layer may be reduced by truncating bits from individual NAL units. Truncation leads to a graceful degradation of the video quality of the reproduced enhancement layer. This concept is known as Fine Granularity Scalability (FGS). In SVC, FGS is provided by a concept known as progressive refinement slices.

3.2. Parameter Set Concept

The parameter set concept is inherited from [H.264]. Please refer to section 1.2 of RFC 3984 for more details.

In SVC, pictures from different layers may use the same sequence or picture parameter set, but may also use different sequence or

picture parameter sets. If different sequence or picture parameter sets are used, then, at any time instant during the decoding process, there may be more than one active sequence or picture parameter set. Any specific active sequence parameter set remains unchanged throughout a coded video sequence in the layer in which the active sequence parameter set is referred to. The active picture parameter set remains unchanged within a coded picture.

3.3. Network Abstraction Layer Unit Header

An SVC NAL unit, i.e., a NAL units of type 20 and 21, consists of a header of four or five bytes and the payload byte string. An SVC NAL unit typically encapsulates VCL data as defined in Annex G of [SVC] but may also contain VCL data compliant to older profiles of [H.264]. A special type of an SVC NAL unit is the suffix NAL unit

that includes descriptive information of a preceding NAL unit.

SVC extends the NAL unit header defined in [H.264] by three or four additional bytes. The header indicates the type of the NAL unit, the (potential) presence of bit errors or syntax violations in the NAL unit payload, information regarding the relative importance of the NAL unit for the decoding process, the layer decoding dependency information, and FGS fragmentation information. This RTP payload specification is designed to be unaware of the bit string in the NAL unit payload.

The NAL unit header co-serves as the payload header of this RTP payload format. The payload of a NAL unit follows immediately.

The syntax and semantics of the NAL unit header are formally specified in [SVC], but the essential properties of the NAL unit header are summarized below.

The first byte of the NAL unit header has the following format (the bit fields are the same as in [H.264] and [RFC3984], while the semantics have changed slightly, in a backward compatible way):


```
+---------------+
|0|1|2|3|4|5|6|7|
+-+-+-+-+-+-+-+-+
|F|NRI|  Type   |
+---------------+
```

F: 1 bit
forbidden_zero_bit. H.264 declares a value of 1 as a syntax violation.

NRI: 2 bits
nal_ref_idc. A value of 00 indicates that the content of the NAL unit is not used to reconstruct reference pictures for inter picture prediction. Such NAL units can be discarded without risking the integrity of the reference pictures in the same layer. Values greater than 00 indicate that the decoding of the NAL unit is required to maintain the integrity of the reference pictures.

Type: 5 bits
nal_unit_type. This component specifies the NAL unit payload type as defined in table 7-1 of [SVC], and later within this memo. For a reference of all currently defined NAL unit types and their semantics, please refer to section 7.4.1 in [SVC].

Previously, NAL unit types 20 and 21 (among others) have been reserved for future extensions. SVC is using these two NAL unit types. They indicate the presence of three or four additional bytes in the NAL unit header. The first three additional bytes are as shown below.

```
+---------------+---------------+---------------+
|0|1|2|3|4|5|6|7|0|1|2|3|4|5|6|7|0|1|2|3|4|5|6|7|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|RR |   PRID    | TL  | DID | QL|B|U|D|G|L| O |E|
+---------------+---------------+---------------+
```

RR: 2 bits
reserved_zero_two_bits. Reserved bits for future extension. RR MUST be zero.

PRID: 6 bits

priority_id. This component specifies a priority identifier for the NAL unit. A lower value of PRID indicates a higher priority.

TL: 3 bits
temporal_id. This component indicates the temporal layer (or frame rate) hierarchy. Informally put, a layer consisted of pictures of a smaller temporal_id value has a smaller frame rate. A given temporal layer typically depends on the lower temporal layers (i.e. the temporal layers with smaller temporal_id values) but never depends on any higher temporal layer.

DID: 3 bits
dependency_id. This component denotes the inter-layer coding dependency hierarchy. At any temporal location, a picture of a smaller dependency_id value may be used for inter-layer prediction for coding of a picture of a larger dependency_id value, while a picture of a larger dependency_id value is disallowed to be used for inter-layer prediction for coding of a picture of a smaller dependency_id value.

QL: 2 bits
quality_id. This component designates the quality level hierarchy of

a progressive refinement (PR) or quality (SNR) enhancement layer slice. At any temporal location and with identical dependency_id value, a picture with quality_id equal to ql uses a picture with quality_id equal to ql-1 for inter-layer prediction.

B: 1 bit
layer_base_flag. A value of 1 indicates that no inter-layer prediction (of coding mode, motion, sample value, and/or residual prediction) is used for the current slice. A value of 0 indicates that inter-layer prediction may be used for the current slice.

U: 1 bit
use_base_prediction_flag. A value of 1 indicates that only the base representations of the reference pictures are used during the inter prediction process of the current slice. A value of 0 indicates that the base representations of the reference pictures are not used during the inter prediction process of the current slice.

D: 1 bit

discardable_flag. A value of 1 indicates that the content of the NAL unit with dependency_id equal to currDependencyId is not used in the decoding process of NAL units with dependency_id larger than currDependencyId. Such NAL units can be discarded without risking the integrity of higher scalable layers with larger values of dependency_id. discardable_flag equal to 0 indicates that the decoding of the NAL unit is required to maintain the integrity of higher scalable layers with larger values of dependency_id.

G: 1 bit
fragmented_flag. A value of 1 indicates that the current NAL unit is a FGS (progressive refinement) slice. A value of 0 indicates that the current NAL unit is not a FGS slice. If quality_id is equal to 0, fragmented_flag shall be equal to 0.

L: 1 bit
last_fragment_flag. When fragmented_flag is equal to 0, the semantics of this component is unspecified. When fragmented_flag is equal to 1, this component, together with fragment_order, specifies whether the current NAL unit is a fragmented FGS slice, and if yes, whether the current NAL unit is the last fragment of the fragmented slice, as follows. When fragment_order is equal to 0 and last_fragment_flag is equal to 1, the current NAL unit is an un-fragmented FGS slice. When fragment_order is greater than 0 and last_fragment_flag is equal to 1, the current NAL unit is the last fragment of a fragmented FGS slice. When last_fragment_flag is equal

to 0, the current NAL unit is a fragment but not the last fragment of a fragmented FGS slice.

O: 2 bits
fragment_order. When fragmented_flag is equal to 0, the semantics of this component is unspecified. When fragmented_flag is equal to 1, this component, together with last_fragment_flag, specifies whether the current NAL unit is a fragmented FGS slice, and if yes, the fragment order, as follows. When fragment_order is equal to 0 and last_fragment_flag is equal to 1, the current NAL unit is an un-fragmented FGS slice. When fragment_order is greater than 0 and last_fragment_flag is equal to 1, the current NAL unit is the last fragment of a fragmented FGS slice, and fragment_order indicates the fragment order. When last_fragment_flag is equal to 0, the current

NAL unit is a fragment but not the last fragment of a fragmented FGS slice, and fragment_order indicates the fragment order.

E: 1 bit
extension_flag. A value of 1 indicates the existence of the last byte, tl0_frame_idx, in the NAL unit header. A value of 0 indicates that tl0_frame_idx is not present in the NAL unit header. Please refer to [SVC] for information in detail about tl0_frame_idx.

This memo introduces the same additional NAL unit types as RFC 3984, which are presented in section 6.3. The NAL unit types defined in this memo are marked as unspecified in [SVC]. Moreover, this specification extends the semantics of F, NRI, PRID, D, TL, DID and QL as described in section 6.4.

4. Scope

This payload specification can only be used to carry the "naked" NAL unit stream over RTP, and not the byte stream format according to Annex B of [SVC]. Likely, the applications of this specification will be in the IP based multimedia communications fields including conversational multimedia, video telephony or video conferencing, Internet streaming and TV over IP.

This specification allows, in a given RTP session, to encapsulate NAL units belonging to
  o the base layer only, detailed specification in [RFC3984], or
  o one or more enhancement layers, or
  o the base layer and one or more enhancement layers

5. Definitions and Abbreviations

5.1. Definitions

5.1.1. Definitions per SVC specification

This document uses the definitions of [SVC]. The following terms, defined in [SVC], are summed up for convenience:

scalable bitstream: A bitstream that uses the scalable extensions defined in Annex G of [SVC], i.e. a bitstream with a base layer and at least one enhancement layer.

suffix NAL unit: A NAL unit that immediately follows another NAL unit in decoding order and contains descriptive information of the preceding NAL unit, which is referred to as the associated NAL unit. A suffix NAL unit shall have nal_ref_idc equal to 20 or 21, shall have dependency_id and quality_level both equal to 0, and shall not contain a coded slice. A suffix NAL unit belongs to the same coded picture as the associated NAL unit. A suffix NAL unit may be used for indicating temporal levels within the base layer.

base layer: The base layer is typically representing the minimal spatial resolution and the minimal fidelity of an SVC bitstream. The base layer must be fully complying with [H.264]. The base layer is independently decodable without the requirement of using any other layer of the SVC bitstream. In SVC context each slice NAL unit in the base layer is associated with a suffix NAL unit, which has a four or five bytes NAL unit header containing all the syntax elements described in section 3.3. The base layer may be temporally scalable.

enhancement layer: An SVC enhancement layer is identified by priority_id, temporal_level, dependency_id, and quality_level as defined in [SVC] and summarized in section 3.3.

access unit: A set of NAL units pertaining to a certain temporal location. An access unit includes the coded slices of all the scalable layers at that temporal location and possibly other associated data, e.g. SEI messages and parameter sets.

coded video sequence: A sequence of access units that consists, in decoding order, of an instantaneous decoding refresh (IDR) access unit followed by zero or more non-IDR access units including all subsequent access units up to but not including any subsequent IDR access unit.

IDR access unit: An access unit in which all the primary coded pictures are IDR pictures. Such an access unit allows for random access to any operation point.

IDR picture: A coded picture with the property that the decoding of this coded picture and all the following coded pictures in decoding order, with the same value of dependency_id, can be performed without inter prediction from any picture prior to the coded picture in decoding order with the same value of dependency_id. Thus an IDR picture allows for random access to the scalable layer, which it belongs to. An IDR picture causes a "reset" in the decoding process of the scalable layer containing the IDR picture.

progressive refinement (PR) slice: A progressive refinement slice is contained in an SVC NAL unit that may be truncated since the end of the slice header for bit-rate and quality reduction. PR slices provide Fine Granularity Scalability (FGS).

5.1.2. Definitions local to this memo

operation point: An operation point of a SVC bitstream represents a certain level of temporal, spatial and quality scalability. An operation point contains all NAL units required for restoring a valid bitstream (conforming to [SVC]) up to a certain SVC layer. The operation point is further described by priority_id, temporal_level, dependency_id, and quality_level values of that layer.

RTP packet stream: A sequence of RTP packets with increasing sequence numbers, identical PT and SSRC, carried in one RTP session. Within the scope of this memo, one RTP packet stream is utilized to transport an integer number of SVC layers.

Session multiplexing: The scalable SVC bitstream is distributed onto different RTP sessions, whereby each RTP session carries a single RTP packet stream. Each RTP session requires a separate signaling and has a separate Timestamp, Sequence Number, and SSRC space. Dependency between sessions MUST be signaled according to [SDPsiglay] and this memo.

5.2. Abbreviations

In addition to the abbreviations defined in [RFC3984], the following ones are defined.

CGS: Coarse Granularity Scalability
FGS: Fine Granularity Scalability

6. RTP Payload Format

6.1. Design Principles

The following design principles have been observed:

- Backward compatibility with RFC 3984 wherever possible.

- As the SVC base layer is H.264/AVC compatible, we assume the base layer (when transmitted in its own session) to be encapsulated using RFC 3984. Requiring this has the desirable side effect that it can be used by RFC 3984 legacy devices.

- MANEs are signaling aware and rely on signaling information. MANEs have state.

- MANEs can terminate RTP sessions, and create different RTP sessions with perhaps modified content. This form of a MANE acts as an RTP mixer.

- MANEs can also act as RTP translators. The perhaps most likely use case is media-aware stream thinning. By using the payload header information identifying layers within an RTP session, MANEs are able to remove packets from the RTP session while otherwise keeping the session intact. This implies rewriting the RTP headers of the outgoing packet stream and rewriting of RTCP Receiver Reports.

- Packet integrity needs to be preserved end-to-end (whereby end-to-end can mean endpoint to endpoint but also endpoint to MANE, if (and only if) the MANE acts as a Mixer).

- In case of layered multicast transmission as motivated in section 13.2, each RTP packet stream in a given session may contain NAL units belong to one or more SVC layer(s) of the same scalable

bitstream. The layers contained within a RTP session may be

identified by using payload header structures as defined in this memo.

6.2. RTP Header Usage

Please see section 5.1 of RFC 3984 [RFC3984]. The following applies in addition.

When layers of an SVC scalable bitstream are transported in more than one RTP session, e.g. in layered multicast for which the use case is given in 13.2, session multiplexing MUST be used only as RTP multiplexing technique.

6.3. Common Structure of the RTP Payload Format

Please see section 5.2 of RFC 3984 [RFC3984].

6.4. NAL Unit Header Usage

The structure and semantics of the NAL unit header were introduced in section 3.3. This section specifies the semantics of F, NRI, PRID, D, TL, DID, QL, B, U, G, L, and O according to this specification.

The semantics of F specified in section 5.3 of [RFC3984] also applies herein.

For NRI, for the bitstream that is compliant with [H.264] and transported using RFC 3984, the semantics specified in section 5.3 of [RFC3984] are applicable, i.e., NRI also indicates the relative importance of NAL units. In SVC context, only the semantics specified in [SVC] are applicable, i.e., NRI does not indicate the relative importance of NAL units.

For PRID, the semantics specified in [SVC] applies. In addition, MANEs implementing unequal error protection may use this information to protect NAL units with smaller PRID values better than those with larger PRID values, for example by including only the more important NAL units in a FEC protection mechanism. The importance for the decoding process decreases as the PRID value increases.

For D, in addition to the semantics specified in [SVC], according to this memo, MANEs may use this information to protect NAL units with D equal to 0 better than NAL units with D equal to 1. Furthermore, based on this information, a MANE or a receiver may determine whether a given NAL unit is required for successfully decoding a certain operation point of the SVC bitstream.

For TL, DID and QL, in addition to the semantics specified in [SVC], according to this memo, values of TL, DID or QL indicate the relative priority in their respective dimension. A lower value of TL, DID or QL indicates a higher priority if the other two components are identical correspondingly. MANEs may use this information to protect more important NAL units better than less important NAL units.

> Informative note: PRID, D, TL, DID, and QL, in combination, provide complete information of the relative priority of a NAL unit compared to any other NAL unit. [Edt. note: examples may be provided in Informative Appendix 13 in future versions.]

For U, in addition to the semantics specified in [SVC], according to this memo, MANEs may use this information to protect NAL units with U equal to 1 (which are referred to as key picture NAL units) better than NAL units with U equal to 0.

6.5. Packetization Modes

Please see section 5.4 of RFC 3984 [RFC3984]. The single NAL unit packetization mode SHALL NOT be used.

> Informative note: The non-interleaved mode allows an application to encapsulate a single NAL unit in a single RTP packet. Historically, the single NAL unit mode has been included into [RFC3984] only for compatibility with ITU-T Rec. H.241 Annex A [H.241]. There is no point in carrying this historic ballast towards a new application space such as the one provided with SVC. More technically speaking, the implementation complexity increase for providing the additional mechanisms of the non-interleaved

mode (namely STAPs) is so minor, and the benefits are so great, that we require STAP implementation.

6.6. Decoding Order Number (DON)

Please see section 5.5 of RFC 3984 [RFC3984]. The following applies in addition.

When different layers of a SVC bitstream are transported in more than one RTP packet stream, the interleaved packetization mode MUST be used, and the DON values of all the NAL units MUST indicate the correct NAL unit decoding order over all the RTP packet streams.

If Session multiplexing is used, each session MUST signal an identical value for the MIME parameters sprop-interleaving-depth, sprop-max-don-diff, sprop-deint-buf-req, and sprop-init-buf-time. Further, these values must be valid for the reception capabilities over all sessions. A receiver MUST signal the same MIME parameter deint-buf-cap for all sessions used for Session multiplexing. [Ed.Note(YkW): I think we need more thinking on the value of the parameters. For example, requiring the parameters be the same for all the RTP streams and clients might be overkill for receivers of only lower layers.]
Edt. Note (StW): In RFC3984, the aforementioned codepoints are optional. It appears that for SVC, when used in conjunction with session mux, they are mandatory. I don't know how to express this in the MIME registration; we'll cross that bridge once we are getting to it.

6.7. Single NAL Unit Packet

Please see section 5.6 of RFC 3984 [RFC3984].

6.8. Aggregation Packets

Please see section 5.7 of RFC 3984 [RFC3984].

6.9. Fragmentation Units (FUs)

Please see section 5.8 of RFC 3984 [RFC3984].

6.10. Payload Content Scalability Information (PACSI) NAL Unit

A new NAL unit type is specified in this memo, and referred to as

payload content scalability information (PACSI) NAL unit. The PACSI NAL unit, if present, MUST be the first NAL unit in an aggregation packet, and it MUST NOT be present in other types of packets. The PACSI NAL unit indicates scalability characteristics that are common for all the remaining NAL units in the payload, thus making it easier for MANEs to decide whether to forward/process/discard the aggregation packet. Senders MAY create PACSI NAL units and receivers MAY ignore them, or use them as hints to enable efficient aggregation packet processing.

Informative note: The NAL unit type for the PACSI NAL unit is selected among those values that are unspecified in the SVC specification and in RFC 3984 -- and therefore are ignored by H.264/AVC or SVC decoders and RFC 3984 receivers. Hence an SVC stream, even when including PACSI NAL units, can be processed with RFC 3984 receivers and H.264/AVC or SVC decoders.

When the first aggregation unit of an aggregation packet contains a PACSI NAL unit, there MUST be at least one additional aggregation unit present in the same packet. The RTP header fields are set according to the remaining NAL units in the aggregation packet.

When a PACSI NAL unit is included in a multi-time aggregation packet, the decoding order number for the PACSI NAL unit MUST be set to indicate that the PACSI NAL unit is the first NAL unit in decoding order among the NAL units in the aggregation packet or the PACSI NAL unit has an identical decoding order number to the first NAL unit in decoding order among the remaining NAL units in the aggregation packet.

The structure of PACSI NAL unit is exactly the same as the four-byte SVC NAL unit header (where E is equal to 0) specified in 3.3, and reproduced here once more for convenience:


```
+---------------+---------------+---------------+---------------+
|0|1|2|3|4|5|6|7|0|1|2|3|4|5|6|7|0|1|2|3|4|5|6|7|0|1|2|3|4|5|6|7|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|F|NRI|  Type   |RR |   PRID    | TL  | DID | QL|B|U|D|G|L| O |E|
+---------------+---------------+---------------+---------------+
```

The values of the fields in PACSI NAL unit MUST be set as follows.

o The F bit MUST be set to 1 if the F bit in at least one remaining NAL unit in the payload is equal to 1. Otherwise, the F bit MUST be set to 0.

o The NRI field MUST be set to the highest value of NRI field among all the remaining NAL units in the payload.

o The Type field MUST be set to 30.

o The RR field MUST be set to 0.

o The PRID field MUST be set to the lowest value of the PRID values associated with all the remaining NAL units in the payload.

o The TL field MUST be set to the lowest value of the TL values associated with all the remaining NAL units in the payload.

o The DID field MUST be set to the lowest value of the DID values associated with all the remaining NAL units in the payload.

o The QL field MUST be set to the lowest value of the QL values associated with all the remaining NAL units in the payload.

o The B bit MUST be set to 1 if the B bit associated with all the remaining NAL units in the payload is equal to 1. Otherwise, the B bit MUST be set to 0.

o The U bit MUST be set to 1 if the U bit associated with all the remaining NAL units in the payload is equal to 1. Otherwise, the

U bit MUST be set to 0.

o The D bit MUST be set to 0 if the D value associated with at least one remaining NAL unit in the payload is equal to 0. Otherwise, the D bit MUST be set to 1.

o The G bit MUST be set to 1 if the G bit associated with at least one of the remaining NAL units in the payload is equal to 1. Otherwise, the G bit MUST be set

to 0.

o The L bit MUST be set to 1 if for any NAL unit having fragmented_flag equal to 1 in the payload, the corresponding NAL unit having the bit L equal to 1 is also in the payload. Otherwise, the bit L MUST be set to 0.

o The O field MUST be set to the lowest value of the O values associated with all the remaining NAL units in the payload.

o The E field or extension_flag field (1 bit) MUST be set to 0.

7. Packetization Rules

Please see section 6 of RFC 3984 [RFC3984]. The following rules apply in addition.

The single NAL unit mode SHALL NOT be used. (See also section 6.5 for the motivation).

When a suffix NAL unit is encapsulated for transmission, it SHOULD be aggregated to the same transmission packet as the NAL unit preceding the suffix NAL unit in decoding order.

Informative note: When either the suffix NAL unit or the associated NAL unit containing an H.264/AVC coded slice is lost, the remaining one would be of no use in SVC context.

When layers of a SVC bitstream are transported in more than one RTP session, the interleaved packetization mode MUST be used.

8. De-Packetization Process (Informative)

Please see section 7 of RFC 3984 [RFC3984]. The following rules apply in addition.

[Edt. Do we need here more information about cross layer DON? TS: Yes, in the next version.]

9. Payload Format Parameters

[Edt. note: this section 9 and its subsections will be updated according to the changes listed below, a little later in the

process. For now, we just list the adjustments necessary, so not to bury any new information in the RFC 3984 text.]

Section 8 of [RFC3984] applies with the following modification.

The sentence

''The parameters are specified here as part of the MIME subtype registration for the ITU-T H.264 | ISO/IEC 14496-10 codec.''

is replaced with

''The parameters are specified here as part of the MIME subtype registration for the SVC codec.''

9.1. MIME Registration

Editor's note: this needs to be updated by copy-pasting the RFC 3984 MIME registration into this document, so to make it self-contained. Will be done later in the process.

The MIME subtype for the SVC codec is allocated from the IETF tree.

The receiver MUST ignore any unspecified parameter.

Media Type name: video

Media subtype name: H.264-SVC

Required parameters: none

OPTIONAL parameters:

The optional MIME parameters specified in [RFC3984] apply, with the following constraints (to be edited in at the appropriate time):

sprop-interleaving-depth:
In case of using Session multiplexing, the same sprop-interleaving-depth value MUST be signaled for all sessions and MUST be valid over all sessions of the multiplex.

sprop-max-don-diff:
In case of using Session multiplexing, the same sprop-max-don-diff value MUST be signaled for all sessions and MUST be valid over all sessions of the multiplex.

sprop-deint-buf-req:
In case of using Session multiplexing, the same sprop-deint-buf-req value MUST be signaled for all sessions and MUST be valid over all sessions of the multiplex.

sprop-init-buf-time:
In case of using Session multiplexing, the same sprop-init-buf-time value MUST be signaled for all sessions and MUST be valid over all sessions of the multiplex.

deint-buf-cap:
In case of using Session multiplexing, the same deint-buf-cap value MUST be signaled by the receiver for all sessions and MUST be valid over all sessions of the multiplex.

In addition the following optional MIME parameters apply:

sprop-scalability-info:
This parameter MAY be used to convey the NAL unit containing the scalability information SEI message as specified in [SVC]. The

parameter MUST NOT be used to indicate codec capability in any capability exchange procedure. The value of the parameter is the base64 representation of the NAL unit containing the scalability information SEI message.

sprop-layer-ids:
This parameter MAY be used to signal the layer identification value(s), expressed by the value of the the second and the third byte of the SVC NAL unit header, for one or more SVC layer(s) conveyed in one RTP session. A layer identification is a three character value base64 coded. If more than one layer is transmitted within one RTP session, the layer identification value of each layer MUST be itemized with decreasing importance for decoding and MUST be comma-separated.

Encoding considerations:
This type is only defined for transfer via RTP (RFC 3550).

Security considerations:
See section 9 of RFC XXXX.

Public specification:

Please refer to section 15 of RFC XXXX.

Additional information:
None

File extensions: none
Macintosh file type code: none
Object identifier or OID: none
Person & email address to contact for further information:
Intended usage: COMMON
Author:
Change controller:
IETF Audio/Video Transport working group
delegated from the IESG.

9.2. SDP Parameters

9.2.1. Mapping of MIME Parameters to SDP

The MIME media type video/SVC string is mapped to fields in the Session Description Protocol (SDP) as follows:

* The media name in the "m=" line of SDP MUST be video.

* The encoding name in the "a=rtpmap" line of SDP MUST be SVC (the MIME subtype).

* The clock rate in the "a=rtpmap" line MUST be 90000.

* The OPTIONAL parameters "profile-level-id", "max-mbps", "max-fs", "max-cpb", "max-dpb", "max-br", "redundant-pic-cap", "sprop-parameter-sets", "parameter-add", "packetization-mode", "sprop-interleaving-depth", "deint-buf-cap", "sprop-deint-buf-req", "sprop-init-buf-time", "sprop-max-don-diff", "max-rcmd-nalu-size'', ''sprop-layer-ids'', and ''sprop-scalability-info'', when present, MUST be included in the "a=fmtp" line of SDP. These parameters are expressed as a MIME media type string, in the form of a semicolon separated list of parameter=value pairs.

9.2.2. Usage with the SDP Offer/Answer Model

TBD.

9.2.3. Usage with Session and SSRC multiplexing

If Session multiplexing is used, the rules on signaling media decoding dependency in SDP as defined in [SDPsiglay] apply.

9.2.4. Usage in Declarative Session Descriptions

TBD.

9.3. Examples

TBD.

9.4. Parameter Set Considerations

Please see section 10 of RFC 3984 [RFC3984].

10. Security Considerations

Please see section 11 of RFC 3984 [RFC3984].

11. Congestion Control

Within any given RTP session carrying payload according to this specification, the provisions of section 12 of RFC 3984 [RFC3984] apply. Reducing the session bandwidth is possible by one or more of the following means, listed in an order that, in most cases, will assure the least negative impact to the user experience:

a) removing some or all bits of a given FGS NAL unit as long as the remaining bits still form a conforming SVC NAL unit. Note: doing so does not reduce the number of NAL units, but the bit rate of the highest enhancement layer. This can be translated into a reduced packet count when aggregating those smaller NAL units into packets small enough to fit the MTU size.
b) stop sending NAL units belonging to the highest enhancement layer(s), when more than one layer is transported in the session.
c) dropping NAL units of the base layer according to their importance for the decoding process, as indicated in the NAL unit's NRI field (this may lead to a non-compliant bitstream, and annoying artifacts)
d) dropping NAL units or entire packets not according to the aforementioned rules (media-unaware stream thinning). This results in the reception of a non-compliant bitstream and, most likely, in very annoying artifacts

Informative note: The discussion above is centered on NAL units and not on packets, primarily because that is the level where senders can meaningfully manipulate the scalable bitstream. The mapping of NAL units to RTP packets is fairly flexible when using aggregation packets. Depending on the nature of the congestion control algorithm, the ''dimension'' of congestion measurement (packet count or bitrate) and reaction to it (reducing packet count or bitrate or both) can be adjusted accordingly.

When multiple sessions are SSRC multiplexed onto the same transport address, a receiver can still calculate and communicate in RTCP-RRs the per-session congestion. However, when it is known that these SSRC-multiplexed sessions originate from the same sender's transport address (a condition henceforth referred to as ''on the same path

All aforementioned means are available to the RTP sender, regardless whether that sender is located in the sending endpoint or in a mixer based MANE.

When a translator-based MANE is employed, then the MANE MAY manipulate the session only on the MANE's outgoing path, so that the sensed end-to-end congestion falls within the permissible envelope. As all translators, in this case the MANE needs to rewrite RTCP RRs to reflect the manipulations it has performed on the session.

12. IANA Consideration

[Edt. Note: A new MIME type should be registered from IANA.]

13. Informative Appendix: Application Examples

13.1. Introduction

Scalable video coding is a concept that has been around at least since MPEG-2 [MPEG2], which goes back as early as 1993. Nevertheless, it has never gained wide acceptance; perhaps partly because applications didn't materialize in the form envisioned during standardization.

MPEG and JVT, respectively, performed a requirement analysis before

the SVC project was launched. Dozens of scenarios have been studied. While some of the scenarios appear not to follow the most basic design principles of the Internet -- and are therefore not appropriate for IETF standardization -- others are clearly in the scope of IETF work. Of these, this draft chooses the following subset for immediate consideration. Note that we do not reference the MPEG and JVT documents directly; partly, because at least the

MPEG documents have a limited lifespan and are not publicly available, and partly because the language used in these documents is inappropriately video centric and imprecise, when it comes to protocol matters.

With these remarks, we now introduce three main application scenarios that we consider as relevant, and that are implementable with this specification.

13.2. Layered Multicast

This well-understood form of the use of layered coding [McCanne/Vetterli] implies that all layers are individually conveyed in their own RTP packet streams, each carried in its own RTP session using the IP (multicast) address and port number as the single demultiplexing point. Receivers ''tune'' into the layers by subscribing to the IP multicast, normally by using IGMP [IGMP].

Layered Multicast has the great advantage of simplicity and easy implementation. However, it has also the great disadvantage of utilizing many different transport addresses. While we consider this not to be a major problem for a professionally maintained content server, receiving client endpoints need to open many ports to IP multicast addresses in their firewalls. This is a practical problem from a firewall/NAT viewpoint. Furthermore, even today IP multicast is not as widely deployed as many wish.

We consider layered multicast an important application scenario for three reasons. First, it is well understood and the implementation constraints are well known. There may well be large scale IP networks outside the immediate Internet context that may wish to employ layered multicast in the future. One possible example could be a combination of content creation and core-network distribution for the various mobile TV services, e.g. those being developed by 3GPP (MBMS) [MBMS] and DVB (DVB-H) [DVB-H].

13.3. Streaming of an SVC scalable stream

In this scenario, a streaming server has a repository of stored SVC coded layers for a given content. At the time of streaming, and according to the capabilities, connectivity, and congestion

situation of the client(s), the streaming server generates and serves a scalable stream. Both unicast and multicast serving is possible. At the same time, the streaming server may use the same repository of stored layers to compose different streams (with a different set of layers) intended for other audiences.

As every endpoint receives only a single SVC RTP session, the number of firewall pinholes can be optimized to one.

The main difference between this scenario and straightforward simulcasting lies in the architecture and the requirements of the streaming server, and is therefore out of the scope of IETF standardization. However, compelling arguments can be made why such a streaming server design makes sense. One possible argument is related to storage space and channel bandwidth. Another is bandwidth adaptivity without transcoding -- a considerable advantage in a congestion controlled network. When the streaming server learns about congestion, it can reduce sending bitrate by choosing fewer layers or utilizing FGS, when composing the layered stream; see section 10. SVC is designed to gracefully support both bandwidth rampdown and bandwidth rampup with a considerable dynamic range. This payload format is designed to allow for bandwidth flexibility in the mentioned sense, both for CGS and FGS layers. While, in theory, a transcoding step could achieve a similar dynamic range, the computational demands are impractically high and video quality is typically lowered -- therefore, few (if any) streaming servers implement full transcoding.

13.4. Multicast to MANE, SVC scalable stream to endpoint

This scenario is a bit more complex, and designed to optimize the network traffic in a core network, while still requiring only a single pinhole in the endpoint's firewall. One of its key applications is the mobile TV market.

Consider a large private IP network, e.g. the core network of 3GPP. Streaming servers within this core network can be assumed to be professionally maintained. We assume that these servers can have many ports open to the network and that layered multicast is a real option. Therefore, we assume that the streaming server multicasts

INTERNET-DRAFT   RTP Payload Format for SVC Video   December 2006

SVC scalable layers, instead of simulcasting different representations of the same content at different bit rates.

Also consider many endpoints of different classes. Some of these endpoints may not have the processing power or the display size to meaningfully decode all layers; other may have these capabilities. Users of some endpoints may not wish to pay for high quality and are happy with a base service, which may be cheaper or even free. Other users are willing to pay for high quality. Finally, some connected users may have a bandwidth problem in that they can't receive the bandwidth they would want to receive -- be it through congestion, connectivity, change of service quality, or for whatever other reasons. However, all these users have in common that they don't want to be exposed too much, and therefore the number of firewall pinholes need to be small.

This situation can be handled best by introducing middleboxes close to the edge of the core network, which receive the layered multicast streams and compose the single SVC scalable bit stream according to the needs of the endpoint connected. These middleboxes are called MANEs throughout this specification. In practice, we envision the MANE to be part of (or at least physically and topologically close to) the base station of a mobile network, where all the signaling and media traffic necessarily are multiplexed on the same physical link. This is why we do not worry too much about decomposition aspects of the MANE as such.

MANEs necessarily need to be fairly complex devices. They certainly need to understand the signaling, so, for example, to associate the PT octet in the RTP header with the SVC payload type.

A MANE may terminate the multicasted layered RTP sessions incoming from the core network side, and create new RTP sessions (perhaps even multicast sessions) to the endpoints connected to them. In RTP terminology, these types of MANEs are RTP mixers. This implies, per RFC 3550, a very loose relationship between the incoming and outgoing RTP sessions. In particular, there is no direct relationship between the incoming and outgoing RTP sequence numbers, RTP timestamps, payload types used, etc.

Mixer-based MANEs are conceptually easy to implement and can offer powerful features, primarily because they necessarily can ''see'' the payload (including the RTP payload headers), utilize the wealth of

layering information available therein, and manipulate it.

While a mixer-based MANE operation in its most trivial form (combining multiple RTP packet streams into a single one) can be implemented comparatively simply -- reordering the incoming packets according to the DON and sending them in the appropriate order -- more complex forms can also be envisioned. For example, a mixer-type MANE can be optimizing the outgoing RTP stream to the MTU size of the outgoing path by utilizing the aggregation and fragmentation mechanisms of this memo.

A MANE can also act as a translator. In this case, we envision its functionality to stream thinning, so to adhere to congestion control principles as discussed in section 11. While the implementation of the forward (media) channel of such a MANE appears to be comparatively simple, the need to rewrite RTCP RRs makes even such a MANE a complex device.

While the implementation complexity of either case of a MANE, as discussed above, is fairly high, the computational demands are comparatively low. In particular, SVC and/or this specification contain means to easily generate the correct inter-layer decoding order of NAL units. It is also simple to identify the fine granularity scalable bits in a given NAL unit. No serious bit-oriented processing is required and no significant state information (beyond that of the signaling and perhaps the SVC sequence parameter sets) need to be kept.

13.5. Scenarios currently not considered for complexity reasons

-- vacat --

13.6. Scenarios currently not considered for being unaligned with IP philosophy

Remarks have been made that the current draft does not take into consideration at least one application scenario which some JVT folks consider important. In particular, their idea is to make the RTP payload format (or the media stream itself) self-contained enough that a stateless, non signaling aware device can ''thin'' an RTP session to meet the bandwidth demands of the endpoint. They call

this device a ''Router'' or ''Gateway'', and sometimes a MANE. Obviously, it's not a Router or Gateway in the IETF sense. To distinguish it from a MANE as defined in RFC 3984 and in this specification, let's call it a MDfH (Magic Device from Heaven).

To simplify discussions, let's assume point-to-point traffic only. The endpoint has a signaling relationship with the streaming server, but it is known that the MDfH is somewhere in the media path (e.g. because the physical network topology ensures this). It has been requested, at least implicitly through MPEG's and JVT's requirements document, that the MDfH should be capable to intercept the SVC scalable bit stream, modify it by dropping packets or parts thereof, and forwarding the resulting packet stream to the receiving endpoint. It has been requested that this payload specification contains protocol elements facilitating such an operation, and the argument has been made that the NRI field of RFC 3984 serves exactly the same purpose.

The authors of this I-D do not consider the scenario above to be aligned with the most basic design philosophies the IETF follows, and therefore have not addressed the comments made (except through this section). In particular, we see the following problems with the MDfH approach):

- As the very minimum, the MDfH would need to know which RTP streams are carrying SVC. We don't see how this could be accomplished but by using a static payload type. None of the IETF defined RTP profiles envision static payload types for SVC, and even the de-facto profiles developed by some application standard organizations (3GPP for example) do not use this outdated concept. Therefore, the MDfH necessarily needs to be at least ''listening'' to the signaling.
- If the RTP packet payload were encrypted, it would be impossible to interpret the payload header and/or the first bytes of the media stream. We understand that there are crypto schemes under

discussion that encrypt only the last n bytes of an RTP payload, but we are more than unsure that this is fully in line with the IETF's security vision.

Even if the above two problems would have been overcome through standardization outside of the IETF, we still foresee serious design flaws:

- An MDfH can't simply dump RTP packets it doesn't want to forward. It either needs to act as a full RTP Translator (implying that it

rewrites RTCP RRs and such), or it needs to patch the RTP sequence numbers to fulfill the RTP specification. Not doing either would, for the receiver, look like the gaps in the sequence numbers occurred due to unintentional erasures, which has interesting effects on congestion control (if implemented), will break pretty much every meta-payload ever developed, and so on. (Many more points could be made here).

- An MDfH also can't ''prune'' FGS packets. Again, doing so would not be compatible with meta payloads, and would mess up RTCP RRs and congestion control (if the congestion control is based on octet count and not on packet count; there are discussions related to the former at least in the context of TFRC).

In summary, based on our current knowledge we are not willing to specify protocol mechanisms that support an operation point that has so little in common with classic RTP use.

13.7. SSRC Multiplexing

The authors have complentated the idea of introducing SSRC multiplexing, i.e. allowing to send multiple RTP packet streams containing layers in the same RTP session, differentiated by SSRC values. Our intention was to minimize the number of firewall pinholes in an endpoint to one, by using MANEs to aggregate multiple outgoing sessions stemming from a server into a single session (with SSRC multiplexed packet streams). We were hoping that would be feasible even with encrypted packets in an SRTP context.

While an implementation along these lines indeed appears to be feasible for the forward media path, the RTCP RR rewrite cannot be implemented in the way necessary for this scheme to work. This

relates to the need to authenticate the RTCP RRs as per SRTP. While the RTCP RR itself does not need to be rewritten by the scheme we envisioned, it's transport addresses needs to be manipulated. This, in turn, is incompatible with the mandatory authetification of RTCP RRs. As a result, there would be an requirement that a MANE needs to be in the RTCP security context of the sessions, which was not envisioned in our use case.

As the envisioned use case cannot be implemented, we refrained to add the considerable document complexity to support SSRC multiplexing herein.

14. Acknowledgements

Funding for the RFC Editor function is currently provided by the Internet Society.

15. References

15.1. Normative References

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.

[MPEG4-10] ISO/IEC International Standard 14496-10:2003.

[H.264] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services", May 2003.

[SDPsiglay] Schierl, T., ''Signaling media decoding dependency in Session Description Protocol (SDP)'', IETF internet draft draft-schierl-mmusic-layered-codec-02, December 2007.

[SVC] Joint Video Team, ''Joint Scalable Video Model 8: Joint Draft 8 with proposed changes'', available from http://ftp3.itu.ch/av-arch/jvt-site /jvt-site/2006_10_Hangzhou/JVT-U202.zip , October 2006.

[RFC3984] Wenger, S., Hannuksela, M, Stockhammer, T, Westerlund, M, Singer, D, ''RTP Payload Format for H.264 Video'', RFC 3984, February 2005.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

15.2. Informative References

[DVB-H] DVB - Digital Video Broadcasting (DVB); DVB-H Implementation Guidelines, ETSI TR 102 377, 2005

[H.241] ITU-T Rec. H.241, ''Extended video procedures and control signals for H.300-series terminals'', May 2006

[IGMP] Cain, B., Deering S., Kovenlas, I., Fenner, B. and Thyagarajan, A., ''Internet Group Management Protocol, Version 3'', RFC 3376, October 2002.

[McCanne/Vetterli] V. Jacobson, S. McCanne and M. Vetterli. Receiver-driven layered multicast. In Proc. of ACM SIGCOMM'96, pages

117--130, Stanford, CA, August 1996.
[MBMS] 3GPP - Technical Specification Group Services and System Aspects; Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs (Release 6), December 2005.
[MPEG2] ISO/IEC International Standard 13818-2:1993.
[SRTP] Baugher, M., McGrew, D, Naslund, M, Carrara, E, Norrman, K, ''The secure real-time transport protocol (SRTP)'', RFC 3711, March 2004.

16. Author's Addresses

Stephan Wenger
Nokia Research Center
P.O. Box 100
FIN-33721 Tampere
Finland
Phone: +358-50-486-0637
Email: stewe@stewe.org

Ye-Kui Wang
Nokia Research Center
P.O. Box 100
FIN-33721 Tampere
Finland
Phone: +358-50-486-7004
Email: ye-kui.wang@nokia.com

Thomas Schierl
Fraunhofer HHI
Einsteinufer 37
D-10587 Berlin
Germany
Phone: +49-30-31002-227
Email: schierl@hhi.fhg.de

17. Intellectual Property Statement

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).
Further, the author Thomas Schierl of Fraunhofer HHI is sponsored by the European Commission under the contract number FP6-IST-0028097, project ASTRALS.

18. RFC Editor Considerations

none

19. Open Issues

1. Packetization rules need work.
   3. Alignment with the SVC specification (ongoing)

20. Changes Log

Version 00

- 29.08.2005, YkW: Initial version
- 29.09.2005, Miska: Reviewed and commented throughout the document
- 05.10.2006, StW: Editorial changes through the document, and formatted the document in RFC payload format style

From -00 to -01

- 04.02.2006, StW: Added details to scope
- 04.02.2006, StW: Added short subsection 6.1 ''Design Principles''
- 04.02.2006, StW: Added section 15, ''Application Examples''
- 06.02 - 03.03.2006, YkW: Various modifications throughout the document

- 13.02.2006 - 03.03.2006 , ThS: Added definitions and additional information to section 3.3, 5.1, 7 and 8, parameters in section 9.1 and added section 14 for NAL unit re-ordering for layered multicast. Further modifications throughout the document

From -01 to -02

- 06.03.2006, StW: Editorial improvements
- 26.05.2006, YkW: Updated NAL unit header syntax and semantics according to the latest draft SVC spec
- 20.06.2006, Miska/YkW: Added section 6.10 ''Payload Content Scalability Information (PACSI) NAL Unit''
- 20.06.2006, YkW: Updated the NAL unit reordering process for layered multicast (removed the old section 14 ''Informative Appendix: NAL Unit Re-ordering for Layered Multicast'' and added the new section 13 ''NAL Unit Reordering for Layered Multicast'')

From -02 to -03

- 05.09.2006, YkW: Updated the NAL unit header syntax, definitions, etc., according to the foreseen July JVT output. Updated possible MANE adaptation operations according to SPID, TL, DID and QL. Clarified the removal of single NAL unit packetiztaion mode. Added the support of SSRC multiplexing in layered multicast.
- 08.09.2006, StW: Editorial changes throughout the document
- 08.09.2006, YkW: Added the packetization rule for suffix NAL unit.
- 19.09.2006, YkW: Moved/updated SSRC multiplexing support to section 6.2 ``RTP header usage''. Moved/updated the cross layer DON constraint to Section 6.6 ``Decoding order number''. Moved/updated the

packetization rule when a SVC bistream is transported over more than one RTP session to Section 7 ``Packetization rules''. Removed Section 13 ''Support of layered multicast''.
- 16.10, TS: Added detailed four-byte NAL unit header description. Change ''AVC'' to ''H.264'' conforming to 3984. Modifications throughout the document. Extended description of 3rd byte of PACSI NAL unit. Corrected terms RTP session and RTP packet stream in case of SSRC multiplexing. Added terms in definition section on RTP multiplexing. Constraints on optional MIME parameters of 3984 for cross-layer DON (DON section and MIME parameters). Copied parts of SI paper regarding mixer, translator and SSRC mux with SRTP to section application examples. Added section on SDP usage with Session and SSRC

multiplexing. Added points in Design principles on translator/mixer and RTP multiplexing. Added additional founding information in Ack-section. Corrected reference for SVC and added reference for generic signaling.
17.10, StW: Fixed many editorials, clarified MANE, mixer, translator and RTP packet stream throughout doc (hopefully consistently)
18.10., removed comments, clarified B-Bit, changed definition of base-layer (do not need to be of the lowest temporal resolution),

From -03 to draft-ietf-avt-rtp-svc-00

- 23.11.06, StW: Editorials throughout the memo
- 23.11.06, StW: removed all occurrences of the security discussions, as they are incorrect. When using SRTP, the RTCP is authenticated, implying that a translator cannot rewrite RTCP RRs, implying that RRs would be incorrect as soon as the session is modified (i.e. packets are being removed), implying that SSRC-mux does not work in multicast.
- 23.11.06, StW: rewrote congestion control
- 23.11.06, StW: removed application scenario related to SRTP, as this does not work (see above
- 23.11.06, StW: added informative reference to H.241
- 27/29.11.06, YkW: editorial changes throughout the document
- 27/29.11.06, YkW: alignment with the SVC specification
- 19.12.06, TS:
  TS: [SVC] is now the complete Joint Draft of H.264
  TS: Removed SSRC Multiplexing
  TS: Changed use cases for MANE as a translator
  TS: Editorials throughout the document, alignment with SVC spec.
- 20-28.12.06, StW/TS/YkW: editorial changes throughout the document