Audio/Video Transport WG
Internet Draft
Intended status: Standards track
Expires: January 2009
July 14, 2008

M.M. Hannuksela
Y.-K. Wang
Nokia

Session Multiplexing for SVC Video
draft-hannuksela-avt-rtp-svc-01.txt

Abstract

This memo describes two alternative methods for decoding order recovery of the Network Abstraction Layer (NAL) units carried in multiple RTP sessions for Scalable Video Coding (SVC), which is defined in Annex G of the ITU-T Recommendation H.264 video codec that is technically identical to Amendment 3 of ISO/IEC International Standard 14496-10. The methods apply when non-interleaved transmission of NAL units using the Single NAL Unit packetization mode or the Non-Interleaved packetization mode defined in RFC 3984 is in use.

Table of Contents

1. Introduction

Section 1 of draft-ietf-avt-rtp-svc-13 applies.

This memo specifies two alternative methods for decoding order recovery of NAL units carried in a non-interleaved manner in multiple RTP sessions, referred to as Multi-Session Transmission (MST). Either of these two introduced MST packetization modes could be used to replace those specified in draft-ietf-avt-rtp-svc-13.

2. Conventions

Section 2 of draft-ietf-avt-rtp-svc-13 applies.

3. Definitions and Abbreviations

3.1. Definitions

3.1.1. Definitions from the SVC Specification

Section 3.1.1 of draft-ietf-avt-rtp-svc-13 applies.

3.1.2. Definitions Specific to This Memo

Section 3.1.2 of draft-ietf-avt-rtp-svc-13 applies with the following addition.

access unit identifier (AUID): A variable that is derived for each access unit when the single NAL unit packetization mode or the non-interleaved packetization mode is in use in Multi-Session Transmission. The value of AUID is identical for all NAL units of an access unit regardless of the session in which the NAL units are conveyed in. The AUID values of consecutive access units differ regardless of which sessions are decoded, but there are no other constraints of AUID values.

3.2. Abbreviations

Section 3.2 of draft-ietf-avt-rtp-svc-13 applies with the following additions.

AUID: Access Unit Identifier

TSD: Timestamp Difference

4. RTP Payload Format

4.1. Design Principles

Section 5.1 of draft-ietf-avt-rtp-svc-13 applies.

4.2. RTP Header Usage

Section 5.2 of draft-ietf-avt-rtp-svc-13 applies.

4.3. Common Structure of the RTP Payload Format

Section 5.3 of draft-ietf-avt-rtp-svc-13 applies.

4.4. NAL Unit Header Usage

Section 5.4 of draft-ietf-avt-rtp-svc-13 applies.

4.5. Packetization Modes

Section 5.4 of RFC 3984 applies when MST is not in use. The packetization modes specified in Section 5.4 of RFC 3984 are also referred to as session packetization modes.

When MST is in use, the following applies in addition.

4.5.1. Packetization Modes for Multi-Session Transmission

This memo specifies two MST packetization modes for non-interleaved MST:

- o Non-interleaved AUID-based mode (NI-A)

- o Non-interleaved timestamp-difference-based mode (NI-TSD)

In the NI-A and NI-TSD modes, NAL units in each RTP session are transmitted in NAL unit decoding order.

NI-A or NI-TSD could be used instead of the MST packetization modes NI-T, NI-C, and NI-TC specified in draft-ietf-avt-rtp-svc-13. The NI-A and NI-TSD modes simplify the packetization rules compared to those of NI-T, NI-C, and NI-TC. In the NI-A and NI-TSD modes, senders need not add NAL units to the stream and receivers need not remove the added NAL units as must be done in the NI-T and NI-TC modes. Moreover, the NI-MTAP packet introduced for NI-T and NI-TC modes is not needed and hence one precious NAL unit type value (the last one left for use in RTP payload specifications after the introduction of the PACSI NAL unit in the SVC draft) is saved for future extensions. The decoding order recovery process for the NI-A and NI-TSD modes does not require the reception and processing of RTCP sender reports, which makes the decoding order recovery process more straightforward compared to that of the NI-T mode.

The operation of the NI-A mode is very similar to the NI-TSD mode - the only difference being how access units are identified. The NI-A mode labels each access unit with an identifier, while the NI-TSD mode identifies access units with their RTP timestamp, which is indicated relative to the current packet in order to avoid dependencies on the random initial RTP timestamp. However, when the NI-TSD mode is in use, the same initial RTP timestamp offset MUST be used in each associated RTP session as proposed in [I-D.lennox-avt-rtp-layered-encoding-timestamps]. As the NI-TSD mode leaves less implementation freedom for senders and hence reduces the likelihood of ill-behaving sender implementations, it is the preferred mode proposed in this memo. However, as the usage of the same initial RTP offset in all sessions as proposed in [I-D.lennox-avt-rtp-layered-encoding-timestamps] has not been agreed yet, we included both NI-A and NI-TSD in this memo.

This memo does not specify any MST mode for interleaved transmission, which would allow transmission of NAL units out of NAL unit decoding order in each RTP session.

The MST packetization mode in use is signaled by the pmode media type parameter or by external means.

The used MST packetization mode governs which session packetization modes are allowed in the involved RTP sessions, which in turn govern which NAL unit types are allowed as RTP payloads.

Table 3.1 summarizes the allowed session packetization modes for the NI-A and NI-TSD MST packetization modes.

Table 3.1 Summary of allowed session packetization modes for the NI-A and NI-TSD MST packetization modes (yes = allowed, no = disallowed)

Session-Specific Mode	Base Session	Enhancement Session
Single NAL Unit Mode	yes	no
Non-Interleaved Mode	yes	yes
Interleaved Mode	no	no

Table 3.2 summarizes the allowed packet payload types for each allowed session packetization mode of the NI-A and NI-TSD MST packetization modes.

Table 3.2 Summary of allowed packet payload types for each session packetization mode of the NI-A and NI-TSD MST packetization modes (yes = allowed, no = disallowed, ig = ignore)

Packet Payload Type	Packet Type	Single NAL Unit Mode	Non-Interleaved Mode
0	undefined	ig	ig
1-23	NAL unit	yes	yes
24	STAP-A	no	yes
25	STAP-B	no	no
26	MTAP16	no	no
27	MTAP24	no	no
28	FU-A	no	yes
29	FU-B	no	no (base session) yes (enh. session)
30	PACSI	yes	yes
31	undefined	ig	ig

Informative note: FU-B are allowed in the enhancement session as specified in Section 4.9.

The packet payload type values indicated as undefined in Table 3.2 are reserved for future extensions. NAL units of those type values SHOULD NOT be sent by a sender (as packet payloads in single NAL unit packets or aggregation units in aggregation packets, or in FU packets) and MUST be ignored by a receiver. Note that NAL unit types 30 and 31 are indicated as undefined in RFC 3984, therefore RFC 3984 receivers MUST ignore NAL units of these types, if present.

4.6. Decoding Order Number (DON)

Section 5.5 of [RFC3984] applies when MST is not in use.

4.7. Identification of Access Units for Decoding Order Recovery in Multi-Session Transmission

The decoding order recovery process in the NI-A and NI-TSD MST packetization modes proposed in this memo consists of three steps. First, a set of candidate access units is formed by including the next access unit in transmission order (in relation to the access unit that has just been processed) in each of the sessions. Second,

for each candidate access unit, the previous access unit in decoding order in the same or a lower session is identified by information in the associated PACSI NAL unit or FU-B NAL unit. In the NI-A mode, the Access Unit Identifier is used for the identification of the previous access unit. In the NI-TSD mode, the signed timestamp difference between the current access unit and the previous access unit in the same or a lower session is indicated. Third, the next access unit in decoding order is the access unit in the highest session among the candidate access units for which the indicated previous access unit is not a candidate access unit.

4.7.1. Access Unit Identifier (AUID) for the NI-A Mode

When the NI-A MST packetization mode is in use, the packetization of each session MUST be as specified in Section 5.1. and the following applies.

The NI-A mode uses two fields, AUID and PAUID, for the recovery of the decoding order of NAL units. AUID and PAUID are conveyed in PACSI NAL units or in FU-B packets. AUID and PAUID MUST be conveyed in at least one PACSI NAL unit or FU-B packet for each access unit in each session.

AUID indicates the access unit identifier. The AUID value for all NAL units having the same NALU-time MUST be identical. The AUID value for consecutive access units in any set of sessions in the session dependency order MUST differ.

PAUID indicates the access unit identifier of the previous access unit in decoding order among the session containing the packet including the PAUID field and the sessions below it in the session dependency hierarchy specified according to [I-D.ietf-mmusic-decoding-dependency].

AUID and PAUID are 8-bit unsigned integers.

4.7.2. Timestamp Difference (TSD) for the NI-TSD Mode

When the NI-TSD MST packetization mode is in use, the packetization of each session MUST be as specified in Section 5.1. and the following applies.

The NI-TSD mode uses the RTP timestamp and one field, TSD, for the recovery of the decoding order of NAL units. TSD is conveyed in PACSI NAL units or in FU-B packets. TSD MUST be conveyed in at least one PACSI NAL unit or FU-B packet for each access unit in each session.

The TSD field SHALL be set as follows:

TSD = (TS(p) - TS(c)) / AUTICK, when abs(TS(p) - TS(c)) <= 2^31

TSD = (TS(p) - 2^32 - TS(c)) / AUTICK, when TS(p) - TS(c) > 2^31

TSD = (2^32 - TS(p) - TS(c)) / AUTICK, when TS(c) - TS(p) > 2^31

where TS(p) is the RTP timestamp of the previous access unit containing NAL units within this session (conveying the TSD field), TS(c) is the RTP timestamp of the current access unit (conveying the TSD field), and AUTICK is the value of the sprop-au-tick media type parameter.

> Informative note: The second and third equation above cover the cases where TS(c) and TS(p), respectively, have wrapped over the maximum value for 32-bit unsigned integer, while the first equation covers the cases where neither of TS(p) and TS(c) have wrapped over.

TSD is a 16-bit signed integer.

4.8. Aggregation Packets

Section 5.6 of draft-ietf-avt-rtp-svc-13 applies.

4.9. Fragmentation Units (FUs)

Section 5.7 of draft-ietf-avt-rtp-svc-13 applies with the following modifications.

When fragmentation units are used in the NI-A mode, FU-B MUST be used in enhancement sessions for the first fragmentation unit of a fragmented NAL unit. The DON field of the FU-B header in enhancement sessions is replaced by the AUID field followed by the PAUID field. The AUID field MUST be equal to the AUID value for the access unit containing the fragmented NAL unit. The semantics of the PAUID field are specified in Section 4.7.1.

When fragmentation units are used in the NI-TSD mode, FU-B MUST be used in enhancement sessions for the first fragmentation unit of a fragmented NAL unit. The DON field of the FU-B header in enhancement sessions is replaced by the TSD field. The semantics of the TSD field are specified in Section 4.7.2.

4.10. Payload Content Scalability Information (PACSI) NAL Unit

Section 5.8 of draft-ietf-avt-rtp-svc-13 applies with the following modifications.

4.10.1. PACSI NAL Unit Modifications for the NI-A Mode

The DONC field is replaced by the AUID field followed by the PAUID field.

The semantics of DONC are removed.

The occurrences of "DONC" are replaced with "AUID and PAUID".

The semantics of AUID and PAUID are specified as follows.

- o When present, the field AUID indicates the access unit identifier for all the NAL units in the aggregation packet (when the PACSI NAL unit is included in an aggregation packet) or the AUID of the next non-PACSI NAL unit in transmission order (when the PACSI NAL unit is included in a single NAL unit packet). The constraints in Section 4.7.1. apply for the AUID.

- o The semantics of the PAUID field are specified in Section 4.7.1.

4.10.2. PACSI NAL Unit Modifications for the NI-TSD Mode

The DONC field is replaced by the TSD field.

The semantics of DONC are removed.

The occurrences of "DONC" are replaced with "TSD".

The semantics of TSD are specified in Section 4.7.2.

5. Packetization Rules

Section 6 of draft-ietf-avt-rtp-svc-13 applies.

5.1. Packetization Rules for Multi-Session Transmission

When MST is used, decoding order recovery for NAL units carried in the associated RTP sessions is needed. The following packetization rules ensure that decoding order of NAL units carried in the associated sessions can be correctly recovered for each of the MST packetization modes according to the de-packetization process specified in Section 6.1. .

5.1.1. NI-A and NI-TSD MST Packetization Rules

When the NI-A or NI-TSD mode is in use, the following applies.

- For each single NAL unit packet containing a non-PACSI NAL unit, if present, the previous packet MUST have the same RTP timestamp as the single NAL unit packet, and the following applies.

  If the NALU-time of the non-PACSI NAL unit is not equal to the NALU-time of the previous non-PACSI NAL unit in decoding order, the previous packet MUST contain a PACSI NAL unit containing the AUID and PAUID fields when the NI-A mode is in use or the TSD field when the NI-TSD mode is in use;

  Otherwise (the NALU-time of the non-PACSI NAL unit is equal to the NALU-time of the previous non-PACSI NAL unit in decoding order), the previous packet MAY contain a PACSI NAL unit containing the AUID and PAUID fields when the NI-A mode is in use or the TSD field when the NI-TSD mode is in use.

- For each STAP-A packet, if present, if the RTP timestamp is different from the RTP timestamp of the previous STAP-A packet, the first NAL unit in the STAP-A packet MUST be a PACSI NAL unit containing the AUID and PAUID fields when the NI-A mode is in use or the TSD field when the NI-TSD mode is in use.

- For each FU-A packet, if present, the previous packet MUST have the same RTP timestamp as the FU-A packet, and the following applies.

  If the FU-A packet is the start of the fragmented NAL unit, the following applies;

    If the NALU-time of the fragmented NAL unit is not equal to the NALU-time of the previous non-PACSI NAL unit in decoding order, the previous packet MUST contain a PACSI NAL unit containing the AUID and PAUID fields when the NI-A mode is in use or the TSD field when the NI-TSD mode is in use;

    Otherwise (the NALU-time of the fragmented NAL unit is equal to the NALU-time of the previous non-PACSI NAL unit in decoding order), the previous packet MAY contain a PACSI NAL unit containing the AUID and PAUID fields when the NI-A mode is in use or the TSD field when the NI-TSD mode is in use.

- For each single NAL unit packet containing a PACSI NAL unit, if present, the PACSI NAL unit MUST contain the AUID and PAUID fields when the NI-A mode is in use or the TSD field when the NI-TSD mode is in use.

5.1.2. Packetization rules for non-VCL NAL units

Section 6.1.4 of draft-ietf-avt-rtp-svc-13 applies.

5.1.3. Packetization rules for Prefix NAL units

Section 6.1.5 of draft-ietf-avt-rtp-svc-13 applies.

6. De-Packetization Process

For single-session transmission, where a single RTP session is used, the de-packetization process specified in Section 7 of [RFC3984] applies.

For multi-session transmission, where more than one RTP sessions are used to receive data from the same SVC bitstream, the de-packetization process is specified in Section 6.1.

6.1. De-Packetization Process for Multi-Session Transmission

6.1.1. Decoding Order Recovery for the NI-A Mode

The following process SHALL be applied when the NI-A mode is in use.

The decoding order recovery SHOULD start from an access unit where NAL units are present for the base session, herein referred to as access unit F. Any packets preceding the first received packet of access unit F in reception order SHOULD be discarded. The decoding order of NAL units of access unit F is specified below.

For subsequent access units to be ordered, the following applies. Let AUID(n) and PAUID(n) be the AUID and PAUID values, respectively, of the first access unit in decoding order containing data in session n. The first access unit in decoding order containing data in session n can be identified by the smallest value of RTP sequence number within session n (taking into account the potential wraparound of RTP sequence numbers) among those packets whose payloads have not been passed to the decoder yet. Let a set of sessions S consist of those values of n for which NAL units are present in the first access unit in decoding order containing data in session n but are not present in a higher session in the same access unit. In other words,

the set of sessions S contains the highest session of those access units that are candidates of being next in decoding order.

The next access unit in decoding order is the access unit with the greatest value of m, where PAUID(m) is not equal to AUID(i), where m is any value within the set of sessions S and i is any value less than m within the set of sessions S. In other words, the next access unit in decoding order is found by investigating the candidate access units in session dependency order from the highest session to the lowest session according to the highest session for which the candidate access units contain NAL units. The next access unit in decoding order is the first access unit in the above investigation order that is not indicated to follow any candidate access unit in a lower session in decoding order. The decoding order of NAL units of the access unit having AUID equal to AUID(m) is specified below.

> Informative note: In practical implementations, the set of sessions S can be formed by considering only those access units that have arrived within a certain inter-session jitter compensation period. Consequently, it may not be necessary to wait access units from all sessions to arrive at a particular time for decoding order recovery.

If several NAL units share the same value of AUID, the order in which NAL units are passed to the decoder is specified as follows:

- Collect all NAL units NU(y) associated with the same value of AUID.

- Place the collected NAL units in the session dependency order specified according to [I-D.ietf-mmusic-decoding-dependency] and then in the consecutive order of appearance within each session into an access unit while satisfying the NAL unit order rules in SVC access units as specified in [SVC] and summarized as an informative algorithm in Section 6.1.3.

6.1.1.1. Example 1 (Informative)

The example shown in Figure 1 refers to three RTP sessions A, B and C containing a multiplexed SVC bitstream. In the example, the dependency signaling [I-D.ietf-mmusic-decoding-dependency] indicates that Session A is the base RTP session, B is the first enhancement RTP session and depends on A, and C is the second RTP enhancement session and depends on A and B. In the example, Session A has the lowest frame rate and Session B and C have the same, but a higher frame rate (using a hierarchical prediction structure). Arbitrary values of AUID values have been used in the example.

Figure 1 shows an example for de-jitter buffering with different jitters present in the sessions, i.e. at buffering startup not all packets with the same timestamp are available in all the de-jittering buffers. Jitter between the sessions is first assumed to be compensated by removing all NAL units preceding NAL unit with AUID equal to 2 (TS[1]).

At the next step, the first access unit with data present in the base session is identified. In this example, it is the access unit with AUID euqal to 4 (TS[8]). The preceding access units (with AUID equal to 2 (TS[1]) and AUID equal to 5 (TS[3])) are removed. NAL units of access unit with AUID equal to 4 (TS[8]) are passed to the decoder in layer dependency order.

The next access unit (with AUID equal to 6 (TS[6])) has NAL units present in each session, hence it is selected as the next access unit to be decoded.

Within independent sessions the next NAL units in decoding order belong to the access unit with AUID equal to 8 (TS[5]) (in sessions B and C) and to access unit AUID equal to 9 (TS[12]) (in session A). As session B and session A are not the highest sessions for the access unit with AUID equal to 8 and 9, respectively, the set of sessions S consists of only one session and the access unit with AUID equal to AUID(C) is selected as the next access unit in decoding order.

The decoding order recovery process is then continues similarly for the following access units.

Decoding order and dependency of NAL units per received RTP session with different jitter in sessions at buffering startup time:

```
C: -------------(2,3)-(5,2)-(4,5)-(6,4)-(8,6)-(7,8)-(9,7)-
     |     |     |     |     |     |     |     |     |
B: -(1,a)-(3,1)-(2,3)-(5,2)-(4,5)-(6,4)-(8,6)-(7,8)-(9,7)-
     |     |                 |     |                 |
A: -------(3,a)-------------(4,3)-(6,4)-------------(9,6)-
----------------------------------------------------------->
TS: [4]   [2]   [1]   [3]   [8]   [6]   [5]   [7]   [12]


Key:
A, B, C               - RTP sessions
'( )'                 - (AUID, PAUID) a=any value in this example
'|'                   - indicates corresponding NAL units of the
                        same access unit AU(TS[..]) in the RTP
                        sessions
Integer values in '[]' - media Timestamp (TS), sampling time as
                        derived from RTP timestamps associated to
                        the access unit AU(TS[..]).
```

Figure 1 Example for MST with different jitter in session at startup

6.1.1.2. Example 2 (Informative)

The example shown in Figure 2 refers to three RTP sessions A, B and C containing a multiplexed SVC bitstream. In the example, the dependency signaling [I-D.ietf-mmusic-decoding-dependency] indicates that Session A is the base RTP session, B is the first enhancement RTP session and depends on A, and C is the second RTP enhancement session and depends on A and B. Sessions A, B and C represent different levels of temporal scalability. Arbitrary values of AUID values have been used in the example. The initial de-jittering is assumed to be tackled similarly as in the previous example and not illustrated in Figure 2.

At the beginning, the first access unit with data present in the base session is identified. In this example, it is the access unit with AUID euqal to 3 (TS[8]). The preceding access unit (with AUID equal to 2 (TS[3]) is removed.

The next NAL units in decoding order belong to access unit with AUID equal to 9, 5, and 1 for session A, B, and C respectively, hence AUID(A)=9, PAUID(A)=3, AUID(B)=5, PAUID(B)=3, AUID(C)=1, PAUID(C)=5. All three sessions are present in the set of sessions S. As PAUID(C) is equal to AUID(B), the access unit with AUID equal to AUID(C) is not selected as the next access unit in decoding order. As PAUID(B) is not equal to AUID(A), the access unit with AUID equal to AUID(B) is selected as the next access unit in decoding order.

The next NAL units in decoding order belong to access unit with AUID equal to 9, 8, and 1 for session A, B, and C respectively, hence AUID(A)=9, PAUID(A)=3, AUID(B)=8, PAUID(B)=9, AUID(C)=1, PAUID(C)=5. All three sessions are present in the set of sessions S. As PAUID(C) is not equal to AUID(B) or AUID(A), the access unit with AUID equal to AUID(C) is selected as the next access unit in decoding order. After that, access unit with AUID equal to 4 is selected similarly as the next in decoding order.

The next NAL units in decoding order belong to access unit with AUID equal to 9, 8, and 7 for session A, B, and C respectively, hence AUID(A)=9, PAUID(A)=3, AUID(B)=8, PAUID(B)=9, AUID(C)=7, PAUID(C)=8. All three sessions are present in the set of sessions S. As PAUID(C) is equal to AUID(B) and PAUID(B) is equal to AUID(A), the access unit with AUID equal to AUID(C) or AUID(B) is not selected as the next access unit in decoding order. As there is no session below session A, the access unit with AUID equal to AUID(A) is selected as the next access unit in decoding order.

The decoding order recovery process is then continues similarly for the following access units.

Decoding order and dependency of NAL units per received RTP session:

```
C: --(2,a)-------------(1,5)-(4,1)-------------(7,8)-(6,7)-

B: --------------(5,3)-------------------(8,9)-------------

A: --------(3,a)-------------------(9,3)-------------------
------------------------------------------------------------>
TS:  [3]   [8]   [6]   [5]   [7]   [12]  [10]  [9]   [11]


Key:
A, B, C                - RTP sessions
'( )'                  - (AUID, PAUID) a=any value in this example
'|'                    - indicates corresponding NAL units of the
                         same access unit AU(TS[..]) in the RTP
                         sessions
Integer values in '[]' - media Timestamp (TS), sampling time as
                         derived from RTP timestamps associated to
                         the access unit AU(TS[..]).
```

Figure 2 Example for MST with different jitter in session at startup

6.1.2. Decoding Order Recovery for the NI-TSD Mode

The following process SHALL be applied when the NI-TSD session-multiplexing packetization mode is in use.

The decoding order recovery SHOULD start from an access unit where NAL units are present for the base session, herein referred to as access unit F. Any packets preceding the first received packet of access unit F in reception order SHOULD be discarded. The decoding order of NAL units of access unit F is specified below.

For subsequent access units to be ordered, the following applies. Let TS(n) and TSD(n) be the RTP timestamp and TSD values, respectively, of the first access unit in decoding order containing data in session n. The first access unit in decoding order containing data in session n can be identified by the smallest value of RTP sequence number within session n (taking into account the potential wraparound of RTP sequence numbers) among those packets whose payloads have not been passed to the decoder yet. Let a set of sessions S consist of those values of n for which NAL units are present in the first access unit in decoding order containing data in session n but are not present in a higher session in the same access unit. In other words, the set of sessions S contains the highest session of those access units that are candidates of being next in decoding order.

The next access unit in decoding order is the access unit with the greatest value of m, where TS(m) + TSD(m) * AUTICK (where AUTICK is the value of the sprop-au-tick media type parameter) is not equal to TS(i), where m is any value within the set of sessions S and i is any value less than m within the set of sessions S. In other words, the next access unit in decoding order is found by investigating the candidate access units in session depedency order from the highest session to the lowest session according to the highest session for which the candidate access units contain NAL units. The next access unit in decoding order is the first access unit in the above investigation order that is not indicated to follow any candidate access unit in a lower session in decoding order. The decoding order of NAL units of the access unit having RTP timestamp equal to TS(m) is specified below.

> Informative note: In practical implementations, the set of sessions S can be formed by considering only those access units that have arrived within a certain inter-session jitter compensation period. Consequently, it may not be necessary to wait access units from all sessions to arrive at a particular time for decoding order recovery.

If several NAL units share the same value of RTP timestamp, the order in which NAL units are passed to the decoder is specified as follows:

- Collect all NAL units NU(y) associated with the same value of RTP timestamp.

- Place the collected NAL units in the session dependency order specified according to [I-D.ietf-mmusic-decoding-dependency] and then in the consecutive order of appearance within each session into an access unit while satisfying the NAL unit order rules in SVC access units as specified in [SVC] and summarized as an informative algorithm in Section 6.1.3.

6.1.2.1. Example 1 (Informative)

The video stream in this example is identical to that of Section 6.1.1.1.

The example shown in Figure 3 refers to three RTP sessions A, B and C containing a multiplexed SVC bitstream. In the example, the dependency signaling [I-D.ietf-mmusic-decoding-dependency] indicates that Session A is the base RTP session, B is the first enhancement RTP session and depends on A, and C is the second RTP enhancement session and depends on A and B. In the example, Session A has the lowest frame rate and Session B and C have the same, but a higher frame rate (using a hierarchical prediction structure).

Figure 3 shows an example for de-jitter buffering with different jitters present in the sessions, i.e. at buffering startup not all packets with the same timestamp are available in all the de-jittering buffers. Jitter between the sessions is first assumed to be compensated by removing all NAL units preceding NAL unit with TS[1].

At the next step, the first access unit with data present in the base session is identified. In this example, it is the access unit with TS[8]. The preceding access units (with TS[1] and TS[3]) are removed. NAL units of access unit with TS[8] are passed to the decoder in layer dependency order.

The next access unit (with TS[6]) has NAL units present in each session, hence it is selected as the next access unit to be decoded.

Within independent sessions the next NAL units in decoding order belong to the access unit with TS[5] (in sessions B and C) and to access unit with TS[12] (in session A). As session B and session A are not the highest sessions for the access unit with TS[5] and TS[12], respectively, the set of sessions S consists of only one session and the access unit with TS[5] is selected as the next access unit in decoding order.

The decoding order recovery process is then continues similarly for the following access units.

Decoding order and dependency of NAL units per received RTP session with different jitter in sessions at buffering startup time:

```
C: -------------(1)---(-2)--(-5)--(2)---(1)---(-2)--(-5)--
     |     |     |     |     |     |     |     |     |
B: -( )---(2)---(1)---(-2)--(-5)--(2)---(1)---(-2)--(-5)--
     |     |                 |     |                 |
A: -------(2)---------------(-6)--(2)---------------(-6)--
-----------------------------------------------------------> 
TS: [4]   [2]   [1]   [3]   [8]   [6]   [5]   [7]   [12]


Key:
A, B, C                - RTP sessions
'( )'                  - (TSD)
'|'                    - indicates corresponding NAL units of the
                         same access unit AU(TS[..]) in the RTP
                         sessions
Integer values in '[]' - media Timestamp (TS), sampling time as
                         derived from RTP timestamps associated to
                         the access unit AU(TS[..]).
```

Figure 3 Example for MST with different jitter in session at startup

6.1.2.2. Example 2 (Informative)

The video stream in this example is identical to that of Section 6.1.1.2.

The example shown in Figure 4 refers to three RTP sessions A, B and C containing a multiplexed SVC bitstream. In the example, the dependency signaling [I-D.ietf-mmusic-decoding-dependency] indicates that Session A is the base RTP session, B is the first enhancement RTP session and depends on A, and C is the second RTP enhancement session and depends on A and B. Sessions A, B and C represent different levels of temporal scalability. The initial de-jittering is assumed to be tackled similarly as in the previous example and not illustrated in Figure 4.

At the beginning, the first access unit with data present in the base session is identified. In this example, it is the access unit with TS[8]. The preceding access unit (with TS[3] is removed.

The next NAL units in decoding order belong to access unit with TS[12], TS[6], and TS[5] for sessions A, B, and C, respectively, hence TS(A)=12, TSD(A)=-4, TS(B)=6, TSD(B)=2, TS(C)=5, and TSD(C)=1. All three sessions are present in the set of sessions S. As TS(C) + TSD(C) = 5 + 1 = 6 = TS(B), the access unit with TS[5] is not selected as the next access unit in decoding order. As TS(B) + TSD(B) = 6 + 2 = 8 is not equal to TS(A), the access unit with TS[6] is selected as the next access unit in decoding order.

The next NAL units in decoding order belong to access unit with TS[12], TS[10], and TS[5] for sessions A, B, and C, respectively, hence TS(A)=12, TSD(A)=-4, TS(B)=10, TSD(B)=2, TS(C)=5, and TSD(C)=1. All three sessions are present in the set of sessions S. As TS(C) + TSD(C) = 5 + 1 = 6 is not equal to TS(A) or TS(B), the access unit with TS[5] is selected as the next access unit in decoding order. After that, access unit with TS[7] is selected similarly as the next in decoding order.

The next NAL units in decoding order belong to access unit with TS[12], TS[10], and TS[9] for sessions A, B, and C, respectively, hence TS(A)=12, TSD(A)=-4, TS(B)=10, TSD(B)=2, TS(C)=9, and TSD(C)=1. All three sessions are present in the set of sessions S. As TS(C) + TSD(C) = 9 + 1 = 10 = TS(B) and TS(B) + TSD(B) = 10 + 2 = 12 = TS(A), the access unit with TS[9] or TS[10] is not selected as the next access unit in decoding order. As there is no session below session A, the access unit with TS[12] is selected as the next access unit in decoding order.

The decoding order recovery process is then continues similarly for the following access units.

Decoding order and dependency of NAL units per received RTP session:

```
C: --(-2)--------------(1)---(-2)--------------(1)---(-2)-

B: --------------(2)---------------------(2)--------------

A: --------(-4)--------------------(-4)-------------------
----------------------------------------------------------->
TS:  [3]   [8]   [6]   [5]   [7]   [12]  [10]  [9]   [11]


Key:
0, 1, 2                - RTP sessions
'( )'                  - (TSD)
'|'                    - indicates corresponding NAL units of the
                         same access unit AU(TS[..]) in the RTP
                         sessions
Integer values in '[]' - media Timestamp (TS), sampling time as
                         derived from RTP timestamps associated to
                         the access unit AU(TS[..]).
```

Figure 4 Example for MST with different jitter in session at startup

6.1.3. Informative Algorithm for NI-A and NI-TSD Decoding Order Recovery within an Access Unit

Section 7.1.1.1 of draft-ietf-avt-rtp-svc-13 applies.

7. Payload Format Parameters

Section 8 of draft-ietf-avt-rtp-svc-13 applies.

7.1. Media Type Registration

Section 8.1 of draft-ietf-avt-rtp-svc-13 applies with the following modifications.

pmode:
This parameter signals the properties of a NAL unit stream carried in more than one RTP session using MST or the capabilities of a receiver implementation. When the value of pmode is equal to "NI-A", the NI-A mode MUST be used. When the value of pmode is equal to "NI-TSD", the NI-TSD mode MUST

be used. This parameter MUST NOT be present, when "packetization-mode" is present.

sprop-au-tick:
This parameter indicates the number of 90000-kHz clock ticks used as a multiplier in the NI-TSD mode. The parameter MUST NOT be present when pmode is not equal to "NI-TSD". If the parameter is not present and the NI-TSD mode is in use, sprop-au-tick is inferred to be equal to 1. The value of sprop-au-tick MUST be a positive integer.

7.2. SDP Parameters

Section 8.2 of draft-ietf-avt-rtp-svc-13 applies.

7.3. Examples

Section 8.3 of draft-ietf-avt-rtp-svc-13 applies.

7.4. Parameter Set Considerations

Section 8.4 of draft-ietf-avt-rtp-svc-13 applies.

8. Security Considerations

Section 9 of draft-ietf-avt-rtp-svc-13 applies.

9. Congestion Control

Section 10 of draft-ietf-avt-rtp-svc-13 applies.

10. IANA Consideration

Section 11 of draft-ietf-avt-rtp-svc-13 applies.

11. Informative Appendix: Application Examples

Section 12 of draft-ietf-avt-rtp-svc-13 applies.

12. References

12.1. Normative References

Section 13.1 of draft-ietf-avt-rtp-svc-13 applies with the following additions.

[I-D.ietf-avt-rtp-svc] Wenger, S., Wang, Y.-K., Schierl, T., and Eleftheriadis, A., "RTP payload format for SVC video", draft-ietf-avt-rtp-svc-13 (work in progress), July 2008.

[I-D.lennox]   Lennox, J., Schierl, T., and Ganesan S., "Real-Time Transport Protocol (RTP) Timestamps for Layered Encodings", draft-lennox-avt-rtp-layered-encoding-timestamps-00, June 2, 2008.

12.2. Informative References

Section 13.2 of draft-ietf-avt-rtp-svc-13 applies.

13. Authors' Addresses

Miska M. Hannuksela
Nokia Research Center
P.O. Box 1000
33721 Tampere
Finland

Phone: +358-7180-73151
EMail: miska.hannuksela@nokia.com

Ye-Kui Wang
Nokia Research Center
P.O. Box 1000
33721 Tampere
Finland

Phone: +358-50-466-7004
EMail: ye-kui.wang@nokia.com