Internet Engineering Task Force
Internet Draft
Document: draft-lshuo-avt-rtp-avsp2-00.txt
Expires: February 2008

L. Huo
Peking University
L. Wang
Beijing Univ. of P&T

August 2007

RTP Payload Format for AVS-P2 Video

Status of this Memo

Copyright Notice

Abstract

This memo specifies an RTP payload format for encapsulating AVS-P2 compressed video bit streams, as defined by the Audio Video Coding Standard Workgroup of China (AVS Workgroup). The payload format has wide applicability, as it supports applications from simple low bit-rate conversational usage, to Internet video streaming with interleaved transmission, to high bit-rate video-on-demand.

Table of Contents

1. Introduction

1.1 Overview of AVS-P2 Video Codec

This memo specifies an RTP payload specification for the video coding standard known as AVS-P2. The official name for AVS-P2 is "Information Technology - Advanced Audio and Video Coding Part 2: Video", which was defined by the Audio Video Coding Standard Workgroup of China (AVS Workgroup), and approved as GB/T 20090.2 -2006 by Standardization Administration of China and enacted on March 1, 2006 [1]. In this memo the AVS-P2 acronym is used for the codec and the standard.

The AVS-P2 video codec has a very broad application range that

covers all forms of digital compressed video from, low bit-rate Internet streaming applications to HDTV broadcast and Digital Cinema applications with nearly lossless coding. The overall performance of AVS-P2 is such that bit rate savings of more than 50% are reported,

when compared against MPEG-2. AVS-P2 has comparable compression performance with that of H.264/AVC£¬ however with a valuable feature of lower computational complexity [9]. AVS-P2 has been adopted by number of applications including Chinese IPTV operators, Mobile TV operators as well as digital terrestrial TV broadcasting operators.

AVS-P2 specification [1] defines the AVS-P2 bit stream syntax and specifies constraints that must be met by AVS-P2 conformant bit streams. It also specifies the complete process required to decode the bit stream. However, it does not specify the AVS-P2 compression algorithm, thus allowing for different ways to implement an AVS-P2 encoder.

AVS-P2 is a hybrid coding based on spatial and temporal prediction, 8x8 transform and entropy coding. It has one profile called Jizhun profile. In this profile, there are 4 levels, which are level 4.0, 4.2, 6.0 and 6.2, respectively.

The AVS-P2 bit stream is defined as a hierarchy of layers. This is conceptually similar to the notion of a protocol stack of networking protocols. The outermost layer is called the video sequence layer. The other layers are, picture, slice, macroblock and block. A video sequence begins with a sequence header, followed by a series of one or more coded pictures. Each picture begins with a picture header, followed by a series of one or more slices. A slice comprises one or more contiguous rows of macroblocks. Each macroblock consists of one 16x16 luma block and two 8x8 chroma blocks for 4:2:0 format and four 8x8 chroma blocks for 4:2:2.

AVS-P2 has Intra picture (I-picture), forward predicted picture (P-picture), and bi-directional predicted picture (B-picture). The prediction reference picture number is maximally two. The predictions are at the integer-pel resolution and quarter-pel resolution. It uses a 4-tap filter for half pel interpolation and a 4-tap filter for quarter pel interpolation. It uses in-loop deblocking. It uses two dimensional context adaptive variable length coding, 19 look-up tables are used. It uses 8x8 intra prediction from the upper row and left column pels, five prediction angles are used.

Each picture can be coded as an I-picture, P-picture, or B-picture. Random accessible point is defined in AVS-P2. The sequence header can occur repeatedly in the AVS-P2 video bit stream before any random access point.

In Jizhun profile, each sequence header, picture header and slice is considered a Coding Data Unit (CDU). A CDU is always byte-aligned and is defined as a unit that can be parsed (i.e., syntax decoded)

independently of other information in the same layer. The beginning of each CDU is signaled by an identifier called Start Code. Macroblocks and blocks are not CDUs and thus do not have a Start Code and are not necessarily byte-aligned.

The Start Code consists of four bytes. The first three bytes are the Start Code Prefix with the fixed value of 0x000001. The fourth byte is called the Start Code Data and it is used to indicate the type of the CDU that follows the Start Code. The Start Code is always byte-aligned and is transmitted in network byte order. To prevent accidental emulation of the Start Code in the coded bit stream, AVS-P2 defines an encapsulation mechanism that uses byte stuffing. There are also other types of CDUs defined in AVS-P2. See Table 1 (in Section 7) of AVS-P2 specification [1] for a complete list of CDUs and their corresponding Start Code Values.

1.2 Conventions used in this document

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 [2].

2. Scope

The applications of this memo include video telephony, video conferencing, Internet media streaming, IPTV, video-on-demand, etc.

3. Definitions and abbreviations

This memo uses definitions and abbreviations of AVS-P2 [1]. Additionally, the following definitions and abbreviations also apply to this specification.

3.1 Definitions

NAL unit (Network Abstract Layer Unit)
Before packetizing an AVS-P2 video bit stream using the RTP Payload Format defined in this memo, firstly the bit stream MUST be transformed into a NAL unit stream, i.e., mapping the CDU data between every two consecutive Start Code prefixes (0x000001) in the AVS-P2 video bit stream into a NAL unit. The detail definitions of NAL unit and the transformation from AVS-P2 video bitstream into NAL unit stream are described in Section 4 of this memo.

NAL unit stream
A sequence composed of one or more NAL units.

NAL unit decoding order
The order of NALUs in a NAL unit stream.

Decoding order number (DON)
A field in the RTP payload structure or a derived variable indicating NAL unit decoding order. Values of DON are in the

range of 0 to 65535, inclusive. After reaching the maximum value, the value of DON wraps around to 0.

Transmission order
The order of packets in ascending RTP sequence number order (in modulo arithmetic). Within an aggregation packet, the NAL unit transmission order is the same as the order of appearance of NAL units in the packet.

Access Unit
A series of NAL units which constitute a frame of coded picture. Besides picture header and slice data, an access unit can also contain other types of coding data. All data within the same access unit MUST have the same timestamp value for RTP packetization.

Media Aware Network Element (MANE)
A network element, such as a middlebox or application layer gateway that is capable of parsing certain aspects of the RTP payload headers or the RTP payload and reacting to the contents.

3.2 Abbreviation

DON:	Decoding Order Number
DONB:	Decoding Order Number Base
DOND:	Decoding Order Number Difference
FEC:	Forward Error Correction
FU:	Fragmentation Unit
MTAP:	Multi-Time Aggregation Packet
MTAP16:	MTAP with 16-bit timestamp offset
MTAP24:	MTAP with 24-bit timestamp offset
MTU:	Maximum Transfer Unit
NAL:	Network Abstraction Layer
NALU:	NAL Unit
PSI£º	Payload Structure Indicator
RTP:	Real-time Transport Protocol
STAP:	Single-Time Aggregation Packet
STAP-A:	STAP type A
STAP-B:	STAP type B

4. NAL Unit

The syntax of NAL unit used in this memo resembles the one defined for H.264/AVC in IETF RFC 3984 [10]. An NAL unit is composed of two parts: NAL unit header and NAL unit data. The NAL unit header consists of exactly one byte, while the NAL unit data consists of a series of one or more bytes.

The conversion process from AVS-P2 video bit stream to NAL unit stream is as follows: first map the CDU data between every two consecutive Start Code Prefixes (0x000001) in the AVS-P2 video bit

stream (including the first Start Code Value but excluding Start Code Prefixes) into the NAL unit data, and then insert a NAL unit header before the NAL unit data according the Start Code Value and its context.

The format of NAL unit header is shown in Figure 1.

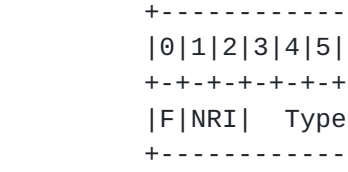


Figure 1. NAL unit header format

The syntax and semantics of the NAL unit are as follows:

F: 1 bit
Forbidden zero bit, its value SHOULD be 0.

NRI: 2 bit
NAL Reference Identification (nal_ref_idc). Value of non-zero means that the data contained in this NAL unit is sequence header or reference frame data. Value of 0 means that the data contained in this NAL unit is not reference frame data. For sequence header NAL unit, nal_ref_idc SHOULD NOT be 0. For a certain frame, if nal_ref_idc of one NAL unit's is 0, then nal_ref_idc of all NAL units in the same frame SHOULD be 0. Nal_ref_idc of NAL units for I frames SHOULD NOT be 0.

Type: 5 bit
NAL unit type (nal_unit_type). The value of this field is decided according to the start code value and the picuture header contained in the following NAL unit data, and their context, as shown in Table 1.

Table 1. Value assignment for the NAL unit type field in NAL unit header

NALU type	Corresponding CDU in NALU data	Reason for type assignment
0	reserved	
1	Sequence header	Start code value is 0xB0
2	Video extension	Start code value is 0xB5
3	User data	Start code value is 0xB2
4	Video edit	Start code value is 0xB7
5	Picture header of I frame	Start code value is 0xB3
6	Picture header of P frame	Start code value is 0xB6£¬and the picture coding type of the picture header is 01b
7	Picture header	Start code value is 0xB6£¬and the

	of B frame	picture coding type of the picture header is 10b
8	Slice data	Start code value is 0x00~0xAF£¬and the

	of I frame	start code value of the last picture header before this NALU is 0xB3
9	Slice data of P frame	Start code value is 0x00~0xAF£¬and the start code value of the last picture header before this NALU is 0xB6, and the picture coding type of the picture header is 01b
10	Slice data of B frame	Start code value is 0x00~0xAF£¬and the start code value of the last picture header before this NALU is 0xB6, and the picture coding type of the picture header is 10b
11-23	reserved	
24-31	undefined	

When the decoder receives NAL unit stream, before decoding, it MUST discard every NAL unit header of an NAL unit, and then insert a Start Code Prefix (0x000001) in the same position to transform the NAL unit stream back into an AVS-P2 video bit stream.

5 RTP Payload Format

5.1 RTP Header Usage

The RTP header format is defined in IETF RFC 3550 [3] as shown in Figure 2.

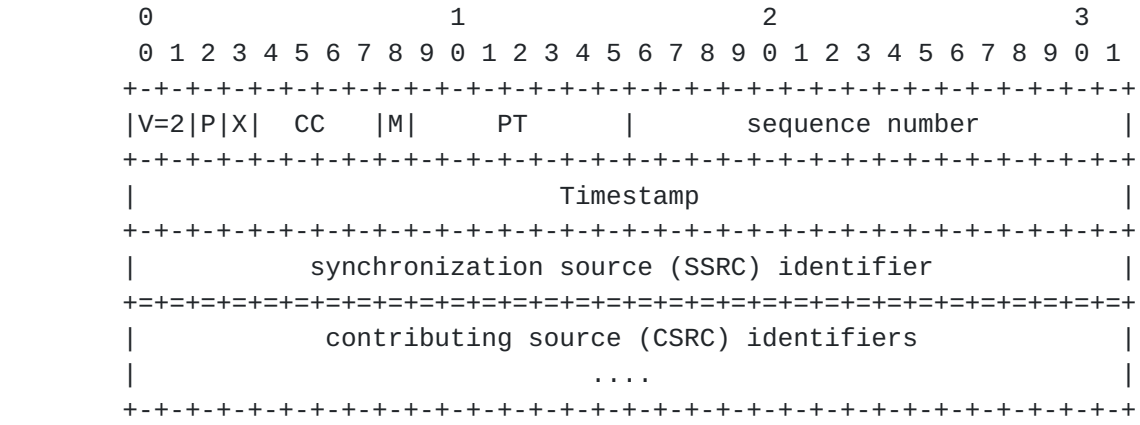


Figure 2. RTP header according to RFC 3550

For the usage of RTP header, this document obeys the same rules define in RFC 3550, except for some enhancements for the M and Timestamp fields:

Marker bit (M): 1 bit
Set for the very last packet of the access unit indicated by the RTP timestamp, in line with the normal use of the M bit in video

formats, to allow an efficient playout buffer handling. For aggregation packets (STAP and MTAP), the marker bit in the RTP header MUST be set to the value that the marker bit of the last

NAL unit of the aggregation packet would have been if it were transported in its own RTP packet. Decoders MAY use this bit as an early indication of the last packet of an access unit, but MUST NOT rely on this property.

Timestamp: 32 bit
The RTP timestamp is set to the sampling timestamp of the content. A 90 kHz clock rate MUST be used, which means that the time unit of RTP timestamp is 1/90000 second. If the NAL unit has no timing properties of its own, such as sequence header NALU, its RTP timestamp is set to the timestamp of the first following coded picture after it. The setting of the RTP Timestamp for MTAPs is defined in section 5.6.2.

5.2 Common structure of the RTP Payload format

The document defines three different basic payload structures, while each may be further divided into different sub-types.

Single NALU packet
Contains one and only one NAL unit in a single RTP payload.

Aggregation packet
Multiple NAL units are aggregated into a single RTP payload. This packet exists in four versions, the Single-Time Aggregation Packet type A (STAP-A), the Single-Time Aggregation Packet type B (STAP-B), Multi-Time Aggregation Packet (MTAP) with 16-bit offset (MTAP16), and Multi-Time Aggregation Packet (MTAP) with 24-bit offset (MTAP24).

Fragmentation Units (FUs)
Used to fragment a single NAL unit over multiple RTP packets. Exists with two versions, FU-A and FU-B.

Different payload structures are identified by the first byte of the RTP payload, which is called PSI (payload structure indication). PSI has the same format with NALU header, as shown in Figure 3.

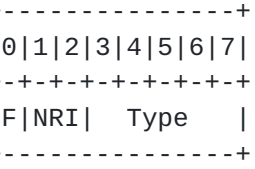


Figure 3: Format of PSI

F: 1bit
Value 0 means that there are no bit errors and other semantic errors in the RTP payload. Value 1 means that there may be bit

errors and other semantic errors in RTP payload. If bit error is found in RTP payload, MANE SHOULD set F to 1.

NRI: 2bit
Besides the rules defined in Section 4 for NALU, NRI value in PSI indicates the relative transmission priority of RTP packet. MANE can use this information to protect more important RTP packet. The highest priority is 11b, followed by 10b, and then 01b, and 00b. The NRI value for sequence header and I pictures SHOULD be set to 11b.

Type: 5bit
Indicate the RTP payload structure, as shown in Table 2.

Table 2. Summary of RTP payload structure types

Type	payload struture	explain	Section
0	Undefined		-
1-23	Single NALU	Single NAL unit packet	5.5
24	STAP-A	Single-time aggregation packet - A	5.6.1
25	STAP-B	Single-time aggregation packet - B	5.6.1
26	MTAP16	Multi-time aggregation packet with 16 bit offset	5.6.2
27	MTAP24	Multi-time aggregation packet with 24 bit offset	5.6.2
28	FU-A	Fragmentation unit - A	5.7
29	FU-B	Fragmentation unit - B	5.7
30-31	Undefined		-

5.3 Packetization Modes

This payload format specifies three cases of packetization modes: Single NAL unit mode, Non-interleaved mode and Interleaved mode. In the Single NAL unit mode or the non-interleaved mode, NAL units are transmitted in NAL unit decoding order. The interleaved mode allows transmission of NAL units out of NAL unit decoding order.

The used packetization mode governs which payload structures are allowed in RTP payloads. The packetization mode in use MAY be signaled by the value of the OPTIONAL packetization-mode media type parameter or by external means. Table 3 summarizes the allowed payload structures for each packetization mode. Some payload structures values are reserved for future extensions.

Table 3. Summary of allowed payload structures for each packetization mode (yes = allowed, no = disallowed, ig = ignore)

PSI Type	payload structure	Single NALU mode	Non-Interleaved mode	Interleaved mode

0	Undefined	ig	ig	ig
1-23	NALU	yes	yes	no

24	STAP-A	no	yes	no
25	STAP-B	no	no	yes
26	MTAP16	no	no	yes
27	MTAP24	no	no	yes
28	FU-A	no	yes	yes
29	FU-B	no	no	yes
30-31	Undefined	ig	ig	ig

5.4 Decoding Order Number

In the interleaved packetization mode, the transmission order of NAL units is allowed to differ from the decoding order of the NAL units. Decoding order number (DON) is a field in the payload structure or a derived variable that indicates the NAL unit decoding order.

The coupling of transmission and decoding order is controlled by the OPTIONAL sprop-interleaving-depth media type parameter as follows. When the value of the OPTIONAL sprop-interleaving-depth parameter is equal to 0 (explicitly or per default) or transmission of NAL units out of their decoding order is disallowed by external means, the transmission order of NAL units MUST conform to the NAL unit decoding order. When the value of the OPTIONAL sprop-interleaving-depth parameter is greater than 0 or transmission of NAL units out of their decoding order is allowed by external means,

- the order of NAL units in an MTAP16 and an MTAP24 is NOT REQUIRED to be the NAL unit decoding order, and

- the order of NAL units generated by decapsulating STAP-Bs, MTAPs, and FUs in two consecutive packets is NOT REQUIRED to be the NAL unit decoding order.

The RTP payload structures for a single NAL unit packet, an STAP-A, and an FU-A do not include DON. STAP-B and FU-B structures include DON, and the structure of MTAPs enables derivation of DON as specified in section 5.6.2.

In the single NAL unit packetization mode, the transmission order of NAL units, determined by the RTP sequence number, MUST be the same as their NAL unit decoding order. In the non-interleaved packetization mode, the transmission order of NAL units in single NAL unit packets, STAP-As, and FU-As MUST be the same as their NAL unit decoding order. The NAL units within an STAP MUST appear in the NAL unit decoding order. Thus, the decoding order is first provided through the implicit order within a STAP, and second provided through the RTP sequence number for the order between STAPs, FUs, and single NAL unit packets.

Signaling of the value of DON for NAL units carried in STAP-B, MTAP,

and a series of fragmentation units starting with an FU-B is specified in sections 5.6.1, 5.6.2, and 5.7, respectively. The DON value of the first NAL unit in transmission order may be set to any

value. Values of DON are in the range of 0 to 65535, inclusive. After reaching the maximum value, the value of DON wraps around to 0.

The decoding order of two NAL units contained in any STAP-B, MTAP, or a series of fragmentation units starting with an FU-B is determined as follows. Let DON(i) be the decoding order number of the NAL unit having index i in the transmission order. Function don_diff(m,n) is specified as follows:

If DON(m) == DON(n), don_diff(m,n) = 0

If (DON(m) < DON(n) and DON(n) - DON(m) < 32768), don_diff(m,n) = DON(n) - DON(m)

If (DON(m) > DON(n) and DON(m) - DON(n) >= 32768), don_diff(m,n) = 65536 - DON(m) + DON(n)

If (DON(m) < DON(n) and DON(n) - DON(m) >= 32768), don_diff(m,n) = - (DON(m) + 65536 - DON(n))

If (DON(m) > DON(n) and DON(m) - DON(n) < 32768), don_diff(m,n) = - (DON(m) - DON(n))

When don_diff(m,n) is equal to 0, then the NAL unit decoding order of the two NAL units can be in either order. A positive value of don_diff(m,n) indicates that the NAL unit having transmission order index n follows, in decoding order, the NAL unit having transmission order index m. A negative value of don_diff(m,n) indicates that the NAL unit having transmission order index n precedes, in decoding order, the NAL unit having transmission order index m.

Values of DON related fields (DON, DONB, and DOND; see section 5.6) MUST be such that the decoding order determined by the values of DON, as specified above, conforms to the NAL unit decoding order. If the order of two NAL units in NAL unit decoding order is switched and the new order does not conform to the NAL unit decoding order, the NAL units MUST NOT have the same value of DON. If the order of two consecutive NAL units in the NAL unit stream is switched and the new order still conforms to the NAL unit decoding order, the NAL units may have the same value of DON. Consequently, NAL units having the same value of DON can be decoded in any order, and two NAL units having a different value of DON SHOULD be passed to the decoder in the order specified above. When two consecutive NAL units in the NAL unit decoding order have a different value of DON, the value of DON for the second NAL unit in decoding order SHOULD be the value of DON for the first, incremented by one.

5.5 Single NAL Unit Packet

The single NAL unit packet MUST contain one and only one NAL unit as defined in Section 4. This means that neither an aggregation packet nor a fragmentation unit can be used within a single NAL unit

packet. In the payload structure of single NAL unit, the first byte of the RTP payload (i.e., PSI) co-serves as the NALU header which is directly followed by the NALU data, as shown in figure 4.

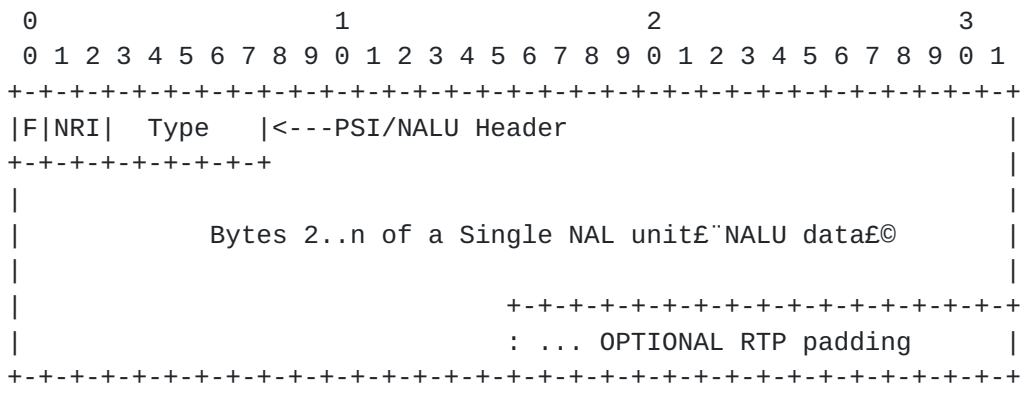


Figure 4. Single NAL Unit Packet RTP payload format

5.6 Aggregation Packets

Two types of aggregation packets are defined in this document:

- Single-time aggregation packet (STAP): aggregates NAL units with identical NALU-time. Two types of STAPs are defined, one without DON (STAP-A) and another including DON (STAP-B).

- Multi-time aggregation packet (MTAP): aggregates NAL units with potentially differing NALU-time. Two different MTAPs are defined, differing in the length of the NAL unit timestamp offset: MTAP16 and MTAP24.

The term NALU-time is defined as the value that the RTP timestamp would have if that NAL unit would be transported in its own RTP packet. Each NAL unit to be carried in an aggregation packet is encapsulated in an aggregation unit. Please see below for the four different aggregation units and their characteristics. Figure 5 shows the structure of the RTP payload format for aggregation packets.

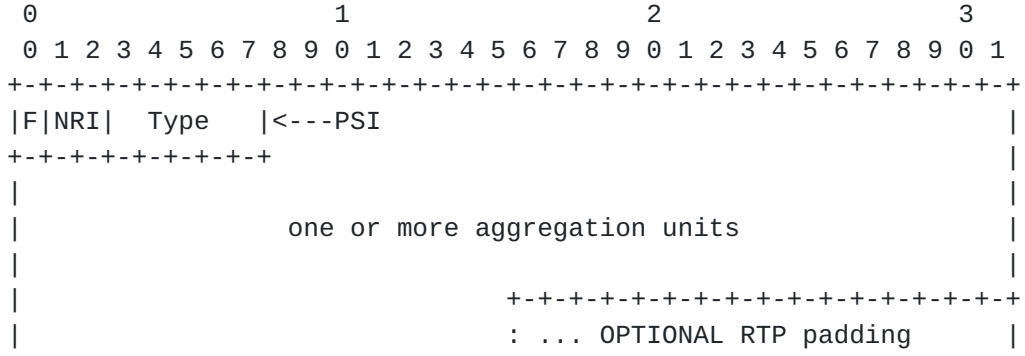


Figure 5. RTP payload format for aggregation packets

The RTP timestamp MUST be set to the earliest of the NALU times of all the NAL units to be aggregated. The type field of PSI MUST be set to the appropriate value, as indicated in Table 4. The F bit of PSI MUST be cleared if all F bits of the aggregated NAL units are zero; otherwise, it MUST be set to 1. The value of NRI in PSI MUST be the maximum of all the NAL units carried in the aggregation packet.

Table 4. Type field of PSI for STAPs and MTAPs

PSI Type	Payload structure	Timestamp offset field length£¨in bits£©	DON related fields £¨DON¡¢DONB¡¢DOND£© present
24	STAP-A	0	NO
25	STAP-B	0	YES
26	MTAP16	16	YES
27	MTAP24	24	YES

The marker bit (M) in the RTP header is set to the value that the marker bit of the last NAL unit of the aggregated packet would have if it were transported in its own RTP packet.

Following PSI, the payload of an aggregation packet consists of one or more aggregation units. See sections 5.6.1 and 5.6.2 for the four different types of aggregation units. An aggregation packet can carry as many aggregation units as necessary; however, the total amount of data in an aggregation packet obviously MUST fit into an IP packet, and the size SHOULD be chosen so that the resulting IP packet is smaller than the MTU size. An aggregation packet MUST NOT contain fragmentation units specified in section 5.7. Aggregation packets MUST NOT be nested; i.e., an aggregation packet MUST NOT contain another aggregation packet.

5.6.1 Single-Time Aggregation Packet

Single-time aggregation packet (STAP) SHOULD be used whenever NAL units are aggregated that all share the same NALU-time. The payload of an STAP-A does not include DON and consists of at least one single-time aggregation unit, as presented in Figure 6. The payload of an STAP-B consists of a 16-bit unsigned decoding order number (DON) (in network byte order) followed by at least one single-time aggregation unit, as presented in Figure 7.

The DON field specifies the value of DON for the first NAL unit in an STAP-B in transmission order. For each successive NAL unit in appearance order in an STAP-B, the value of DON is equal to (the value of DON of the previous NAL unit in the STAP-B + 1) % 65536,

in which '%' stands for the modulo operation.

A single-time aggregation unit consists of 16-bit unsigned size

information (in network byte order) that indicates the size of the following NAL unit in bytes (excluding these two octets, but including the NAL unit type octet of the NAL unit), followed by the NAL unit itself, including its NAL unit type byte. A single-time aggregation unit is byte aligned within the RTP payload, but it may not be aligned on a 32-bit word boundary. Figure 8 presents the structure of the single-time aggregation unit.

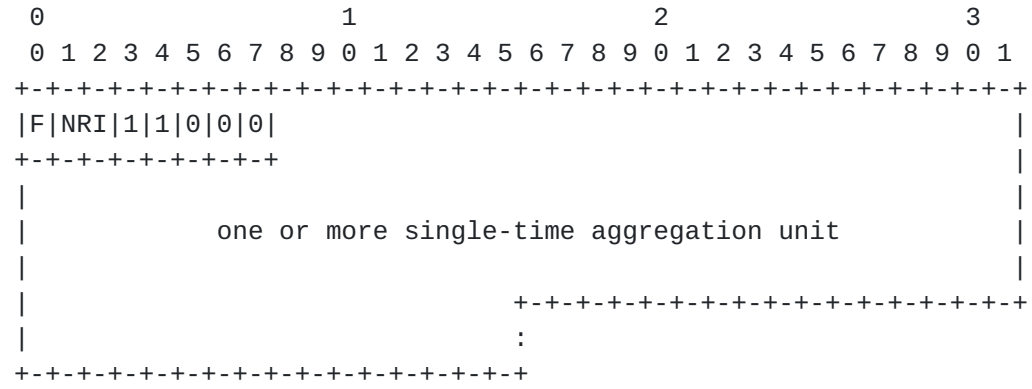


Figure 6. Payload format for STAP-A

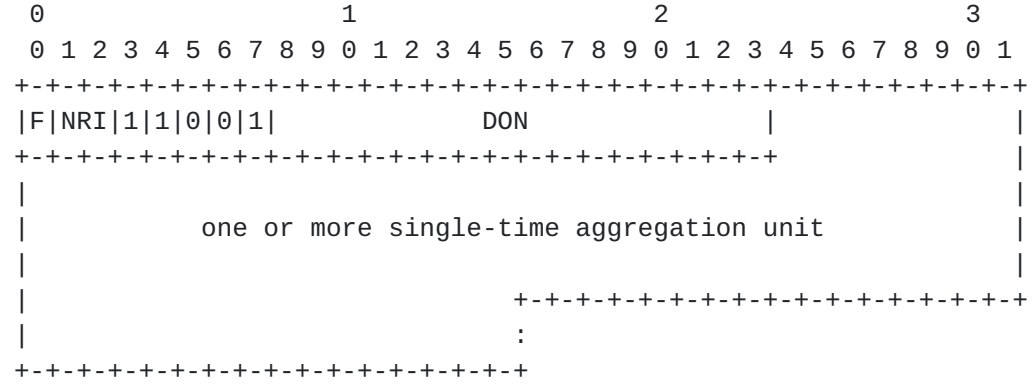


Figure 7. Payload format for STAP-B

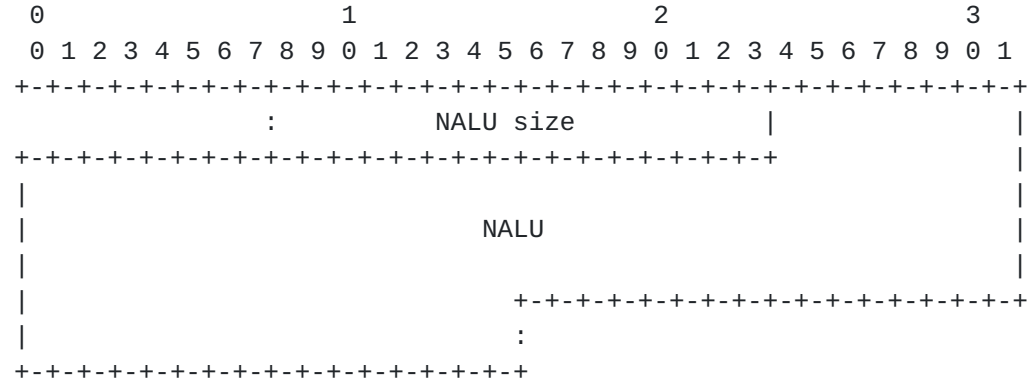


Figure 8. Structure for Single-Time aggregation unit

Figure 9 shows an example of an RTP packet that contains an STAP-A. The STAP contains two single-time aggregation units,

labeled as 1 and 2 in the figure.

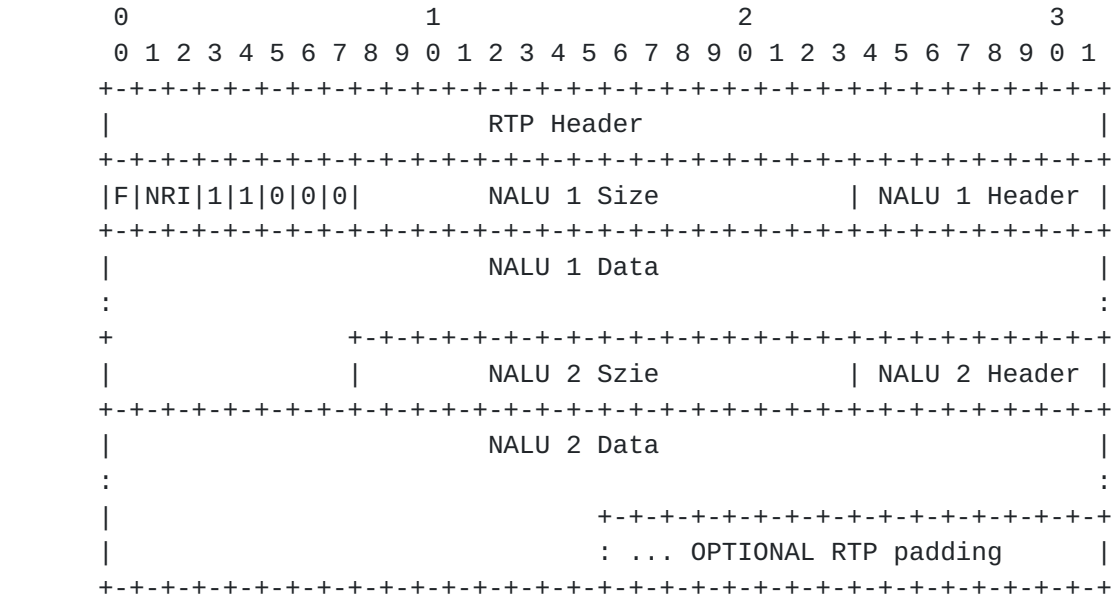


Figure 9. An example of an RTP packet including an STAP-A and two single-time aggregation units

Figure 10 shows an example of an RTP packet that contains an STAP-B. The STAP contains two single-time aggregation units, labeled as 1 and 2 in the figure.

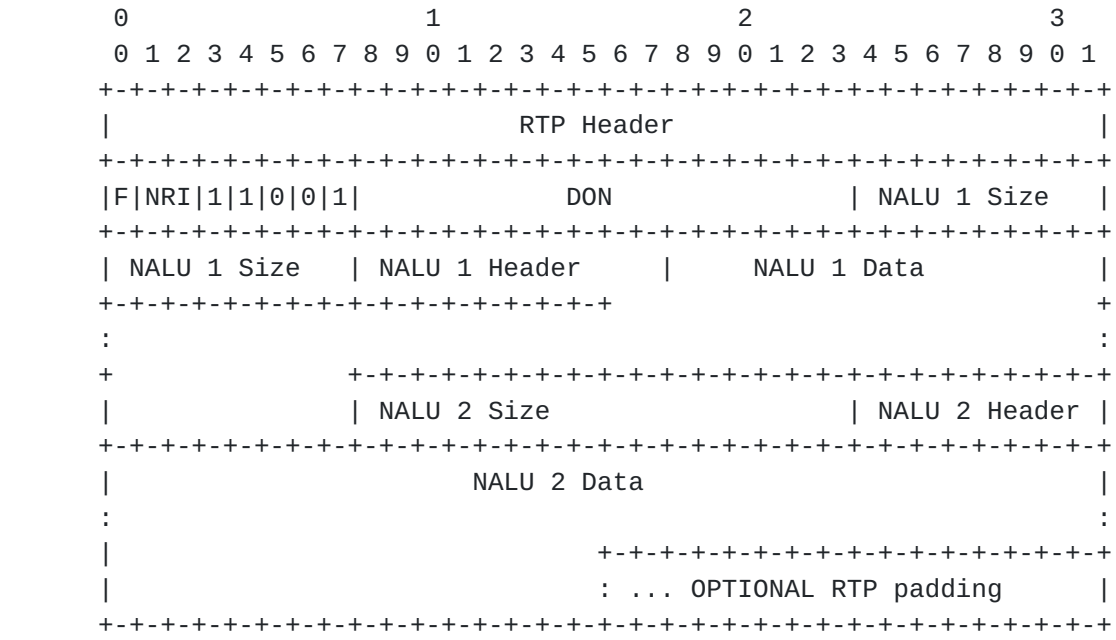


Figure 10. An example of an RTP packet including an STAP-B and two single-time aggregation units

5.6.2 Multi-Time Aggregation Packets (MTAPs)

The NAL unit payload of MTAPs consists of a 16-bit unsigned decoding

order number base (DONB) (in network byte order) and one or more multi-time aggregation units, as shown in Figure 11. DONB MUST contain the value of DON for the first NAL unit in the NAL unit decoding order among the NAL units of the MTAP.

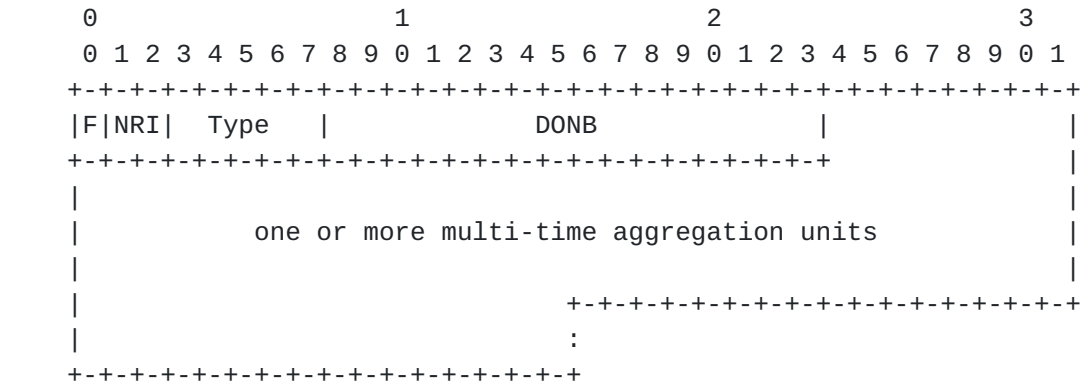
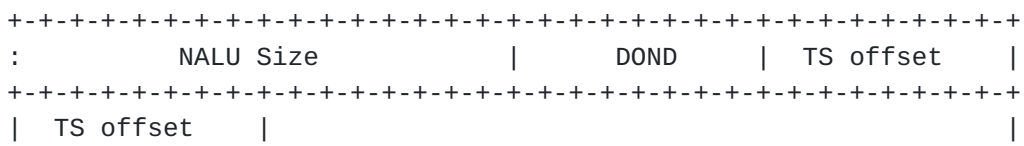


Figure 11. MTAP payload format

Two different multi-time aggregation units are defined in this specification. Both of them consist of 16 bits unsigned size information of the following NAL unit (in network byte order), an 8-bit unsigned decoding order number difference (DOND), and n bits (in network byte order) of timestamp offset (TS offset) for this NAL unit, whereby n can be 16 or 24. The choice between the different MTAP types (MTAP16 and MTAP24) is application dependent: the larger the timestamp offset is, the higher the flexibility of the MTAP, but the transport efficiency is lower.

The structure of the multi-time aggregation units for MTAP16 and MTAP24 are presented in Figures 12 and 13, respectively. The starting or ending position of an aggregation unit within a packet is NOT REQUIRED to be on a 32-bit word boundary. The DON of the following NAL unit is equal to (DONB + DOND) % 65536, in which % denotes the modulo operation. This memo does not specify how the NAL units within an MTAP are ordered, but, in most cases, NAL unit decoding order SHOULD be used.

The timestamp offset field MUST be set to a value equal to the value of the following formula: If the NALU-time is larger than or equal to the RTP timestamp of the packet, then the timestamp offset equals (the NALU-time of the NAL unit - the RTP timestamp of the packet). If the NALU-time is smaller than the RTP timestamp of the packet, then the timestamp offset is equal to the NALU-time + (2^32 - the RTP timestamp of the packet).


NALU Size
DOND
TS offset
TS offset

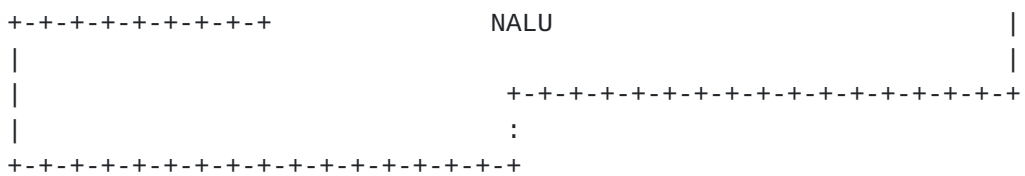


Figure 12. Multi-time aggregation unit for MTAP16

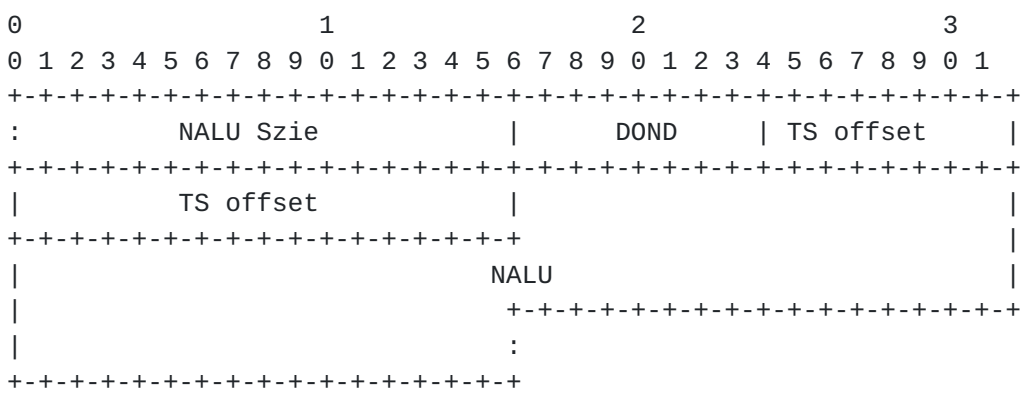
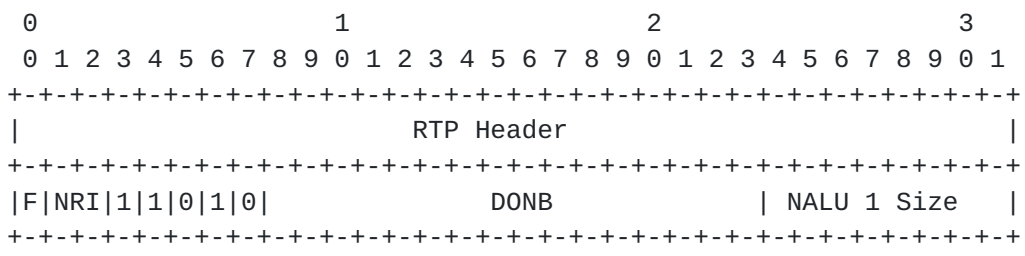
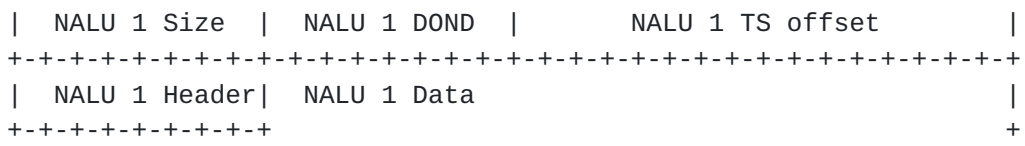


Figure 13. Multi-time aggregation unit for MTAP24

For the "earliest" multi-time aggregation unit in an MTAP the timestamp offset MUST be zero. Hence, the RTP timestamp of the MTAP itself is identical to the earliest NALU-time. The "earliest" multi-time aggregation unit means the one that would have the smallest extended RTP timestamp among all the aggregation units of an MTAP if the aggregation units were encapsulated in single NAL unit packets. An extended timestamp is a timestamp that has more than 32 bits and is capable of counting the wraparound of the timestamp field, thus enabling one to determine the smallest value if the timestamp wraps. Such an "earliest" aggregation unit may not be the first one in the order in which the aggregation units are encapsulated in an MTAP. The "earliest" NAL unit need not be the same as the first NAL unit in the NAL unit decoding order either.

Figure 14 presents an example of an RTP packet that contains a multi-time aggregation packet of type MTAP16 that contains two multi-time aggregation units, labeled as 1 and 2 in the figure.




| NALU 1 Size | NALU 1 DOND | NALU 1 TS offset |
| NALU 1 Header | NALU 1 Data |

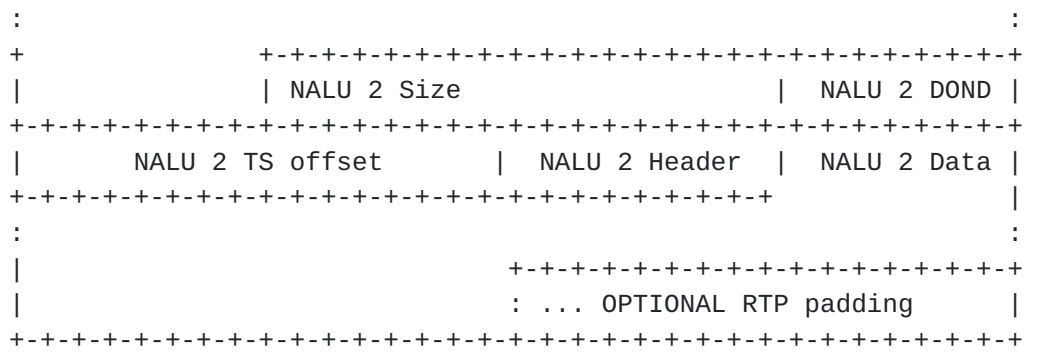


Figure 14. An RTP packet including a multi-time aggregation packet of type MTAP16 and two multi-time aggregation units

Figure 15 presents an example of an RTP packet that contains a multi-time aggregation packet of type MTAP24 that contains two multi-time aggregation units, labeled as 1 and 2 in the figure.

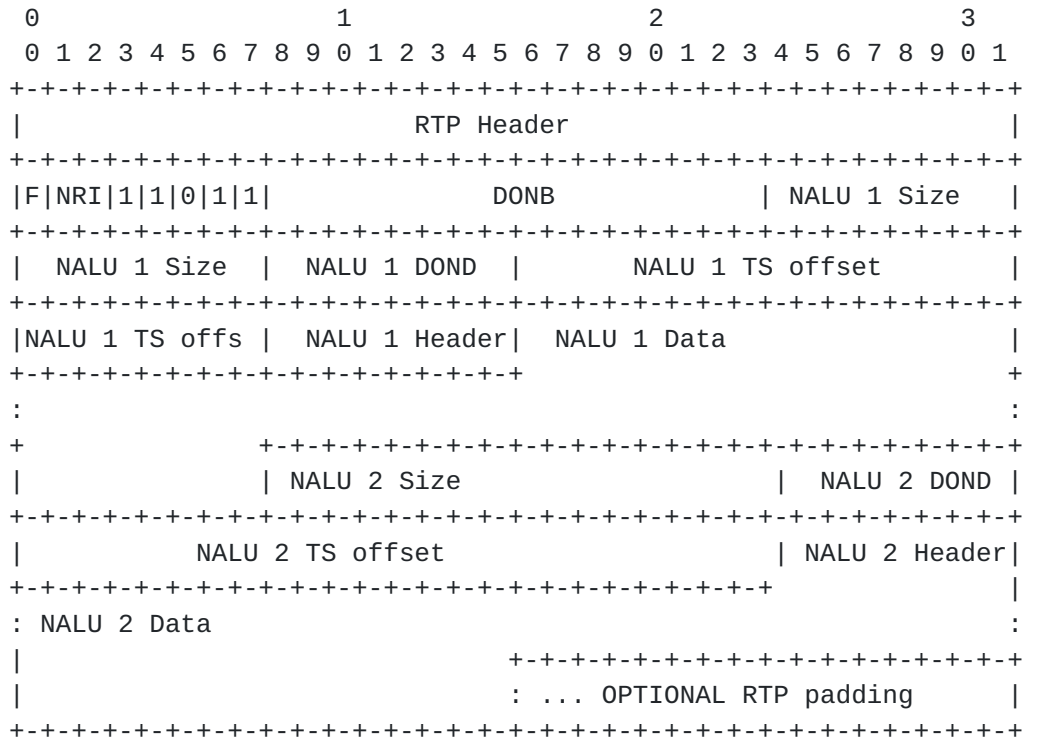


Figure 15. An RTP packet including a multi-time aggregation packet of type MTAP24 and two multi-time aggregation units

5.7 Fragmentation Units(FUs)

This payload type allows fragmenting a NAL unit into several RTP packets. Doing so on the application layer instead of relying on lower layer fragmentation (e.g., by IP) ,The payload format is

capable of transporting NAL units bigger than 64 kbytes over an IPv4 network that may be present in prerecorded video, particularly in High Definition formats.

Fragmentation is defined only for a single NAL unit and not for any aggregation packets. A fragment of a NAL unit consists of an integer number of consecutive octets of that NAL unit. Each octet of the NAL unit MUST be part of exactly one fragment of that NAL unit. Fragments of the same NAL unit MUST be sent in consecutive order with ascending RTP sequence numbers, (with no other RTP packets within the same RTP packet stream being sent between the first and last fragment. Similarly, a NAL unit MUST be reassembled in RTP sequence number order.

A NAL unit fragmented is called fragmented NAL unit. STAPs and MTAPs MUST NOT be fragmented. FUs MUST NOT be nested; i.e., an FU MUST NOT contain another FU.

The RTP timestamp of an RTP packet carrying an FU is set to the NALU time of the fragmented NAL unit.

Figure 16 presents the RTP payload format for FU-As. An FU-A consists of a PSI (1 byte), a fragmentation unit header (1 byte), and a fragmentation unit payload.

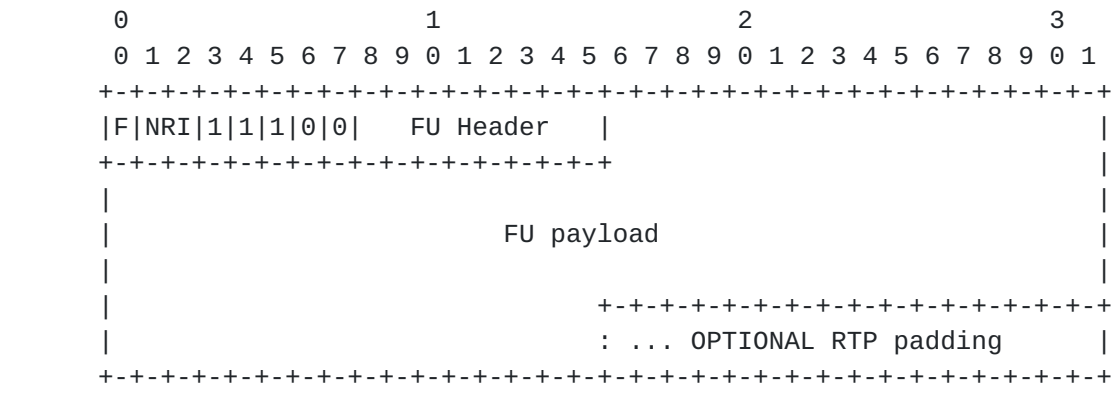
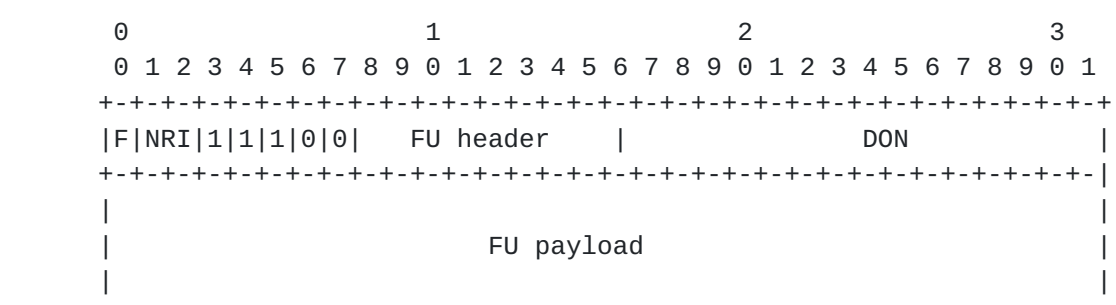


Figure 16. RTP payload for FU-A

Figure 17 presents the RTP payload format for FU-Bs. An FU-B consists of a PSI (1 byte), a fragmentation unit header (1 byte), a decoding order number (DON, in network byte order), and a fragmentation unit payload.



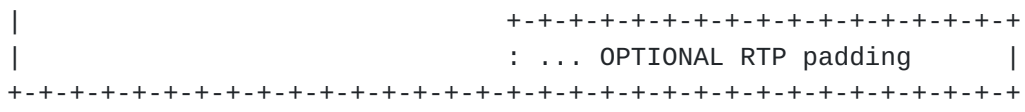

: ... OPTIONAL RTP padding

Figure 17. RTP payload for FU-B

NAL unit type FU-B MUST be used in the interleaved packetization mode for the first fragmentation unit of a fragmented NAL unit. NAL unit type FU-B MUST NOT be used in any other case. In other words, in the interleaved packetization mode, each NALU that is fragmented has an FU-B as the first fragment, followed by one or more FU-A fragments.

Figure 18 shows the format of the FU header:

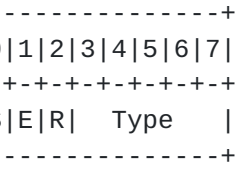


Figure 18. FU header format

S: 1 bit
When set to one, the Start bit indicates the start of a fragmented NAL unit. When the following FU payload is not the start of a fragmented NAL unit payload, the Start bit is set to zero.

E: 1 bit
When set to one, the End bit indicates the end of a fragmented NAL unit, i.e., the last byte of the payload is also the last byte of the fragmented NAL unit. When the following FU payload is not the last fragment of a fragmented NAL unit, the End bit is set to zero.

R: 1 bit
The Reserved bit MUST be equal to 0 and MUST be ignored by the receiver.

Type: 5 bits
The NAL unit payload type as defined in Table 1 of Section 4.

The value of DON in FU-Bs is selected as described in section 5.4.

A fragmented NAL unit MUST NOT be transmitted in one FU; i.e., the Start bit and End bit MUST NOT both be set to one in the same FU header.

The FU payload consists of fragments of the payload of the fragmented NAL unit so that if the fragmentation unit payloads of consecutive FUs are sequentially concatenated, the payload of the fragmented NAL unit can be reconstructed. The NAL unit type octet

of the fragmented NAL unit is not included as such in the fragmentation unit payload, but rather the information of the NAL unit type octet of the fragmented NAL unit is conveyed in F and NRI

fields of the FU indicator octet of the fragmentation unit and in the type field of the FU header. A FU payload MAY have any number of octets and MAY be empty.

If a fragmentation unit is lost, the receiver SHOULD discard all following fragmentation units in transmission order corresponding to the same fragmented NAL unit.

A receiver in an endpoint or in a MANE MAY aggregate the first n-1 fragments of a NAL unit to an (incomplete) NAL unit, even if fragment n of that NAL unit is not received. In this case, the forbidden_zero_bit of the NAL unit MUST be set to one to indicate a syntax violation.

6. Packetization Rules

6.1 Common Packetization Rules

All senders MUST enforce the following packetization rules regardless of the packetization mode in use:

- Coded picture header or slice NAL units belonging to the same coded picture (and thus sharing the same RTP timestamp value) MAY be sent in any order permitted by the applicable profiles defined in AVS-P2; However, for delay-critical systems, they SHOULD be sent in their original coding order to minimize the delay.

- Sequence headers are handled in accordance with the rules and recommendations given in section 7.3.

- Senders (include MANE) MUST NOT duplicate any NAL unit except for sequence header or picture header NAL units. Sequence header NAL units MUST NOT be duplicated to affect any active sequence header. Duplicated Picture header NAL units MUST be followed by the picture's slice NAL units (but MAY not be the first slice of the picture). Duplication SHOULD be performed on the application layer and not by duplicating RTP packets (with identical sequence numbers).

Senders using the non-interleaved mode and the interleaved mode MUST enforce the following packetization rule:

- MANEs MAY convert many single NAL unit packets into one aggregation packet, convert an aggregation packet into several single NAL unit packets, or mix both concepts, in an RTP translator. The RTP translator SHOULD take into account at least the following parameters: path MTU size, unequal protection mechanisms, bearable latency of the system, and buffering capabilities of the receiver.

6.2 Single NAL Unit Mode

This mode is in use when the value of the OPTIONAL packetization-mode media type parameter is equal to 0, the packetization-mode is not present, or no other packetization mode is signaled by external means. All receivers MUST support this mode. Only single NAL unit packets MAY be used in this mode. STAPs, MTAPs, and FUs MUST NOT be used. The transmission order of single NAL unit packets MUST comply with the NAL unit decoding order.

6.3 Non-Interleaved Mode

This mode is in use when the value of the OPTIONAL packetization-mode media type parameter is equal to 1 or the mode is turned on by external means. Only single NAL unit packets, STAP-As, and FU-As MAY be used in this mode. STAP-Bs, MTAPs, and FU-Bs MUST NOT be used. The transmission order of NAL units MUST comply with the NAL unit decoding order.

6.4 Interleaved Mode

This mode is in use when the value of the OPTIONAL packetization-mode media type parameter is equal to 2 or the mode is turned on by external means. Some receivers MAY support this mode. Only STAP-Bs, MTAPs, FU-As, and FU-Bs MAY be used. Single NAL unit packets and STAP-As MUST NOT be used. The transmission order of packets and NAL units is constrained as specified in section 5.4.

7. Payload Format Parameters

This section specifies the media type parameters that MAY be used to select optional features of the payload format and certain features of the bitstream. The parameters are specified here as part of the media subtype registration for the AVS-P2 video specification. A mapping of the parameters into the Session Description Protocol (SDP) [4] is also provided for applications that use SDP. Equivalent parameters could be defined elsewhere for use with control protocols that do not use media type parameters or SDP.

Some parameters provide a receiver with the properties of the stream that will be sent. The name of all these parameters starts with "sprop" for stream properties. Some of these "sprop" parameters are limited by other payload or codec configuration parameters. The media sender selects all "sprop" parameters rather than the receiver. This uncommon characteristic of the "sprop" parameters MAY NOT be compatible with some signaling protocol concepts, in which case the use of these parameters SHOULD be avoided.

7.1 Media Type Registration

This registration uses the template defined in IETF RFC 4288 [5] and follows IETF RFC 3555 [6].

The media subtype for the AVS-P2 video is allocated from the IETF tree. The receiver MUST ignore any unspecified parameter.

Media Type name:
video

Media subtype name:
AVS1-P2

Required parameters:
none

Optional parameters:

profile-level-id:
A base16 [7] (hexadecimal) representation of the following two bytes in the sequence header of AVS-P2: profile_id and level_id.

If the profile-level-id parameter is used to indicate properties of a AVS-P2 bit stream, it indicates the profile and level that has to support in order to comply with when it decodes the stream.

If the profile-level-id parameter is used for capability exchange or session setup procedure, it indicates the profile that the codec supports and the highest level supported for the signaled profile.

If no profile-level-id is present, the Jizhun Profile without additional constraints at Level 4.0 MUST be implied.

max-mbps, max-fs, max-dpb, and max-br:
These parameters MAY be used to signal the capabilities of a receiver implementation. These parameters MUST NOT be used for any other purposes. The profile-level-id parameter MUST be present in the same receiver capability description that contains any of these parameters. The level conveyed in the value of the profile-level-id parameter MUST be such that the receiver is fully capable of supporting. These four parameters MAY be used to indicate capabilities of the receiver that extend the required capabilities of the signaled level, as specified below.

When more than one parameter from the four is present, the receiver MUST support all signaled capabilities simultaneously. For example, if both max-mbps and max-br are present, the signaled level with the extension of both

the frame rate and bit rate is supported by the receiver.
That is, the receiver is able to decode bit stream in which
the macroblock processing rate is up to max-mbps (inclusive),

the bit rate is up to max-br (inclusive), the coded picture buffer size is derived as specified in the semantics of the max-br parameter below, and other properties comply with the level specified in the value of the profile-level-id parameter.

A receiver MUST NOT signal values of max-mbps, max-fs, max-dpb, and max-br that meet the requirements of a higher level, referred to as level A herein, compared to the level specified in the value of the profile-level-id parameter, if the receiver can support all the properties of level A.

max-mbps:
The value of max-mbps is an integer indicating the maximum macroblock processing rate in units of macroblocks per second. The max-mbps parameter signals that the receiver is capable of decoding video at a higher rate than is REQUIRED by the signaled level conveyed in the value of the profile-level-id parameter. When max-mbps is signaled, the receiver MUST be able to decode AVS-P2 bit streams that conform to the signaled level, with the exception that the value of maximum microblocks per second in Table B.4 and B.5 of AVS-P2 [1] for the signaled level is replaced with the value of max-mbps. The value of max-mbps MUST be greater than or equal to the value of maximum microblocks per second for the level given in Table B.4 and B.5 of AVS-P2. Senders MAY use this knowledge to send a given size at a higher frame rate than is indicated in the signaled level.

max-fs:
The value of max-fs is an integer indicating the maximum frame size in units of macroblocks. The max-fs parameter signals that the receiver is capable of decoding larger picture sizes than are REQUIRED by the signaled level conveyed in the value of the profile-level-id parameter. When max-fs is signaled, the receiver MUST be able to decode bit streams that conform to the signaled level, with the exception that the value of maximum macroblocks per frame in Table B.4 and B.5 of AVS-P2 for the signaled level is replaced with the value of max-fs. The value of max-fs MUST be greater than or equal to the value of maximum macroblocks per frame for the level given in Table B.4 and B.5 of AVS-P2. Senders MAY use this knowledge to send larger pictures at a proportionally lower frame rate than is indicated in the signaled level.

max-dpb:
The value of max-dpb is an integer indicating the maximum decoded picture buffer size in units of 1000 bits. The

max-dpb parameter signals that the receiver has more memory than the minimum amount of decoded picture buffer memory required by the signaled level conveyed in the value of the profile-level-id parameter. When max-dpb is signaled, the

receiver MUST be able to decode bit streams that conform to the signaled level, with the exception that the value of BBV buffer size in Table B.4 and B.5 of AVS-P2 for the signaled level is replaced with the value of 1000*(max-dpb). The value of 1000*(max-dpb) MUST be greater than or equal to the value of BBV buffer size for the level given in Table B.4 and B.5 of AVS-P2. Senders MAY use this knowledge to construct coded streams with improved compression compared to BBV buffer size of the signaled profile.

max-br:
The value of max-br is an integer indicating the maximum video bit rate in units of 1000 bits per second. The max-br parameter signals that the video decoder of the receiver is capable of decoding video at a higher bit rate than is required by the signaled level conveyed. When max-br is signaled, the video codec of the receiver MUST be able to decode bit streams that conform to the signaled level, conveyed in the profile-level-id parameter, with the exception that the value of maximum bit rate in Table B.4 and B.5 of AVS-P2 for the signaled level is replaced with 1000*(max-br). The value of 1000*(max-br) MUST be greater than or equal to the value of maximum bit rate for the signaled level given in TableB-3, B-4. Senders MAY use this knowledge to send higher bitrate video as allowed in the level definition to achieve improved video quality.

sprop-parameter-sets:
This parameter MAY be used to convey any sequence header bit stream. The parameter MUST NOT be used to indicate codec capability in any capability exchange procedure. The value of the parameter is the base64 [7] representation of the sequence header bit stream. The headers are conveyed in decoding order, and a comma is used to separate any pair of headers in the list.

parameter-add:
This parameter MAY be used to signal whether the receiver of this parameter is allowed to add headers in its signaling response using the sprop-parameter-sets parameter. The value of this parameter is either 0 (deny) or 1 (allowing). If the parameter is not present, its value MUST be 1.

packetization-mode:
This parameter signals the properties of an RTP payload type or the capabilities of a receiver implementation. Only a single configuration point can be indicated; thus, when capabilities to support more than one packetization-mode are

declared, multiple configuration points (RTP payload types) must be used. When the value of packetization-mode is equal to 0 or packetization-mode is not present, the single NAL mode, as defined in section 6.2. When the value of

packetization-mode is equal to 1, the non-interleaved mode, as defined in section 6.3, MUST be used. When the value of packetization-mode is equal to 2, the interleaved mode, as defined in section 6.4 , MUST be used. The value of packetization mode MUST be an integer in the range of 0..2, inclusive..

sprop-interleaving-depth:
This parameter MUST NOT be present when packetization-mode is not present or the value of packetization-mode is equal to 0 or 1. This parameter MUST be present when the value of packetization-mode is equal to 2.

This parameter signals the properties of a NAL unit stream. It specifies the maximum number of NAL units that precede any NAL unit in the NAL unit stream in transmission order and follow the NAL unit in decoding order. Consequently, it is guaranteed that receivers can reconstruct NAL unit decoding order when the buffer size for NAL unit decoding order recovery is at least the value of sprop-interleaving-depth + 1 in terms of NAL units. The value of sprop-interleaving-depth MUST be an integer in the range of 0 to 32767, inclusive.

sprop-deint-buf-req:
This parameter MUST NOT be present when packetization-mode is not present or the value of packetization-mode is equal to 0 or 1. It MUST be present when the value of packetization-mode is equal to 2.

sprop-deint-buf-req signals the required size of the deinterleaving buffer for the NAL unit stream. The value of the parameter MUST be greater than or equal to the maximum buffer occupancy (in units of bytes) required in such a deinterleaving buffer that is specified in section 11.2.

The value of sprop-deint-buf-req must be an integer in the range of 0 to 4294967295, inclusive.

deint-buf-cap:
This parameter signals the capabilities of a receiver implementation and indicates the amount of deinterleaving buffer space in units of bytes that the receiver has available for reconstructing the NAL unit decoding order. A receiver is able to handle any stream for which the value of the sprop-deint-buf-req parameter is smaller than or equal to this parameter.

If the parameter is not present, then a value of 0 MUST be

used for deint-buf-cap. The value of deint-buf-cap MUST be an integer in the range of 0 to 4294967295, inclusive.

sprop-init-buf-time:

This parameter MAY be used to signal the properties of a NAL unit stream. The parameter MUST NOT be present, if the value of packetization-mode is equal to 0 or 1.

The parameter signals the initial buffering time that a receiver MUST buffer before starting decoding to recover the NAL unit decoding order from the transmission order. The parameter is the maximum value of (transmission time of a NAL unit - decoding time of the NAL unit), assuming reliable and instantaneous transmission, the same timeline for transmission and decoding, and that decoding starts when the first packet arrives.

An example of specifying the value of spropinit-buf-time follows. A NAL unit stream is sent in the following interleaved order, in which the value corresponds to the decoding time and the transmission order is from left to right:

0 2 1 3 5 4 6 8 7 ...

Assuming a steady transmission rate of NAL units, the transmission times are:

0 1 2 3 4 5 6 7 8 ...

Subtracting the decoding time from the transmission time column-wise results in the following series:

0 -1 1 0 -1 1 0 -1 1 ...

Thus, in terms of intervals of NAL unit transmission times, the value of sprop-init-buf-time in this example is 1. The parameter is coded as a non-negative base10 integer representation in clock ticks of a 90-kHz clock. If the parameter is not present, then no initial buffering time value is defined. Otherwise the value of sprop-initbuf-time MUST be an integer in the range of 0 to 4294967295, inclusive.

In addition to the signaled sprop-init-buftime, receivers SHOULD take into account the transmission delay jitter buffering, including buffering for the delay jitter caused by any network elements.

sprop-max-don-diff:

This parameter MAY be used to signal the properties of a NAL unit stream. It MUST NOT be used to signal transmitter or receiver or codec capabilities. The parameter MUST NOT be present if the value of packetization-mode is equal to 0 or 1.

sprop-max-don-diff is an integer in the range of 0 to 32767, inclusive. If sprop-max-don-diff is not present, the value of the parameter is unspecified. sprop-maxdon-diff is calculated

as follows:

sprop-max-don-diff = max{AbsDON(i) -AbsDON(j)},
for any i and any j>i,

where i and j indicate the index of the NAL unit in the transmission order and AbsDON denotes a decoding order number of the NAL unit that does not wrap around to 0 after 65535. In other words, AbsDON is calculated as follows:

Let m and n be consecutive NAL units in transmission order. For the very first NAL unit in transmission order (whose index is 0), AbsDON(0) = DON(0). For other NAL units, AbsDON is calculated as follows:

If DON(m) == DON(n), AbsDON(n) = AbsDON(m)

If (DON(m) < DON(n) and DON(n) - DON(m) < 32768),
AbsDON(n) = AbsDON(m) + DON(n) - DON(m)

If (DON(m) > DON(n) and DON(m) - DON(n) >= 32768),
AbsDON(n) = AbsDON(m) + 65536 - DON(m) + DON(n)

If (DON(m) < DON(n) and DON(n) - DON(m) >= 32768),
AbsDON(n) = AbsDON(m) - (DON(m) + 65536 - DON(n))

If (DON(m) > DON(n) and DON(m) - DON(n) < 32768),
AbsDON(n) = AbsDON(m) - (DON(m) - DON(n))

where DON(i) is the decoding order number of the NAL unit having index i in the transmission order. The decoding order number is specified in section 5.5.

max-rcmd-nalu-size:
This parameter MAY be used to signal the capabilities of a receiver. The parameter MUST NOT be used for any other purposes. The value of the parameter indicates the largest NALU size in bytes that the receiver can handle efficiently. The parameter value is a recommendation, not a strict upper boundary. The sender MAY create larger NALUs but must be aware that the handling of these may come at a higher cost than NALUs conforming to the limitation.

The value of max-rcmd-nalu-size MUST be an integer in the range of 0 to 4294967295, inclusive. If this parameter is not specified, no known limitation to the NALU size exists. Senders still have to consider the MTU size available between the sender and the receiver and SHOULD run MTU discovery for this purpose.

Encoding considerations:

This media type is framed and contains binary data.

Security considerations:
See Section 8 of RFC xxxx.

Interoperability considerations:
None.

Public specification:
RFC xxxx.

Applications that use this media type:
Video telephone, video conferencing, Internet media streaming, IPTV, video-on-demand, etc.

Additional information:
None.

Person and email address to contact for further information:
lshuo@jdl.ac.cn

Intended usage:
COMMON.

Restrictions on usage:
This media type depends on RTP framing; therefore, it is only defined for transfer via RTP (IETF RFC 3550).

File extensions:
None.

Macintosh file type code:
None.

Object identifier or OID:
None.

Author:
lshuo@jdl.ac.cn

Change controller:
IETF Audio/Video Transport Working Group delegated from the IESG.

7. 2 SDP Parameters

7.2.1 Mapping of Media Type Parameters to SDP

The media type string "video/AVS1-P2" is mapped to fields in the Session Description Protocol (SDP) [4] as follows:

- The media name in the "m=" line of SDP MUST be video (the type name).

- The encoding name in the "a=rtpmap" line of SDP MUST be AVS1-P2 (the subtype name).

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

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

An example of media representation in SDP is as follows (Baseline Profile, Level 6.0):

```
m=video 49170 RTP/AVP 98
a=rtpmap:98 AVS1-P2/90000
a=fmtp:98 profile-level-id=2040; sprop-parameter-sets=[SH#0]
```

where [SH#0] means a base64 expression of sequence header.

7.2.2 Usage with the SDP Offer/Answer Model

When AVS-P2 is offered over RTP using SDP in an Offer/Answer model [8] for negotiation for unicast usage, the following limitations and rules apply:

- The parameters identifying a AVS1-P video media format are "profile-level-id" , "packetization-mode" and "sprop-deint-buf-req" (if "packetization-mode" is equal to 2). These three parameters MUST be used symmetrically, which means the answerer MUST either maintain all configuration parameters or remove the media format (payload type) completely, if one or more of the parameter values are not supported.

To simplify handling and matching of these configurations, the same RTP payload type number used in the offer SHOULD also be used in the answer, as specified in [8]. An answer MUST NOT contain a payload type number used in the offer unless the configuration ("profile-level-id", "packetization-mode", and if present "sprop-deint-buf-req") is the same as in the offer.

- The parameters "sprop-parameter-sets", "sprop-deint-buf-req", "sprop-interleaving -depth" , "sprop-max-don-diff", and "sprop-init-buf-time" describe the properties of the AVS-P2 bit stream that the offerer or answerer is sending for this media format configuration. This differs from the normal usage

of the Offer/Answer parameters: normally such parameters declare the properties of the stream that the offerer or the answerer is able to receive. When dealing with AVS-P2, the offerer assumes

that the answerer will be able to receive media encoded using the configuration being offered.

- The capability parameters "max-mbps", "max-fs", "max-dpb", "max-br", and "max-rcmd-nalu-size" MAY be used to declare further capabilities. Their interpretation depends on the direction attribute. When the direction attribute is sendonly, then the parameters describe the limits of the RTP packets and the NAL unit stream that the sender is capable of producing. When the direction attribute is sendrecv or recvonly, then the parameters describe the limitations of what the receiver accepts.

- As specified above, an offerer has to include the size of the deinterleaving buffer in the offer for an interleaved AVS-P2 stream. To enable the offerer and answerer to inform each other about their capabilities for deinterleaving buffering, both parties are RECOMMENDED to include "deint-buf-cap". This information MAY be used when the value for "sprop-deint-buf-req" is selected in a second round of offerand answer. For interleaved streams, it is also RECOMMENDED to consider offering multiple payload types with different buffering requirements when the capabilities of the receiver are unknown.

- The "sprop-parameter-sets" parameter is used as described above. In addition, an answerer MUST maintain all sequence headers received in the offer in its answer. Depending on the value of the "parameter-add" parameter, different rules apply: If "parameter-add" is 0, the answer MUST NOT add any additional headers. If "parameter-add" is 1, the answerer, in its answer, MAY add additional headers to the "sprop-parameter-sets" parameter. The answerer MUST also, independent of the value of "parameter-add", accept to receive a video stream using the sprop-parameter-sets it declared in the answer.

For streams being delivered over multicast, the following rules apply in addition:

- The stream properties parameters "sprop-parameter-sets", "sprop-deint-buf-req", "sprop-interleaving-depth", "sprop-max-don-diff", and "sprop-init-buf-time" MUST NOT be changed by the answerer. Thus, a payload type can either be accepted unaltered or removed.

- The receiver capability parameters "max-mbps", "max-fs", "max-dpb", "max-br", and "max-rcmd-nalu-size" MUST be supported by the answerer for all streams declared assendrecv or recvonly; otherwise, the media format is removed, or the session rejected.

Below are the complete lists of how the different parameters SHALL be interpreted in the different combinations of offer or answer and direction attribute.

- In offers and answers for which "a=sendrecv" or no direction attribute is used, or in offers and answers for which "a=recvonly" is used, the following interpretation of the parameters MUST be used.

  Declaring actual configuration or properties for receiving:
  - profile-level-id
  - packetization-mode

  Declaring actual properties of the stream to be sent (applicable only when "a=sendrecv" or no direction attribute is used):
  - sprop-deint-buf-req
  - sprop-interleaving-depth
  - sprop-parameter-sets
  - sprop-max-don-diff
  - sprop-init-buf-time

  Declaring receiver implementation capabilities:
  - max-mbps
  - max-fs
  - max-dpb
  - max-br
  - deint-buf-cap
  - max-rcmd-nalu-size

  Declaring how Offer/Answer negotiation SHALL be performed:
  - parameter-add

- In an offer or answer for which the direction attribute "a=sendonly" is included for the media stream, the following interpretation of the parameters MUST be used:

  Declaring actual configuration and properties of stream proposed to be sent:
  - profile-level-id
  - packetization-mode
  - sprop-deint-buf-req
  - sprop-max-don-diff
  - sprop-init-buf-time
  - sprop-parameter-sets
  - sprop-interleaving-depth

  Declaring the capabilities of the sender when it receives a stream:
  - max-mbps
  - max-fs
  - max-dpb
  - max-br

- deint-buf-cap
- max-rcmd-nalu-size

Declaring how Offer/Answer negotiation SHALL be performed:

- parameter-add

Furthermore, the following considerations are necessary:

- Parameters used for declaring receiver capabilities are in general downgradable, i.e., they express the upper limit for a sender's possible behavior. Thus a sender MAY select to set its encoder using only lower/lesser or equal values of these parameters. "sprop-parameter-sets" MUST NOT be used in a sender's declaration of its capabilities, as the limits of the values that are carried inside the parameter sets are implicit with the profile and level used.

- Parameters declaring a configuration point are not downgradable, with the exception of the level part of the "profile-level-id" parameter. This expresses values a receiver expects to be used and must be used verbatim on the sender side.

- When a sender's capabilities are declared, and non-downgradable parameters are used in this declaration, then these parameters express a configuration that is acceptable. In order to achieve high interoperability levels, it is often advisable to offer multiple alternative configurations; e.g., for the packetization mode. It is impossible to offer multiple configurations in a single payload type. Thus, when multiple configuration offers are made, each offer requires its own RTP payload type associated with the offer.

- A receiver SHOULD understand all media type parameters, even if it only supports a subset of the payload format's functionality. This ensures that a receiver is capable of understanding when an offer to receive media can be downgraded to what is supported by the receiver of the offer.

- An answerer MAY extend the offer with additional media format configurations. However, to enable their usage, in most cases a second offer is required from the offerer to provide the stream properties parameters that the media sender will use. This also has the effect that the offerer has to be able to receive this media format configuration, not only to send it.

- If an offerer wishes to have non-symmetric capabilities between sending and receiving, the offerer has to offer different RTP sessions; i.e., different "m=" lines declared as "recvonly" and "sendonly", respectively.

7.2.3 Usage in Declarative Session Descriptions

When AVS-P2 video over RTP is offered with SDP in a declarative

style, as in RTSP [11] or SAP [12], the following considerations are necessary.

- All parameters capable of indicating the properties of both an AVS-P2 bit stream and a receiver are used to indicate the properties of an AVS-P2 bit stream. For example, in this case, the parameter "profile-level-id" declares the values used by the stream, instead of the capabilities of the sender. This results in that the following interpretation of the parameters MUST be used:

  Declaring actual configuration or properties:
  - profile-level-id
  - sprop-parameter-sets
  - packetization-mode
  - sprop-interleaving-depth
  - sprop-deint-buf-req
  - sprop-max-don-diff
  - sprop-init-buf-time

  Not usable:
  - max-mbps
  - max-fs
  - max-dpb
  - max-br
  - redundant-pic-cap
  - max-rcmd-nalu-size
  - parameter-add
  - deint-buf-cap

- A receiver of the SDP is REQUIRED to support all parameters and values of the parameters provided; otherwise, the receiver MUST reject (RTSP) the session. It falls on the creator of the session to use values that are expected to be supported by the receiving application.

7.3 Considerations for Sequence Header

The sequence headers play a vital rule for the operations of AVS1-P2 video codec. Due to their importance for the decoding process, lost or erroneously transmitted sequence headers can hardly be concealed locally at the receiver. A reference to a corrupt header has normally fatal results to the decoding process. Corruption could occur, for example, due to the erroneous transmission or loss of a header data structure, or due to the untimely transmission of a header update. Therefore, the following recommendations are provided as a guideline for the implementer of the RTP sender:

Sequence header NALUs can be transported using three different principles:

A. Using a session control protocol (out-of-band) prior to the actual RTP session.

B. Using a session control protocol (out-of-band) during an ongoing

RTP session.

C. Within the RTP stream in the payload (in-band) during an ongoing RTP session.

It is necessary to implement principles A and B within a session control protocol. Principle C is supported by the RTP payload format defined in this document.

Principle A SHOULD be used for the transmission of initial sequence header of the whole sequence. Principle B SHOULD be used for update of in-band sequence header. Principle C SHOULD be used for update of in-band sequence header.

During a session, the sequence header SHUOLD be transmitted out-of-band using principle A, and updated using principles B or C. At least one sequence header MAY be useful using out-of-band transmission of initial sequence header, and update when new header is coming.

If principle B is used for updating sequence headers, it is impossible to ensure the synchronization between the sequence header and the in-band transmitted NAL units. This will cause confusion in both senders and receivers. Therefore it is RECOMMENDED to only use principle C to update the sequence header.

8. Security Considerations

RTP packets using the payload format defined in this document are subject to the security considerations discussed in IETF RFC 3550, and in any appropriate RTP profile (for example, IETF RFC 3551 [13]). This implies that confidentiality of the media streams is achieved by encryption; for example, IETF RFC 3711 [14]. Because the data compression used with this payload format is applied end-to-end, any encryption needs to be performed after compression.

A potential denial-of-service threat exists for data encodings using compression techniques that have non-uniform receiver-end computational load. The attacker can inject pathological datagrams into the stream that are complex to decode and that cause the receiver to be overloaded. AVS-P2 is particularly vulnerable to such attacks, as it is extremely simple to generate user-data that affect the decoding process of future bit stream. Therefore, the usage of data origin authentication and data integrity protection of at least the RTP packet is RECOMMENDED; for example, IETF RFC 3711.

Note that the appropriate mechanism to ensure confidentiality and integrity of RTP packets and their payloads is very dependent on the application and on the transport and signaling protocols employed.

Thus, although IETF RFC 3711 is given as an example above, other possible choices exist.

End-to-End security with either authentication, integrity or confidentiality protection will prevent a MANE from performing media-aware operations other than discarding complete packets. And in the case of confidentiality protection it will even be prevented from performing discarding of packets in a media aware way. To allow any MANE to perform its operations, it will be REQUIRED to be a trusted entity which is included in the security context establishment.

9. Congestion Control

Congestion control for RTP SHALL be used in accordance with RFC 3550, and with any applicable RTP profile; e.g., IETF RFC 3551. An additional requirement if best-effort service is being used is: users of this payload format MUST monitor packet loss to ensure that the packet loss rate is within acceptable parameters. Packet loss is considered acceptable if a TCP flow across the same network path, and experiencing the same network conditions, would achieve an average throughput, measured on a reasonable timescale, that is not less than the RTP flow is achieving. This condition can be satisfied by implementing congestion control mechanisms to adapt the transmission rate, or the number of layers subscribed for a layered multicast session, or by arranging for a receiver to leave the session if the loss rate is unacceptably high.

The bit rate adaptation necessary for obeying the congestion control principle is easily achievable when real-time encoding is used. However, when pre-encoded content is being transmitted, bandwidth adaptation requires the availability of more than one coded representation of the same content, at different bit rates, or the existence of non-reference pictures in the bitstream. The switching between the different representations can normally be performed in the same RTP session; e.g., in the I-Frames. Only when non-downgradable parameters (such as the profile part of the profile/level ID) are REQUIRED to be changed does it become necessary to terminate and re-start the media stream. This may be accomplished by using a different RTP payload type.

10. IANA Considerations

Apply to IANA for registering one new media type; see section 7.1.

11. De-Packetization Process (Informative)

The de-packetization process is implementation dependent. Therefore, the following description SHOULD be seen as an example of a suitable implementation. Other schemes may be used as well. Optimizations relative to the described algorithms are likely

possible. Section 11.1 presents the de-packetization process for the single NAL unit and non-interleaved packetization modes, whereas section 11.2 describes the process for the interleaved mode. Section 11.3 includes additional decapsulation guidelines for receivers.

All normal RTP mechanisms related to buffer management apply. In particular, duplicated or outdated RTP packets (as indicated by the RTP sequences number and the RTP timestamp) are removed. To determine the exact time for decoding, factors such as a possible intentional delay to allow for proper inter-stream synchronization must be factored in.

11.1. Single NAL Unit and Non-Interleaved Mode

The receiver includes a receiver buffer to compensate for transmission delay jitter. The receiver stores incoming packets in reception order into the receiver buffer. Packets are decapsulated in RTP sequence number order. If a decapsulated packet is a single NAL unit packet, the NAL unit contained in the packet is passed directly to the decoder. If a decapsulated packet is an STAP-A, the NAL units contained in the packet are passed to the decoder in the order in which they are encapsulated in the packet. If a decapsulated packet is an FU-A, all the fragments of the fragmented NAL unit (if exists) are concatenated and passed to the decoder.

11.2. Interleaved Mode

The general concept behind these de-packetization rules is to reorder NAL units from transmission order to the NAL unit decoding order.

The receiver includes a receiver buffer, which is used to compensate for transmission delay jitter. In this section, the receiver operation is described under the assumption that there is no transmission delay jitter. To make a difference from a practical receiver buffer that is also used for compensation of transmission delay jitter, the receiver buffer is here called the deinterleaving buffer in this section.

11.2.1 Size of the Deinterleaving Buffer

When SDP Offer/Answer model or any other capability exchange procedure is used in session setup, the properties of the received stream SHOULD be such that the receiver capabilities are not exceeded. In the SDP Offer/Answer model, the receiver can indicate its capabilities to allocate a deinterleaving buffer with the deint-buf-cap media type parameter. The sender indicates the requirement for the deinterleaving buffer size with the sprop-deint-buf-req parameter. It is therefore RECOMMENDED to set the deinterleaving buffer size, in terms of number of bytes, equal to or greater than the value of sprop-deint-buf-req parameter.

When a declarative session description is used in session setup,

the sprop-deint-buf-req parameter signals the requirement for the deinterleaving buffer size. It is therefore RECOMMENDED to set the deinterleaving buffer size, in terms of number of bytes, equal to or

greater than the value of sprop-deint-buf-req parameter.

11.2.2 Deinterleaving Process

There are two buffering states in the receiver: initial buffering and buffering while playing. Initial buffering occurs when the RTP session is initialized. After initial buffering, decoding and playback is started, and the buffering-while-playing mode is used.

Regardless of the buffering state, the receiver stores incoming NAL units, in reception order, in the deinterleaving buffer as follows. NAL units of aggregation packets are stored in the deinterleaving buffer individually. The value of DON is calculated and stored for all NAL units.

The receiver operation is described below with the help of the following functions and constants:

- Function AbsDON is specified in section 7.1.
- Function don_diff is specified in section 5.4.
- Constant N is the value of the OPTIONAL sprop-interleaving-depth media type parameter (see section 7.1) incremented by 1.

Initial buffering lasts until one of the following conditions is fulfilled:

- There are N NAL units in the deinterleaving buffer.
- If sprop-max-don-diff is present, don_diff(m,n) is greater than the value of sprop-max-don-diff, in which n corresponds to the NAL unit having the greatest value of AbsDON among the received NAL units and m corresponds to the NAL unit having the smallest value of AbsDON among the received NAL units.
- Initial buffering has lasted for the duration equal to or greater than the value of the OPTIONAL sprop-init-buf-time parameter.

The NAL units to be removed from the deinterleaving buffer are determined as follows:

- If the deinterleaving buffer contains at least N NAL units, NAL units are removed from the deinterleaving buffer and passed to the decoder in the order specified below until the buffer contains (N-1) NAL units.
- If sprop-max-don-diff is present, all NAL units m for which don_diff(m,n) is greater than sprop-max-don-diff are removed from

the deinterleaving buffer and passed to the decoder in the order specified below. Herein, n corresponds to the NAL unit having the greatest value of AbsDON among the received NAL units and m

correnspondes to the being measured NAL units.

The order in which NAL units, whitch is removed from the deinterleaving buffer, are passed to the decoder is specified as follows:

- Let PDON be a variable that is initialized to 0 at the beginning of an RTP session.

- For each NAL unit associated with a value of DON, a DON distance is calculated as follows: If the value of DON of the NAL unit is larger than the value of PDON, the DON distance is equal to DON - PDON. Otherwise, the DON distance is equal to 65535 - PDON + DON + 1.

- NAL units are delivered to the decoder in ascending order of DON distance. If several NAL units share the same value of DON distance, they can be passed to the decoder in any order.

- When the number of NAL units have been only (N-1), the value of PDON is set to the value of DON for the last NAL unit passed to the decoder.

11.3. Additional De-Packetization Guidelines

The following additional de-packetization rules may be used to implement an operational AVS-P2 Video de-packetizer:

- RTP receivers (e.g., in gateways) may identify lost coded slice data partitions A (DPAs). If a lost DPA is found, a gateway may decide not to send the corresponding coded slice data partitions, as their information is meaningless for AVS-P2 Video decoders. In this way a MANE can reduce network load by discarding useless packets without parsing a complex bitstream.

- Receivers having to discard packets or NALUs SHOULD first discard all packets/NALUs in which the value of the NRI field of the NAL unit type octet is equal to 0. This will minimize the impact on user experience and keep the reference pictures intact. If more packets have to be discarded, then packets with a numerically lower NRI value SHOULD be discarded before packets with a numerically higher NRI value. However, discarding any packets with an NRI bigger than 0 very likely leads to decoder drift and SHOULD be avoided.

12. References

12.1 Normative references

[1] Standardization Administration of China, "GB/T 20090.2-2006, Information technology - Advanced coding of audio and video, Part 2: Video", March, 2006.

[2] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
[3] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.
[4] Handley, M. and V. Jacobson, "SDP: Session Description Protocol", RFC 2327, April 1998.
[5] Freed, N. and Klensin, J., "Media Type Specifications and Registration Procedures", BCP 13, RFC 4288, December 2005.
[6] Casner, S. and P. Hoschka, "MIME Type Registration of RTP Payload Formats", RFC 3555, July 2003.
[7] Josefsson, S., Ed., "The Base16, Base32, and Base64 Data Encodings", RFC 3548, July 2003.
[8] Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model with Session Description Protocol (SDP)", RFC 3264, June 2002.

12.2 Informative references

[9] Wang X.F., and Zhao D.B., "Performance Comparison of AVS and H.264/AVC Video Coding Standards," Journal of Computer Science & Technology. Vol. 21, No. 3, pp310-314, May 2006.
[10]Wenger S., Hannuksela M.M., Stockhammer T., Westerlund M., and Singer D., "RTP Payload Format for H.264 Video", RFC 3984, February 2005.
[11]Schulzrinne, H., Rao, A., and R. Lanphier, "Real Time Streaming Protocol (RTSP)", RFC 2326, April 1998.
[12] Handley, M., Perkins, C., and E. Whelan, "Session Announcement Protocol", RFC 2974, October 2000.
[13]Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, July 2003.
[14]Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.

Author's Addresses

Longshe Huo
Peking University
School of EE & CS
#5 YiHeYuan Road, Haidian District
Beijing, 100871
P.R. China
Email: lshuo@jdl.ac.cn

Lei Wang
Beijing Univ. of P&T
School of Telecom Engineering

Beijing University of Posts and Telecommunications
#10 XiTuCheng Road, Haidian District
Beijing, 100876
P.R. China

Phone: +861062282720
Email: wanglei_elf@bbn.cn

IPR Notices

Full Copyright Statement