

Audio/Video Transport Payloads WG
Internet Draft
Intended status: Standards track
Expires: December 2012

Y.-K. Wang
Qualcomm Inc.
T. Schierl
Fraunhofer HHI
R. Skupin
Fraunhofer HHI
P. Yue
Huawei Technologies
June 25, 2012

**RTP Payload Format for MVC Video
draft-ietf-payload-rtp-mvc-02.txt**

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on December 25, 2012.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the [Trust Legal Provisions](#) and are provided without warranty as described in the Simplified BSD License.

Abstract

This memo describes an RTP payload format for Multiview Video Coding (MVC), the multiview extension of the ITU-T Recommendation H.264 video codec that is technically identical to ISO/IEC International Standard 14496-10. The RTP payload format allows for packetization of one or more Network Abstraction Layer (NAL) units, produced by the video encoder, in each RTP payload. The payload format can be applied in RTP based 3D video transmissions such as such as 3D video streaming, free-viewpoint video, and 3DTV.

Table of Contents

- [1. Introduction.....3](#)
- [1.1. The MVC Codec.....4](#)
- [1.1.1. Overview.....4](#)
- [1.1.2. Parameter Set Concept.....4](#)
- [1.1.3. Network Abstraction Layer Unit Header.....5](#)
- [1.2. Overview of the Payload Format.....7](#)
- [1.2.1. Design Principles.....8](#)
- [1.2.2. Transmission Modes and Packetization Modes.....8](#)
- [2. Conventions.....8](#)
- [3. Definitions and Abbreviations.....9](#)
- [3.1. Definitions.....9](#)
- [3.1.1. Definitions per MVC specification.....9](#)
- [3.1.2. Definitions Specific to this memo.....10](#)
- [3.1. Abbreviations.....11](#)
- [4. MVC RTP Payload Format.....11](#)
- [4.1. RTP Header Usage.....11](#)
- [4.2. Common Structure of the RTP Payload Format.....11](#)
- [4.3. NAL Unit Header Usage.....11](#)
- [4.4. Packetization Modes.....12](#)
- 4.4.1. Packetization Modes for single-session transmission.12
- [4.4.2. Packetization Modes for multi-session transmission..13](#)
- [4.5. Aggregation Packets.....13](#)
- [4.6. Fragmentation Units \(FUs\).....13](#)
- 4.7. Payload Content Scalability Information (PACSI) NAL Unit for MVC.....13
- 4.8. Non-Interleaved Multi-Time Aggregation Packets (NI-MTAPs)17
- 4.9. Cross-Session DON (CS-DON) for multi-session transmission17
- [5. Packetization Rules.....17](#)
- [6. De-Packetization Process \(Informative\).....19](#)
- [7. Payload Format Parameters.....19](#)
- [7.1. Media Type Registration.....19](#)

| | | |
|------------------------|--|--------------------|
| 7.2. | SDP Parameters..... | 24 |
| 7.2.1. | Mapping of Payload Type Parameters to SDP..... | 24 |
| 7.2.2. | Usage with the SDP Offer/Answer Model..... | 24 |
| 7.2.3. | Usage with multi-session transmission..... | 24 |
| 7.2.4. | Usage in Declarative Session Descriptions..... | 24 |
| 7.3. | Examples..... | 24 |
| 7.4. | Parameter Set Considerations..... | 24 |
| 8. | Security Considerations..... | 24 |
| 9. | Congestion Control..... | 25 |
| 10. | IANA Considerations..... | 25 |
| 11. | Acknowledgments..... | 25 |
| 12. | References..... | 25 |
| 12.1. | Normative References..... | 25 |
| 12.2. | Informative References..... | 26 |
| | Author's Addresses..... | 26 |
| 13. | Open issues..... | 27 |
| 14. | Changes Log..... | 28 |

[1. Introduction](#)

This memo specifies an RTP [[RFC3550](#)] payload format for the multiview extension of the H.264/AVC video coding standard, also known as Multiview Video Coding (MVC). MVC is specified in Annex H of ITU-T Rec. H.264 [[H.264](#)] | ISO/IEC 14496 Part 10 [[MPEG4-10](#)].

MVC covers a wide range of 3D video applications, including 3D video streaming, free-viewpoint video as well as 3DTV.

This memo follows a backward compatible enhancement philosophy, by keeping as close an alignment to the H.264/AVC payload format [[RFC6184](#)] as possible. It documents the enhancements relevant from an RTP transport viewpoint, and defines signaling support for MVC, including a new media subtype name.

Due to the similarity between MVC and Scalable Video Coding (SVC), as defined in Annex G of H.264 [[H.264](#)], in system and transport aspects, this memo reuses the design principles as well as many features of the SVC RTP payload draft [[RFC6190](#)].

[Ed.Note(TS):Need text on session multiplexing and on the relation of this draft to [[RFC6190](#)] here.]

[1.1. The MVC Codec](#)

[1.1.1. Overview](#)

MVC provides multi-view video bitstreams. An MVC bitstream contains a base view conforming to at least one of the profiles of H.264/AVC defined in Annex A of [\[H.264\]](#), and one or more non-base views. To enable high compression efficiency, coding of a non-base view can utilize other views for inter-view prediction, thus its decoding relies on the presence of the views it depends on. Each coded view itself may be temporally scalable. Besides temporal scalability, MVC also supports view scalability, wherein a subset of the encoded views can be extracted, decoded and displayed, whenever it is desired by the application.

The concept of video coding layer (VCL) and network abstraction layer (NAL) is inherited from H.264/AVC. The VCL contains the signal processing functionality of the codec; mechanisms such as transform, quantization, motion-compensated prediction, loop filtering and inter-view prediction. The NAL encapsulates each slice generated by the VCL into one or more NAL units. Please consult [RFC 6184](#) for a more in-depth discussion of the NAL unit concept. MVC specifies the decoding order of NAL units.

In MVC, one access unit contains all NAL units pertaining to one output time instance for all the views. Within one access unit, the coded representation of each view, also named as view component, consists of one or more slices.

The concept of temporal scalability is not newly introduced by SVC or MVC, as profiles defined in Annex A of [\[H.264\]](#) already support it. In [\[H.264\]](#), sub-sequences have been introduced in order to allow optional use of temporal layers. SVC extended this approach by advertising the temporal scalability information within the NAL unit header or prefix NAL units, both were inherited to MVC.

[1.1.2. Parameter Set Concept](#)

The parameter set concept was first specified in [\[H.264\]](#). Please refer to [section 1.2 of \[RFC6184\]](#) for more details. SVC introduced some new parameter set mechanisms. MVC has inherited the parameter set concept from [\[H.264\]](#).

In particular, a different type of sequence parameter set (SPS), which is referred to as subset SPS, using a different NAL unit type than "the old SPS" specified in [\[H.264\]](#) is used for non-base views, while the base view still uses "the old SPS". Slices from different

views would be able to use either 1) the same sequence or picture parameter set, or 2) different sequence or picture parameter sets.

The inter-view dependency and the decoding order of all the encoded views are indicated in a new syntax structure, the SPS MVC extension, included in each subset SPS.

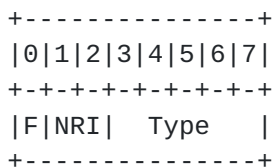
1.1.3. Network Abstraction Layer Unit Header

An MVC NAL unit of type 20 or 14 consists of a header of four octets and the payload byte string. MVC NAL units of type 20 are coded slices of non-base views. A special type of an MVC NAL unit is the prefix NAL unit (type 14) that includes descriptive information of the associated H.264/AVC VCL NAL unit (type 1 or 5) that immediately follows the prefix NAL unit.

MVC extends the one-byte H.264/AVC NAL unit header by three additional octets. The header indicates the type of the NAL unit, information regarding the relative importance of the NAL unit for the decoding process, the view identification information, the temporal layer identification information, and other fields as discussed below.

The syntax and semantics of the NAL unit header are formally specified in [H.264], but the essential properties of the NAL unit header are summarized below.

The first byte of the NAL unit header has the following format (the bit fields are the same as defined for the one-byte H.264/AVC NAL unit header, while the semantics of some fields have changed slightly, in a backward compatible way):



F: 1 bit

forbidden_zero_bit. H.264/AVC declares a value of 1 as a syntax violation.

NRI: 2 bits

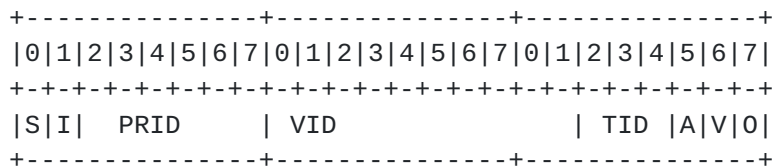
nal_ref_idc. A value of 00 indicates that the content of the NAL unit is not used to reconstruct reference pictures for future

prediction. Such NAL units can be discarded without risking the integrity of the reference pictures in the same view. A value higher than 00 indicates that the decoding of the NAL unit is required to maintain the integrity of reference pictures in the same view, or that the NAL unit contains parameter sets.

Type: 5 bits

nal_unit_type. This component specifies the NAL unit type.

In H.264/AVC, NAL unit types 14 and 20 are reserved for future extensions. MVC uses these two NAL unit types. NAL unit type 14 is used for prefix NAL unit, and NAL unit type 20 is used for coded slice of non-base view. NAL unit types 14 and 20 indicate the presence of three additional octets in the NAL unit header, as shown below.



S: 1 bit

svc_extention_flag. MUST be equal to 0 in MVC context. In the context of Scalable Video Coding (SVC), the flag must be equal to 1.

I: 1 bit

non_idr_flag. This component specifies whether the access unit the NAL unit belongs to is an IDR access unit (when equal to 0) or not (when equal to 1), as specified in [H.264].

PRID: 6 bits

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

VID: 10 bits

view_id. This component specifies the view identifier of the view the NAL unit belongs to.

TID: 3 bits

temporal_id. This component specifies the temporal layer (or frame rate) hierarchy. Informally put, a temporal layer consisting of view component with a less temporal_id corresponds to a lower frame rate. A given temporal layer typically depends on the lower temporal layers (i.e. the temporal layers with less temporal_id values) but never depends on any higher temporal layer (i.e. a temporal layer with a greater temporal_id value).

A: 1 bit

anchor_pic_flag. This component specifies whether the access unit the NAL unit belongs to is an anchor access unit (when equal to 1) or not (when equal to 0), as specified in [H.264].

V: 1 bit

inter_view_flag. This component specifies whether the view component is used for inter-view prediction (when equal to 1) or not (when equal to 0).

O: 1 bit

reserved_one_bit. Reserved bit for future extension. R shall be equal to 1. Receivers SHOULD ignore the value of reserved_zero_one_bit. This memo reuses the same additional NAL unit types introduced in RFC 6190, which are presented in section 4.2. In addition, this memo introduces one more NAL unit type, 30, as specified in section 4.7. These NAL unit types are marked as unspecified in [H.264] and intentionally reserved for use in systems specifications like this memo. Moreover, this specification extends the semantics of F, NRI, PRID, TID, A, and I as described in section 4.3.

1.2. Overview of the Payload Format

This payload specification can only be used to carry the "naked" NAL unit stream over RTP, and not the byte stream format according to Annex B of [H.264]. Likely, the applications of this specification will be in the IP based multimedia communications fields including 3D video streaming over IP, free-viewpoint video over IP, and 3DTV over IP.

This specification allows, in a given RTP packet stream, to encapsulate NAL units belonging to

- o the base view only, detailed specification in [[RFC6184](#)], or
- o one or more non-base views, or
- o the base view and one or non-base views

[Ed.Note(Ykw): To be extended to allow separate carriage of different temporal layers in different RTP packet streams as in [[RFC6190](#)].]

1.2.1. Design Principles

The following design principles have been observed:

- o Backward compatibility with [[RFC6184](#)] wherever possible.
- o As the MVC base view is H.264/AVC compatible, the base view or any H.264/AVC compatible subset of it, when transmitted in its own RTP packet stream, MUST be encapsulated using [[RFC6184](#)]. Requiring this has the desirable side effect that the transmitted data can be received by [[RFC6184](#)] receivers and decoded by H.264/AVC decoders.
- o Media-Aware Network Elements (MANEs) as defined in [[RFC6184](#)] are signaling aware and rely on signaling information. MANEs have state.
- o MANEs can aggregate multiple RTP streams, possibly from multiple RTP sessions.
- o MANEs can perform media-aware stream thinning. By using the payload header information identifying Layers within an RTP session, MANEs are able to remove packets from the incoming RTP packet stream. This implies rewriting the RTP headers of the outgoing packet stream and rewriting of RTCP Receiver Reports.

1.2.2. Transmission Modes and Packetization Modes

Please see [section 1.2.2 of \[RFC6190\]](#).

2. Conventions

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

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

3. Definitions and Abbreviations

3.1. Definitions

3.1.1. Definitions per the MVC Specification

This document uses the definitions of [H.264]. The following terms, defined in [H.264], are summed up for convenience:

access unit: A set of NAL units always containing exactly one primary coded picture with one or more view components. In addition to the primary coded picture, an access unit may also contain one or more redundant coded pictures, one auxiliary coded picture, or other NAL units not containing slices or slice data partitions of a coded picture. The decoding of an access unit always results in one decoded picture. All slices or slice data partitions in an access unit have the same value of picture order count.

anchor access unit: An access unit in which the primary coded picture is an anchor picture.

anchor picture: A coded picture in which all slices may reference only slices within the same access unit, i.e., inter-view prediction may be used, but no inter prediction is used, and all following coded pictures in output order do not use inter prediction from any picture prior to the coded picture in decoding order. The value of `anchor_pic_flag` is equal to 1 for all the prefix NAL units (when present) and all the slice extension NAL units that are contained in an anchor picture.

base view: A bitstream subset that contains all the NAL units with the `nal_unit_type` syntax element equal to 1, 5 or 14 of the bitstream and does not contain any NAL unit with the `nal_unit_type` syntax element equal to 15, or 20 and conforms to one or more of the profiles specified in Annex A of [H.264].

prefix NAL unit: A NAL unit with `nal_unit_type` equal to 14 that immediately precedes a NAL unit with `nal_unit_type` equal to 1, 5, or 12. The NAL unit that succeeds the prefix NAL unit is also referred to as the associated NAL unit. The prefix NAL unit

contains data associated with the associated NAL unit, which are considered to be part of the associated NAL unit.

target output view: A view that is targeted for output.

view component: An access unit subset containing only NAL units that share to the same view identifier.

3.1.2. Definitions Specific to this Memo

MVC NAL unit: A NAL unit of NAL unit type 14 or 20 as specified in Annex H of [[H.264](#)]. An MVC NAL unit has a four-byte NAL unit header.

operation point: An operation point of an MVC bitstream represents a certain level of temporal and view scalability. An operation point contains only those NAL units required for a valid bitstream to represent a certain subset of views at a certain temporal level. An operation point is described by the view_id values of the subset of views, and the highest temporal_id.

multi-session transmission: The transmission mode in which the MVC bitstream is transmitted over multiple RTP sessions, with each stream having the same SSRC. These multiple RTP streams can be associated using the RTCP CNAME, or explicit signalling of the SSRC used. Dependency between RTP sessions MUST be signaled according to [[RFC5583](#)] and this memo.

single-session transmission: The transmission mode in which the MVC bitstream is transmitted over a single RTP session, with a single SSRC and separate timestamp and sequence number spaces.

cross-session decoding order number (CS-DON): A derived variable indicating NAL unit decoding order number over all NAL units within all the session-multiplexed RTP sessions that carry the same MVC bitstream.

[Ed.Note(TS):Need more definitions here.]

3.1. Abbreviations

In addition to the abbreviations defined in [\[RFC6184\]](#), the following ones are defined.

MVC: Multiview Video Coding
CS-DON: Cross-Session Decoding Order Number
MST: multi-session transmission
PACSI: Payload Content Scalability Information
SST: single-session transmission

4. MVC RTP Payload Format

4.1. RTP Header Usage

Please see [section 5.1 of \[RFC6184\]](#).

4.2. Common Structure of the RTP Payload Format

Please see [section 5.2 of \[RFC6184\]](#).

4.3. NAL Unit Header Usage

The structure and semantics of the NAL unit header were introduced in [section 1.1.3](#). This section specifies the semantics of F, NRI, PRID, TID, A and I according to this specification.

Note that, in the context of this section, "protecting a NAL unit" means any RTP or network transport mechanism that could improve the probability of success delivery of the packet conveying the NAL unit, including applying a QoS-enabled network, forward error correction (FEC), retransmissions, and advanced scheduling behavior, whenever possible.

The semantics of F specified in [section 5.3 of \[RFC6184\]](#) also applies herein.

For NRI, for a bitstream conforming to one of the profiles defined in Annex A of [\[H.264\]](#) and transported using [\[RFC6184\]](#), the semantics specified in [section 5.3 of \[RFC6184\]](#) are applicable, i.e., NRI also indicates the relative importance of NAL units. In MVC context, in addition to the semantics specified in Annex H of [\[H.264\]](#) are applicable, NRI also indicate the relative importance of NAL units within a view. MANES MAY use this information to protect more important NAL units better than less important NAL units.

[Ed.Note(Ykw): "MVC context" to be clearly specified.]

For PRID, the semantics specified in Annex H of [H.264] applies. Note that MANEs implementing unequal error protection MAY use this information to protect NAL units with smaller PRID values better than those with larger PRID values, for example by including only the more important NAL units in a forward error correction (FEC) protection mechanism. The importance for the decoding process decreases as the PRID value increases.

For TID, in addition to the semantics specified in Annex H of [H.264], according to this memo, values of TID indicate the relative importance. A lower value of TID indicates a higher importance for NAL units within a view. MANEs MAY use this information to protect more important NAL units better than less important NAL units.

For A, in addition to the semantics specified in Annex H of [H.264], according to this memo, MANEs MAY use this information to protect NAL units with A equal to 1 better than NAL units with A equal to 0. MANEs MAY also utilize information of NAL units with A equal to 1 to decide when to forward more packets for an RTP packet stream. For example, when it is sensed that view switching has happened such that the operation point has changed, MANEs MAY start to forward NAL units for a new target view only after forwarding a NAL unit with A equal to 1 for the new target view.

For I, in addition to the semantics specified in Annex H of [H.264], according to this memo, MANEs MAY use this information to protect NAL units with I equal to 1 better than NAL units with I equal to 0. MANEs MAY also utilize information of NAL units with I equal to 1 to decide when to forward more packets for an RTP packet stream. For example, when it is sensed that view switching has happened such that the operation point has changed, MANEs MAY start to forward NAL units for a new target view only after forwarding a NAL unit with I equal to 1 for the new target view.

4.4. Packetization Modes

[Ed.Note(TS): Need to add text from [RFC6190] to this section with respect to MVC.]

4.4.1. Packetization Modes for Single-Session Transmission

This section will address the issues of [section 4.5.1](#) and 5.1 of [\[RFC6190\]](#).

4.4.2. Packetization Modes for Multi-Session Transmission

This section will address the issues of [section 4.5.2](#) and 5.2 of [\[RFC6190\]](#).

4.5. Aggregation Packets

This section will address the issues of [section 4.7 of \[RFC6190\]](#).

4.6. Fragmentation Units (FUs)

This section will address the issues of [section 4.8 of \[RFC6190\]](#).

4.7. Payload Content Scalability Information (PACSI) NAL Unit for MVC

A new NAL unit type is specified in this memo, and referred to as payload content scalability information (PACSI) NAL unit. The PACSI NAL unit, if present, MUST be the first NAL unit in an aggregation packet, and it MUST NOT be present in other types of packets. The PACSI NAL unit indicates view and temporal scalability information and other characteristics that are common for all the remaining NAL units in the payload of the aggregation packet. Furthermore, a PACSI NAL unit MAY include a DONC field and contain zero or more SEI NAL units. PACSI NAL unit makes it easier for MANEs to decide whether to forward/process/discard the aggregation packet containing the PACSI NAL unit. Senders MAY create PACSI NAL units and receivers MAY ignore them, or use them as hints to enable efficient aggregation packet processing. Note that the NAL unit type for the PACSI NAL unit is selected among those values that are unspecified in [\[H.264\]](#) and [\[RFC6184\]](#).

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

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

The structure of a PACSI NAL unit is as follows. The first four octets are exactly the same as the four-byte MVC NAL unit header as discussed in [section 4.3](#). They are followed by two always present octet, two optional octets, and zero or more SEI NAL units, each SEI

NAL unit preceded by a 16-bit unsigned size field (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 SEI NAL unit). Figure 1 illustrates the PACSI NAL unit structure and an example of a PACSI NAL unit containing two SEI NAL units.

The bits P, C, S, and E are specified only if the bit X is equal to 1. The T bit MUST NOT be equal to 1 if the aggregation packet containing the PACSI NAL unit is not an STAP-A packet. The T bit MAY be equal to 1 if the aggregation packet containing the PACSI NAL unit is an STAP-A packet. The field DONC MUST NOT be present if the T bit is equal to 0, and MUST be present if the T bit is equal to 1.

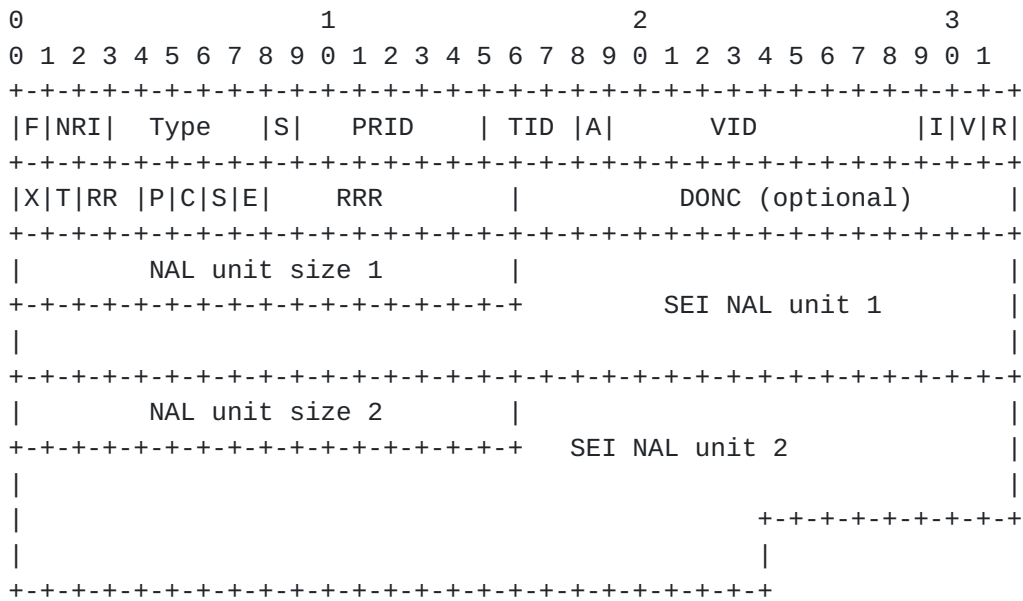


Figure 1. PACSI NAL unit structure

The values of the fields in PACSI NAL unit MUST be set as follows. The term "target NAL units" are used in the semantics of some fields. The target NAL units are such NAL units contained in the aggregation packet, but not included in the PACSI NAL unit, that are within the access unit to which the first NAL unit following the PACSI NAL unit in the aggregation packet belongs.

- o The F bit MUST be set to 1 if the F bit in at least one of the remaining NAL units in the aggregation packet is equal to 1. Otherwise, the F bit MUST be set to 0.

- o The NRI field MUST be set to the highest value of NRI field among all the remaining NAL units in the aggregation packet.
- o The Type field MUST be set to 30.
- o The S bit MUST be set to 1.
- o The PRID field MUST be set to the lowest value of the PRID values of all the remaining NAL units in the aggregation packet.
- o The TID field MUST be set to the lowest value of the TID values of all the remaining NAL units with the lowest value of VID in the aggregation packet.
- o The A bit MUST be set to 1 if the A bit of at least one of the remaining NAL units in the aggregation packet is equal to 1. Otherwise, the A bit MUST be set to 0.
- o The VID field MUST be set to the lowest value of the VID values of all the remaining NAL units in the aggregation packet.
- o The I bit MUST be set to 1 if the I bit of at least one of the remaining NAL units in the aggregation packet is equal to 1. Otherwise, the I bit MUST be set to 0.
- o The V bit MUST be set to 1 if the V bit of at least one of the remaining NAL units in the aggregation packet is equal to 1. Otherwise, the A bit MUST be set to 0.
- o The R bit MUST be set to 0. Receivers SHOULD ignore the value of R.
- o If the X bit is equal to 1, the bits P, C, S, and E are specified as below. Otherwise, the bits P, C, S, and E are unspecified, and receivers MUST ignore these bits. The X bit SHOULD be identical for all the PACSI NAL units involved in all the RTP sessions conveying an MVC bitstream.
- o The RR field MUST be set to '00' (in binary form). Receivers SHOULD ignore the value of RR.
- o If the T bit is equal to 1, the OPTIONAL field DONC MUST be present and specified as below. Otherwise, the field DONC MUST NOT be present.
- o The P bit MUST be set to 1 if all the remaining NAL units in the aggregation packet are with `redundant_pic_cnt` higher than 0, i.e.

the slices are redundant slices. Otherwise, the P bit MUST be set to 0.

Informative note: The P bit indicates whether the packet can be discarded because it contains only redundant slice NAL units. Without this bit, the corresponding information can be concluded from the syntax element `redundant_pic_cnt`, which is buried in the variable-length coded slice header.

o The C bit MUST be set to 1 if the target NAL units belong to an access unit for which the view components are intra coded. Otherwise, the C bit MUST be set to 0. The C bit SHOULD be identical for all the PACSI NAL units for which the target NAL units belong to the same access unit.

Informative note: The C bit indicates whether the packet contains intra slices which may be the only packets to be forwarded for a fast forward playback, e.g. when the network condition is extremely bad.

o The S bit MUST be set to 1, if the first VCL NAL unit, in transmission order, of the view component containing the first NAL unit following the PACSI NAL unit in the aggregation packet is present in the aggregation packet. Otherwise, the S bit MUST be set to 0.

o The E bit MUST be set to 1, if the last VCL NAL unit, in transmission order, of the view component containing the first NAL unit following the PACSI NAL unit in the aggregation packet is present in the aggregation packet. Otherwise, the E field MUST be set to 0.

Informative note: The S or E bit indicates whether the first or last slice, in transmission order, of a view component is in the packet, to enable a MANE to detect slice loss and take proper action such as requesting a retransmission as soon as possible, as well as to allow an efficient playout buffer handling similarly as the M bit in the RTP header. The M bit in the RTP header still indicates the end of an access unit, not the end of a view component.

o The RRR field MUST be set to '00000000'(in binary form). Receivers SHOULD ignore the value of RRR.

o When present, the field DONC indicates the CL-DON value for the first NAL unit in the STAP-A in transmission order.

SEI NAL units included in the PACSI NAL unit, if any, MUST contain a subset of the SEI messages associated with the access unit of the first NAL unit following the PACSI NAL unit within the aggregation packet.

Informative note: Senders may repeat such SEI NAL units in the PACSI NAL unit the presence of which in more than one packet is essential for packet loss robustness. Receivers may use the repeated SEI messages in place of missing SEI messages.

An SEI message SHOULD NOT be included in a PACSI NAL unit and included in one of the remaining NAL units contained in the same aggregation packet.

4.8. Non-Interleaved Multi-Time Aggregation Packets (NI-MTAPs)

This section will address the issues of [section 4.7.1 of \[RFC6190\]](#).

4.9. Cross-Session DON (CS-DON) for Multi-Session Transmission

This section will address the issues of [section 4.11 of \[RFC6190\]](#).

5. Packetization Rules

[Ed.Note(TS): We need to adjust this section with respect to [\[RFC6190\]](#).]

[Section 6 of \[RFC6184\]](#) applies. The following rules apply in addition.

All receivers MUST support the single NAL unit packetization mode to provide backward compatibility to endpoints supporting only the single NAL unit mode of [RFC 3984](#). However, the single NAL unit packetization mode SHOULD NOT be used whenever possible, because encapsulating NAL units of small sizes, e.g. small NAL units containing parameter sets, SEI messages or prefix NAL units, in their own packets is typically less efficient because of the relatively big overhead.

All receivers MUST support the non-interleaved packetization mode.

Informative note: The non-interleaved mode allows an application to encapsulate a single NAL unit in a single RTP packet. Historically, the single NAL unit mode has been included into [\[RFC6184\]](#) only for compatibility with ITU-T Rec. H.241 Annex A [\[H.241\]](#). There is no point in carrying this historic ballast towards a new application space such as the one provided with

MVC. More technically speaking, the implementation complexity increase for providing the additional mechanisms of the non-interleaved mode (namely STAP-A and FU-A) is minor, and the benefits are great, that STAP-A implementation is required.

A NAL unit of small size SHOULD be encapsulated in an aggregation packet together with one or more other NAL units. For example, non-VCL NAL units such as access unit delimiter, parameter set, or SEI NAL unit are typically small.

A prefix NAL unit SHOULD be aggregated to the same packet as the associated NAL unit following the prefix NAL unit in decoding order.

When the first aggregation unit of an aggregation packet contains a PACSI NAL unit, there MUST be at least one additional aggregation unit present in the same packet.

When an MVC bitstream is transported in more than one RTP session, the following applies.

- o Interleaved mode SHOULD be used for all the RTP sessions.
- o An RTP session that does not use interleaved mode SHOULD be constrained as follows.
 - Non-interleaved mode MUST be used.
 - STAP-A MUST be used, and any other type of packets MUST NOT be used.
 - Each STAP-A MUST contain a PACSI NAL unit and the DONC field MUST be present in the PACSI NAL unit.

Informative note: The motivation for these constraints is to allow the use of non-interleaved mode for the session conveying the H.264/AVC compatible view, such that [RFC 3984](#) receivers without interleaved mode implementation can subscribe to the base view session.

Non-VCL NAL units SHOULD be conveyed in the same session as the associated VCL NAL units. To meet this, SEI messages that are contained in scalable nesting SEI message and are applicable to more than one session SHOULD be separated and contained into multiple scalable nesting SEI messages. The DON values MUST indicate the cross-layer decoding order number values as if all these SEI messages were in separate scalable nesting SEI messages and

contained in the beginning of the corresponding access units as specified in [[H.264](#)].

6. De-Packetization Process (Informative)

For a single RTP session, the de-packetization process specified in [section 7 of \[RFC6184\]](#) applies.

For receiving more than one of multiple RTP sessions conveying a scalable bitstream, an example of a suitable implementation of the de-packetization process is to be specified similarly as what will be finally included in [[RFC6190](#)].

7. Payload Format Parameters

This section specifies the 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 type registration for the MVC codec. A mapping of the parameters into the Session Description Protocol (SDP) [[RFC4566](#)] is also provided for applications that use SDP. Equivalent parameters could be defined elsewhere for use with control protocols that do not use SDP.

7.1. Media Type Registration

The media subtype for the MVC codec is allocated from the IETF tree.

The receiver MUST ignore any unspecified parameter.

Informative note: Requiring ignoring unspecified parameter allows for backward compatibility of future extensions. For example, if a future specification that is backward compatible to this specification specifies some new parameters, then a receiver according to this specification is capable of receiving data per the new payload but ignoring those parameters newly specified in the new payload specification. This sentence is also present in [RFC 3984](#).

Media Type name: video

Media subtype name: H264-MVC

The media subtype "H264" MUST be used for RTP streams using [RFC 3984](#), i.e. not using any of the new features introduced by this specification compared to [RFC 3984](#). For RTP streams using any of the new features introduced by this specification compared to RFC

3984, the media subtype "H264-MVC" SHOULD be used, and the media subtype "H264" MAY be used. Use of the media subtype "H264" for RTP streams using the new features allows for [RFC 3984](#) receivers to negotiate and receive H.264/AVC or MVC streams packetized according to this specification, but to ignore media parameters and NAL unit types it does not recognize.

Required parameters: none

OPTIONAL parameters:

In the following definitions of parameters, "the stream" or "the NAL unit stream" refers to all NAL units conveyed in the current RTP session in SST, and all NAL units conveyed in the current RTP session and all NAL units conveyed in other RTP sessions that the current RTP session depends on in MST.

profile-level-id:

sprop-view-scalability-info:

This parameter MAY be used to convey the NAL unit containing the view scalability information SEI message as specified in Annex H of [\[H.264\]](#). This parameter MAY be used to signal the contained target temporal level and target output views of an MVC bitstream. The parameter MUST NOT be used to indicate codec capability in any capability exchange procedure. The value of the parameter is the base64 [\[RFC4648\]](#) representation of the NAL unit containing the view scalability information SEI message. If present, the NAL unit MUST contain only one SEI message that is a view scalability information SEI message.

This parameter MAY be used in an offering or declarative SDP message to indicate what temporal level and output views (operation points) can be provided. A receiver MAY indicate its choice of one operation point using the optional media type parameter sprop-view-operation-point-id.

sprop-view-operation-point-id:

This parameter MAY be used to signal a receiver's choice of the offers or declared operation points using sprop-view-scalability-info or sprop-view-operation-point-info. The

value of `sprop-view-operation-point-id` is a base16 representation of the `operation_point_id[i]` syntax element in the view scalability information SEI message as specified in Annex H of [H.264] or operation-point-ID contained in `sprop-view-operation-point-info`.

`sprop-view-operation-point-info`

This parameter MAY be used to describe the operation points of an RTP session. The value of this parameter consists of a comma-separated list of view-operation-point-description vector. The values given by the view-operation-point-description vectors are the same as, or are derived from, the values that would be given for an operation point in the view scalability information SEI message as specified in Annex H of [H.264], where the term operation point in the view scalability information SEI message refers to those NAL units required for a valid bitstream to represent a certain subset of views at a certain temporal level. An operation point is described by the `view_id` values of the subset of views, and the highest `temporal_id`.

Each view-operation-point-description vector has variable number (depends on the number of view-IDs) of elements, provided as a comma-separated list of values as defined below. The first value of the view-operation-point-description vector is preceded by a '<', and the last value of the view-operation-point-description vector is followed by a '>'. If the `sprop-view-operation-point-info` is followed by exactly one view-operation-point-description vector, this describes the highest operation point contained in the RTP session. If there are two or more view-operation-point-description vectors, the first describes the lowest and the last describes the highest operation point contained in the RTP session.

The values given by the operation-point-description vector are as follows, in the order listed:

- operation-point-ID: This value specifies the identifier of the operation point, which is identical to the `operation_opint_id` that would be indicated (for the same values of a list of views and the highest `temporal_id`) in the view scalability information SEI message. This field MAY be empty, indicating that the value is unspecified. When there are multiple view-operation-point-description

vectors with operation-point-ID, the values of operation-point-ID do not need to be consecutive.

- temporal-ID: This value specifies the maximum value of `temporal_id` of the NAL units in the representation of the current operation point.
- num-target-output-views-minus1: This value plus 1 specifies the number of target output views for the current operation point.
- view-ID: each of this parameter specifies the identifier of a target output view for the current operation point. The number of this parameter depends on `num-target-output-views-minus1`.
- profile-level-ID: This value specifies the profile-level-idc of the operation point in the base16 format. The default sub-profile or default level indicated by the parameter `profile-level-ID` in the `sprop-view-operation-point-info` vector SHALL be equal to or lower than the default sub-profile or default level indicated by `profile-level-id`, which may be either present or the default value is taken. This field MAY be empty, indicating that the value is unspecified.
- avg_bitrate: This value specifies the average bit rate of the representation of the current operation point. This parameter is given as an integer in kbits per second over the entire stream. Note that this parameter is provided in the view scalability information SEI message in bits per second and calculated over a variable time window. This field MAY be empty, indicating that the value is unspecified.
- max_bitrate: This value specifies the maximum bitrate of the operation point. This parameter is given as an integer in kbits per second and describes the maximum bitrate per each one-second window. Note that this parameter is provided in the view scalability information SEI message in bits per second and is calculated over a variable time window. This field MAY be empty, indicating that the value is unspecified.
- avg_frm_rate: This value specifies the average frame rate of the operation point. This value is given as an integer

in frames per 256 seconds. The field MAY be empty, indicating that the value is unspecified.

Similarly to sprop-view-scalability-info, this parameter MAY be used in an offering or declarative SDP message to indicate what temporal level and output views (operation points) can be provided. A receiver MAY indicate its choice of the highest layer it wants to send and/or receive using the optional media type parameter sprop-view-operation-point-id.

[Editors' note: more parameters to be added]

Encoding considerations:

This type is only defined for transfer via RTP ([RFC 3550](#)).

Security considerations:

See [section 10](#) of RFC XXXX.

Public specification:

Please refer to RFC XXXX and its [section 14](#).

Additional information: none

File extensions: none

Macintosh file type code: none

Object identifier or OID: none

Person & email address to contact for further information:

Intended usage: COMMON

Author: NN

Change controller:

IETF Audio/Video Transport working group delegated from the IESG.

[7.2.](#) SDP Parameters

[7.2.1.](#) Mapping of Payload Type Parameters to SDP

The media type video/H264-MVC string is mapped to fields in the Session Description Protocol (SDP) as follows:

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

The encoding name in the "a=rtpmap" line of SDP MUST be H264-MVC (the media subtype).

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

The OPTIONAL parameters, when present, MUST be included in the "a=fmtp" line of SDP. These parameters are expressed as a media type string, in the form of a semicolon separated list of parameter=value pairs.

[7.2.2.](#) Usage with the SDP Offer/Answer Model

TBD.

[7.2.3.](#) Usage with Multi-Session Transmission

If multi-session transmission is used, the rules on signaling media decoding dependency in SDP as defined in [\[RFC5583\]](#) apply.

[7.2.4.](#) Usage in Declarative Session Descriptions

TBD.

[7.3.](#) Examples

TBD.

[7.4.](#) Parameter Set Considerations

Please see [section 10 of \[RFC6184\]](#).

[8.](#) Security Considerations

Please see [section 11 of \[RFC6184\]](#).

9. Congestion Control

TBD.

10. IANA Considerations

Request for media type registration to be added.

11. Acknowledgments

The work of Thomas Schierl has been supported by the European Commission under contract number FP7-ICT-248036, project COAST.

This document was prepared using 2-Word-v2.0.template.dot.

12. References

12.1. Normative References

- [H.264] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services", January 2012.
- [RFC6190] Wenger, S., Wang, Y. -K., Schierl, T. and A. Eleftheriadis, "RTP payload format for SVC video", [RFC 6190](#), May 2011.
- [RFC5583] Schierl, T., and Wenger, S., "Signaling media decoding dependency in the Session Description Protocol (SDP)", [RFC 5583](#), July 2009.
- [MPEG4-10]
ISO/IEC International Standard 14496-10:2012.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC3548] Josefsson, S., "The Base16, Base32, and Base64 Data Encodings", [RFC 3548](#), July 2003.
- [RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and Jacobson, V., "RTP: A Transport Protocol for Real-Time Applications", STD 64, [RFC 3550](#), July 2003.
- [RFC6184] Wang, Y.-K., Kristensen, T., Jesup, R., "RTP Payload Format for H.264 Video", [RFC 6184](#), May 2011.

[RFC4566] Handley, M., Jacobson, V., and Perkins, C., "SDP: Session Description Protocol", [RFC 4566](#), July 2006.

12.2. Informative References

- [DVB-H] DVB - Digital Video Broadcasting (DVB); DVB-H Implementation Guidelines, ETSI TR 102 377, 2005.
- [H.241] ITU-T Rec. H.241, "Extended video procedures and control signals for H.300-series terminals", May 2006.
- [IGMP] Cain, B., Deering S., Kovenlas, I., Fenner, B., and Thyagarajan, A., "Internet Group Management Protocol, Version 3", [RFC 3376](#), October 2002.
- [McCanne] McCanne, S., Jacobson, V., and Vetterli, M., "Receiver-driven layered multicast", in Proc. of ACM SIGCOMM'96, pages 117--130, Stanford, CA, August 1996.
- [MBMS] 3GPP - Technical Specification Group Services and System Aspects; Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs (Release 6), December 2005.
- [MPEG2] ISO/IEC International Standard 13818-2:1993.
- [RFC3450] Luby, M., Gemmell, J., Vicisano, L., Rizzo, L., and Crowcroft, J., "Asynchronous layered coding (ALC) protocol instantiation", [RFC 3450](#), December 2002.

Author's Addresses

Ye-Kui Wang
Qualcomm Incorporated
5775 Morehouse Drive
San Diego, CA 92121
USA
Phone: +1-858-651-8345
EMail: yekuiw@qualcomm.com

Thomas Schierl
Fraunhofer HHI
Einsteinufer 37
D-10587 Berlin
Germany
Phone: +49-30-31002-227
EMail: ts@thomas-schierl.de

Robert Skupin
Fraunhofer HHI
Einsteinufer 37
D-10587 Berlin
Germany
Phone: +49-30-314-21700
EMail: robert.skupin@hhi.fraunhofer.de

Peiyu Yue
Huawei Technologies
Huawei Base
101 Software Avenue, Yuhua District
Nanjing, Jiangsu 210012
China

Phone: +86-25-56620258
Email: yuepeiyu@huawei.com

13. Open Issues

- The use of CL-DON for session reordering allows also for interleaved transmission with non-interleaved packetization mode. There should be a clear separation between both tools. This issue should be handled the same way as for the SVC payload draft.
- Since SVC session multiplexing (multi source transmission(MST)) is cleared, it would be great to just reference the MST sections in [\[RFC6190\]](#). Since the text in sections [6](#) and [7](#) of [\[RFC6190\]](#) is currently very SVC specific, the authors would have to try to rewrite these sections in a more generic way. If this is not possible, we need to copy text from [\[RFC6190\]](#) with respect to MVC.
- The structure of this document should be aligned with recently finished [RFC6190](#).
- This document is not intended to be a delta document in respect to [RFC6190](#).
- The PASCII definition in this document differs from the definition in [RFC6190](#)

14. Changes Log

Initial version 00

10 November 2007: YkW
Initial version

12 November 2007: TS
- Added definition of "Session multiplexing"
- Added the reference of [I-D.[draft-ietf-mmusic-decoding-dependency](#)], and its reference in [section 9.2.3](#)

12 November 2007: YkW
- Added the reference of [I-D.[draft-ietf-avt-svc](#)] and its reference in [section 1](#).
- Added in sections [3.1](#) and [3.2](#) paragraphs regarding inter-view prediction

From [draft-wang-avt-rtp-mvc-00](#) to [draft-wang-avt-rtp-mvc-01](#)

18 February 2008: YkW
- Alignment to the latest MVC draft in JVT-Z209 and version 07 of [I-D.[draft-ietf-avt-svc](#)].

25 February 2008: TS

- Minor modifications and updates throughout the document

- Added open issue on clear separation between "decoding order recovery" and "interleaving"

From [draft-wang-avt-rtp-mvc-01](#) to [draft-wang-avt-rtp-mvc-02](#)

09 July 2008: TS

- Minor modifications and updates throughout the document

- Added open issue

- NAL unit header alignment with MVC spec

- [Section 6](#). References corresponding sections in [[RFC3984](#)] and [I-D.[draft-ietf-avt-svc](#)].

- TBD: [Section 7](#), we may align [I-D.[draft-ietf-avt-svc](#)] in a way that SVC is not mentioned in this paragraphs, so that we can reference them from this document.

21 August 2008:

- Minor modifications, editing and adding notes throughout the document.
- Updated references

From [draft-wang-avt-rtp-mvc-02](#) to [draft-wang-avt-rtp-mvc-03](#)

04 February 2009: YkW

- Updated author's address.

04 February 2009: YkW

- Updated the boiler template.

From [draft-wang-avt-rtp-mvc-03](#) to [draft-wang-avt-rtp-mvc-04](#)

22 October 2009: YkW

- Updated author's address and the boiler template (added the last sentence in Copyright Notice).

From [draft-wang-avt-rtp-mvc-04](#) to [draft-wang-avt-rtp-mvc-05](#)

22 April 2010: YkW

- To keep the draft alive, no change other than version number etc.

From [draft-wang-avt-rtp-mvc-05](#) to [draft-ietf-avt-rtp-mvc-00](#)

28 April 2010: YkW

- No change other than version number etc.

From [draft-ietf-avt-rtp-mvc-00](#) to [draft-ietf-avt-rtp-mvc-01](#)

8/9 October 2010:

- YkW: Updated the NAL unit header syntax and semantics in [section 3.3](#) per the latest MVC specification.

- TS: Minor edits

From [draft-ietf-avt-rtp-mvc-01](#) to [draft-ietf-payload-rtp-mvc-00](#)

14 March 2011: YkW

- Minor changes such as updates of some references the work group name from AVT to AVT Payload, etc.

From [draft-ietf-payload-rtp-mvc-00](#) to [draft-ietf-payload-rtp-mvc-01](#)

1 September 2011: RS

- Added some definitions
- Started structural alignment with [RFC 6190](#)
- Reference updates: ([RFC3984](#) -> [RFC6184](#)), (I-D.[draft-ietf-avt-rtp-svc](#) -> [RFC6190](#))

From [draft-ietf-payload-rtp-mvc-01](#) to [draft-ietf-payload-rtp-mvc-02](#)

- Added definitions of some media type parameters (PY)
- Minor changes throughout the document, including updating of some definitions and references (YkW)