Video Services Forum (VSF) Technical Recommendation TR-03

Transport of Uncompressed Elementary Stream Media over IP

November 12, 2015

VSF_TR-03_2015-11-12

http://www.videoservicesforum.org

INTELLECTUAL PROPERTY RIGHTS

THE POSSIBILITY HAS BEEN RAISED TO THE VSF THAT THE PRACTICE OR IMPLEMENTATION OF THIS RECOMMENDATION MAY INVOLVE THE USE OF A THIRD PARTY INTELLECTUAL PROPERTY RIGHT. HOWEVER, THE OWNER OF SUCH RIGHT HAS NOT AS YET ASSERTED SUCH RIGHT, NOR HAS THE VSF AS YET CONTACTED SUCH THIRD PARTY.

VSF TAKES NO POSITION CONCERNING THE EVIDENCE, VALIDITY OR APPLICABILITY OF THIS OR ANY OTHER INTELLECTUAL PROPERTY RIGHTS, WHETHER ASSERTED BY VSF MEMBERS OR OTHERS OUTSIDE OF THE RECOMMENDATION DEVELOPMENT PROCESS. IF A PATENT HOLDER ASSERTS SUCH RIGHTS AND FILES A STATEMENT WITH THE VSF OF WILLINGNESS TO GRANT A LICENSE UNDER SUCH RIGHTS ON REASONABLE AND NONDISCRIMINATORY TERMS AND CONDITIONS TO APPLICANTS DESIRING TO OBTAIN SUCH A LICENSE, THEN DETAILS MAY BE OBTAINED FROM THE VSF.

LIMITATION OF LIABILITY

VSF SHALL NOT BE LIABLE FOR ANY AND ALL DAMAGES, DIRECT OR INDIRECT, ARISING FROM OR RELATING TO ANY USE OF THE CONTENTS CONTAINED HEREIN, INCLUDING WITHOUT LIMITATION ANY AND ALL INDIRECT, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES (INCLUDING DAMAGES FOR LOSS OF BUSINESS, LOSS OF PROFITS, LITIGATION, OR THE LIKE), WHETHER BASED UPON BREACH OF CONTRACT, BREACH OF WARRANTY, TORT (INCLUDING NEGLIGENCE), PRODUCT LIABILITY OR OTHERWISE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THE FOREGOING NEGATION OF DAMAGES IS A FUNDAMENTAL ELEMENT OF THE USE OF THE CONTENTS HEREOF, AND THESE CONTENTS WOULD NOT BE PUBLISHED BY VSF WITHOUT SUCH LIMITATIONS.

Executive Summary

The carriage of live professional media over IP is expected to enhance the flexibility and agility of the video production plant. Its advantages include agnosticism to resolution, bit depth, and frame rate, compatibility with network interfaces on commodity Ethernet switches and commodity servers, flexible association of elementary streams into desired groups of media, and a natural mechanism to coordinate with network-based registration and discovery of devices, streams, and media capabilities.

This Technical Recommendation differs from current practices that bind video, audio, and ancillary data into the Serial Digital Interface (SDI) multiplex before encapsulation in IP, and instead provides for the carriage of video, audio, and ancillary data as separate elementary streams that can be individually forwarded through a network.

Table of Contents

1. Introduction (Informative)

The Video Services Forum Studio Video over IP (SVIP) Activity Group was tasked to develop or recommend a standard for Video over IP without SDI encapsulation.

To achieve this goal, The SVIP Activity Group studied and documented the requirements for Video over IP/Ethernet within the broadcast plant, including the following elements: Video, Audio, Ancillary Data, groups, Timing, Sequencing, Identities, and Latency.

The SVIP Activity Group researched current and proposed solutions, and then developed a gap analysis between requirements and existing solutions.

This Technical Recommendation references what the Activity Group feels are the best available existing standards to achieve its goal, with a minimum amount of invention. It is possible that a more perfect mechanism may be developed in the future.

This Technical Recommendation specifically describes a mechanism for the interoperable exchange of uncompressed audio, video, and ancillary data streams over an IP networked infrastructure. This includes a mechanism to describe synchronized groups of streams.

The SVIP Activity Group determined that its goals would be best met by adopting solutions that fit into the architecture of the Internet Engineering Task Force (IETF) Real Time Protocol (RTP). This architecture allows for a high level of flexibility; elementary media streams may be individually forwarded through the network based on their packet headers, and is extensible for a mix of uncompressed payloads as well as different types of compressed payloads not defined in this Technical Recommendation.

Future Technical Recommendations from the SVIP Activity Group may describe mechanisms for stream and device identity and registration.

Activity Group chairperson:
Michael Bany
DVBLInk
mbany@dvblink.net

1.1 Contributors

Merrick Ackermans, Turner Broadcasting
Ayman Alkhatib, Evertz
Nestor Amaya, Coveloz
Jeff Beachy, SDVI Corp.
David Berry, Sony
Nelson Botsford, Imagine Communications
Tommy Bouchard, Radio-Canada
Peter Brightwell, BBC
Richard Bullock, Ericsson
Richard Cartwright, Snell Advanced Media
Peter Chave, Cisco
Pierre Costa, AT&T
John Dale, Media Links
Daniel Miner, DirecTV
Subha Dhesikan, Cisco
Jack Douglas, PacketStorm
Bob Edge, Bob Edge TV Consulting, Inc.
Thomas Edwards, FOX
Eric Fankhauser, Evertz
Paul Gardiner, Sony
Brad Gilmer, Gilmer & Associates, Inc.
Ira Goldstone, FOX
Kevin Gross, AVA Networks
Richard Hastie, Mellanox
Inge Hillestad, Nevion
Rakesh Jalali, Evertz
Ali Jerbi, Cisco
Larry Kaplan, SDVI Corp.
Brian Keane, Aperi Corp.
Chin Koh, Nevion
Toshiaki Kojima, Sony
Al Kovalick, Media Systems Consulting
Andrew Krupiczka, ESPN

Yoji Kudo, Media Links
Ammar Latif, Cisco
Philippe Lemonnier, IRT B-Com
Marc Levy, Macnica Americas
Luann Linnebur, Ericsson
John Mailhot, Imagine Communications
Chuck Meyer, Grass Valley
Kazu Nakamura, Media Links
Kazuki Narita, Media Links
Andrew Osmond, Aperi Corp.
Carl Ostrom, System Resource
Gerard Phillips, Snell Advanced Media
Valerie Popie, EVS
Steve Posick, ESPN
Robert Pot, Grass Valley
Félix Poulin, EBU
Andy Rayner, Nevion
Robert Ruhl, Operations Manager, VSF
Patrick Scully, Ciena
Feng Shi, Emcore
Wes Simpson, Telecom Product Consulting
John Snow, Xilinx
Emory Strilkauskas, ESPN
Rainer Sturm, Arkona
Tom Sun, Xilinx
Christopher Taylor, Coveloz
Patrick Waddell, Harmonic Inc.
Robert Wadge, BBC
Leigh Whitcomb, Imagine Communications
John Williamson, Cisco
Gene Zimmerman, Cobalt Digital, Inc

1.2 About the Video Services Forum

The Video Services Forum, Inc. (www.videoservicesforum.org) is an international association dedicated to video transport technologies, interoperability, quality metrics and education. The VSF is composed of service providers, users and manufacturers. The organization's activities include:

- providing forums to identify issues involving the development, engineering, installation, testing and maintenance of audio and video services;
- exchanging non-proprietary information to promote the development of video transport service technology and to foster resolution of issues common to the video services industry;
- identification of video services applications and educational services utilizing video transport services;
- promoting interoperability and encouraging technical standards for national and international standards bodies.

The VSF is an association incorporated under the Not For Profit Corporation Law of the State of New York. Membership is open to businesses, public sector organizations and individuals worldwide. For more information on the Video Services Forum, contact Bob Ruhl, Operations Manager, Video Services Forum, +1 609 410 6767, bob.ruhl1@verizon.net.

2. Conformance Notation

Normative text is text that describes elements of the design that are indispensable or contains the conformance language keywords: "shall", "should", or "may". Informative text is text that is potentially helpful to the user, but not indispensable, and can be removed, changed, or added editorially without affecting interoperability. Informative text does not contain any conformance keywords.

All text in this document is, by default, normative, except: the Introduction, any section explicitly labeled as "Informative" or individual paragraphs that start with "Note:"

The keywords "shall" and "shall not" indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted.

The keywords, "should" and "should not" indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.

The keywords "may" and "need not" indicate courses of action permissible within the limits of the document.

The keyword “reserved” indicates a provision that is not defined at this time, shall not be used, and may be defined in the future. The keyword “forbidden” indicates “reserved” and in addition indicates that the provision will never be defined in the future.

A conformant implementation according to this document is one that includes all mandatory provisions ("shall") and, if implemented, all recommended provisions ("should") as described. A conformant implementation need not implement optional provisions ("may") and need not implement them as described.

Unless otherwise specified, the order of precedence of the types of normative information in this document shall be as follows: Normative prose shall be the authoritative definition; Tables shall be next; followed by formal languages; then figures; and then any other language forms.

3. Normative References

[1] AES67-2013 “AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability”

[2] IEEE 1588-2008 “IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems”

[3] IETF RFC 768 “User Datagram Protocol”

[4] IETF RFC 791 “Internet Protocol”

[5] IETF RFC 1112 “Host Extensions for IP Multicasting”

[6] IETF RFC 2974 “Session Announcement Protocol”

[7] IETF RFC 3190 “RTP Payload Format for 12-bit DAT Audio and 20- and 24-bit Linear Sampled Audio”

[8] IETF RFC 3550 “RTP: A Transport Protocol for Real-Time Applications”

[9] IETF RFC 3551 "RTP Profile for Audio and Video Conferences with Minimal Control"

[10] IETF RFC 4175 "RTP Payload Format for Uncompressed Video"

[11] IETF RFC 4566 "SDP: Session Description Protocol"

[12] IETF RFC 5888 "The Session Description Protocol (SDP) Grouping Framework"

[13] IETF RFC 7104 "Duplication Grouping Semantics in the Session Description Protocol"

[14] IETF RFC 7273 "RTP Clock Source Signalling"

[15] IETF draft-ietf-payload-rtp-ancillary "RTP Payload for SMPTE ST 291 Ancillary Data"

[16] ITU-R Recommendation BT.601 "Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios"

[17] ITU-R Recommendation BT.709 "Parameter values for the HDTV standards for production and international programme exchange"

[18] ITU-R Recommendation BT.1886 "Reference electro-optical transfer function for flat panel displays used in HDTV studio production"

[19] ITU-R Recommendation BT.2020 "Parameter values for ultra-high definition television systems for production and international programme exchange"

[20] SMPTE ST 291-1:2011 "Ancillary Data Packet and Space Formatting"

[21] SMPTE ST 2059-1:2015 "Generation and Alignment of Interface Signals to the SMPTE Epoch"

[22] SMPTE ST 2059-2:2015 "SMPTE Profile for Use of IEEE-1588 Precision Time Protocol in Professional Broadcast Applications"

[23] SMPTE ST 2084:2014 "High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays"

4. Informative References

[24] AES3-2009 (r2014) "AES standard for digital audio engineering - Serial transmission format for two-channel linearly represented digital audio data"

[25] AES10-2008 (r2014) “AES Recommended Practice for Digital Audio Engineering - Serial Multichannel Audio Digital Interface (MADI)”

[26] IETF RFC 7272 “Inter-Destination Media Synchronization (IDMS) Using the RTP Control Protocol (RTCP)”

[27] SMPTE ST 259:2008 “For Television — SDTV Digital Signal/Data — Serial Digital Interface”

[28] SMPTE ST 292-1:2012 “1.5 Gb/s Signal/Data Serial Interface”

[29] SMPTE ST 428-12:2013 “D-Cinema Distribution Master Common Audio Channels and Soundfield Groups”

[30] SMPTE ST 2022-6:2012 “Transport of High Bit Rate Media Signals over IP Networks (HBRMT)”

[31] SMPTE ST 2022-7:2013 “Seamless Protection Switching of SMPTE ST 2022 IP Datagrams”

[32] SMPTE ST 2031:2015 “Carriage of DVB/SCTE VBI Data in VANC”

[33] SMPTE ST 2067-8:2013 “Interoperable Master Format — Common Audio Labels”

5. Abbreviations

ANC: Ancillary Data, see definition.

SDI: Serial Digital Interface, generically referring to both the standard definition (SMPTE ST 259 [27]) and the high-definition (SMPTE ST 292-1 [28]) interfaces.

RTP: Real Time Protocol, as per IETF RFC 3550 [8].

6. Definitions

Ancillary Data — Society of Motion Picture and Television Engineers (SMPTE) Ancillary data (ANC), as defined by SMPTE ST 291-1 [20].

Note: ANC can carry an extensible range of data types, including time code, KLV metadata, Closed Captioning, Subtitles, Teletext essence, and the Active Format Description (AFD). ANC data packet payload definitions for a specific application are specified by a SMPTE Standard, Recommended Practice, Registered Disclosure Document, or by a document generated by another organization, a company, or an individual (an Entity). SMPTE

	Registered ANC packet types can be found on the SMPTE Registry for Data Identification Word Assignments at: http://www.smpte-ra.org/S291/S291_reg.html
Device	A media processor that has receivers and/or senders.
Elementary stream	A stream that only contains only one kind of data, e.g. audio, video or ancillary data.
Groups	Two or more RTP sessions that are intended for synchronized presentation.
Receiver	A consumer of a single RTP elementary stream.
RTP session	An association among a set of participants communicating with RTP. Note: A participant may be involved in multiple RTP sessions at the same time. In a multimedia session, each medium is typically carried in a separate RTP session. A participant distinguishes multiple RTP sessions by reception of different sessions using different pairs of destination transport addresses, where a pair of transport addresses comprises one network address plus a pair of ports, one for RTP and (optionally) one for RTCP.
Sender	A producer of a single RTP elementary stream.

7. System Overview (Informative)

This TR proposes that time-related essence (video, audio and ancillary data) is carried over an IP network as separate elementary RTP streams. Traditionally, professional video systems have used the SDI (Serial Digital Interface) to carry uncompressed video in the active video region, as well as embedded PCM audio and other ancillary data in the ancillary regions of the raster. SDI can be carried in its entirety over IP/RTP as a single stream through the use of SMPTE ST 2022-6 [30]. However, the ability to separate the audio, video, and ancillary data as separately network forwarded streams provides for greater flexibility in the production of media.

It should be noted that the elementary media streams carried using the techniques described in this Technical Recommendation may be converted from/to SDI through the use of IP/SDI gateways, or these elementary media streams may be created and processed independently of an SDI infrastructure in an all-IP fashion.

A system is built from devices that have senders and/or receivers. Streams of RTP packets flow from senders to receivers. If a sender emits packets with a multicast destination address, multiple receivers can receive the stream over the network. Also elementary RTP streams can be sent in a unicast stream from one sender to one receiver.

Devices may be adapters that convert from/to existing standard interfaces like SDI, AES3 [24] or AES10 (MADI) [25], or they may be processors that receive one or more streams, transform them in some way and transmit the resulting stream(s). Cameras and monitors may transmit and receive elementary RTP streams directly through an IP-connected interface, eliminating the need for SDI.

Inter-stream synchronization relies on timestamps in the RTP packets that are sourced by the sender from a common reference clock. The reference clock is distributed to all participating senders and receivers via PTP (IEEE 1588 [2]). The synchronization mechanism described is identical to that used for audio in AES67 [1], widening its application to other essence types.

Synchronization at the receiving device is achieved by the comparison of RTP timestamps with the common reference clock. The timing relationship between different streams is determined by their relationship to the reference clock.

This Technical Recommendation describes the method for transporting video and audio as uncompressed essence, but the same principles may be applied to compressed video and audio (not defined in this Technical Recommendation) by substituting the appropriate standard RTP payload.

An end-user or service provider of broadcast transmission services can utilize devices that implement this Technical Recommendation (TR) for unidirectional transport of various media signals over IP.

8. Synchronization

Synchronization of device reference clocks shall be achieved using IEEE 1588-2008 Precision Time Protocol (PTP) [2].

The IEEE 1588-2008 "Default PTP profile for use with the delay request-response mechanism" Version 1.0 (Profile identifier: 00-1B-19-00-01-00) shall be supported, however the SMPTE ST 2059-2 PTP Profile [22] should be supported.

9. Media Clocks

The media clock is used by senders to sample media and by receivers when recovering digital media streams. The media clock shall advance at an exact rate as specified below with respect to the reference clock. The media clock and the reference clock shall share the IEEE 1588 [2] and SMPTE ST 2059-1 [21] epochs. Digital media to be carried on the network shall be sampled according to the media clock.

For audio, the rate of the media clock shall be the same as the audio sampling frequency, and the supported media clock rate shall be 48 kHz. A 96 kHz media clock rate may be supported.

For video and ancillary data, the rate of the media clock shall be 90 kHz.

Note: A 90 kHz clock rate is the same as used by MPEG Transport Streams, and this is in general use worldwide to deliver video with a wide range of frame rates. Video and ancillary data streams that operate at a 60/1.001 Hz frame rate will find that frame times do not have an integer relationship to a 90 kHz clock, and typical implementations clock 60/1.001 Hz frame times by alternating between increments of 1501 and 1502.

RTP clocks (used to determine RTP timestamps) shall operate with a constant offset with respect to the media clock. The offset shall be conveyed through session description (as per section 13.2) on a per-stream basis.

10. Link Offset (Informative)

Link offset is the difference in time between when media enters the RTP sender and when it leaves the RTP receiver. Link offset must be sufficient to accommodate network delay and delay variation as well as any buffering or processing in the sender and receiver.

Link offset is determined by the receiver and is expected to be a constant and knowable value. Some receivers may support a configurable link offset such that inter-stream synchronization (e.g. lip sync) can be achieved either through manual configuration or automated means such as those described in RFC 7272 [26]. Each sender-receiver pair may have a different link offset value.

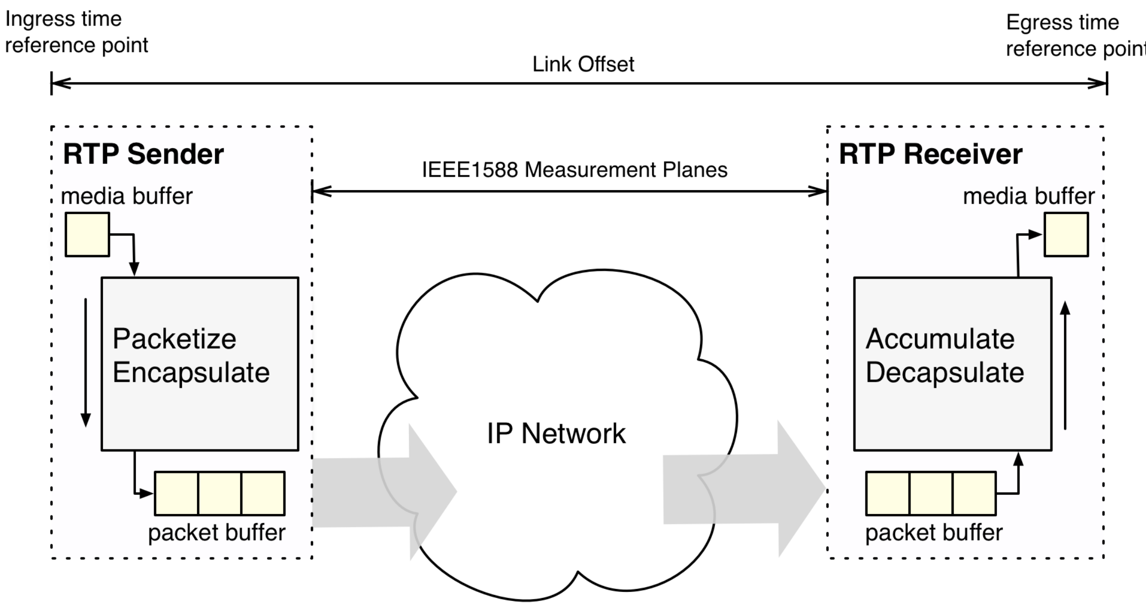

Figure 1: Example of Link Offset and Reference Points (Informative)

11. Media Transport

As these systems are expected to reside on private networks and IP address exhaustion is not an immediate concern in this context, media packets shall be transported using IP version 4 as defined in RFC 791 [4].

Note: provisions have been made in this work to allow future support for IPv6 in a straightforward manner.

Devices shall use Real-time Transport Protocol as defined in RFC 3550 [8]. Devices shall operate in accordance with RTP Profile for Audio and Video Conferences with Minimal Control as defined in RFC 3551 [9]. Devices shall use UDP as defined in RFC 768 [3] for transport of RTP.

Receivers are not required to reassemble fragmented IP packets. Therefore senders shall ensure that IP packets are emitted with a size that does not exceed the allowable maximum transmission unit (MTU) for the network application. Senders shall use a maximum 1440 octet UDP datagram size.

Senders and receivers shall support both multicast IP packets as specified in RFC 1112 [5] and unicast IP as specified in RFC 791 [4].

12. Video

Video streams shall be carried as per RFC 4175 [10]. Video scan line numbers should start at 1. There is no requirement for video scan line numbers to start at the first active video line number of an SDI raster. Only the active samples are included in the RTP payload: inactive samples and the contents of horizontal and vertical blanking shall not be transported.

Note: SMPTE ST 2031 [32] allows for the carriage of generic DVB/SCTE VBI Data in SMPTE 291-1 Ancillary Data if required.

Note: Devices that provide a gateway between RFC 4175 and SDI ought to be aware of the provision regarding starting video scan line numbers. Devices can determine when this provision is in effect by examining the line number of the first packet of a new video field or frame. The first packet of a new video field or frame can be identified by examining RTP headers.

For RTP timestamp purposes, when interfacing with video signals specified in SMPTE ST 2059-1, the time of the Alignment Point of the video frame defined by SMPTE ST 2059-1 shall be used as the sampling instant of the frame in progressive systems, or the sampling instant of the first field of a frame in interlaced systems.

Note: The sampling instant of the second field of a frame in interlaced systems is often taken as the point midway between the two adjoining frames.

YCbCr 4:2:2 10-bit Pgroup Example (Informative)

A key concept in RFC 4175 is the pixel group size, or "pgroup". The pgroup parameter is formally defined as “the size in octets of the smallest grouping of pixels such that 1) the grouping comprises an integer number of octets; and 2) if color sub-sampling is used, samples are only shared within the grouping.”

RFC 4175 states for YCbCr 4:2:2 format video: “Samples are packed in order Cb0-Y0-Cr0-Y1 for both interlaced and progressive scan lines. For 8-, 10-, 12-, or 16-bit samples, the pgroup is formed from two adjacent pixels (4, 5, 6, or 8 octets, respectively).”

Specifically for YCbCr 4:2:2 10-bit, the pgroup is 5 octets long, and its contents are shown in Figure 2.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Cb0        |         Y0        |        Cr0        |Y1 |  <- Y1(9:8), 2 x MSB
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Y1      | <- Y1(7:0), 8 x LSB
+-+-+-+-+-+-+-+-+
```

Figure 2: Example of YCbCr 4:2:2 10-bit Pgroup (Informative)

12.1 Audio

Audio streams shall be carried as per AES67 [1], with the following constraints:

1. Audio receivers are not required to support SIP for connection management.
2. Audio senders shall use the 24-bit linear encoding format ("L24") defined in RFC 3190 [7] clause 4.
3. Audio senders shall support 48 kHz sampling for audio.
4. Audio senders may use 96 kHz sampling for audio.

SDP descriptions of audio may include the optional parameter "channel-order" (as per RFC 3190 [7]). The "SVIP" convention of channel-order shall be defined as per Appendix A – SVIP Audio Channel Order.

12.2 Ancillary Data

Ancillary Data shall be carried as per draft-ietf-payload-rtp-ancillary [15] (or as per an IETF RFC that results from that draft).

13. Session Description

13.1 General Session Description Requirements

Individual streams or groups of streams meant for synchronized simultaneous playback shall be described using the Session Description Protocol of RFC 4566 [11].

A Session Description that includes multiple media lines ("m" lines) intended for synchronized play out shall (as per RFC 5888 [12]) have those "m" lines identified by a media stream identification attribute ("mid") and shall have a group session level attribute ("group") that specifies the Lip Synchronization ("LS") semantic.

Audio streams shall abide by the Session Description requirements of AES67 [1], Clause 8.1.

13.2 Stream Media Clock Descriptions

All descriptions shall have a media-level "mediaclk" attribute as per RFC 7273 [14]. The "direct" reference should be used. All descriptions shall also have a media-level "ts-refclk" attribute as per RFC 7273 [14].

13.3 Colorimetry and EOTF

Video streams that utilize Rec. ITU-R BT.2020 [19] shall specify the "a=fmtp" parameter "colorimetry" with the value "BT.2020".

Video streams that utilize the Electro-Optical Transfer Function (EOTF) of Rec. ITU-R BT.1886 [18] ("gamma") may specify the "a=fmtp" parameter "EOTF" with the value "BT.1886". If there is no "a=fmtp" parameter "EOTF" present, then receivers shall assume that the EOTF is that of Rec. ITU-R BT.1886 [18].

Video streams that utilize an Electro-Optical Transfer Function (EOTF) of SMPTE ST 2084 [23] shall specify the "a=fmtp" parameter "EOTF" with the value "SMPTE2084".

13.4 SDP for Duplicated RTP Streams

Duplicated RTP streams may be used for redundant transmission to achieve hitless failover. Duplicated RTP streams that use the mechanisms of Separate Source Addresses (RFC 7104 [13], Section 4.1) or Separate Destination Addresses (RFC 7104 [13], Section 4.2) shall signal the RTP duplication using the session level group ("group") attribute of RFC 5888 [12] and the duplication grouping ("DUP") semantics of RFC 7104 [13].

Note: SMPTE ST 2022-7 [31] also allows for RTP streams with duplicated source and destination addresses on separate physical networks; such a mechanism can not be represented with SDP.

13.5 SDP Example (Informative)

```
v=0
o=- 123456 11 IN IP4 192.168.1.1
s=Professional Networked Media Test
i=A test of video, audio, and ANC
t=0 0
a=group:LS V1 A1 M1
a=recvonly
m=video 50000 RTP/AVP 96
c=IN IP4 239.0.0.1/32
a=rtpmap:96 raw/90000
a=fmtp:96 sampling=YCbCr-4:2:2; width=1280; height=720; depth=10;
        colorimetry=BT.2020;EOTF=SMPTE2084
a=ts-refclk:ptp=IEEE1588-2008:39-A7-94-FF-FE-07-CB-D0:0
a=mediaclk:direct=2216659908
a=mid:V1
m=audio 50010 RTP/AVP 97
c=IN IP4 239.0.0.2/32
a=rtpmap:97 L24/48000/6
a=ptime:0.250
a=ts-refclk:ptp=IEEE1588-2008:39-A7-94-FF-FE-07-CB-D0:0
a=mediaclk:direct=963214424
a=fmtp:97 channel-order=SVIP.(L,C,R,Lrs,Rrs,LFE)
a=mid:A1
m=video 50020 RTP/AVP 98
c=IN IP4 239.0.0.3/32
a=rtpmap:98 smpte291/90000
a=ts-refclk:ptp=IEEE1588-2008:39-A7-94-FF-FE-07-CB-D0:0
a=mediaclk:direct=2216659908
a=mid:M1
```

This SDP file defines a session whose session ID is "123456" and version "11", and the session was created from IP 192.168.1.1. The session name is "Professional Networked Media Test", and session information is "A test of video, audio, and ANC." The timing line "t=0 0" indicates a session of unbounded length. A lip sync ("LS") grouping exists between media stream IDs ("mid") V1, A1, and M1.

The video media description indicates a Media Type "video", UDP port 50000, and dynamic RTP payload type 96. The video connection data indicates multicast IP on 239.0.0.1 with TTL 32. The attribute "rtpmap" indicates that the payload type 96 will have Media Subtype "raw" with a 90 kHz RTP clock. The attribute "fmtp" specifies the format specific parameters 4:2:2, 1280x720, 10-bit, ITU-R Rec. BT.2020 color space, and the SMPTE ST 2084 EOTF. The "ts-refclk" attribute points to an IEEE 1588 clock with grandmaster clock ID "39-A7-94-FF-FE-07-CB-D0" and PTP domain "0". The “mediaclk:direct” attribute sets the RTP timestamp as “2216659908” at the PTP epoch. The video is labeled with media stream ID “V1”.

The audio media description indicates Media Type “audio”, UDP port 50010, dynamic payload type 97, a multicast IP address of 239.0.0.2, Media Subtype “L24” (24-bit PCM) with 48 kHz sampling and 6 channels. The attribute “ptime” indicates a packet time (the length of time in milliseconds represented by the media in a packet) of 250 µs. Like the video media description, the audio media is linked to a PTP clock and the RTP timestamp at the PTP epoch is specified. The channel order is a typical 5.1 surround mix. The audio is labeled with media stream ID “A1”.

The final media description is of SMPTE ST 291-1 ancillary data carried as dynamic payload type 98 on IP 239.0.0.3 port 50020, and also linked to a PTP clock. The ancillary data is labeled with media stream ID “M1”.

14. Session Announcement

RTP sessions may be announced using the Session Announcement Protocol (SAP) of RFC 2974 [6]. SAP messages shall use the “application/sdp” payload type field, and shall contain an SDP description. SAP announcements should be sent to multicast IP address 224.2.127.254 port 9875 as a default unless equipment has been configured in another fashion.

Note: The use of SAP announcements using this default multicast address allows for rapid awareness of available streams by receivers without pre-configuration. It should also be noted that SAP is an experimental RFC, and it may have scaling problems in large implementations.

Appendix A – SVIP Audio Channel Order

Note: The Audio Channel Symbols in this section have been developed from SMPTE ST 428-12 [29] and SMPTE ST 2067-8 [33].

As per RFC 3190 [7], the optional SDP parameter channel-order may be used. The syntax is:

```
a=fmtp:<payload type> channel-order=<convention>.<order>
```

The convention "SVIP" shall be defined as an ordered list of Audio Channel Symbols The Audio Channel Symbols shall be contained within parenthesis and are separated by a comma, for example:

```
a=fmtp:101 channel-order=SVIP.(L, C, R, Lss, Rss, Lrs, Rrs, LFE)
```

The SVIP convention Audio Channel Symbols shall be:

L Left - Intended to drive the Left loudspeaker

R Right - Intended to drive the Right loudspeaker

C Center - Intended to drive the Center loudspeaker

LFE LFE - Intended to drive the Screen Low Frequency Effects loudspeaker

Ls Left Surround - Intended to drive the Left Surround

Rs Right Surround - Intended to drive the Right Surround

Lss Left Side Surround - Intended to drive the Left Side Surround

Rss Right Side Surround - Intended to drive the Right Side Surround

Lrs Left Rear Surround - Intended to drive the Left Rear Surround loudspeaker(s)

Rrs Right Rear Surround - Intended to drive the Right Rear Surround loudspeaker(s)

Lc Left Center - Intended to drive the Left Center loudspeaker

Rc Right Center - Intended to drive the Right Center loudspeaker

Cs Center Surround - Intended to drive the Center Surround loudspeaker

HI Hearing Impaired - A dedicated Audio Channel optimizing dialog intelligibility for the hearing impaired. This may carry a special dialog centric mix, i.e. a mix in which the dialog is predominate and dynamic range compression may be employed

VIN Visually Impaired – Narrative. A dedicated narration channel describing the main picture events for the visually impaired

M1 Mono One - A single channel of monaural audio, which can function on its own or be used in a Dual Mono Soundfield Group

M2 Mono Two - A second single channel of monaural audio, identical to the first, which can function on its own or be used in a Dual Mono Soundfield Group

Lt Left Total - Matrix encoded left channel of an Lt-Rt pair. This indicates that multiple channel of audio have been matrix encoded into two channels. If decoded, this will typically drive the Left, Center and Surround loudspeakers. If not decoded, it will typically drive only the left loudspeaker

Rt Right Total - Matrix encoded right channel of an Lt-Rt pair. This indicates that multiple channel of audio have been matrix encoded into two channels. If decoded, this will typically drive the Right, Center and Surround loudspeakers. If not decoded, it will typically drive only the right loudspeaker

Lst Left Surround Total - Matrix encoded left surround channel of an Lst-Rst pair. This indicates that Left Surround, Center Surround and Right Surround have been matrix encoded into two channels. If decoded, this will typically drive the Left Surround and Center Surround loudspeakers. If not decoded, it will typically drive only the Left Surround loudspeaker(s)

Rst Right Surround Total - Matrix encoded right surround channel of an Lst-Rst pair. This indicates that Left Surround, Center Surround and Right Surround have been matrix encoded into two channels. If decoded, this will typically drive the Right Surround and Center Surround loudspeakers. If not decoded, it will typically drive only the Right Surround loudspeaker(s)

S Surround - A single channel that Intended to drive one or more surround loudspeakers. A.k.a. "Mono Surround"

NSC[CHNUM] - Numbered Source Channel [CHNUM] . A single channel of audio that is ordered between 001 and 127 and is intended to play out in a Discrete Numbered Sources Soundfield Group; for example NSC025.

Note: As per RFC 4566 [11], values in SDP are case sensitive.

Appendix B – Interoperability Profiles (Informative)

This TR defines a set of interoperability profiles, which are defined for the purpose of interoperability testing, and only represent a subset of the capability of devices that operate under this Technical Recommendation.

Table 1 – Interoperability Profiles For Video

Profile	Video
1 "SD"	576i/25 480i/29.97
2 "HD"	720p/50 720p/59.94 1080i/25 1080i/29.97
3 "1080p24/25"	1080p/23.98 1080p/24 1080p/25
4 "1080p50/60"	1080p/50 1080p/59.94
5 "UHDTV1"	2160p/50 2160p/60 2160p/59.94

The video of these profiles is 4:2:2 color subsampled and 10 bits of depth.
Signaled colorimetry includes: ITU-R Rec. 601, ITU-R Rec. 709, and ITU-R Rec. 2020.
RTP timestamps for these video profiles use a 90 kHz clock.

Interoperability Profiles for Audio:

Mono:
    Carriage of 1 channel (M1) in a single elementary stream

Stereo:
    Carriage of 2 channels (L,R) in a single elementary stream

5.1:
    Carriage of 6 channels (L,R,C,Ls,Rs,LFE) in a single elementary stream

Audio uses 48 kHz sampling and 24 bits per sample.

Interoperability Profiles for Ancillary Data:

Captions:
    SMPTE 334-1 Carriage of CEA 608 and CEA 708 Captions in VANC
    SMPTE RDD-8 Carriage of WST (teletext) subtitles in VANC

SMPTE ST 2031 Carriage of DVB/SCTE VBI Data in VANC

AFD:

SMPTE ST 2016 Active Format Description & Bar Data in VANC

SCTE 104:

SMPTE ST 2010 Vertical Ancillary Data Mapping of ANSI/SCTE 104 Messages

ATC:

SMPTE 12-2 Time Code in the Ancillary Data Space

Appendix C – Asynchronous Media Clocks (Informative)

Systems operating under this Technical Recommendation may need to accept inputs from other media systems that do not share synchronized clocks. RFC 7273 [14] provides two mechanisms to signal asynchronously clocked media in SDP. The first is an "Asynchronously Generated Media Clock" where a sending media system has a local clock that has no clear synchronization to the network clock of the receiving media system. An asynchronous media clock may be explicitly signaled with the attribute:

```
a=mediaclk:sender
```

Also if there is no signaling of the media clock source at all, the asynchronously generated media clock is assumed.

The second method is a "Stream-Referenced Media Clock". A group of related RTP streams may be synchronized to a specific master RTP stream clock. In this case the RTP clock master stream's SSRC is indicated in an SDP attribute, and linked via a media clock tag to other streams. The master stream SDP attributes would be of the form:

```
a=ssrc:<ssrc> mediaclk:id=src:<media-clktag> sender
a=mediaclk:id=src:<media-clktag> sender
```

RTP streams slaved to the master stream use an attribute of the form:

```
a=mediaclk:id=<media-clktag> sender
```

Also if a receiver sees the signaling of "a=ts-refclk" which does not refer to a reference clock that is available to the receiver, it may need to handle that stream as an asynchronous stream.

This Technical Recommendation provides no guidance as to how asynchronously clocked streams are to be processed by the receiving media system.