Preamble to Video Services Forum (VSF) Technical Recommendation TR-06-4 Part 4

January 29, 2024

The Reliable Internet Stream Transport (RIST) project was initiated as an Activity Group under the auspices of the Video Services Forum in 2017. The RIST Protocol is defined by TR-06-1 (RIST Simple Profile, published in 2018 and updated in 2020), TR-06-2 (RIST Main Profile, published in 2020 and updated in 2021 and 2022), and TR-06-3 (RIST Advanced Profile, published in 2021 and updated in 2022).

The TR-06-4 series of recommendations define ancillary features for the RIST protocol that are applicable to multiple profiles. This series includes:

- TR-06-4 Part 1, Source Adaptation, published in 2022.
- TR-06-4 Part 2, Use of Wireguard VPN in RIST Devices, published in 2023.
- TR-06-4 Part 3, RIST Relay, published in 2023.
- TR-06-4 Part 5, RIST Multicast Discovery, published in 2023.

This document is TR-06-4 Part 4, RIST Decoder Synchronization. When multiple encoders are transmitting to multiple decoders, it may be useful to synchronize the decoder playback (in sports or worship applications, for example). This means that video frames that enter the encoders at the same time are required to be played by the decoders at the same time. Since such signals are often transmitted over IP (through the Internet or a private network) using RIST, TR-06-4 Part 4 defines extensions to RIST to provide such synchronization.

Work continues within the group towards developing additional RIST specifications that include additional features. As the Activity Group develops and reaches consensus on new functions and capabilities, these documents will also be released in support of the RIST effort. For additional information about the RIST Activity group, or to find out about participating in the development of future specifications, please visit http://vsf.tv/RIST.shtml.

Video Services Forum (VSF)
Technical Recommendation TR-06-4
Part 4

Reliable Internet Stream Transport (RIST)
RIST Decoder Synchronization

Approved January 29, 2024

INTELLECTUAL PROPERTY RIGHTS

THIS RECOMMENDATION IS BEING OFFERED WITHOUT ANY WARRANTY WHATSOEVER, AND IN PARTICULAR, ANY WARRANTY OF NONINFRINGEMENT IS EXPRESSLY DISCLAIMED. ANY USE OF THIS RECOMMENDATION SHALL BE MADE ENTIRELY AT THE IMPLEMENTER'S OWN RISK, AND NEITHER THE FORUM, NOR ANY OF ITS MEMBERS OR SUBMITTERS, SHALL HAVE ANY LIABILITY WHATSOEVER TO ANY MPLEMENTER OR THIRD PARTY FOR ANY DAMAGES OF ANY NATURE WHATSOEVER, DIRECTLY OR INDIRECTLY, ARISING FROM THE USE OF THIS RECOMMENDATION.

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

When multiple encoders are transmitting multiple signals to multiple decoders, it may be useful to synchronize the decoder playback (in sports or worship applications, for example). This means that video frames that enter the encoders at the same time will be played by the decoders at the same time. Since such signals are often transmitted over IP (through the Internet or a private network) using RIST, this Technical Recommendation defines extensions to RIST to provide such synchronization.

Recipients of this document are invited to submit technical comments. The VSF also requests that recipients notify us of any relevant patent claims or other intellectual property rights of which they may be aware, that might be infringed by any implementation of the Recommendation set forth in this document, and to provide supporting documentation.

Table of Contents

1 Introduction (Informative)

As broadcasters and others increasingly utilize unconditioned Internet circuits to transport high-quality video, the demand grows for systems that can compensate for the packet losses and delay variation that often affect these streams. A variety of solutions are currently available on the market; however, incompatibilities exist between devices from different suppliers.

The Reliable Internet Stream Transport (RIST) project was launched specifically to address the lack of compatibility between devices, and to define a set of interoperability points using existing or new standards and recommendations.

In addition to reliably transporting video over the Internet or a private IP network, some applications have a requirement of synchronized playback. Assume that there are N encoders/senders and M ≥ N decoders/receivers connected by an IP network. Each decoder is receiving and playing a stream from one of the encoders. The requirement is that frames of video that enter the N encoders “at the same time” are played by the M decoders “at the same time”.

“At the same time” is left vague on purpose. It is usually understood to be “within a frame” – typically a number small enough as not to be perceptible by an average human being.

This Technical Recommendation defines extensions to RIST Simple Profile (VSF TR-06-1) and RIST Advanced Profile (VSF TR-06-3) to provide this synchronization.

1.1 Contributors

The following individuals participated in the Video Services Forum RIST working group that developed this technical recommendation.

Merrick Ackermans (CBS/Paramount)	Sergio Ammirata (SipRadius/AMMUX)	Paul Atwell (Media Transport Solutions)
Ronald Fellman (QVidium)	Michael Firth (Nevion)	Oded Gants (Zixi)
Brian Keane (Net Insight)	Holger Klaas (Nevion)	Ciro Noronha (Cobalt Digital)
Adi Rozenberg (AlvaLinks)	Wes Simpson (LearnIPVideo)	Thomas True (Nvidia)

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 or this document, please call +1 929-279-1995 or e-mail opsmgr@videoservicesforum.org.

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 and 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 References

VSF TR-06-1:2020, Reliable Internet Stream Transport (RIST) Protocol Specification – Simple Profile

VSF TR-06-2:2023, Reliable Internet Stream Transport (RIST) Protocol Specification – Main Profile

VSF TR-06-3:2022, Reliable Internet Stream Transport (RIST) Protocol Specification – Advanced Profile

VSF TR-06-3:2023, Reliable Internet Stream Transport (RIST) – Advanced Profile Levels Annex

VSF TR-06-4 Part 3, Reliable Internet Stream Transport (RIST) – RIST Relay

IETF RFC 3550, RTP: a Transport Protocol for Real-Time Applications

IETF RFC 5905, Network Time Protocol Version 4: Protocol and Algorithms Specification

IETF RFC 7826, Real-Time Streaming Protocol Version 2.0

ISO/IEC 13818-1:2023, Generic coding of moving pictures and associated audio information, Part 1: Systems

Any mention of references throughout the remainder of this document refers to the versions described here, unless explicitly stated otherwise.

4 Receiver/Decoder Synchronization Protocol

4.1 General Overview (Informative)

Fundamentally, decoder synchronization is achieved by making the end-to-end delay (i.e., the interval between the time a frame of video enters the encoder and the time the same frame is displayed at the decoder) the same between all encoder/decoder pairs. If this can be achieved, then all decoders are synchronized, and it is not necessary to synchronize (genlock) the video coming into the encoders; moreover, if there is a common time base available to all encoders and decoders, neither the encoders nor the decoders even need to be co-located.

To provide such a solution, the following functions are needed:

1. A common time base between all the encoders and all the decoders.
2. A means for each decoder to determine how long has it been since each frame has been captured.

3. A common agreed end-to-end delay between all the decoders, which must be no less than the worst-case end-to-end delay between all encoders and decoders.

Using these functions, each decoder adds an additional "Sync Delay" so that the Total End-to-End Delay for all encoder/decoder chains is exactly the same. This is illustrated in Figure 1; each decoder adjusts the Sync Delay so that all Encoder/Decoder pairs have the exact same delay.

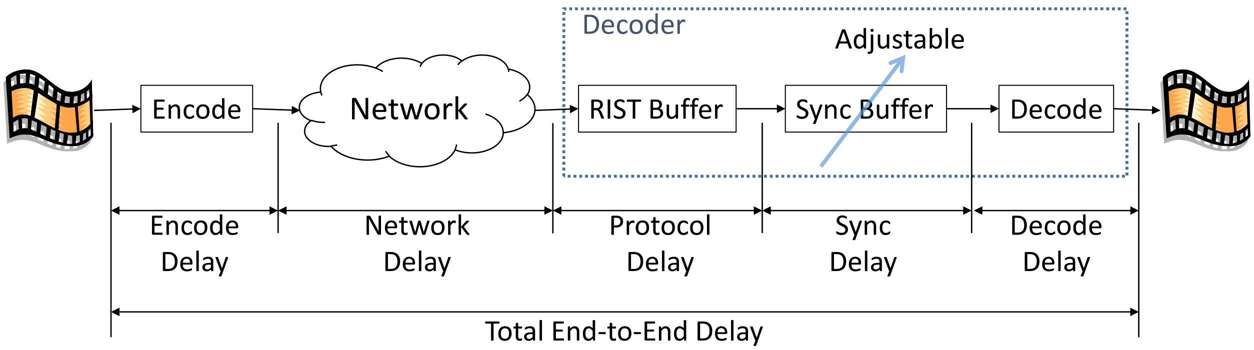

Figure 1: End-to-End Delay Components

For the purposes of this Specification, the required functions are chosen as follows:

1. The common time base between all the encoders and decoders is chosen to be the Network Time Protocol (NTP), defined in RFC 5905.
2. This Technical Recommendation defines the protocols between the encoders and decoders to allow the decoders to determine how long has it been since each frame has been captured.
3. Each decoder must be configured with a target end-to-end delay that is no less than the worst-case encoder/decoder delay. Determination and dissemination of such value is out of the scope of this Specification and is left at the discretion of the implementer. For example, implementers may expose this as a configuration item, and have decoders alarm if they cannot achieve the configured delay.

RIST uses RTP for media transmission, and all RTP packets include a 32-bit Timestamp field. RFC 7826 ties the RTP Timestamp with a corresponding NTP timestamp to provide synchronization between different media flows (e.g., the video and audio streams out of an RTSP camera). This Technical Recommendation uses the same mechanism to achieve synchronization across different devices.

4.2 Encoder/Sender Operation

For the purposes of this Technical Recommendation, a "program essence" is defined as a set of zero or more audio, video, and/or metadata component essences that are intended to be played synchronously. The synchronization between component essences of a program is achieved by attaching periodic timestamps from a common time base to each component essence, and by including periodic samples of the time base so that a receiver can recover this time base. For the purposes of this Technical Recommendation, "Media Timestamps" are the samples of this common time base. For Transport Streams compliant with ISO/IEC 13818-1, the "program essence" corresponds to "program (system)" as defined in that standard, the time base corresponds to the System Time Clock (STC), and the "Media Timestamps" correspond to the Program Clock Reference (PCR) values. Figure 2 shows a reference model for the Encoder/Sender operation. Encoder/Sender devices compliant with this Technical Recommendation shall support one program essence per RTP flow. Devices may support transmission of multiple program essences per RTP flow as described in Section 6. The Encoder/Sender shall operate as follows:

- Every time the Encoder/Sender captures a block of media for compression, it shall simultaneously capture the Media Timestamp associated with that block of media and the current NTP timestamp and store these two values. If we denote the last captured Media Timestamp by M_L and the corresponding NTP timestamp by N_L, the Encoder/Sender will always store the last (M_L, N_L) values. If the encoder is generating a Transport Stream as per ISO/IEC 13818-1, the captured Media Timestamp corresponds to the value of the Presentation Time-Stamp (PTS) for block of media just captured.
  Note: if the encoder is capturing multiple essence types (i.e., audio, video, and synchronized data), it is sufficient to use only one essence type for this purpose. It is recommended that video be used if present.
- It is assumed that the Media Timestamp is inserted periodically by the encoder in the program essence. Whenever the Encoder/Sender produces an RTP packet that contains a Media Timestamp, it shall calculate the corresponding NTP timestamp and store this calculated timestamp together with the RTP Timestamp of the packet. The NTP timestamp shall be calculated as follows:
  M_L: Last Media Timestamp
  N_L: NTP timestamp corresponding to the last Media Timestamp
  M_P: Media timestamp of the current packet (see Figure 2)
  f_M: Frequency of the Media Timestamp
  T_P: RTP timestamp of the current packet (see Figure 2)
  N_P: Calculated NTP timestamp of the Media Timestamp of the current packet:

$$N_P = N_L + \frac{M_P - M_L}{f_M}$$

  The Encoder/Sender shall store the (T_P, N_P) values every time it sends a packet that contains a Media Timestamp. If the Sender is transmitting multiple video essences, the stored (T_P, N_P) value shall correspond to the first Media Timestamp present in the packet. Note: If the encoder is generating a Transport Stream, the Media Timestamp is the PCR.

- The Encoder/Sender shall transmit the (T_P, N_P) values periodically using the mechanisms described later in this Technical Recommendation.
  Note: It is necessary to send this information periodically because there is no requirement that the media frequency be locked to NTP.

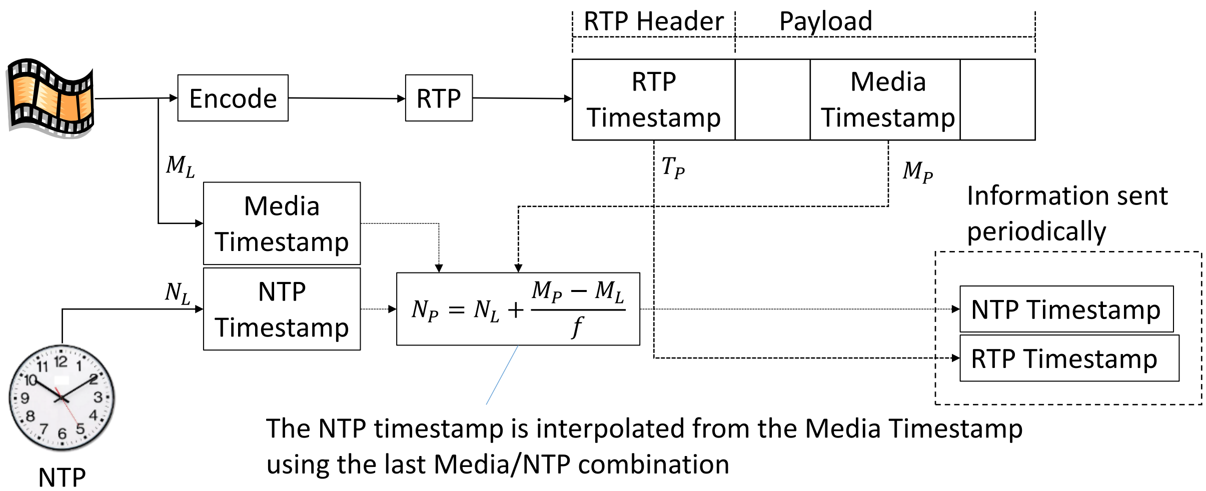

Figure 2: Encoder/Sender Operation

The format of the periodic message from the encoders/senders to the decoders/receivers is shown in Figure 3. The message is divided into two parts:

- **Required Fields:** these fields shall always be included in the periodic message.
- **Additional Fields:** these fields are included in the periodic message depending on the characteristics of the sender, as described later in this document.

The sender shall set the Required Fields in Figure 3 as follows:

- **NTP timestamp (64 bits):** this field shall be set by the Encoder/Sender to the calculated full 64-bit NTP timestamp N_P calculated as described above.
- **RTP timestamp (32 bits):** this field shall be set to the Encoder/Sender to the RTP timestamp T_P of the RTP packet containing the Media Timestamp M_P used in the computation of N_P, as described above.

The sender shall include the Additional Fields from Figure 3 if it is a Gateway that is separate from the encoders and this Gateway is internally generating the periodic synchronization message independently of the original encoders. The fields shall be set by the sender as follows:

- **G (1 bit):** The sender shall set this field to “1” if it is a gateway separate from the encoders, internally generating the Synchronization Message. Receivers shall assume that G is “0” if the additional fields are not included in the message.
- **Reserved (31 bits):** Senders compliant with this Technical Recommendation shall set these bits to “0”, and receivers compliant with this Technical Recommendation shall ignore them.

Required Fields

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             NTP timestamp, most significant word              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             NTP timestamp, least significant word             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         RTP timestamp                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Additional Fields

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|G|                          Reserved                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: Periodic Synchronization Message

4.2.1 RIST Simple Profile Operation

RIST Simple Profile (VSF TR-06-1) requires senders to periodically transmit either empty RTCP Receiver Reports (RR) or Sender Reports (SR), with the interval between two successive RTCP packets of 100 milliseconds or less. Encoder/Senders compliant with this Technical Recommendation shall always send SR RTCP packets and use the NTP and RTP timestamp fields in the SR packet to transmit the latest (T_P, N_P) value, as indicated in Figure 4. These are the Required Fields in Figure 3.

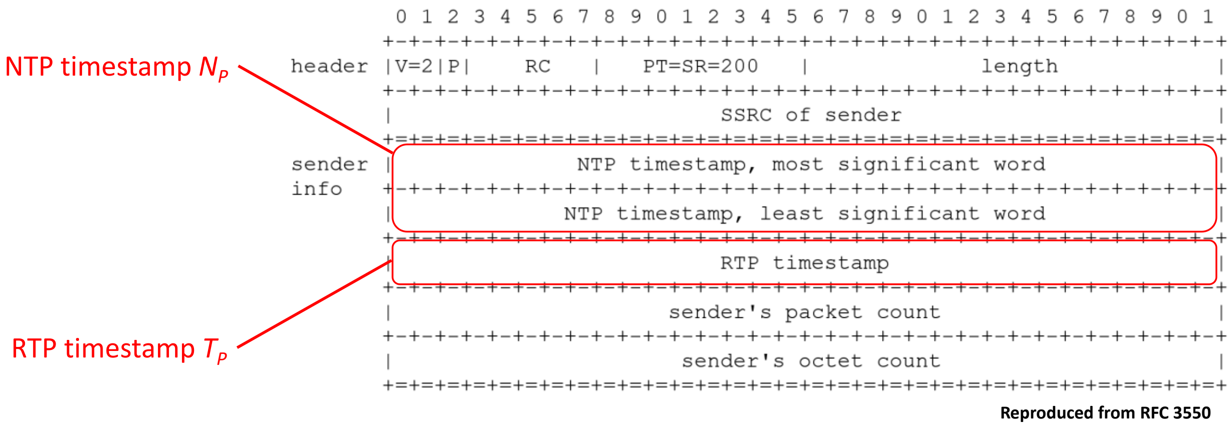

Figure 4: Decoder Synchronization fields in the SR RTCP Message

If the sender is required to send the Additional Fields from Figure 3, it shall do so as a profile-specific extension to the SR message, as shown in Figure 5.

```
         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
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
header  |V=2|P|    RC   |   PT=SR=200   |             length            |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
        |                         SSRC of sender                        |
        +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
sender  |              NTP timestamp, most significant word             |
info    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
        |             NTP timestamp, least significant word             |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
        |                         RTP timestamp                         |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
        |                     sender's packet count                     |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
        |                      sender's octet count                     |
        +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
report  :                              ...                              :
blocks  :                                                               :
        +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
        |    profile-specific extension: Additional Fields (4 bytes)    |
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: Including Additional Fields in the SR Message

Note: there are small interpretation changes between this Technical Recommendation and RFC 3550 regarding the NTP timestamp and the RTP timestamp in the SR packet. Table 1 summarizes these differences.

Table 1: Differences between this Technical Recommendation and RFC 3550

This Specification	RFC 3550
NTP timestamp is required	NTP timestamp is optional, can be set to zero
NTP timestamp must come from an actual NTP server	NTP timestamp can be device's wall clock
NTP timestamp corresponds to frame capture time, as described in this document	NTP timestamp corresponds to SR message transmission time
RTP timestamp corresponds to timestamp of the packet carrying the frame, as described in this document	RTP timestamp corresponds to the same point in time as the NTP timestamp

4.2.2 RIST Main Profile Operation

RIST Main Profile (VSF TR-06-2) operates as a tunnel for RIST Simple Profile flows. Therefore, the operation described in Section 4.2.1 shall apply to RIST Main Profile.

4.2.3 RIST Advanced Profile Operation

RIST Advanced Profile (VSF TR-06-3) has several tunnel levels (see the Advanced Profiles Levels Annex): IPv4-Tunnel, IPv6-Tunnel, Layer2-Tunnel, and Main-Profile-Tunnel. These levels can carry one or more RIST Simple Profile flows, and, for these cases, the operation described in Section 4.2.1 also shall apply.

RIST Advanced Profile also includes Media levels, using the Direct Payload format. These levels are RTP flows without RTCP, and thus require an alternate mechanism to carry the synchronization information. Therefore, this Technical Recommendation defines a new RIST Advanced Profile Control Message for the Media levels, as indicated in Table 2.

Table 2: Advanced Profile Control Index Values

Control Index	Message Type	Mandatory
0x0000	NACK Bitmask	
0x0001	NACK Range	
0x0002-0x0003	TR-06-4 Part 1 Link Quality Reports	
0x0004	**Encoder/Sender Synchronization Message**	
0x0005-0x000F	Reserved for future NACK messages	
0x0010	RTT Echo Request	
0x0011	RTT Echo Response	Yes
0x0012-0x001F	Reserved for future RTT messages	
0x0020	ST 2022-5 FEC Row Packet	
0x0021	ST 2022-5 FEC Column Packet	
0x0022	ST 2022-1 FEC Row Packet	
0x0023	ST 2022-1 FEC Column Packet	
0x0024-0x002F	Reserved for future FEC messages	
0x0030-0x77FF	Reserved for future control messages	
0x7800-0x7FFF	Reserved for private vendor use	
0x8000	RIST Main Profile Keep-Alive message	Yes
0x8001	Flow Attribute message	
0x8002-0x800F	Reserved for future tunnel messages	
0x8010	Advanced Profile SRP authentication for PSK sessions	
0x8011	PSK Future Nonce Announcement Message	
0x8012-0x801F	Reserved for future authentication messages	
0x8020	Control Message Unsupported Response	
0x8021-0x802F	Reserved for future error messages	
0x8030-0x804F	TR-06-4 Part 3 RIST Relay Messages	
0x8050-0xF7FF	Reserved for future control messages	
0xF800-0xFFFF	Reserved for private vendor use	

The format of the Encoder/Sender Synchronization Message is shown in Figure 6. The Encoder/Sender shall set the (T_P, N_P) values as indicated in the figure and shall send this message periodically, with an interval no greater than 100 milliseconds.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Control Index = 0x0004    |          Length=12 or 16      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            NTP timestamp, most significant word               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             NTP timestamp, least significant word             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         RTP timestamp                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|G|                           Reserved                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 6: Encoder/Sender Synchronization Message

The encoder/sender may omit the Additional Fields section of the message in Figure 6 if G=0. The receiver shall derive the presence or absence of the Additional Fields section by inspecting the Length field. If Length=12, the message does not contain the Additional Fields section; if Length=16, it does.

If the RIST Advanced Profile Encoder/Sender is sending multiple Direct Payload flows multiplexed using the Flow ID field, the RTP header for the Encoder/Sender Synchronization Message shall include the Flow ID of the flow it refers to.

4.3 Decoder/Receiver Operation

The SR RTCP message (Section 4.2.1) or the Encoder/Sender Synchronization Message (Section 4.2.3) provide the Decoder/Receiver with the relationship between the Media Timestamp and the NTP timestamp. The Decoder/Receiver shall operate as follows:

- When the Decoder/Receiver receives the SR RTCP message or the Encoder/Sender Synchronization Message, the Decoder/Receiver shall store the received (T_P, N_P) values from the message. The Decoder/Receiver should validate the NTP timestamp N_P and should discard invalid values, including timestamps for RTP packets that do not contain a Media Timestamp for the relevant video essence. If the decoder supports multiple essences, it shall identify the essence for the received (T_P, N_P) by looking up the first Media Timestamp in the received RTP packet with RTP timestamp T_P[1]. It is legal to have the timestamp N_P be in the past, but not more than the configured Total End-to-End Delay shown in Figure 1.
- When the Decoder/Receiver receives an RTP media packet containing a Media timestamp M_C with an RTP timestamp T_C, it shall compute and store an (M_C, N_C) value defined as follows:

  T_P: RTP timestamp from the last received SR/Synchronization packet for this essence

[1] For Multi-Program Transport Streams, the decoder finds the first PCR-bearing transport packet in the RTP packet payload, and uses the Packet Identifier (PID) of that transport packet to identify the program it belongs to.

N_P: NTP timestamp from the last received SR/Synchronization packet for this essence
M_C: Received Media timestamp
T_C: RTP timestamp of the packet carrying the Media timestamp M_C
f_R: RTP timestamp frequency for this media type[2]
N_C: Interpolated NTP timestamp for the received Media timestamp, calculated as follows:

$$N_C = N_P + \frac{T_C - T_P}{f_R}$$

- When the Decoder/Receiver is ready to play out a frame of video with timestamp M_V, it shall use the latest (M_C, N_C) value to play this frame at NTP time N_V calculated as follows:
  M_C: Last received Media timestamp
  N_C: Interpolated NTP timestamp for the last received Media Timestamp
  M_V: Media timestamp of a frame of video ready to be played out
  f_M: Frequency of the Media Timestamp
  D: Configured Total End-to-End Delay (see Figure 1)
  N_V: NTP time at which the frame with timestamp M_V is to be played out

$$N_V = N_C + \frac{M_V - M_C}{f_M} + D$$

  If the calculated time N_V is in the past, this is an error condition, and the Decoder/Receiver should raise an alarm. Other actions in this situation are left to the discretion of the implementer.

Note: in a well-designed system, the Decoder/Receiver will recover the media clock from the incoming stream, and that media clock will be locked to the Encoder/Sender media clock. Once the Decoder/Receiver delays the first frame and plays it at the calculated time, subsequent frames are likely to be played at the correct times as well, and the synchronization messages can be used to verify ongoing synchronization. The mechanisms to handle error conditions are left to the discretion of the implementer.

5 Gateway Considerations

For the purposes of this section, a gateway is defined as a device or process that communicates with encoders and decoders over an actual or virtual network. Moreover, a gateway offers protocol translation between RIST and other protocols.

[2] This frequency is 90 kHz for Transport Streams and 1 MHz for RIST Advanced Profile.

Two cases are envisioned:

1. **Gateway Receiver:** the gateway receives one or more RIST streams from encoders compliant with this Technical Recommendation and transmits the content to one or more decoders using a different protocol.
2. **Gateway Transmitter:** the gateway receives one or more streams from encoders using some non-RIST protocol and transmits the content using RIST to one or more decoders compliant with this Technical Recommendation.

5.1 Gateway Receiver Case (Informative)

A gateway can align (synchronize) the output streams to the decoders using the algorithms described in Section 4.3. More specifically, packets on the egress of the gateway correspond to timestamps equivalent in time.

Synchronized playback as defined in this Technical Recommendation can only be achieved if the following conditions are all true:

- There is negligible or constant network delay between the gateway and the downstream decoders.
- All the downstream decoders have the exact same decoding delay.
- All the downstream decoders have consistent and repeatable decoding delay regardless of where in the bitstream they start receiving.

5.2 Gateway Transmitter Case

In this case, the gateway does not have any knowledge of the time elapsed since each encoder captured the incoming frames. The gateway can generate the synchronization data described in Section 4.2 using the arrival time of the bitstream as the synchronization point and shall include the Additional Fields section of the message with G=1.

Synchronized playback as defined in this Technical Recommendation can only be achieved if the following conditions are all true:

- There is negligible or constant network delay between the gateway and the encoders.
- All the encoders have the exact same encoding delay and generate a bitstream with equivalent timing parameters.

Decoder/Receivers should provide a warning to the user if they receive Synchronization Messages with G=1.

6 Multiple Essence Support

An encoder/sender may support transmission of more than one video essence multiplexed into a single RTP flow (e.g., a Multi-Program Transport Stream). If synchronization of such video essences is desired, the encoder/sender may use the technique described in this section to achieve such synchronization. The technique can be used to synchronize the various video essences in the RTP stream, as well as video essences from other RTP streams.

For encoder/senders supporting multiple essences, the (T_P, N_P) value stored shall correspond to the first Media Timestamp present in the packet, and the encoder/sender shall cache one (T_P, N_P) value per essence. The encoder/sender shall ensure that (T_P, N_P) values for all essences are no older than 100 milliseconds and may need to split RTP payloads containing more than one Media Timestamp into multiple RTP packets to meet this requirement. All (T_P, N_P) values shall be transmitted at least once every 100 milliseconds, and the encoder/sender may combine multiple SR packets into a compound RTCP packet to reduce overhead.

7 RIST Relay Considerations (Informative)

Encoder/Senders compliant with this Technical Recommendation can use the services of a RIST Relay compliant with TR-06-4 Part 3 to reach one or more Decoder/Receivers. If the RIST Relay is operating transparently, and not doing any sort of protocol conversion, the mechanisms in this Technical Recommendation will work, and no additional changes are needed in the Relay itself.

However, if the RIST Relay is operating in one of the proxy cases described in TR-06-4 Part 3 Section 6, or if it is doing protocol conversion, it will need to explicitly relay the timestamp information described in this document. This may involve conversion between the formats in Sections 4.2.1 and 4.2.3. The actual data can be simply copied, but the messages will need to be created. RIST Relay implementers are advised to consider this feature and add support for this conversion.