[MS-H264PF]: RTP Payload Format for H.264 Video Streams Extensions

Revision Summary

Date	Revision History	Revision Class	Comments
01/20/2012	0.1	New	Released new document.
04/11/2012	0.1	No change	No changes to the meaning, language, or formatting of the technical content.

Table of Contents

1 Introduction

The RTP Payload Format for H264 Streams Extensions protocol describes the payload format to carry real-time video streams in the payload of the Real-Time Transport Protocol (RTP). It is used to transmit and receive real-time video streams in two-party peer-to-peer calls and in multi-party conference calls.

Sections 1.8, 2, and 3 of this specification are normative and can contain the terms MAY, SHOULD, MUST, MUST NOT, and SHOULD NOT as defined in RFC 2119. Sections 1.5 and 1.9 are also normative but cannot contain those terms. All other sections and examples in this specification are informative.

1.1 Glossary

The following terms are defined in [MS-GLOS]:

network byte order

The following terms are defined in [MS-OFCGLOS]:

forward error correction (FEC)
Real-Time Transport Protocol (RTP)

The following terms are specific to this document:

MAY, SHOULD, MUST, SHOULD NOT, MUST NOT: These terms (in all caps) are used as described in [RFC2119]. All statements of optional behavior use either MAY, SHOULD, or SHOULD NOT.

1.2 References

References to Microsoft Open Specifications documentation do not include a publishing year because links are to the latest version of the documents, which are updated frequently. References to other documents include a publishing year when one is available.

1.2.1 Normative References

We conduct frequent surveys of the normative references to assure their continued availability. If you have any issue with finding a normative reference, please contact dochelp@microsoft.com. We will assist you in finding the relevant information. Please check the archive site, http://msdn2.microsoft.com/en-us/library/E4BD6494-06AD-4aed-9823-445E921C9624, as an additional source.

[ISO/IEC14496-10:2010] ISO/IEC, "Information technology -- Coding of audio-visual objects", Part 10: Advanced Video Coding, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=56538

[MS-RTP] Microsoft Corporation, "Real-time Transport Protocol (RTP) Extensions".

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997, http://www.rfc-editor.org/rfc/rfc2119.txt

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and Jacobson, V., "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003, http://www.ietf.org/rfc/rfc3550.txt

[RFC5109] A. Li, Ed., "RTP Payload Format for Generic Forward Error Correction", December 2007, http://www.ietf.org/rfc/rfc5109.txt

[RFC6184] Wang, Y. K., Even, R., Kristensen, T. et al., "RTP Payload Format for H.264 Video", May 2011, http://www.ietf.org/rfc/rfc6184.txt

[RFC6190] Wenger, S., Wang, Y. K., Schierl, T., et al., "RTP Payload Format for Scalable Video Coding", May 2011, http://www.ietf.org/rfc/rfc6190.txt

1.2.2 Informative References

[MS-GLOS] Microsoft Corporation, "Windows Protocols Master Glossary".

[MS-OFCGLOS] Microsoft Corporation, "Microsoft Office Master Glossary".

1.3 Overview

The RTP Payload Format for H264 Streams Extensions protocol specifies a payload format to transport an H.264 bitstream using Real-Time Transport Protocol (RTP).

The syntax of this protocol follows the definition in [RFC6190] with the following extensions:

1. Customized PACSI packet is used to signal the stream layout and video frame cropping information.
2. Simulcast streams are supported. A sender capable of simulcast can send the same video coded sequence in different video resolutions and different video codecs at the same time.

1.4 Relationship to Other Protocols

This protocol carries H.264 bitstream, described in [ISO/IEC14496-10:2010], as a payload, and in turn is carried as a payload in RTP, as described in [MS-RTP].

1.5 Prerequisites/Preconditions

This protocol specifies only the payload format for H.264 video streams. This protocol requires the establishment of an RTP stream, a mechanism to obtain H.264 video access units for it to packetize, and a mechanism to render H.264 video access units that it has depacketized.

Higher layers are required to provide H.264 access units.

1.6 Applicability Statement

This protocol is only applicable for transporting video access units encoded using the H.264 codec.

1.7 Versioning and Capability Negotiation

This protocol has the following versioning constraints:

- **Supported Transports:** This protocol uses **RTP** as its transport as discussed in section 2.1

1.8 Vendor-Extensible Fields

None.

1.9 Standards Assignments

None.

2 Messages

2.1 Transport

This protocol is a payload for the [MS-RTP] transport protocol and therefore relies on RTP for providing the means to transport its payload over the network.

2.2 Message Syntax

The Network Abstraction Layer (NAL) unit format, transmission mode, and packetization mode are the same as defined in [RFC6184] and [RFC6190] with a few extensions.

The Payload Content Scalability Information (PACSI) packet, specified in [RFC6190], MAY be extended by incorporating one or more customized Supplemental Enhancement Information (SEI) NAL units. This protocol defines two types of SEI messages:

1. Stream Layout SEI Message
2. Cropping Info SEI Message

All fields in the messages specified in this protocol are in **Network Byte Order** unless explicitly called out.

2.2.1 Namespaces

None.

2.2.2 RTP Header Usage

The syntax of the RTP header is specified in [MS-RTP] section 2.2.1. The fields of the fixed RTP header have their usual meaning with the following additional notes:

Marker (M): This bit MUST be set to 1 if the RTP packet contains the last packet of a layer of an access unit. The RTP packet MAY be a VCL NAL unit, as defined in [RFC6184] section 4.1, or an H.264 FEC packet associated with one or more VCL NAL units.

Timestamp: The syntax of this field is defined in [RFC3550], section 5.1. The sampling clock frequency MUST be 90000 Hz. All RTP packets of the same access unit of a simulcast stream MUST carry the same timestamp. The timestamps of two different simulcast streams are not required to be equal, even if the RTP packets contain VCL NAL units for the same coded picture.

2.2.3 Transmission Mode

The syntax of transmission mode follows the syntax defined in [RFC6190].

This protocol only supports Multiple-Session Transmission (MST).

2.2.4 Packetization Mode

The syntax of packetization mode used in this protocol follows the syntax defined in [RFC6184] and [RFC6190].

This protocol only supports NI-TC packetization mode.

2.2.5 NAL Unit Usage

The syntax of the NAL unit format and the meaning of the NAL unit header fields as defined in [RFC6184] and [RFC6190] with the following additional notes:

- PACSI NAL unit MUST be present in each layer in each access unit. It MUST be the first NAL unit of the layer. The PACSI NAL unit MAY be aggregated with NAL units into one STAP-A NAL unit. In that case it MUST be the first NAL unit present in the aggregated STAP-A NAL unit.
- PACSI NAL unit MUST not be fragmented.
- When a NAL unit is larger than the MTU size, it MUST be fragmented into multiple FU-A NAL units.
- Multiple small NAL units of the same layer of the same access unit MAY be aggregated into one STAP-A NAL unit. The size of STAP-A NAL unit MUST not exceed the MTU size.
- All other NAL unit types are passed to the decoder without any processing.

2.2.6 Stream Layout SEI Message

The stream layout is a structure that describes information about all layers present in the current simulcast streams. This provides a reliable way for the receiver to retrieve the information about the simulcast streams without waiting to receive NAL units from all layers.

This protocol defines a User Data Unregistered SEI message as the stream layout message.

The syntax of the User Data Unregistered SEI message followed in this protocol is as defined in [ISO/IEC14496-10:2010] Annex D.

The stream layout SEI message MUST be embedded in a PACSI NAL unit. The PACSI NAL containing the stream layout SEI message MAY be present in any layer and MAY not be followed by any VCL NAL unit.

The format of stream layout SEI message is defined as follows:

										1										2										3	
0	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	20	1	2	3	4	5	6	7	8	9	30	1
F	NRI		Type					payloadType								payloadSize								uuid_iso_iec_11578							
uuid_iso_iec_11578																															
uuid_iso_iec_11578																															
uuid_iso_iec_11578																															
uuid_iso_iec_11578																								LPB0							
LPB1								LPB2								LPB3								LPB4							

LPB5	LPB6	LPB7	MaxPRID	R	P
LDSize	Layer Description				
More Layer Description …					

F (1 bit): A forbidden_zero_bit, as specified in [RFC6184], section 1.1.3.

NRI (2 bit): A nal_ref_idc, as defined in [RFC6184], section 1.1.3.

Type (5 bit): A nal_unit_type, as defined in [RFC6184], section 1.1.3. MUST set to 6.

payloadType (variable): A SEI payload type. MUST set to 5 to indicate User Data Unregistered SEI message. The syntax used by this protocol is as defined in [ISO/IEC14496-10:2010], section 7.3.2.3.1.

payloadSize (variable): SEI payload size. The syntax that is used by this protocol for this field is the same as defined in [ISO/IEC14496-10:2010], section 7.3.2.3.1. The payloadSize value is the size of the stream layout SEI message excluding the F, NRI, Type, payloadType, and payloadSize fields.

uuid_iso_iec_11578 (16 bytes): A uuid to indicate the SEI message is the stream layout. The value MUST be set to {0x139FB1A9-446A-4DEC-8CBF-65B1E12D2CFD}.

LPB0 (1 byte): A layer presence byte #0. The value of each bit indicates whether the layer identified the corresponding PRID (priority ID) is present in the current simulcast streams. The value of 1 means the layer is present. The value of 0 means the layer is not present. The less significant bit corresponds to a smaller PRID. This byte corresponds to layer PRID 0~7.

LPB1 (1 byte): Layer presence byte #1. This byte corresponds to layer PRID 8~15.

LPB2 (1 byte): Layer presence byte #1. This byte corresponds to layer PRID 16~23.

LPB3 (1 byte): Layer presence byte #1. This byte corresponds to layer PRID 24~31.

LPB4 (1 byte): Layer presence byte #1. This byte corresponds to layer PRID 32~39.

LPB5 (1 byte): Layer presence byte #1. This byte corresponds to layer PRID 40~47.

LPB6 (1 byte): Layer presence byte #1. This byte corresponds to layer PRID 48~55.

LPB7 (1 byte): Layer presence byte #1. This byte corresponds to layer PRID 56~63.

MaxPRID (6 bits): The maximum PRID set to 1 in LPB0~LPB7.

R (1 bit): Reserved bit. Must set to 0.

P (1 bit): A layer description presence flag. The value of 1 indicates the Layer description table and layer description size (LDSize) field are present. The value of 0 indicates not present.

LDSize (1 byte): A layer description size. The value indicates the size of layer description table. MUST be larger than or equal to 16.

Layer Description (variable): The size is defined by LDSize field. If P field is 1, there MUST be at least one and there MAY be more than one Layer Descriptions. If there is no need for Layer Description, P field MUST be set to 0 and LDSize and Layer Description fields MUST not be present.

The format of Layer Description is defined as follows:

												1													2						3
0	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	20	1	2	3	4	5	6	7	8	9	30	1
Coded Width																				Coded Height											
Display Width																				Display Height											
Bitrate																															
FPSIdx					LT				PRID							CB			R		R2										

Coded Width (2 bytes): A coded width of the resolution of the coded picture.

Coded Height (2 bytes): A coded height of the resolution of the coded picture.

Display Width (2 bytes): A display height of the resolution of the coded picture.

Display Height (2 bytes): A display height of the resolution of the coded picture.

Bitrate (4 bytes): The target bitrate of the coded NAL units in this layer, in bits-per-second unit.

FPSIdx (5 bits): Index of a predefined frame rate table. Each FPSIdx value corresponds to one frame rate value. The frame rate value is the target frame rate represented by this layer. The frame rate table is defined as follows:

FPSIdx	Frame Rate (fps)
0	7.5
1	12.5
2	15
3	25
4	30
5	50
6	60

Only **FPSIdx** values 0~6 are defined.

LT (3 bits): A layer type.

0: Base layer.

1: Temporal layer.

2: Rewritable CGS layer.

3: Non-Rewritable CGS layer.

4: MGS layer.

5: Spatial layer.

6~7: Reserved for future use.

PRID (6 bits): The priority ID associated with this layer. It is used to determine the RTP stream that the Layer Description describes. The PRID value of Layer Description MUST match the priority ID of the SVC stream.

CB (1 bit): Constrained baseline profile. The value of 1 indicates the layer is coded in constrained baseline profile. The value of 0 indicates the layer MAY not be coded in constrained baseline profile.

R (1 bit): Reserved bit for future use. The sender MUST set it to 0. The receiver SHOULD ignore it.

R2 (2 bytes): Reserved for future use. The sender MUST set it to 0. The receiver SHOULD ignore it.

2.2.6.1 Stream Layout Types

There are two types of stream layout SEI messages. They differ by the presence of the **Layer Description** field.

1. Full stream layout.

   The **P** field MUST be "1". There MUST be at least one **Layer Description** field. For each bit of value "1" in the layer presence bytes, there MUST be a **Layer Description** field with PRID as the layer index.

2. Update stream layout.

   The **P** field MUST be "0". The **LDSize** and **Layer Description** fields are not present.

A full stream layout defines the complete stream layout information of simulcast streams. It specifies how many layers are contained in the simulcast streams and the information about each layer. There MUST be one full stream layout PACSI inserted prior to any other NAL unit. More full stream layout PACSIs MAY be inserted as needed.

An update stream layout defines an update to a previous full stream layout or update stream layout. An update stream layout does not contain any detailed **Layer Description** field, so it MUST only be used to indicate the adding or removing of layers from the simulcast streams through changing the bit values of the layer presence bytes. A layer presence bit MUST not be set to "1" if the value of the same bit is "0" in the most recent full stream layout. This means that an update stream layout MUST not contain a new layer that is not defined in the previous full stream layout.

2.2.7 Cropping Info SEI Message

The cropping info is a structure that describes the frame coordinates of one or multiple cropping windows. Cropping windows are regions of interest that may be decided by a face tracker or by senders' preference at runtime. The provision of cropping info provides a reliable way for the receiver to properly crop a window out of the original frame without loss of important information when necessary.

This protocol defines a User Data Unregistered SEI message as the cropping info message.

The syntax of the User Data Unregistered SEI message is defined in [ISO/IEC14496-10:2010] Annex D.

The cropping info SEI message MAY be embedded in a PACSI NAL unit. The PACSI NAL containing the cropping info SEI message MUST be present in the base layer and MAY not be followed by any VCL NAL unit.

The format of cropping info SEI message is defined as follows:

<table>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>2</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>3</td><td></td></tr>
<tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>0</td><td>1</td></tr>
<tr><td>F</td><td colspan="2">NRI</td><td colspan="5">Type</td><td colspan="8">payloadType</td><td colspan="8">payloadSize</td><td colspan="8">uuid_iso_iec_11578</td></tr>
<tr><td colspan="32">uuid_iso_iec_11578</td></tr>
<tr><td colspan="32">uuid_iso_iec_11578</td></tr>
<tr><td colspan="32">uuid_iso_iec_11578</td></tr>
<tr><td colspan="24">uuid_iso_iec_11578</td><td colspan="8">numOfCropData</td></tr>
<tr><td colspan="8">crop_info_type</td><td colspan="8">frame_crop_confidence_level1</td><td colspan="16">frame_crop_left_offset1</td></tr>
<tr><td colspan="16">frame_crop_right_offset1</td><td colspan="16">frame_crop_top_offset1</td></tr>
<tr><td colspan="16">frame_crop_bottom_offset1</td><td colspan="8">frame_crop_confidence_level2</td><td colspan="8">frame_crop_left_offset2</td></tr>
<tr><td colspan="8">frame_crop_left_offset2</td><td colspan="16">frame_crop_right_offset2</td><td colspan="8">frame_crop_top_offset2</td></tr>
<tr><td colspan="8">frame_crop_top_offset2</td><td colspan="16">frame_crop_bottom_offset2</td><td colspan="8">…</td></tr>
<tr><td colspan="8">frame_crop_confidence_levelN</td><td colspan="16">frame_crop_left_offsetN</td><td colspan="8">frame_crop_right_offsetN</td></tr>
<tr><td colspan="8">frame_crop_right_offsetN</td><td colspan="16">frame_crop_top_offsetN</td><td colspan="8">frame_crop_bottom_offsetN</td></tr>
</table>

frame_crop_bottom_offset N	

F bit (1 bit): See the definition in [RFC6190].

NRI (2 bit): See the definition in [RFC6190].

Type (5 bit): See the definition in [RFC6190].

payloadType (1 Byte): Must be 5. See the definition in [ISO/IEC14496-10:2010].

payloadSize (21 + 17 * N Byte): N is numOfCropData. See the definition in [ISO/IEC14496-10:2010].

uuid_iso_iec_11578 (16 bytes): [Guid("BB7FC1A0-6986-4052-90F0-0929217539CF")]

numOfCropData(1 byte): Total number of cropping info set in the message.

crop_info_type (1 byte): Must be zero.

frame_crop_confidence_level1(1 byte): Specify the confidence level, quantified with a value between 0 and 100 with higher value meaning higher confidence, of the first cropping window. The cropping info confidence level is decided by the crop info processor. The value zero indicates the cropping confidence is indeterminate.

frame_crop_right_offset1 (2 bytes): Specify the offset from the right edge of the first cropping window to the right edge of the rectangular region of the coded video sequence, in terms of pixels.

frame_crop_left_offset1 (2 bytes): Specify the offset from the left edge of the first cropping window to the left edge of the rectangular region of the coded video sequence, in terms of pixels.

frame_crop_top_offset1 (2 bytes): Specify the offset from the top edge of the first cropping window to the top edge of the rectangular region of the coded video sequence, in terms of pixels.

frame_crop_bottom_offset1 (2 bytes): Specify the offset from the bottom edge of the first cropping window to the bottom edge of the rectangular region of the coded video sequence, in terms of pixels.

crop_confidence_level2 (1byte): Specify the confidence level, quantified with a value between 0 and 100 with higher value meaning higher confidence, of the second cropping window (if present). The cropping info confidence level is decided by the crop info processor. The value zero indicates the cropping confidence is indeterminate.

frame_crop_right_offset2 (2 bytes): Specify the offset from the right edge of the second cropping window (if present) to the right edge of the rectangular region of the coded video sequence, in terms of pixels.

frame_crop_left_offset2 (2 bytes): Specify the offset from the left edge of the second cropping window (if present) to the left edge of the rectangular region of the coded video sequence, in terms of pixels.

frame_crop_top_offset2 (2 bytes): Specify the offset from the top edge of the second cropping window (if present) to the top edge of the rectangular region of the coded video sequence, in terms of pixels.

frame_crop_bottom_offset2 (2 bytes): Specify the offset from the bottom edge of the second cropping window (if present) to the bottom edge of the rectangular region of the coded video sequence, in terms of pixels.

crop_confidence_levelN(1byte): Specify the confidence level, quantified with a value between 0 and 100 with higher value meaning higher confidence, of the nth cropping window (if present). The cropping info confidence level is decided by the crop info processor. The value zero indicates the cropping confidence is indeterminate.

frame_crop_right_offsetN(2 bytes): Specify the offset from the right edge of the nth cropping window (if present) to the right edge of the rectangular region of the coded video sequence, in terms of pixels.

frame_crop_left_offsetN(2 bytes): Specify the offset from the left edge of the nth cropping window (if present) to the left edge of the rectangular region of the coded video sequence, in terms of pixels.

frame_crop_top_offsetN(2 bytes): Specify the offset from the top edge of the nth cropping window (if present) to the top edge of the rectangular region of the coded video sequence, in terms of pixels.

frame_crop_bottom_offsetN(2 bytes): Specify the offset from the bottom edge of the nth cropping window (if present) to the bottom edge of the rectangular region of the coded video sequence, in terms of pixels.

The following diagram is an example of the offsets defined earlier in this section.

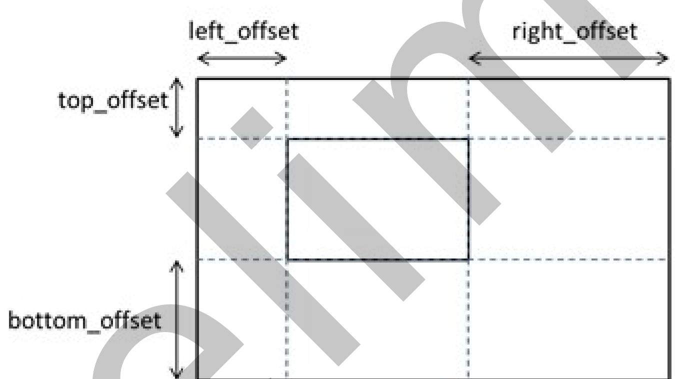

Figure 1: Offsets example

2.2.8 H264 Forward Error Correction (FEC) Payload Format

Forward Error Correct (**FEC**) is to use redundant packets to reduce the effect of network packet loss. This protocol uses the FEC payload format defined in [RFC5109] with some extensions.

2.2.8.1 H264 FEC Packet Structure

The FEC packet is constructed by placing an FEC header, FEC level header, FEC level extension header, and FEC level payload into the RTP payload, as shown in the following table.

There is an extra FEC level extension header in this protocol that differs from the FEC level extension header, as specified by [RFC5109].

Only one FEC level is allowed, which is similar to FEC level 0, as specified in [RFC5109].

RTP Header (12 bytes or more)
FEC Header (10 bytes)
FEC Level Header
FEC Level Extension Header
FEC Level Payload

2.2.8.1.1 RTP Header for FEC Packets

FEC packets MUST be added to the end of H.264 data packets for each layer for each access unit.

All the fields in the RTP header of FEC packets are used according to [MS-RTP], with some additional notes:

Marker: This field MUST set to 1 for the last FEC packet of each layer of each access unit. MUST set to 0 otherwise.

SSRC: The SSRC value MUST be the same as the data packets protected by this FEC packet.

Sequence Number: FEC packets MUST share the same numbering space as the data packets it protects.

Timestamp: The timestamp MUST be set to the same value as the data packets protects by this FEC packet.

Payload Type: The payload type for the FEC packets is determined through dynamic, out of band means. The payload type for FEC packets MUST be different from the payload type for the data packets.

2.2.8.1.2 FEC Header for FEC Packets

The FEC header format in this protocol adheres to the FEC header format as specified in [RFC5109] with the exception that it defines a SN Offset field instead of a SN base field.

										1										2										3	
0	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	20	1	2	3	4	5	6	7	8	9	30	1
E	L	P	X	CC				M	PT Recovery							SN Offset															

TS Recovery	
Length Recovery	

E (1 bit): An Extension flag. It MUST be set to 1.

L (1 bit): A Long mask flag, as defined in [RFC5109], section 7.3.

P (1 bit): A P recovery field, as defined in [RFC5109], section 7.3.

X (1 bit): An X recovery field, as defined in [RFC5109], section 7.3.

CC (4 bits): A CC recovery field, as defined in [RFC5109], section 7.3.

M (1 bit): An M recovery field, as defined in [RFC5109], section 7.3.

PT Recovery (7 bits): A PT recovery field, as defined in [RFC5109], section 7.3.

SN Offset (2 bytes): A sequence number (SN) offset field. The value is the difference between the RTP sequence number of the FEC packet and the lowest RTP sequence number of the data packets the FEC packet protects. It is the unsigned 32-bit value calculated by subtracting the lowest RTP sequence number from the RTP sequence number of the FEC packet with sequence number wraparound considered.

TS Recovery (4 bytes): A timestamp (TS) recovery field, as defined in [RFC5109], section 7.3.

Length Recovery (2 bytes): Used to determine the length of any recovered packets. This field is as defined in [RFC5109], section 7.3, with the exception that it does not include the length of CSRC list, extension, and padding.

2.2.8.1.3 FEC Level Header for FEC Packets

The format of FEC level header, as defined in [RFC5109], section 7.4.

										1										2										3	
0	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	20	1	2	3	4	5	6	7	8	9	30	1
Protection Length																mask															
mask cont. (present only when L = 1)																															

The protection length field and the mask field have the same meaning as defined in [RFC5109].

2.2.8.1.4 FEC Level Extension Header

This protocol defines a FEC level extension header between the FEC level header and FEC level payload.

										1					
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**0**	**1**	**2**	**3**	**4**	**5**
V	C	HR1	HR2	Reserved				FEC Count				FEC Index			

V (1 bit): Version field. This field MUST be set to 0.

C (1 bit): C field. This field MUST be set to 0.

HR1 (1 bit): Header recovery bit1 field. It is used to store the one bit in the FEC protection string.

HR2 (1 bit): Header recovery bit2 field. It is used to store the one bit in the FEC protection string.

Reserved (4 bits): Reserved for future use. It MUST be set to 0.

FEC Count (4 bits): The number of FEC packets generated by a protection operation.

FEC Index (4 bits): The index of an FEC packet in all the FEC packets generated by a protection operation. The value ranges from 0 to 15. At most 15 FEC packets CAN be generated by one protection operation.

2.3 Directory Service Schema Elements

None.

3 Protocol Details

3.1 Sender Details

This section covers the role of the sender of H.264 NAL units.

3.1.1 Abstract Data Model

None.

3.1.2 Timers

None.

3.1.3 Initialization

None.

3.1.4 Higher-Layer Triggered Events

3.1.4.1 Send an H.264 NAL Unit

Whenever higher layers send an H.264 NAL unit and the NAL unit is very small, the sender is not required to send it out right away. Instead, it MAY buffer the NAL unit, wait to receive the next NAL unit from higher layers, and aggregate the next NAL unit with the previously received or aggregated NAL units into a new STAP-A aggregation packet.

Whenever higher layers send an H.264 NAL unit and the NAL unit does not fit into one RTP packet because the NAL unit size is larger than the Maximum Transport Unit (MTU) size), the NAL unit MUST be fragmented into multiple FU-A NAL units.

3.1.5 Message Processing Events and Sequencing Rules

3.1.5.1 Packetization Rules

The sender applies the same packetization rule as defined in [RFC6184] section 6 with the following additional notes:

1. The sender MUST work in non-interleaved mode. Only single NAL unit packets, STAP-A units, and FU-A units are allowed. STAP-B, MTAP, and NI-MTAP MUST not be used.
2. One RTP session MUST only carry NAL units for the same layer. This means all PACSI units and type 20 units (coded slice in scalable extension) MUST have the same PRID value in the same coded sequence.
3. PACSI MUST be the first NAL unit on its own, or the STAP-A packet containing PACSI MUST be the first NAL unit of a layer in an access unit when PACSI is aggregated. A PACSI MUST be the first NAL unit in the aggregated STAP-A packet when the PACSI NAL unit is aggregated. There MUST be only one PACSI NAL unit in any layer in any access unit.
4. The stream layout SEI message MUST be contained in a PACSI NAL unit.

3.1.5.2 Generation of Forward Error Correction (FEC) Packet

The FEC packets are generated by applying a protection operation on protected bit strings derived from the protected media RTP packets.

3.1.5.2.1 Generation of the FEC Header, FEC Level Extension Header and FEC Level Header

For each media packet, the protected bit string (64 bit long) for the generation of the FEC header is formed by concatenating the following fields together in the order specified:

- 2 bits of value 0 (2 bits)
- Padding bit of the media packet (1 bit)
- Extension bit of the media packet (1 bit)
- 4 bits of value 0 (4 bits)
- Marker bit of the media packet (1 bit)
- PT field of the media packet (7 bits)
- 32 bits of value 0 (32 bits)
- Unsigned network-ordered 16-bit representation of the media payload. It MUST not include the 12-byte fixed RTP header, CSRC list, extension, and padding.

Then the FEC bit strings are formed by applying protection operation on the protected bit strings. There MAY be multiple FEC bit strings generated by one protection operation. Each FEC bit string is used to form a FEC packet.

The FEC header and the FEC level extension header of each FEC packet is generated from the corresponding FEC bit string as follows:

The first (most significant) bit in the FEC bit string is written into the HR1 field of the FEC level extension header. The next bit in the FEC bit string is written into the HR2 field of the FEC level extension header. The next bit in the FEC bit string is written into the P recovery field of the FEC header. The next bit in the FEC bit string is written into the X recovery field of the FEC header. The next 4 bits in the FEC bit string are written into the CC recovery field of the FEC header. The next bit in the FEC bit string is written into the M recovery field of the FEC header. The next 7 bits in the FEC bit string are written into the PT recovery field of the FEC header. The next 32 bits in the FEC bit string are written into the TS recovery field of the FEC header. The next 16 bits are written into the length recovery field in the FEC header.

The extension (E) field of the FEC header MUST be set to 1. The long mask (L) field MUST be set to 1 if the number of protected media packets is larger than 16 and MUST be set to 0 otherwise.

The SN Offset field of the FEC header is set to an unsigned network-ordered value calculated by subtracting the smallest RTP sequence number from the FEC packet sequence number.

The V field, C field, and Reserved field of the FEC level extension header MUST be set to 0.

The FEC Count field of the FEC level extension header MUST be set to the number of FEC packets generated by the protection operation. The FEC Index field of the FEC level extension header MUST be set to the index of the FEC packet starting from 0.

The protection length field and the mask field of the FEC level header MUST be written as defined in [RFC5109].

3.1.5.2.2 FEC Protection Operation Algorithms

This protocol defines two types of algorithms for FEC protection operation.

- Exclusive OR (XOR)

  XOR is used when the number FEC packet applied to protect one or more media packets is 1.

- Reed-Solomon

  The detail of this algorithm is beyond the discussion of this protocol.

3.1.5.3 Signalling of Simulcast

Multiple simulcast streams MAY be encoded using the same camera feed at the same time. Each simulcast stream MAY be decoded by itself and MAY contain multiple layers, including one base layer and 0 or more enhancement layers.

The sender signals the simulcast by using stream layout PACSI NAL units. The format of a stream layout PACSI is defined in section 2.2.6 in this protocol.

A full stream layout (section 2.2.6.1) PACSI NAL unit MUST be sent as the very first NAL unit.

Whenever one or more layers are removed, an update stream layout (section 2.2.6.1) or a full stream layout PACSI NAL unit MUST be sent. The update stream layout PACSI NAL unit SHOULD be used if no layer information has changed since the previous full stream layout and there is no need for the layer presence bytes to be present.

Whenever one or more layers are added and the layer descriptions are not changed, an update stream layout or a full stream layout PACSI NAL unit MUST be sent. For the same reason noted in the previous paragraph, the update stream layout PACSI NAL unit is preferred.

For other changes in the simulcast streams, a full stream layout PACSI NAL unit MUST be sent. This includes a new layer that is never present before it is added or any **Layer Description** field changes.

A stream layout PACSI MAY be sent in any layer. There MAY not be any VCL data NAL units following it in the same access unit. A stream layout PACSI MUST be sent prior to the change it signals. This means that when a layer is added into the simulcast streams, the stream layout PACSI NAL unit MUST be sent before any other NAL unit of the new layer is sent. After a stream layout PACSI NAL unit is sent with a layer removed from the simulcast streams, no NAL unit of the removed layer is allowed to be sent. This is needed so that the receiver has timely, accurate information about the simulcast streams.

3.1.5.3.1 RTVideo Simulcast Stream

An RTVideo simulcast stream is also allowed to coexist with H.264 simulcast streams. Only at most one RTVideo simulcast stream is allowed. The RTVideo simulcast stream is encoded using the same camera feed as other H.264 simulcast streams. An RTVideo simulcast stream is sent on its own RTP stream with its own SSRC.

The presence of RTVideo simulcast stream is not signaled through the stream layout PACSI. Instead, the receiver MUST recognize an RTVideo simulcast stream by looking at RTP payload type in the RTP packets. The RTVideo payload type is negotiated through SDP, as defined in [MS-SDP].

A mixer receivers simulcast streams MAY decide to only forward one of the simulcast streams to the downstream endpoints. It MUST find the presence by looking at the stream layout PACSI and forward one of the H.264 simulcast streams if present. If there is no H.264 simulcast stream present, then it MUST forward RTVideo stream.

3.1.6 Timer Events

None.

3.1.7 Other Local Events

None.

3.2 Receiver Details

This section covers the role of receiver of H.264 NAL units.

3.2.1 Abstract Data Model

None.

3.2.2 Timers

None.

3.2.3 Initialization

None.

3.2.4 Higher-Layer Triggered Events

None.

3.2.5 Message Processing Events and Sequencing Rules

3.2.5.1 DePacketization Rules

When receiving an H.264 RTP packet, the de-packetization process specified in [RFC6184] section 7.1 applies.

Additionally,

- Each layer associated with the same priority ID has its own receiver buffer.
- If the first packet in a RTP stream identified by a SSRC of the same timestamp is neither a PACSI NAL unit nor an STAP-A packet with the PACSI NAL unit as the first NAL unit, then all the packets of the same SSRC and the same timestamp MUST be discarded.

- If a packet is received and it is not stream layout PACSI NAL unit and is not STAP-A packet containing stream layout PACSI NAL unit, it MUST be discarded when any of the following conditions is met:
  - There is no stream layout PACSI NAL unit received
  - The Layer Presence bit in the previous stream layout is not set or there is no Layer Description associated with the priority ID of the NAL unit.

3.2.5.2 Recovery Procedures

When the loss of media packets happens, the FEC packets CAN be used to recover the lost media packets.

This protocol only allows FEC level 0 and only allows the recovery the full media packet. Partial media packet recovery is not supported.

When an FEC packet is received, check whether any media packet protected by the FEC packet is lost and whether the received FEC packets are enough to recover the loss. In general the lost media packets can be recovered if the number of FEC packets is equal to or larger than the number of lost media packets.

3.2.5.2.1 Recovery of the RTP Header

For each received media packet, the protected bit string is formed as specified in section 3.1.5.2.1.

For each received FED packet, the FEC bit string (64 bit long) is formed by concatenating the following fields together in the order specified:

- HR1 field of the FEC level extension header (1 bit)
- HR2 field of the FEC level extension header (1 bit)
- P recovery field of the FEC header (1 bit)
- X recovery field of the FEC header (1 bit)
- CC recovery field of the FEC header (4 bits)
- M recovery field of the FEC header (1 bit)
- PT recovery field of the FEC header (7 bits)
- TS recovery field of the FEC header (32 bits)
- Length recovery field of the FEC header

The protected bit strings of the media packets with the corresponding indices, and the FEC bit strings of the FEC packets with the corresponding indices are passed to the recovery operation to generate the recovery bit strings for the lost media packets.

For each recovery bit string, the RTP header of the lost media packet is recovered as follows:

1. Create a new RTP packet.
2. Set the version field in the new packet to 2. Skip the first 2 bits in the recover bit string.

3. Set the Padding bit in the new packet to the next bit in the recovery bit string.

4. Set the Extension bit in the new packet to the next bit in the recovery bit string.

5. Set the CC field in the new packet to the CC field of the FEC packet. Skip the next 4 bits in the recovery bit string.

6. Set the marker bit in the new packet to the next bit in the recovery bit string.

7. Set the payload type in the new packet to the next 7 bits in the recovery bit string.

8. Set the SN field in the new packet to SNi. The lost packet corresponds to Bit I in the mask field of the FEC header. Then SNi = SN of the FEC packet – SN Offset field + i.

9. Set the TS field to the TS field of the FEC packet. Skip the next 32 bits in the recovery bit string.

10. Set the SSRC of the new packet to the SSRC of the FEC packet.

11. Set the CSRC list of the new packet to be the CSRC list of the FEC packet.

12. Read the next 16 bits in the recovery bit string. It's the length of payload of the new packet in network order.

3.2.5.2.2 Recovery of the RTP Payload

For each received media packet, the protected bit string is the RTP payload exclude the 12-byte fixed RTP header and CSRC list.

For each received FEC packet, the FEC bit string is the FEC level 0 payload.

The protection length field of the FEC level 0 header is the length of the FEC bit string in byte.

If any protected bit string is shorter than protection length, then it MUST be padded to the protection length with 0 at the end.

After the recovery operation is complete, the length of RTP payload of the recovered media packet MAY be shorter than the protection length and MUST be set to the value obtained in section 3.2.5.2.1.

3.2.6 Timer Events

None.

3.2.7 Other Local Events

None.

4 Protocol Examples

4.1 Stream Layout SEI Message Example

Stream Layout SEI message bytes in network byte order:

0x06, 0x05, 0x3a, 0x13, 0x9f, 0xb1, 0xa9, 0x44, 0x6a, 0x4d, 0xec, 0x8c, 0xbf, 0x65, 0xb1, 0xe1, 0x2d, 0x2c, 0xfd, 0x00, 0x00, 0x00, 0x00, 0x00, 0x00, 0x00, 0x03, 0xe5, 0x10

The Stream Layout SEI contains fields of the following values:

F=0, NRI=0, Type=6 (SEI)

payloadType=5 (user data unregistered user SEI), payloadSize=0x3a

uuid_iso_iec_11578 = 0x13, 0x9f, 0xb1, 0xa9, 0x44, 0x6a, 0x4d, 0xec, 0x8c, 0xbf, 0x65, 0xb1, 0xe1, 0x2d, 0x2c, 0xfd

LPB1~6=0, LPB7=03(PRID 56 and PRID 57 are present in this message)

MaxPRID=57, R= 0, P=1(Layer description table is present)

LDsize=0x10

Followed by Layer Description for PRID 56 in network byte order:

0x05, 0x00, 0x02, 0xd0, 0x05, 0x00, 0x02, 0xd0, 0x00, 0x16, 0xe3, 0x60, 0x10, 0xe0, 0x00, 0x00,

The Layer Description for PRID 56 contains fields of the following values:

Coded Width=1280(0x500), Coded Height=720(0x2d0)

Display Width=1280(0x500), Display Height=720(0x2d0)

Bitrate=1500000(0x0016e360)

FPSIndex=2,LayerType=0,PRID=56, CB=0, R=0, R2=0

Followed by Layer Description for PRID 57 in network byte order:

0x05, 0x00, 0x02, 0xd0, 0x05, 0x00, 0x02, 0xd0, 0x00, 0x0f, 0x42, 0x40, 0x21, 0xe4, 0x00, 0x00

The Layer Description for PRID 57 contains fields of the following values:

Coded Width=1280(0x500), Coded Height=720(0x2d0)

Display Width=1280(0x500), Display Height=720(0x2d0)

Bitrate=1000000(0x000f4240)

FPSIndex=4,LayerType=1,PRID=57, CB=0, R=0, R2=0

4.2 Cropping Info SEI Message Example

Cropping Info SEI message bytes in network byte order:

0x06, 0x05, 0x1b, 0xbb, 0x7f, 0xc1, 0xa0, 0x69, 0x86, 0x40, 0x52, 0x90, 0xf0, 0x09, 0x29, 0x21, 0x75, 0x39, 0xcf, 0x01, 0x00, 0xff, 0x01, 0x18, 0x01, 0x18, 0x00, 0x00, 0x00, 0x00

The Cropping Info SEI message contains fields of the following values:

F=0, NRI=0, Type=6(SEI)

payloadType=5 (user data unregistered user SEI), payloadSize=0x1b

uuid_iso_iec_11578=0xbb, 0x7f, 0xc1, 0xa0, 0x69, 0x86, 0x40, 0x52, 0x90, 0xf0, 0x09, 0x29, 0x21, 0x75, 0x39, 0xcf

numOfCropData=1, crop_info_type=0

frame_crop_confidence_level1=255,

frame_crop_left_offset=280(0x118)

frame_crop_right_offset=280(0x118)

frame_crop_top_offset=0

frame_crop_bottom_offset=0

4.3 H264 Forward Error Correction Example

H264 FEC Payload Header bytes in network byte order:

0x80,0x00,0x00,0x07,0x00,0x00,0x00,0x00,0x03,0x7B

The payload header contains fields of the following values:

E=1, L=0, P=0,X=0,CC=0,M=0,PT Recovery=0,SN offset=7,TS Recovery=0,Length Recovery=37B

Followed by FEC Level Header bytes in network byte order:

0x03, 0x68, 0xFC, 0x00

The FEC Level Header contains fields of the following values:

Protection Length=0x368, Mask=0xFC00

Followed by FEC Level Extension Header bytes in network byte order:

0x00, 0x10

The FEC Level Extension Header contains fields of the following values:

V=0, C=0, HR1=0, HR2=0, Reserved=0, FEC Count=1, FEC Index=0;

Followed by FEC payload:

0x64, 0x05, 0xD5, 0xA8 … (total 872 bytes)

5 Security

5.1 Security Considerations for Implementers

None.

5.2 Index of Security Parameters

None.

6 Appendix A: Product Behavior

The information in this specification is applicable to the following Microsoft products or supplemental software. References to product versions include released service packs:

- Microsoft® Lync Server 15 Technical Preview
- Microsoft® Lync 15 Technical Preview

Exceptions, if any, are noted below. If a service pack or Quick Fix Engineering (QFE) number appears with the product version, behavior changed in that service pack or QFE. The new behavior also applies to subsequent service packs of the product unless otherwise specified. If a product edition appears with the product version, behavior is different in that product edition.

Unless otherwise specified, any statement of optional behavior in this specification that is prescribed using the terms SHOULD or SHOULD NOT implies product behavior in accordance with the SHOULD or SHOULD NOT prescription. Unless otherwise specified, the term MAY implies that the product does not follow the prescription.

7 Change Tracking

No table of changes is available. The document is either new or has had no changes since its last release.

8 Index