International Telecommunication Union

# ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

# H.241

(09/2005)

SERIES H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS

Infrastructure of audiovisual services – Communication procedures

# Extended video procedures and control signals for H.300-series terminals

ITU-T Recommendation H.241

ITU-T

ITU-T H-SERIES RECOMMENDATIONS

**AUDIOVISUAL AND MULTIMEDIA SYSTEMS**



*For further details, please refer to the list of ITU-T Recommendations.*

## ITU-T Recommendation H.241

# Extended video procedures and control signals for H.300-series terminals


**Summary**

This Recommendation defines the use of advanced video codecs, including ITU-T Rec. H.264, in H.310, H.320, H.321, H.322, H.323 and H.324 terminals. It also defines generic extended signalling for use with all video codecs in the H.300-series terminals.

This revised version (2005) incorporates the contents of ITU-T Rec. H.241 (2003), Corrigendum 1 (March 2004) and Amendment 1 (January 2005), and adds the new MaxStaticMBPS parameter for ITU-T Rec. H.264 and support for H.264 packetization according to RFC 3984.



**Source**

ITU-T Recommendation H.241 was approved on 13 September 2005 by ITU-T Study Group 16 (2005-2008) under the ITU-T Recommendation A.8 procedure.



**Keywords**

capability exchange, commands, H.264, H.310, H.320, H.321, H.322, H.323, H.324, signalling, video, video codec, video coding, videoconferencing, videotelephony.

## FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

## NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g. interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

## INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

# CONTENTS

# ITU-T Recommendation H.241

## Extended video procedures and control signals for H.300-series terminals

## 1 Scope

This Recommendation defines the procedures for use of advanced video codecs, including ITU-T Rec. H.264, with H.300-series terminals, including H.310, H.320, H.321, H.322, H.323 and H.324. Such procedures include control, indication, capability exchange, and transport mechanisms.

Additionally, this Recommendation defines generic extended video control, indication, and capability signalling, applicable for use with all video codecs used in H.300-series multimedia terminals.

This revised version (2005) incorporates the contents of H.241 (2003), Corrigendum 1 (March 2004) and Amendment 1 (January 2005) which enables signalling of the new H.264 Profiles, and updates Table 5/H.241 to enable the signalling of the new H.264 Level 1b, introduced in H.264 (2005). It also adds the new MaxStaticMBPS parameter for ITU-T Rec. H.264 and support for H.264 packetization according to RFC 3984.

## 2 References

### 2.1 Normatives references

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.


– ITU-T Recommendation H.221 (2004), *Frame structure for a 64 to 1920 kbit/s channel in audiovisual teleservices*.

– ITU-T Recommendation H.230 (2004), *Frame-synchronous control and indication signals for audiovisual systems*.

– ITU-T Recommendation H.239 (2005), *Role management and additional media channels for H.300-series terminals*.

– ITU-T Recommendation H.242 (2004), *System for establishing communication between audiovisual terminals using digital channels up to 2 Mbit/s*.

– ITU-T Recommendation H.245 (2005), *Control protocol for multimedia communication*.

– ITU-T Recommendation H.261 (1993), *Video codec for audiovisual services at* $p \times 64$ *kbit/s*.

– ITU-T Recommendation H.263 (2005), *Video coding for low bit rate communication*.

– ITU-T Recommendation H.264 (2005), *Advanced video coding for generic audiovisual services*.

– ITU-T Recommendation H.310 (1998), *Broadband audiovisual communication systems and terminals*.

– ITU-T Recommendation H.320 (2004), *Narrow-band visual telephone systems and terminal equipment*.

– ITU-T Recommendation H.323 (2003), *Packet-based multimedia communications systems*.

– ITU-T Recommendation H.324 (2005), *Terminal for low bit-rate multimedia communication*.

– IETF RFC 3550 (2003), *RTP: A Transport Protocol for Real-Time Applications*.

– IETF RFC 3984 (2005), *RTP Payload Format for H.264 Video.*

### 2.2 Informative references

– ITU-T Recommendation H.263 Appendix III (2001), *Examples for H.263 encoder/decoder implementations.*

## 3 Definitions

This Recommendation defines the following term:

**3.1 terminal**: A terminal is any endpoint and may be a user's terminal or some other communication system such as an MCU or an information server.

## 4 Abbreviations

This Recommendation uses the following abbreviations:

| | |
|---|---|
| AL-SDU | Adaptation Layer Service Data Unit (see ITU-T Rec. H.324) |
| ASN.1 | Abstract Syntax Notation One (see ITU-T Rec. H.245) |
| BAS | Bit-rate Allocation Signal (see ITU-T Rec. H.221) |
| C&I | Control & Indication |
| IDR | Instantaneous Decoding Refresh (see ITU-T Rec. H.264) |
| MBE | Multiple Byte Extension (see ITU-T Rec. H.221) |
| OID | Object Identifier (see ITU-T Rec. H.245) |

## 5 Conventions

### 5.1 System terminology

In order to simplify references, this Recommendation refers to two classes of signalling systems for H.300-series terminals.

– "BAS-based systems" refers to those systems that use signalling in the H.221 BAS channel; these include H.320, H.321 and H.322 systems.

– "H.245-based systems" refers to those systems that use signalling according to ITU-T Rec. H.245; these include H.310, H.323 and H.324 systems.

### 5.2 Message names

In this Recommendation signalling messages which are common to both H.245 and BAS signalling systems are referred to by their names as given in Annex A/H.245, except in cases where their use in the unique BAS signalling environment are described. Message names are presented in **bold font** to distinguish them from the other text of this Recommendation.

Table 1 provides a reference for corresponding H.245 and H.242/H.230 messages mentioned in this Recommendation.

**Table 1/H.241 – Corresponding H.245 and BAS video signals**

| H.245 name | H.230 mnemonic |
|---|---|
| h263Options.customPictureClockFrequency | ØCPCF |
| h263Options.customPictureFormat | ØCSFMT |
| h263Options.customPictureFormat | ØCPAR |
| h263VideoCapability.enhancementLayerInfo | ØSCLPREF |
| lostPartialPicture | lostPartialPicture |
| lostPicture | lostPicture |
| recoveryReferencePicture | recoveryReferencePicture |
| videoBadMBs | VBMBC |
| videoFastUpdateGOB | videoFastUpdateGOB |
| videoFastUpdateMB | videoFastUpdateMB |
| videoFastUpdatePicture | VCU |
| videoFreezePicture | VCF |
| videoNotDecodedMBs | videoNotDecodedMBs |
| videoSendSyncEveryGOB | ØGHOP |
| videoSendSyncEveryGOBCancel | Øcancel-GHOP |

### 5.3 Requirement terminology

In this Recommendation the following conventions are used:

– "Shall" indicates a mandatory requirement.

– "Should" indicates a suggested but optional course of action.

– "May" indicates an optional course of action rather than a recommendation that something take place.

## 6 Commands and indications

### 6.1 C&I applicable to all video codecs

For further study.

### 6.2 C&I for use with ITU-T Rec. H.264

The following C&I signals shall not be used with regard to any channel operating according to ITU-T Rec. H.264:

– BAS signals **ØCPCF, ØCSFMT, ØCPAR, ØSCLPREF**

– **lostPartialPicture**

– **lostPicture**

– **recoveryReferencePicture**

– **videoBadMBs**

– **videoFastUpdateGOB**

– **videoFastUpdateMB**

– **videoNotDecodedMBs**
– **videoSendSyncEveryGOB**
– **videoSendSyncEveryGOBCancel**

NOTE 1 – The above signals are either specific to ITU-T Rec. H.263, or have parameters which do not correspond to H.264 structures or value ranges. Replacement signals which could be used either with ITU-T Rec. H.264, or in a generic form for any video codec, are for further study.

All other C&I not mentioned in this clause shall be used as specified elsewhere.

NOTE 2 – For example, the use of videoIndicateReadyToActivate and corresponding BAS signal VIR is not affected by this Recommendation.

### 6.2.1 videoFreezePicture command in ITU-T Rec. H.264

When a video decoder according to ITU-T Rec. H.264 receives a **videoFreezePicture** command, it shall freeze its displayed picture until one of the following events:

a) A recovery point signalled in a recovery point SEI message (D.2.7/H.264) is reached.

b) Reception of an IDR picture.

c) A timeout period of at least 6 seconds has elapsed since the **videoFreezePicture** command was received.

### 6.2.2 videoFastUpdatePicture command in ITU-T Rec. H.264

When a video encoder according to ITU-T Rec. H.264 receives a **videoFastUpdatePicture** command, the encoder shall enter the fast update mode by using one of the procedures specified in 6.2.2.1 or 6.2.2.2. The procedure in 6.2.2.1 is the preferred response in a lossless transmission environment. Both procedures satisfy the requirement to enter the fast update mode for H.264 video encoding.

NOTE 1 – The procedures re-initialize a H.264 decoder completely such that valid video frames will be decoded. Such re-initialization is effective regardless of whether or not the decoder was previously decoding any video stream from any endpoint.

The procedure should be accomplished as quickly as possible, but the re-initialization video stream shall be completely transmitted within 3 seconds of receiving the **videoFastUpdatePicture** command.

NOTE 2 – The 3-second requirement is needed to avoid timeout of the 6-second timer associated with the **videoFreezePicture** command, taking into account network and system latencies and possible cascaded MCUs. The **videoFreezePicture** command is used by MCUs as part of the video switching procedure (see 6.1.1/H.243).

#### 6.2.2.1 IDR procedure to respond to videoFastUpdatePicture

This clause gives one possible way to respond to **videoFastUpdatePicture**.

The encoder shall, in the order presented here:

1) Immediately prepare to send an IDR picture (see clause 3/H.264).

2) Send a H.264 sequence parameter set corresponding to the IDR picture to be sent. The encoder may optionally also send other parameter sets.

3) Send a H.264 picture parameter set corresponding to the IDR picture to be sent. The encoder may optionally also send other parameter sets.

4) Send the IDR picture.

5) From this point forward in time, send or re-send any other sequence or picture parameter sets, not sent in this procedure, prior to their reference by any H.264 slice, regardless of whether such parameter sets were previously sent prior to receiving the **videoFastUpdatePicture** command. Such parameter sets may be sent all at once (within

the limits of ITU-T Rec. H.264), one at a time as needed, or in any combination of these methods. Parameter sets may be re-sent at any time for redundancy.

#### 6.2.2.2 Gradual recovery procedure to respond to videoFastUpdatePicture

This clause gives one possible way to respond to **videoFastUpdatePicture**.

The encoder shall, in the order presented here:

1) Send a recovery point SEI message (D.2.7/H.264).
2) Repeat any sequence and picture parameter sets that were sent before the recovery point SEI message, prior to their reference in a H.264 slice.

The encoder shall ensure that the decoder has access to all reference pictures for inter prediction of pictures at or after the recovery point in output order. For example, the encoder may mark all reference pictures as "unused for reference" by issuing a memory_management_control_operation equal to 5 (see 8.2.5/H.264).

The value of the recovery_frame_cnt syntax element in the recovery point SEI message shall be such that the time between the reception of the **videoFastUpdatePicture** command and completing the transmission of the access unit including the recovery point as specified in D.2.7/H.264 is less than or equal to 3 seconds.

Re-sending of parameter sets may be done all at once (within the limits of ITU-T Rec. H.264), one at a time as needed, or in any combination of these methods. Parameter sets may be re-sent at any time for redundancy.

### 6.2.3 Recovery point SEI message

H.264 video decoders in H.300-series terminals shall support reception of the recovery point SEI message (see D.2.7/H.264) and identify the signalled recovery point.

Upon reception of a recovery point SEI message, the decoder shall continue to decode until the recovery point regardless of apparent errors in the stream such as reference to absent pictures, and should not send a **videoFastUpdatePicture** command in response to such apparent error.

If a **videoFreezePicture** is in force the decoder shall not display the decoded pictures, and shall continue to display the previously frozen picture. If the broken_link_flag in the recovery point SEI message is set, the decoder may choose not to display decoded pictures until the recovery point is reached.

If the decoder detects bitstream corruption between the SEI message and the recovery point in decoding order, a **videoFastUpdatePicture** command should be sent.

### 6.2.4 H.264-on BAS command

For BAS-based systems, the H.264-on BAS command defined in ITU-T Rec. H.221 shall be used to signal that video according to ITU-T Rec. H.264 is being transmitted. This command shall be used analogously to the BAS command H.261-on. Video shall occupy the same capacity as stipulated in ITU-T Rec. H.221 for the case of H.261 video.

# 7 Transport of coded video in H.300-series systems

## 7.1 Transport of H.264 video streams

Regardless of which H.300-series system is in use (ITU-T Recs H.310, H.320, H.321, H.322, H.323, or H.324), all H.264 encoders should take the Maximum Transmission Unit (MTU) size of IP networks into account when choosing the maximum length of H.264 NAL Units, as H.323 gateways may be used to transport these streams on IP networks.

To be transported in a maximum-length RTP packet according to H.323, H.264 NAL Units should be less than 64 000 bytes long. This value allows a substantial margin for packet header information.

To avoid IP-layer packet fragmentation (which may increase header overhead and the probability of loss due to errors), H.264 NAL Units should be substantially shorter than the MTU size of the network. For example, on an Ethernet network with a 1472 byte MTU, a 1200 byte NAL Unit allows for addition of considerable header overhead without exceeding the MTU size of the network.

### 7.1.1 Parameter set transmission

H.264 parameter set information shall be transmitted in-band to the H.264 video stream (see the Notes in 7.4.1.2.1/H.264).

Terminals sending H.264 video shall transmit each sequence or picture parameter set at a time prior to its reference by any H.264 slice. These parameter sets may be re-sent at any time for redundancy.

NOTE – There is no requirement that parameter sets must be transmitted each time they are about to be referenced by a H.264 slice. The transmission may take place at any time prior to the reference. Ordinarily, many H.264 slices will refer to the same parameter set, with the parameter set being sent only once.

### 7.1.2 Use of H.264 in BAS-based systems

When carried in a BAS signalling-based system, the H.264 video shall make use of the byte stream format given in Annex B/H.264.

The resulting byte stream shall be transmitted using the framing and forward error correction method given in 5.4/H.261. This procedure is the same as that used for ITU-T Recs H.261 and H.263.

Terminals encoding H.264 video may insert fill bits using the fill indicator (Fi) as described in 5.4.3/H.261.

NOTE – Insertion of such fill may be useful, for example, to reduce the effective coded video data rate within the video channel to avoid exceeding a H.264 decoder's maximum video bitrate (MaxBR) as given in Annex A/H.264.

### 7.1.3 Transport of H.264 streams in H.310 systems

In H.310 systems, the H.264 video shall make use of the byte stream format given in Annex B/H.264. H.264 shall be used without BCH error correction and without error correction framing.

### 7.1.4 Transport of H.264 streams in H.323 systems

In H.323 systems, H.264 shall be used without BCH error correction and without error correction framing. H.323 systems shall not make use of the byte stream format given in Annex B/H.264.

All H.323 systems that support H.264 shall support carriage of the H.264 video stream according to Annex A, and shall signal this in their capability set by including MediaPacketizationCapability.rtpPayload.Type.payloadDescriptor.oid, with the OID having the value {itu-t(0) recommendation(0) h(8) 241 specificVideoCodecCapabilities(0) h264(0) iPpacketization(0) h241AnnexA(0)}.

H.323 systems that support H.264 should also support RFC 3984's non-interleaved mode and may support RFC 3984's interleaved mode, in addition to Annex A.

The capability of using RFC 3984 non-interleaved mode shall be signalled by including a MediaPacketizationCapability.rtpPayloadType.payloadDescriptor.oid, with the OID having the value {itu-t(0) recommendation(0) h(8) 241 specificVideoCodecCapabilities(0) h264(0) iPpacketization(0) RFC3984NonInterleaved(1)}.

The capability of using RFC 3984 interleaved mode shall be signalled by including a MediaPacketizationCapability.rtpPayloadType.payloadDescriptor.oid, with the OID having the value {itu-t(0) recommendation(0) h(8) 241 specificVideoCodecCapabilities(0) h264(0) iPpacketization(0) RFC3984Interleaved(2)}.

NOTE 1 – Since RFC 3984's single NAL unit mode and Annex A are technically identical, the codepoints above permit the use of all packetization modes of RFC 3984.

A sender which signals one of these packetization modes in its Open Logical Channel message shall transmit video according to the corresponding mode of RFC 3984 or Annex A.

In RFC 3984 interleaved mode, senders and receivers need to have a common understanding of the required buffer sizes for the interleaving buffer. Unless signalled explicitly, these buffer sizes shall take the following values:

| | |
|---|---|
| sprop-interleaving-depth | 80 |
| sprop-deint-buf-req | 65536 |

The explicit signalling of these parameters is for further study.

NOTE 2 – See RFC 3984 section 8.1 for a description of both parameters. The values given are sufficient to support macroblock-line interleaved packetization of video signals with 1080 lines at 8 Mbit/s. See III.2.3.1 of Appendix III/H.263 for a discussion of macroblock-line interleaved packetization.

#### 7.1.5 Transport of H.264 streams in H.324 systems

In H.324 systems, H.264 shall be used without BCH error correction and without error correction framing, and shall make use of the byte stream format given in Annex B/H.264.

H.264 encoders shall align the Annex B/H.264 start code prefix for the first NAL unit of each access unit with the start of an AL-SDU.

## 8 Capability exchange signalling

### 8.1 General

Terminals which display received video shall be capable of displaying any picture format and frame rate for which they signal the capability. The format used to display such received video streams is not required to match the exact format transmitted.

NOTE – For example, a video conferencing system which decodes ITU-T Rec. H.264 at a given Profile and Level must display any picture format and frame rate allowed by that Profile and Level.

If during an ongoing connection a terminal which is transmitting video receives a changed capability set, the terminal shall adapt its video coding method to conform with all the limitations signalled in the received capability set.

### 8.2 Signalling of H.245 generic parameters in BAS-based systems

This Recommendation signals a subset of H.245 **GenericParameter** structures in BAS channel MBE messages. These messages shall be carried in BAS-based systems using the procedures of Annex A/H.239. These procedures avoid emulation of the MBE BAS code.

### 8.3 H.264 capabilities

#### 8.3.1 General

H.300-series terminals may optionally support video according to ITU-T Rec. H.264.

The H.264 capability set is structured as a list of one or more H.264 capabilities, each of which includes:

- Profile (mandatory);
- Level (mandatory);
- Zero or more optional parameters.

These capabilities indicate the ability to decode using one or more H.264 Profiles. The exact syntax and semantics are given in the subclauses below. In the case of H.245-based systems, each capability is contained in a **GenericCapability** structure. For BAS-based systems, all capabilities are carried in a single MBE message.

The bitrate made available for a video stream by an H.300-series system may be less than the maximum video bitrate which decoders are required to support by Annex A/H.264. Terminals are not required to decode video streams which they do not receive.

#### 8.3.1.1 Optional parameters

For each H.264 capability, optional parameters may be signalled. These parameters permit a terminal to signal that, in addition to meeting the support requirements for the signalled Profile and Level, the terminal has additional capabilities. Such additional capabilities in decoders may permit encoders to send a video stream which takes advantage of these capabilities.

Terminals shall not signal a set of optional parameters indicating the practical capability to fully support a given Level, without also signalling support for that Level.

The optional parameters are:

1) CustomMaxMBPS – If present, indicates the decoder has a higher processing rate capability.
2) CustomMaxFS – If present, indicates the decoder can decode larger picture (frame) sizes.
3) CustomMaxDPB – If present, indicates the decoder has additional decoded picture buffer memory.
4) CustomMaxBRandCPB – If present, indicates the decoder can decode a higher video bitrate and has a correspondingly larger coded picture buffer.
5) MaxStaticMBPS – If present, indicates the maximum number of macroblocks per second the decoder could process in the hypothetical case that all macroblocks are static macroblocks (see 8.3.2.8).
6) max-rcmd-nal-unit-size – If present, indicates the maximum recommended NAL unit size in bytes. Encoders may exceed this size, but inefficiencies or an increased chance of loss due to errors might result (see 8.3.2.9).
7) max-nal-unit-size – If present, indicates the maximum NAL unit size, in bytes, that the receiver can process. The encoder shall not exceed this size (see 8.3.2.10).

If these parameters are present, the signalled values replace the MaxMBPS, MaxFS, MaxDPB, MaxBR, and MaxCPB values, respectively, in Table A-1/H.264 for the given Profile and Level, and indicate that in addition to fully conforming with the Profile and Level requirements, these additional capabilities are available at the decoder.

These optional parameters permit, for example, support of 1024 × 768 × 3 Hz while using Level 2 (CIF/30 Hz), a common mode for videoconferencing systems.

NOTE – The use of these optional parameters to signal decoder capabilities does not alter the requirement of ITU-T Rec. H.264 that the **level_idc** syntax element, set by the encoder in the video bitstream, indicate an Annex A/H.264 Level with which the bitstream fully conforms. The use of these optional parameters permits the encoder to send bitstreams with a Level higher than the Level capability of the decoder, if the bitstream exceeds the decoder's Level capability only within the limits of these optional parameters. To maximize interoperability, encoders should set **level_idc** to indicate the lowest Level of Annex A/H.264 that the bitstream fully conforms to.

All H.300-series systems which support H.264 shall support Baseline Profile, Level 1, in addition to any other Profiles, Levels or optional parameters.

### 8.3.2 H.264 generic capabilities for H.245

This clause defines the Generic Capabilities for H.264 in the H.245 signalling system.

If a terminal has the capability to decode according to more than one H.264 Profile with different Levels capabilities (for example, Baseline Profile at Level 3 and Extended Profile at Level 2) or with different optional parameters for each Profile, this may be signalled by a separate Generic Capability for each supported Profile.

NOTE – Parameter identifier value 0 is not defined, and should not be defined in the future. This value is reserved so that it can be used in the equivalent BAS signalling-based system MBE message as a demarcation between separate capabilities within the single MBE message, as defined in Annex A/H.239.

#### 8.3.2.1 H.264 capability identifier

See Table 2.

**Table 2/H.241 – H.264 capability identifier**

| | |
|---|---|
| Capability name | ITU-T Rec. H.241 H.264 Video Capabilities |
| Capability identifier type | standard |
| Capability identifier value | {itu-t(0) recommendation(0) h(8) 241 specificVideoCodecCapabilities(0) h264(0) generic-capabilities(1)} |
| maxBitRate | This field shall be included, in units of 100 bit/s |
| collapsing | This field shall contain the H.264 Capability Parameters as given below. |
| nonCollapsing | This field shall not be included. |
| nonCollapsingRaw | This field shall not be included. |
| transport | This field shall not be included. |

#### 8.3.2.2 H.264 profile parameter

See Table 3.

**Table 3/H.241 – H.264 capability parameter – Profile**

| Parameter name | Profile |
|---|---|
| Parameter description | This parameter is a Boolean array.<br>If bit 2 (value 64) is 1, this indicates the Baseline Profile.<br>If bit 3 (value 32) is 1, this indicates the Main Profile.<br>If bit 4 (value 16) is 1, this indicates the Extended Profile.<br>If bit 5 (value 8) is 1, this indicates the High Profile.<br>If bit 6 (value 4) is 1, this indicates the High 10 Profile.<br>If bit 7 (value 2) is 1, this indicates the High 4:2:2 Profile.<br>If bit 8 (value 1) is 1, this indicates the High 4:4:4 Profile.<br>All other bits are reserved, shall be set to 0, and shall be ignored by receivers.<br>In a decoder capability, for each bit set to 1, this means that the terminal is capable of decoding the indicated Profile(s) using the Level and other optional parameters in this Generic Capability.<br>In an OpenLogicalChannel message, for each bit set to 1, this means that the logical channel contents obey all constraints of the indicated Profile(s).<br>NOTE 1 – If in the future more H.264 Profiles are defined than the number of reserved bits can accommodate, additional Profiles could be signalled by allocating another parameter for additional Profiles.<br>NOTE 2 – Bit 1 remains reserved since, if the three high-order bits of this parameter are set, this could create an unintentional emulation of the MBE BAS code in ITU-T Rec. H.230. |
| Parameter identifier value | 41 |
| Parameter status | Mandatory.<br>This parameter shall appear exactly once in each Generic Capability. |
| Parameter type | booleanArray |
| Supersedes | This field shall not be included. |

### 8.3.2.3 H.264 level parameter

The Level parameter signals the H.264 level.

**Table 4/H.241 – H.264 capability parameter – Level**

| | |
|---|---|
| Parameter name | Level |
| Parameter description | Signals a value according to Table 5, indicating the H.264 Level. All other values are reserved and shall not be transmitted.<br>Terminals that receive this signal with a Level parameter value less than the lowest Level parameter value shown in Table 5 shall ignore this capability parameter.<br>NOTE – Such values are reserved for future use.<br>For all other received Level parameter values, the terminal shall interpret the signalled H.264 Level number as the H.264 Level number corresponding to the highest Level parameter value in Table 5 which is less than or equal to the received Level parameter value. |
| Parameter identifier value | 42 |
| Parameter status | Mandatory.<br>This parameter shall appear exactly once in each Generic Capability. |
| Parameter type | unsignedMin |
| Supersedes | This field shall not be included. |

**Table 5/H.241 – Level parameter values**

| Level parameter value | H.264 Level number |
|---|---|
| 15 | 1 |
| 19 | 1b |
| 22 | 1.1 |
| 29 | 1.2 |
| 36 | 1.3 |
| 43 | 2 |
| 50 | 2.1 |
| 57 | 2.2 |
| 64 | 3 |
| 71 | 3.1 |
| 78 | 3.2 |
| 85 | 4 |
| 92 | 4.1 |
| 99 | 4.2 |
| 106 | 5 |
| 113 | 5.1 |

NOTE – Table 5 and this parameter description are constructed such that new H.264 Levels defined in the future, which are strictly between or above existing Levels, may be inserted into Table 5 in the future. If new Levels are defined which do not meet these constraints, they may be inserted below the lowest existing Level. In that case, new rules for interpreting such Level parameter values will be needed.

### 8.3.2.4 H.264 CustomMaxMBPS processing rate parameter

The optional CustomMaxMBPS parameter permits a decoder to signal that it is capable of decoding video at a higher rate than required by the signalled Level. Encoders may use this knowledge, for example, to send pictures of a given size at a higher frame rate.

**Table 6/H.241 – H.264 capability parameter – CustomMaxMBPS**

| | |
|---|---|
| Parameter name | CustomMaxMBPS |
| Parameter description | CustomMaxMBPS is the maximum macroblock processing rate, in units of 500 macroblocks per second.<br>This optional parameter, when present, may be considered by the encoder to replace the MaxMBPS value in Table A-1/H.264 for the signalled Level. The value of (CustomMaxMBPS × 500) shall not be less than the value MaxMBPS for the Level given in Table A-1/H.264. |
| Parameter identifier value | 3 |
| Parameter status | Optional.<br>This parameter shall appear at most once in each Generic Capability. |
| Parameter type | unsignedMin |
| Supersedes | This field shall not be included. |

#### 8.3.2.5 H.264 CustomMaxFS frame size parameter

The optional CustomMaxFS parameter permits a decoder to signal that it is capable of decoding larger picture sizes than required by the signalled Level. Encoders may use this knowledge, for example, to send larger pictures at a proportionally lower frame rate.

**Table 7/H.241 – H.264 capability parameter – CustomMaxFS**

| | |
|---|---|
| Parameter name | CustomMaxFS |
| Parameter description | CustomMaxFS is the maximum frame size, in units of 256 luma macroblocks.<br>This optional parameter, when present, shall be considered to replace the MaxFS value in Table A.1/H.264 for the signalled Level. The value of (CustomMaxFS × 256) shall not be less than the value MaxFS for the Level given in Table A.1/H.264. |
| Parameter identifier value | 4 |
| Parameter status | Optional.<br>This parameter shall appear at most once in each Generic Capability. |
| Parameter type | unsignedMin |
| Supersedes | This field shall not be included. |

#### 8.3.2.6 H.264 CustomMaxDPB memory parameter

The optional CustomMaxDPB parameter permits a decoder to signal that it has more than the minimum amount of decoded picture buffer memory required by the signalled Level. Encoders may use this knowledge to construct coded video streams with improved compression.

A system which signals CustomMaxDPB shall be capable of storing the following number of decoded frames in its decoded picture buffer:

$$\mathrm{Min}(32768 \times \mathrm{CustomMaxDPB} \div (\mathrm{PicWidthInMbs} \times \mathrm{FrameHeightInMbs} \times 256 \times \mathrm{ChromaFormatFactor}), 16)$$

PicWidthInMbs, FrameHeightInMbs, and ChromaFormatFactor are defined in ITU-T Rec. H.264.

**Table 8/H.241 – H.264 capability parameter – CustomMaxDPB**

| | |
|---|---|
| Parameter name | CustomMaxDPB |
| Parameter description | CustomMaxDPB is the maximum decoded picture buffer size, in units of 32 768 bytes.<br>This optional parameter, when present, shall be considered to replace the MaxDPB value in Table A-1/H.264 for the signalled Level. The value of (CustomMaxDPB × 32 768) shall not be less than the value (MaxDPB × 1024) for the Level given in Table A-1/H.264. |
| Parameter identifier value | 5 |
| Parameter status | Optional.<br>This parameter shall appear at most once in each Generic Capability. |
| Parameter type | unsignedMin |
| Supersedes | This field shall not be included. |

#### 8.3.2.7 H.264 CustomMaxBRandCPB bitrate and coded picture buffer size parameter

The optional CustomMaxBRandCPB parameter permits a decoder to signal that it is capable of decoding video streams of higher bitrate, and that it has a correspondingly larger coded picture buffer, than required by the signalled Level. Encoders may use this knowledge, for example, to send higher bitrate video to achieve improved video quality.

**Table 9/H.241 – H.264 capability parameter – CustomMaxBRandCPB**

| | |
|---|---|
| Parameter name | CustomMaxBRandCPB |
| Parameter description | CustomMaxBRandCPB is the maximum video bitrate. The maximum coded picture buffer (CPB) size is derived from the maximum video bitrate.<br>The units for maximum video bitrate are 25 000 bit/s for the VCL HRD parameters (see A.3.1 item i/H.264) and 30 000 bit/s for the NAL HRD parameters (see A.3.1 item j/H.264).<br>The CPB size shall be derived as equal to the MaxCPB for the signalled Level (see Table A-1/H.264), multiplied by the ratio of the signalled maximum bitrate to the MaxBR for the signalled level.<br>For example, if a terminal signals Level 1.2 with CustomMaxBRandCPB equal to 62, this indicates a maximum video bitrate of 1.550 Mbit/s for VCL HRD parameters, a maximum video bitrate of 1.860 Mbit/s for NAL HRD parameters, and a CPB size of 4 036 458 bits ((62 × 25 000)/384 000) × 1000 × 1000.<br>This optional parameter, when present, shall be considered to replace the MaxBR and MaxCPB values in Table A-1/H.264 for the signalled Level. The bit rate signalled by the CustomMaxBRandCPB parameter shall not be less than the maximum bit rate given in the MaxBR column of Table A-1/H.264, for the Level signalled. |
| Parameter identifier value | 6 |

**Table 9/H.241 – H.264 capability parameter – CustomMaxBRandCPB**

| | |
|---|---|
| Parameter status | Optional.<br>This parameter shall appear at most once in each Generic Capability. |
| Parameter type | unsignedMin |
| Supersedes | This field shall not be included. |

#### 8.3.2.8 H.264 MaxStaticMBPS processing rate parameter

The optional MaxStaticMBPS parameter permits a decoder to signal that it is capable of decoding video containing static macroblocks at a higher rate than required by the signalled Level. Encoders may use this knowledge, for example, to send pictures of a given size at a higher frame rate.

In the H.264 context, static macroblocks are defined as macroblocks for which all of the following conditions are fulfilled:

1) codedBlockPatternLuma and CodedBlockPatternChroma, when assigned a value in H.264, are both equal to 0;
2) either of the following conditions are fulfilled:
   a) mb_type is equal to P_Skip or P_L0_16 × 16 and weighted_pred_flag is not equal to 1, or,
   b) mb_type is equal to B_Skip, B_Direct_16 × 16, B_L0_16 × 16, or B_L1_16 × 16 and weighted_bipred_idc is not equal to 1;
3) only a single list X for X = 0 or 1 (List 0 or List 1) is used in the inter prediction process for the macroblock, within which the values of mvLX[0], mvLX[1], and refIdxLX are all equal to 0;
4) either of the following conditions are fulfilled:
   a) the macroblock is a frame macroblock and the reference index value 0 refers to the immediately-preceding frame or complementary field pair in decoding order and the immediately-preceding picture in decoding order is not a non-paired field;
   b) the macroblock is a field macroblock and the reference index value 0 refers to the immediately-preceding field of the same parity in decoding order.

   NOTE – The conditions specified above result in a decoding process for macroblocks consisting of copying samples from the same position as the current macroblock in the preceding reference picture in decoding order. The conditions specified above also identify only those macroblocks for which at most one motion vector difference is present in the bitstream.-

All other macroblocks are non-static macroblocks.

**Table 9b/H.241 – H.264 capability parameter – MaxStaticMBPS**

| | |
|---|---|
| Parameter name | MaxStaticMBPS |
| Parameter description | MaxStaticMBPS is the maximum number of static macroblocks per second the decoder can process under the assumption that all macroblocks are static macroblocks, in units of 500 macroblocks per second.<br>When this optional parameter is present, the value of MaxMBPS in Table A-1/H.264 for the signalled Level should be considered by the encoder to be equal to the result of the following procedure:<br>1) If the optional parameter CustomMaxMBPS is signalled, set a variable *MaxMacroblocksPerSecond* equal to the value (CustomMaxMBPS × 500). Otherwise, set *MaxMacroblocksPerSecond* equal to the value of MaxMBPS for the Level given in Table A-1/H.264.<br>2) Set a variable $P_{non\text{-}static}$ to the proportion of non-static macroblocks in picture n.<br>3) Set a variable $P_{static}$ to the proportion of static macroblocks in picture n.<br>4) The value of MaxMBPS in Table A-1/H.264 for the signalled Level should be considered by the encoder to be equal to:<br>$$\frac{1}{\frac{P_{non\text{-}static}}{MaxMacroblocksPerSecond}+\frac{P_{static}}{MaxStaticMBPS\times 500}}$$<br>The encoder should recompute this value for each picture.<br>The value of (MaxStaticMBPS × 500) shall not be less than the value MaxMBPS for the Level given in Table A-1/H.264, and if CustomMaxMBPS is signalled, shall not be less than the value (CustomMaxMBPS × 500).<br>The computed value of MaxMBPS should be used by the encoder to determine the minimum interval between picture n and picture n+1, as specified in the references to MaxMBPS in Annex A/H.264. |
| Parameter identifier value | 7 |
| Parameter status | Optional.<br>This parameter shall appear at most once in each Generic Capability. |
| Parameter type | unsignedMin |
| Supersedes | This field shall not be included. |

#### 8.3.2.8.1 Use of H.264 MaxStaticMBPS example (informative)

This clause does not form an integral part of this Recommendation.

For example, suppose a Level 1.2 capable decoder (MaxMBPS = 6000) with a signalled MaxStaticMBPS value of 120 (a processing rate of 60,000 static macroblocks per second) is receiving XGA video (1024 × 768 luma samples per picture), which contains 3072 luma macroblocks per picture, and that only a mouse cursor is moving in the video scene. (This example assumes the decoder has a CustomMaxFS value that permits this picture size.)

Suppose further that encoding the mouse cursor region requires only 4 macroblocks in a particular picture, so all other macroblocks can be static macroblocks. The procedure described above yields a MaxMBPS of 59,305 macroblocks per second (1÷((4÷3072)÷6000)+(((3072 – 4)÷3072)÷60 000)).

This would permit the encoder to generate the next picture after an interval of 51.8 ms (3072÷59,305), corresponding to an instantaneous frame rate of 19.3 Hz (59,305÷3072), compared to an interval of 512 ms (3072÷6000), corresponding to an instantaneous frame rate of only 2.0 Hz without the use of MaxStaticMBPS.

#### 8.3.2.8.2 Determination of H.264 MaxStaticMBPS value (informative)

This clause does not form an integral part of this Recommendation. It provides informative guidance regarding considerations for the determination of the MaxStaticMBPS value for a given decoder implementation.

Practical decoder implementations make use of a wide variety of hardware and software architectures, and there may be no single method of determining a decoder's value of MaxStaticMBPS that is appropriate in all cases; determination of this value is left to the implementer.

One possible method is described here solely as an example:

1) given an implementation which can decode sequences containing only non-static macroblocks at a rate of $R_{decode}$ macroblocks per second.

2) An encoded test video sequence with a known number of macroblocks ($N$), proportion of static macroblocks ($P_{static}$), and of non-static macroblocks ($P_{non\text{-}static} = 1 - P_{static}$) can be decoded, and the time required to decode either each picture or the entire sequence measured by experiment ($T_{decode}$ in seconds).

3) The rate at which static macroblocks can be decoded (StaticMBPS) can be calculated by:

$$\text{StaticMBPS} = P_{\text{static}} \div (T_{\text{decode}} \div N - P_{\text{non-static}} \div R_{\text{decode}})$$

   This procedure can be repeated with different test sequences containing different proportions of static and non-static macroblocks and different picture sizes.

4) The values of StaticMBPS obtained can be plotted against varying values of $P_{static}$ and picture size tested, and interpolation applied between the test points. (Note that in many decoder implementation architectures, the plot of StaticMBPS vs. picture size will form a curve.)

5) The lowest value of StaticMBPS obtained on the plots could then be used as the value of MaxStaticMBPS.

In some decoder implementation architectures, the decoding rate is affected by a deblocking filter computation where static and non-static macroblocks are adjacent to each other. To take account of this factor, worst-case test patterns of static and non-static macroblocks can be used.

#### 8.3.2.9 H.264 max-rcmd-nal-unit-size

The value of this parameter indicates the largest NAL unit size in bytes that the receiver can handle efficiently. The parameter value is a recommendation, not a strict upper boundary. The sender may create larger NAL units but implementers should consider that inefficiencies or an increased chance of loss due to errors might result.

**Table 9c/H.241 – H.264 capability parameter – max-rcmd-nal-unit-size**

| | |
|---|---|
| Parameter name | max-rcmd-nal-unit-size |
| Parameter description | The value of this parameter indicates the largest NAL unit size in bytes that the receiver is able to handle efficiently. The parameter may have values in the range of 0 to 4 294 967 295, inclusive. |
| Parameter identifier value | 8 |
| Parameter status | Optional.<br>This parameter shall appear at most once in each Generic Capability. |
| Parameter type | Integer |
| Supersedes | This field shall not be included. |

#### 8.3.2.10 H.264 max-nal-unit-size

The value of this parameter indicates the largest NAL unit size in bytes that the receiver is able to handle at all. The sender shall not create NAL units larger than this size.

In the absence of this signal, senders shall not create NAL units larger than 1400 bytes when using the Interleaved or Non-Interleaved packetization modes. When operating in the Annex A packetization mode, senders should not create NAL units larger than 1400 bytes.

**Table 9d/H.241 – H.264 Capability Parameter – max-nal-unit-size**

| | |
|---|---|
| Parameter name | max-nal-unit-size |
| Parameter description | The value of this parameter indicates the largest NAL unit size in bytes that the receiver can process. The parameter can have values in the range of 0 to 4 294 967 295, inclusive. |
| Parameter identifier value | 9 |
| Parameter status | Optional.<br>This parameter shall appear at most once in each Generic Capability. |
| Parameter type | unsigned32Min |
| Supersedes | This field shall not be included. |

### 8.3.3 H.264 capabilities for BAS-based systems

#### 8.3.3.1 H.320 video algorithm hierarchy

The H.320 enhanced video algorithm hierarchy in Annex A/H.320 is not extended for H.264. No relative level in the hierarchy is specified for H.264 with regard to other video codecs.

#### 8.3.3.2 H.264 capabilities MBE message format

For H.264 operation, the capabilities exchange is handled by an MBE message (see 2.2.3/H.230). This MBE message uses the type identification byte <H.264> (see Table 2/H.230). A terminal shall signal the H.264 capability by including within its capability set the message:

$$\{ \text{Start-MBE} / N / \text{<H.264>} / B_1 / \ldots / B_{N-1} \}$$

The H.264 capability MBE bytes $B_1$ through $B_{N-1}$ may contain one or more encoding or decoding capabilities for H.264.

Each capability, which corresponds to a single H.245 **GenericCapability** message, consists of the mandatory Profile and Level parameters, and an optional set of zero or more **parameterIdentifier**/**parameterValue** pairs from the set of H.264 generic capability parameters defined in 8.3.2. These pairs are carried in the format given in 8.2 above.

Encoder capabilities are for further study.

The first two bytes of each decoder capability within the MBE shall contain the H.264 Profile parameter, followed by the H.264 Level parameter, as defined in Tables 3 and 4. No parameter identifier is included in the MBE, as these mandatory parameters are identified by their position in the decoder capability string.

Following the Profile and Level parameters, zero or more **parameterIdentifier**/**parameterValue** pairs containing the optional CustomMaxMBPS, CustomMaxFS, CustomMaxDPB, and CustomMaxBRandCPB parameters may be included, according to the syntax and semantics given for these parameters. The set of **parameterIdentifier**/**parameterValue** pairs may appear in any order within the capability.

If the H.264 capability MBE contains more than one capability, the second and succeeding capabilities within the MBE message shall be demarcated by a single byte of value zero immediately before the start of each succeeding capability.

NOTE – This zero byte appears in the position where a Parameter ID would otherwise appear. Since the H.264 generic capabilities do not define a parameter with a **parameterIdentifier** value of zero, no confusion results.

Receivers shall ignore the value of any **parameterValue** following an undefined **parameterIdentifier**.

Table 10 below gives an example of an MBE with a single decoder capability indicating Baseline Profile, Level 3.1, with a CustomMaxMBPS parameter of 246 000 macroblocks/second:

**Table 10/H.241 – Baseline profile example MBE**

| MBE | Value | Description |
|---|---|---|
| **Byte 1** | Start-MBE | Start of MBE. From H.230 |
| **Byte 2** | 6 | Number of bytes to follow |
| **Byte 3** | <H.264> | Indicates H.264 MBE. From H.230 |
| **Byte 4** | 64 | Profile parameter – indicates Baseline Profile |
| **Byte 5** | 71 | Level parameter – indicates Level 3.1 |
| **Byte 6** | 3 | Parameter ID – CustomMaxMBPS |
| **Byte 7** | 172 | Lowest 6 bits of 492 (equals 246 000/500), ORed with 128 |
| **Byte 8** | 7 | Remaining 7 bits of 492 |

Table 11 gives an example H.264 capability MBE for a system that supports two capabilities:

- Baseline Profile, Level 2.2; and
- Main Profile, Level 2, with CustomMaxFS supporting 800 × 600 SVGA format and CustomMaxMBPS supporting this format at a rate of 10 frames per second.

**Table 11/H.241 – Two profile example MBE**

| MBE | Value | Description |
|---|---|---|
| **Byte 1** | Start-MBE | Start of MBE. From H.230 |
| **Byte 2** | 10 | Number of bytes to follow |
| **Byte 3** | <H.264> | Indicates H.264 MBE. From H.230 |
| **Byte 4** | 32 | Profile parameter – indicates Main Profile |
| **Byte 5** | 43 | Level parameter – indicates Level 2 |
| **Byte 6** | 4 | Parameter ID – CustomMaxFS |
| **Byte 7** | 8 | Indicates 2048 macroblock frame size (1900 needed for 800 × 600) |
| **Byte 8** | 3 | Parameter ID – CustomMaxMBPS |
| **Byte 9** | 38 | Indicates 19 000 macroblocks/sec processing rate |
| **Byte 10** | 0 | Demarcates start of new capability |
| **Byte 11** | 64 | Profile parameter – indicates Baseline Profile |
| **Byte 12** | 57 | Level parameter – indicates Level 2.2 |

# Annex A

# H.264 transport for H.323

## A.1 Introduction

All details required to implement the H.264 RTP payload format for H.264 are contained in this annex and its references.

Readers should note that this annex is not the complete and primary specification of the RTP payload specification for H.264; please refer to the appropriate IETF RFC for this informative reference. This annex is intended only for use with ITU-T Rec. H.241.

Readers should also note that the terminology used in this annex differs somewhat from that used in the body of ITU-T Rec. H.241 and other ITU-T Recommendations according to Table A.1:

**Table A.1/H.241**

| ITU-T Recommendation term (H.241 and others in the H.323 framework) | Clause A.2/H.241 (RTP payload spec for H.264) Term |
|---|---|
| Annex A | Specification or document |
| may | MAY |
| shall | MUST |
| shall not | MUST NOT |
| should | SHOULD |
| should not | SHOULD NOT |

BCH forward error correction and the byte stream format of Annex B/H.264 shall not be used for H.323 transport.

## A.2 RTP payload format for H.264 video

### A.2.1 RTP header usage

The format of the RTP header is specified in RFC 3550 and is reprinted in Figure A.1 for convenience. This payload format uses the fields of the header in a manner consistent with that specification.

<table>
<tr><td>0</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 colspan="2">V=2</td><td>P</td><td>X</td><td colspan="4">CC</td><td>M</td><td colspan="7">PT</td><td colspan="16">Sequence number</td></tr>
<tr><td colspan="32">Timestamp</td></tr>
<tr><td colspan="32">Synchronization source (SSRC) identifier</td></tr>
<tr><td colspan="32">Contributing source (CSRC) identifiers<br>....</td></tr>
</table>

**Figure A.1/H.241 – RTP header according to RFC 3550**

The RTP header information shall be set as follows:

Version (V): 2 bits

Set to 2 according to RFC 3550.

Padding (P): 1 bit

Used according to RFC 3550.

Extension (X): 1 bit

Specified in the RTP profile in use.

CSRC count (CC): 4 bits

Used according to RFC 3550.

Marker bit (M): 1 bit

Set for the very last packet of the Access Unit indicated by the RTP timestamp, in line with the normal use of the M bit and to allow efficient playout buffer handling. Decoders MAY use this bit as an early indication of the last packet of a coded picture, but MUST not rely on this property because the last packet of the picture may get lost, and because future backward compatible extensions of this payload spec will allow for packet aggregation that does not necessarily preserve the M bit values for all NALUs.

Payload type (PT): 7 bits

The assignment of an RTP payload type for this new packet format is outside the scope of this annex, and will not be specified here. It is expected that the RTP profile under which this payload format is being used will assign a payload type for this encoding or specify that the payload type is to be bound dynamically.

Sequence number (SN): 16 bits

Increased by one for each sent packet. Set to a random value during startup as per RFC 3550.

Timestamp: 32 bits

The RTP timestamp shall be set to the sampling timestamp of the content. If the NALU has no timing properties of its own (for example, parameter set and SEI NAL units), the RTP timestamp shall be set to the RTP timestamp of the primary coded picture that is associated with the same access unit as the NALU according to 7.4.1.2.3/H.264.

Synchronization source (SSRC) identifier: 32 bits

Used according to RFC 3550.

Contributing source (CSRC) identifiers: 0 to 15 items, 32 bits each

Used according to RFC 3550.

### A.2.2 Simple packet

The RTP payload of a Simple Packet according to this specification shall consist of one NALU as depicted in Figure A.2. The type of the NALU MUST be one of those specified in ITU-T Rec. H.264. A NALU stream composed by de-encapsulating Simple Packets in RTP sequence number order MUST conform to the NAL unit decoding order according to 7.4.1.2/H.264.

```
 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
+---------------------------------------------------------------+
|                                                               |
|                             NALU                              |
|                                                               |
|                               +-------------------------------+
|                               |    ... Optional RTP padding   |
+-------------------------------+-------------------------------+
```

**Figure A.2/H.241 – RTP payload format for simple packet**

## A.3 Packetization rules

- VCL NALUs specified according to 7.4.1/H.264 (i.e., NALUs containing a coded slice or a coded slice data partition) belonging to the same picture (and hence share the same RTP timestamp value) MAY be sent in any order permitted by the applicable profile defined in ITU-T Rec. H.264, although, for delay critical systems, they SHOULD be sent in their original coding order to minimize the delay. Note that the coding order is not necessarily the scan order, but the order the NAL packets become available to the RTP stack.
- Packets containing SEI NALUs MAY be sent anytime allowed by ITU-T Rec. H.264.
- Parameter set NALUs MUST NOT be sent in an RTP session whose Parameter Sets were already changed by control protocol messages during the lifetime of the RTP session. If parameter set NALUs are allowed by this condition, they MAY be sent at any time.
- All NALU types MAY be mixed freely, provided that above rules are obeyed. In particular, it is allowed to mix coded slices and coded slice data partitions when allowed by the applicable profile defined in ITU-T Rec. H.264.

## A.4 De-packetization process (non-normative)

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

Since the packetization rules above already enforce a standard-compliant NALU stream when processing received RTP packets in their original order, the most straightforward de-packetization scheme is to re-order the RTP packets according to their sequence numbers and forward the RTP payload to the decoder.

The following additional de-packetization rules MAY be used to implement a more optimized operational JVT de-packetizer:

- Intelligent RTP receivers (e.g., in gateways) MAY identify lost NALUs of type "coded slice data partition A" (DPA). If a lost DPA is found, the associated coded slice data partition B (DPB) and coded slice data partition C (DPC) NALUs are meaningless to the decoder and MAY be discarded. Gateways, for example, MAY decide not to forward the DPB and DPC NALUs in this case, to ease the network load.
- Receivers MAY discard all packets that have a value of nal_ref_idc equal to 0. However, it is preferable to process those packets if possible, because the user experience may suffer if the packets are discarded.

# Appendix I

## ASN.1 OIDs defined in this Recommendation

| OID | Clause reference |
|---|---|
| {itu-t(0) recommendation(0) h(8) 241 specificVideoCodecCapabilities(0) h264(0) iPpacketization(0) h241AnnexA(0)} | 7.1.4 |
| {itu-t(0) recommendation(0) h(8) 241 specificVideoCodecCapabilities(0) h264(0) iPpacketization(0) RFC3984NonInterleaved(1)} | 7.1.4 |
| {itu-t(0) recommendation(0) h(8) 241 specificVideoCodecCapabilities(0) h264(0) iPpacketization(0) RFC3984Interleaved(2)} | 7.1.4 |
| {itu-t(0) recommendation(0) h(8) 241 specificVideoCodecCapabilities(0) h264(0) generic-capabilities(1)} | 8.3.2.1 |

## SERIES OF ITU-T RECOMMENDATIONS

| | |
|---|---|
| Series A | Organization of the work of ITU-T |
| Series D | General tariff principles |
| Series E | Overall network operation, telephone service, service operation and human factors |
| Series F | Non-telephone telecommunication services |
| Series G | Transmission systems and media, digital systems and networks |
| **Series H** | **Audiovisual and multimedia systems** |
| Series I | Integrated services digital network |
| Series J | Cable networks and transmission of television, sound programme and other multimedia signals |
| Series K | Protection against interference |
| Series L | Construction, installation and protection of cables and other elements of outside plant |
| Series M | Telecommunication management, including TMN and network maintenance |
| Series N | Maintenance: international sound programme and television transmission circuits |
| Series O | Specifications of measuring equipment |
| Series P | Telephone transmission quality, telephone installations, local line networks |
| Series Q | Switching and signalling |
| Series R | Telegraph transmission |
| Series S | Telegraph services terminal equipment |
| Series T | Terminals for telematic services |
| Series U | Telegraph switching |
| Series V | Data communication over the telephone network |
| Series X | Data networks, open system communications and security |
| Series Y | Global information infrastructure, Internet protocol aspects and next-generation networks |
| Series Z | Languages and general software aspects for telecommunication systems |

Printed in Switzerland
Geneva, 2006