ITU - Telecommunications Standardization Sector Document Q11-A-31

STUDY GROUP 16 Filename: q11a31.doc

Circuit-switched network multimedia systems and terminals Experts Group

Portland, 24-27 June 1997

Question: Q.11/16

Source: Sen-ching Samson Cheung (sccheung@vtel.com, 1-408-922-5489)

Michael Chen (cpchen@vtel.com, 1-408-922-4701)

Dan Klenke (dank@vtel.com, 1-408-922-5521)

VTEL Inc.

350 E. Plumeria Drive, San Jose, CA 95134, U.S.A.

Title: Support of H.263+ on H.320 (Proposal)

1. Introduction

This contribution proposes modifications to the H.320 suite of standards in support of recent enhancements to the video coding Recommendation, H.263.

H.320 was first issued in 1989. At that time, it specified H.261 as its only video coding algorithm. In 1995, H.320 was revised to include, among other things, two additional video algorithms; H.262 (MPEG-2 Video) and H.263, which was originally developed for the GSTN terminal, H.324. In March of 1997, the ITU-T Study Group 16 determined a revision of H.263 that included a number of annexes defining new optional modes of operation. (These new options are known colloquially as H.263+. For convenience, that terminology will be used here.) This document proposes extensions to the current video capabilities signaling such that H.320 terminals will be able to take advantage of the improved H.263.

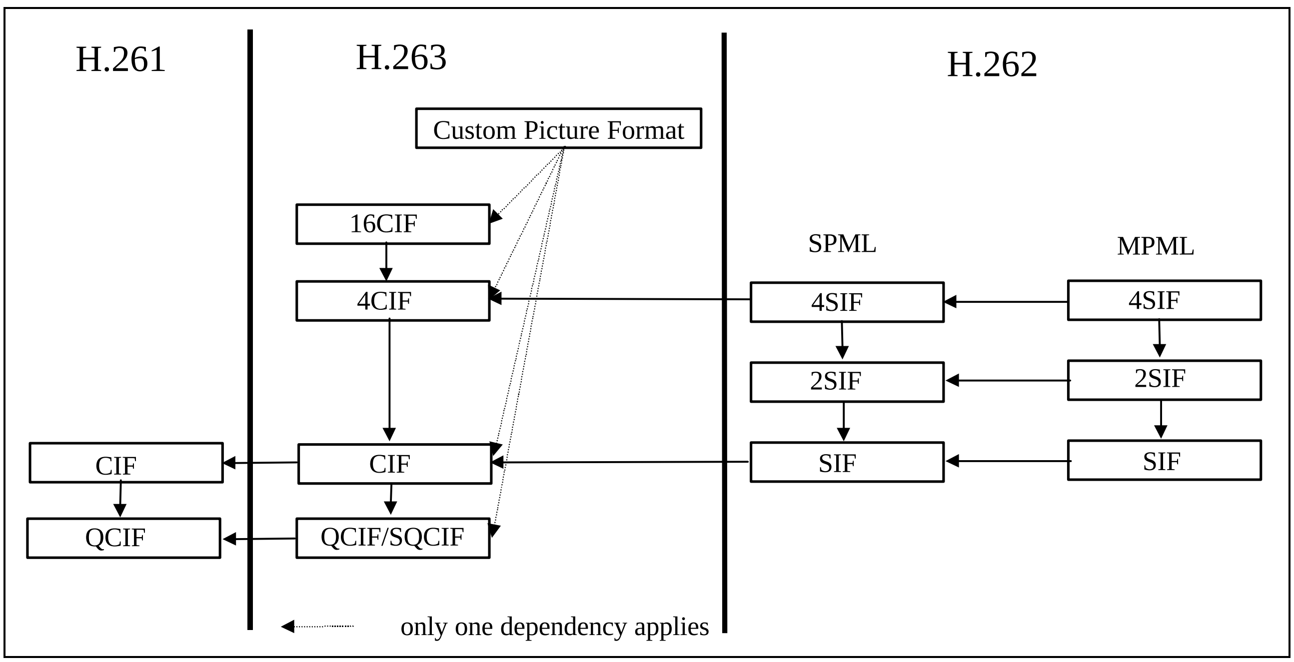
Due to the scope of applications of H.320 terminals and the inherent constraints of ISDN networks, it may not be necessary, or even practical, to support all the new features of H.263+. For example, Chroma Keying (H.263/Annex L), Scalability (H.263/Annex O), and some forms of Reference Picture Selection (H.263/Annex N) require manipulation of multiple channels, a feature which is not easily supported by H.320. A primary goal of this proposal is to adopt as many features as possible without adding too much delay to the capability exchange process. A firm requirement of this proposal is to remain fully compatible with existing H.320 terminals using any of the H.261, H.262, and H.263 Recommendations.

Enabling support of H.263+ options in H.320 terminals will provide the large community of video conferencing users a reasonable path to improved video quality while maintaining the infrastructure (including multipoint connectivity) which has been developed over several years. It will also add to the likelihood of high-quality interoperability between the current generation H.320 systems and the next generation of H.323 and H.324I systems.

2. Enhanced Video Algorithm Hierarchy

The Enhanced Video Algorithm Hierarchy defined in H.320 Annex A[1] is used to provide a guaranteed level of interoperability among all enhanced H.320 video codecs. The new features introduced by H.263+ will be treated as new options requiring only the support of baseline H.263 and H.261. The support of H.262 will not depend on the support of any H.263+ features.

When a particular resolution of a higher level algorithm is supported, the Hierarchy defines an *equivalent resolution* in lower level algorithms which are required to be supported at a higher frame rate (or a smaller MPI). H.263+ introduces Custom Picture Format which can be used to represent almost continuous ranges of resolutions. Following the convention adopted in H.245[2], the Custom Picture Format will be signaled as a range where every resolution in between will be supported (see next Section.) The *equivalent resolution* of a Custom Picture Format will be the largest H.263 format (QCIF, CIF, 4CIF, 16CIF) just smaller than the upper bound of the custom picture format range. For example, if the custom picture format range is [176..528] x [144..432], then the equivalent resolution is CIF (352 x 288). This equivalent resolution needs to be explicitly defined in the H.263 capset (or implicity defined by declaring a higher resolution.) The following diagram depicts the proposed Hierarchy among all the video algorithms.



In addition to picture format, proper hierarchy structure should be defined for Custom Picture Clock Frequency, Pixel Aspect Ratio and Scalability. The details of how these capabilities are signaled can be found in section 3. If Custom Picture Clock Frequency is used, standard PCF (29.97Hz) must also be supported with MPI, measured in seconds, no larger than the MPI declared. For example, if Custom Picture Frequency of 25 Hz is used and the declared MPI is 2, then the codec should also be able to run at MPI of 1 using the standard PCF. For Pixel Aspect Ratio, if non-CIF pixel aspect ratio is supported, CIF pixel aspect ratio must also be supported. These kinds of backward compatibilities do not need to be explicitly defined. Also, these rules are only applicable to standard picture format. The hierarchy structure of scalability will be defined in section 3.

3. Video Capability Set

The capabilities exchange of H.263+ will be embedded in the same MBE message used for the current H.262/H.263[3]. All the specifications of H.263+ capabilities will come after those of H.263 and H.262, starting with the extension code 01111111. All the H.263+ capabilities should be used only to enhance defined H.263 capabilities. No new resolutions should be defined except for Custom Picture Format which will inherit all the capabilities from its equivalent resolution in H.263.

H.263+ introduces twelve new annexes and other features such as custom picture format and picture clock frequency. Some of these features need multiple bits to fully specify them. On the other hand, H.221 only allocates 50 Bytes per second for capability exchange. Obviously, defining all the H.263+ capabilities in series within the MBE message will certainly cause too much delay in many situations. Our approach is to use the leading first byte to indicate the presence of four different feature sets in H.263+. Using the proposed structure in Appendix A, the following describes the four feature sets:

1. Custom Picture Format (2-4 bytes) - The custom picture format is represented as a range with the lower bound smaller than or equal to the upper bound in both dimensions. If the upper and lower bounds are equal, only one frame size needs to be specified. Only eight bits are used to specify each dimension instead of nine bits as in H.263+. Therefore, to maintain roughly the same maximum frame size allowed in H.263+, the actual dimension specified must be multiples of eight.
2. Custom Picture Clock Frequency (2-3 bytes) - It includes
   - custom picture clock frequency,
   - MPI,
   - [optional] HRD-B,
   - [optional] BPPmaxKB.
3. Profiles or Core technology options (0-2 Bytes) - One of the working items for Q.15/SG16 is to define appropriate profiles to simplify the usage of various features offered by H.263+[4]. These profiles will likely contain commonly used options (most of the coding efficiency options) and network specific options (such as Reference Picture Selection Mode). With the risk of being slightly presumptuous, we have grouped the following features of H.263+ into this category:
   - Annex I : Advanced Intra Coding mode,
   - Annex J : Deblocking Filter mode,
   - Annex M : Improved PB-frames mode,
   - Annex N : Reference Picture Selection mode – only videomux mode is allowed.
   - Annex Q : Reduced-Resolution Update mode,
   - Annex R : Independently Segmented Decoding mode,
   - Annex S : Alternative Inter VLC mode,
   - Annex T : Modified Quantization mode.

   The proposed scheme also allows users to bypass the profiles and signal each option separately.
4. Enhancement Options (1-10 bytes) - There are other options in H.263+ which are application specific but may not be used very often. We have called them collectively Enhancement Options. They include the following:
   - Interlaced format support – This option is being proposed in the first Q15/16 meeting[5].
   - Pixel Aspect Ratio – PAR is equal to the Display Aspect Ratio (DAR) multiplied by the ratio between frame height and frame width. In H.263+, 20 bits are needed to fully specify a PAR. In order to reduce the number of bits spent in the capset and based on the fact that most display devices use either square pixel, 4:3 DAR or 16:9 DAR, we used a PAR-DAR mixed representation method : If 4:3 or 16:9 DAR is indicated, the PAR declared will be based on the upper bound of the custom source format range, if available, or the standard source format declared.
   - Annex K : Slice Structured Mode,
   - Annex P : Reference Picture Resampling Mode.

- Annex L : Supplemental Enhancement Information – We have grouped all the functionalities provided in this annex into three different capabilities:
  - *Picture Freezing and Tagging Capability* which includes Full-Picture Freeze Request, Partial-Picture Freeze Request, Resizing Partial-Picture Freeze Request, full-Picture Snapshot Tag, Partial-Picture Snapshot Tag, Video Segment Start Tag and End Tags.
  - *Progressive Refinement* which includes Progressive Refinement Segment Start and End Tags. Unlike the counterpart in H.245, the refinement segment must follow the agreed-upon capability of the rest of the video.
  - *Chroma Keying* is a subset of what can be done in Annex L. Both the background and foreground material must be embedded in the same bitstream. Such a proposal is currently under consideration by Q.15/16[6].
- Annex O : Temporal, SNR, and Spatial Scalability Mode – A complete description of how scalability can be adopted in H.320 network is beyond the scope of this proposal. Instead, we will provide a high level description of what is necessary. Hopefully this will stimulate further work in this area. As H.320 lacks the flexible logical channel structure as in H.323 or H.324, specifying the structure of the layered video poses an unique challenge to H.320 terminals. To simplify the problem, the proposed scheme only allows hierarchical scalability in which decoding any layer will require decoding of all the lower SNR and spatial scalable layers. As temporal layers are not used in decoding any upper layers, their presence is optional. This eliminates the need to specify the dependency graph between different layers.

  It also seems reasonable to enforce all the layers to use the same capsets consistent with their own kind of scalability except for the MPI and the picture format. The latter two capabilities are determined by further restricting the enhancement from one layer to the other to be one of the following:
  - Spatial Scalable Enhancement Layer whose picture format is twice in each dimension of the immediately lower layer with the same MPI.
  - SNR Enhancement Layer with same picture format and MPI as the immediately lower layer.
  - Temporal Enhancement Layer with one or two B-pictures. The picture format remains the same while the MPI is equal to that of the base layer divided by two (for one B-picture) or three (for two B-pictures).

  The source format and MPI specified in the capability exchange refer to the highest layer and each enhancement layer will be specified starting from the lowest. For example, a three-layer scalable bitstream with a base layer of QCIF-15fps, temporal scalable layer with one B-picture in the middle and spatial scalable layer of CIF on top will have a capset like this : {CIF + MPI_1 + Scalable_Layer_2 + temporal_1_Bpicture + spatial}. As there are two scalable layers with a disposable middle temporal layer, there are four possible layer combinations including the base layer. To figure out the capset for each feasible structure, we can start by assuming that the base layer has picture format a and MPI b. The second temporal layer has one B-picture so the MPI becomes $b/2$. The highest layer is a spatial enhancement layer so the picture format becomes $2a$ which is CIF and MPI remains $b/2$ which is 1. Thus, a is QCIF and b is 2. Hence, the capsets for the four possible structures are : {QCIF + MPI_2}, {QCIF + MPI_1 + Scalable_Layer_1 + temporal_1_Bpicture}, {CIF + MPI_2 + Scalable_Layer_1 + spatial} and {CIF + MPI_1 + Scalable_Layer_2 + temporal_1_Bpicture + spatial} . The only remaining parameter is to specify the bandwidth requirement of each layer. There are two methods to represent this. One is to use to one of the allowable ISDN-rates: 64, 128, 192, 256, 320, 384, 768, 1152, 1536, 1920 kbps. The other way is to specify the relative bandwidth with respect to the immediately lower layer. This is useful to carry scalable layers below 64 kbps.

Finally when a scalable structure is defined in the capset, the codec is expected to support non-scalable coding as well. The required non-scalable picture formats are those of the spatial enhancement layers as well as the base layer. As for MPI, the required MPI for each picture format will be that of the spatial layer or the temporal layer immediately higher if it exists. Using the above example, the codec will need to support QCIF at MPI_1 and CIF at MPI_2.

4. Video Commands

The following commands should be added to support new H.263+ functions:

progressiveRefinementStart-doOnePicture
- Start progressive refinement on a still picture.

progressiveRefinementStart-doContinuousPictures
- Start progressive refinement on moving pictures.

progressiveRefinementAbort
- Stop progressive refinement and resume video.

Appendix A: An Example of H.263+ Video Capabilities in H.320

The followings describe a structure for H.263+ capabilities in an MBE message. The total size is between 2 to 21 bytes. The notation *Byte a[b-c]* indicates the field specified in Byte *a* from bit *b* to bit *c*.

Bit Position	Value	Specification
Byte 0 (Mandatory)		
0-7	7Fh	Extension codeword. If the same codeword, not belonging to the structure of the defined capset, is encountered again in the same MBE message, the data following the codeword will be ignored.
Byte 1 (Mandatory)		
0-2		Source Format
	000	H.263_QCIF/SQCIF – enhancements defined for QCIF/SQCIF resolution.
	001	H.263_CIF – enhancements defined for CIF resolution.
	010	H.263_4CIF – enhancements defined for 4CIF resolution.
	011	H.263_16CIF – enhancements defined for 16CIF resolution.
	100	Presence of Custom Picture Format. Two distinct bounds
	101	Presence of Custom Picture Format. Equal bounds.
	110-111	Reserved
3	0/1	Presence of Custom Picture Clock Frequency.
4-6	x	Profiles - 000 is reserved to allow separate signaling of individual options and 111 is reserved for a one-byte profile extension.
7	0/1	Presence of Functional Enhancement Options
Byte 2 Custom Source Format (Present when Byte1[0-2]=100 or Byte1[0-2]=101)		
0-7	x	FrameHeight/8-1 for the upper picture format bound; valid range is [0,143].
Byte 3 Custom Source Format (Present when Byte1[0-2]=100 or Byte1[0-2]=101)		
0-7	x	FrameWidth/8-1 for the upper picture format bound; valid range is [0,255].
Byte 4 Custom Source Format (Present when Byte1[0-2]=100)		
0-7	x	FrameHeight/8-1 for the lower picture format bound; valid range is [Byte3[0-7],143].
Byte 5 Custom Source Format (Present when Byte1[0-2]=100)		
0-7	x	FrameWidth/8-1 for the lower picture format bound; valid range is [Byte4[0-7],255].
Byte 6 (Present when Byte1[3]=1) **Picture Clock Frequency = 1 800 000/ (Clock Conversion Factor * Clock Divisor)**		
0		Clock Conversion Factor
	0	1000
	1	1001
2-7	x	Clock Divisor; 0000 is forbidden.
Byte 7 Picture Clock Frequency (Present when Byte1[3]=1)		
0-5	x	(x+1)/PCF is the MPI; valid range is [0,63].
6	0/1	Specify HRD-B
7	0/1	Specify BPPmaxKB
Byte 8 Picture Clock Frequency (Present when Byte1[3]=1)		
0-3	x	HRD-B as defined in H.242.
4-7	x	BPPmaxKB as defined in H.242.
Byte 9 Profiles / Core Technology Options (Present when Byte1[4-6]=111)		
0-7	x	Profile extension
Byte 9 Profiles / Core Technology Options (Present when Byte1[4-6]=000)		
0	0/1	Presence of next Core Technology option byte

2	0/1	Annex I : Advanced Intra Coding mode
3	0/1	Annex J : Deblocking Filter mode
4	0/1	Annex M : Improved PB-frames mode
5	0/1	Annex Q : Reduced-Resolution Update mode
6	0/1	Annex S : Alternative Inter VLC Mode
7	0/1	Annex T : Modified Quantization Mode
Byte 10 Profiles / Core Technology Options (Presence when Byte1[4-6]=000 and Byte9[0]=1)		
0	0/1	Presence of next core technology option byte
1	0/1	Annex R : Independently Segmented Decoding Mode.
2	0/1	Annex N : Reference Picture Selection Mode -- Only videoMux mode allowed because the lack of separate backchannel.
3-4		meaningful only when Byte10[2]=1
	00	ackMessageOnly
	01	nackMessageOnly
	10	ackOrNackMessageOnly
	11	ackAndNackMessage
5-7	x	meaingful only when Byte10[2]=1; It represents the number of additional picture memory - 1. Valid range is [0-7].
Byte 11 Enhancement Options (Present when Byte1[7]=1)		
0	0/1	Presence of next Enhancement Option byte
1	0/1	Presence of Scalability Capability
2	0/1	Annex K : Slice Structure Mode
3-4		meaningful only when Annex K is used.
	00	Rectangular Slice; Receive/Transmit in order.
	01	Rectangular Slice; Receive/Transmit out of order.
	10	NonRectangular Slice; Receive/Transmit in order.
	11	NonRectangular Slice; Receive/Transmit out of order.
5	0/1	Interlaced Display Support
6-7	00	Square Pixel Aspect Ratio
	01	4:3 Display Aspect Ratio with the PAR defined by the declared standard format or the upper bound in the custom source format range.
	10	16:9 Display Aspect Ratio with the PAR defined by the declared standard format or the upper bound in the custom source format range.
	11	Any Pixel Aspect Ratio.
Byte 12 Enhancement Options (present when Byte11[0]=1)		
0	0/1	Presence of next Enhancement option byte (Fixed at zero for current revision.)
1	0/1	Picture Freeze and Tagging in Annex L , Supplemental Enhancement.
2	0/1	Progressive Refinement in Annex L. The refinement segment must follow the same capability of the video channel.
3	0/1	Chroma Keying in Annex L with both foreground and background material present in the same bitstream.
4	0/1	Annex P : Reference Picture Resampling.
5-7		meaningful only when Annex P is used.
	000	Dynamic Picture Resizing By Four
	001	Dynamic Picture Resizing
	010	Dynamic Warping
	011-111	Reserved
Byte 13 Scalability (Present when Byte11[1]=1)		
0-3	x	Number of Scalable Layers. Valid Range : [1-10]
4-5		Enhancement Layer 1
	00	Spatial Scalable Layer

	01	SNR Scalable Layer
	10	Temporal Scalable with one B-picture
	11	Temporal Scalable with two B-pictures
6-7	x	Enhancement Layer 2 (same definition as Byte13[4-5])
Byte 14 Scalability (Present when Byte11[1]=1)		
0-3	x	Maximum Bandwidth for Enhancement Layer 1 0000 64kbps 1000 1536kbps 0001 128kbps 1001 4 0010 192kbps 1010 2 0011 256kbps 1011 3/2 0100 320kbps 1100 same 0101 384kbps 1101 3/4 0110 768kbps 1110 2/4 0111 1152kbps 1111 1/4
4-7	x	Maximum Bandwidth for Enhancement Layer 2 (same definition as Byte14[0-3])
Byte 15 Scalability (Present when Byte13[0-3] >2)		
0-1	x	Enhancement Layer 3(same definition as Byte13[4-5])
2-3	x	Enhancement Layer 4
4-5	x	Enhancement Layer 5
7-8	x	Enhancement Layer 6
Byte 16 Scalability (Present when Byte13[0-3]>2)		
0-3	x	Maximum Bandwidth for Enhancement Layer 3 (same definition as Byte14[0-3])
4-7	x	Maximum Bandwidth for Enhancement Layer 4
Byte 17 Scalability (Present when Byte13[0-3]>4)		
0-3	x	Maximum Bandwidth for Enhancement Layer 5
4-7	x	Maximum Bandwidth for Enhancement Layer 6
Byte 18 Scalability (Present when Byte13[0-3] >6)		
0-1	x	Enhancement Layer 7 (same definition as Byte13[4-5])
2-3	x	Enhancement Layer 8
4-5	x	Enhancement Layer 9
7-8	x	Enhancement Layer 10
Byte 19 Scalability (Present when Byte13[0-3]>6)		
0-3	x	Maximum Bandwidth for Enhancement Layer 7
4-7	x	Maximum Bandwidth for Enhancement Layer 8
Byte 20 Scalability (Present when Byte13[0-3]>8)		
0-3	x	Maximum Bandwidth for Enhancement Layer 9
4-7	x	Maximum Bandwidth for Enhancement Layer 10

[1] Recommendation H.320. June 1996.
[2] Recommendation H.245, Version 3. March 1997
[3] Recommendation H.242. June 1996.
[4] Gary Sullivan, Tom Geary. Prelimanary Meeting Announcement of the First Meeting of the Experts Groups for SG15 Question 11 & 15. Email to itu-adv-video on 10 May 1997.
[5] Yi-Tong Tse, Q15-A-17: Support of Interlaced Scan Format in H.263+.
[6] Peter List. Email to itu-adv-video on 12 May 1997.