Audio Video Transport Group
Internet Draft
Document: draft-basso-avt-videoconreq-02.txt

A. Basso
NMS Communications
O.Levin
Microsoft
N. Ismail
Cisco Systems

Expires: April 2005
October 2004

Requirements for transport of video control commands

Status of this Memo

Abstract

A variety of video communication services such as video conferencing and video messaging rely on the capability of video encoders and decoders to respond to control commands. This document outlines this set of commands as well as the requirements for their transport.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [1].

Table of Contents

What is new from version 01

1. Document updated to be conformant to Guidelines to Authors of Internet-Drafts
2. Section 5.2.4 RateNotify command removed.
3. Section 7.1 Clarified Reliable versus unreliable delivery.
4. Added Section 7.2 Transport alternatives
5. Section 7.4 Relation with signaling. Removed
6. Section 9.8 Interoperability with other protocols. Removed

7. Section 9.9 MUST has been changed to SHOULD.

8. Updated references

What is new from version 00

1. Added boilerplate text.
2. Sec. 3: Clarification of terminology.
3. Sec. 6 : clarification the reference to IETF protocols only.
4. Harmonization with H.241.

1. Introduction

A variety of video communication services such as video conferencing and video messaging rely on the capability of video encoders and decoders to respond to control commands. This document outlines a generic set of commands applicable to a variety of video codecs as well as the requirements for their transport.

2. Background

RTP [9] is the protocol of choice for the delivery of real time media. RTCP, the companion control protocol, allows some form of monitoring of the media delivery. An enhanced RTCP feedback scheme enabling a generic decoder to provide hints to the corresponding encoder in case of network losses has been described in [7]. Similar solutions were provided for specific coding schemes such ad H.261 [3] H.263 [4] and MPEG-4 [5].
Currently, there is no standard protocol support that allows a given application to exchange control commands with a given codec.

3. Video coding

In current coding schemes such as H.261 [2], H.263 [3], MPEG-1, 2,4 [5], H.264 [6] pictures can be coded with various modalities i.e. intra o predicted pictures. Furthermore pictures can be used as references in the decoding process or not.
More precisely, intra pictures are pictures that can be decoded without first decoding any other picture. Predicted (or non-intra) pictures may require data from one or more previously decoded pictures in order to be decoded. A reference picture is a picture that is stored in the decoder for use as a reference in the decoding process of some subsequent picture in the video bitstream. Finally a non-reference picture is a picture that is not used as a reference for the decoding process of any other picture in the bitstream. The concepts of intra versus non-intra and reference versus non-reference

are independent. A particular picture can in general be any one of the four types, intra reference, non-intra reference, intra non-reference, non-intra non-reference.

Furthermore video pictures are not coded as a whole but are partitioned in small blocks called macrobolocks (MB) and every MB is individually coded. MBs are grouped together in sets of variable size. Such sets are called, in dependence of the coding standard, slices or Group of Blocks (GoBs).Such sets of MB can be scattered in the picture.

4. Use Cases

This section describes use cases of codec control commands.

1. A use case includes an RTP video mixer composing multiple encoded video sources into a single encoded video stream. Each time a video source is to be added to the video composition, the RTP mixer needs to request an encoded reference picture from the video source or a specific area of the picture defined by one or more slices.

2. Another use case includes an RTP video mixer that receives multiple encoded RTP video streams from conference participants and dynamically selects one of the streams to be included in its output RTP stream. For every new video stream selected, the mixer will request a intra picture from the remote source in order for the receiving endpoints to be able to decode and display the output stream smoothly when the switch occurs. The video mixer in this scenario will stop the delivery of the current RTP stream and it will wait for the intra picture from the source before it switches to that source.

3. Another use case includes a given application that needs to signal to the remote encoder a request of change in the coding strategy asking to deliver video pictures at a lower frame rate but with better picture quality or vice versa. Such requests may be based on input from the end user.

4. Another use case includes an application that has became aware of packet losses and in order to mitigate their effect requests an intra picture from the remote encoder. This will stop the spatial and temporal propagation of coding errors inherent to commonly used predictive video coding schemes. It is also possible to obtain random access recovery without a fast update. This is sometimes called "gradual decoder refresh". See for example the recovery point SEI message in H.264/AVC [6].

5. Another use case includes a video mixer that switches its output stream to a new video source. The video mixer will instruct the

receiving endpoints by means of a codec control command to complete the decoding of the current picture and then wait for a new video reference picture. Concurrently, the video mixer requests a reference picture from the new video source and immediately switches to the new source. Once the new source receives the request for the reference picture and acts on it, the receiving endpoints will restart decoding and displaying the new picture.
The main benefit of this method as opposed to the video mixer stopping video transmission of the new source until it detects a new reference picture, as in use case 2, is that the video mixer does not have to discover the beginning of a reference picture. This can simplify the video mixer task especially in the case in which the picture has multiple reference pictures.

6. Another use case includes a video mixer that dynamically selects one of the received video streams to be sent out to participants and tries to provide the highest bit rate possible to all participants while minimizing stream transrating. One way of achieving this is for the mixer to setup sessions with endpoints using the maximum bit rate accepted by that endpoint and by the call admission method used by the mixer.
By means of commands that allow flow control, the mixer can then reduce the maximum bit rate sent by endpoints to the lowest common denominator of all received streams. As the lowest common denominator changes due to endpoints joining or leaving, the mixer can adjust the limits to which endpoints can send their streams to match the new limit.
The mixer then would request a new maximum bit rate, which is equal or less than the maximum bit-rate negotiated at session setup, for a specific media stream, and the remote endpoint can respond with the actual bit-rate that it can support.

5. Codec Commands

The ensemble of commands described in this section is divided into two sets. The first set includes commands that are sent to decoders typically to control the presentation of the content. The second set includes commands that are sent to remote encoders.

5.1 Decoder Control Commands

1. VideoFreezePicture

It instructs the video decoder to complete the decoding of the current video picture and subsequently display it until a timeout period is elapsed or the receipt of a message that indicates the release of the frozen picture and resume normal decoding and

presentation. Note that the freeze picture release command is part of

the H.261, H.262, H.263 and H.264 bitstreams. Coding schemes that support picture freeze release in their bitstreams, MUST use freeze release to signal the remote end to resume decoding.
H.264 specifies a timeout period of at least 6 seconds from the receipt of the VideoFreezePicture. See use case 5 for an example of how such command might be used.

5.2 Encoder Control Commands

1. videoFastUpdatePicture

A "fast update", also known as an "instantaneous decoder refresh", involves sending an intra picture to a decoder and thereafter refraining from using any picture sent prior to that intra picture as a reference for the decoding process of any subsequent picture sent in the stream.
The videoFastUpdatePicture command instructs the video encoder to complete the encoding of the current video picture and to generate a full intra picture at the earliest opportunity. The evaluation of such opportunity includes the current encoder coding strategy and the current available network resources. An H.264 encoder can react to a VideoFastUpdatePicture command with an IDR procedure or a gradual recovery procedure as specified in [10]

Intra pictures, independently from the instant in time when they are encoded, are in general several times larger in size than predicted pictures. Thus in scenarios in which the available bandwidth is small the use of a intra picture implies a delay that is significantly longer than the typical picture duration.

2. VideoTemporalSpatialTradeOff(index)

It instructs the video encoder to change its trade-off between temporal and spatial resolution. Index assumes values from O to 31 to indicate monotonically a desire for higher frame rate.
In general the encoder reaction time may be significantly longer than the typical picture duration.

3. RateRequest(MaxBitrate)

It instructs the far-end encoder to change the maximum bit rate of the given media stream being transmitted. MaxBitRate indicates, in units of 100 bit/s, the new requested maximum bit rate for the associated media stream. The new requested bit rate has to be equal to or less than the bit rate negotiated during session setup.

6. General requirements

6.1 Reuse of Existing Protocols

The codec control messages should be transported using an already existing transport protocol whenever possible. The transport protocol should allow at a minimum the leveraging of its security elements.

6.2 Maintain Existing Protocol Integrity

In meeting the requirement of Section 7, the codec control transport mechanism MUST NOT break existing protocols or cause backward compatibility problems.

6.3 Avoid Duplicating Existing Protocols

The codec control mechanism SHOULD NOT duplicate the functionality of existing IETF protocols. The focus of codec control is new functionality not addressed by existing IETF protocols or extending existing IETF protocols within the structures of the requirement in Section 7. Where an existing IETF protocol can be gracefully extended to support codec control requirements, such extensions are acceptable alternatives for meeting the requirements.

6.4 Efficiency

The codec control transport mechanism SHOULD employ protocol elements known to result in efficient operation. Techniques to be considered include re-use of transport connections across sessions i.e. codec control messages that controls different media sessions may be aggregated on one codec control transport channel and piggybacking of responses on requests in the reverse direction

7. Codec Control Requirements

7.1 Reliable and Unreliable Delivery.

The commands VideoPictureFreeze and VideoTemporalSpatialTradeOff and the commands relative to flow control as RateRequest require a reliable delivery.

The command videoFastUpdatePicture implies a specific modification of the media, which is delivered in an unreliable fashion. Given that the delivery of the media is unreliable, the sender cannot rely on the fact that the request has been safely delivered but needs to assure that the requested modification of the data (i.e., insertion of a reference picture) is received before taking any action. Thus the receiver has always to "observe" the incoming data for the requested change independently of the method of delivery of the videoFastUpdatePicture command. VideoFastUpdatePicture can be thus

delivered over an unreliable channel. If the expected change in the media does not happen the command will be retransmitted.

7.2 Transport alternatives

Commands such VideoTemporalSpatialTradeOff and RateRequest relative to flow control can be interpreted as changes of a given presentation description and potentially carried via existing protocols such SDP. This is not the case of the VideoFastUpdatePicture and VideoPictureFreeze commands.

7.3 Capability description

The capability of codec control for each supported message should be described and negotiated, for example using SDP offer/answer, for both senders and receivers during session setup. The transport protocol used for the delivery of codec control messages should also be specified as of session setup.

7.4 Relation with media session

The delivery channel of the codec control messages must be associated with the media session it controls. Using one codec control channel per media session and associating the two channels during session setup could achieve this purpose. Alternatively one media control channel could be used for multiple media sessions. In this case the controlled media session MUST be identified in each codec control message.

The transport channel of the codec control messages should follow a similar path to that of the media session it controls. Inter-operability with other standards for codec control delivery might cause a deviation from this requirement.

7.5 Bidirectional transport

Messages can be originated from receivers as well as a senders thus the transport mechanism must allow bi-directional exchange of messages.

7.6 Extensibility

Codec control message syntax should be extensible to easily support the addition of new control messages.

7.7 Unicast and Multicast Support

The codec control transport MUST work and scale for media sessions that use point-to-point unicast.

The codec control transport MUST work and scale for media sessions that use SSM (Source Specific Multicast) and has a small to moderate group size.

The codec control transport will not address ASM (Any Source Multicast) media sessions in which media sources are not known until they start transmission.

7.8 Timely delivery

For some video services the ability to transmit codec control commands in a timely fashion is essential to the delivery of a high quality user experience. The delay introduced by the transport protocol SHOULD be negligible with respect of the time constants of the delivered media stream.

8. Security Considerations

<TODO>

9. IANA Considerations

<TODO>

10. Acknowledgments

The authors would like to acknowledge the comments from around the Community in helping refine this document. Particular recognition goes to Roni Evens.

11. Copyright Notice

12. Informative References

1 S. Bradner, "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997

2 S. Bradner "IETF Rights in Contributions" RFC 3667 February 2004

3 ITU-T Recommendation H.261 (1993), Video codec for audiovisual services at p . 64 kbit/s.

4 ITU-T Recommendation H.263 (1998), Video coding for low bit rate communication.

5 ISO/IEC 14496-2:2001/Amd.1:2002, "Information technology - Coding of audio-visual objects - Part2: Visual", 2001.

6 Joint Video Team of ITU-T and ISO/IEC JTC 1, Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC), Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G050, March 2003.

7 J. Ott et al. Extended RTP Profile for RTCP-based Feedback (RTP/AVPF), draft-ietf-avt-rtcp-feedback-09.txt, August 2004, IETF Draft. Work in progress.

8 T. Turletti and C. Huitema, "RTP Payload Format for H.261 Video Streams, RFC 2032, October 1996.

9 H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-time Applications", RFC3550 STD 64 July 2003.

10 ITU-T Recommendation H.241 (07/2003), Extended video procedures and control signals for H.300-series terminals.

11 S. Bradner "Intellectual Property rights in IETF Technology" RFC3668, February 2004

13. IPR Notices

The IETF takes no position regarding the validity or scope of any

Author's Addresses

Andrea Basso
NMS Communications
200 Shultz Drive
Red Bank, NJ 07701 USA
Phone: (732) 936-2118
Email: andrea_basso@nmss.com

Orit Levin
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052 USA
EMail: oritl@microsoft.com

Nermeen Ismail
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706, USA
Phone: +1 408 853 8714
Email: nismail@cisco.com