Network Working Group
Request for Comments: 2032
Category: Standards Track

T. Turletti
MIT
C. Huitema
Bellcore
October 1996

RTP Payload Format for H.261 Video Streams

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Table of Contents

1. Abstract

This memo describes a scheme to packetize an H.261 video stream for transport using the Real-time Transport Protocol, RTP, with any of the underlying protocols that carry RTP.

This specification is a product of the Audio/Video Transport working group within the Internet Engineering Task Force. Comments are solicited and should be addressed to the working group's mailing list at rem-conf@es.net and/or the authors.

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

2. Purpose of this document

The ITU-T recommendation H.261 [6] specifies the encodings used by ITU-T compliant video-conference codecs. Although these encodings were originally specified for fixed data rate ISDN circuits, experiments [3],[8] have shown that they can also be used over packet-switched networks such as the Internet.

The purpose of this memo is to specify the RTP payload format for encapsulating H.261 video streams in RTP [1].

3. Structure of the packet stream

3.1. Overview of the ITU-T recommendation H.261

The H.261 coding is organized as a hierarchy of groupings. The video stream is composed of a sequence of images, or frames, which are themselves organized as a set of Groups of Blocks (GOB). Note that H.261 "pictures" are referred as "frames" in this document. Each GOB holds a set of 3 lines of 11 macro blocks (MB). Each MB carries information on a group of 16x16 pixels: luminance information is specified for 4 blocks of 8x8 pixels, while chrominance information is given by two "red" and "blue" color difference components at a resolution of only 8x8 pixels. These components and the codes representing their sampled values are as defined in the ITU-R Recommendation 601 [7].

This grouping is used to specify information at each level of the hierarchy:

- At the frame level, one specifies information such as the delay from the previous frame, the image format, and various indicators.

- At the GOB level, one specifies the GOB number and the default quantifier that will be used for the MBs.

- At the MB level, one specifies which blocks are present and which did not change, and optionally a quantifier and motion vectors.

Blocks which have changed are encoded by computing the discrete cosine transform (DCT) of their coefficients, which are then quantized and Huffman encoded (Variable Length Codes).

The H.261 Huffman encoding includes a special "GOB start" pattern, composed of 15 zeroes followed by a single 1, that cannot be imitated by any other code words. This pattern is included at the beginning of

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

each GOB header (and also at the beginning of each frame header) to mark the separation between two GOBs, and is in fact used as an indicator that the current GOB is terminated. The encoding also includes a stuffing pattern, composed of seven zeroes followed by four ones; that stuffing pattern can only be entered between the encoding of MBs, or just before the GOB separator.

3.2. Considerations for packetization

H.261 codecs designed for operation over ISDN circuits produce a bit stream composed of several levels of encoding specified by H.261 and companion recommendations. The bits resulting from the Huffman encoding are arranged in 512-bit frames, containing 2 bits of synchronization, 492 bits of data and 18 bits of error correcting code. The 512-bit frames are then interlaced with an audio stream and transmitted over px64 kbps circuits according to specification H.221 [5].

When transmitting over the Internet, we will directly consider the output of the Huffman encoding. All the bits produced by the Huffman encoding stage will be included in the packet. We will not carry the 512-bit frames, as protection against bit errors can be obtained by other means. Similarly, we will not attempt to multiplex audio and video signals in the same packets, as UDP and RTP provide a much more efficient way to achieve multiplexing.

Directly transmitting the result of the Huffman encoding over an unreliable stream of UDP datagrams would, however, have poor error resistance characteristics. The result of the hierachical structure of H.261 bit stream is that one needs to receive the information present in the frame header to decode the GOBs, as well as the information present in the GOB header to decode the MBs. Without precautions, this would mean that one has to receive all the packets that carry an image in order to properly decode its components.

If each image could be carried in a single packet, this requirement would not create a problem. However, a video image or even one GOB by itself can sometimes be too large to fit in a single packet. Therefore, the MB is taken as the unit of fragmentation. Packets must start and end on a MB boundary, i.e. a MB cannot be split across multiple packets. Multiple MBs may be carried in a single packet when they will fit within the maximal packet size allowed. This practice is recommended to reduce the packet send rate and packet overhead.

To allow each packet to be processed independently for efficient resynchronization in the presence of packet losses, some state information from the frame header and GOB header is carried with each

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

packet to allow the MBs in that packet to be decoded. This state information includes the GOB number in effect at the start of the packet, the macroblock address predictor (i.e. the last MBA encoded in the previous packet), the quantizer value in effect prior to the start of this packet (GQUANT, MQUANT or zero in case of a beginning of GOB) and the reference motion vector data (MVD) for computing the true MVDs contained within this packet. The bit stream cannot be fragmented between a GOB header and MB 1 of that GOB.

Moreover, since the compressed MB may not fill an integer number of octets, the data header contains two three-bit integers, SBIT and EBIT, to indicate the number of unused bits in the first and last octets of the H.261 data, respectively.

4. Specification of the packetization scheme

4.1. Usage of RTP

The H.261 information is carried as payload data within the RTP protocol. The following fields of the RTP header are specified:

- The payload type should specify H.261 payload format (see the companion RTP profile document RFC 1890).

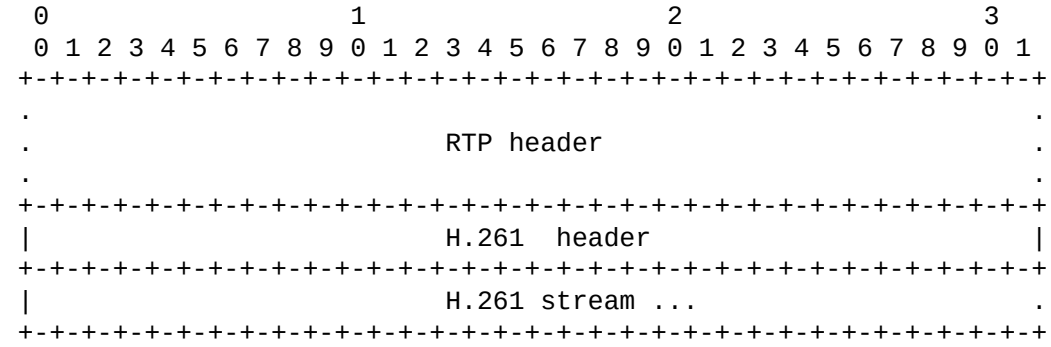
- The RTP timestamp encodes the sampling instant of the first video image contained in the RTP data packet. If a video image occupies more than one packet, the timestamp will be the same on all of those packets. Packets from different video images must have different timestamps so that frames may be distinguished by the timestamp. For H.261 video streams, the RTP timestamp is based on a 90kHz clock. This clock rate is a multiple of the natural H.261 frame rate (i.e. 30000/1001 or approx. 29.97 Hz). That way, for each frame time, the clock is just incremented by the multiple and this removes inaccuracy in calculating the timestamp. Furthermore, the initial value of the timestamp is random (unpredictable) to make known-plaintext attacks on encryption more difficult, see RTP [1]. Note that if multiple frames are encoded in a packet (e.g. when there are very little changes between two images), it is necessary to calculate display times for the frames after the first using the timing information in the H.261 frame header. This is required because the RTP timestamp only gives the display time of the first frame in the packet.

- The marker bit of the RTP header is set to one in the last packet of a video frame, and otherwise, must be

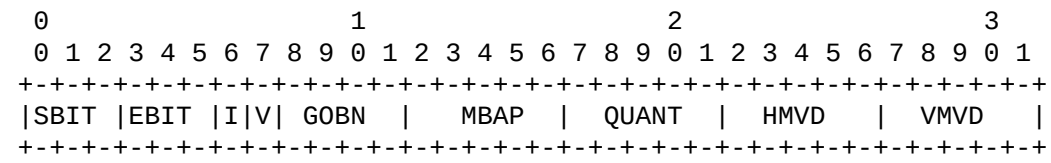
DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

zero. Thus, it is not necessary to wait for a following packet (which contains the start code that terminates the current frame) to detect that a new frame should be displayed.

The H.261 data will follow the RTP header, as in:



The H.261 header is defined as following:



The fields in the H.261 header have the following meanings:

Start bit position (SBIT): 3 bits
Number of most significant bits that should be ignored in the first data octet.

End bit position (EBIT): 3 bits
Number of least significant bits that should be ignored in the last data octet.

INTRA-frame encoded data (I): 1 bit
Set to 1 if this stream contains only INTRA-frame coded blocks. Set to 0 if this stream may or may not contain INTRA-frame coded blocks. The sense of this bit may not change during the course of the RTP session.

Motion Vector flag (V): 1 bit
Set to 0 if motion vectors are not used in this stream. Set to 1 if motion vectors may or may not be used in this stream. The sense of this bit may not change during the course of the session.

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.