Network Working Group
Request for Comments: 3016
Category: Standards Track

Y. Kikuchi
Toshiba
T. Nomura
NEC
S. Fukunaga
Oki
Y. Matsui
Matsushita
H. Kimata
NTT

November 2000

RTP Payload Format for MPEG-4 Audio/Visual Streams

Status of this Memo

Copyright Notice

Abstract

This document describes Real-Time Transport Protocol (RTP) payload formats for carrying each of MPEG-4 Audio and MPEG-4 Visual bitstreams without using MPEG-4 Systems. For the purpose of directly mapping MPEG-4 Audio/Visual bitstreams onto RTP packets, it provides specifications for the use of RTP header fields and also specifies fragmentation rules. It also provides specifications for Multipurpose Internet Mail Extensions (MIME) type registrations and the use of Session Description Protocol (SDP).

1. Introduction

The RTP payload formats described in this document specify how MPEG-4 Audio [3][5] and MPEG-4 Visual streams [2][4] are to be fragmented and mapped directly onto RTP packets.

These RTP payload formats enable transport of MPEG-4 Audio/Visual streams without using the synchronization and stream management functionality of MPEG-4 Systems [6]. Such RTP payload formats will be used in systems that have intrinsic stream management functionality and thus require no such functionality from MPEG-4 Systems. H.323 terminals are an example of such systems, where MPEG-4 Audio/Visual streams are not managed by MPEG-4 Systems Object Descriptors but by H.245. The streams are directly mapped onto RTP packets without using MPEG-4 Systems Sync Layer. Other examples are SIP and RTSP where MIME and SDP are used. MIME types and SDP usages of the RTP payload formats described in this document are defined to directly specify the attribute of Audio/Visual streams (e.g., media type, packetization format and codec configuration) without using MPEG-4 Systems. The obvious benefit is that these MPEG-4 Audio/Visual RTP payload formats can be handled in an unified way together with those formats defined for non-MPEG-4 codecs. The disadvantage is that interoperability with environments using MPEG-4 Systems may be difficult, other payload formats may be better suited to those applications.

The semantics of RTP headers in such cases need to be clearly defined, including the association with MPEG-4 Audio/Visual data elements. In addition, it is beneficial to define the fragmentation rules of RTP packets for MPEG-4 Video streams so as to enhance error resiliency by utilizing the error resilience tools provided inside the MPEG-4 Video stream.

1.1 MPEG-4 Visual RTP payload format

MPEG-4 Visual is a visual coding standard with many new features: high coding efficiency; high error resiliency; multiple, arbitrary shape object-based coding; etc. [2]. It covers a wide range of bitrates from scores of Kbps to several Mbps. It also covers a wide variety of networks, ranging from those guaranteed to be almost error-free to mobile networks with high error rates.

With respect to the fragmentation rules for an MPEG-4 Visual bitstream defined in this document, since MPEG-4 Visual is used for a wide variety of networks, it is desirable not to apply too much restriction on fragmentation, and a fragmentation rule such as "a single video packet shall always be mapped on a single RTP packet" may be inappropriate. On the other hand, careless, media unaware fragmentation may cause degradation in error resiliency and bandwidth efficiency. The fragmentation rules described in this document are flexible but manage to define the minimum rules for preventing meaningless fragmentation while utilizing the error resilience functionalities of MPEG-4 Visual.

The fragmentation rule recommends not to map more than one VOP in an RTP packet so that the RTP timestamp uniquely indicates the VOP time framing. On the other hand, MPEG-4 video may generate VOPs of very small size, in cases with an empty VOP (vop_coded=0) containing only

VOP header or an arbitrary shaped VOP with a small number of coding blocks. To reduce the overhead for such cases, the fragmentation rule permits concatenating multiple VOPs in an RTP packet. (See fragmentation rule (4) in section 3.2 and marker bit and timestamp in section 3.1.)

While the additional media specific RTP header defined for such video coding tools as H.261 or MPEG-1/2 is effective in helping to recover picture headers corrupted by packet losses, MPEG-4 Visual has already error resilience functionalities for recovering corrupt headers, and these can be used on RTP/IP networks as well as on other networks (H.223/mobile, MPEG-2/TS, etc.). Therefore, no extra RTP header fields are defined in this MPEG-4 Visual RTP payload format.

1.2 MPEG-4 Audio RTP payload format

MPEG-4 Audio is a new kind of audio standard that integrates many different types of audio coding tools. Low-overhead MPEG-4 Audio Transport Multiplex (LATM) manages the sequences of audio data with relatively small overhead. In audio-only applications, then, it is desirable for LATM-based MPEG-4 Audio bitstreams to be directly mapped onto the RTP packets without using MPEG-4 Systems.

While LATM has several multiplexing features as follows;

- Carrying configuration information with audio data,
- Concatenation of multiple audio frames in one audio stream,
- Multiplexing multiple objects (programs),
- Multiplexing scalable layers,

in RTP transmission there is no need for the last two features. Therefore, these two features MUST NOT be used in applications based on RTP packetization specified by this document. Since LATM has been developed for only natural audio coding tools, i.e., not for synthesis tools, it seems difficult to transmit Structured Audio (SA) data and Text to Speech Interface (TTSI) data by LATM. Therefore, SA data and TTSI data MUST NOT be transported by the RTP packetization in this document.

For transmission of scalable streams, audio data of each layer SHOULD be packetized onto different RTP packets allowing for the different layers to be treated differently at the IP level, for example via some means of differentiated service. On the other hand, all configuration data of the scalable streams are contained in one LATM configuration data "StreamMuxConfig" and every scalable layer shares the StreamMuxConfig. The mapping between each layer and its configuration data is achieved by LATM header information attached to the audio data. In order to indicate the dependency information of the scalable streams, a restriction is applied to the dynamic assignment rule of payload type (PT) values (see section 4.2).

For MPEG-4 Audio coding tools, as is true for other audio coders, if the payload is a single audio frame, packet loss will not impair the decodability of adjacent packets. Therefore, the additional media specific header for recovering errors will not be required for MPEG-4 Audio. Existing RTP protection mechanisms, such as Generic Forward Error Correction (RFC 2733) and Redundant Audio Data (RFC 2198), MAY be applied to improve error resiliency.

2. 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 [7].

3. RTP Packetization of MPEG-4 Visual bitstream

This section specifies RTP packetization rules for MPEG-4 Visual content. An MPEG-4 Visual bitstream is mapped directly onto RTP packets without the addition of extra header fields or any removal of Visual syntax elements. The Combined Configuration/Elementary stream mode MUST be used so that configuration information will be carried to the same RTP port as the elementary stream. (see 6.2.1 "Start codes" of ISO/IEC 14496-2 [2][9][4]) The configuration information MAY additionally be specified by some out-of-band means. If needed for an H.323 terminal, H.245 codepoint "decoderConfigurationInformation" MUST be used for this purpose. If needed by systems using MIME content type and SDP parameters, e.g., SIP and RTSP, the optional parameter "config" MUST be used to specify the configuration information (see 5.1 and 5.2).

When the short video header mode is used, the RTP payload format for H.263 SHOULD be used (the format defined in RFC 2429 is RECOMMENDED, but the RFC 2190 format MAY be used for compatibility with older implementations).

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|X|  CC   |M|     PT      |       sequence number         | RTP
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           timestamp                           | Header
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           synchronization source (SSRC) identifier            |
+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
|            contributing source (CSRC) identifiers             |
|                             ....                              |
+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
|                                                               | RTP
|       MPEG-4 Visual stream (byte aligned)                     | Pay-
|                                                               | load
|                               +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                               :...OPTIONAL RTP padding        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1 - An RTP packet for MPEG-4 Visual stream

3.1 Use of RTP header fields for MPEG-4 Visual

Payload Type (PT): The assignment of an RTP payload type for this new packet format is outside the scope of this document, and will not be specified here. It is expected that the RTP profile for a particular class of applications will assign a payload type for this encoding, or if that is not done then a payload type in the dynamic range SHALL be chosen by means of an out of band signaling protocol (e.g., H.245, SIP, etc).

Extension (X) bit: Defined by the RTP profile used.

Sequence Number: Incremented by one for each RTP data packet sent, starting, for security reasons, with a random initial value.

Marker (M) bit: The marker bit is set to one to indicate the last RTP packet (or only RTP packet) of a VOP. When multiple VOPs are carried in the same RTP packet, the marker bit is set to one.

Timestamp: The timestamp indicates the sampling instance of the VOP contained in the RTP packet. A constant offset, which is random, is added for security reasons.

- When multiple VOPs are carried in the same RTP packet, the timestamp indicates the earliest of the VOP times within the VOPs carried in the RTP packet. Timestamp information of the rest of

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.

WHAT WILL YOU BUILD? | sales@docketalarm.com | 1-866-77-FASTCASE

fastcase® Smarter legal research.