RTP Retransmission Payload Format

1 Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt
The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

2 Abstract

This document describes new RTP payload formats to enable multiple and optional selective retransmissions in RTP. They are especially applicable to environments where low delay RTCP feedback is available, such as described in [4]. These payload formats can be used to separate the media stream according to prioritization of packets or according to the status of the transmission (i.e., transmission or retransmission). This serves as a retransmission framework, where we give examples how this can be used for different retransmission schemes.

Table of Contents

1 Status of this Memo 1
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].

Introduction to the problem

The Real-Time Transport Protocol (RTP) [3] was designed as an Internet protocol to transmit real-time (or near real-time) data over multicast or unicast. Even though the multicast capability is a strong feature of RTP, a common use of it is non-interactive unicast transmission of media streams. This application, further referred to as media streaming, has lower delay requirements. A constant offset delay, in which already received packets are stored at the client, is tolerable.

Typically RTP is run over the User Datagram Protocol (UDP). While the unreliable UDP performs good for reliable channels, the quality of the received media stream is bad, when transmitted over error-prone links. Especially compressed streams, such as MPEG video-data, are highly susceptible to packet loss, which will lead to a bad video quality at the client. E.g. in a wireless environment with an unstable transmission quality (bit error rate (BER) about 10E-3..10E-6) a sufficient quality of video and voice transmitted by RTP cannot be achieved.

The scenario for such a media streaming over an error-prone link might be the one in Figure 1. Media files are stored on a media server which is connected via an internet link to a mobile network.
To achieve a better quality of the received media stream, especially when transmitted over error-prone links, retransmissions of lost packets of the stream seems to be a possible solution. However the quite low delay requirements of media streaming applications and the bandwidth limitations of the wireless link have to be considered. In wireless systems bandwidth is a scarce resource and efficient use of this resource is of vital importance. Hence reliability for the complete media stream might not be achievable.

Furthermore in compressed media streams, e.g. MPEG video stream, not every frame is equally important. In this example some frames contain data that other data depends upon, which makes these frames more important. Current protocols applying retransmission for reliable transmission (e.g. TCP) make no use of this characteristic of compressed media streams.

Therefore a retransmission scheme that retransmits selected packets, but if necessary these packets several times seems to be the best solution and can substantially enhance the quality of the received media stream.

In the next section a framework is outlined which is based on the payload formats defined in sections 6 and 7. This framework enables us to do multiple selective retransmissions as we describe in sections 10 and 11.

Outline of the solution

Much work has been done recently to make RTP more reliable. As one result a new feedback timing behavior for unicast sessions or small multicast groups is proposed [4], which includes a framework of how to use this feedback. This document include the ability to indicate the detection of packet loss from the receiver to the sender. With this functionality it is possible for the server to do retransmissions for lost packets. However this scheme would be very general and simple and lacks the possibility to do selective or multiple retransmissions effectively.
The payload formats presented in this document are meant as a general solution to enable retransmissions in RTP. Based on this scheme we present two examples how the payload formats could be used to realize first a retransmissions scheme which concentrates on bandwidth efficiency within the media stream and second a somehow more complex retransmission scheme for special applications that indicates the priority of the transmitted media in-band. It is described in detail in section 11.

The payload formats are not meant to replace a media RTP payload format, but are used additionally. This means that e.g. an RTP retransmission packet could be composed as follows:

```
+-----------------------------------------------------+
|                                                     |
|              RTP header with PT=RTX/LP                |
|                                                     |
+-----------------------------------------------------+
|                                                     |
|RTX/SEL Payload Type specific header with PT=media_pt|
|                                                     |
|                                                     |
+-----------------------------------------------------+
|                                                     |
|                     Media Payload                     |
|                                                     |
```

The sender of RTP packets which wants to use the retransmission capabilities has different possibilities at transmitting RTP packets. For packets that are transmitted for the first time, the sender MAY use the Selective (SEL) payload format. If so, it has the possibility to signal the packet’s priority and the sequence number of the last high priority packet in band. This can be used at the receiver of the packets to detect if the lost packet was high or low priority. Low priority packets are usually not subject to retransmissions and thus no retransmission request has to be sent for low priority packets.

Retransmissions SHOULD be transmitted with the Retransmission (RTX) payload format, because the RTX payload format specific header in combination with the usual RTP header contains all information that is needed at the receiver to detect lost retransmissions and reconstruct the original packet (e.g. original sequence number or timestamp).

6 Retransmission Payload Format

Retransmissions in RTP SHOULD be sent with the RTX payload format. This enables multiple retransmissions. The syntax of the header is as follows:

```
0 1 2 3 4 5 6 7
+---+---+---+---+---+---+---+---+
```
Extended Delta Sequence Number (X): 1 bit
This bit indicates the used length of the Delta RTP Sequence Number. For X=0 the least significant 8 bits of the Delta RTP Sequence Number are transmitted, for X=1 24 bits are transmitted.

Payload Type of original packet: 7 bit
This field contains the original payload type information, i.e. the payload type of the payload that is transported in this packet.

(LSB) Delta RTP Sequence Number: 8 bit
This field contains the difference of this packet’s RTP sequence number to the RTP sequence number of the original transmission of this packet’s payload, calculated modulo 2^32, i.e. Delta SN = (SN(this) − SN(original)) mod 2^32. If the number of bits that are needed to encode the difference exceeds 8, X=1 MUST be set and the 16 most significant bits of the 24 bit Delta RTP Sequence Number are transmitted in the next field.

MSB Delta RTP Sequence Number: 16 bit
For X=1 this field carries the 16 most significant bits of the Delta RTP Sequence Number (see also "(LSB) Delta RTP Sequence Number").

RTP Retransmission Payload Format

Selective Payload Format

In certain scenarios, e.g. environments with a very low bit rate feedback channel, it is necessary to limit the feedback that is sent. This can be done by the means of this selective (SEL) payload type. If packets of a low priority are lost, the receiver does not need to send feedback.

Still these packets are needed to detect the loss of high priority packets, and therefore they contain the sequence number of the last important packet that was sent. If the receiver gets a low priority packet, it has only to verify if it has received the indicated last important packet. Only for these packets retransmission requests should be sent. The syntax of the packet is as follows:

```
0 1 2 3 4 5 6 7
+---------------+---------------+---------------+---------------+---------------+---------------+---------------+---------------+
|DTI| PT of original packet |
+---------------+---------------+---------------+---------------+---------------+---------------+---------------+---------------+
|PRI| Delta SN |
+---------------+---------------+---------------+---------------+---------------+---------------+---------------+---------------+
```

July 2001
DTI (Delta Time Indicator): 1 bit
DTI=1 indicates that the delta time field is present (i.e. the payload format header has a size of 4 bytes. DTI=0 means that the Delta Time field is not used and thus the header has a size of 2 bytes.

Payload Type of original packet: 7 bit
This field contains the original payload type information, i.e. the payload type of the payload that is transported in this packet.

PRI (Priority): 1 bit
This field indicates the priority of the packet. PRI=1 means a high priority packet, where PRI=0 is sent for low priority packets.

Delta SN (Delta Sequence Number): 7 bit
The offset in sequence number space to the previous important packet. It is calculated as follows: DSN = SN_this - SN_important.

Delta Time: 15 bit
This field indicates the difference between the time stamp of the last important RTP packet (that should be retransmitted if lost) and that of the current RTP packet. It can be used by the receiver to compute the timestamp of a lost important packet.

RTP Retransmission Payload Format                         July 2001

The default unit of this field is [ms] and by default the following table is used to encode the Delta Time values.

<table>
<thead>
<tr>
<th>Delta Time [ms]</th>
<th>bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>more than 32766</td>
<td>0111 1111 1111 1111</td>
</tr>
<tr>
<td>32766</td>
<td>0111 1111 1111 1110</td>
</tr>
<tr>
<td>32765</td>
<td>0111 1111 1111 1101</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>2</td>
<td>0000 0000 0000 0010</td>
</tr>
<tr>
<td>1</td>
<td>0000 0000 0000 0001</td>
</tr>
<tr>
<td>0</td>
<td>0000 0000 0000 0000</td>
</tr>
<tr>
<td>-1</td>
<td>1111 1111 1111 1111</td>
</tr>
<tr>
<td>-2</td>
<td>1111 1111 1111 1110</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>-32766</td>
<td>1000 0000 0000 0010</td>
</tr>
<tr>
<td>-32767</td>
<td>1000 0000 0000 0001</td>
</tr>
<tr>
<td>less than -32767</td>
<td>1000 0000 0000 0000</td>
</tr>
</tbody>
</table>

It is possible to define other units or coding tables, however, how to specify and negotiate this is outside the scope of this document.

8 Feedback

Feedback is an important part of any retransmission scheme. To be
efficient, also under near real-time conditions, it might be needed to receive this feedback with the lowest delay that is possible. The basic RTP set constraints about feedback timing rules, which might not be suitable for a retransmission scheme under real-time or near real-time conditions.

To enable low delay feedback a new extended RTP profile was defined, which redefines these timing rules according to the group size of the multicast session. In unicast sessions it is possible to send feedback virtually immediately. Hence this profile provides the means to do enable low delay feedback which is needed for the retransmission schemes.

In the Low Delay Feedback profile RTCP feedback messages are defined, which are suited for retransmission requests or acknowledgements. Generally a simple feedback message, called general NACK is suitable to be used as a retransmission request. It is defined in [4] as follows:

<table>
<thead>
<tr>
<th>V=2</th>
<th>P</th>
<th>E</th>
<th>FMT=1</th>
<th>PT=RTPFB</th>
<th>length</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSRC of packet sender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSRC of media source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PID</td>
<td>BLP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The PID is used to indicate a lost packet for which the retransmission is requested and the BLP might be used to indicate several lost packets in one NACK message. A detailed description of this and other low delay feedback messages can be found in [4].

Concentration Control

In general retransmission schemes contradict congestion control. If packets are lost, which is generally accepted to be an indication of congestion, the sender increases the network load by sending additional retransmissions. While this behavior cannot be accepted in best effort environments such as the Internet, it is reasonable in quality of service (QoS) controlled networks, e.g by reserving a certain bit rate for the RTP session. However the used bit rate MUST be monitored to control that the maximum rate is by no means exceeded.

Even though retransmission algorithms normally contradict congestion control they could enhance the quality of service, in environments where losses occur not only due to congestion but also (or mainly)
due to bit errors on the link.

Taking this information into account the following rule MUST be followed, if the retransmissions are used in an best effort environment (e.g. the Internet).

At reception of a retransmission request or detecting packet loss by other means, the sender MUST NOT increase the sending bit rate on mid term timescales. Retransmissions MAY be sent within the current bit rate, e.g. by skipping or delaying the transmission of other scheduled packets.


10 A Multiple Retransmission Algorithm

This section gives an examples how the RTX payload format is used to enable multiple retransmissions in RTP.

RTP Retransmission Payload Format                         July 2001

The sender transmits the media data in normal RTP packets using an RTP payload format for that media. If packets get lost the receiver will detect that by the gap in the sequence number. To signal lost packets the NACK message, described in section 8 is used. The PID field is used to specify a lost packet. The default method, which is to use the RTP sequence number to indicate the lost packet, is used. The BLP field can be used to indicate several lost packets in one NACK message. The detailed semantic is described in [4].

At receiving a retransmission request the sender decides, according to the priority of the packet, the network conditions (see section 9) and maybe other factors, if the packet is to be retransmitted. In case of a positive decision, it uses the RTX payload format (see section 6) to retransmit the packet. The sequence number of the RTP packet is increased by one over the previous packet sent. This enables the receiver to detect also a lost retransmission.

Example of a retransmission packet:

```plaintext
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|V=2|P|X| CC=0  |    PT=RTX     |      RTP sequence number      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        RTP timestamp = timestamp of original transmission                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         SSRC of media source          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|   PT=???    |     DSN=12    |                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+                               +
:                                                               :
:                         Media Payload                         :
:                                                               :
```

11 A Selective Retransmission Algorithm
A special scenario, where the previously described multiple retransmission algorithm does not work well, is if the feedback channel has a very limited capacity. Possible scenarios are small multicast groups, where it is according to the defined timing rules for RTCP feedback seldom possible to send immediate feedback, or the feedback channel is mainly used for other purposes (e.g. wireless control channel) and has a very low bit rate available for RTCP feedback. In this scenarios it is very important to limit the feedback to the very minimum, which would mean that feedback is only sent in form of retransmission requests for really important data.

RTP Retransmission Payload Format

To solve the above described requirements on the retransmission scheme, it is necessary that the receiver knows the priority of the lost packets, to decide if a retransmission request should be sent or not. To do so at least two solutions apply: First the important and non-important packets are sent in different streams, using a different sequence number space. Second the selective packet format is used.

There are scenarios where the first solution is applicable and thus should be used, because it uses less overhead. However there are drawbacks with this solution. Given that only very few high priority packets are sent, e.g. once every 2 seconds, the time between loosing a packet and detection of this packet loss is at least this inter-packet time.

Hence in scenarios as described above, where feedback can be sent only very infrequently, and thus only really important packets are retransmitted, a solution which uses the selective payload format might be the only possible solution at all. The mechanism works as described in the following.

The sender sends all media data in RTP packets using the SEL payload type plus additional the media payload type. The priority field is set according to the packet’s priority. In all packets the Delta SN field is used to signal the delta to the sequence number of the previous high priority packet. The receiver can use this information to check if the last high priority packet was received correctly or not. If this packet was not received or he cannot reliable detect the priority of lost packets (if many packets were lost in a row), a retransmission request is sent.

The receiver can also reconstruct the timestamp of the lost packet if the sender used the Delta Time field. From this time information it is possible to calculate the time after which the packet cannot be used any more (display time). In this case the client should maintain a round trip time (RTT) measurement and compare the calculated display time with the estimated time when the retransmission could arrive. If the retransmission would arrive already outdated, a retransmission request should not be sent.

At receiving a retransmission request from the receiver, the server has to make the same decision as described in the previous section, i.e. it has to judge according to the packets priority (which is most probably quite high, because a retransmission request for it was sent), network status etc. if the packet should be
retransmitted. If so it uses the retransmission packet type, to enable multiple retransmissions as described in the previous section. However the SEL payload type is additionally used to enable the priority detection at the client.

RTP Retransmission Payload Format  
July 2001

12 Security Considerations

T.B.D.

13 Acknowledgements

For the work on this retransmission framework, we received a lot of help and advice from Colin Perkins. Especially the mechanism of sending retransmissions with a new sequence number in the same sequence number space, using a different payload type was his idea.

14 Intellectual property considerations

Matsushita has filed patent applications that might possibly have technical relation to this contribution. If part(s) of the contribution by Matsushita employee is (are) included in a standard and Matsushita has patents and/or patent application(s) that are essential to implementation of such included part(s) in said standard, Matsushita is prepared to grant - on the basis of reciprocity (grantback) - a license on such included part(s) on reasonable, non-discriminatory terms and conditions (in accordance with paragraph 10.3.3 of the RFC 2026).

15 References

1  Bradner, S., "The Internet Standards Process -- Revision 3", BCP 9, RFC 2026, October 1996.

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


16 Author’s Addresses

Akihiro Miyazaki
Tsugihiko Ohno
Mitsubishi Electric Corporation
Information Technology R&D Center
5-1-1 Ofuna, Kamakura, Kanagawa, 247-8501, Japan
Tel. +81-467-41-2103
Fax. +81-467-41-2137
Mail tohno@isl.melco.co.jp