Synchronized Playback in Rapid Acquisition of Multicast Sessions
draft-yang-avt-rtp-synced-playback-04.txt

Abstract

When watching the same IPTV channel, different TV sets may not render the same picture and the associated audio at the same moment. The variation in end-to-end delay that resulted in such asynchronous playback between users is referred to as inter-user playback delay. Unicast based rapid acquisition of multicast RTP sessions (RAMS) as specified in [I-D.ietf-avt-rapid-acquisition-for-rtp] is an important technique in achieving fast channel switching in IPTV as well as other multicast applications. In addition, RAMS also significantly relaxes the requirement of relatively short random access point period in encoding of video streams in multicast applications, thus allowing significantly improved compression efficiency. However, on the other hand, the use of RAMS increases inter-user playback delay, which makes users receiving the same multicast session playback the same content asynchronously. This document specifies a mechanism to help reduce inter-user playback delay caused by the use of RAMS.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

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.
Table of Contents

1. Introduction ................................................. 3
2. Conventions .................................................. 4
3. Definitions .................................................. 5
4. Reducing Inter-User Playback Delay due to RAMS ............. 5
5. Message Extensions .......................................... 8
   5.1. Extension to RAMS-R ................................. 8
   5.2. Extension to RAMS-I ................................. 9
6. Security Considerations ...................................... 10
7. IANA Considerations ......................................... 10
8. The example of the duration speedup time ................... 11
9. Acknowledgements ........................................... 11
10. Normative References ....................................... 11
Authors’ Addresses .............................................. 12
1. Introduction

The Internet-Draft "Unicast-Based Rapid Acquisition of Multicast RTP Sessions" in [I-D.ietf-avt-rapid-acquisition-for-rtp] presents a method based on unicast burst stream for rapid acquisition of multicast RTP sessions (RAMS), thus to reduce the waiting time for the so-called Reference Information (RI). This method is effective in reducing channel switching and tune-in delay in multicast applications, such as IPTV. The RI typically starts at an access unit that is a random access point. In RAMS, RTP receivers start playback from the random access point from which the RI starts when they switch from one multicast session to another. As the unicast burst stream starts from the RI and is transmitted as fast as possible and faster than the media rate, on average the receiver can start processing the unicast burst stream faster, because it does not need to wait for the next random access point, as it does if it directly joined the multicast group.

Another important benefit brought by RAMS is that significantly improved coding efficiency for video streams is possible. In conventional multicast applications, video streams must be encoded with frequent random access points, e.g. 0.5 to 1 second, to allow new receivers to tune-in or existing users to switching from another multicast session. Random access points typically contain intra-coded pictures, for which the compression efficiency is significantly, e.g., several to ten times, lower than inter-coded pictures. Therefore, the less the random access points, the higher the coding efficiency. When RAMS is in use, random access period length is not the decision factor for the tune-in or channel switching delay. This means that the video random access point frequency can be significantly reduced, leading to significantly improved compression efficiency.

In a video communication system, an end-to-end delay is unavoidable. Typical end-to-end delay components are transmission delay, receiver buffering delay (which also handles transmission jitter), decoding buffering delay, output buffering delay, and other processing delays. Use of RAMS causes additional much longer end-to-end delay, the amount of which is equal to the time difference between the latest RTP timestamp of primary multicast packets buffered in the RS when the unicast burst starts and the RTP timestamp of the starting point of the unicast burst.

In multicast applications, receivers receiving the same multicast session can have different end-to-end delays and thus the playback of the same content among the receivers is not synchronized. In this document, a delay in playback time of the same content between any two users is referred to as inter-user playback delay (IUPD). This
document further refers to the typical end-to-end delay (when RAMS is not in use) as Common End-to-end Delay (CED). Among the delay components of CED, jitter buffer delay is the main part. Typical set-top box de-jitter buffers can store 100-500 ms (of SDTV) video, so network jitter must be within these limits and delay variation beyond these limits will manifest itself as loss [TR-126]. Compared to jitter buffer delay under the level of several hundreds of milliseconds, the acquisition of Reference Information by RAMS may result in longer and more serious delay which depends on the period of random access points, which sometimes are, vaguely, referred to as Group of Pictures (GOP) length (distance between key or I-frames). This document refers to this additional longer and more serious end-to-end delay caused by the use of RAMS as Rapid Acquisition Playback Delay (RAPD), and is specific to eliminate the RAPD to reduce the IUPD. The methods employed to solve CED is out of scope for this document.

Due to IUPD, different users may watch different pictures from different TV sets when watching the same IPTV content. In some application scenarios, e.g., remote education or a discussion room for an ongoing TV program in a social network, different users may be discussing the same content received through multicast. In this case, an obvious playback synchronization loss due to excessive inter-user playback delay can generate bad user experience.

As mentioned above, a disadvantage of RAMS is that the use of the technique causes RAPD for each user. Receivers are not synchronized in sending RAMS requests. Regardless of when the receiver starts the RAMS request (i.e. joins the program), the playback will start from a previous random access point. Given the starting point, the closer to the next random access point the receiver joins the program, the longer the RAP will be. Thus, the use of RAMS increases IUPD.

When obvious IUPD affects user experiences in above application scenarios, some actions must be taken. One way is to constrain the use of long random access period length, which significantly hurts video compression efficiency. In this document, we describe some mechanisms to eliminate the RAPD to reduce IUPD due to the use of RAMS and to allow the use of long random access period length for improved compression efficiency at the same time.

2. Conventions

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 [RFC2119].
3. Definitions

This document uses the following acronyms and definitions frequently:

Inter-User Playback Delay (IUPD): The playback delay between different users, for the same content transmitted on the same multicast session, due to variations in end-to-end delays.

Common End-to-end Delay (CED): The end-to-end delay when RAMS is not in use, which consists of transmission delay, receiver buffering delay (which also handles transmission jitter), decoding buffering delay, output buffering delay, and other processing delays.

Rapid Acquisition Playback Delay (RAPD): The additional end-to-end delay introduced by the use of RAMS. The value is equal to the time difference between the latest RTP timestamp of primary multicast packets buffered in the RS when the unicast burst starts and the RTP timestamp of the starting point of the unicast burst.

4. Reducing Inter-User Playback Delay due to RAMS

Inter-User Playback Delay (IUPD) is caused by the variation in end-to-end delay, which includes Common End-to-End Delay (CED) and Rapid Acquisition Playback Delay (RAPD). CED is primarily determined by jitter buffering delay which is in the range from 100 ms to 500 ms [TR-126]. In practice, IUPD caused by CED, and primarily by jitter buffer delay variations, should be under the level of tens of milliseconds. In contrast to this small range of IUPD values caused by CED, the IUPD caused by RAPD can be much higher, up to a few seconds or even to ten seconds, as long period of random access points can be used for significantly improved video compression efficiency when RAMS is in use.

To lower IUPD the receiver can speed up video playback until it catches up with the primary multicast stream. The procedures are outlined below.

The playback synchronization of receivers uses the Primary Multicast Stream as a reference point to address RAPD due to the use of RAMS. Each receiver keeps the synchronization with the primary multicast stream so that all receivers can keep an approximate synchronization in playback. The advantage is that it does not need every receiver to get playback synchronization information of other receivers, and thus scalability is not an issue as the procedure does not depend on the number of receivers.

The mechanism for speeding up video frames uses a method of skipping
video frame. In other words, one frame per an interval of some video frames is skipped until the number of extra video frames has been skipped. This way, the additional end-to-end delay, RAPD, can be compensated, and at the same time a smooth playback is ensured.

The mechanism involves the following changes to the RAMS method:
1) When the RTP Receiver (RR) sends a rapid acquisition request for the new multicast RTP session, the request MAY contain additional information indicating whether RR supports inter-user playback delay reduction.

2) When the Retransmission Sever (RS) receives the RAMS-R message and decides to accept it, RS MAY include the following additional information in the RAMS-I message to RR:

   a) N, the playback delay reduction target in number of frame durations; and

   b) V, recommended interval, in frames, between two continuous events for skipping of one frame.

When the RAMS-R message indicates that RR supports inter-user playback delay reduction, RS SHOULD include the above information in the RAMS-I message.

Retransmission server (RS) can precisely determine the value of N by detecting the time of RAMS-R, as the amount of additional end-to-end delay introduced by the use of RAMS, i.e. RAPD, is known to RS. The value is equal to the time difference between the latest RTP timestamp of primary multicast packets buffered in the RS when the unicast burst starts and the RTP timestamp of the starting point of the unicast burst. The value V is a recommended skip frame interval and the value must be chosen such that there is no noticeable audio distortion. For a video frame rate of 30 frames per second, typically when V is greater than 15 there is no noticeable audio distortion even when just speeding up the audio without using any specific algorithm to fix the audio pitch.

3) When RR receives an RAMS-I message containing the above information, it SHOULD speed up media rendering during playback taking into account the information as follows. During the speedup playback, after each V frames, one frame is skipped as if it was not present, and the presentation time of each remaining frame is shifted earlier by one frame duration, until totally N frames have been skipped. Receivers will playback the media content with its original speed after totally N frames have been skipped. Note that decoding remains the same as if speedup playback was not in use.

Besides the above mechanism, RS can use selective transmission of packets in the beginning of the unicast burst, by taking advantage of the temporal scalability of video bitstreams.
Video bitstreams are typically temporally scalable, in the sense that a portion of the bitstream can be extracted and decoded with a lower frame rate. Conventionally video coding standards like MPEG-2 can realize temporal scalability using different types of frames, i.e. I frame, P frame, and B frame, wherein I frames only consist of the lowest temporal layer (corresponding to the lowest frame rate), P frames only consist of the middle temporal layer, and B frames only consist of the highest temporal layer. Decoding of one temporal layer requires the presence of the temporal layer itself and all the lower temporal layers, but not the presence of any higher layer.

H.264/AVC and its scalable extension SVC (Scalable Video Coding) support advanced temporal scalability thanks to the flexible inter prediction scheme and flexible reference frame management scheme. It should be noted that in H.264/AVC and its extensions, B frames themselves may be reference frames, which may be used by other frames for inter prediction, therefore the discarding of which may lead to problems in decoding of the rest of the bitstream.

The RS MAY transmit the unicast burst as follows. In the beginning of the unicast burst, the RS discards one or more of the highest temporal layers and transmits only the remaining lower temporal layers. After a certain point, all temporal layers are transmitted. This would speed up the acquisition of the multicast session under the same unicast burst rate, at the cost of lower initial frame rate. At the same time, if the playback of the initial stream with lower frame rate is sped up, inter-user playback delay can be reduced.

5. Message Extensions

This section defines the extensions to RAMS-R and RAMS-I messages for inter-user playback delay reduction.

5.1. Extension to RAMS-R

The RAMS Request message is identified by SFMT=1. This message is used by RTP_Rx to request rapid acquisition for a primary multicast RTP session, or one or more primary multicast streams belonging to the same primary multicast RTP session. The FCI field MUST contain only one RAMS Request. The FCI field has the structure depicted in Figure 1.
5.2. Extension to RAMS-I

The RAMS Information message is identified by SFMT=2. This message is used to describe the unicast burst that will be sent for rapid acquisition. It also includes other useful information for RTP_Rx as described below. The FCI field MUST contain only one RAMS Information. The FCI field has the structure depicted in Figure 2.

Figure 1: FCI field syntax for the RAMS Request message

- Request for Inter-user Playback Delay Reduction (0 bits): TLV element that indicates that RTP_Rx is only requesting the inter-user playback delay reduction for the desired primary multicast stream(s). If this TLV element exists in the RAMS-R message, RS SHALL send the value N - the playback delay reduction target in number of frame durations and the value V - the recommended interval, in frames, between two continuous events for skipping of one frame. Note that this TLV element does not carry a Value field. Thus, the Length field MUST be set to zero.

Type: 6
<table>
<thead>
<tr>
<th>SFMT=2</th>
<th>MSN</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: FCI field syntax for the RAMS Information message

- The value \(N\) (16 bits) - the playback delay reduction target in number of frame durations: TLV element that denotes the amount of frame delay which is equal to the frame difference between the latest RTP timestamp of primary multicast packets buffered in the RS when the unicast burst starts and the RTP timestamp of the starting point of the unicast burst.

If the request for inter-user playback delay reduction has been accepted, RS MUST send this field at least once, so that RTP\_Rx knows to speed up media rendering during playback. Otherwise, RS ignore the request for inter-user playback delay reduction.

Type: 36

- The value \(V\) (8 bits) - recommended interval, in frames, between two continuous events for skipping of one frame: Optional TLV element that denotes the recommended interval for skipping of one frame. During the speedup playback of RTP\_Rx, after each \(V\) frames, one frame is skipped as if it was not present, and the presentation time of each remaining frame is shifted earlier by one frame duration, until totally \(N\) frames have been skipped. Receivers will playback the media content with its original speed after totally \(N\) frames have been skipped.

Type: 37

6. Security Considerations

TBD.

7. IANA Considerations

TBD.
8. The example of the duration speedup time

The following example demonstrates the speedup time. We consider a video frame sequence with a Frames per Second (FPS) equal to 30. The length of the first Group of Pictures (GOP) is 160 frames. I1B2B3P4B5B6P7B8B9P10B11B12P13B14B15P16B17B18P19B20P22B23B24P25B26B27P28B29B30P31B32...B158B159P160I161... If a receiver requests the RAMS and the unicast burst starts when the primary multicast stream is at the 120th frame which is equivalent to 4 seconds, that is, the value N (as specified in section 4) is equal to 120; the RS recommends that V is equal to 15. The receiver will skip one frame per an interval of 15 frames when receiving media stream. It will skip two frames per second because the FPS is equal to 30. In this example, B15 and B30 frames are skipped; the 45th and 60th frames also are skipped, as so on. When B15 is skipped by the receiver, the next P16 frame will replace B15 frame to be rendered without any frame delay. Similarly the B30, the 45th frame and the 60th frame are replaced by P31, the 46th and the 61st frames to be rendered, and so on. The FPS during the speedup is the same as the original FPS. In the first second, the receiver will play back up to B32 frame, that is, it speeds up by two frames. After two seconds, the receiver will play back up to the 64th frame and speeds up by four frames, and so on. In one minute, it will speed up by 120 frames (120/2=60 sec). In this example of 4 seconds delay, the receiver will need one minute to speed up during the playback and catch up with the primary multicast stream time.

Actually, in our commercial deployment for high definition most of RAMS delay are about 3-5 seconds, so that we achieve synchronization in about 1 minute.

9. Acknowledgements

The authors thank the following individuals for their contributions, comments and suggestions to this document: Roni Even, Colin Perkins, Ali C. Begen (abegen) and Bill Ver Steeg (versteb).

10. Normative References

[I-D.ietf-avt-rapid-acquisition-for-rtp]

[I-D.ietf-avt-rtcp-guidelines]
Ott, J. and C. Perkins, "Guidelines for Extending the RTP
Control Protocol (RTCP)

draft-ietf-avt-rtcp-guidelines-03 (work in progress), February 2010.


Authors' Addresses

Peilin Yang
Huawei Technologies Co., Ltd.
No.91, Baixia Road, Nanjing 210001
P.R.China

Phone: +86-25-84565881
Email: yangpeilin@huawei.com
Ye-Kui Wang  
Huawei Technologies Co., Ltd.  
400 Somerset Corporate Blvd, Suite 602  
Bridgewater, NJ 08807  
USA  

Phone: +1-908-541-3518  
Email: yekuiwang@huawei.com