Improved Rapid Acquisition of Multicast Sessions

draft-wang-avt-rtp-improved-rams-01.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79. This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

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.

This Internet-Draft will expire on April 26, 2010.
Abstract

This document describes two improvements to unicast based rapid acquisition of multicast RTP sessions (RAMS) described in [I-D.ietf-avt-rapid-acquisition-for-rtp]. One improved method allows a receiver to simultaneously request a unicast burst stream and join the multicast group, and then select one of the two streams to be first processed. Another method proposes to optionally convey a parameter called unicast backspace in the RAMS-I message. This parameter can be used by RR to determine when to join the primary multicast session.

Table of Contents

1. Introduction .................................................. 3
   1.1. Improvement - RR simultaneously request a unicast burst stream and join the multicast group ....................... 3
   1.2. Improvement - RR determines the primary multicast session join time using parameter called Unicast Backspace..... 5
2. Conventions .................................................. 6
3. Improved Rapid Acquisition of Multicast RTP Session ........ 7
   3.1. RR simultaneously request a unicast burst stream and join the multicast group .............................................. 7
   3.2. RR determines the primary multicast session join time using parameter called Unicast Backspace .......................... 9
      3.2.1. Unicast Backspace Field .................................. 9
      3.2.2. Example Message Flow .................................... 10
4. Security Considerations ..................................... 11
5. IANA Considerations ........................................... 11
6. Acknowledgements ............................................. 12
7. References ..................................................... 12
8. Authors’ Addresses ............................................. 12
1. Introduction

The basic idea of unicast based rapid acquisition of multicast RTP sessions (RAMS) in [I-D.ietf-avt-rapid-acquisition-for-rtp] is as follows. Instead of joining the multicast group, the receiver first requests a unicast stream, which is transmitted at a rate faster than the multicast stream rate. The unicast stream starts from a random access point, which contains the so-called Reference Information (RI) or the start of the RI, and thus can be processed without waiting for the next random access point, and consequently the delay for acquisition of the multicast stream is reduced. This method is effective in reducing channel switching and tune-in delay in multicast applications, such as IPTV.

Another important benefit of 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. per 0.5 to 1 second, to allow new receivers to tune-in or existing users to switch 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 no longer affects the tune-in or channel switching delay. This means that the video random access point frequency can be significantly reduced, which means significantly improved compression efficiency.

1.1. Improvement - RR simultaneously request a unicast burst stream and join the multicast group

Nevertheless, RAMS has the following constraints:

- At some point, if the receiver directly joins the multicast group, the received multicast stream may start exactly at a random access point or at a point that is very close to the next random access point. In this case, the use of the RAMS method actually increases the acquisition delay. Unfortunately, the receiver has no means to know how far the next random access point in the multicast stream is before it joins the multicast group, otherwise it could choose to either use RAMS or directly join the multicast group, whichever is better.
The request for the unicast burst stream may be lost. Even if the request is retransmitted and finally received, the acquisition delay gets increased.

Part or all of the RI may be lost. Again, the lost information could be retransmitted, which increases the acquisition delay. If the RI or part of it gets finally lost in any case, the received data in the unicast burst stream cannot be processed until the next random access point. In this case, the acquisition delay would also be lower if the receiver had chosen to directly join the multicast group without trying to request the unicast burst stream. Unfortunately, there is no way for the receiver to know what loss could happen before the occurrence of the loss. To solve the above problems, this document proposes an improved RAMS method, wherein the receiver simultaneously requests the unicast burst stream and joins the multicast group. When both the unicast burst stream and the multicast stream are transmitted to the receiver, the downlink bandwidth needed is at least twice as high as the media rate (including the overhead due to network protocol headers). Different than the original RAMS method in [I-D.ietf-avt-rapid-acquisition-for-rtp], herein the unicast burst stream does not need to be transmitted at a rate higher than the media rate, though a higher rate can still be used if possible.

The proposed scheme is rationalized based on the following observations.
A user who has a certain network accessing bandwidth can typically upgrade the access bandwidth without upgrade of the network accessing equipment. This means that the network access equipment can typically support higher bandwidth. On the other hand it may not be a major issue for the network provider to send data to the user at a bandwidth higher than the single stream to the user, unless there is limitation on the data sent by the network provider. In the context of unicast based rapid acquisition of multicast streams, as the data of the unicast burst stream is just a copy of the multicast stream, there is no additional cost for the data itself (as content) if both streams are obtained. Therefore, it is feasible both technically and economically (maybe at some added cost) for the user to simultaneously receive the unicast burst stream and the multicast stream. In particular, as the acquisition process involving receiving both streams is short, in the magnitude of a few seconds, it is likely that network providers, which in many cases are also content providers and services providers, are willing to allow the momentary use of higher bandwidth than usual, either freely to users (i.e. users still pay for the usual bandwidth) or under a new contract that explicitly covers such momentary uses of higher bandwidth. Their reasoning will be to supply better user experience when they compete with other delivery technologies.

In a digital home with multiple TVs and possibly other connected equipments such as PCs, more than one TV program on different TVs may be watched simultaneously in addition to other network uses under one network accessing contact. In such a common scenario, the bandwidth available can easily be as at least twice high as the media rate.

Note that there will still be cases wherein available bandwidth is not enough for transmission of both the unicast burst stream and the multicast stream simultaneously. Therefore, it should still be allowed for receivers to switch between the proposed improved method, the original RAMS method, and the conventional method (i.e. directly joining the multicast group).

1.2. Improvement - RR determines the primary multicast session join time using parameter called Unicast Backspace

Per [I-D.ietf-avt-rapid-acquisition-for-rtp], RS may change the unicast burst rate during the acquisition process, decided by itself or based on some feedback information from RR. RS may therefore recalculate the earliest multicast join time and send the
revised values to RR using additional RAMS-I messages. If such an 
RAMS-I message gets lost, RR would fail to update its multicast 
join time accordingly and therefore join the multicast session at 
an inappropriate time. This increases the chance and amount of 
overlap or gap between the unicast burst and primary multicast 
stream.

To increase the accuracy of multicast join time deduction during 
the fast acquisition, this document proposes to convey a field 
called unicast backspace in the first RAMS-I message which is as a 
response to the acceptance of the first RAMS-R message. Unicast 
backspace is the RTP time stamp difference between the first 
unicast burst packet to be sent and the latest primary multicast 
packet in the buffer of RS. The RTP timestamps values for these 
packets are taken from the primary multicast stream. The value of 
unicast backspace indicates the amount of data that would be 
additional transmitted to RR compared to when the RAMS process is 
not in use. This value also indicates the additional end-to-end 
delay introduced by the RAMS process.

Unicast backspace is a constant during a fast acquisition process 
once the start point of the unicast burst is decided by RS. This 
value, in conjunction with some parameters that are known to RR, 
e.g. burst rate, playback rate, amount of buffered data, and 
initial playback delay, can be used by RR to heuristically 
determine when to launch multicast join process by sending an SFGMP 
join message. RR therefore may not count on RAMS-I to notify the 
SFGMP join time update due to burst rate change during the rapid 
acquisition and therefore may alleviate the RS implementation 
complexity for re-deducing multicast join time. And this way, the 
amount of protocol signaling interaction between RS and RR may be 
reduced. Furthermore, when unicast backspace value is known, RR may 
remove or reduce the additional end-to-end delay introduced by the 
RAMS process by appropriately controlling the playback process, 
such that the playback can be synchronized with other users of the 
same primary multicast session.

Note that how RR uses this field to deduce when RR launch the 
multicast join process is out of scope of this document.

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 RFC 2119 
[RFC2119].
3. Improved Rapid Acquisition of Multicast RTP Sessions

3.1. RR simultaneously request a unicast burst stream and join the multicast group

The improved RAMS method is described by the following steps (referring to Figure 1, based on the same terminology in [I-D.ietf-avt-rapid-acquisition-for-rtp]):

1) Request: RR sends an RAMS-R message to the RS to request for the unicast burst session of the new multicast RTP session, and an SFGMP (i.e. IGMP or MLD) Join message to the Router to join the new multicast session. Preferably, the RAMS-R and SFGMP Join messages are sent simultaneously. However, the RAMS-R and SFGMP Join messages can also be sent at different temporal locations but should be reasonably close to each other, and there is no restriction to the temporal order of sending the two messages.

2) Response: RS receives the RAMS-R message and decides whether to accept it or not. RS sends one or more RAMS-Information (RAMS-I) messages to the receiver. If the request is accepted by RS, it starts sending the unicast burst stream. The unicast burst can be sent to receiver either after or together with RAMS-I message. At the same time or at a slightly different time, the multicast stream starts to flow to RR from the Multicast Source.

3) Updated Request or Updated Response: RR MAY send updated RAMS-R messages to RS. RS MAY send updated RAMS-I messages to RR.
Figure 1 Flow of the improved RAMS method
4) Process: For the received unicast stream and the received multicast stream, the one that can be processed first is chosen. A stream can only be processed from a correctly received random access point. If both the received unicast stream and the received multicast stream contain correctly received random access points, then either can be chosen. If the multicast stream is chosen, then the unicast stream is terminated immediately. In this case, all received packets of the multicast stream are processed (depacketized and decoded). Otherwise (i.e. the unicast stream is chosen), the unicast stream is terminated when it catches the beginning of the received multicast stream, i.e. after the packet in the unicast stream having OSN equal to SNm-1 is received, where OSN stands for Original Sequence Number as defined in [I-D.ietf-avt-rapid-acquisition-for-rtp], SNm is the Sequence Number (SN) of the first packet in the received multicast stream. In this case, all received packets of the unicast stream having OSN equal to or greater than SNm are discarded, and other received packets of the unicast stream and all received packets of the multicast stream are processed.

5) Terminate: In the above step, the unicast stream is terminated by sending an RAMS-T message to RS. To terminate the unicast stream immediately, RR sends an RAMS-T message without an RTP sequence number. To instruct RS to terminate the unicast stream after catching the multicast stream, RR SHALL include in the RAMS-T message the sequence number of the first received RTP packet in the received multicast stream. When RR is receiving the unicast stream, and if RR becomes no longer interested in the multicast stream, RR sends a RTCP BYE message to RS to terminate the RTP retransmission session (which contains both the unicast burst stream as well as the retransmission stream for lost packets).

3.2. RR determines the primary multicast session join time using parameter called Unicast Backspace

3.2.1. Unicast Backspace Field

Unicast backspace (32 bits): Optional TLV element. This field gives the RTP timestamp difference between the first unicast burst packet to be sent and the latest primary multicast packet in the buffer. The RTP timestamps values for these packets are taken from the primary multicast stream. The value of unicast backspace indicates the amount of data that would be additional transmitted
to RR compared to when the RAMS process is not in use. This value also indicates the additional end-to-end delay introduced by the RAMS process. This field SHOULD be conveyed in the first RAMS-I message and MAY be conveyed in other RAMS-I messages. RR MAY use this field to determine when to join the multicast session by sending an SFGMP join message. RR MAY also remove or reduce the additional end-to-end delay due to the use of RAMS, as indicated by the unicast backspace field, by appropriately controlling the playback process, such that the playback can be synchronized with other users of the same primary multicast session.

### 3.2.2. Example Message Flow

Figure 2 depicts an example of messaging flow for RR using backspace field to heuristically determine the multicast join time.

```
+-----------+   +----------------+   +----------+   +----------+
| Multicast |   | Retransmission |   |          |   |    RTP   |
| Source    |   |     Server     |   |  Router  |   |  Receiver|
|           |   |      (RS)      |   |          |   |    (RR)  |
+-----------+   +----------------+   +----------+   +----------+

-- RTP Multicast ------------------->|                |
-- RTP Multicast ->
<"""""""""""" RTCP RAMS-R """"
"" (RTCP RAMS-I with """""""
unicast backspace)
.. Unicast RTP Burst .............>
<""""""""(RTCP RAMS-R)''
<~ SFGMP Join ~~
```

Figure 2 Message Flow for heuristic Multicast Join Time Determination

""""> Unicast RTCP Messages
The example is described by the following steps:

1. RR sends a RTCP RAMS-R message to RS to request a rapid acquisition process.

2. Upon accepting of the RAMS-R message, RS sends an RTCP RAMS-I message with unicast backspace field to RR.

3. RS transmit Unicast RTP Burst to RR.

4. During the burst, RR uses the value of unicast backspace to determine when to send SFGMP Join message to join the primary multicast group. For example, RR (periodically) estimates the buffer occupation in terms of media timestamp offset between the first and the last RTP packets in the buffer (called B for example). If the difference between the unicast backspace value and B is less than or equal to a pre-defined threshold value, RR begins to join the multicast session. Unicast backspace is a determinate value and is always fixed during the burst. The amount of data in the buffer is known by RR. The threshold value can be derived as equal to

\[
\frac{(b_r - p_r)}{p_r} \times \text{multicast}_\text{join}_\text{latency}
\]

Where \(b_r\) is the burst rate, \(p_r\) is the playback rate, and \text{multicast}_\text{join}_\text{latency} is the delay between RR sending SFGMP join message and RR getting the first primary multicast packet.

This way, RR would not rely on RAMS-I to revise the earliest multicast join time after the burst rate is changed by RS.

4. Security Considerations

TBD.

5. IANA Considerations

TBD.
6. Acknowledgements

TBD.

This document was prepared using 2-Word-v2.0.template.dot.

7. References

[I-D.ietf-avt-rapid-acquisition-for-rtp]
Steeg, B., Begen, A., Caenegem, T., and Z. Vax, "Unicast-Based Rapid Acquisition of Multicast RTP Sessions",


8. Authors’ Addresses

YeKui Wang
Huawei Technologies Co., Ltd.
400 Somerset Corp Blvd, Suite 602
Bridgewater, NJ 08807
USA

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

Jinliang Dai
Huawei Technologies Co., Ltd.
BuildingNO.17, Zpark
Hai-Dian District, Beijing
P. R. China

Phone: +86-10-82829100
EMail: daijinliang@huawei.com
Jiangping Feng  
Huawei Technologies Co., Ltd.  
Building NO.17, Zpark  
Haidian District, Beijing  
P. R. China  
Phone: +86-10-82829092  
EMail: fengjiangping@huawei.com

Zi-Xuan Zou  
Huawei Technologies Co., Ltd.  
Huawei Industrial Base, Longgang, B1-4  
P. R. China  
Phone: +86-0755-28789364  
EMail: tendyntu@huawei.com