AVT B. VerSteeg
Internet-Draft A. Begen
Intended status: Standards Track Cisco
Expires: August 22, 2010 T. VanCaenegem
Alcatel-Lucent
Z. Vax
Microsoft Corporation
February 18, 2010

Unicast-Based Rapid Acquisition of Multicast RTP Sessions
draft-ietf-avt-rapid-acquisition-for-rtp-07

Abstract

When an RTP receiver joins a multicast session, it may need to acquire and parse certain Reference Information before it can process any data sent in the multicast session. Depending on the join time, length of the Reference Information repetition (or appearance) interval, size of the Reference Information as well as the application and transport properties, the time lag before an RTP receiver can usefully consume the multicast data, which we refer to as the Acquisition Delay, varies and may be large. This is an undesirable phenomenon for receivers that frequently switch among different multicast sessions, such as video broadcasts.

In this document, we describe a method using the existing RTP and RTCP protocol machinery that reduces the acquisition delay. In this method, an auxiliary unicast RTP session carrying the Reference Information to the receiver precedes/accompanies the multicast stream. This unicast RTP flow may be transmitted at a faster than natural birate to further accelerate the acquisition. The motivating use case for this capability is multicast applications that carry real-time compressed audio and video. However, the proposed method can also be used in other types of multicast applications where the acquisition delay is long enough to be a problem.

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.

This Internet-Draft will expire on August 22, 2010.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the BSD License.

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.

Table of Contents

1. Introduction

Most multicast flows carry a stream of inter-related data. Certain information must first be acquired by the receivers to start processing any data sent in the multicast session. This document refers to this information as Reference Information. The Reference Information is conventionally sent periodically in the multicast session (although its content may change over time) and usually consists of items such as a description of the schema for the rest of the data, references to which data to process, encryption information including keys, as well as any other information required to process the data in the multicast stream [IC2009].

Real-time multicast applications require the receivers to buffer data. The receiver may have to buffer data to smooth out the network jitter, to allow loss-repair methods such as Forward Error Correction and retransmission to recover the missing packets, and to satisfy the data processing requirements of the application layer.

When a receiver joins a multicast session, it has no control over what point in the flow is currently being transmitted. Sometimes the receiver may join the session right before the Reference Information is sent in the session. In this case, the required waiting time is usually minimal. Other times, the receiver may join the session right after the Reference Information has been transmitted. In this case, the receiver has to wait for the Reference Information to appear again in the flow before it can start processing any multicast data. In some other cases, the Reference Information is not contiguous in the flow but dispersed over a large period, which forces the receiver to wait for all of the Reference Information to arrive before starting to process the rest of the data.

The net effect of waiting for the Reference Information and waiting for various buffers to fill up is that the receivers may experience significantly large delays in data processing. In this document, we refer to the difference between the time an RTP receiver joins the multicast session and the time the RTP receiver acquires all the necessary Reference Information as the Acquisition Delay. The acquisition delay may not be the same for different receivers; it usually varies depending on the join time, length of the Reference Information repetition (or appearance) interval, size of the Reference Information as well as the application and transport properties.

The varying nature of the acquisition delay adversely affects the receivers that frequently switch among multicast sessions. In this specification, we address this problem for RTP-based multicast applications and describe a method that uses the fundamental tools

offered by the existing RTP and RTCP protocols [RFC3550]. In this method, either the multicast source (or the distribution source in a source-specific multicast (SSM) session) retains the Reference Information for a period after its transmission, or an intermediary network element (that we refer to as Retransmission Server) joins the multicast session and continuously caches the Reference Information as it is sent in the session and acts as a feedback target (See [I-D.ietf-avt-rtcpssm]) for the session. When an RTP receiver wishes to join the same multicast session, instead of simply issuing a Source Filtering Group Management Protocol (SFGMP) Join message, it sends a request to the feedback target for the session and asks for the Reference Information. The Retransmission Server starts a new unicast RTP (retransmission) session and sends the Reference Information to the RTP receiver over that session. If there is spare bandwidth, the Retransmission Server may burst the Reference Information faster than its natural rate. As soon as the receiver acquires the Reference Information, it can join the multicast session and start processing the multicast data. A simplified network diagram showing this method through an intermediary network element is depicted in Figure 1.

This method potentially reduces the acquisition delay. We refer to this method as Unicast-based Rapid Acquisition of Multicast RTP Sessions. A primary use case for this method is to reduce the channel-change times in IPTV networks where compressed video streams are multicast in different SSM sessions and viewers randomly join these sessions.

```
                                       -----------------------
                                 +--->|     Intermediary      |
                                 |    |    Network Element    |
                                 | ...|(Retransmission Server)|
                                 | :   -----------------------
                                 | :
                                 | v
           -----------         ----------             ----------
          | Multicast |       |          |---------->| Joining  |
          |  Source   |------>|  Router  |..........>|   RTP    |
          |           |       |          |           | Receiver |
           -----------         ----------             ----------
                                   |
                                   |                  ----------
                                   +---------------->| Existing |
                                                     |    RTP   |
                                                     | Receiver |
                                                      ----------


          -------> Multicast RTP Flow
          .......> Unicast RTP Flow
```

Figure 1: Rapid acquisition through an intermediary network element

A principle design goal in this solution is to use the existing tools in the RTP/RTCP protocol family. This improves the versatility of the existing implementations, and promotes faster deployment and better interoperability. To this effect, we use the unicast retransmission support of RTP [RFC4588] and the capabilities of RTCP to handle the signaling needed to accomplish the acquisition.

1.1. Acquisition of RTP Streams vs. RTP Sessions

By the definition given in [RFC3550], an RTP session may involve one or more RTP streams each identified with a unique SSRC. All RTP streams within a single RTP session are sent towards the same transport address, i.e., they share the same destination IP address and port. In RTP jargon, these streams are said to be SSRC-multiplexed. On the other hand, an SSM session is uniquely identified by its source address and destination group address. However, it may carry one or more RTP sessions, each associated with a different destination port. Consequently, while it is not very practical, it is still possible for an SSM session to carry multiple RTP sessions each carrying multiple SSRC-multiplexed RTP streams.

Developing a protocol that can jointly handle the rapid acquisition of all of the RTP sessions in an SSM session is neither practical nor

necessary. Rather, in this specification we focus on developing a protocol that handles the rapid acquisition of a single RTP session (called primary multicast RTP session) carrying one or more RTP streams (called primary multicast streams). If desired, multiple instances of this protocol may be run in parallel to acquire multiple RTP sessions simultaneously.

When an RTP receiver requests the Reference Information from the Retransmission Server, it may opt to rapidly acquire a specific subset of the available RTP streams in the primary multicast RTP session. Alternatively, it may request the rapid acquisition of all of the RTP streams in that RTP session. Regardless of how many RTP streams are requested by the RTP receiver or how many will be actually sent by the Retransmission Server, only one unicast RTP (retransmission) session will be established by the Retransmission Server serving as the feedback target for that RTP session. The RTP receiver multiplexes this unicast RTP session with the primary multicast RTP session it receives as part of the SSM session. If the RTP receiver wants to rapidly acquire multiple RTP sessions simultaneously, separate unicast RTP (retransmission) sessions will be established for each of them.

1.2. Outline

In the rest of this specification, we have the following outline: In Section 4, we describe the delay components in generic multicast applications. Section 5 presents an overview of the protocol design considerations for rapid acquisition. We provide the protocol details of the rapid acquisition method in Section 6 and Section 7. Section 8 and Section 9 discuss the SDP signaling issues with examples and NAT-related issues, respectively. Finally, Section 10 discusses the security considerations.

Section 3 provides a list of the definitions frequently used in this document.

2. Requirements Notation

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:

(Primary) SSM (or Multicast) Session: The multicast session to which RTP receivers can join at a random point in time.

Primary Multicast RTP Session: The multicast RTP session an RTP receiver is interested in acquiring rapidly. A primary SSM session may carry multiple multicast RTP sessions, but only one of them can be the primary from the viewpoint of rapid acquisition.

Primary Multicast (RTP) Streams: The RTP stream(s) carried in the primary multicast RTP session.

Source Filtering Group Management Protocol (SFGMP): Following the definition in [RFC4604], SFGMP refers to the Internet Group Management Protocol (IGMP) version 3 [RFC3376] and the Multicast Listener Discovery Protocol (MLD) version 2 [RFC3810] in the IPv4 and IPv6 networks, respectively. However, the rapid acquisition method introduced in this document does not depend on a specific version of either of these group management protocols. In the remainder of this document, SFGMP will refer to any group management protocol that has Join and Leave functionalities.

Feedback Target (FT): Unicast RTCP feedback target as defined in [I-D.ietf-avt-rtcpssm]. FT_Ap denotes a specific feedback target running on a particular address and port.

Retransmission (Burst) Packet: An RTP packet that is formatted as defined in [RFC4588].

Reference Information: The set of certain media content and metadata information that is sufficient for an RTP receiver to start usefully consuming a media stream. The meaning, format and size of this information are specific to the application and are out of scope of this document.

Preamble Information: A more compact form of the whole or a subset of the Reference Information transmitted out-of-band.

(Unicast) Burst (Stream): A unicast stream of RTP retransmission packets that enable an RTP receiver to rapidly acquire the Reference Information associated with a primary multicast stream. Each burst stream is identified by its SSRC identifier that is unique in the primary multicast RTP session. The burst streams are typically transmitted at an accelerated rate.

Retransmission Server (RS): The RTP/RTCP endpoint that can generate the retransmission packets and the burst streams. RS may also generate other non-retransmission packets to aid the rapid acquisition process.

4. Elements of Delay in Multicast Applications

In an any-source (ASM) or a source-specific (SSM) multicast delivery system, there are three major elements that contribute to the overall acquisition delay when an RTP receiver switches from one multicast session to another one. These are:

- Multicast switching delay
- Reference Information latency
- Buffering delays

Multicast switching delay is the delay that is experienced to leave the current multicast session (if any) and join the new multicast session. In typical systems, the multicast join and leave operations are handled by a group management protocol. For example, the receivers and routers participating in a multicast session may use the Internet Group Management Protocol (IGMP) version 3 [RFC3376] or the Multicast Listener Discovery Protocol (MLD) version 2 [RFC3810]. In either of these protocols, when a receiver wants to join a multicast session, it sends a message to its upstream router and the routing infrastructure sets up the multicast forwarding state to deliver the packets of the multicast session to the new receiver. Depending on the proximity of the upstream router, the current state of the multicast tree, the load on the system and the protocol implementation, the join times vary. Current systems provide join latencies usually less than 200 milliseconds (ms). If the receiver had been participating in another multicast session before joining the new session, it needs to send a Leave message to its upstream router to leave the session. In common multicast routing protocols, the leave times are usually smaller than the join times, however, it is possible that the Leave and Join messages may get lost, in which case the multicast switching delay inevitably increases.

Reference Information latency is the time it takes the receiver to acquire the Reference Information. It is highly dependent on the proximity of the actual time the receiver joined the session to the next time the Reference Information will be sent to the receivers in the session, whether the Reference Information is sent contiguously or not, and the size of the Reference Information. For some multicast flows, there is a little or no interdependency in the data, in which case the Reference Information latency will be nil or negligible. For other multicast flows, there is a high degree of interdependency. One example of interest is the multicast flows that carry compressed audio/video. For these flows, the Reference Information latency may become quite large and be a major contributor to the overall delay. Refer to [I-D.begen-avt-rtp-mpeg2ts-preamble]

for details.

The buffering component of the overall acquisition delay is driven by the way the application layer processes the payload. In many multicast applications, an unreliable transport protocol such as UDP [RFC0768] is often used to transmit the data packets, and the reliability, if needed, is usually addressed through other means such as Forward Error Correction (e.g., [I-D.ietf-fecframe-interleaved-fec-scheme]) and retransmission. These loss-repair methods require buffering at the receiver side to function properly. In many applications, it is also often necessary to de-jitter the incoming data packets before feeding them to the application. The de-jittering process also increases the buffering delays. Besides these network-related buffering delays, there are also specific buffering needs that are required by the individual applications. For example, standard video decoders typically require an amount, sometimes a significant amount, of coded video data to be available in the pre-decoding buffers prior to starting to decode the video bitstream.

5. Protocol Design Considerations and Their Effect on Resource Management for Rapid Acquisition

Rapid acquisition is an optimization of a system that must continue to work correctly and properly whether or not the optimization is effective, or even fails due to lost control and feedback messages, congestion, or other problems. This is fundamental to the overall design requirements surrounding the protocol definition and to the resource management schemes to be employed together with the protocol (e.g., QoS machinery, server load management, etc). In particular, the system needs to operate within a number of constraints:

o First, a rapid acquisition operation must fail gracefully. The user experience must, except perhaps in pathological circumstances, be not significantly worse for trying and failing to complete rapid acquisition compared to simply joining the multicast session.

o Second, providing the rapid acquisition optimizations must not cause collateral damage to either the multicast session being joined, or other multicast sessions sharing resources with the rapid acquisition operation. In particular, the rapid acquisition operation must avoid interference with the multicast session that may be simultaneously being received by other hosts. In addition, it must also avoid interference with other multicast sessions sharing the same network resources. These properties are possible, but are usually difficult to achieve.

One challenge is the existence of multiple bandwidth bottlenecks between the receiver and the server(s) in the network providing the rapid acquisition service. In commercial IPTV deployments, for example, bottlenecks are often present in the aggregation network connecting the IPTV servers to the network edge, the access links (e.g., DSL, DOCSIS) and in the home network of the subscribers. Some of these links may serve only a single subscriber, limiting congestion impact to the traffic of only that subscriber, but others can be shared links carrying multicast sessions of many subscribers. Also note that the state of these links may be varying over time. The receiver may have knowledge of a portion of this network, or may have partial knowledge of the entire network. The methods employed by the devices to acquire this network state information is out of scope for this document. The receiver should be able to signal the server with the bandwidth that it believes it can handle. The server also needs to be able to rate limit the flow in order to stay within the performance envelope that it knows about. Both the server and receiver need to be able to inform the other of changes in the requested and delivered rates. However, the protocol must be robust in the presence of packet loss, so this signaling must include the appropriate default behaviors.

A second challenge is that for some uses (e.g., high-bitrate video) the unicast burst bitrate is high while the flow duration of the unicast burst is short. This is because the purpose of the unicast burst is to allow the RTP receiver to join the multicast quickly and thereby limit the overall resources consumed by the burst. Such high-bitrate, short-duration flows are not amenable to conventional admission control techniques. For example, end-to-end per-flow signaled admission control techniques such as RSVP have too much latency and control channel overhead to be a good fit for rapid acquisition. Similarly, using a TCP (or TCP-like) approach with a 3-way handshake and slow-start to avoid inducing congestion would defeat the purpose of attempting rapid acquisition in the first place by introducing many round-trip times (RTT) of delay.

These observations lead to certain unavoidable requirements and goals for a rapid acquisition protocol. These are:

- The protocol must be designed to allow a deterministic upper bound on the extra bandwidth used (compared to just joining the multicast session). A reasonable size bound is e*B, where B is the nominal bandwidth of the primary multicast streams, and e is an excess-bandwidth coefficient. The total duration of the unicast burst must have a reasonable bound; long unicast bursts devolve to the bandwidth profile of multi-unicast for the whole system.

- o The scheme should minimize (or better eliminate) the overlap of the unicast burst and the primary multicast stream. This minimizes the window during which congestion could be induced on a bottleneck link compared to just carrying the multicast or unicast packets alone.

- o The scheme must minimize (or better eliminate) any gap between the unicast burst and the primary multicast stream, which has to be repaired later, or in the absence of repair, will result in loss being experienced by the application.

In addition to the above, there are some other protocol design issues to be considered. First, there is at least one RTT of "slop" in the control loop. In starting a rapid acquisition burst, this manifests as the time between the client requesting the unicast burst and the burst description and/or the first unicast burst packets arriving at the receiver. For managing and terminating the unicast burst, there are two possible approaches for the control loop: The receiver can adapt to the unicast burst as received, converge based on observation and explicitly terminate the unicast burst with a second control loop exchange (which takes a minimum of one RTT, just as starting the unicast burst does). Alternatively, the server generating the unicast burst can pre-compute the burst parameters based on the information in the initial request and tell the receiver the burst duration.

The protocol described in the next section allows either method of controlling the rapid acquisition unicast burst.

6. Rapid Acquisition of Multicast RTP Sessions

We start this section with an overview of the rapid acquisition of multicast sessions (RAMS) method.

6.1. Overview

[I-D.ietf-avt-rtcpssm] specifies an extension to the RTP Control Protocol (RTCP) to use unicast feedback in an SSM session. It defines an architecture that introduces the concept of Distribution Source, which - in an SSM context - distributes the RTP data and redistributes RTCP information to all RTP receivers. This RTCP information is retrieved from the Feedback Target, to which RTCP unicast feedback traffic is sent. The feedback target MAY be implemented in one or more entities different from the Distribution Source, and different RTP receivers MAY use different feedback targets.

This document builds further on these concepts to reduce the acquisition delay when an RTP receiver joins a multicast session at a random point in time by introducing the concept of the Burst Source and new RTCP feedback messages. The Burst Source has a cache where the most recent packets from the primary multicast RTP session are continuously stored. When an RTP receiver wants to receive a primary multicast stream prior to joining the SSM session, it may first request a unicast burst from the Burst Source. In this burst, the packets are formatted as RTP retransmission packets [RFC4588] and carry the Reference Information. This information allows the RTP receiver to start usefully consuming the RTP packets sent in the primary multicast RTP session.

Using an accelerated bitrate (as compared to the nominal bitrate of the primary multicast stream) for the unicast burst implies that at a certain point in time, the payload transmitted in the unicast burst is going to be the same as the payload in the associated multicast stream, i.e., the unicast burst will catch up with the primary multicast stream. At this point, the RTP receiver no longer needs to receive the unicast burst and can join the SSM session. This method is referred to as the Rapid Acquisition of Multicast Sessions (RAMS).

This document proposes extensions to [RFC4585] for an RTP receiver to request a unicast burst as well as for additional control messaging that can be leveraged during the acquisition process.

6.2. Message Flows

Figure 2 shows the main entities involved in rapid acquisition and the message flows. They are

- o Multicast Source
- o Feedback Target (FT)
- o Burst/Retransmission Source
- o RTP Receiver (RTP_Rx)

```
      -----------                                      --------------
     |           |----------------------------------->|              |
     |           |.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.>|              |
     |           |                                    |              |
     | Multicast |         ----------------           |              |
     |  Source   |        | Retransmission |          |              |
     |           |------->|  Server  (RS)  |--------->|              |
     |           |.-.-.-.>|                |.-.-.-.-.>|              |
     |           |        |  ------------  |          |              |
      -----------         | |  Feedback  | |<.=.=.=.=.|              |
                          | |   Target   | |<~~~~~~~~~| RTP Receiver |
                          |  ------------  |          |   (RTP_Rx)   |
                          |                |          |              |
                          |  ------------  |          |              |
                          | | Burst  and | |<~~~~~~~~>|              |
                          | |  Retrans.  | |.........>|              |
                          | |   Source   | |<.=.=.=.=>|              |
                          |  ------------  |          |              |
                          |                |          |              |
                           ----------------            --------------


     -------> Multicast RTP Flow
     .-.-.-.> Multicast RTCP Flow
     .=.=.=.> Unicast RTCP Reports
     ~~~~~~~> Unicast RTCP Feedback Messages
     .......> Unicast RTP Flow
```

Figure 2: Flow diagram for unicast-based rapid acquisition

The feedback target (FT) is the entity as defined in [I-D.ietf-avt-rtcpssm], to which RTP_Rx sends its RTCP feedback messages indicating packet loss by means of an RTCP NACK message or indicating RTP_Rx's desire to rapidly acquire the primary multicast RTP session by means of an RTCP feedback message defined in this document. While the Burst/Retransmission Source is responsible for responding to these messages and for further RTCP interaction with RTP_Rx in the case of a rapid acquisition process, it is assumed in the remainder of the document that these two logical entities (FT and Burst/Retransmission Source) are combined in a single physical entity and they share state. In the remainder of the text, the term Retransmission Server (RS) will be used whenever appropriate, to refer to the combined functionality of the FT and Burst/Retransmission Source.

However, it must be noted that only FT is involved in the primary multicast RTP session, whereas the Burst/Retransmission Source transmits the unicast burst and retransmission packets both formatted

as RTP retransmission packets [RFC4588] in a single separate unicast RTP retransmission session to each RTP_Rx. In the retransmission session, as in any other RTP session, RS and RTP_Rx regularly send the periodic sender and receiver reports, respectively.

The unicast burst is triggered by an RTCP feedback message that is defined in this document based on the common packet format provided in [RFC4585], whereas an RTP retransmission is triggered by an RTCP NACK message defined in [RFC4585]. In the RTP/AVPF profile, there are certain rules that apply to scheduling of both of these messages, which are detailed in Section 3.5 of [RFC4585]. One of the rules states that in a multi-party session such as the SSM sessions we are considering in this specification, an RTP receiver cannot send an RTCP feedback message for a minimum of one second period after joining the session (i.e., Tmin=1.0 second). While this rule has the goal of avoiding problems associated with flash crowds in typical multi-party sessions, it defeats the purpose of rapid acquisition. Furthermore, when RTP receivers delay their messages requesting a burst by a deterministic or even a random amount, it still does not make a difference since such messages are not related and their timings are independent from each other. Thus, in this specification we initialize Tmin to zero and allow the RTP receivers to send a burst request message right at the beginning. It should, however, be emphasized that for the subsequent messages during rapid acquisition, the timing rules of [RFC4585] still apply.

Figure 3 depicts an example of messaging flow for rapid acquisition. The RTCP feedback messages are explained below. The optional messages are indicated in parentheses and they may or may not be present during rapid acquisition. Note that in a scenario where rapid acquisition is performed by a feedback target co-located with the media sender, the same method (with the identical message flows) still applies.

```
                  -------------------------
                 | Retransmission  Server  |
    -----------  |  ------   ------------  |   --------    ------------
   | Multicast | | |  FT  | | Burst/Ret. | |  |        |  |    RTP     |
   |  Source   | | |      | |   Source   | |  | Router |  |  Receiver  |
   |           | |  ------   ------------  |  |        |  |  (RTP_Rx)  |
    -----------  |      |         |        |   --------    ------------
     |            -------------------------       |                |
     |                  |         |               |                |
     |-- RTP Multicast ---------->--------------->|                |
     |-. RTCP Multicast -.-.-.-.->-.-.-.-.-.-.-.->|                |
     |                  |         |               |                |
     |                  |         |********************************|
```

```
     |                  |         |*      PORT MAPPING SETUP      *|
     |                  |         |********************************|
     |                  |         |               |                |
     |                  |<~~~~~~~~~~~~~~~~~~~~~~~~~ RTCP RAMS-R ~~~|
     |                  |         |               |                |
     |                  |         |********************************|
     |                  |         |* UNICAST SESSION  ESTABLISHED *|
     |                  |         |********************************|
     |                  |         |               |                |
     |                  |         |~~~ RTCP RAMS-I ~~~~~~~~~~~~~~~>|
     |                  |         |               |                |
     |                  |         |... Unicast RTP Burst .........>|
     |                  |         |               |                |
     |                  |<~~~~~~~~~~~~~~~~~~~~~~~~ (RTCP RAMS-R) ~~|
     |                  |         |               |                |
     |                  |         |~~ (RTCP RAMS-I) ~~~~~~~~~~~~~~>|
     |                  |         |               |                |
     |                  |         |               |                |
     |                  |         |               |<= SFGMP Join ==|
     |                  |         |               |                |
     |-- RTP Multicast ------------------------------------------->|
     |-. RTCP Multicast -.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.-.>|
     |                  |         |               |                |
     |                  |         |               |                |
     |                  |         |<~~~~~~~~~~~~~~~ RTCP RAMS-T ~~~|
     |                  |         |               |                |
     |                  |         |               |                |
     |                  |<~~~~~~~~~~~~~~~~~~~~~~~~~~ (RTCP NACK) ~~|
     |                  |         |               |                |
     |                  |         |               |                |
     |                  |         |...(Unicast Retransmissions)...>|
     |                  |         |               |                |
     :                  :         :               :                :
     :                  :         :               :                :
     |                  |         |<.=.= Unicast RTCP Reports .=.=>|
     :                  :         :               :                :
     :                  :         :               :                :
     |                  |         |               |                |


   -------> Multicast RTP Flow
   .-.-.-.> Multicast RTCP Flow
   .=.=.=.> Unicast RTCP Reports
   ~~~~~~~> Unicast RTCP Feedback Messages
   =======> SFGMP Messages
   .......> Unicast RTP Flow
```

Figure 3: Message flows for unicast-based rapid acquisition

This document defines the expected behaviors of RS and RTP_Rx. It is instructive to have independently operating implementations on RS and RTP_Rx that request the burst, describe the burst, start the burst, join the multicast session and stop the burst. These implementations send messages to each other, but there must be provisions for the cases where the control messages get lost, or re-ordered, or are not being delivered to their destinations.

The following steps describe rapid acquisition in detail:

1. Port Mapping Setup: For the primary multicast RTP session, the RTP and RTCP destination ports are declaratively specified (Refer to Section 8 for examples in SDP). However, in the unicast RTP retransmission session, RTP_Rx often needs to choose its receive ports for RTP and RTCP. Since this unicast session is established after RTP_Rx sends its rapid acquisition request and it is received by RS in the primary multicast RTP session, RTP_Rx MUST setup the port mappings between the unicast and multicast sessions and send this mapping information to RS before it sends its request so that RS knows how to communicate with RTP_Rx.

   The details of this setup procedure and other NAT-related issues are left to Section 9 to keep the present discussion focused on the RAMS message flows.

2. Request: RTP_Rx sends a rapid acquisition request for the primary multicast RTP session to the feedback target address of that session. The request contains the SSRC identifier of RTP_Rx and may contain the media sender SSRC identifier(s) associated with the desired primary multicast stream(s). This RTCP feedback message is defined as the RAMS-Request (RAMS-R) message and may contain parameters that constrain the burst, such as the buffer and bandwidth limits.

   Before joining the SSM session, RTP_Rx learns the addresses for the multicast source, group and RS by out-of-band means. If RTP_Rx desires to rapidly acquire only a subset of the primary multicast streams available in the primary multicast RTP session, the SSRC identifiers for the desired RTP streams MUST also be obtained out-of-band, since no RTP packets have been received yet for those streams. Based on this information, RTP_Rx populates the desired SSRC(s) in its request message.

   When RS successfully receives the RAMS-R message, it responds to it by accepting or rejecting the request. Right before RS sends any RTP or RTCP packet(s) described below, it establishes the unicast RTP retransmission session.

3. Response: RS sends RAMS-Information (RAMS-I) message(s) to RTP_Rx to convey the status for the burst(s) requested by RTP_Rx. The RAMS-I message is sent by the Burst/Retransmission Source logical entity that is part of RS.

   In cases where the primary multicast RTP session associated with FT_Ap on which the RAMS-R message was received contains only a single primary multicast stream, RS SHALL always use the SSRC of the RTP stream associated with FT_Ap in the RAMS-I message(s) regardless of the media sender SSRC specified in the RAMS-R message. In such cases the 'ssrc' attribute MAY be omitted from the media description. If the requested SSRC and the actual media sender SSRC do not match, RS SHOULD explicitly populate the correct media sender SSRC in the initial RAMS-I message.

   FT_Ap could also be associated with an RTP session that carries two or more primary multicast streams. If RTP_Rx will issue a collective request to receive the whole primary multicast RTP session, it does not need the 'ssrc' attributes to be described in the media description. Note that if FT_Ap is associated with two or more RTP sessions, RTP_Rx's request will be ambiguous. Thus, each FT_Ap MUST be associated with a single RTP session.

   If RTP_Rx is willing to rapidly acquire only a subset of the primary multicast streams, the RAMS-R message MUST explicitly list the media sender SSRCs. Upon receiving such a message, RS MAY accept the request for only the media sender SSRC(s) that matched one of the RTP streams it serves. It MUST reject all other requests with the appropriate response code.

   * Reject Responses: RS MUST take into account any limitations that MAY have been specified by RTP_Rx in the RAMS-R message when making a decision regarding the request. If RTP_Rx has requested to acquire the whole primary multicast RTP session but RS cannot provide a rapid acquisition service for any of the primary multicast streams, RS MUST reject the request via a single RAMS-I message with a collective reject response code and whose media sender SSRC field is set to one of SSRCs served by this FT_Ap. Upon receiving this RAMS-I message, RTP_Rx abandons the rapid acquisition attempt and may immediately join the multicast session by sending an SFGMP Join message towards its upstream multicast router.

     In all other cases, RS MUST send a separate RAMS-I message with the appropriate response code for each primary multicast stream that has been requested by RTP_Rx but cannot be served by RS.

* Accept Responses: RS MUST send a separate RAMS-I message with the appropriate response code for each primary multicast stream that has been requested by RTP_Rx and will be served by RS. Such RAMS-I messages comprise fields that can be used to describe the individual unicast burst streams.

  A particularly important field carries the RTP sequence number of the first packet transmitted in the respective RTP stream to allow RTP_Rx to detect any missing initial packet(s). Note that the first RTP packet transmitted in an RTP stream is not necessarily a burst packet. It could be a payload-specific RTP packet, which is payload-type-multiplexed with the burst packets (See [I-D.begen-avt-rtp-mpeg2ts-preamble] for an example). When RS accepts the request, this field MUST be populated in the RAMS-I message and the initial RAMS-I message SHOULD precede the unicast burst or be sent at the start of the burst so that RTP_Rx may quickly detect any missing initial packet(s).

Where possible, it is RECOMMENDED to include all RAMS-I messages in the same compound RTCP packet. However, it is possible that the RAMS-I message for a primary multicast stream may get delayed or lost, and RTP_Rx may start receiving RTP packets before receiving a RAMS-I message. Thus, RTP_Rx SHOULD NOT make protocol dependencies on quickly receiving the initial RAMS-I message. For redundancy purposes, it is RECOMMENDED that RS repeats the RAMS-I messages multiple times as long as it follows the RTCP timer rules defined in [RFC4585].

4. Unicast Burst: For the primary multicast stream(s) for which the request is accepted, RS starts sending the unicast burst(s) that comprises one or more RTP retransmission packets. The burst packet(s) are sent by the Burst/Retransmission Source logical entity. In addition, there MAY be optional payload-specific information that RS chooses to send to RTP_Rx. Such an example is discussed in [I-D.begen-avt-rtp-mpeg2ts-preamble] for transmitting the payload-specific information for MPEG2 Transport Streams.

5. Updated Request: RTP_Rx MAY send an updated RAMS-R message (to the FT entity of RS) with a different value for one or more fields of an earlier RAMS-R message. Upon receiving an updated request, RS may use the updated values for sending/shaping the burst, or refine the values and use the refined values for sending/shaping the burst. Subsequently, RS MAY send an updated RAMS-I message to indicate the changes it made.

However, the updated RAMS-I message may get lost. It is also possible that the updated RAMS-R message could have been lost. Thus, RTP_Rx SHOULD NOT make protocol dependencies on quickly (or ever) receiving an updated RAMS-I message, or assume that RS will honor the requested changes.

RTP_Rx may be in an environment where the available resources are time-varying, which may or may not deserve sending a new updated request. Determining the circumstances where RTP_Rx should or should not send an updated request and the methods that RTP_Rx can use to detect and evaluate the time-varying available resources are not specified in this document.

6. Updated Response: RS may send more than one RAMS-I messages, e.g., to update the value of one or more fields in an earlier RAMS-I message. The updated RAMS-I messages may or may not be a direct response to a RAMS-R message. RS may also send updated RAMS-I messages to signal RTP_Rx in real time to join the multicast session. RTP_Rx depends on RS to learn the join time, which can be conveyed by the first RAMS-I message, or can be sent/revised in a later RAMS-I message. If RS is not capable of determining the join time in the initial RAMS-I message, it MUST send another RAMS-I message (with the join time information) later.

7. Multicast Join Signaling: The RAMS-I message allows RS to signal explicitly when RTP_Rx SHOULD send the SFGMP Join message. If the request is accepted, this information MUST be conveyed in at least one RAMS-I message and its value MAY be updated by subsequent RAMS-I messages. If RTP_Rx has received multiple RAMS-I messages, it SHOULD use the information from the most recent RAMS-I message.

   There may be missing packets if RTP_Rx joins the multicast session too early or too late. For example, if RTP_Rx starts receiving the primary multicast stream while it is still receiving the unicast burst at a high excess bitrate, this may result in an increased risk of packet loss. Or, if RTP_Rx joins the multicast session some time after the unicast burst is finished, there may be a gap between the burst and multicast data (a number of RTP packets may be missing). In both cases, RTP_Rx may issue retransmissions requests (via RTCP NACK messages) [RFC4585] to the FT entity of RS to fill the gap. RS may or may not respond to such requests. When it responds and the response causes significant changes in one or more values reported earlier to RTP_Rx, an updated RAMS-I should be sent to RTP_Rx.

8. Multicast Receive: After the join, RTP_Rx starts receiving the primary multicast stream(s).

9. Terminate: RS may know when it needs to ultimately stop the unicast burst based on its parameters. However, RTP_Rx may need to ask RS to terminate the burst prematurely or at a specific sequence number. For this purpose, it uses the RAMS-Termination (RAMS-T) message. A separate RAMS-T message is sent for each primary multicast stream served by RS unless an RTCP BYE message has been sent as described in Step 10. For the burst requests that were rejected by RS, there is no need to send a RAMS-T message.

   If RTP_Rx wants to terminate a burst prematurely, it SHALL send a plain RAMS-T message for the particular primary multicast stream, and upon receiving this message RS MUST terminate the unicast burst. If RTP_Rx requested to acquire the entire primary multicast RTP session but wants to terminate this request before it learns the individual media sender SSRC(s) via RAMS-I message(s), it cannot use RAMS-T message(s) and thus MUST send an RTCP BYE message to terminate the request.

   Otherwise, the default behavior for RTP_Rx is to send a RAMS-T message right after it joined the multicast session and started receiving multicast packets. In that case, RTP_Rx SHALL send a RAMS-T message with the sequence number of the first RTP packet received in the primary multicast stream, and RS SHOULD terminate the respective burst after it sends the unicast burst packet whose Original Sequence Number (OSN) field in the RTP retransmission payload header matches this number minus one.

   RTP_Rx MUST send at least one RAMS-T message for each primary multicast stream served by RS (if an RTCP BYE message has not been issued yet as described in Step 10). The RAMS-T message(s) MUST be addressed to the Burst/Retransmission Source logical entity. Against the possibility of a message loss, it is RECOMMENDED that RTP_Rx repeats the RAMS-T messages multiple times as long as it follows the RTCP timer rules defined in [RFC4585].

10. Terminate with RTCP BYE: When RTP_Rx is receiving one or more burst streams, if RTP_Rx becomes no longer interested in acquiring any of the primary multicast streams, RTP_Rx SHALL issue an RTCP BYE message for the RTP retransmission session and another RTCP BYE message for the primary multicast RTP session. These RTCP BYE messages are sent to the Burst/Retransmission Source and FT logical entities, respectively.

Upon receiving an RTCP BYE message, the Burst/Retransmission Source logical entity MUST terminate the rapid acquisition operation, and cease transmitting any further burst packets and retransmission packets. If support for [RFC5506] has been signaled, the RTCP BYE message MAY be sent in a reduced-size RTCP packet. Otherwise, Section 6.1 of [RFC3550] mandates the RTCP BYE message always to be sent with a sender or receiver report in a compound RTCP packet (If no data has been received, an empty receiver report MUST be still included). With the information contained in the receiver report, RS can figure out how many duplicate RTP packets have been delivered to RTP_Rx (Note that this will be an upper-bound estimate as one or more packets might have been lost during the burst transmission). The impact of duplicate packets and measures that can be taken to minimize the impact of receiving duplicate packets will be addressed in Section 6.4.

Note that an RTCP BYE message issued for the RTP retransmission session terminates the whole session and ceases transmitting any further packets in that RTP session. Thus, in this case there is no need for sending explicit RAMS-T messages, which would only terminate their respective bursts.

For the purpose of gathering detailed information about RTP_Rx's rapid acquisition experience, [I-D.begen-avt-rapid-sync-rtcp-xr] defines an RTCP Extended Report (XR) Block. This report is designed to be payload-independent, thus, it can be used by any multicast application that supports rapid acquisition. Support for this XR report is, however, OPTIONAL.

6.3. Synchronization of Primary Multicast Streams

When RTP_Rx acquires multiple primary multicast streams, it may need to synchronize them for the playout. This synchronization is traditionally achieved by the help of the RTCP sender reports [RFC3550]. If the playout will start before RTP_Rx has joined the multicast session, RTP_Rx must receive the information reflecting the synchronization among the primary multicast streams early enough so that it can play out the media in a synchronized fashion. However, this would require RS to cache the sender reports sent in the primary multicast RTP session(s), and piggyback the latest synchronization information on its own sender report and send an early sender report in the unicast RTP retransmission session. This issue and its implications are discussed in detail in [I-D.ietf-avt-rapid-rtp-sync].

An alternative approach is to use the RTP header extension mechanism [RFC5285] and convey the synchronization information in a header

extension as defined in [I-D.ietf-avt-rapid-rtp-sync].

[RFC4588] says that retransmission packets SHOULD carry the same header extension carried in the header of the original RTP packets. Thus, as long as the multicast source emits a header extension with the synchronization information frequently enough, there is no additional task that needs to be carried out by RS. The synchronization information will be sent to RTP_Rx along with the burst packets. The frequent header extensions sent in the primary multicast RTP sessions also allow rapid synchronization of the RTP streams for the RTP receivers that do not support RAMS or that directly join the multicast session without running RAMS. Thus, in RAMS applications, it is RECOMMENDED that the multicast sources frequently send synchronization information by using header extensions following the rules presented in [I-D.ietf-avt-rapid-rtp-sync]. It should be noted that the regular sender reports are still sent in the unicast session by following the rules of [RFC3550].

6.4. Burst Shaping and Congestion Control in RAMS

This section provides informative guidelines about how RS can shape the transmission of the unicast burst and how congestion can be dealt within the RAMS process.

A higher bitrate for the unicast burst naturally conveys the Reference Information and media content to RTP_Rx faster. This way, RTP_Rx can start consuming the data sooner, which results in a faster acquisition. A higher bitrate also represents a better utilization of RS resources. As the burst may continue until it catches up with the primary multicast stream, the higher the bursting bitrate, the less data RS needs to transmit. However, a higher bitrate for the burst also increases the chances for congestion-caused packet loss. Thus, as discussed in Section 5, there must be an upper bound on the bandwidth used by the burst.

When RS transmits the burst, it should take into account all available information to prevent any packet loss that may take place during the bursting as a result of buffer overflow on the path between RS and RTP_Rx and at RTP_Rx itself. The bursting bitrate may be determined by taking into account the following information, when available:

a. Information obtained via the RAMS-R message, such as Max RAMS Buffer Fill Requirement and/or Max Receive Bitrate (See Section 7.2).

b. Information obtained via RTCP receiver reports provided by RTP_Rx in the retransmission session, allowing in-session bitrate adaptations for the burst. When these receiver reports indicate packet loss, this may indicate a certain congestion state in the path from RS to RTP_Rx.

c. Information obtained via RTCP NACKs provided by RTP_Rx in the primary multicast RTP session, allowing in-session bitrate adaptations for the burst. Such RTCP NACKs are transmitted by RTP_Rx in response to packet loss detection in the burst. NACKs may indicate a certain congestion state on the path from RS to RTP_Rx.

d. There may be other feedback received from RTP_Rx, e.g., in the form of ECN-CE markings [I-D.westerlund-avt-ecn-for-rtp] that may influence in-session bitrate adaptation.

e. Information obtained via updated RAMS-R messages, allowing in-session bitrate adaptations, if supported by RS.

f. Transport protocol-specific information. For example, when DCCP is used to transport the RTP burst, the ACKs from the DCCP client can be leveraged by the RS / DCCP server for burst shaping and congestion control.

g. Pre-configured settings for each RTP_Rx or a set of RTP_Rxs that indicate the upper-bound bursting bitrates for which no packet loss will occur as a result of congestion along the path of RS to RTP_Rx. For example, in managed IPTV networks, where the bottleneck bandwidth along the end-to-end path is known and where the network between RS and this link is provisioned and dimensioned to carry the burst streams, the bursting bitrate does not exceed the provisioned value. These settings may also be dynamically adapted using application-aware knowledge.

RS chooses the initial burst bitrate as follows:

o When using RAMS in environments as described in (g), RS MUST transmit the burst packets at an initial bitrate higher than the nominal bitrate, but within the engineered or reserved bandwidth limit.

o When RS cannot determine a reliable bitrate value for the unicast burst (through a or g), RS should choose an appropriate initial bitrate not above the nominal bitrate and increase it gradually unless a congestion is detected.

In both cases, during the burst transmission RS MUST continuously

monitor for packet losses as a result of congestion by means of one or more of the mechanisms described in (b,c,d,e,f). When RS relies on RTCP receiver reports, sufficient bandwidth must be provided to RTP Rx for RTCP transmission. To achieve a reasonable fast adaptation against congestion, it is recommended that RTP_Rx sends a receiver report at least once every two RTTs between RS and RTP_Rx. Although the specific heuristics and algorithms that deduce a congestion state and how subsequently RS should act are outside the scope of this specification, the following two practices are recommended:

o Upon detection of a significant packet loss, which RS attributes to congestion, RS should decrease the burst bitrate. The rate by which RS increases and decreases the bitrate for the burst may be determined by a TCP-friendly bitrate adaptation algorithm for RTP over UDP , or in the case of (f) by the congestion control algorithms defined in DCCP [I-D.ietf-dccp-rtp].

o If the congestion is persistent and RS has to reduce the burst bitrate to a point where the RTP Rx buffer may underrun or the burst will consume too much RS resources, RS should terminate the burst and transmit a RAMS-I message to RTP Rx with the appropriate response code. It is then up to RTP Rx to decide when to join the multicast session.

In case there is no congestion experienced during the burst, a specific situation occurs near the end of the unicast burst, when RS has almost no more additional data to sustain the relatively higher burst bitrate, thus, the upper-bound burst bitrate automatically gets limited by the nominal bitrate of the primary multicast stream. During this time frame, RTP_Rx eventually needs to join the multicast session. This means that both the burst packets and the multicast packets may be simultaneously received by RTP_Rx for a period of time, enhancing the risk of congestion again.

Since RS signals RTP_Rx when it should send the SFGMP Join message, RS may have a rough estimate of when RTP_Rx will start receiving multicast packets in the SSM session. RS may keep on sending burst packets but should reduce the bitrate accordingly at the appropriate instant by taking the bitrate of the whole SSM session into account. If RS ceases transmitting the burst packets before the burst catches up, any gap resulting from this imperfect switch-over by RTP_Rx can be later repaired by requesting retransmissions for the missing packets from RS. The retransmissions may be shaped by RS to make sure that they do not cause collateral loss in the primary multicast RTP session and the RTP retransmission session.

6.5. Failure Cases

In the following, we examine the implications of losing the RAMS-R, RAMS-I or RAMS-T messages and other failure cases.

When RTP_Rx sends a RAMS-R message to initiate a rapid acquisition but the message gets lost and RS does not receive it, RTP_Rx will get neither a RAMS-I message, nor a unicast burst. In this case, RTP_Rx MAY resend the request when it is eligible to do so based on the RTCP timer rules defined in [RFC4585]. Or, after a reasonable amount of time, RTP_Rx may time out (based on the previous observed response times) and immediately join the SSM session.

In the case RTP_Rx starts receiving a unicast burst but it does not receive a corresponding RAMS-I message within a reasonable amount of time, RTP_Rx may either discard the burst data or decide not to interrupt the unicast burst, and be prepared to join the SSM session at an appropriate time it determines or as indicated in a subsequent RAMS-I message (if available). To minimize the chances of losing the RAMS-I messages, it is RECOMMENDED that RS repeats the RAMS-I messages multiple times based on the RTCP timer rules defined in [RFC4585].

In the failure cases where the RAMS-R message is lost and RTP_Rx gives up, or the RAMS-I message is lost, RTP_Rx MUST still terminate the burst(s) it requested by following the rules described in Section 6.2.

In the case a RAMS-T message sent by RTP_Rx does not reach its destination, RS may continue sending burst packets even though RTP_Rx no longer needs them. In such cases, it is RECOMMENDED that RTP_Rx repeats the RAMS-T message multiple times based on the RTCP timer rules defined in [RFC4585]. In the worst case, RS MUST be provisioned to deterministically terminate the burst when it can no longer send the burst packets faster than it receives the primary multicast stream packets.

Section 6.3.5 of [RFC3550] explains the rules pertaining to timing out an SSRC. When RS accepts to serve the requested burst(s) and establishes the retransmission session, it should check the liveness of RTP_Rx via the RTCP messages and reports RTP_Rx sends. The default rules explained in [RFC3550] apply in RAMS as well.

7. Encoding of the Signaling Protocol in RTCP

This section defines the formats of the RTCP transport-layer feedback messages that are exchanged between the Retransmission Server (RS)

and RTP Receiver (RTP_Rx) during rapid acquisition. These messages are referred to as the RAMS Messages. They are payload-independent and MUST be used by all RTP-based multicast applications that support rapid acquisition regardless of the payload they carry.

Payload-specific feedback messages are not defined in this document. However, further optional payload-independent and payload-specific information can be included in the exchange.

The common packet format for the RTCP feedback messages is defined in Section 6.1 of [RFC4585]. Each feedback message has a fixed-length field for version, padding, feedback message type (FMT), payload type (PT), length, SSRC of packet sender, SSRC of media sender as well as a variable-length field for feedback control information (FCI).

In the RAMS messages, the PT field is set to RTPFB (205) and the FMT field is set to RAMS (6). Individual RAMS messages are identified by a sub-field called Sub Feedback Message Type (SFMT). Any Reserved field SHALL be set to zero and ignored.

Depending on the specific scenario and timeliness/importance of a RAMS message, it may be desirable to send it in a reduced-size RTCP packet [RFC5506]. However, unless support for [RFC5506] has been signaled, compound RTCP packets MUST be used by following [RFC3550] rules.

Following the rules specified in [RFC3550], all integer fields in the messages defined below are carried in network-byte order, that is, most significant byte (octet) first, also known as big-endian. Unless otherwise noted, numeric constants are in decimal (base 10).

7.1. Extensions

To improve the functionality of the RAMS method in certain applications, it may be desirable to define new fields in the RAMS Request, Information and Termination messages. Such fields MUST be encoded as TLV elements as described below and sketched in Figure 4:

- o Type: A single-octet identifier that defines the type of the parameter represented in this TLV element.

- o Length: A two-octet field that indicates the length (in octets) of the TLV element excluding the Type and Length fields, and the 8-bit Reserved field between them. Note that this length does not include any padding that is required for alignment.

- o Value: Variable-size set of octets that contains the specific value for the parameter.

In the extensions, the Reserved field SHALL be set to zero and ignored. If a TLV element does not fall on a 32-bit boundary, the last word MUST be padded to the boundary using further bits set to zero.

In a RAMS message, any vendor-neutral or private extension MUST be placed after the mandatory fields (if any). The extensions MAY be placed in any order. The support for extensions is OPTIONAL.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Reserved    |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:                             Value                             :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 4: Structure of a TLV element

7.1.1. Vendor-Neutral Extensions

If the goal in defining new TLV elements is to extend the functionality in a vendor-neutral manner, they MUST be registered with IANA through the guidelines provided in Section 11.5.

The current document defines several vendor-neutral extensions in the subsequent sections.

7.1.2. Private Extensions

It is desirable to allow vendors to use private extensions in a TLV format. For interoperability, such extensions MUST NOT collide with each other.

A certain range of TLV Types (between - and including - 128 and 254) is reserved for private extensions (Refer to Section 11.5). IANA management for these extensions is unnecessary and they are the responsibility of individual vendors.

The structure that MUST be used for the private extensions is depicted in Figure 5. Here, the enterprise numbers are used from http://www.iana.org/assignments/enterprise-numbers. This will ensure the uniqueness of the private extensions and avoid any collision.

```
      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
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |      Type     |   Reserved    |             Length            |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |                       Enterprise Number                       |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     :                             Value                             :
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 5: Structure of a private extension

7.2. RAMS Request

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.

Unless signaled otherwise, a RAMS-R message is used to request a single primary multicast stream whose SSRC is indicated in the media sender SSRC field of the message header. In cases where RTP_Rx does not know the media sender SSRC, it MUST set that field to its own SSRC.

If RTP_Rx wants to request two or more primary multicast streams or all of the streams in the primary multicast RTP session, RTP_Rx MUST provide explicit signaling as described below and set the media sender SSRC field to its own SSRC to minimize the chances of accidentally requesting a wrong primary multicast stream.

The FCI field MUST contain only one RAMS Request. The FCI field has the structure depicted in Figure 6.

```
      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
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |    SFMT=1     |                    Reserved                   |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     :      Optional TLV-encoded Fields (and Padding, if needed)     :
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 6: FCI field syntax for the RAMS Request message

- Requested Media Sender SSRC(s): Optional TLV element that lists the media sender SSRC(s) requested by RTP_Rx. If this TLV element does not exist in the RAMS-R message, it means that RTP_Rx is only

interested in a single primary multicast stream whose media sender SSRC is already specified in the header of the RAMS-R message. However, if this TLV element exists, RS MUST ignore the media sender SSRC specified in the header of the RAMS-R message. If this TLV element exists but the Length field is set to zero, meaning that no media sender SSRC is listed, it means that RTP_Rx is requesting to rapidly acquire the entire primary multicast RTP session. Otherwise, RTP_Rx lists the individual media sender SSRCs in this TLV element and sets the Length field of the TLV element to 4*n, where n is the number of SSRC entries.

Type: 1

- Min RAMS Buffer Fill Requirement (32 bits): Optional TLV element that denotes the minimum milliseconds of data that RTP_Rx desires to have in its buffer before allowing the data to be consumed by the application.

  RTP_Rx may have knowledge of its buffering requirements. These requirements may be application and/or device specific. For instance, RTP_Rx may need to have a certain amount of data in its application buffer to handle transmission jitter and/or to be able to support error-control methods. If RS is told the minimum buffering requirement of the receiver, it may tailor the burst(s) more precisely, e.g., by choosing an appropriate starting point. The methods used by RTP_Rx to determine this value are application specific, and thus, out of the scope of this document.

  If specified, the amount of backfill that will be provided by the unicast bursts and any payload-specific information MUST NOT be smaller than the specified value since it will not be able to build up the desired level of buffer at RTP_Rx and may cause buffer underruns.

  Type: 2

- Max RAMS Buffer Fill Requirement (32 bits): Optional TLV element that denotes the maximum milliseconds of data that RTP_Rx can buffer without losing the data due to buffer overflow.

  RTP_Rx may have knowledge of its buffering requirements. These requirements may be application or device specific. For instance, one particular RTP_Rx may have more physical memory than another RTP_Rx, and thus, can buffer more data. If RS knows the buffering ability of the receiver, it may tailor the burst(s) more precisely. The methods used by the receiver to determine this value are application specific, and thus, out of scope.

If specified, the amount of backfill that will be provided by the unicast bursts and any payload-specific information MUST NOT be larger than this value since it may cause buffer overflows at RTP_Rx.

Type: 3

- Max Receive Bitrate (64 bits): Optional TLV element that denotes the maximum bitrate (in bits per second) that the RTP receiver can process the aggregation of the unicast burst(s) and any payload-specific information that will be provided by RS. The limits may include local receiver limits as well as network limits that are known to the receiver.

  If specified, the total bitrate of the unicast burst(s) plus any payload-specific information MUST NOT be larger than this value since it may cause congestion and packet loss.

  Type: 4

- Request for Preamble Only (0 bits): Optional TLV element that indicates that RTP_Rx is only requesting the preamble information for the desired primary multicast stream(s). If this TLV element exists in the RAMS-R message, RS SHOULD NOT send any burst packets other than the preamble packets. Note that this TLV element does not carry a Value field. Thus, the Length field MUST be set to zero.

  Type: 5

The semantics of the RAMS-R feedback message is independent of the payload type.

7.3. RAMS Information

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.

A separate RAMS-I message with the appropriate media sender SSRC and response code is sent by RS for each primary multicast stream that has been requested by RTP_Rx.

The FCI field MUST contain only one RAMS Information. The FCI field has the structure depicted in Figure 7.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    SFMT=2     |      MSN      |           Response            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
:      Optional TLV-encoded Fields (and Padding, if needed)     :
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 7: FCI field syntax for the RAMS Information message

- o Message Sequence Number (8 bits) : Mandatory field that denotes the sequence number of the RAMS-I message for the particular media sender SSRC specified in the message header. The MSN value SHALL be set to zero only when a new RAMS request is received. During rapid acquisition, the same RAMS-I message MAY be repeated for redundancy purposes without incrementing the MSN value. If an updated RAMS-I message will be sent (either with a new information or an updated information), the MSN value SHALL be incremented by one. In the MSN field, the regular wrapping rules apply.

- o Response (16 bits): Mandatory field that denotes the RS response code for this RAMS-I message. This document defines several initial response codes and registers them with IANA. If a new vendor-neutral response code will be defined, it MUST be registered with IANA through the guidelines specified in Section 11.6. If the new response code is intended to be used privately by a vendor, there is no need for IANA management. Instead, the vendor MUST use the private extension mechanism (Section 7.1.2) to convey its message and MUST indicate this by putting zero in the Response field.

- o Media Sender SSRC (32 bits): Optional TLV element that specifies the media sender SSRC of the unicast burst stream. While this information is already available in the message header, it may be useful to repeat it in an explicit field. For example, if FT_Ap that received the RAMS-R message is associated with a single primary multicast stream but the requested media sender SSRC does not match the SSRC of the RTP stream associated with this FT_Ap, RS SHOULD include this TLV element in the initial RAMS-I message to let RTP_Rx know that the media sender SSRC has changed. If the two SSRCs match, there is no need to include this TLV element.

  Type: 31

- o RTP Seqnum of the First Packet (16 bits): TLV element that specifies the RTP sequence number of the first packet that will be sent in the respective RTP stream. This allows RTP_Rx to know whether one or more packets sent by RS have been dropped at the

beginning of the stream. If RS accepts the RAMS request, this element MUST exist. If RS rejects the RAMS request, this element SHALL NOT exist.

Type: 32

- Earliest Multicast Join Time (32 bits): TLV element that specifies the delta time (in ms) between the arrival of the first RTP packet in the RTP stream (which could be a burst packet or a payload-specific packet) and the earliest time instant when RTP_Rx SHOULD send an SFGMP Join message to join the multicast session. A zero value in this field means that RTP_Rx may send the SFGMP Join message right away.

  If the RAMS request has been accepted, RS MUST send this field at least once, so that RTP_Rx knows when to join the multicast session. If the burst request has been rejected as indicated in the Response field, this field MAY be omitted or set to zero. In that case, it is up to RTP_Rx when or whether to join the multicast session.

  It should be noted that when RS serves two or more bursts and sends a separate RAMS-I message for each burst, the join times specified in these RAMS-I messages should correspond to more or less the same time instant, and RTP_Rx sends the SFGMP Join message based on the earliest join time.

  Type: 33

- Burst Duration (32 bits): Optional TLV element that denotes the duration of the burst, i.e., the delta difference between the first and the last burst packet, that RS is planning to send (in ms) in the respective RTP stream. In the absence of additional stimulus, RS will send a burst of this duration. However, the burst duration may be modified by subsequent events, including changes in the primary multicast stream and reception of RAMS-T messages.

  Note that RS MUST terminate the flow in a deterministic timeframe, even if it does not get a RAMS-T or a BYE from RTP_Rx. It is OPTIONAL to send this field in a RAMS-I message when the burst request is accepted. If the burst request has been rejected as indicated in the Response field, this field MAY be omitted or set to zero.

  Type: 34

o Max Transmit Bitrate (64 bits): Optional TLV element that denotes the maximum bitrate (in bits per second) that will be used by RS for the RTP stream associated with this RAMS-I message.

Type: 35

The semantics of the RAMS-I feedback message is independent of the payload type.

The initial RAMS-I message SHOULD precede the unicast burst or be sent at the start of the burst. Subsequent RAMS-I message(s) MAY be sent during the unicast burst and convey changes in any of the fields.

7.4. RAMS Termination

The RAMS Termination message is identified by SFMT=3.

The RAMS Termination is used to assist RS in determining when to stop the burst. A separate RAMS-T message is sent by RTP_Rx for each primary multicast stream that has been served by RS. Each of these RAMS-T messages has the appropriate media sender SSRC populated in its message header.

If RTP_Rx wants RS to stop a burst prematurely, it sends a plain RAMS-T message as described below. Upon receiving this message, RS stops the respective burst immediately. If RTP_Rx wants RS to terminate all of the bursts, it should send all of the respective RAMS-T messages in a single compound RTCP packet.

The default behavior for RTP_Rx is to send a RAMS-T message right after it joined the multicast session and started receiving multicast packets. In that case, RTP_Rx includes the sequence number of the first RTP packet received in the primary multicast stream in the RAMS-T message. With this information, RS can decide when to terminate the unicast burst.

The FCI field MUST contain only one RAMS Termination. The FCI field has the structure depicted in Figure 8.

```
      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
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |    SFMT=3     |                   Reserved                    |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     :      Optional TLV-encoded Fields (and Padding, if needed)     :
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 8: FCI field syntax for the RAMS Termination message

- Extended RTP Seqnum of First Multicast Packet (32 bits): Optional TLV element that specifies the extended RTP sequence number of the first packet received from the SSM session for a particular primary multicast stream. The low 16 bits contain the sequence number of the first packet received from the SSM session, and the most significant 16 bits extend that sequence number with the corresponding count of sequence number cycles, which may be maintained according to the algorithm in Appendix A.1 of [RFC3550].

  Type: 61

The semantics of the RAMS-T feedback message is independent of the payload type.

8. SDP Signaling

8.1. Definitions

The syntax of the 'rtcp-fb' attribute has been defined in [RFC4585]. Here we add the following syntax to the 'rtcp-fb' attribute (the feedback type and optional parameters are all case sensitive):

(In the following ABNF [RFC5234], fmt, SP and CRLF are used as defined in [RFC4566].)

```
      rtcp-fb-syntax = "a=rtcp-fb:" rtcp-fb-pt SP rtcp-fb-val CRLF

      rtcp-fb-pt         = "*"   ; wildcard: applies to all formats
                         / fmt   ; as defined in SDP spec

      rtcp-fb-val        = "nack" SP "ssli"
```

The following parameter is defined in this document for use with 'nack':

- 'ssli' stands for Stream Synchronization Loss Indication and indicates the use of RAMS messages as defined in Section 7.

This document also defines a new media-level SDP attribute ('rams-updates') that indicates whether RS supports updated request messages or not. This attribute is used in a declarative manner. If RS supports updated request messages and this attribute is included in the SDP description, RTP_Rx may send updated requests. RS may or may not be able to accept value changes in every field in an updated RAMS-R message. However, if the 'rams-updates' attribute is not included in the SDP description, RTP_Rx SHALL NOT send updated requests (RTP_Rx MAY still repeat its initial request without changes, though).

8.2. Requirements

The use of SDP to describe the RAMS entities normatively requires the support for:

- The SDP grouping semantics [I-D.ietf-mmusic-rfc3388bis]

- The RTP/AVPF profile [RFC4585]

- The RTP retransmissions [RFC4588]

- The RTCP extensions for SSM sessions with unicast feedback [I-D.ietf-avt-rtcpssm]

The support for the source-specific media attributes [RFC5576] may be required in some deployments as described below.

8.3. Examples

This section provides a declarative SDP [RFC4566] example for enabling rapid acquisition of multicast RTP sessions.

In the example shown Figure 9, we have a primary multicast (source) stream and a unicast retransmission stream. The source stream is multicast from a distribution source (with a source IP address of 198.51.100.1) to the multicast destination address of 233.252.0.2 and port 41000. A Retransmission Server including feedback target functionality (with an address of 192.0.2.1 and port of 41001) is specified with the 'rtcp' attribute. The RTP receiver(s) can report missing packets on the source stream to the feedback target and request retransmissions. In the RAMS context, the parameter 'rtx-time' specifies the time in milliseconds that the Retransmission Server keeps an RTP packet in its cache available for retransmission (measured from the time the packet was received by the Retransmission

Server).

In this example, both the conventional retransmission and rapid acquisition support are enabled. This is achieved by the "a=rtcp-fb:98 nack ssli" line. Note that this SDP includes the "a=sendonly" line for the media description of the retransmission stream and is for the Retransmission Server (RS). Its counterpart for the RTP Receiver (RTP_Rx) includes the "a=recvonly" line as shown in Figure 10.

When an RTP receiver asks for rapid acquisition before it joins a primary multicast RTP session, it sends a RAMS-R message to the feedback target of that primary multicast RTP session. If FT_Ap is associated with only one RTP stream, the RTP receiver does not need to learn the SSRC of that stream via an out-of-band method. If RS accepts the request, it will send an RAMS-I message with the correct SSRC identifier. If FT_Ap is associated with a multi-stream RTP session and the RTP receiver is willing to request rapid acquisition for the entire session, the RTP receiver again does not need to learn the SSRCs via an out-of-band method. However, if the RTP receiver is willing to request a particular subset of the primary multicast streams, it must learn their SSRC identifiers and list them in the RAMS-R message. Since this RTP receiver has not yet received any RTP packets for the primary multicast stream(s), the RTP receiver must in this case learn the SSRC value(s) from the 'ssrc' attribute of the media description. In addition to the SSRC value, the 'cname' source attribute must also be present in the SDP description [RFC5576].

Note that listing the SSRC values for the primary multicast streams in the SDP file does not create a problem in SSM sessions when an SSRC collision occurs. This is because in SSM sessions, an RTP receiver that observed an SSRC collision with a media sender MUST change its own SSRC [I-D.ietf-avt-rtcpssm] by following the rules defined in [RFC3550].

A feedback target that receives a RAMS-R feedback message becomes aware that the prediction chain at the RTP receiver side has been broken or does not exist anymore. If the necessary conditions are satisfied (as outlined in Section 7 of [RFC4585]) and available resources exist, RS may react to the RAMS-R message by sending any transport-layer and payload-specific feedback message(s) and starting the unicast burst.

```
v=0
o=ali 1122334455 1122334466 IN IP4 rams.example.com
s=Rapid Acquisition Example
t=0 0
a=group:FID 1 2
a=rtcp-unicast:rsi
m=video 41000 RTP/AVPF 98
i=Primary Multicast Stream
c=IN IP4 233.252.0.2/255
a=source-filter: incl IN IP4 233.252.0.2 198.51.100.1
a=recvonly
a=rtpmap:98 MP2T/90000
a=rtcp:41001 IN IP4 192.0.2.1
a=rtcp-fb:98 nack
a=rtcp-fb:98 nack ssli
a=ssrc:123321 cname:iptv-ch32@rams.example.com
a=rams-updates
a=mid:1
m=video 41002 RTP/AVPF 99
i=Unicast Retransmission Stream (Ret. and Rapid Acq. Support)
c=IN IP4 192.0.2.1
a=sendonly
a=rtpmap:99 rtx/90000
a=rtcp:41003
a=fmtp:99 apt=98; rtx-time=5000
a=mid:2
```

Figure 9: Example SDP for RS when RAMS support is enabled

```
v=0
o=ali 1122334455 1122334466 IN IP4 rams.example.com
s=Rapid Acquisition Example
t=0 0
a=group:FID 1 2
a=rtcp-unicast:rsi
m=video 41000 RTP/AVPF 98
i=Primary Multicast Stream
c=IN IP4 233.252.0.2/255
a=source-filter: incl IN IP4 233.252.0.2 198.51.100.1
a=recvonly
a=rtpmap:98 MP2T/90000
a=rtcp:41001 IN IP4 192.0.2.1
a=rtcp-fb:98 nack
a=rtcp-fb:98 nack ssli
a=ssrc:123321 cname:iptv-ch32@rams.example.com
a=rams-updates
a=mid:1
m=video 41002 RTP/AVPF 99
i=Unicast Retransmission Stream (Ret. and Rapid Acq. Support)
c=IN IP4 192.0.2.1
a=recvonly
a=rtpmap:99 rtx/90000
a=rtcp:41003
a=fmtp:99 apt=98; rtx-time=5000
a=mid:2
```

Figure 10: Example SDP for RTP_Rx when RAMS support is enabled

In this section, we considered the simplest scenario where the primary multicast RTP session carried only one stream and the RTP receiver wanted to rapidly acquire this stream only. Best practices for scenarios where the primary multicast RTP session carries two or more streams or the RTP receiver wants to acquire one or more streams from multiple primary multicast RTP sessions at the same time are presented in [I-D.begen-avt-rams-scenarios].

9. NAT Considerations

For a variety of reasons, one or more NAPT devices (hereafter simply called NAT) may exist between RTP_Rx and RS. NATs have a variety of operating characteristics for UDP traffic [RFC4787]. For a NAT to permit traffic from RS to arrive at RTP_Rx, the NAT(s) must first either:

a. See UDP traffic sent from RTP_Rx (which is on the 'inside' of the NAT) to RS (which is on the 'outside' of the NAT). This traffic is sent to the same transport address as the subsequent response traffic, or;

b. Be configured to forward certain ports (e.g., using HTML configuration, UPnP IGD [UPnP-IGD], DLNA [DLNA]). Details of this are out of scope of this document.

For both (a) and (b), RTP_Rx is responsible for maintaining the NAT's state if it wants to receive traffic from the RS on that port. For (a), RTP_Rx MUST send UDP traffic to keep the NAT binding alive, at least every 30 seconds [RFC4787]. Note that while (a) is more like an automatic/dynamic configuration, (b) is more like a manual/static configuration.

When RTP_Rx sends a RAMS-R message in the primary multicast RTP session and the request is received by RS, a new unicast RTP retransmission session will be established between RS and RTP_Rx.

While the ports on the RS side are already signaled via out-of-band means (e.g., SDP), RTP_Rx may need to convey to RS the RTP and RTCP ports it wants to use on its side for the new session. Since there are two RTP sessions involved during this process and one of them is established upon a feedback message sent in the other one, this requires an explicit port mapping method. This problem equally applies to scenarios where the RTP media is multicast in an SSM session, and an RTP receiver requests retransmission from a local repair server by using the RTCP NACK messages for the missing packets and the repair server retransmits the requested packets over a unicast session. Thus, instead of laying out a specific solution for the RAMS applications, a general solution is introduced in [I-D.begen-avt-ports-for-ucast-mcast-rtp].

Applications using RAMS MUST support this solution both on the RS and RTP_Rx side to allow RTP receivers to use their desired ports and to support RAMS behind NAT devices.

10. Security Considerations

Applications that are using RAMS make heavy use of the unicast feedback mechanism described in [I-D.ietf-avt-rtcpssm] and the payload format defined in [RFC4588]. Thus, these applications are subject to the general security considerations discussed in [I-D.ietf-avt-rtcpssm] and [RFC4588]. In this section, we give an overview of the guidelines and suggestions described in these specifications from a RAMS perspective. We also discuss the security

considerations that explicitly apply to applications using RAMS.

First of all, much of the session description information is available in the SDP descriptions that are distributed to the media senders, Retransmission Servers and RTP receivers. Adequate security measures are RECOMMENDED to ensure the integrity and authenticity of the SDP descriptions so that transport addresses of the media senders, distribution sources, feedback targets as well as other session-specific information can be authenticated.

Compared to an RTCP NACK message that triggers one or more retransmissions, a RAMS Request (RAMS-R) message may trigger a new burst stream to be sent by the Retransmission Server. Depending on the application-specific requirements and conditions existing at the time of the RAMS-R reception by the Retransmission Server, the resulting burst stream may contain potentially a large number of retransmission packets. Since these packets are sent at a faster than the nominal rate, RAMS consumes more resources on the Retransmission Server, the RTP receiver and the network. This particularly makes denial-of-service attacks more intense, and hence, more harmful than attacks that target ordinary retransmission sessions.

Following the suggestions given in [RFC4588], counter-measures SHOULD be taken to prevent tampered or spoofed RTCP packets. Tampered RAMS-R messages may trigger inappropriate burst streams or alter the existing burst streams in an inappropriate way. For example, if the Max Receive Bitrate field is altered by a tampered RAMS-R message, the updated burst may overflow the buffer on the receiver side, or oppositely, may slow down the burst to the point that it becomes useless. Tampered RAMS Termination (RAMS-T) messages may terminate valid burst streams pre-maturely resulting in gaps in the received RTP packets. RAMS Information (RAMS-I) messages contain fields that are critical for the success of the RAMS operation. Any tampered information in the RAMS-I message may easily cause the RTP receiver to make wrong decisions. Consequently, the RAMS operation may fail.

While most of the denial-of-service attacks can be prevented by the integrity and authenticity checks enabled by SRTP, an attack can still be started by legitimate endpoints that send several valid RAMS-R messages to a particular feedback target in a synchronized fashion and very short amount of time. Since a RAMS operation may temporarily consume a large amount of resources, a series of the RAMS-R messages may temporarily overload the Retransmission Server. In these circumstances, the Retransmission Server may, for example, reject incoming RAMS requests until its resources become available again. One means to ameliorate this threat is to apply a per-endpoint policing mechanism on the incoming RAMS requests. A

reasonable policing mechanism should consider application-specific requirements and minimize false negatives.

In addition to the denial-of-service attacks, man-in-the-middle and replay attacks can also be harmful. However, RAMS itself does not bring any new risks or threats other than the ones discussed in [I-D.ietf-avt-rtcpssm].

[RFC4588] RECOMMENDS that the cryptography mechanisms are used for the retransmission payload format to provide protection against known plaintext attacks. As discussed in [RFC4588], the retransmission payload format sets the timestamp field in the RTP header to the media timestamp of the original packet and this does not compromise the confidentiality. Furthermore, if cryptography is used to provide security services on the original stream, then the same services, with equivalent cryptographic strength, MUST be provided on the retransmission stream per [RFC4588].

11. IANA Considerations

The following contact information shall be used for all registrations in this document:

Ali Begen
abegen@cisco.com

170 West Tasman Drive
San Jose, CA 95134 USA

Note to the RFC Editor: In the following, please replace "XXXX" with the number of this document prior to publication as an RFC.

11.1. Registration of SDP Attributes

This document registers a new attribute name in SDP.

```
SDP Attribute ("att-field"):
Attribute name:     rams-updates
Long form:          Support for Updated RAMS Request Messages
Type of name:       att-field
Type of attribute:  Media level
Subject to charset: No
Purpose:            See this document
Reference:          [RFCXXXX]
Values:             None
```

11.2. Registration of SDP Attribute Values

This document registers a new value for the 'nack' attribute to be used with the 'rtcp-fb' attribute in SDP. For more information about 'rtcp-fb', refer to [RFC4585].

```
Value name:     ssli
Long name:      Stream Synchronization Loss Indication
Usable with:    nack
Reference:      [RFCXXXX]
```

11.3. Registration of FMT Values

Within the RTPFB range, the following format (FMT) value is registered:

```
Name:           RAMS
Long name:      Rapid Acquisition of Multicast Sessions
Value:          6
Reference:      [RFCXXXX]
```

11.4. SFMT Values for RAMS Messages Registry

This document creates a new sub-registry for the sub-feedback message type (SFMT) values to be used with the FMT value registered for RAMS messages. The registry is called the SFMT Values for RAMS Messages Registry. This registry is to be managed by the IANA according to the Specification Required policy of [RFC5226].

The length of the SFMT field in the RAMS messages is a single octet, allowing 256 values. The registry is initialized with the following entries:

Value	Name	Reference
0	Reserved	[RFCXXXX]
1	RAMS Request	[RFCXXXX]
2	RAMS Information	[RFCXXXX]
3	RAMS Termination	[RFCXXXX]
4-254		Specification Required
255	Reserved	[RFCXXXX]

The SFMT values 0 and 255 are reserved for future use.

Any registration for an unassigned SFMT value MUST contain the following information:

- o Contact information of the one doing the registration, including at least name, address, and email.

- o A detailed description of what the new SFMT represents and how it shall be interpreted.

Note that new RAMS functionality should be introduced by using the extension mechanism within the existing RAMS message types not by introducing new message types unless it is absolutely necessary.

11.5. RAMS TLV Space Registry

This document creates a new IANA TLV space registry for the RAMS extensions. The registry is called the RAMS TLV Space Registry. This registry is to be managed by the IANA according to the Specification Required policy of [RFC5226].

The length of the Type field in the TLV elements is a single octet, allowing 256 values. The Type values 0 and 255 are reserved for future use. The Type values between (and including) 128 and 254 are reserved for private extensions.

The registry is initialized with the following entries:

Type	Description	Reference
0	Reserved	[RFCXXXX]
1	Requested Media Sender SSRC(s)	[RFCXXXX]
2	Min RAMS Buffer Fill Requirement	[RFCXXXX]
3	Max RAMS Buffer Fill Requirement	[RFCXXXX]
4	Max Receive Bitrate	[RFCXXXX]
5	Request for Preamble Only	[RFCXXXX]
6-30		Specification Required
31	Media Sender SSRC	[RFCXXXX]
32	RTP Seqnum of the First Packet	[RFCXXXX]
33	Earliest Multicast Join Time	[RFCXXXX]
34	Burst Duration	[RFCXXXX]
35	Max Transmit Bitrate	[RFCXXXX]
36-60		Specification Required
61	Extended RTP Seqnum of First Multicast Packet	[RFCXXXX]
62-127		Specification Required
128-254		No IANA Maintenance
255	Reserved	[RFCXXXX]

Any registration for an unassigned Type value MUST contain the following information:

o Contact information of the one doing the registration, including at least name, address, and email.

o A detailed description of what the new TLV element represents and how it shall be interpreted.

11.6. RAMS Response Code Space Registry

This document creates a new IANA TLV space registry for the RAMS response codes. The registry is called the RAMS Response Code Space Registry. This registry is to be managed by the IANA according to the Specification Required policy of [RFC5226].

The length of the Response field is two octets, allowing 65536 codes. However, the response codes have been classified and registered following an HTTP-style code numbering in this document. New response codes SHALL follow the guidelines below:

Code Level	Purpose
1xx	Informational
2xx	Success
3xx	Redirection
4xx	RTP Receiver Error
5xx	Retransmission Server Error

The Response code 65536 is reserved for future use.

The registry is initialized with the following entries:

Code	Description	Reference
0	A private response code is included in the message	[RFCXXXX]
100	Parameter update for RAMS session	[RFCXXXX]
200	RAMS request has been accepted	[RFCXXXX]
201	Unicast burst has been completed	[RFCXXXX]
400	Invalid RAMS-R message syntax	
401	Invalid min buffer requirement in RAMS-R message	[RFCXXXX]
402	Invalid max buffer requirement in RAMS-R message	[RFCXXXX]
403	Invalid max bitrate requirement in RAMS-R message	[RFCXXXX]
500	An unspecified RS internal error has occurred	[RFCXXXX]
501	RS has no bandwidth to start RAMS session	[RFCXXXX]
502	Burst is terminated due to network congestion	[RFCXXXX]
503	RS has no CPU available to start RAMS session	[RFCXXXX]
504	RAMS functionality is not available on RS	[RFCXXXX]
505	RAMS functionality is not available for RTP_Rx	[RFCXXXX]
506	RAMS functionality is not available for the requested multicast stream	[RFCXXXX]
507	RS has no valid starting point available for the requested multicast stream	[RFCXXXX]
508	RS has no reference information available for the requested multicast stream	[RFCXXXX]
509	RS has no RTP stream matching the requested SSRC	[RFCXXXX]
510	RAMS request to acquire the entire session has been denied	[RFCXXXX]
511	Only the preamble information is sent	[RFCXXXX]
512	RAMS request has been denied due to a policy	[RFCXXXX]

Any registration for an unassigned Response code MUST contain the

following information:

- o Contact information of the one doing the registration, including at least name, address, and email.

- o A detailed description of what the new Response code describes and how it shall be interpreted.

12. Contributors

Dave Oran and Magnus Westerlund have contributed significantly to this specification by providing text and solutions to some of the issues raised during the development of this specification.

13. Acknowledgments

The following individuals have reviewed the earlier versions of this specification and provided helpful comments: Colin Perkins, Joerg Ott, Roni Even, Dan Wing, Tony Faustini, Peilin Yang, Jeff Goldberg, Muriel Deschanel, Orit Levin, Guy Hirson, Tom Taylor, Xavier Marjou, Ye-Kui Wang, Zixuan Zou, Ingemar Johansson, Haibin Song, Ning Zong, Jonathan Lennox and Sean Sheedy.

14. Change Log

14.1. draft-ietf-avt-rapid-acquisition-for-rtp-07

The following are the major changes compared to version 06:

- o Congestion control considerations text has been added to Section 6.4.

14.2. draft-ietf-avt-rapid-acquisition-for-rtp-06

The following are the major changes compared to version 05:

- o Comments from WGLC have been addressed. See the mailing list for the list of changes.

- o Support for multi-stream RTP sessions has been added.

- o NAT section has been revised.

14.3. draft-ietf-avt-rapid-acquisition-for-rtp-05

The following are the major changes compared to version 04:

- Editorial changes throughout the document.

14.4. draft-ietf-avt-rapid-acquisition-for-rtp-04

The following are the major changes compared to version 03:

- Clarifications for the definition of RS.
- Response codes have been defined.

14.5. draft-ietf-avt-rapid-acquisition-for-rtp-03

The following are the major changes compared to version 02:

- Clarifications for the RAMS-I message.
- Type values have been assigned.

14.6. draft-ietf-avt-rapid-acquisition-for-rtp-02

The following are the major changes compared to version 01:

- Port mapping discussion has been removed since it will be discussed in a separate draft.
- Security considerations section has been added.
- Burst shaping section has been completed.
- Most of the outstanding open issues have been addressed.

14.7. draft-ietf-avt-rapid-acquisition-for-rtp-01

The following are the major changes compared to version 00:

- Formal definitions of vendor-neutral and private extensions and their IANA registries have been added.
- SDP examples were explained in more detail.
- The sub-FMT field has been introduced in the RAMS messages for message type identification.

- Some terminology has been fixed.

- NAT considerations section has been added.

14.8. draft-ietf-avt-rapid-acquisition-for-rtp-00

This is a resubmission of version 03 as a WG item.

14.9. draft-versteeg-avt-rapid-synchronization-for-rtp-03

The following are the major changes compared to version 02:

- The title and message names have been changed.

- RTCP message semantics have been added. RAMS protocol has been revised to handle updated requests and responses.

- Definitions have been revised.

- RTP/RTCP muxing reference has been added.

14.10. draft-versteeg-avt-rapid-synchronization-for-rtp-02

The following are the major changes compared to version 01:

- The discussion around MPEG2-TS has been moved to another document.

- The RAMS-R, RAMS-I and RAMS-T messages have been extensively modified and they have been made mandatory.

- IANA Considerations section has been updated.

- The discussion of RTCP XR report has been moved to another document.

- A new section on protocol design considerations has been added.

14.11. draft-versteeg-avt-rapid-synchronization-for-rtp-01

The following are the major changes compared to version 00:

- The core of the rapid synchronization method is now payload-independent. But, the draft still defines payload-specific messages that are required for enabling rapid synch for the RTP flows carrying MPEG2-TS.

- RTCP APP packets have been removed, new RTCP transport-layer and payload-specific feedback messages have been defined.

- o The step for leaving the current multicast session has been removed from Section 6.2.

- o A new RTCP XR (Multicast Join) report has been defined.

- o IANA Considerations section have been updated.

- o Editorial changes to clarify several points.

15. References

15.1. Normative References

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, July 2003.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3376] Cain, B., Deering, S., Kouvelas, I., Fenner, B., and A. Thyagarajan, "Internet Group Management Protocol, Version 3", RFC 3376, October 2002.

[RFC3810] Vida, R. and L. Costa, "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", RFC 3810, June 2004.

[RFC4604] Holbrook, H., Cain, B., and B. Haberman, "Using Internet Group Management Protocol Version 3 (IGMPv3) and Multicast Listener Discovery Protocol Version 2 (MLDv2) for Source-Specific Multicast", RFC 4604, August 2006.

[RFC4566] Handley, M., Jacobson, V., and C. Perkins, "SDP: Session Description Protocol", RFC 4566, July 2006.

[I-D.ietf-mmusic-rfc3388bis]
Camarillo, G. and H. Schulzrinne, "The SDP (Session Description Protocol) Grouping Framework", draft-ietf-mmusic-rfc3388bis-04 (work in progress), November 2009.

[RFC4585] Ott, J., Wenger, S., Sato, N., Burmeister, C., and J. Rey, "Extended RTP Profile for Real-time Transport Control Protocol (RTCP)-Based Feedback (RTP/AVPF)", RFC 4585, July 2006.

[RFC4588] Rey, J., Leon, D., Miyazaki, A., Varsa, V., and R.

Hakenberg, "RTP Retransmission Payload Format", RFC 4588, July 2006.

[I-D.ietf-avt-rtcpssm]
Ott, J. and J. Chesterfield, "RTCP Extensions for Single-Source Multicast Sessions with Unicast Feedback", draft-ietf-avt-rtcpssm-19 (work in progress), November 2009.

[RFC5576] Lennox, J., Ott, J., and T. Schierl, "Source-Specific Media Attributes in the Session Description Protocol (SDP)", RFC 5576, June 2009.

[RFC5234] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, January 2008.

[RFC5506] Johansson, I. and M. Westerlund, "Support for Reduced-Size Real-Time Transport Control Protocol (RTCP): Opportunities and Consequences", RFC 5506, April 2009.

[RFC5285] Singer, D. and H. Desineni, "A General Mechanism for RTP Header Extensions", RFC 5285, July 2008.

[I-D.ietf-avt-rapid-rtp-sync]
Perkins, C. and T. Schierl, "Rapid Synchronisation of RTP Flows", draft-ietf-avt-rapid-rtp-sync-09 (work in progress), January 2010.

[I-D.begen-avt-ports-for-ucast-mcast-rtp]
Begen, A. and B. Steeg, "Port Mapping Between Unicast and Multicast RTP Sessions", draft-begen-avt-ports-for-ucast-mcast-rtp-02 (work in progress), February 2010.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

15.2. Informative References

[RFC0768] Postel, J., "User Datagram Protocol", STD 6, RFC 768, August 1980.

[I-D.begen-avt-rams-scenarios]
Begen, A., "Considerations for RAMS Scenarios", draft-begen-avt-rams-scenarios-00 (work in progress), October 2009.

[I-D.begen-avt-rtp-mpeg2ts-preamble]
Begen, A. and E. Friedrich, "RTP Payload Format for MPEG2-TS Preamble", draft-begen-avt-rtp-mpeg2ts-preamble-04 (work in progress), December 2009.

[I-D.begen-avt-rapid-sync-rtcp-xr]
Begen, A. and E. Friedrich, "Multicast Acquisition Report Block Type for RTCP XR", draft-begen-avt-rapid-sync-rtcp-xr-03 (work in progress), October 2009.

[I-D.westerlund-avt-ecn-for-rtp]
Westerlund, M., Johansson, I., Perkins, C., and K. Carlberg, "Explicit Congestion Notification (ECN) for RTP over UDP", draft-westerlund-avt-ecn-for-rtp-02 (work in progress), October 2009.

[I-D.ietf-fecframe-interleaved-fec-scheme]
Begen, A., "RTP Payload Format for 1-D Interleaved Parity FEC", draft-ietf-fecframe-interleaved-fec-scheme-09 (work in progress), January 2010.

[RFC4787] Audet, F. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", BCP 127, RFC 4787, January 2007.

[I-D.ietf-dccp-rtp]
Perkins, C., "RTP and the Datagram Congestion Control Protocol (DCCP)", draft-ietf-dccp-rtp-07 (work in progress), June 2007.

[UPnP-IGD]
Forum, UPnP., "Universal Plug and Play (UPnP) Internet Gateway Device (IGD)", November 2001.

[DLNA] , DLNA., "http://www.dlna.org/home".

[IC2009] Begen, A., Glazebrook, N., and W. VerSteeg, "Reducing Channel Change Times in IPTV with Real-Time Transport Protocol (IEEE Internet Computing)", May 2009.

Authors' Addresses

Bill VerSteeg
Cisco
5030 Sugarloaf Parkway
Lawrenceville, GA 30044
USA

Email: billvs@cisco.com

Ali Begen
Cisco
170 West Tasman Drive
San Jose, CA 95134
USA

Email: abegen@cisco.com

Tom VanCaenegem
Alcatel-Lucent
Copernicuslaan 50
Antwerpen,   2018
Belgium

Email: Tom.Van_Caenegem@alcatel-lucent.be

Zeev Vax
Microsoft Corporation
1065 La Avenida
Mountain View, CA 94043
USA

Email: zeevvax@microsoft.com