AVT
Internet-Draft
Intended status: Standards Track
Expires: March 3, 2011

B. VerSteeg
A. Begen
Cisco
T. VanCaenegem
Alcatel-Lucent
Z. Vax
Microsoft Corporation
August 30, 2010

Unicast-Based Rapid Acquisition of Multicast RTP Sessions
draft-ietf-avt-rapid-acquisition-for-rtp-14

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 can 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 can be transmitted at a faster than natural bitrate 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 in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

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."

This Internet-Draft will expire on March 3, 2011.

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 Simplified 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 4
1.1. Acquisition of RTP Streams vs. RTP Sessions 6
1.2. Outline 7
2. Requirements Notation 7
3. Definitions 7
4. Elements of Delay in Multicast Applications 9
5. Protocol Design Considerations and Their Effect on Resource Management for Rapid Acquisition 10
6. Rapid Acquisition of Multicast RTP Sessions (RAMS) 12
6.1. Overview 12
6.2. Message Flows 13
6.3. Synchronization of Primary Multicast Streams 23
6.4. Burst Shaping and Congestion Control in RAMS 24
6.5. Failure Cases 27
7. Encoding of the Signaling Protocol in RTCP 28
7.1. Extensions 29
7.1.1. Vendor-Neutral Extensions 30
7.1.2. Private Extensions 30
7.2. RAMS Request 30
7.3. RAMS Information 33
7.4. RAMS Termination 36
8. SDP Signaling 37
8.1. Definitions 37
8.2. Requirements 37
8.3. Example and Discussion 38
9. NAT Considerations 41
10. Security Considerations 42
11. IANA Considerations 44
11.1. Registration of SDP Attributes 44
11.2. Registration of SDP Attribute Values 44
11.3. Registration of FMT Values 44
11.4. SFMT Values for RAMS Messages Registry 45
11.5. RAMS TLV Space Registry 45
11.6. RAMS Response Code Space Registry 46
11.6.1. Response Code Definitions 49
12. Contributors 50
13. Acknowledgments 51
14. References 51
14.1. Normative References 51
14.2. Informative References 53
Authors' Addresses 54

1. Introduction

Most multicast flows carry a stream of inter-related data. The receivers need to acquire certain information 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 can 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 might 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 might 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 can 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 might 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. While this problem equally applies to both any-source (ASM) and source-

specific (SSM) multicast applications, in this specification we address it for the SSM-based multicast applications by describing 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 an 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 [RFC5760]) 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 might 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.

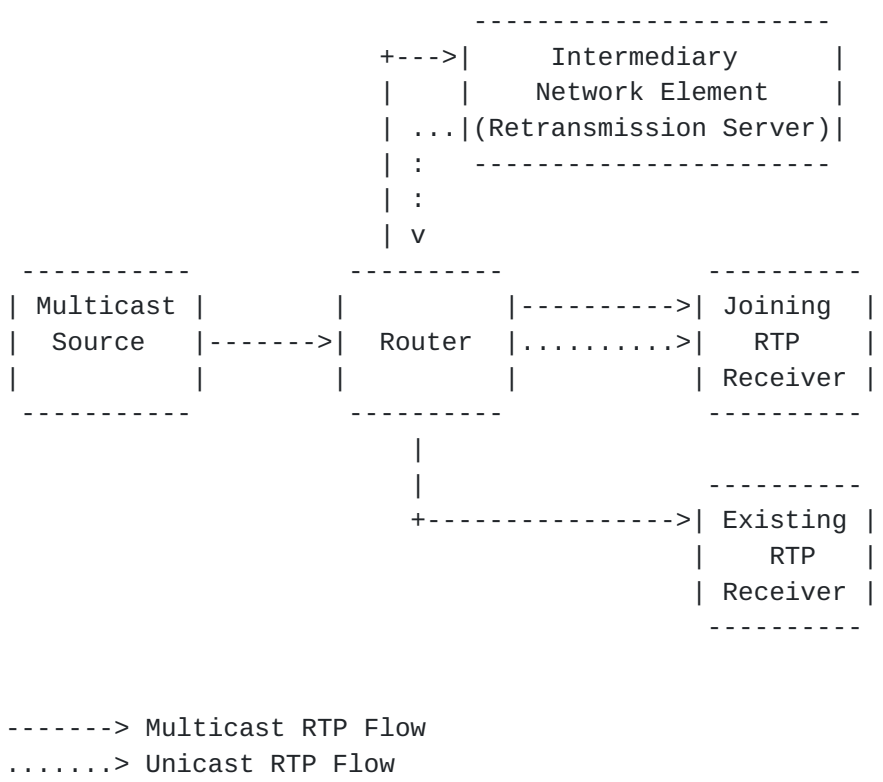


Figure 1: Rapid acquisition through an intermediary network element

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

In this memo we describe a protocol that handles the rapid acquisition of a single multicast 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 can opt to rapidly acquire a specific subset of the available RTP streams in the primary multicast RTP session. Alternatively, the RTP receiver can 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 session will be established by the retransmission server. This unicast RTP session is separate from the associated primary multicast RTP session. As a result, there are always two different RTP sessions in a single instance of the rapid acquisition protocol: (i) the primary multicast RTP session with its associated unicast feedback and (ii) the unicast RTP session.

If the RTP receiver wants to rapidly acquire multiple RTP sessions simultaneously, separate unicast RTP 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. A primary SSM session can carry multiple RTP streams.

Primary Multicast RTP Session: The multicast RTP session an RTP receiver is interested in acquiring rapidly. From the RTP receiver's viewpoint, the primary multicast RTP session has one associated unicast RTCP feedback stream to a Feedback Target, in addition to the primary multicast RTP stream(s).

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 [RFC5760]. FT_Ap denotes a specific feedback target running on a particular address and port.

Retransmission (or Burst) Packet: An RTP packet that is formatted as defined in Section 4 of [RFC4588]. The payload of a retransmission or burst packet comprises the retransmission payload header followed by the payload of the original RTP packet.

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 (or Retransmission) RTP Session: The unicast RTP session used to send one or more unicast burst RTP streams and their associated RTCP messages. The terms "burst RTP session" and "retransmission RTP session" can be used interchangeably.

(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 Synchronization Source (SSRC) identifier that is unique in the primary multicast RTP session. Following the session-multiplexing guidelines in [RFC4588], each unicast burst stream will use the same SSRC and CNAME as its primary multicast RTP stream.

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

4. Elements of Delay in Multicast Applications

In 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 can 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 might 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 can become quite large and be a major contributor to the overall delay.

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 up to a few seconds, 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

This section is for informational purposes and does not contain requirements for implementations.

Rapid acquisition is an optimization of a system that is expected to 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:

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

- 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 might 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 might 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 can vary over time. The receiver might have knowledge of a portion of this network, or might 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.

- 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.

- 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 (RAMS)

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

6.1. Overview

[RFC5760] 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. One or more entities different from the Distribution Source MAY implement the feedback target functionality, 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, it can first request a unicast burst from the Burst Source before it joins the SSM session. In this burst, the packets are formatted as RTP retransmission packets [RFC4588] and carry 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 defines 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

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

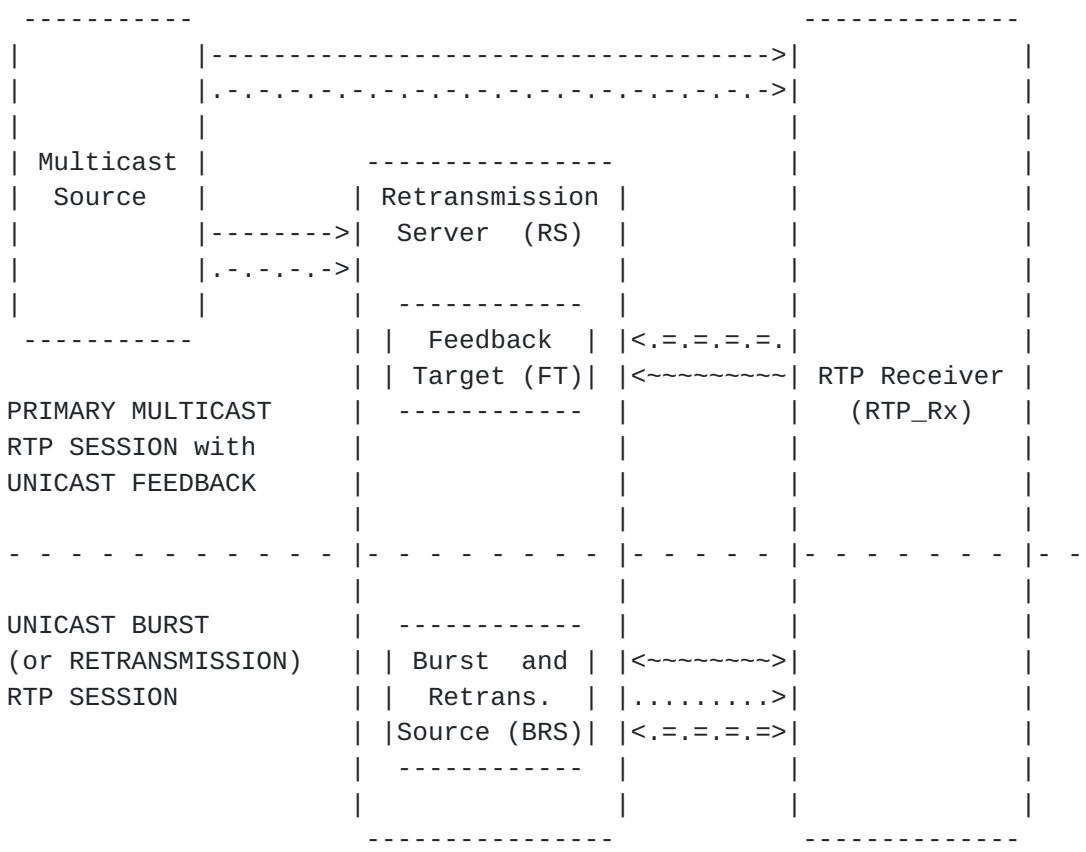


Figure 2: Flow diagram for unicast-based rapid acquisition

The feedback target (FT) is the entity as defined in [RFC5760], to which the 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 (BRS) is responsible for responding to these messages and for further RTCP interaction with the 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 BRS) are combined in a single physical entity and they share state. In the remainder of the text, the term Retransmission Server (RS) is used whenever appropriate, to refer to this single physical entity.

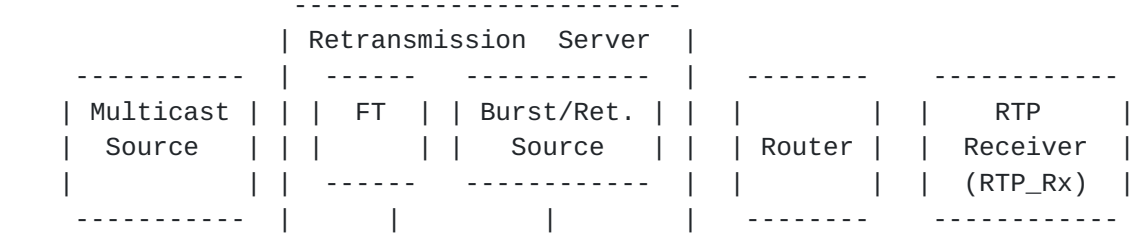
The FT is involved in the primary multicast RTP session and receives

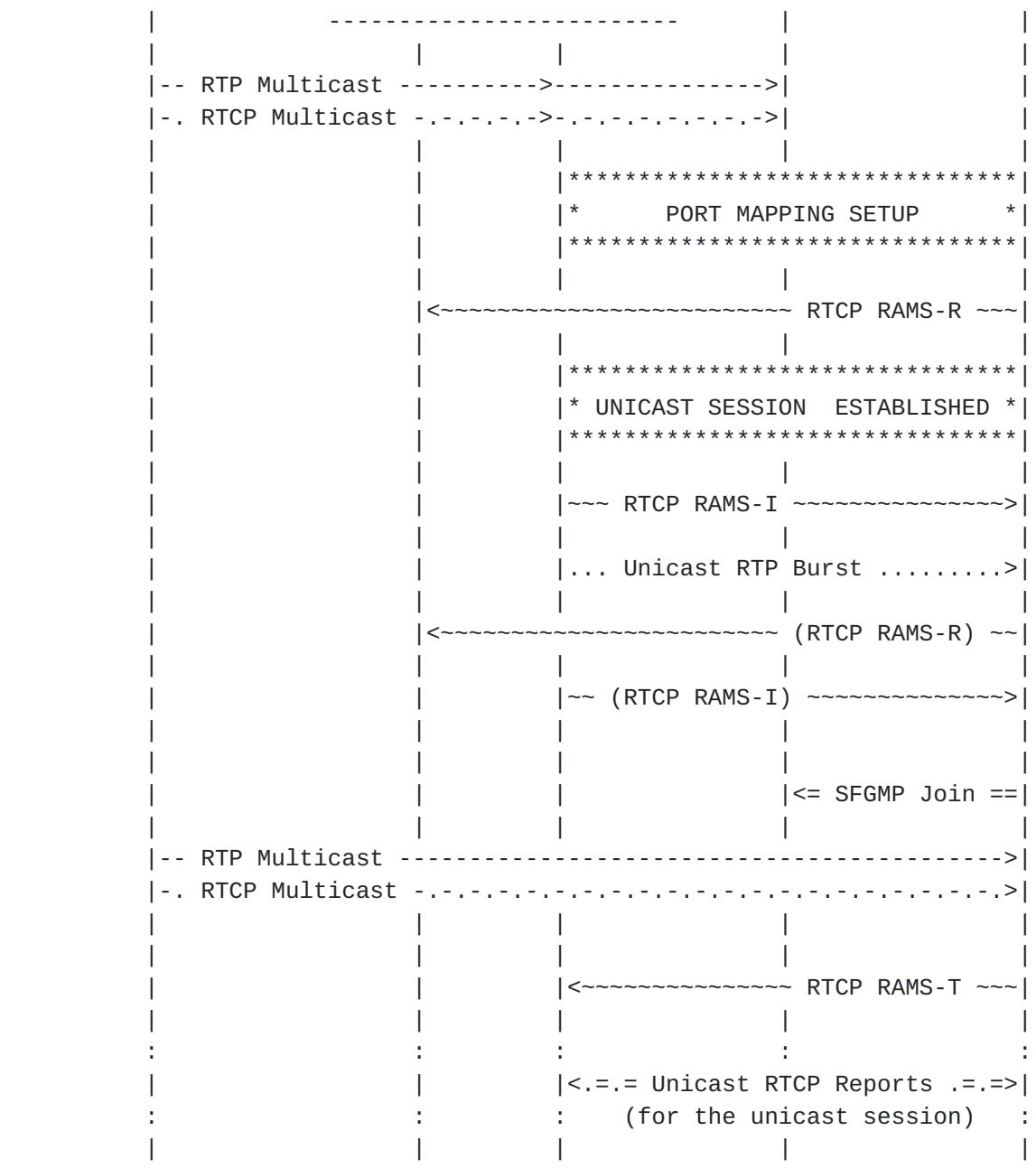
unicast feedback for that session as in [RFC5760]. The BRS is involved in the unicast burst (or retransmission) RTP session and transmits the unicast burst and retransmission packets formatted as RTP retransmission packets [RFC4588] in a single separate unicast RTP session to each RTP_Rx. In the unicast burst RTP session, as in any other RTP session, the BRS and RTP_Rx regularly send the periodic sender and receiver reports, respectively.

The unicast burst is started by an RTCP feedback message that is defined in this document based on the common packet format provided in [RFC4585]. An RTP retransmission is triggered by an RTCP NACK message defined in [RFC4585]. Both of these messages are sent to the FT of the primary multicast RTP session, and can start the unicast burst/retransmission RTP session.

In the RTP/AVPF profile, there are certain rules that apply to scheduling of both of these messages sent to the FT in the primary multicast RTP session, and these 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_Rx 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. For the subsequent messages (e.g., updated requests) 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 might or might not be present during rapid acquisition. 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.





-------> 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 the RS and RTP_Rx.
~~~~~~~

It is instructive to consider independently operating implementations on the 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 provisions are needed 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, the RTP_Rx needs to choose its RTP and RTCP receive ports for the unicast burst RTP session. Since this unicast session is established after the RTP_Rx has sent a RAMS-Request (RAMS-R) message as unicast feedback in the primary multicast RTP session, the RTP_Rx MUST first setup the port mappings between the unicast and multicast sessions and send this mapping information to the FT along with the RAMS-R message so that the BRS knows how to communicate with the RTP_Rx.

   The details of this setup procedure are explained in [I-D.ietf-avt-ports-for-ucast-mcast-rtp]. Other NAT-related issues are left to Section 9 to keep the present discussion focused on the RAMS message flows.

2. Request: the RTP_Rx sends a rapid acquisition request (RAMS-R) for the primary multicast RTP session to the unicast feedback target of that session. The request contains the SSRC identifier of the RTP_Rx (aka SSRC of packet sender) and can contain the media sender SSRC identifier(s) of the primary multicast stream(s) of interest (aka SSRC of media source). The RAMS-R message can contain parameters that constrain the burst, such as the buffer and bandwidth limits.

   Before joining the SSM session, the RTP_Rx learns the addresses for the multicast source, group and RS by out-of-band means. If the RTP_Rx desires to rapidly acquire only a subset of the primary multicast streams available in the primary multicast RTP session, the RTP_Rx MUST also acquire the SSRC identifiers for the desired RTP streams out-of-band. Based on this information, the RTP_Rx populates the desired SSRC(s) in the RAMS-R message.

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

3. Response: The BRS sends RAMS-Information (RAMS-I) message(s) to the RTP_Rx to convey the status for the burst(s) requested by the RTP_Rx.

   If the primary multicast RTP session associated with the FT_Ap (a specific feedback target running on a particular address and port) on which the RAMS-R message was received contains only a single primary multicast stream, the BRS SHALL always use the SSRC of the RTP stream associated with the FT_Ap in the RAMS-I message(s) regardless of the media sender SSRC requested 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, the BRS MUST explicitly populate the correct media sender SSRC in the initial RAMS-I message (See Section 7.3).

   The FT_Ap could also be associated with an RTP session that carries two or more primary multicast streams. If the 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.

   If the FT_Ap is associated with two or more RTP sessions, RTP_Rx's request will be ambiguous. To avoid any ambiguity, each FT_Ap MUST be only associated with a single RTP session.

   If the RTP_Rx is willing to rapidly acquire only a subset of the primary multicast streams, the RTP_Rx MUST list all the media sender SSRC(s) denoting the stream(s) it wishes to acquire in the RAMS-R message. Upon receiving such a message, the BRS MAY accept the request for all or a subset of the media sender SSRC(s) that matched the RTP stream(s) it serves. The BRS MUST reject all other requests with an appropriate response code.

   * Reject Responses: The BRS MUST take into account any limitations that may have been specified by the RTP_Rx in the RAMS-R message when making a decision regarding the request. If the RTP_Rx has requested to acquire the whole primary multicast RTP session but the BRS cannot provide a rapid acquisition service for any of the primary multicast streams, the BRS MUST reject the request via a single RAMS-I message with a collective reject response code, which is defined as 510 in Section 11.6, and whose media sender SSRC field is set to one of SSRCs served by this FT_Ap. Upon receiving this RAMS-I message, the RTP_Rx abandons the rapid acquisition attempt and can immediately join the multicast session by sending an SFGMP Join message towards its upstream multicast

router.

In all other cases, the BRS MUST send a separate RAMS-I message with the appropriate 5xx-level response code (as defined in Section 11.6) for each primary multicast stream that has been requested by the RTP_Rx but cannot be served by the BRS. There could be multiple reasons why the BRS has rejected the request, however, the BRS chooses the most appropriate response code to inform the RTP_Rx.

Upon receiving a reject response indicating a transient problem such as insufficient BRS or network resources, if the RTP_Rx wants to retry sending the same request, it MUST follow the RTCP timer rules of [RFC4585] to allow the transient problems go away. However, if the reject response indicates a non-transient problem (such as the ones reported by response codes 504, 505 and 506), the RTP_Rx MUST NOT attempt a retry.

The BRS can employ a policing mechanism against the broken RTP_Rx implementations that are not following these rules. See Section 10 for details.

* Accept Responses: The BRS MUST send at least one separate RAMS-I message with the appropriate response code (either zero indicating a private response or response code 200 indicating success as listed in Section 11.6) for each primary multicast stream that has been requested by the RTP_Rx and will be served by the BRS. Such RAMS-I messages comprise fields that can be used to describe the individual unicast burst streams. When there is a RAMS-R request for multiple primary multicast streams, the BRS MUST include all the individual RAMS-I messages corresponding to that RAMS-R request in the same compound RTCP packet if these messages fit in the same packet.

  The RAMS-I message carries the RTP sequence number of the first packet transmitted in the respective RTP stream to allow the RTP_Rx to detect any missing initial packet(s). When the BRS accepts a request for a primary multicast stream, this field MUST always be populated in the RAMS-I message(s) sent for this particular primary multicast stream. It is RECOMMENDED that the BRS sends a RAMS-I message at the start of the burst so that the RTP_Rx can quickly detect any missing initial packet(s).

It is possible that the RAMS-I message for a primary multicast

stream can get delayed or lost, and the RTP_Rx can start receiving RTP packets before receiving a RAMS-I message. An RTP-RX implementation MUST NOT assume it will quickly receive the initial RAMS-I message. For redundancy purposes, it is RECOMMENDED that the BRS 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, the BRS starts sending the unicast burst(s) that comprises one or more RTP retransmission packets sent in the unicast burst RTP session. In addition, in some applications the BRS can send preamble information data to the RTP_Rx in addition to the requested burst to prime the media decoder(s). However, for the BRS to send the preamble information in a particular format, explicit signaling from the RTP_Rx is required. In other words, the BRS MUST NOT send preamble information in a particular format unless the RTP_Rx has signaled support for that format in the RAMS-R message through a public or private extension as defined in Section 7.1.

   The format of this preamble data is RTP-payload specific and not specified here.

5. Updated Request: The RTP_Rx MAY send an updated RAMS-R message (as unicast feedback in the primary multicast RTP session) with a different value for one or more fields of an earlier RAMS-R message. If there is already a burst planned for or being transmitted to a particular RTP_Rx for a particular stream, the newly arriving RAMS-R is an updated request; otherwise, it is a new request. It is also possible that the RTP_Rx has sent an improperly formatted RAMS-R message or used an invalid value in the RAMS-R message. If notified by the BRS using a 4xx-level response code (as defined in Section 11.6), the RTP_Rx MAY re-send its corrected request only after following the timing rules of [RFC4585].

   The BRS determines the identity of the requesting RTP_Rx by looking at its canonical name identifier (CNAME item in the SDES source description). Thus, the RTP_Rx MUST choose a globally unique CNAME identifier. Different such ways are provided in [I-D.ietf-avt-rtp-cnames]. In addition to one or more fields with updated values, an updated RAMS-R message may also include the fields whose values did not change.

   Upon receiving an updated request, the BRS can use the updated values for sending/shaping the burst, or refine the values and use the refined values for sending/shaping the burst.

Subsequently, the BRS MAY send an updated RAMS-I message in the unicast burst RTP session to indicate the changes it made.

It is an implementation-dependent decision for an RTP_RX whether and when to send an updated request.

6. Updated Response: The BRS can send more than one RAMS-I messages in the unicast burst RTP session, e.g., to update the value of one or more fields in an earlier RAMS-I message. The updated RAMS-I messages might or might not be a direct response to a RAMS-R message. The BRS can also send updated RAMS-I messages to signal the RTP_Rx in real time to join the SSM session, when the BRS had already sent an initial RAMS-I message, e.g., at the start of the burst. The RTP_Rx depends on the BRS 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 the BRS is not capable of determining the join time in the initial RAMS-I message, the BRS MUST send another RAMS-I message (with the join time information) later.

7. Multicast Join Signaling: The RAMS-I message allows the BRS to signal explicitly when the RTP_Rx needs to send the SFGMP Join message. The RTP_Rx SHOULD use this information from the most recent RAMS-I message unless it has more accurate information. 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.

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

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

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

   If the RTP_Rx wants to terminate a burst prematurely, it SHALL send a RAMS-T message for the SSRC of the primary multicast stream it wishes to terminate. This message is sent in the unicast burst RTP session. Upon receiving this message, the BRS MUST terminate the unicast burst. If the 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) or RTP packets, the RTP_Rx cannot use RAMS-T message(s) and thus MUST send an RTCP BYE message in the unicast burst RTP session to terminate the request.

   Otherwise, the default behavior for the RTP_Rx is to send a RAMS-T message in the unicast burst RTP session immediately after it joins the multicast session and started receiving multicast packets. In that case, the RTP_Rx SHALL send a RAMS-T message with the sequence number of the first RTP packet received in the primary multicast stream. Then, the BRS MUST 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.

   If an RTCP BYE message has not been issued yet as described in Step 10, the RTP_Rx MUST send at least one RAMS-T message for each primary multicast stream served by the BRS. The RAMS-T message(s) MUST be sent to the BRS in the unicast burst RTP session. Against the possibility of a message loss, it is RECOMMENDED that the RTP_Rx repeats the RAMS-T messages multiple times as long as it follows the RTCP timer rules defined in [RFC4585].

   The binding between RAMS-T and ongoing bursts is achieved through RTP_Rx's CNAME identifier, and packet sender and media sender SSRCs. Choosing a globally unique CNAME makes sure that the RAMS-T messages are processed correctly.

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

    Upon receiving an RTCP BYE message, the BRS 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, the RS can figure out how many duplicate RTP packets have been delivered to the 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.

    Since an RTCP BYE message issued for the unicast burst RTP session terminates that session and ceases transmitting any further packets in that session, 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.ietf-avt-multicast-acq-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 an RTP_RX acquires multiple primary multicast streams, it might need to synchronize them for playout. This synchronization is achieved by the help of the RTCP sender reports [RFC3550]. If the playout will start before the RTP_Rx has joined the multicast session, the RTP_Rx needs to 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.

The suggested 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]. Per [RFC4588] "if the original RTP header carried an RTP header extension, the retransmission packet SHOULD carry the same header extension." 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 the BRS. The synchronization information will be sent to the 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]. 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 the BRS can shape the transmission of the unicast burst and how congestion can be dealt within the RAMS process. When the RTP_Rx detects that the unicast burst is causing severe congestion, it can prefer to send a RAMS-T message immediately to stop the unicast burst.

A higher bitrate for the unicast burst naturally conveys the Reference Information and media content to the RTP_Rx faster. This way, the RTP_Rx can start consuming the data sooner, which results in a faster acquisition. A higher bitrate also represents a better utilization of the BRS resources. As the burst may continue until it catches up with the primary multicast stream, the higher the bursting bitrate, the less data the BRS 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 has to be an upper bound on the bandwidth used by the burst.

When the BRS transmits the unicast burst, it is expected to take into account all available information to prevent any packet loss that might take place during the bursting as a result of buffer overflow on the path between the RS and RTP_Rx and at the RTP_Rx itself. The bursting bitrate can 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 the RTP_Rx in the retransmission session, allowing in-session bitrate adaptations for the burst. When these receiver reports indicate packet loss, this can indicate a certain congestion state in the path from the RS to the RTP_Rx.

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

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

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

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 BRS / 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 the 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 the RS and this link is provisioned and dimensioned to carry the burst streams, the bursting bitrate does not exceed the provisioned value. These settings can also be dynamically adapted using application-aware knowledge.

The BRS chooses the initial burst bitrate as follows:

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

- When the BRS cannot determine a reliable bitrate value for the unicast burst (using a through g), it is desirable that the BRS chooses an appropriate initial bitrate not above the nominal bitrate and increases it gradually unless a congestion is detected.

In both cases, during the burst transmission the BRS 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 the BRS relies on RTCP receiver reports, sufficient bandwidth needs to be provided to RTP Rx for RTCP transmission in the unicast burst RTP session. To achieve a reasonable fast adaptation against congestion, it is recommended that the RTP_Rx sends a receiver report at least once every two RTTs between the RS and RTP_Rx. Although the specific heuristics and algorithms that deduce a congestion state and how subsequently the BRS acts are outside the scope of this specification, the following two methods are the best practices:

- Upon detection of a significant amount of packet loss, which the BRS attributes to congestion, the BRS decreases the burst bitrate. The rate by which the BRS increases and decreases the bitrate for the burst can 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 [RFC5762].

- If the congestion is persistent and the BRS has to reduce the burst bitrate to a point where the RTP Rx buffer might underrun or the burst will consume too many resources, the BRS terminates the burst and transmits 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.

Even though there is no congestion experienced during the burst, congestion may anyway arise near the end of the burst as the RTP_Rx eventually needs to join the multicast session. During this brief period both the burst packets and the multicast packets can be simultaneously received by the RTP_Rx, thus enhancing the risk of congestion.

Since the BRS signals the RTP_Rx when the BRS expects the RTP_Rx to send the SFGMP Join message, the BRS can have a rough estimate of when the RTP_Rx will start receiving multicast packets in the SSM session. The BRS can keep on sending burst packets but reduces the bitrate accordingly at the appropriate instant by taking the bitrate of the whole SSM session into account. If the BRS ceases transmitting the burst packets before the burst catches up, any gap resulting from this imperfect switch-over by the RTP_Rx can be later repaired by requesting retransmissions for the missing packets from the RS. The retransmissions can be shaped by the BRS to make sure that they do not cause collateral loss in the primary multicast RTP session and the unicast burst RTP 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 the RTP_Rx sends a RAMS-R message to initiate a rapid acquisition but the message gets lost and the FT does not receive it, the RTP_Rx will get neither a RAMS-I message, nor a unicast burst. In this case, the 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, the RTP_Rx can time out (based on the previous observed response times) and immediately join the SSM session.

In the case the RTP_Rx starts receiving a unicast burst but it does not receive a corresponding RAMS-I message within a reasonable amount of time, the RTP_Rx can 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). If the network is subject to packet loss, it is RECOMMENDED that the BRS 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 the RTP_Rx gives up, or the RAMS-I message is lost, the 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 the RTP_Rx does not reach its destination, the BRS can continue sending burst packets even though the RTP_Rx no longer needs them. In such cases, it is RECOMMENDED that the RTP_Rx repeats the RAMS-T message multiple times based on the RTCP timer rules defined in [RFC4585]. The BRS MUST be provisioned to 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 the BRS accepts to serve the requested burst(s) and establishes the retransmission session, the BRS needs to check the liveness of the RTP_Rx via the RTCP messages and reports the 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] and is also sketched in Figure 4.

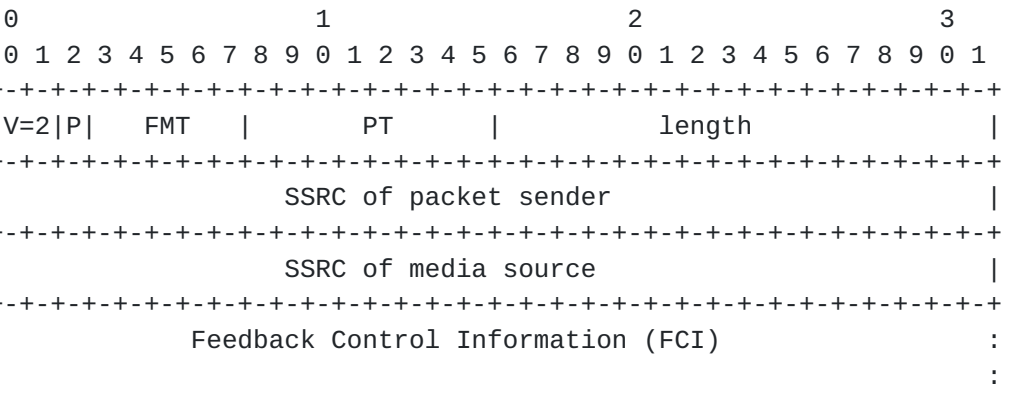


Figure 4: The common packet format for the RTCP feedback messages

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 can 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 stated, numeric constants are in decimal (base 10).

7.1. Extensions

To improve the functionality of the RAMS method in certain applications, it can be desirable to define new fields in the RAMS Request, Information and Termination messages. Such fields MUST be encoded as Type-Length-Value (TLV) elements as described below and sketched in Figure 5:

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

- 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. This length does not include any padding that is required for alignment.

- 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 and mandatory TLV elements (if any). The extensions MAY be placed in any order. In a RAMS message, multiple TLV elements with the same Type value MUST NOT exist.

The support for extensions (unless specified otherwise) is OPTIONAL.

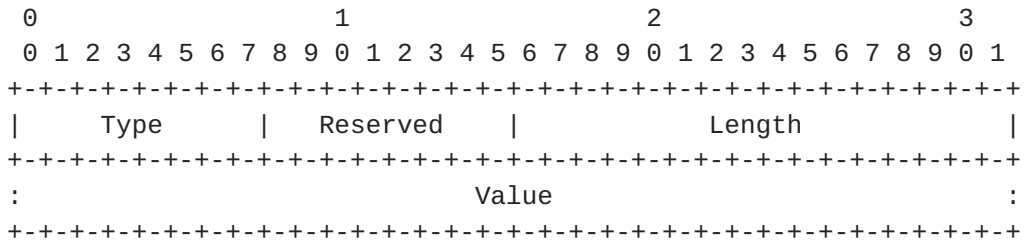


Figure 5: Structure of a TLV element

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 6. 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.

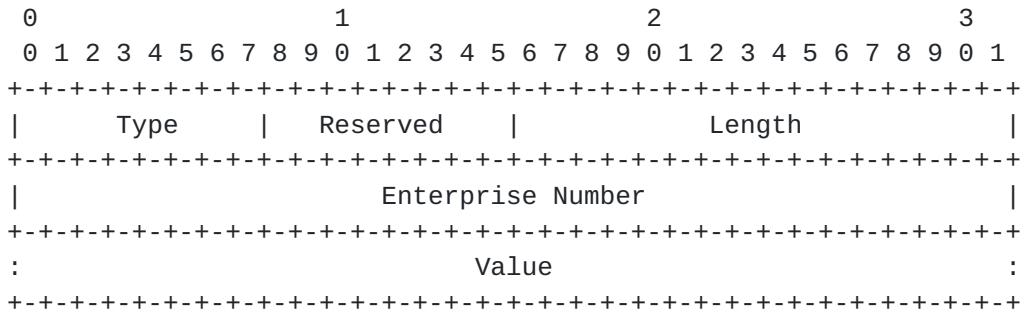


Figure 6: Structure of a private extension

7.2. RAMS Request

The RAMS Request (RAMS-R) message is identified by SFMT=1. This message is sent as unicast feedback in the primary multicast RTP session by the RTP_Rx to request rapid acquisition of a primary multicast RTP session, or one or more primary multicast streams belonging to the same primary multicast RTP session. In the RAMS-R message, the RTP_Rx MUST set both the packet sender SSRC and media sender SSRC fields to its own SSRC since the media sender SSRC may not be known. The RTP_Rx MUST provide explicit signaling as

described below to request stream(s). This minimizes 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 7.

The semantics of the RAMS-R message is independent of the payload type carried in the primary multicast RTP session.

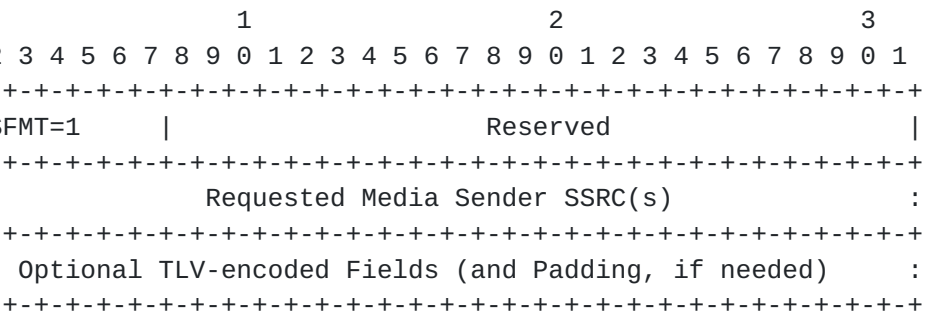


Figure 7: FCI field syntax for the RAMS Request message

o Requested Media Sender SSRC(s): Mandatory TLV element that lists the media sender SSRC(s) requested by the RTP_Rx. The BRS MUST ignore the media sender SSRC specified in the header of the RAMS-R message.

  If the Length field is set to zero (i.e., no media sender SSRC is listed), it means that the RTP_Rx is requesting to rapidly acquire the entire primary multicast RTP session. Otherwise, the 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

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

  The RTP_Rx can have knowledge of its buffering requirements. These requirements can be application and/or device specific. For instance, the RTP_Rx might 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 the BRS is told the minimum buffering requirement of the receiver, the BRS can tailor the burst(s) more precisely, e.g., by choosing an appropriate

starting point. The methods used by the 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. Otherwise, the backfill will not be able to build up the desired level of buffer at the RTP_Rx and can cause buffer underruns.

Type: 2

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

  The RTP_Rx can have knowledge of its buffering requirements. These requirements can be application or device specific. For instance, one particular RTP_Rx might have more physical memory than another RTP_Rx, and thus, can buffer more data. If the BRS knows the buffering ability of the receiver, the BRS can 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. Otherwise, the backfill may cause buffer overflows at the RTP_Rx.

  Type: 3

- Max Receive Bitrate (64 bits): Optional TLV element that denotes the maximum bitrate (in bits per second) at which the RTP_Rx can process the aggregation of the unicast burst(s) and any payload-specific information that will be provided by the BRS. The limits can 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. Otherwise, congestion and packet loss may occur.

  Type: 4

- Preamble-only Allowed (0 bits): Optional TLV element that indicates that the RTP_Rx is willing to receive only the preamble information for the desired primary multicast stream(s) in case

the BRS cannot send the unicast burst stream(s) (In such cases, the BRS uses the response code 511 in the RAMS-I message as defined in Section 11.6). Note that the RTP_Rx signals the particular preamble format(s) it supports through a public or a private extension in the same RAMS-R message.

If this TLV element does not exist in the RAMS-R message, the BRS MUST NOT respond to the RAMS-R message by sending the preamble information only. Since this TLV element does not carry a Value field, the Length field MUST be set to zero.

Type: 5

7.3. RAMS Information

The RAMS Information (RAMS-I) 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 the RTP_Rx as described below.

A separate RAMS-I message with the appropriate response code is sent in the unicast burst RTP session by the BRS for each primary multicast stream that has been requested by the RTP_Rx. In each of these RAMS-I messages, the media sender SSRC and packet sender SSRC fields are both set to the SSRC of the BRS, which equals the SSRC of the respective primary multicast stream.

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

The semantics of the RAMS-I message is independent of the payload type carried in the primary multicast RTP session.

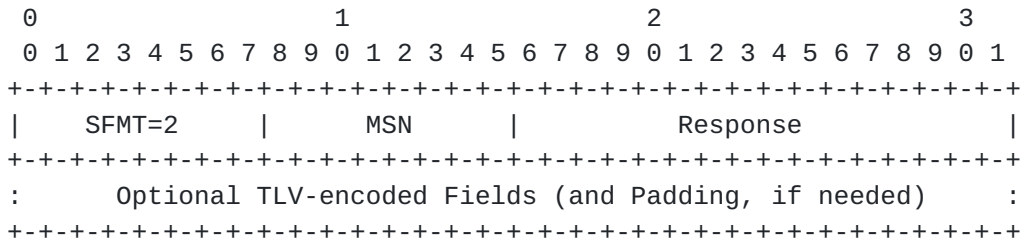


Figure 8: FCI field syntax for the RAMS Information message

A RAMS-I message has the following fields:

- 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.

- Response (16 bits): Mandatory field that denotes the response code for this RAMS-I message. This document defines several initial response codes in Section 11.6 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. When the RTP_Rx receives an RAMS-I message with a private response code that it does not understand, the RTP_Rx still needs to process the TLV elements it understands.

The following TLV elements have been defined for the RAMS-I messages:

- 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 can be useful to repeat it in an explicit field. If the 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, the BRS includes this TLV element in the initial RAMS-I message to let the 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

- 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 unicast RTP stream. This allows the RTP_Rx to know whether one or more packets sent by the BRS have been dropped at the beginning of the stream. If the BRS accepts the RAMS request, this element exists. If the BRS 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 unicast RTP stream (which could be a burst packet or a payload-specific packet) and the earliest time instant when an RTP_Rx MAY send an SFGMP Join message to join the multicast session. A zero value in this field means that the RTP_Rx MAY send the SFGMP Join message right away.

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

  When the BRS 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 the 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 time the burst will last (in ms), i.e., the delta difference between the expected transmission times of the first and the last burst packets that the BRS is planning to send in the respective unicast RTP stream. In the absence of additional stimulus, the BRS will send a burst of this duration. However, the burst duration can be modified by subsequent events, including changes in the primary multicast stream and reception of RAMS-T messages.

  The BRS MUST terminate the flow in the timeframe indicated by this burst duration or the duration established by those subsequent events, even if it does not get a RAMS-T or a BYE message from the 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

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

  Type: 35

7.4. RAMS Termination

The RAMS Termination (RAMS-T) message is identified by SFMT=3.

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

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

The default behavior for the RTP_Rx is to send a RAMS-T message immediately after it joined the multicast session and started receiving multicast packets. In that case, the 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, the BRS 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 9.

The semantics of the RAMS-T message is independent of the payload type carried in the primary multicast RTP session.

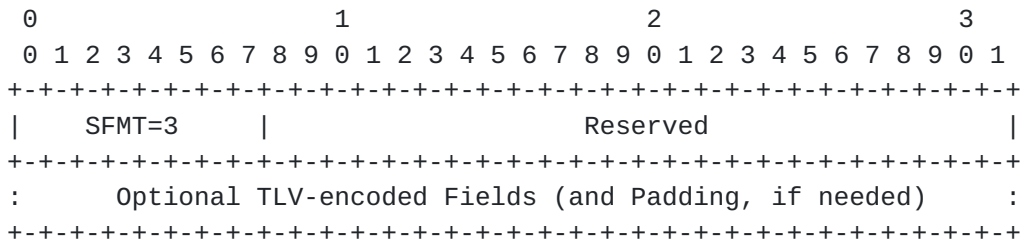


Figure 9: FCI field syntax for the RAMS Termination message

o 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 can be maintained according to the algorithm in Appendix A.1 of

[RFC3550].

Type: 61

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], rtcp-fb-nack-param is used as defined in [RFC4566].)

```
rtcp-fb-nack-param =/ SP "rai"
```

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

- 'rai' stands for Rapid Acquisition 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 the BRS supports updated request messages or not. This attribute is used in a declarative manner and no Offer/Answer Model behavior is specified. If the BRS supports updated request messages and this attribute is included in the SDP description, the RTP_Rx can send updated requests. The BRS might or might 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, the RTP_Rx SHALL NOT send updated requests. The 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 framework and flow identification (FID) semantics [RFC5888]

- The RTP/AVPF profile [RFC4585]

- The RTP retransmission payload format [RFC4588]
- The RTCP extensions for SSM sessions with unicast feedback [RFC5760]
- The multicast RTCP port attribute [I-D.ietf-avt-rtcp-port-for-ssm]
- Multiplexing RTP and RTCP on a single port on both endpoints in the unicast session[RFC5761]

The support for the source-specific media attributes [RFC5576] can also be needed.

8.3. Example and Discussion

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

```
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=rtpmap:98 MP2T/90000
a=multicast-rtcp:42000
a=rtcp:43000 IN IP4 192.0.2.1
a=rtcp-fb:98 nack
a=rtcp-fb:98 nack rai
a=ssrc:123321 cname:iptv-ch32@rams.example.com
a=rams-updates
a=mid:1
m=video 51000 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-mux
a=fmtp:99 apt=98;rtx-time=5000
a=mid:2
```

Figure 10: Example SDP for a single-channel RAMS scenario

In this example SDP description, 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. The corresponding RTCP traffic is multicast to the same multicast destination address at port 42000. Here, we are assuming that the multicast RTP and RTCP ports are carefully chosen so that different RTP and RTCP streams do not collide with each other.

The feedback target (FT_Ap) residing on the RS (with an address of 192.0.2.1) at port 43000 is declared with the "a=rtcp" line [RFC3605]. The support for the conventional retransmission is indicated through the "a=rtcp-fb:98 nack" line. The RTP receiver(s) can report missing packets on the source stream to the feedback target and request retransmissions. The SDP above includes the "a=sendonly" line for the media description of the retransmission stream since the retransmissions are sent in only one direction.

The support for rapid acquisition is indicated through the "a=rtcp-fb:98 nack rai" line. The parameter 'rtx-time' applies to both the conventional retransmissions and rapid acquisition. However, when rapid acquisition is enabled, the standard definition for the parameter 'rtx-time' given in [RFC4588] is not followed. Instead, when rapid acquisition support is enabled, 'rtx-time' specifies the time in milliseconds that the BRS keeps an RTP packet in its cache available for retransmission (measured from the time the packet was received by the BRS, not from the time indicated in the packet timestamp).

Once an RTP_Rx has acquired an SDP description, it can ask for rapid acquisition before it joins a primary multicast RTP session. To do so, it sends a RAMS-R message to the feedback target of that primary multicast RTP session. If the FT_Ap is associated with only one RTP stream, the RTP_Rx does not need to learn the SSRC of that stream via an out-of-band method. If the BRS accepts the rapid acquisition request, it will send an RAMS-I message with the correct SSRC identifier. If the FT_Ap is associated with a multi-stream RTP session and the RTP_Rx is willing to request rapid acquisition for the entire session, the RTP_Rx again does not need to learn the SSRCs via an out-of-band method. However, if the RTP_Rx intends 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_Rx has not yet received any RTP packets for the primary multicast stream(s), the RTP_Rx must in this case learn the SSRC value(s) from the 'ssrc' attribute of the media description [RFC5576]. In addition to the SSRC value, the 'cname' source attribute must also be present in the SDP description [RFC5576].

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_Rx that observed an SSRC collision with a media sender must change its own SSRC [RFC5760] by following the rules defined in [RFC3550].

A feedback target that receives a RAMS-R message becomes aware that the RTP_Rx wants to rapidly catch up with a primary multicast RTP session. If the necessary conditions are satisfied (as outlined in Section 7 of [RFC4585]) and available resources exist, the BRS can react to the RAMS-R message by sending any transport-layer (and optional payload-specific, when allowed) feedback message(s) and starting the unicast burst.

In this section, we considered the simplest scenario where the primary multicast RTP session carried only one stream and the RTP_Rx 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_Rx 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) can exist between the RTP_Rx and RS. NATs have a variety of operating characteristics for UDP traffic [RFC4787]. For a NAT to permit traffic from the BRS to arrive at the RTP_Rx, the NAT(s) must first either:

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

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

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

When the RTP_Rx sends a request (RAMS-R) message to the FT as unicast feedback in the primary multicast RTP session, and the request is received by the FT, a new unicast burst RTP session will be established between the BRS and RTP_Rx.

While the FT and BRS ports on the RS are already signaled via out-of-band means (e.g., SDP), the RTP_Rx needs to convey the RTP and RTCP ports it wants to use on its side for the new session. Since there are two RTP sessions (one multicast and one unicast) 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.

Applications using RAMS MUST support the solution presented in [I-D.ietf-avt-ports-for-ucast-mcast-rtp] both on the RS and RTP_Rx side to allow RTP receivers to use their desired ports and to support RAMS behind NAT devices. The port mapping message exchange needs to take place whenever the RTP_Rx intends to make use of the RAMS protocol for rapidly acquiring a specific multicast RTP session, prior to joining the associated SSM session.

10. Security Considerations

Applications that are using RAMS make heavy use of the unicast feedback mechanism described in [RFC5760], the payload format defined in [RFC4588] and the port mapping solution specified in [I-D.ietf-avt-ports-for-ucast-mcast-rtp]. Thus, these applications are subject to the general security considerations discussed in [RFC5760], [RFC4588] and [I-D.ietf-avt-ports-for-ucast-mcast-rtp]. 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 protected.

Compared to an RTCP NACK message that triggers one or more retransmissions, a RAMS Request (RAMS-R) message can 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 can potentially contain 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 servers, RTP receivers and the network. In particular, this can make 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 can 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 can overflow the buffer at the receiver side, or oppositely, can slow down the burst to the point that it becomes useless. Tampered RAMS Termination (RAMS-T) messages can terminate valid burst streams prematurely resulting in gaps in the received RTP packets. RAMS Information (RAMS-I) messages contain fields that are critical for a successful rapid acquisition. Any tampered information in the RAMS-I message can easily cause an RTP receiver to make wrong decisions. Consequently, the RAMS operation can fail.

While most of the denial-of-service attacks can be prevented by the integrity and authenticity checks enabled by Secure RTP (SRTP) [RFC3711], 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 can temporarily consume a large amount of resources, a series of the RAMS-R messages can temporarily overload the retransmission server. In these circumstances, the retransmission server can, 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 [RFC5760].

[RFC4588] RECOMMENDS that the cryptography mechanisms are used for the retransmission payload format to provide protection against known plain-text 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].

To protect the RTCP messages from man-in-the-middle and replay attacks, the RTP receivers and retransmission server SHOULD perform a DTLS-SRTP handshake [RFC5764] over the RTCP channel. Because there is no integrity-protected signaling channel between an RTP receiver and the retransmission server, the retransmission server MUST maintain a list of certificates owned by legitimate RTP receivers, or their certificates MUST be signed by a trusted Certificate Authority. Once the DTLS-SRTP security is established, non-SRTCP-protected messages received from a particular RTP receiver are ignored by the retransmission server. To reduce the impact of DTLS-SRTP overhead when communicating with different feedback targets on the same retransmission server, it is RECOMMENDED that RTP receivers and the retransmission server both support TLS Session Resumption without Server-side State [RFC5077]. To help scale SRTP to handle many RTP receivers asking for retransmissions of identical data, implementors may consider using the same SRTP key for SRTP data sent to the receivers [I-D.ietf-avt-srtp-ekt] and consider the security of such SRTP key sharing.

11. IANA Considerations

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

Ali Begen
abegen@cisco.com

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 the 'rtcp-fb' attribute, refer to Sections 4.2 and 6.2 of [RFC4585].

```
Value name:     rai
Long name:      Rapid Acquisition 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	Assignable - Specification Required	
255	Reserved	[RFCXXXX]

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

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

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

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

New RAMS functionality is intended to 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	Assignable - 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	Assignable - Specification Required	
61	Extended RTP Seqnum of First Multicast Packet	[RFCXXXX]
62-127	Assignable - Specification Required	
128-254	No IANA Maintenance	
255	Reserved	[RFCXXXX]

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

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

- 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 should be classified following the guidelines below:

Code Level	Purpose
1xx	Informational
2xx	Success
3xx	Redirection
4xx	RTP Receiver (RTP_Rx) Error
5xx	Burst/Retransmission Source (BRS) Error

The Response code 65535 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	[RFCXXXX]
401	Invalid min buffer requirement in RAMS-R message	[RFCXXXX]
402	Invalid max buffer requirement in RAMS-R message	[RFCXXXX]
403	Insufficient max bitrate requirement in RAMS-R message	[RFCXXXX]
404	Invalid RAMS-T message syntax	[RFCXXXX]
500	An unspecified BRS internal error has occurred	[RFCXXXX]
501	BRS has insufficient bandwidth to start RAMS session	[RFCXXXX]
502	Burst is terminated due to network congestion	[RFCXXXX]
503	BRS has insufficient CPU cycles to start RAMS session	[RFCXXXX]
504	RAMS functionality is not available on BRS	[RFCXXXX]
505	RAMS functionality is not available for RTP_Rx	[RFCXXXX]
506	RAMS functionality is not available for the requested multicast stream	[RFCXXXX]
507	BRS has no valid starting point available for the requested multicast stream	[RFCXXXX]
508	BRS has no reference information available for the requested multicast stream	[RFCXXXX]
509	BRS 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]

The definitions for these codes are provided in Section 11.6.1.

Any registration for an unassigned Response code needs to contain the following information:

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

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

11.6.1. Response Code Definitions

1xx-Level Response Codes: These response codes are sent for informational purposes.

- 100: This is used when the BRS wants to update a value that was sent earlier to the RTP_Rx.

2xx-Level Response Codes: These response codes are sent to indicate success.

- 200: This is used when the server accepts the RAMS request.

- 201: This is used when the unicast burst has been completed and the BRS wants to notify the RTP_Rx.

4xx-Level Response Codes: These response codes are sent to indicate that the message sent by the RTP_Rx is either improperly formatted or contains an invalid value. The rules the RTP_Rx needs to follow upon receiving one of these response codes are outlined in Step 5 in Section 6.2. The 4xx-level response codes are also used as status codes in the Multicast Acquisition Report Block [I-D.ietf-avt-multicast-acq-rtcp-xr] in order to collect RTP_Rx-based error conditions.

- 400: This is used when the RAMS-R message is improperly formatted.

- 401: This is used when the minimum RAMS buffer fill requirement value indicated in the RAMS-R message is invalid.

- 402: This is used when the maximum RAMS buffer fill requirement value indicated in the RAMS-R message is invalid.

- 403: This is used when the maximum receive bitrate value indicated in the RAMS-R message is insufficient.

- 404: This is used when the RAMS-T message is improperly formatted.

5xx-Level Response Codes: These response codes are sent to indicate an error on the BRS side. The rules the RTP_Rx needs to follow upon receiving one of these response codes are outlined in Step 3 in Section 6.2 and Section 7.2. The 5xx-level response codes are also used as status codes in the Multicast Acquisition Report Block [I-D.ietf-avt-multicast-acq-rtcp-xr] in order to collect BRS-based error conditions.

- 500: This is used when the BRS has experienced an internal error and cannot accept the RAMS request.
- 501: This is used when the BRS does not have enough bandwidth to send the unicast burst stream.
- 502: This is used when the BRS terminates the unicast burst stream due to network congestion.
- 503: This is used when the BRS does not have enough CPU resources to send the unicast burst stream.
- 504: This is used when the BRS does not support sending a unicast burst stream.
- 505: This is used when the requesting RTP_Rx is not eligible to receive a unicast burst stream.
- 506: This is used when RAMS functionality is not enabled for the requested multicast stream.
- 507: This is used when the BRS cannot find a valid starting point for the unicast burst stream satisfying the RTP_Rx's requirements.
- 508: This is used when the BRS cannot find the essential reference information for the requested multicast stream.
- 509: This is used when the BRS cannot match the requested SSRC to any of the streams it is serving.
- 510: This is used when the BRS cannot serve the requested entire session.
- 511: This is used when the BRS sends only the preamble information but not the whole unicast burst stream.
- 512: This is used when the RAMS request is denied due to a policy specified for the requested multicast stream, requesting RTP_Rx or this particular BRS.

12. Contributors

Dave Oran, Magnus Westerlund and Colin Perkins 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: 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, Jose Rey, Sean Sheedy and Keith Drage.

14. References

14.1. 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.

[RFC5888] Camarillo, G. and H. Schulzrinne, "The Session Description Protocol (SDP) Grouping Framework", RFC 5888, June 2010.

[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.

[RFC5760] Ott, J., Chesterfield, J., and E. Schooler, "RTP Control Protocol (RTCP) Extensions for Single-Source Multicast Sessions with Unicast Feedback", RFC 5760, February 2010.

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

[RFC3605] Huitema, C., "Real Time Control Protocol (RTCP) attribute in Session Description Protocol (SDP)", RFC 3605, October 2003.

[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-12 (work in progress), July 2010.

[RFC5761] Perkins, C. and M. Westerlund, "Multiplexing RTP Data and Control Packets on a Single Port", RFC 5761, April 2010.

[I-D.ietf-avt-rtcp-port-for-ssm] Begen, A., "RTP Control Protocol (RTCP) Port for Source-Specific Multicast (SSM) Sessions", draft-ietf-avt-rtcp-port-for-ssm-02 (work in progress), August 2010.

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

[RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.

[RFC5764] McGrew, D. and E. Rescorla, "Datagram Transport Layer Security (DTLS) Extension to Establish Keys for the Secure

Real-time Transport Protocol (SRTP)", RFC 5764, May 2010.

[RFC5077] Salowey, J., Zhou, H., Eronen, P., and H. Tschofenig, "Transport Layer Security (TLS) Session Resumption without Server-Side State", RFC 5077, January 2008.

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

14.2. 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.ietf-avt-rtp-cnames]
Begen, A., Perkins, C., and D. Wing, "Guidelines for Choosing RTP Control Protocol (RTCP) Canonical Names (CNAMEs)", draft-ietf-avt-rtp-cnames-01 (work in progress), August 2010.

[I-D.ietf-avt-multicast-acq-rtcp-xr]
Begen, A. and E. Friedrich, "Multicast Acquisition Report Block Type for RTP Control Protocol (RTCP) Extended Reports (XRs)", draft-ietf-avt-multicast-acq-rtcp-xr-01 (work in progress), May 2010.

[I-D.ietf-avt-ecn-for-rtp]
Westerlund, M., Johansson, I., Perkins, C., and K. Carlberg, "Explicit Congestion Notification (ECN) for RTP over UDP", draft-ietf-avt-ecn-for-rtp-02 (work in progress), July 2010.

[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.

[RFC5762] Perkins, C., "RTP and the Datagram Congestion Control

Protocol (DCCP)", RFC 5762, April 2010.

[I-D.ietf-avt-srtp-ekt]
McGrew, D., Andreasen, F., Wing, D., and K. Fischer, "Encrypted Key Transport for Secure RTP", draft-ietf-avt-srtp-ekt-01 (work in progress), July 2010.

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

[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
181 Bay Street
Toronto, ON M5J 2T3
Canada

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