AVTCORE WG
Internet-Draft
Updates: 3550 (if approved)
Intended status: Standards Track
Expires: January 10, 2013

M. Westerlund
Ericsson
C. Perkins
University of Glasgow
J. Lennox
Vidyo

July 9, 2012

Multiple Media Types in an RTP Session
draft-westerlund-avtcore-multi-media-rtp-session-00

Abstract

This document specifies how an RTP session can contain media streams with media from multiple media types such as audio, video, and text. This has been restricted by the RTP Specification, and thus this document updates RFC 3550 to enable this behavior for applications that satisfy the applicability for using multiple media types in a single RTP session.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

When the Real-time Transport Protocol (RTP) [RFC3550] was designed, close to 20 years ago, IP networks were very different compared to the ones in 2012 when this is written. The almost ubiquitous deployment of Network Address Translators (NAT) and Firewalls has increased the cost and likely-hood of communication failure when using many different transport flows. Thus there exists a pressure to reduce the number of concurrent transport flows.

RTP [RFC3550] as defined recommends against having multiple media types, like audio and video, in the same RTP session. The motivation for this is dependent on particular usage or dependencies on lower layer Quality of Service (QoS). When these aren't present, there are no strong RTP reasons for not allowing multiple media types in one RTP session. However, the Session Description Protocol (SDP) [RFC4566], as one of the dominant signalling method for establishing RTP session, has enforced this rule, by not allowing multiple media types for a given receiver destination or set of ICE candidates, which is the most common method to determine which RTP session the packets are intended for.

The fact that these limitations have been in place for so long a time, in addition to RFC 3550 being written without fully considering multiple media types in an RTP session, does result in a number of considerations being needed. This document provides such considerations regarding applicability as well as functionality, including normative specification of behavior.

First, some basic definitions are provided. This is followed by a background that discusses the motivation in more detail. A overview of the solution of how to provide multiple media types in one RTP session is then presented. Next is the formal applicability this specification have followed by the normative specification. This is followed by a discussion how some RTP/RTCP Extensions should function in the case of multiple media types in one RTP session. A specification of the requirements on signalling from this specification and a look how this is realized in SDP using Bundle [I-D.ietf-mmusic-sdp-bundle-negotiation]. The document ends with the security considerations.

2. Definitions

2.1. Requirements Language

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].

2.2. Terminology

The following terms are used with supplied definitions:

Endpoint: A single entity sending or receiving RTP packets. It may be decomposed into several functional blocks, but as long as it behaves as a single RTP stack entity it is classified as a single endpoint.

Media Stream: A sequence of RTP packets using a single SSRC that together carries part or all of the content of a specific Media Type from a specific sender source within a given RTP session.

Media Type: Audio, video, text or application whose form and meaning are defined by a specific real-time application.

RTP Session: As defined by [RFC3550], the endpoints belonging to the same RTP Session are those that share a single SSRC space. That is, those endpoints can see an SSRC identifier transmitted by any one of the other endpoints. An endpoint can receive an SSRC either as SSRC or as CSRC in RTP and RTCP packets. Thus, the RTP Session scope is decided by the endpoints' network interconnection topology, in combination with RTP and RTCP forwarding strategies deployed by endpoints and any interconnecting middle nodes.

3. Motivation

This section discusses in more detail the main motivations why allowing multiple media types in the same RTP session is suitable.

3.1. NAT and Firewalls

The existence of NATs and Firewalls at almost all Internet access has had implications on protocols like RTP that were designed to use multiple transport flows. First of all, the NAT/FW traversal solution one uses needs to ensure that all these transport flows are established. This has three different impacts:

1. Increased delay to perform the transport flow establishment

2. The more transport flows, the more state and the more resource consumption in the NAT and Firewalls. When the resource consumption in NAT/FWs reaches their limits, unexpected behaviors usually occur.

3. More transport flows means a higher risk that some transport flow fails to be established, thus preventing the application to communicate.

Using fewer transport flows reduces the risk of communication failure, improved establishment behavior and less load on NAT and Firewalls.

3.2. No Transport Level QoS

Many RTP-using applications don't utilize any network level Quality of Service functions. Nor do they expect or desire any separation in network treatment of its media packets, independent of whether they are audio, video or text. When an application has no such desire, it doesn't need to provide a transport flow structure that simplifies flow based QoS.

3.3. Architectural Equality

For applications that don't desire any type of different treatment, neither on the transport level nor in RTP or RTCP reporting, using the same RTP session for both media types appears a reasonable choice. The architecture should be neutral to media type, rather look at what it provides based on the application users choice. Therefore this bias should be removed and let the application designer make the choice if they need multiple RTP sessions or not based on other aspects.

4. Overview of Solution

The goal of the solution is to enable having one or more RTP sessions, where each RTP session may contain two or more media types. This includes having multiple RTP sessions containing a given media type, for example having three sessions containing video and audio.

The solution is quite straightforward. The first step is to override the SHOULD and SHOULD NOT language of the RTP specification [RFC3550]. This is done by appropriate exception clauses given that this specification is followed.

Within an RTP session where multiple media types have been configured for use, a SSRC may only send one media type during its lifetime. Different SSRCs must be used for the different media sources, the same way multiple media sources of the same media type already have to do. The payload type will inform a receiver which media type the SSRC is being used for. Thus the payload type must be unique across all of the payload configurations independent of media type that may be used in the RTP session.

Some few extra considerations within the RTP sessions also needs to be considered. RTCP bandwidth and regular reporting suppression (AVPF and SAVPF) should be considered to be configured. Certain payload types like FEC also need additional rules.

The final important part of the solution to this is to use signalling and ensure that agreement on using multiple media types in an RTP session exists, and how that then is configured. Thus document documents some existing requirements, while an external reference defines how this is accomplished in SDP.

5. Applicability

This specification has limited applicability and any one intending to use must ensure that their application and usage meets the below criteria for usage.

5.1. Usage of the RTP session

Before choosing to use this specification, an application implementer needs to ensure that they don't have a need for different RTP sessions between the media types for some reason. The main rule is that if one expects to have equal treatment of all media packets, then this specification might be suitable. The equal treatment include anything from network level up to RTCP reporting and feedback. The document Guidance on RTP Multiplexing Architecture [I-D.westerlund-avtcore-multiplex-architecture] gives more detailed guidance on aspects to consider when choosing how to use RTP and specifically sessions. RTP-using applications that need or would prefer multiple RTP sessions, but do not require the functionalities or behaviors that multiple transport flows give, can consider using Multiple RTP Sessions on a Single Lower-Layer Transport [I-D.westerlund-avtcore-transport-multiplexing].

5.2. Signalled Support

Usage of this specification is not compatible with anyone following RFC 3550 and intending to have different RTP sessions for each media

type. Therefore there must be mutual agreement to use multiple media types in one RTP session by all participants within an RTP session. This agreement must in most cases be determined using signalling.

This requirement can be a problem for signalling solutions that can't negotiate with all participants. For declarative signalling solutions, mandating that the session is using multiple media types in one RTP session can be a way of attempting to ensure that all participants in the RTP session follow the requirement. However, for signalling solutions that lack methods for enforcing that a receiver supports a specific feature, this can still cause issues.

5.3. Homogeneous Multi-party

In multiparty communication scenarios it is important to separate two different cases. One case is where the RTP session contains multiple participants in a common RTP session. This occurs for example in Any Source Multicast (ASM) and Transport Translator topologies as defined in RTP Topologies [RFC5117]. It may also occur in some implementations of RTP mixers that share the same SSRC/CSRC space across all participants. The second case is when the RTP session is terminated in a middlebox and the other participants sources are projected or switched into each RTP session and rewritten on RTP header level including SSRC mappings.

For the first case, with a common RTP session or at least shared SSRC/CSRC values, all participants in multiparty communication are required to support multiple media types in an RTP session. An participant using two or more RTP sessions towards a multiparty session can't be collapsed into a single session with multiple media types. The reason is that in case of multiple RTP sessions, the same SSRC value can be use in both RTP sessions without any issues, but when collapsed to a single session there is an SSRC collision. In addition some collisions can't be represented in the multiple separate RTP sessions. For example, in a session with audio and video, an SSRC value used for video will not show up in the Audio RTP session at the participant using multiple RTP sessions, and thus not trigger any collision handling. Thus any application using this type of RTP session structure must have a homogeneous support for multiple media types in one RTP session, or be forced to insert a translator node between that participant and the rest of the RTP session.

For the second case of separate RTP sessions for each multiparty participant and a central node it is possible to have a mix of single RTP session users and multiple RTP session users as long as one is willing to remap the SSRCs used by a participant with multiple RTP sessions into non-used values in the single RTP session SSRC space for each of the participants using a single RTP session with multiple

media types. It can be noted that this type of implementation is required to understand any type of RTP/RTCP extension being used in the RTP sessions to correctly be able to translate them between the RTP sessions.

5.4. Reduced number of Payload Types

An RTP session with multiple media types in it have only a single 7-bit Payload Type range for all its payload types. Within the 128 available values, only 96 or less if "Multiplexing RTP Data and Control Packets on a Single Port" [RFC5761] is used, all the different RTP payload configurations for all the media types must fit. For most applications this will not be a real problem, but the limitation exists and could be encountered.

5.5. Stream Differentiation

If network level differentiation of the media streams of different media types are desired using this specification can cause severe limitations. All media streams in an RTP session, independent of the media type, will be sent over the same underlying transport flow. Any flow-based Quality of Service (QoS) mechanism will be unable to provide differentiated treatment between different media types, e.g. to prioritize audio over video. If that is desired, separate RTP sessions over different underlying transport flows needs to be used. Any marking-based QoS scheme like DiffServ is not affected unless a network ingress marks based on flows.

5.6. Non-compatible Extensions

There exist some RTP and RTCP extensions that rely on the existence of multiple RTP sessions. If the goal of using an RTP session with multiple media types is to have only a single RTP session, then these extensions can't be used. If one has no need to have different RTP sessions for the media types but is willing to have multiple RTP sessions, one for the main media transmission and one for the extension, they can be used. It should be noted that this assumes that it is possible to get the extension working when the related RTP session contains multiple media types.

Identified RTP/RTCP extensions that require multiple RTP Sessions are:

RTP Retransmission: RTP Retransmission [RFC4588] has a session multiplexed mode. It also has a SSRC multiplexed mode that can be used instead. So use the mode that is suitable for the RTP application.

XOR-Based FEC: The RTP Payload Format for Generic Forward Error Correction [RFC5109] and its predecessor [RFC2733] requires a separate RTP session unless the FEC data is carried in RTP Payload for Redundant Audio Data [RFC2198] which has another set of restrictions.

Note that the Source-Specific Media Attributes [RFC5576] specification defines an SDP syntax (the "FEC" semantic of the "ssrc-group" attribute) to signal FEC relationships between multiple media streams within a single RTP session. However, this can't be used as the FEC repair packets are required to have the same SSRC value as the source packets being protected. [RFC5576] does not normatively update and resolve that restriction.

6. RTP Session Specification

This section defines what needs to be done or avoided to make an RTP session with multiple media types function without issues.

6.1. RTP Session

Section 5.2 of "RTP: A Transport Protocol for Real-Time Applications" [RFC3550] states:

> For example, in a teleconference composed of audio and video media encoded separately, each medium SHOULD be carried in a separate RTP session with its own destination transport address.
>
> Separate audio and video streams SHOULD NOT be carried in a single RTP session and demultiplexed based on the payload type or SSRC fields.

This specification changes both of these sentences. The first sentence is changed to:

> For example, in a teleconference composed of audio and video media encoded separately, each medium SHOULD be carried in a separate RTP session with its own destination transport address, unless specification [RFCXXXX] is followed and the application meets the applicability constraints.

The second sentence is changed to:

> Separate audio and video streams SHOULD NOT be carried in a single RTP session and demultiplexed based on the payload type or SSRC fields, unless multiplexed based on both SSRC and payload type and usage meets what Multiple Media Types in an RTP Session [RFCXXXX]

specifies.

RFC-Editor Note: Please replace RFCXXXX with the RFC number of this specification when assigned.

TBD: Discussion of the motivations in Section 5.2 of the RTP Specification [RFC3550].

6.2. Sender Source Restrictions

A SSRC in the RTP session MUST only send one media type (audio, video, text etc.) during the SSRC's lifetime. The main motivation is that a given SSRC has its own RTP timestamp and sequence number spaces. The same way that you can't send two streams of encoded audio on the same SSRC, you can't send one audio and one video encoding on the same SSRC. Each media encoding when made into an RTP stream needs to have the sole control over the sequence number and timestamp space. If not, one would not be able to detect packet loss for that particular stream. Nor can one easily determine which clock rate a particular SSRCs timestamp shall increase with.

6.3. Payload Type Applicability

Most Payload Types have a native media type, like an audio codec is natural belonging to the audio media type. However, there exist a number of RTP payload types that don't have a native media type. For example, transport robustification mechanisms like RTP Retransmission [RFC4588] and Generic FEC [RFC5109] inherit their media type from what they protect. RTP Retransmission is explicitly bound to the payload type it is protecting, and thus will inherit it. However Generic FEC is a excellent example of an RTP payload type that has no natural media type. The media type for what it protects is not relevant as it is the recovered RTP packets that have a particular media type, and thus Generic FEC is best categorized as an application media type.

The above discussion is relevant to what limitations exist for RTP payload type usage within an RTP session that has multiple media types. When it comes to Generic FEC, is an configured payload type allowed to be used to protect both audio SSRCs and Video SSRCs? Note a particular SSRC carrying Generic FEC will clearly only protect a specific SSRC and thus that instance is bound to the SSRC's media type. For this specific case, it appears possible to have one be applicable to both. However, in cases when the signalling is setup to enable fallback to using separate RTP sessions, then using a different media type, e.g. application, than the media being protected can create issues.

TBD: What recommendations are needed here?

6.4. RTCP

All SSRCs in an RTP session fall under the same set of RTCP configuration parameters, such as the RR and RS bandwidth and the trr-int parameter if AVPF or SAVPF is used. This means that at least the regular reporting period by, and on, a source will be equal, independent of the media type for that source. This should in most cases not be an issue, but it may result in more frequent reporting than is considered necessary for a particular media type or set of media sources. Having multiple media types in one RTP session also results in more SSRCs being present in this RTP session. This increasing the amount of cross reporting between the SSRCs. From an RTCP perspective, two RTP sessions with half the number of SSRCs in each will be slightly more efficient. If someone needs either the higher efficiency due to the lesser number of SSRCs or the fact that one can't tailor RTCP usage per media type, they need to use independent RTP sessions.

When it comes to handling multiple SSRCs in an RTP session there is a clarification under discussion in Real-Time Transport Protocol (RTP) Considerations for Multi-Stream Endpoints [I-D.lennox-avtcore-rtp-multi-stream]. When it comes to configuring RTCP the need for regular periodic reporting needs to be weighted against any feedback or control messages being sent. The applications using AVPF or SAVPF are RECOMMENDED to consider setting trr-int parameter to a value suitable for the applications needs, thus potentially reducing the need for regular reporting and thus releasing more bandwidth for use for feedback or control.

Another aspect of an RTP session with multiple media types is that the used RTCP packets, RTCP Feedback Messages, or RTCP XR metrics used may not be applicable to all media types. Instead all RTP/RTCP endpoints need to correlate the media type of the SSRC being referenced in an messages/packet and only use those that apply to that particular SSRC and its media type. Signalling solutions may have shortcomings when it comes to indicate that a particular set of RTCP reports or feedback messages only apply to a particular media type within an RTP session.

7. Extension Considerations

This section discusses the impact on some RTP/RTCP extensions due to usage of multiple media types in on RTP session. Only extensions where something worth noting has been included.

7.1. RTP Retransmission

SSRC-multiplexed RTP retransmission [RFC4588] is actually very straightforward. Each retransmission RTP payload type is explicitly connected to an associated payload type. If retransmission is only to be used with a subset of all payload types, this is not a problem, as it will be evident from the retransmission payload types which payload types that have retransmission enabled for them.

Session-multiplexed RTP retransmission is also possible to use where an retransmission session contains the retransmissions of the associated payload types in the source RTP session. The only difference to previously is that the source RTP session is one which contains multiple media types. Thus it is even more likely that only a subset of the source RTP session's payload types and SSRCs are actually retransmitted.

Open Issue: When using SDP to signal retransmission for one RTP session with multiple media types and one RTP session for the retransmission data will cause a situation where one will have multiple m= lines grouped using FID and the ones belonging to respective RTP session being grouped using BUNDLE. This usage may contradict both the FID semantics [RFC5888] and an assumption in the RTP retransmission specification [RFC4588].

7.2. Generic FEC

TBW:

8. Signalling

The Signalling requirements

Establishing an RTP session with multiple media types requires signalling. This signalling needs to fulfill the following requirements:

1. Ensure that any participant in the RTP session is aware that this is an RTP session with multiple media types.

2. Ensure that the payload types in use in the RTP session are using unique values, with no overlap between the media types.

3. Configure the RTP session level parameters, such as RTCP RR and RS bandwidth, AVPF trr-int, underlying transport, the RTCP extensions in use, and security parameters, commonly for the RTP session.

4. RTP and RTCP functions that can be bound to a particular media type should be reused when possible also for other media types, instead of having to be configured for multiple code-points. Note: In some cases one will not have a choice but to use multiple configurations.

8.1. SDP-Based Signalling

The signalling of multiple media types in one RTP session in SDP is specified in "Multiplexing Negotiation Using Session Description Protocol (SDP) Port Numbers" [I-D.ietf-mmusic-sdp-bundle-negotiation].

9. IANA Considerations

This document makes no request of IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

10. Security Considerations

Having an RTP session with multiple media types doesn't change the methods for securing a particular RTP session. One possible difference is that the different media have often had different security requirements. When combining multiple media types in one session, their security requirements must also be combined by selecting the most demanding for each property. Thus having multiple media types may result in increased overhead for security for some media types to ensure that all requirements are meet.

Otherwise, the recommendations for how to configure and RTP session do not add any additional requirements compared to normal RTP, except for the need to be able to ensure that the participants are aware that it is a multiple media type session. If not that is ensured it can cause issues in the RTP session for both the unaware and the aware one. Similar issues can also be produced in an normal RTP session by creating configurations for different end-points that doesn't match each other.

11. Acknowledgements

The authors would like to thank Christer Holmberg for the feedback on the document.

12. References

12.1. Normative References

[I-D.ietf-mmusic-sdp-bundle-negotiation]
Holmberg, C. and H. Alvestrand, "Multiplexing Negotiation Using Session Description Protocol (SDP) Port Numbers", draft-ietf-mmusic-sdp-bundle-negotiation-00 (work in progress), February 2012.

[I-D.lennox-avtcore-rtp-multi-stream]
Lennox, J. and M. Westerlund, "Real-Time Transport Protocol (RTP) Considerations for Multi-Stream Endpoints", July 2012.

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

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

12.2. Informative References

[I-D.westerlund-avtcore-multiplex-architecture]
Westerlund, M., Burman, B., and C. Perkins, "RTP Multiplexing Architecture", draft-westerlund-avtcore-multiplex-architecture-01 (work in progress), March 2012.

[I-D.westerlund-avtcore-transport-multiplexing]
Westerlund, M. and C. Perkins, "Multiple RTP Sessions on a Single Lower-Layer Transport", draft-westerlund-avtcore-transport-multiplexing-02 (work in progress), March 2012.

[RFC2198] Perkins, C., Kouvelas, I., Hodson, O., Hardman, V., Handley, M., Bolot, J., Vega-Garcia, A., and S. Fosse-Parisis, "RTP Payload for Redundant Audio Data", RFC 2198, September 1997.

[RFC2733] Rosenberg, J. and H. Schulzrinne, "An RTP Payload Format for Generic Forward Error Correction", RFC 2733, December 1999.

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

[RFC4588] Rey, J., Leon, D., Miyazaki, A., Varsa, V., and R. Hakenberg, "RTP Retransmission Payload Format", RFC 4588, July 2006.

[RFC5109] Li, A., "RTP Payload Format for Generic Forward Error Correction", RFC 5109, December 2007.

[RFC5117] Westerlund, M. and S. Wenger, "RTP Topologies", RFC 5117, January 2008.

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

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

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

Authors' Addresses

Magnus Westerlund
Ericsson
Farogatan 6
SE-164 80 Kista
Sweden

Phone: +46 10 714 82 87
Email: magnus.westerlund@ericsson.com

Colin Perkins
University of Glasgow
School of Computing Science
Glasgow G12 8QQ
United Kingdom

Email: csp@csperkins.org

Jonathan Lennox
Vidyo, Inc.
433 Hackensack Avenue
Seventh Floor
Hackensack, NJ 07601
US

Email: jonathan@vidyo.com