

Internet Engineering Task Force
Internet-Draft
Intended status: BCP
Expires: September 15, 2011

G. Hellstrom
Omnitor
A. van Wijk
Real-Time Text Taskforce (R3TF)
March 14, 2011

**Text media handling in RTP based real-time conferences
draft-hellstrom-text-conference-04**

Abstract

This memo specifies methods for text media handling in multi-party calls, where the text is carried by the RTP protocol. Real-time text is carried in a time-sampled mode according to [RFC 4103](#). Centralized multi-party handling of real-time text is achieved through a media control unit coordinating multiple RTP text streams into one single stream RTP session, identifying each stream with its own CSRC. Identification for the streams are provided through the RTCP messages. This mechanism enables the receiving application to present the received real-time text medium in different ways according to user preferences. Some presentation related features are also described explaining suitable variations of transmission and presentation of text. Call control features are described for the SIP environment, while the transport mechanisms should be suitable for any IP based call control environment using RTP transport. Two alternative methods using a single RTP stream and source identification inline in the text stream are also described.

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 September 15, 2011.

Copyright Notice

Copyright (c) 2011 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.

Table of Contents

1.	Introduction	3
1.1.	Requirements Language	3
2.	Centralized conference model	3
2.1.	Coordination of text RTP streams	4
2.2.	Session control of multi-party sessions	4
3.	Identification of the source of text	5
4.	Presentation of multi-party text	5
4.1.	Associating identities with text streams	6
5.	Transmission of text from each user	6
6.	Presentation level source indicator	6
7.	Mixing for conference-unaware user agents	7
8.	IANA Considerations	7
9.	Security Considerations	8
10.	Congestion considerations	8
11.	Normative References	8
	Authors' Addresses	9

1. Introduction

Real-time text is a medium in real-time conversational sessions. Text entered by participants in a session is transmitted in a time-sampled fashion, so that no specific user action is needed to cause transmission. This gives a direct flow of text that is suitable in a real-time conversational setting. The real-time text medium can be combined with other media in multimedia sessions.

A number of multimedia sessions can be combined in a multi-party session. This memo specifies how the real-time text streams are handled in such multi-party sessions.

The description is mainly focused on the transport level, but also describes a few presentation level features.

Transport of real-time text is specified in [RFC 4103](#) [[RFC4103](#)] RTP Payload for text conversation. It makes use of [RFC 3550](#) [[RFC3550](#)] Real Time Protocol, for transport, and is usually used in the SIP Session Initiation Protocol [RFC 3261](#) [[RFC3261](#)] environment, even if it is also used in other call control environments. Call control aspects in this specification are explained with examples from SIP. The specifications about how to handle multi-party text transport, identification and presentation are valid also for other call control environments where RTP and RTCP are used.

A very brief overview of functions for both real-time and messaging text handling in multi-party sessions is described in [RFC 4597](#) [[RFC4579](#)] Conferencing Scenarios. This specification builds on that description and indicates what existing protocol mechanisms should be used to implement multi-party handling of text in real-time sessions.

1.1. 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 [RFC 2119](#) [[RFC2119](#)].

2. Centralized conference model

In the centralized conference model, one function co-ordinates the sessions with participants in the multi-party session. This function also controls media mixer functions for the media appearing in the session. The central function is common for control of all media, while the media mixers may work differently for each medium.

The central function is called the Focus UA and may be co-located in

an advanced terminal including multi-party control functions, or it may be located in a separate location. Many variants exist for setting up sessions including the multipoint control centre, It is not within scope of this description to describe these, but rather the media specific handling in the mixer required to handle multi-party calls.

The main principle for handling real-time text media in a centralized conference is that one RTP session for real-time text is established between the multipoint media control centre and each participant who is going to have real-time text exchange with the others.

2.1. Coordination of text RTP streams

The preferred way of coordinating text RTP streams is that within each RTP session, text from all participants are transmitted from the media mixer in the same RTP stream, thus all using the same destination address/port combination, and the same RTP SSRC as described in [Section 7.1](#) and 7.3 of RTP [RFC 3550](#) [[RFC3550](#)] about the Mixer function. The source of the primary text in each RTP packet is identified by the CSRC parameter, containing the SSRC of the initial source of text.

The mixer MUST NOT transmit redundant levels of text from one source together with primary text from another source. Thus, when there is text available for primary or redundant transmission from more than one source, the mixer MUST buffer text from other sources until all the redundant transmissions of a packet from one selected source has been transmitted. Without this restriction, there would be no way to decide with what source to associate text recovered from the redundant information in case of packet loss.

The identification of the source is made through the RTCP SDES CNAME and NAME packets as described in RTP[RFC3550].

This method enables the receiver to freely select display characteristics of the text conversation.

2.2. Session control of multi-party sessions

General session control aspects for multi-party sessions are described in [RFC 4575](#) [[RFC4575](#)] A Session Initiation Protocol (SIP) Event Package for Conference State, and [RFC 4579](#) [[RFC4579](#)] Session Initiation Protocol (SIP) Call Control - Conferencing for User Agents. The nomenclature of these specifications are used here.

The procedures for the mixer-based model shall only be applied if a capability exchange for mixer-based real-time text transmission has

been completed. Capability for the mixer-based model is indicated by both the focus and the user agent by the media tag `rtt-mixer=rtp-mix`

3. Identification of the source of text

The Focus UA co-ordinates the media flow. Real-time text media from different sources are combined in one text media session by the Focus UA. The main principle is that the Focus UA SHOULD act as an RTP Mixer as described in RTP [Section 7.1 \[RFC3550\]](#).

The RTP text stream from each participant who transmits text is allocated one unique CSRC. The CSRC is used by the receiver to identify text packets originating from one source. Each RTP packet MUST contain text from only one source.

The redundancy mechanism for increased robustness used by the [RFC 4103](#) transport makes use of the RTP sequence number for detection of loss. The RTP Mixer mechanism maintains a separate CSRC for each source RTP stream in the combined RTP session. Therefore the RTP Mixer mechanism can be used for conveying text from multiple sources to one destination, with maintained possibility to detect and recover loss and identify text from the different sources.

As soon as a new member is added to the RTP session, its characteristics shall be transmitted in RTCP SDES CNAME and NAME reports according to [section 6.5 in RFC 3550](#).

The RTCP SDES report, SHOULD contain identification of the source represented by the CSRC identifier. This identification MUST contain the CNAME field and MAY contain the NAME field and other defined fields of the SDES report.

A focus UA SHOULD primarily convey SDES information received from the sources of the session members. When such information is not available, the focus UA SHOULD compose CSRC, CNAME and NAME information from available information from the SIP session with the participant.

4. Presentation of multi-party text

All session participants MUST observe the CSRC field of incoming text RTP packets, and make note of what source they came from in order to be able to present text in a way that makes it easy to read text from each participant in a session, and get information about the source of the text.

4.1. Associating identities with text streams

A source identity SHOULD be composed from available information sources and displayed together with the text as indicated in ITU-T T.140 Appendix[T.140].

The source should primarily be the NAME field from incoming SDES packets. If this information is not available, and the session is a two-party session, then the T.140 source identity SHOULD be composed from the SIP session participant information. For multi-party sessions the source identity may be composed by local information if sufficient information is not available in the session.

Applications may abbreviate the presented source identity to a suitable form for the available display.

5. Transmission of text from each user

UAs participating in sessions with real-time text, SHOULD send SDES packets in RTCP giving values to appropriate identification fields.

The CNAME field SHALL be included in SDES packets.

The NAME field should be given a value that is suitable as an identifier of text from the user of the UA.

6. Presentation level source indicator

In certain application environments, it may be known to be unsuitable to use the CSRC identification on the RTP level as the base for identifying the source of text. In such cases, an inline coding of the source of text SHOULD be applied in the data stream itself, and an RTP mixer function normally without CSRC identification used for coordinating the sources of text into one RTP stream.

The support of this mixer type is indicated by the SIP header rtt-mix=t140, both by the focus and the user agent.

Information uniquely identifying each user in the multi-party session SHALL then be placed as the parameter value "cn" in the T.140 application protocol function with the function code "c". The identifier shall thus be formatted like this: SOS c cn field contents ST, where SOS and ST are coded as specified in ITU-T T.140 [T.140]. The cn parameter shall be kept short so that it can be repeated in the transmission without concerns for network load.

The information otherwise conveyed in the NAME field of an SDES packet SHOULD then be placed as the parameter value in the T.140 application protocol function with the function code "n".

A T.140 application protocol function with the function code "c" MUST be included in the text in the beginning of text when the source of the text changes. A T.140 application protocol function with the function code "c" MAY be repeated in the text from the same the source. A T.140 application protocol function with the function code "n" MAY be included in the text to further provide identification of the transmitting party. This information SHOULD also be provided in the SDES name field. A receiving UA SHOULD separate text from the different sources and identify and display them accordingly.

In this case, the mixer can use the redundancy transmission function of [RFC 4103](#) without restrictions.

7. Mixing for conference-unaware user agents

Multi-party real-time text contents can be transmitted to conference-unaware user agents if source labeling and formatting of the text is performed by a mixer. This method has the limitations that the format of source identification is purely controlled by the mixer, and that only one source at a time is allowed to present in real-time. Other sources need to be stored temporarily waiting for an appropriate moment to switch the source of transmitted text.

This method is used when no exchange of the rtt-mixer media tag has occurred in the session setup. Support for the method can however be expressed by the focus by the SIP media tag rtt-mixer=text-mixer.

8. IANA Considerations

This document Introduces the SIP media tag rtt-mixer, with a comma-separated parameter list containing the following possible values:

rtp-mixer

t140-mixer

text-mixer

rtp-mixer indicates capability for using the RTP-mixer based presentation of multi-party text. t140-mixer indicates capability for using the T.140 control code source indicators in a mixer. text-mixer indicates capability for using text-level control over formatting and

presentation of multi-party text presentation.

9. Security Considerations

The security considerations valid for [RFC 4103](#) and [RFC 3550](#) are valid also for the multi-party sessions with text.

10. Congestion considerations

The congestion considerations described in [RFC 4103](#) are valid also for multi-party use of the real-time text RTP transport. A risk for congestion may appear if a number of conference participants are active transmitting text simultaneously, because this multi-party transmission method does not allow multiple sources of text to contribute to the same packet.

In situations of risk for congestion, the Focus UA MAY combine packets from the same source to increase the transmission interval per source up to one second. Local conference policy in the Focus UA may be used to decide on which streams shall be selected for such transmission frequency reduction.

11. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", [RFC 3261](#), June 2002.
- [RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, [RFC 3550](#), July 2003.
- [RFC4103] Hellstrom, G. and P. Jones, "RTP Payload for Text Conversation", [RFC 4103](#), June 2005.
- [RFC4575] Rosenberg, J., Schulzrinne, H., and O. Levin, "A Session Initiation Protocol (SIP) Event Package for Conference State", [RFC 4575](#), August 2006.
- [RFC4579] Johnston, A. and O. Levin, "Session Initiation Protocol (SIP) Call Control - Conferencing for User Agents",

[BCP 119](#), [RFC 4579](#), August 2006.

[T.140] ITU-T, "Protocol for multimedia application text conversation", 1998,
<<http://www.itu.int/rec/T-REC-T.140/en>>.

Authors' Addresses

Gunnar Hellstrom
Omnitor
Box 92054
Stockholm SE-120 06
SE

Phone: +46 858900056
Fax: +46 858900051
Email: gunnar.hellstrom@omnitor.se
URI: www.omnitor.se

Arnoud van Wijk
Real-Time Text Taskforce (R3TF)
NL

Fax: +31 412614000
Email: arnoud@realtimetext.org
URI: www.realtimetext.org

