Network Working Group G. Hellstrom
Internet-Draft Gunnar Hellstrom Accessible Communication
Updates: RFC 4102, RFC 4103 (if approved) 9 April 2020
Intended status: Standards Track
Expires: 11 October 2020

RTP-mixer formatting of multi-party Real-time text
draft-hellstrom-avtcore-multi-party-rtt-source-03

Abstract

Real-time text mixers for multi-party sessions need to identify the source of each transmitted group of text so that the text can be presented by endpoints in suitable grouping with other text from the same source.

Regional regulatory requirements specify provision of real-time text in multi-party calls. RFC 4103 mixer implementations can use traditional RTP functions for source identification, but the mixer source switching performance is limited when using the default transmission with redundancy.

An enhancement for RFC 4103 real-time text mixing is provided in the present specification, suitable for a centralized conference model that enables source identification and efficient source switching. The intended use is for real-time text mixers and multi-party-aware participant endpoints. The mechanism builds on use of the CSRC list in the RTP packet.

A capability exchange is specified so that it can be verified that a participant can handle the multi-party coded real-time text stream. The capability is indicated by an sdp media attribute "rtt-mix".

A brief description about how a mixer can format text for the case when the endpoint is not multi-party aware is also provided.

Status of This Memo

Table of Contents

1. Introduction

RFC 4103[RFC4103] specifies use of RFC 3550 RTP [RFC3550] for transmission of real-time text (RTT) and the "text/t140" format. It also specifies a redundancy format "text/red" for increased robustness. RFC 4102 [RFC4102] registers the "text/red" format. Regional regulatory requirements specify provision of real-time text in multi-party calls.

The redundancy scheme enables efficient transmission of redundant text in packets together with new text. However the redundant header format has no source indicators for the redundant transmissions. An assumption has had to be made that the redundant parts in a packet are from the same source as the new text. The recommended transmission is one new and two redundant generations of text (T140blocks) in each packet and the recommended transmission interval is 300 ms.

A mixer, selecting between text input from different sources and transmitting it in a common stream needs to make sure that the receiver can assign the received text to the proper sources for presentation. Therefore, without any extra rule for source identification, the mixer needs to stop sending new text from that source and then make sure that all text so far has been sent with all intended redundancy levels (usually two) before switching source. That causes the very long time of one second to switch between transmission of text from one source to text from another source. Both the total throughput and the switching performance in the mixer is too low for most applications.

A more efficient source identification scheme requires that each redundant T140block has its source individually preserved. The present specification introduces a source indicator by specific rules for populating the CSRC-list in the RTP-packet.

A negotiation mechanism is also introduced for verification that the receiver is able to handle the multi-party coded stream.

A fall-back mixing procedure is specified for cases when the negotiation fails.

2. Nomenclature

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].

The terms SDES, CNAME, NAME, SSRC, CSRC, CSRC list, CC are explained in [RFC3550]

The term "T140block" is defined in RFC 4103 [RFC4103] to contain one or more T.140 code elements.

3. Intended application

The scheme for identification of source of redundant transmissions is intended for transmission from entities taking the mixer role in centralised mixing configurations for RTT. It is intended for reception by both endpoints and mixers.

4. Use of fields in the RTP packets

RFC 4103[RFC4103] specifies use of RFC 3550 RTP[RFC3550], and a redundancy format "text/red" for increased robustness. This specification updates RFC 4102[RFC4102] and RFC 4103[RFC4103] by introducing a rule for populating and using the CSRC-list in the RTP packet in order to enhance the performance in multi-party RTT sessions.

When transmitted from a mixer, the first member in the CSRC-list

SHALL contain the SSRC of the source of the primary T140block in the packet. The second and further members in the CSRC-list SHALL contain the SSRC of the source of the first, second, etc redundant generations of T140blocks included in the packet. (the recommended level of redundancy is to use one primary and two redundant generations of T140blocks.) In some cases, a primary or redundant T140block is empty, but is still represented by a member in the redundancy header. For such cases, the corresponding CSRC-list member MUST also be included.

The CC field SHALL show the number of members in the CSRC list.

Note: This specification departs from section 4 of RFC 2198 [RFC2198] which associates the whole of the CSRC-list with the primary data and assumes that the same list applies to reconstructed redundant data. In the present specification a T140block is associated with exactly one CSRC list member as described above. Also RFC 2198 [RFC2198] anticipates infrequent change to CSRCs; implementers should be aware that the order of the CSRC-list according to this specification will

vary during transitions between transmission from the mixer of text originated by different participants.

The picture below shows a typical RTP packet with multi-party RTT contents and coding according to the present specification.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |V=2|P|X| CC=3  |M|  "RED" PT   |   sequence number of primary  |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |               timestamp of primary encoding "P"               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |           synchronization source (SSRC) identifier            |
   +=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+=+
   |  CSRC list member 1 = SSRC of source of "P"                   |
   |  CSRC list member 2 = SSRC of source of "R1"                  |
   |  CSRC list member 3 = SSRC of source of "R2"                  |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |1|   T140 PT   |  timestamp offset of "R2" | "R2" block length |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |1|   T140 PT   |  timestamp offset of "R1" | "R1" block length |
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|   T140 PT   | "R2" T.140 encoded redundant data             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+---------------+
|   |  "R1" T.140 encoded redundant data       |               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+         +-+-+-+-+
|              "P" T.140 encoded primary data          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: text/red packet with sources indicated in the CSRC-list.

5. Actions at transmission by a mixer

A "text/red" transmitter is usually sending packets at a regular transmission interval as long as there is something (new or redundant T140blocks) to transmit. The default transmission interval for point-to-point operation is 300 ms.

For multi-party operation, the transmission interval from mixers SHOULD be decreased to 100 ms for periods when there is text available for transmission from more than two sources. It is also allowed for the mixer to send a packet as soon as text has been received from a source as long as the maximum number of characters per second indicated by the recipient is not exceeded, and also the number of packets sent per second to a recipient is kept under a specified number. This number SHALL be 10 if no other limit is applied for the application. The mixer has its own SSRC, and its own RTP sequence number series. At time of transmission, the mixer SHALL

populate the RTP packet with a T140block combined from all T140blocks queued for transmission originating from next source in turn for getting its text transmitted as long as this is not in conflict with the allowed number of characters per second. This T140block shall be placed in the primary area of the packet. The SSRC of its source shall be placed as the first member in the CSRC-list. The current time SHALL be inserted in the timestamp.

If no unsent T140blocks were available at this time, but T140blocks are available which have not yet been sent the full intended number of redundant transmissions, then the primary T140block is composed of an empty T140block, and populated (without taking up any length) in a packet for transmission. The SSRC of the mixer is included in the first place of the CSRC-list.

The primary T140block, in the latest transmission is used to populate the first redundant T140block, and its source is placed as the second member of the CSRC-list. The first redundant T140block from the latest transmission is placed as the second redundant T140block, and the corresponding CSRC placed in its place in the CSRC-list.

Usually this is the level of redundancy used. If a higher number of redundancy is used, then the procedure is maintained until all available redundant levels of T140blocks and their sources are placed in the packet. If a receiver has negotiated a lower number of text/red generations, then that level shall be the maximum used by the transmitter.

The timer offset values are inserted in the redundancy header, with the time offset from when the corresponding T140block was sent as original.

The number of members in the CSRC list shall be placed in the "CC" header field. Only mixers place values >0 in the "CC" field.

When there is no new T140block to transmit, and no redundant T140block that has not been retransmitted the intended number of times, the transmission process can stop until either new T140blocks arrive, or a keep-alive method calls for transmission of keep-alive packets.

6. Actions at reception

The enhanced "text/red" receiver included in an endpoint with presentation functions will receive RTP packets in the single stream from the mixer, and shall distribute the T140blocks for presentation in presentation areas for each source. Other tasks for receivers, such as gateways or chained mixers are also feasible, and requires

consideration if the stream shall just be forwarded, or a distribution based on different sources is needed.

If the "CC" field value of a received packet is >1, it indicates the used level of redundancy for the current packet and that the enhanced packet format is used. If the CC field value is 0 or 1, it indicates that the enhanced format is not used. If the CC value is 1, the CSRC indicates the source of primary data.

The RTP sequence numbers of the received packets SHALL be monitored for gaps and packets out of order.

As long as the sequence is correct, each packet SHALL be unpacked in order. The T140blocks SHALL be extracted from the primary area, and the corresponding SSRC SHALL be extracted from the first position in the CSRC list and used for assigning the new T140block to the correct presentation area (or correspondingly).

If a sequence number gap appears and is still there after some defined time for jitter resolution, T140data SHALL be recovered from redundant data. If the gap is wider than the number of generations of redundant T140blocks in the packet, then a t140block SHALL be created with a marker for text loss [T140ad1] and assigned to the SSRC of the transmitter as a general input from the mixer.

Then, the T140blocks in the received packet SHALL be retrieved beginning with the highest redundant generation, grouping them with the corresponding SSRC from the CSRC-list and assigning them to the presentation areas per source. Finally the primary T140block SHALL be retrieved from the packet and similarly its source retrieved from the first position in the CSRC-list, and then assigned to the corresponding presentation area for that source.

If the sequence number gap was equal to or less than the number of redundancy generations in the received packet, a missing text marker SHALL NOT be inserted, and instead the T140blocks and their SSRCs fully recovered from the redundancy information and the CSRC-list in the way indicated above.

Unicode character BOM is sometimes used as a filler or keep alive by transmission implementations. These SHALL be deleted on reception.

Note that empty T140blocks are sometimes included in the packets. They just do not provide any contents.

7. RTCP considerations

A mixer SHALL send RTCP reports with SDES, CNAME and NAME information about the sources in the multi-party call. This makes it possible for participants to compose a suitable label for text from each source.

8. Chained operation

By strictly applying the rules for CSRC-list population by all conforming devices, mixers MAY be arranged in chains.

9. Usage without redundancy

The CSRC list member SHALL be used as source indicator also for cases when the "text/t140" format is used from a mixer. That MAY be the case when robustness in transmission is provided by some other means than by redundancy and the "text/red" format. All aspects of this memo SHALL be applied except the redundant generations in transmission.

For the use case without redundancy using the "text/t140" format, the "CC" field in the RTP packet shall have the value 1, and the CSRC list SHALL contain one member when sent from a mixer.

The "text/red" format SHOULD be used unless some other protection against packet loss is utilized, for example a reliable network or transport.

10. Use with SIP centralized conferencing framework

The SIP conferencing framework, mainly specified in RFC 4353[RFC4353], RFC 4579[RFC4579] and RFC 4575[RFC4575] is suitable for coordinating sessions including multi-party RTT. The RTT stream between the mixer and a participant is one and the same during the conference. Participants get announced by notifications when participants are joining or leaving, and further user information may be provided. The SSRC of the text to expect from joined users can be included in a notification. This can be used both for security purposes and for translation to a label for presentation to other users.

Note: The CSRC-list in an RTP packet only includes participants who's text is included in one or more text blocks. It is not the same as the list of participants in a conference. With audio and video media, the CSRC-list would often contain all participants who are not muted whereas text participants that don't type are completely silent and thus are not represented in RTP packet CSRC-lists.

11. SDP Capability negotiation

There are RTT implementations which implement RFC 4103 [RFC4103] but not the present specification for real-time awareness. Sending mixed text according to the present specification to a device implementing only RFC 4103 [RFC4103] would lead to unreadable presented text. Therefore, in order to negotiate RTT mixing capability according to the present specification, all devices supporting the present specification for multi-party aware participants SHALL include an sdp media attribute "rtt-mix" in the sdp, indicating this capability in offers and answers. Multi-party streams using the coding of the present specification intended for multi-party aware endpoints MUST NOT be sent to devices which have not indicated the "rtt-mix" capability.

Implementations not understanding this parameter MUST ignore it according to common SDP rules.

The sdp media attribute defined here, is named "rtt-mix". It has no parameters. It is intended to be used in "text" media descriptions with "text/red" and "text/t140" formats. It indicates capability to use source indications in the CSRC list according to the present specification. It also indicates ability to receive 10 real-time text packets per second.

Syntax:

a=rtt-mix

The attribute is used in offer/answer procedures in a declarative way. Both parties express their capability to use sources in the CSRC list as specified in the present specification.

A party who has expressed the "rtt-mix" capability MUST populate the CSRC-list according to the present specification if it acts as an rtp-mixer and sends to a party who has expressed the "rtt-mix" capability.

A party who has expressed the "rtt-mix" capability MUST interpret the contents of the CSRC-list according to the present specification in received rtp packets from parties who have expressed "rtt-mix" capability .

A party MUST NOT transmit packets with the format for multi-party aware participants according to the present specification to a party who has not expressed "rtt-mix" capability.

A party performing as a mixer, which has expressed the "rtt-mix" capability, but not received "rtt-mix" capability indication in a session with a participant SHOULD, if nothing else is specified for the application, format transmitted text to that participant to be suitable to present on a multi-party unaware endpoint as further specified in section Section 14.2.

12. Examples

This example shows a symbolic flow of packets from a mixer with loss and recovery. A, B and C are sources of RTT. M is the mixer. P indicates primary data. R1 is first redundant generation data and R2 is second redundant generation data. A1, B1, A2 etc are text chunks (T140blocks) received from the respective sources. X indicates dropped packet between the mixer and a receiver.

```
|----------------|
|Seq no 1        |
|CSRC list A,M,B |
|R2 B99          |
|R1: Empty       |
|P: A1           |
|----------------|
```

Assuming that earlier packets were received in sequence, text A1 is received from packet 1 and assigned to reception area A.

```
|----------------|
|Seq no 2        |
|CC=3            |
|CSRC list C,A,M |
|R2 Empty        |
|R1: A1          |
|P: C1           |
|----------------|
```
Text C1 is received from packet 2 and assigned to reception area C.

```
X----------------|
X Seq no 3       |
```

```
   X CC=3            |
   X CSRC list A,C,A|
   X R2: A1          |
   X R1: C1          |
   X P: A2           |
   X----------------|
   Packet 3 is assumed to be dropped in network problems
```



```
   X----------------|
   X Seq no 4        |
   X CC=3            |
   X CSRC list B,A,C|
   X R2: C1          |
   X R1: A2          |
   X P: B1           |
   X----------------|
   Packet 4 is assumed to be dropped in network problems

   X----------------|
   X Seq no 5        |
   X CC=3            |
   X CSRC list A,B,A|
   X R2: A2          |
   X R1: B1          |
   X P: A3           |
   X----------------|
   Packet 5 is assumed to be dropped in network problems

   |----------------|
   |Seq no 6        |
   |CC=3            |
   |CSRC list C,A,B |
   |R2: B1          |
   |R1: A3          |
   |P: C2           |
   |----------------|
   Packet 6 is received. The latest received sequence number was 2.
   Recovery is therefore tried for 3,4,5. But there is no coverage
   for seq no 3. A missing text mark (U'FFFD) is created and
   appended to the mixer reception area.
```

For seqno 4, text B1 is recovered and appended to reception area B. For seqno 5, text A3 is recovered and appended to reception area A. Primary text C2 is received and appended to reception area C.

With only one or two packets lost, there would not be any need to create a missing text marker, and all text would be recovered.

It will be a design decision how to present the missing text markers assigned to the mixer as a source.

13. Performance considerations

This specification allows new text from one source per packet. Packets SHOULD be transmitted with timed intervals. The default transmission interval is 300 ms for RFC 4103[RFC4103], which is suitable for transmission from single sources. However when more sources contribute to the flow, a shorter transmission interval MAY be applied. The transmission interval is therefore RECOMMENDED to be 100 ms for mixers when there is text from more than two sources available for transmission to the same recipient.It is also allowed for the mixer to send a packet as soon as text has been received from a source as long as the maximum number of characters per second indicated by the recipient is not exceeded, and also the number of packets sent per second to a recipient is kept under a number. In order to achieve good performance, a receiver for multi-party calls SHOULD declare a sufficient CPS value in SDP for the number of allowable characters per second. These characteristics provide for smooth flow of text with acceptable latency from at least 5 sources simultaneously.

The default maximum rate of reception of real-time text is in RFC 4103 [RFC4103] specified to be 30 characters per second. The value MAY be modified in the CPS parameter of the FMTP attribute in the media section for RFC 4103. A mixer combining real-time text from a number of sources may have a higher combined flow of text coming from the sources. Endpoints SHOULD therefore specify a suitable higher value for the CPS parameter, corresponding to its real reception capability. A value for CPS of 150 is RECOMMENDED because it would be suitable for reception in most multi-party sessions, and a realistic value for the capacity in most implementation environments. See RFC 4103 [RFC4103] for the format and use of the CPS parameter.

14. Presentation level considerations

ITU-T T.140 [T140] provides the presentation level requirements for the RFC 4103 [RFC4103] transport. T.140 [T140] has functions for erasure and other formatting functions and has the following general statement for the presentation:

"The display of text from the members of the conversation should be arranged so that the text from each participant is clearly readable, and its source and the relative timing of entered text is visualized in the display. Mechanisms for looking back in the contents from the current session should be provided. The text should be displayed as soon as it is received."

Strict application of T.140 [T140] is of essence for the interoperability of real-time text implementations and to fulfill the

intention that the session participants have the same information of the text contents of the conversation without necessarily having the exact same layout of the conversation. This also includes the ability to ignore optional presentation control codes not supported by a receiving application.

T.140 [T140] specifies a set of presentation control codes to include in the stream. Some of them are optional. Implementations MUST be able to ignore optional control codes that they do not support.

There is no strict "message" concept in real-time text. Line Separator SHALL be used as a separator allowing a part of received text to be grouped in presentation. The characters "CRLF" may be used by other implementations as replacement for Line Separator. The "CRLF" combination SHALL be erased by just one erasing action, just as the Line Separator. Presentation functions are allowed to group text for presentation in smaller groups than the line separators imply and present such groups with source indication together with text groups from other sources (see the following presentation examples). Erasure has no specific limit by any delimiter in the text stream.

14.1. Presentation by multi-party aware endpoints

A multi-party aware receiving party, presenting real-time text MUST separate text from different sources and present them in separate presentation areas. The receiving party MAY separate presentation of parts of text from a source in readable groups based on other criteria than line separator and merge these groups in the presentation area when it benefits the user to most easily find and read text from the different participants. The criteria MAY e.g. be a received comma, full stop, or other phrase delimiters, or a long pause.

When text is received from multiple original sources, the presentation SHOULD provide a view where text is added in multiple places simultaneously.

If the presentation presents text from different sources in one common area, the presenting endpoint SHOULD insert text from the local user ended at suitable points merged with received text to indicate the relative timing for when the text groups were completed. In this presentation mode, the receiving endpoint SHALL present the source of the different groups of text.

A view of a three-party RTT call in chat style is shown in this example .


```
 ________________________________________________
|                                                | |
|[Alice] Hi, Alice here.                         | |
|                                                | |
|[Bob] Bob as well.                              | |
|                                                | |
|[Eve] Hi, this is Eve, calling from Paris.      | |
|      I thought you should be here.             | |
|                                                | |
|[Alice] I am coming on Thursday, my             | |
|      performance is not until Friday morning.| |
```

```
         |                                              | |
         |[Bob] And I on Wednesday evening.             | |
         |                                              | |
         |[Alice] Can we meet on Thursday evening?      | |
         |                                              | |
         |[Eve] Yes, definitely. How about 7pm.         | |
         |     at the entrance of the restaurant        | |
         |     Le Lion Blanc?                           | |
         |[Eve] we can have dinner and then take a walk | |
         |                                              | |
         | <Eve-typing> But I need to be back to        | |
         |    the hotel by 11 because I need            | |
         |                                              |-|
         | <Bob-typing> I wou                           |-|
         |______________________________________________|v|
         | of course, I underst                           |
         |________________________________________________|
```

Figure 1: Example of a three-party call with chat style.

Figure 2: Example of a three-party RTT call presented in chat style.

Other presentation styles than the chat style may be arranged.

This figure shows how a coordinated column view MAY be presented.


```
 _____________________________________________________________________
|       Bob          |         Eve           |         Alice          |
|--------------------|-----------------------|------------------------|
|                    |                       |I will arrive by TGV.   |
```

```
|My flight is to Orly|                      |Convenient to the main |
|                    |Hi all, can we plan   |station.               |
|                    |for the seminar?      |                       |
|Eve, will you do    |                      |                       |
|your presentation on|                      |                       |
|Friday?             |Yes, Friday at 10.    |                       |
|Fine, wo            |                      |We need to meet befo   |
|_____________________________________________________________________|
```

Figure 3: An example of a coordinated column-view of a three-party session with entries ordered vertically in approximate time-order.

14.2. Multi-party mixing for multi-party unaware endpoints

When the mixer has indicated multi-party capability in an sdp negotiation, but the multi-party capability negotiation fails with an endpoint, then the mixer SHOULD compose a best-effort presentation of multi-party real-time text in one stream intended to be presented by an endpoint with no multi-party awareness.

This presentation format has functional limitations and SHOULD be used only to enable participation in multi-party calls by legacy deployed endpoints.

The principles and procedures below do not specify any new protocol elements or behaviors. They are instead composed from the information in ITU-T T.140 [T140] and an ambition to provide a best effort presentation on an endpoint which has functions only for two-party calls.

The mixer mixing for multi-party unaware endpoints SHALL compose a simulated limited multi-party RTT view suitable for presentation in one presentation area. The mixer SHALL group text in suitable groups and prepare for presentation of them by inserting a new line betwwen them if the transmitted text did not already end with a new line. A presentable label SHOULD be composed and sent for the source initially in the session and after each source switch. With this procedure the time for source switching is depending on the actions of the users. In order to expedite source switch, a user can for example end its turn with a new line.

14.2.1. Actions by the mixer at reception from the call participants

When text is received by the mixer from the different participants, the mixer SHALL recover text from redundancy if any packets are lost. The mark for lost text [T140ad1] SHOULD be inserted in the stream if unrecoverable loss appears. Any Unicode BOM characters, possibly used for keep-alive shall be deleted. The time of arrival of text SHALL be stored together with the received text from each source in a queue for transmission to the recipients.

14.2.2. Actions by the mixer for transmission to the recipients

The following procedure SHOULD be applied for each recipient of multi-part text from the mixer.

The text for transmission SHOULD be formatted by the mixer for each receiving user for presentation in one single presentation area. Text received from a participant SHOULD NOT be included in transmission to that participant. When there is text available for transmission from the mixer to a receiving party from more than one participant, the mixer SHOULD switch between transmission of text from the different sources at suitable points in the transmitted stream.

When switching source, the mixer SHOULD insert a line separator if the already transmitted text did not end with a new line (line separator or CRLF). A label SHOULD be composed from information in the CNAME and NAME fields in RTCP reports from the participant to have its text transmitted, or from other session information for that user. The label SHOULD be delimited by suitable characters (e.g. '[ ]') and transmitted. The CSRC SHOULD indicate the selected source. Then text from that selected participant SHOULD be transmitted until a new suitable point for switching source is reached.

Seeking a suitable point for switching source SHOULD be done when there is older text waiting for transmission from any party than the age of the last transmitted text. Suitable points for switching are:

* A completed phrase ended by comma

* A completed sentence

* A new line (line separator or CRLF)

* A long pause (e.g. > 10 seconds) in received text from the currently transmitted source

* If text from one participant has been transmitted with text from other sources waiting for transmission for a long time (e.g. > 1 minute) and none of the suitable points for switching has occurred, a source switch MAY be forced by the mixer at next word delimiter, and also if even a word delimiter does not occur within a time (e.g. 15 seconds) after the scan for word delimiter started.

When switching source, the source which has the oldest text in queue SHOULD be selected to be transmitted. A character display count SHOULD be maintained for the currently transmitted source, starting at zero after the label is transmitted for the currently transmitted source.

There SHOULD be a storage for the latest control code for Select Graphic Rendition (SGR) from each source. If there is an SGR code stored for the current source before the source switch is done, a reset of SGR shall be sent by the sequence SGR 0 [009B 0000 006D] after the new line and before the new label during a source switch. See SGR below for an explanation. This transmission does not influence the display count. If there is an SGR code stored for the new source after the source switch, that SGR code SHOULD be transmitted to the recipient before the label. This transmission does not influence the display count.

14.2.3. Actions on transmission of text

Text from a source sent to the recipient SHOULD increase the display count by one per transmitted character.

14.2.4. Actions on transmission of control codes

The following control codes specified by T.140 require specific actions. They SHOULD cause specific considerations in the mixer. Note that the codes presented here are expressed in UCS-16, while transmission is made in UTF-8 transform of these codes.

BEL 0007 Bell Alert in session, provides for alerting during an active session. The display count SHOULD not be altered.

NEW LINE 2028 Line separator. Check and perform a source switch if appropriate. Increase display count by 1.

CR LF 000D 000A A supported, but not preferred way of requesting a new line. Check and perform a source switch if appropriate. Increase display count by 1.

INT ESC 0061 Interrupt (used to initiate mode negotiation procedure). The display count SHOULD not be altered.

SGR 009B Ps 006D Select graphic rendition. Ps is rendition parameters specified in ISO 6429. The display count SHOULD not be altered. The SGR code SHOULD be stored for the current source.

SOS 0098 Start of string, used as a general protocol element introducer, followed by a maximum 256 bytes string and the ST. The display count SHOULD not be altered.

ST 009C String terminator, end of SOS string. The display count SHOULD not be altered.

ESC 001B Escape - used in control strings. The display count SHOULD not be altered for the complete escape code.

Byte order mark FEFF Zero width, no break space, used for synchronization and keep-alive. SHOULD be deleted from incoming streams. Shall be sent first after session establishment to the recipient. The display count shall not be altered.

Missing text mark FFFD Replacement character, marks place in stream of possible text loss. SHOULD be inserted by the reception procedure in case of unrecoverable loss of packets. The display count SHOULD be increased by one when sent as for any other character.

SGR If a control code for selecting graphic rendition (SGR), other than reset of the graphic rendition (SGR 0) is sent to a recipient, that control code shall also be stored for the source in the storage for SGR. If a reset graphic rendition (SGR 0) originated from a source is sent, then the SGR storage for that source shall be cleared. The display count shall not be increased.

BS 0008 Back Space, intended to erase the last entered character by a source. Erasure by backspace cannot always be performed as the erasing party intended. If an erasing action erases all text up to the end of the leading label after a source switch, then the mixer must not transmit more backspaces. Instead it is RECOMMENDED that a letter "X" is inserted in the text stream for each backspace as an indication of the intent to erase more. A new line is usually coded by a Line Separator, but the character combination "CRLF" MAY be used instead. Erasure of a new line is in both cases done by just one erasing action (Backspace). If the display count has a positive value it is decreased by one when the BS is sent. If the display count is at zero, it is not altered.

14.2.5. Packet transmission

A mixer transmitting to a multi-party unaware terminal SHOULD send primary data only from one source per packet. The SSRC SHOULD be the SSRC of the mixer. The CSRC list SHOULD contain one member and be the SSRC of the source of the primary data.

14.2.6. Functional limitations

When a multi-party unaware endpoint presents a conversation in one display area in a chat style, it inserts source indications for remote text and local user text as they are merged in completed text groups. When an endpoint using this layout receives and presents text mixed for multi-party unaware endpoints, there will be two levels of source indicators for the received text; one generated by the mixer and inserted in a label after each source switch, and another generated by the receiving endpoint and inserted after each switch between local and remote source in the presentation area. This will waste display space and look inconsistent to the reader.

This fact, combined with the slowness in source switching and the limited support of erasure makes it strongly RECOMMENDED to implement multi-party awareness in RTT endpoints. The use of the mixing method for multi-party-unaware endpoints should be left for use with endpoints which are impossible to upgrade to become multi-party aware.

14.2.7. Example views of presentation on multi-party unaware endpoints

The following pictures are examples of the view on a participant's display for the multi-party-unaware case.

```
 _________________________________________________
|       Conference        |          Alice        |
|_________________________|_______________________|
|                         |I will arrive by TGV.  |
|[Bob]:My flight is to    |Convenient to the main |
|Orly.                    |station.               |
|[Eve]:Hi all, can we     |                       |
|plan for the seminar.    |                       |
|                         |                       |
|[Bob]:Eve, will you do   |                       |
|your presentation on     |                       |
|Friday?                  |                       |
|[Eve]:Yes, Friday at 10.|                       |
|[Bob]: Fine, wo          |We need to meet befo   |
|_________________________|_______________________|
```


Figure 4: Alice who has a conference-unaware client is receiving the multi-party real-time text in a single-stream. This figure shows how a coordinated column view MAY be presented on Alice's device.

```
 _________________________________________________
|                                                 |^|
|[Alice] Hi, Alice here.                          | |
|                                                 | |
|[mix][Bob] Bob as well.                          | |
|                                                 | |
|[Eve] Hi, this is Eve, calling from Paris        | |
|      I thought you should be here.              | |
|                                                 | |
|[Alice] I am coming on Thursday, my              | |
|      performance is not until Friday morning.| |
|                                                 | |
|[mix][Bob] And I on Wednesday evening.           | |
|                                                 | |
|[Eve] we can have dinner and then walk           | |
```

```
   |                                               | |
   |[Eve] But I need to be back to                 | |
   |    the hotel by 11 because I need             |-|
   |                                               |-|
   |_______________________________________________|v|
   | of course, I underst                              |
   |___________________________________________________|
```

Figure 5: An example of a view of the multi-party unaware presentation in chat style. Alice is the local user.

15. Gateway Considerations

Multi-party RTT sessions may involve gateways of different kinds. Gateways involved in setting up sessions SHALL correctly reflect the multi-party capability or unawareness of the combination of the gateway and the remote endpoint beyond the gateway.

One case that may occur is a gateway to PSTN for communication with textphones (e.g. TTYs). Textphones are limited devices with no multi-party awareness, and it SHOULD therefore be suitable for the gateway to not indicate multi-party awareness for that case. Another solution is that the gateway indicates multi-party capability towards the mixer, and includes the multi-party mixer function for multi-party unaware endpoints itself. This solution makes it possible to make adaptations for the functional limitations of the textphone (TTY).

16. Congestion considerations

The congestion considerations and recommended actions from RFC 4103 [RFC4103] are valid also in multi-party situations.

17. Acknowledgements

18. IANA Considerations

[RFC EDITOR NOTE: Please replace all instances of RFCXXXX with the RFC number of this document.]

IANA is asked to register the new sdp attribute "rtt-mix".

Contact name:	IESG
Contact email:	iesg@ietf.org
Attribute name:	rtt-mix
Attribute syntax	a=rtt-mix
Attribute semantics	See RFCXXXX section 11.
Attribute value	-
Usage level:	media
Purpose:	Indicate support for the rtp-mixer format for real-time text transmission
O/A procedure	Declarative
Mux Category	normal
Reference:	RFCXXXX

19. Security Considerations

The RTP-mixer model requires the mixer to be allowed to decrypt, pack and encrypt secured text from the conference participants. Therefore the mixer needs to be trusted. This is similar to the situation for central mixers of audio and video.

The requirement to transfer information about the user in RTCP reports in SDES, CNAME and NAME fields for creation of labels may have privacy concerns as already stated in RFC 3550 [RFC3550], and may be restricted of privacy reasons. The receiving user will then get a more symbolic label for the source.

20. Change history

20.1. Changes from version -02 to -03

Changed company and e-mail of the author.

Changed title to "RTP-mixer formatting of multi-party Real-time text" to better match contents.

Check and modification where needed of use of RFC 2119 words SHALL etc.

More about the CC value in sections on transmitters and receivers so that 1-to-1 sessions do not use the mixer format.

Enhanced section on presentation for multi-party-unaware endpoints

A paragraph recommending CPS=150 inserted in the performance section.

20.2. Changes from version -01 to -02

In Abstract and 1. Introduction: Introduced wording about regulatory requirements.

In section 5: The transmission interval is decreased to 100 ms when there is text from more than one source to transmit.

In section 11 about sdp negotiation, a SHOULD-requirement is introduced that the mixer should make a mix for multi-party unaware endpoints if the negotiation is not successful. And a reference to a later chapter about it.

The presentation considerations chapter 14 is extended with more information about presentation on multi-party aware endpoints, and a new section on the multi-party unaware mixing with low functionality but SHOULD a be implemented in mixers. Presentation examples are added.

A short chapter 15 on gateway considerations is introduced.

Clarification about the text/t140 format included in chapter 10.

This sentence added to the chapter 10 about use without redundancy. "The text/red format SHOULD be used unless some other protection against packet loss is utilized, for example a reliable network or transport."

Note about deviation from RFC 2198 added in chapter 4.

In chapter 9. "Use with SIP centralized conferencing framework" the following note is inserted: Note: The CSRC-list in an RTP packet only includes participants who's text is included in one or more text blocks. It is not the same as the list of participants in a conference. With audio and video media, the CSRC-list would often contain all participants who are not muted whereas text participants that don't type are completely silent and so don't show up in RTP packet CSRC-lists.

20.3. Changes from version -00 to -01

Editorial cleanup.

Changed capability indication from fmtp-parameter to sdp attribute "rtt-mix".

Swapped order of redundancy elements in the example to match reality.

Increased the SDP negotiation section

21. References

21.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC2198] Perkins, C., Kouvelas, I., Hodson, O., Hardman, V., Handley, M., Bolot, J.C., Vega-Garcia, A., and S. Fosse-Parisis, "RTP Payload for Redundant Audio Data", RFC 2198, DOI 10.17487/RFC2198, September 1997, <https://www.rfc-editor.org/info/rfc2198>.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, DOI 10.17487/RFC3550, July 2003, <https://www.rfc-editor.org/info/rfc3550>.

[RFC4102] Jones, P., "Registration of the text/red MIME Sub-Type", RFC 4102, DOI 10.17487/RFC4102, June 2005, <https://www.rfc-editor.org/info/rfc4102>.

[RFC4103] Hellstrom, G. and P. Jones, "RTP Payload for Text Conversation", RFC 4103, DOI 10.17487/RFC4103, June 2005, <https://www.rfc-editor.org/info/rfc4103>.

[T140] ITU-T, "Recommendation ITU-T T.140 (02/1998), Protocol for multimedia application text conversation", February 1998, <https://www.itu.int/rec/T-REC-T.140-199802-I/en>.

[T140ad1] ITU-T, "Recommendation ITU-T.140 Addendum 1 - (02/2000), Protocol for multimedia application text conversation", February 2000, <https://www.itu.int/rec/T-REC-T.140-200002-I!Add1/en>.

21.2. Informative References

[RFC4353] Rosenberg, J., "A Framework for Conferencing with the Session Initiation Protocol (SIP)", RFC 4353, DOI 10.17487/RFC4353, February 2006, <https://www.rfc-editor.org/info/rfc4353>.

[RFC4575] Rosenberg, J., Schulzrinne, H., and O. Levin, Ed., "A Session Initiation Protocol (SIP) Event Package for Conference State", RFC 4575, DOI 10.17487/RFC4575, August 2006, <https://www.rfc-editor.org/info/rfc4575>.

[RFC4579] Johnston, A. and O. Levin, "Session Initiation Protocol (SIP) Call Control - Conferencing for User Agents", BCP 119, RFC 4579, DOI 10.17487/RFC4579, August 2006, <https://www.rfc-editor.org/info/rfc4579>.

Author's Address

Gunnar Hellstrom
Gunnar Hellstrom Accessible Communication
Esplanaden 30
SE-13670 Vendelso
Sweden

Email: gunnar.hellstrom@ghaccess.se