Network Working Group	X. Duan
Internet-Draft	S. Wang
Intended status: Standards Track	China Mobile Communications
Expires: July 25, 2010	Corporation
	M. Westerlund
	K. Hellwig
	I. Johansson
	Ericsson AB
	January 21, 2010

TOC

RTP Payload format for GSM-HR
draft-ietf-avt-rtp-gsm-hr-03

Abstract

This document specifies the payload format for packetization of the GSM Half-Rate speech codec data into the Real-time Transport Protocol (RTP). The payload format supports transmission of multiple frames per payload and packet loss robustness methods using redundancy.

Status of this Memo

Copyright Notice

Table of Contents

---

1. Introduction

TOC

This document specifies the payload format for packetization of GSM Half Rate (GSM-HR) codec (3GPP, “Specification : 3GPP TS 46.002 http://

www.3gpp.org/ftp/Specs/archive/46_series/46.002/46002-700.zip,” June 2007.) [TS46.002] encoded speech signals into the Real-time Transport Protocol (RTP) [RFC3550] (Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications,” July 2003.). The payload format supports transmission of multiple frames per payload and packet loss robustness methods using redundancy.

This document starts with conventions, a brief description of the codec, and the payload formats capabilities. The payload format is specified in Section 5 (Payload format). Examples can be found in Section 6 (Examples). The media type and its mappings to SDP, usage in SDP offer/answer is then specified. The document ends with considerations around congestion control and security.

This document registers a media type (audio/gsm-hr-08) for the Real-time Transport protocol (RTP) payload format for the GSM-HR codec.

Note: This format is not compatible with the one that was drafted back in 1999 to 2000 in the Internet drafts: draft-ietf-avt-profile-new-05 to draft-ietf-avt-profile-new-09. A later version of the AVP profile draft was published as RFC 3551 without any specification of the GSM-HR payload format. To avoid a possible conflict with this older format, the media type of the payload format specified in this document has a media type name that is different from (audio/gsm-hr).

---

2. Conventions

TOC

This document uses the normal IETF bit-order representation. Bit fields in figures are read left to right and then down. The left most bit in each field is the most significant. The numbering starts from 0 and ascends, where bit 0 will be the most significant.

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 (Bradner, S., “Key words for use in RFCs to Indicate Requirement Levels,” March 1997.) [RFC2119].

---

3. GSM Half Rate

TOC

The Global System for Mobile Communication (GSM) network provides with mobile communication services for nearly 3 billion users (status 2008). The GSM Half Rate Codec (GSM-HR) is one of the speech codecs that are used in GSM networks. GSM-HR denotes the Half-Rate speech codec as specified in [TS46.002] (3GPP, “Specification : 3GPP TS 46.002 http://www.3gpp.org/ftp/Specs/archive/46_series/46.002/46002-700.zip,” June 2007.).

Note: for historical reasons these 46-series specifications are internally referenced as 06-series. A simple mapping applies, for example 46.020 is referenced as 06.20 and so on.
The GSM-HR codec has a frame length of 20 ms, with narrowband speech sampled at 8 kHz, i.e. 160 samples per frame. Each speech frame is compressed into 112 bits of speech parameters, which is equivalent to a bit rate of 5.6 kbit/s. Speech pauses are detected by a standardized Voice Activity Detection (VAD). During speech pauses the transmission of speech frames is inhibited. Silence Descriptor (SID) frames are transmitted at the end of a talk spurt and about every 480ms during speech pauses to allow for a decent Comfort Noise (CN) quality at receiver side.
The SID frame generation in the GSM radio network is determined by the GSM mobile station and the GSM radio subsystem. SID frames come during speech pauses in uplink from the mobile station about every 480ms. In downlink to the mobile station, when they are generated by the encoder of the GSM radio subsystem, SID frames are sent every 20ms to the GSM base station, which then picks only one every 480ms for downlink radio transmission. For other applications, like transport over IP, it is more appropriate to send the SID frames less often than every 20ms, but 480 ms may be too sparse. We recommend as a compromise that a GSM-HR encoder outside of the GSM radio network (i.e. not in the GSM mobile station and not in the GSM radio subsystem, but for example in the media gateway of the core network) should generate and send SID frames every 160ms.

---

4. Payload format Capabilities

TOC

This RTP payload format carries one or more GSM-HR encoded frames, either full voice or silence descriptor (SID), representing a mono speech signal. To maintain synchronization or express not sent or lost frames it has the capability to indicate No_Data frames.

4.1. Use of Forward Error Correction (FEC)

TOC

Generic forward error correction within RTP is defined, for example, in RFC 5109 [RFC5109] (Li, A., “RTP Payload Format for Generic Forward Error Correction,” December 2007.). Audio redundancy coding is defined in RFC 2198 [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,” September 1997.). Either scheme can be used to add redundant information to the RTP packet stream and make it more resilient to packet losses, at the expense of a higher bit rate. Please see either RFCs for a discussion of the implications of the higher bit rate to network congestion.

In addition to these media-unaware mechanisms, this memo specifies an optional to use GSM-HR specific form of audio redundancy coding, which may be beneficial in terms of packetization overhead. Conceptually, previously transmitted transport frames are aggregated together with new ones. A sliding window can be used to group the frames to be sent in each payload. Figure 1 (An example of redundant transmission) below shows an example.

```
 --+--------+--------+--------+--------+--------+--------+--------+--
   | f(n-2) | f(n-1) |  f(n)  | f(n+1) | f(n+2) | f(n+3) | f(n+4) |
 --+--------+--------+--------+--------+--------+--------+--------+--

    <---- p(n-1) ---->
             <----- p(n) ----->
                      <---- p(n+1) ---->
                               <---- p(n+2) ---->
                                        <---- p(n+3) ---->
                                                 <---- p(n+4) ---->
```

Figure 1: An example of redundant transmission

Here, each frame is retransmitted once in the following RTP payload packet. f(n-2)...f(n+4) denote a sequence of audio frames, and p(n-1)...p(n+4) a sequence of payload packets.

The mechanism described does not really require signaling at the session setup. However, signalling has been defined to allow for the sender to voluntarily bounding the buffering and delay requirements. If nothing is signalled the use of this mechanism is allowed and unbounded. For a certain timestamp, the receiver may receive multiple copies of a frame containing encoded audio data. The cost of this scheme is bandwidth and the receiver delay necessary to allow the redundant copy to arrive.

This redundancy scheme provides a functionality similar to the one described in RFC 2198, but it works only if both original frames and redundant representations are GSM-HR frames. When the use of other media coding schemes is desirable, one has to resort to RFC 2198.

The sender is responsible for selecting an appropriate amount of redundancy based on feedback about the channel conditions, e.g., in the RTP Control Protocol (RTCP) [RFC3550] (Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications,” July 2003.) receiver reports. The sender is also responsible for avoiding congestion, which may be exacerbated by redundancy (see Section 9 (Congestion Control) for more details).

5. Payload format

TOC

The format of the RTP header is specified in [RFC3550] (Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications,” July 2003.). This payload format uses the fields of the header in a manner consistent with that specification.

The duration of one speech frame is 20 ms. The sampling frequency is 8kHz, corresponding to 160 speech samples per frame. An RTP packet may contain multiple frames of encoded speech or SID parameters. Each packet covers a period of one or more contiguous 20 ms frame intervals. During silence periods no speech packets are sent, however SID packets are transmitted every now and then.

To allow for error resiliency through redundant transmission, the periods covered by multiple packets MAY overlap in time. A receiver MUST be prepared to receive any speech frame multiple times. A given frame MUST NOT be encoded as speech frame in one packet and as SID frame or as No_Data frame in another packet. Furthermore, a given frame MUST NOT be encoded with different voicing modes in different packets.

The rules regarding maximum payload size given in Section 3.2 of [RFC5405] (Eggert, L. and G. Fairhurst, “Unicast UDP Usage Guidelines for Application Designers,” November 2008.) SHOULD be followed.

5.1. RTP Header Usage

TOC

The RTP timestamp corresponds to the sampling instant of the first sample encoded for the first frame in the packet. The timestamp clock frequency SHALL be 8000 Hz. The timestamp is also used to recover the correct decoding order of the frames.

The RTP header marker bit (M) SHALL be set to 1 whenever the first frame carried in the packet is the first frame in a talkspurt (see definition of the talkspurt in section 4.1 of [RFC3551] (Schulzrinne, H. and S. Casner, “RTP Profile for Audio and Video Conferences with Minimal Control,” July 2003.)). For all other packets the marker bit SHALL be set to zero (M=0).

The assignment of an RTP payload type for the format defined in this memo is outside the scope of this document. The RTP profiles in use currently mandates binding the payload type dynamically for this payload format.

The remaining RTP header fields are used as specified in RFC 3550 [RFC3550] (Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications,” July 2003.).

5.2. Payload Structure

TOC

The complete payload consists of a payload table of contents (ToC) section, followed by speech data representing one or more speech frames, SID frames or No_Data frames. The following diagram shows the general payload format layout:

```
+-------------+-----------------------
| ToC section | speech data section ...
+-------------+----------------------
```

Figure 2: General payload format layout

Each ToC element is one octet and corresponds to one speech frame, the number of ToC elements is thus equal to the number of speech frames (including SID frames and No_Data frames). Each ToC entry represents a consecutive speech or SID or No_Data frame. The timestamp value for ToC element (and corresponding speech frame data) N within the payload is (RTP timestamp field + (N-1)*160) mod 2^32 . The format of the ToC element is as follows.

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|F| FT  |R R R R|
+-+-+-+-+-+-+-+-+
```

Figure 3: The TOC element

F: Follow flag, 1 denotes that more ToC elements follow, 0 denotes the last ToC element.

R: Reserved bits, MUST be set to zero and MUST be ignored by receiver.

FT: Frame type

> 000 = Good Speech frame
> 001 = Reserved
> 010 = Good SID frame
> 011 = Reserved
> 100 = Reserved
> 101 = Reserved
> 110 = Reserved
> 111 = No_Data frame

The length of the payload data depends on the frame type:

> **Good Speech frame:** The 112 speech data bits are put in 14 octets.
>
> **Good SID frame:** The 33 SID data bits are put in 14 octets, as in case of Speech frames, with the unused 79 bits set all to “1”.
>
> **No data frame:** Length of payload data is zero octets.

Frames marked in the GSM radio subsystem as “Bad Speech frame”, “Bad SID frame” or “No_Data frame” are not sent in RTP packets in order to save bandwidth. They are marked as “No_Data frame”, if they occur within an RTP packet that carries more than one speech frame, SID frame or No_Data frame.

5.2.1. Encoding of Speech Frames

TOC

The 112 bits of GSM-HR-coded speech (b1…b112) are defined in TS 46.020, Annex B [TS46.020] (3GPP, “Specification : 3GPP TS 46.020 http://www.3gpp.org/ftp/Specs/archive/46_series/46.002/46020-700.zip,” June 2007.), in the order of occurrence. The first bit (b1) of the first parameter is placed in bit 0 (the MSB) of the first octet (octet 1) of the payload field; the second bit is placed in bit 1 of the first octet and so on. The last bit (b112) is placed in the LSB (bit 7) of octet 14.

5.2.2. Encoding of Silence Description Frames

TOC

The GSM-HR Codec applies a specific coding for silence periods in so called SID frames. The coding of SID frames is based on the coding of speech frames by using only the first 33 bits for SID parameters and by setting the remaining 79 bits all to “1”.

5.3. Implementation Considerations

TOC

An application implementing this payload format MUST understand all the payload parameters that is defined in this specification. Any mapping of the parameters to a signaling protocol MUST support all parameters. So an implementation of this payload format in an application using SDP is required to understand all the payload parameters in their SDP-mapped form. This requirement ensures that an implementation always can decide whether it is capable of communicating when the communicating entities support this version of the specification.

5.3.1. Transmission of SID frames

TOC

When using this RTP payload format the sender SHOULD generate and send SID frames every 160ms, i.e. every 8th frame, during silent periods. Other SID transmission intervals may occur due to gateways to other systems that uses other transmission intervals.

5.3.2. Receiving Redundant Frames

TOC

The reception of redundant audio frames, i.e. more than one audio frame from the same source for the same time slot, MUST be supported by the implementation.

5.3.3. Decoding Validation

TOC

If the receiver finds a mismatch between the size of a received payload and the size indicated by the ToC of the payload, the receiver SHOULD discard the packet. This is recommended because decoding a frame parsed from a payload based on erroneous ToC data could severely degrade the audio quality.

---

6. Examples

TOC

A few examples to highlight the payload format.

6.1. 3 frames

TOC

A basic example of the aggregation of 3 consecutive speech frames into a single frame.

```
The first 24 bits are ToC elements.
Bit 0 is '1' as another ToC element follow.
Bits 1..3 is 000 = Good speech frame
Bits 4..7 is 0000 = Reserved
Bits 8 is '1' as another ToC element follow.
Bits 9..11 is 000 = Good speech frame
Bits 12..15 is 0000 = Reserved
Bit 16 is '0', no more ToC element follows
Bits 17..19 is 000 = Good speech frame
Bits 20..23 is 0000 = Reserved
```

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|1|0 0 0|0 0 0 0|1|0 0 0|0 0 0 0|0|0 0 0|0 0 0 0|b1           b8|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+             +
|b9   Frame 1                                                 b40|
+                                                               +
|b41                                                          b72|
+                                                               +
|b73                                                         b104|
+               +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|b105       b112|b1                                           b24|
+-+-+-+-+-+-+-+-+                                               +
|b25  Frame 2                                                 b56|
+                                                               +
|b57                                                          b88|
+                                               +-+-+-+-+-+-+-+-+
|b89                                        b112|b1           b8|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+             +
|b9   Frame 3                                                 b40|
+                                                               +
|b41                                                          b72|
+                                                               +
|b73                                                         b104|
+               +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|b105       b112|
+-+-+-+-+-+-+-+-+
```

TOC

6.2. 3 Frames with lost frame in the middle

An example of payload carrying 3 frames where the middle one is No_Data, for example due to loss prior to transmission by the RTP source.

```
        The first 24 bits are ToC elements.
        Bit 0 is '1' as another ToC element follow.
        Bits 1..3 is 000 = Good speech frame
        Bits 4..7 is 0000 = Reserved
        Bits 8 is '1' as another ToC element follow.
        Bits 9..11 is 111 = No_Data frame
        Bits 12..15 is 0000 = Reserved
        Bit 16 is '0', no more ToC element follows
        Bits 17..19 is 000 = Good speech frame
        Bits 20..23 is 0000 = Reserved


         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
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
        |1|0 0 0|0 0 0 0|1|1 1 1|0 0 0 0|0|0 0 0|0 0 0 0|b1           b8|
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+               +
        |b9   Frame 1                                                b40|
        +                                                               +
        |b41                                                         b72|
        +                                                               +
        |b73                                                        b104|
        +               +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
        |b105       b112|b1                                          b24|
        +-+-+-+-+-+-+-+-+                                               +
        |b25  Frame 3                                                b56|
        +                                                               +
        |b57                                                         b88|
        +                                               +-+-+-+-+-+-+-+-+
        |b89                                        b112|
        +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

---

7. Payload Format Parameters

TOC

This RTP payload format is identified using the media type "audio/gsm-hr-08", which is registered in accordance with [RFC4855] (Casner, S., “Media Type Registration of RTP Payload Formats,” February 2007.) and using the template of [RFC4288] (Freed, N. and J. Klensin, “Media Type Specifications and Registration Procedures,” December 2005.). Note: Media subtype names are case-insensitive.

7.1. Media Type Definition

TOC

The media type for the GSM-HR codec is allocated from the IETF tree since GSM-HR is a well know speech codec. This media type registration covers real-time transfer via RTP. The media subtype name contains "-08" to avoid potential conflict with any earlier drafts of GSM-HR RTP payload types that aren't bit compatible.
Note, reception of any unspecified parameter MUST be ignored by the receiver to ensure that additional parameters can be added in the future.
Type name: audio
Subtype name: GSM-HR-08
Required parameters: none
Optional parameters:

> **max-red:** The maximum duration in milliseconds that elapses between the primary (first) transmission of a frame and any redundant transmission that the sender will use. This parameter allows a receiver to have a bounded delay when redundancy is used. Allowed values are integers between 0 (no redundancy will be used) and 65535. If the parameter is omitted, no limitation on the use of redundancy is present.
>
> **ptime:** see [RFC4566] (Handley, M., Jacobson, V., and C. Perkins, “SDP: Session Description Protocol,” July 2006.).
>
> **maxptime:** see [RFC4566] (Handley, M., Jacobson, V., and C. Perkins, “SDP: Session Description Protocol,” July 2006.).

Encoding considerations:

> This media type is framed and binary, see section 4.8 in RFC4288 (Freed, N. and J. Klensin, “Media Type Specifications and Registration Procedures,” December 2005.) [RFC4288].

Security considerations:

> See Section 10 (Security Considerations) of RFCXXXX.

Interoperability considerations:
Published specification:

> RFC XXXX, 3GPP TS 46.002

Applications that use this media type:

Real-time audio applications like voice over IP and teleconference.

Additional information: none
Person & email address to contact for further information:

Ingemar Johansson <ingemar.s.johansson@ericsson.com>

Intended usage: COMMON
Restrictions on usage:

This media type depends on RTP framing, and hence is only defined for transfer via RTP [RFC3550] (Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications,” July 2003.). Transport within other framing protocols is not defined at this time.

Author:

Xiaodong Duan <duanxiaodong@chinamobile.com>

Shuaiyu Wang <wangshuaiyu@chinamobile.com>

Magnus Westerlund <magnus.westerlund@ericsson.com>

Ingemar Johansson <ingemar.s.johansson@ericsson.com>

Karl Hellwig <karl.hellwig@ericsson.com>

Change controller:

IETF Audio/Video Transport working group delegated from the IESG.

7.2. Mapping to SDP

TOC

The information carried in the media type specification has a specific mapping to fields in the Session Description Protocol (SDP) [RFC4566] (Handley, M., Jacobson, V., and C. Perkins, “SDP: Session Description Protocol,” July 2006.), which is commonly used to describe RTP sessions. When SDP is used to specify sessions employing the GSM-HR codec, the mapping is as follows:

* The media type ("audio") goes in SDP "m=" as the media name.

* The media subtype (payload format name) goes in SDP "a=rtpmap" as the encoding name. The RTP clock rate in "a=rtpmap" MUST be 8000, and the encoding parameters (number of channels) MUST either be explicitly set to 1 or omitted, implying a default value of 1.

- The parameters "ptime" and "maxptime" go in the SDP "a=ptime" and "a=maxptime" attributes, respectively.
- Any remaining parameters go in the SDP "a=fmtp" attribute by copying them directly from the media type parameter string as a semicolon-separated list of parameter=value pairs.

7.2.1. Offer/Answer Considerations

TOC

The following considerations apply when using SDP Offer-Answer procedures to negotiate the use of GSM-HR payload in RTP:

- The SDP offerer and answerer MUST generate GSM-HR packets as described by the offered parameters.
- In most cases, the parameters "maxptime" and "ptime" will not affect interoperability; however, the setting of the parameters can affect the performance of the application. The SDP offer-answer handling of the "ptime" parameter is described in [RFC3264] (Rosenberg, J. and H. Schulzrinne, “An Offer/Answer Model with Session Description Protocol (SDP),” June 2002.). The "maxptime" parameter MUST be handled in the same way.
- The parameter "max-red" is a stream property parameter. For sendonly or sendrecv unicast media streams, the parameter declares the limitation on redundancy that the stream sender will use. For recvonly streams, it indicates the desired value for the stream sent to the receiver. The answerer MAY change the value, but is RECOMMENDED to use the same limitation as the offer declares. In the case of multicast, the offerer MAY declare a limitation; this SHALL be answered using the same value. A media sender using this payload format is RECOMMENDED to always include the "max-red" parameter. This information is likely to simplify the media stream handling in the receiver. This is especially true if no redundancy will be used, in which case "max-red" is set to 0.
- Any unknown media type parameter in an offer SHALL be removed in the answer.

TOC

7.2.2. Declarative SDP Considerations

In declarative usage, like SDP in RTSP [RFC2326] (Schulzrinne, H., Rao, A., and R. Lanphier, “Real Time Streaming Protocol (RTSP),” April 1998.) or SAP [RFC2974] (Handley, M., Perkins, C., and E. Whelan, “Session Announcement Protocol,” October 2000.), the parameters SHALL be interpreted as follows:

- *The stream property parameter ("max-red") is declarative, and a participant MUST follow what is declared for the session. In this case it means that the receiver MUST be prepared to allocate buffer memory for the given redundancy. Any transmissions MUST NOT use more redundancy then what has been declared. More than one configuration may be provided if necessary by declaring multiple RTP payload types; however, the number of types should be kept small.

- *Any "maxptime" and "ptime" values should be selected with care to ensure that the session's participants can achieve reasonable performance.

---

8. IANA Considerations

TOC

One media type (audio/gsm-hr-08) has been defined and needs registration in the media types registry; see Section 7.1 (Media Type Definition).

---

9. Congestion Control

TOC

The general congestion control considerations for transporting RTP data apply; see RTP [RFC3550] (Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, “RTP: A Transport Protocol for Real-Time Applications,” July 2003.) and any applicable RTP profile like AVP [RFC3551] (Schulzrinne, H. and S. Casner, “RTP Profile for Audio and Video Conferences with Minimal Control,” July 2003.).
The number of frames encapsulated in each RTP payload highly influences the overall bandwidth of the RTP stream due to header overhead constraints. Packetizing more frames in each RTP payload can reduce the number of packets sent and hence the header overhead, at the expense of increased delay and reduced error robustness. If forward error correction (FEC) is used, the amount of FEC-induced redundancy needs to be regulated such that the use of FEC itself does not cause a congestion problem.

10. Security Considerations

TOC

RTP packets using the payload format defined in this specification are subject to the security considerations discussed in the RTP specification (Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," July 2003.) [RFC3550] , and in any applicable RTP profile. The main security considerations for the RTP packet carrying the RTP payload format defined within this memo are confidentiality, integrity and source authenticity. Confidentiality is achieved by encryption of the RTP payload. Integrity of the RTP packets through suitable cryptographic integrity protection mechanism. Cryptographic system may also allow the authentication of the source of the payload. A suitable security mechanism for this RTP payload format should provide confidentiality, integrity protection and at least source authentication capable of determining if an RTP packet is from a member of the RTP session or not.

Note that the appropriate mechanism to provide security to RTP and payloads following this memo may vary. It is dependent on the application, the transport, and the signalling protocol employed. Therefore a single mechanism is not sufficient, although if suitable the usage of SRTP (Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)," March 2004.) [RFC3711] is recommended. Other mechanism that may be used are IPsec (Kent, S. and K. Seo, "Security Architecture for the Internet Protocol," December 2005.) [RFC4301] and TLS (Dierks, T. and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2," August 2008.) [RFC5246] (RTP over TCP), but also other alternatives may exist.

This RTP payload format and its media decoder do not exhibit any significant non-uniformity in the receiver-side computational complexity for packet processing, and thus are unlikely to pose a denial-of-service threat due to the receipt of pathological data. Nor does the RTP payload format contain any active content.

11. Acknowledgements

TOC

The author would like to thank Xiaodong Duan, Shuaiyu Wang, Rocky Wang and Ying Zhang for their initial work in this area. Many thanks also go to Tomas Frankkila for useful input and comments.

12. References

TOC

12.1. Normative References

TOC

[RFC2119]	Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," BCP 14, RFC 2119, March 1997 (TXT, HTML, XML).
[RFC3264]	Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model with Session Description Protocol (SDP)," RFC 3264, June 2002 (TXT).
[RFC3550]	Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," STD 64, RFC 3550, July 2003 (TXT, PS, PDF).
[RFC3551]	Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control," STD 65, RFC 3551, July 2003 (TXT, PS, PDF).
[RFC4566]	Handley, M., Jacobson, V., and C. Perkins, "SDP: Session Description Protocol," RFC 4566, July 2006 (TXT).
[RFC5405]	Eggert, L. and G. Fairhurst, "Unicast UDP Usage Guidelines for Application Designers," BCP 145, RFC 5405, November 2008 (TXT).
[TS46.002]	3GPP, "Specification : 3GPP TS 46.002 http://www.3gpp.org/ftp/Specs/archive/46_series/46.002/46002-700.zip," June 2007.
[TS46.020]	3GPP, "Specification : 3GPP TS 46.020 http://www.3gpp.org/ftp/Specs/archive/46_series/46.002/46020-700.zip," June 2007.

12.2. Informative References

TOC

[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 (TXT, HTML, XML).
[RFC2326]	Schulzrinne, H., Rao, A., and R. Lanphier, "Real Time Streaming Protocol (RTSP)," RFC 2326, April 1998 (TXT).
[RFC2974]	Handley, M., Perkins, C., and E. Whelan, "Session Announcement Protocol," RFC 2974, October 2000 (TXT).
[RFC3711]	Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)," RFC 3711, March 2004 (TXT).
[RFC4288]	

	Freed, N. and J. Klensin, “Media Type Specifications and Registration Procedures,” BCP 13, RFC 4288, December 2005 (TXT).
[RFC4301]	Kent, S. and K. Seo, “Security Architecture for the Internet Protocol,” RFC 4301, December 2005 (TXT).
[RFC4855]	Casner, S., “Media Type Registration of RTP Payload Formats,” RFC 4855, February 2007 (TXT).
[RFC5109]	Li, A., “RTP Payload Format for Generic Forward Error Correction,” RFC 5109, December 2007 (TXT).
[RFC5246]	Dierks, T. and E. Rescorla, “The Transport Layer Security (TLS) Protocol Version 1.2,” RFC 5246, August 2008 (TXT).

Authors' Addresses

TOC

	Xiaodong Duan
	China Mobile Communications Corporation
	53A, Xibianmennei Ave., Xuanwu District
	Beijing, 100053
	P.R. China
Phone:	
Fax:	
Email:	duanxiaodong@chinamobile.com
URI:	
	Shuaiyu Wang
	China Mobile Communications Corporation
	53A, Xibianmennei Ave., Xuanwu District
	Beijing, 100053
	P.R. China
Phone:	
Fax:	
Email:	wangshuaiyu@chinamobile.com
URI:	
	Magnus Westerlund
	Ericsson AB
	Farogatan 6
	Stockholm, SE-164 80
	Sweden
Phone:	+46 8 719 0000
Fax:	
Email:	magnus.westerlund@ericsson.com
URI:	
	Karl Hellwig

	Ericsson AB
	Kackertstrasse 7-9
	52072 Aachen
	Germany
Phone:	+49 2407 575-2054
Email:	karl.hellwig@ericsson.com
	Ingemar Johansson
	Ericsson AB
	Laboratoriegrand 11
	SE-971 28 Lulea
	SWEDEN
Phone:	+46 73 0783289
Email:	ingemar.s.johansson@ericsson.com