RTP Payload Format for GSM-HR Speech Codecs
draft-wang-avt-rtp-gsm-hr-00.txt

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with Section 6 of BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

This Internet-Draft will expire on August 17 2008.

Copyright Notice

Copyright (C) The IETF Trust (2008).

Abstract

This document registers a media type for the Real-time Transport protocol (RTP) payload format, which is used for the Group Special Mobile half-rate speech transcoding.
1. Introduction

Global System for Mobile Communication (GSM) network has been widely deployed in the last several years to provide mobile communication services. GSM half rate codec (GSM-HR) is one of the compressed speech codecs which are used for the basic speech service in the GSM mobile networks.

GSM-HR denotes GSM 06.20 half-rate speech transcoding, specified in ETS 300 969 which is available from ETSI at the address given in RFC 3551 [1] Section 4.5.8. This codec has a frame length of 112 bits. For transmission in RTP, each codec frame is packed into a 14 octet (112 bit). The packing is specified in ETSI Technical Specification TS 101 318.

This document registers a media type for the Real-time Transport protocol (RTP) payload format for the GSM-HR codec to enabling the use of the codec in the Voice over IP (VoIP) application.

1.1. Terminology

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 [2].
2. Background for GSM HR Speech codes over RTP

2.1. GSM HR Codes

As defined in the GSM standard, the GSM physical layer is presented as a combination of FDMA and TDMA. Based on the frequency division principle, the entire GSM band is divided into several 200-KHz channels which named as absolute radio frequency channel number in the Standard (ARFCN for short). Each channel is divided into 8 timeslots according to the time division principle. Hence, a GSM physical channel in fact corresponds to one timeslot on a certain channel.

Timeslots are repeated according to specific disciplinarian, bringing the concepts of frames and multi-frames. In GSM Standard, the channels bearing voice and data services are repeated in a period of 26 frames. In words, the TDMA multiframe of a traffic channel (TCH) consists of 26 frames, with the duration of 120 ms.

Among the 26 frames on a full rate speech channel (TCH/FS) or an enhanced full rate speech channel (TCH/EFS), 24 are used to bear speech data, one (Frame 13) to transmit the channel associated signaling on a slow associated control channel (SACCH), and one (Frame 26) is idle.

When the system adopts half rate speech channel (TCH/HS), no change will bring to the multi-frame structure of the air interface. In this case, the odd frames in the multi-frame are allocated to one user and even frames to another, while the original Frame 13 serves as the SACCH of the first user and the idle Frame 26 as the SACCH of the second user. In this way, a channel previously only capable of bearing one TCH/FS or TCH/EFS can now bear two TCH/HSs. The capacity of the channel doubles the original.

2.2. A interface over IP

A interface is the interface between MSC and BSC.

```
+----------+                         +----------+
|          |      A interface        |          |
|   BSC    |<----------------------->|   MSC    |
|          |           TDM           |          |
+----------+                         +----------+
```

Figure 1 A interface (Legacy network).
According to the existing protocol, the A interface user plane is only based on TDM. A interface over IP is a technique trend in wireless network evolution, which can construct high bandwidth, high efficiency and low cost basic networks. For A interface over IP, control plane signalling over IP (SIGTRAN) has been introduced in 3GPP Release 7 while certain features (e.g. MSC in Pool and Layered Architecture) require an intermediate signalling network for best performance. In order to take full advantage of IP based technologies the protocols of A interface user plane should be adapted for IP based transport.

A interface over IP can also simplify the implementation of MSCs in a pool. Furthermore, UTRAN network and more advanced RAN can use a common IP backhaul with GERAN. One of the main advantages of having IP based A-interface for the user plane is a much more flexible network design between the BSS and the CS core.

IP hardware in the nodes and IP site and backbone infrastructure can be shared by the A-interface control plane and the user plane. A separation of the signaling network from the user plane can be achieved by using technologies like VLAN tagging, virtual routing etc. This will allow the operator to abolish TDM hardware and TDM infrastructure and by that reduce OPEX and CAPEX.

Further on in most of the current networks, both BSS and CN have transcoding functionality, i.e. Transcoder in BSS and Media Gateway (MGW) in CN. Some core networks have been upgraded to convey compressed speech over IP transport. In this case, removing TC from BSS and transfer compressed speech over A interface will reduce cost of transcoder device, reduce cost of transport resource and improve voice quality by implementing TrFO.

Figure 2 Architecture for A interface over IP.
2.3. GSM HR Speech codes over IP Scenarios

There is an enormous amount of transcoder resources installed in today’s GSM radio networks. Therefore the "final solution" in the standard shall be flexible and allow the use of transcoders placed in the BSS and/or in the CS Core Network. In addition, e.g. for the purpose of migrating the A interface from a TDM to an IP interface, both TDM and IP based A interface should be supported concurrently, at least for the migration phase.

Target solution, the solution yields to align the 2G network architecture with the 3G CS core network architecture, deviates from the current BSS architecture, where today transcoders are functionally integrated into the BSS. Allowing placing transcoders in the CS core network will impacts the functional division between Base Station System (BSS) and Core Network. This solution allows carrying compressed speech in an efficient way across the A-interface. In contrast to TFO the compressed speech is formatted directly and there is no PCM stream in parallel. This will reduce the overall need for transcoders in the solution and it will improve the end-to-end delay. But it will require additional transcoder resources (e.g. for transcoding in all Mobile-to-PSTN calls) and possibly new transcoder types (e.g. GSM_HR) within the Core Network.

In this case, GSM-HR needs to be transferred over IP based A interface if GSM-HR is applied in the air interface.

```
+-------------+                         +----------+
|             |   A interface over IP   |          |
|    BSC      |<----------------------->|   MGW    |
|  without TC |                         |  with TC |
+-------------+                         +----------+

Figure 3 Deployment Scenario: compressed speech(such as GSM-HR) is transferred over IP based A interface.
3. GSM HR codes RTP Payload Formats

3.1. Payload Structure

The GSM half rate codec has frame length of 112 bits. Every frame is encoded into one 14 octet (112 bit) buffer. There shall be no signature.

The complete payload consists of a payload header, a payload table of contents, and speech data representing one or more speech frame-blocks. The following diagram shows the general payload format layout:

```
+----------------+-------------------+----------------
| payload header | table of contents | speech data ... |
+----------------+-------------------+----------------
```

Figure 4 payload structure.

The bits in RTP buffer are numbered from r1 (the most significant bit of first octet) to r112 (the least significant bit of last octet). Within the GSM 06.20 codec parameter bits are numbered in big-endian manner.

1 Encoding of speech frames

There are two alternative formats of a ETS 300 969 [6] speech frames. The first form is for codec mode 0 (unvoiced speech), the second form is for modes 1, 2 and 3 (voiced speech).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>No. of bits</th>
<th>Bit No.(MSB-LSB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R0</td>
<td>5</td>
<td>r1 - r5</td>
</tr>
<tr>
<td>LPC1</td>
<td>11</td>
<td>r6 - r16</td>
</tr>
<tr>
<td>LPC2</td>
<td>9</td>
<td>r17 - r25</td>
</tr>
<tr>
<td>LPC3</td>
<td>8</td>
<td>r26 - r33</td>
</tr>
<tr>
<td>Parameter</td>
<td>No. of bits</td>
<td>Bit No. (MSB-LSB)</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>INT_LPC</td>
<td>1</td>
<td>r34</td>
</tr>
<tr>
<td>MODE</td>
<td>2</td>
<td>r35 - r36</td>
</tr>
<tr>
<td>CODE1_1</td>
<td>7</td>
<td>r37 - r43</td>
</tr>
<tr>
<td>CODE2_1</td>
<td>7</td>
<td>r44 - r50</td>
</tr>
<tr>
<td>GSP0_1</td>
<td>5</td>
<td>r51 - r55</td>
</tr>
<tr>
<td>CODE1_2</td>
<td>7</td>
<td>r56 - r62</td>
</tr>
<tr>
<td>CODE2_2</td>
<td>7</td>
<td>r63 - r69</td>
</tr>
<tr>
<td>GSP0_2</td>
<td>5</td>
<td>r70 - r74</td>
</tr>
<tr>
<td>CODE1_3</td>
<td>7</td>
<td>r75 - r81</td>
</tr>
<tr>
<td>CODE2_3</td>
<td>7</td>
<td>r82 - r88</td>
</tr>
<tr>
<td>GSP0_3</td>
<td>5</td>
<td>r89 - r93</td>
</tr>
<tr>
<td>CODE1_4</td>
<td>7</td>
<td>r94 - r100</td>
</tr>
<tr>
<td>CODE2_4</td>
<td>7</td>
<td>r101 - r107</td>
</tr>
<tr>
<td>GSP0_4</td>
<td>5</td>
<td>r108 - r112</td>
</tr>
</tbody>
</table>

Table 1: The order of GSM 06.20 half rate speech codec parameters in RTP buffer (MODE=0)
Table 2: The order of GSM 06.20 half rate speech codec parameters in RTP buffer (MODE=1, 2 or 3)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>No. of bits</th>
<th>Value (Hex)</th>
</tr>
</thead>
<tbody>
<tr>
<td>INT_LPC</td>
<td>1</td>
<td>r34</td>
</tr>
<tr>
<td>MODE</td>
<td>2</td>
<td>r35 - r36</td>
</tr>
<tr>
<td>LAG_1</td>
<td>8</td>
<td>r37 - r44</td>
</tr>
<tr>
<td>CODE1</td>
<td>9</td>
<td>r45 - r53</td>
</tr>
<tr>
<td>GSP0_1</td>
<td>5</td>
<td>r54 - r58</td>
</tr>
<tr>
<td>LAG_2</td>
<td>4</td>
<td>r59 - r62</td>
</tr>
<tr>
<td>CODE2</td>
<td>9</td>
<td>r63 - r71</td>
</tr>
<tr>
<td>GSP0_2</td>
<td>5</td>
<td>r72 - r76</td>
</tr>
<tr>
<td>LAG_3</td>
<td>4</td>
<td>r77 - r80</td>
</tr>
<tr>
<td>CODE3</td>
<td>9</td>
<td>r81 - r89</td>
</tr>
<tr>
<td>GSP0_3</td>
<td>5</td>
<td>r90 - r94</td>
</tr>
<tr>
<td>LAG_4</td>
<td>4</td>
<td>r95 - r98</td>
</tr>
<tr>
<td>CODE4</td>
<td>9</td>
<td>r99 - r107</td>
</tr>
<tr>
<td>GSP0_4</td>
<td>5</td>
<td>r108 - r112</td>
</tr>
</tbody>
</table>

2 Encoding of silence indication frames

The half-rate codec SID frame is encoded according to the ETS 300 971 [7].

A SID frame is identified by a SID codeword consisting of 79 bits which are all 1. The parameters in table 3 have to be set as shown in order to mark a frame as a SID frame.

Parameter       No. of bits Value (Hex)
-------------------------------
INT_LPC         1           116

Table 2: The order of GSM 06.20 half rate speech codec parameters in RTP buffer (MODE=1, 2 or 3)
MODE              2           316  
LAG_1             8           FF16  
CODE1             9           1FF16  
GSP0_1           5            1F16  
LAG_2             4           F16  
CODE2             9           1FF16  
GSP0_2           5            1F16  
LAG_3             4           F16  
CODE3             9           1FF16  
GSP0_3           5            1F16  
LAG_4             9           1FF16  
GSP0_4           5            1F16  

Table 3: SID codeword for half rate speech codec

3.2. Registration of Media Type GSM-HR

Type name: audio

Subtype name: GSM-HR

Required parameters: none

Optional parameters:

ptime: the recommended length of time (in milliseconds) represented by the media in a packet and the default value is 20 milliseconds. See Section 6 of RFC 4566 [3].

Encoding considerations:

This media type is framed binary data (see Section 4.8 in RFC4288 [4]).
Security consideration:

This media type does not carry active content. It does transfer compressed data.

Interoperability considerations: none

Published specification: RFC XXXX

Applications that use this media type:

Audio and video streaming and conferencing tools

Additional information: none

Person & email address to contact for further information:

Xiaodong Duan duanxiaodong@chinamobile.com

Intended usage: COMMON

Restrictions on usage:

This media type depends on RTP framing, and hence is only defined for transfer via RTP (RFC 3550 [5]). Transfer within other framing protocols is not defined at this time.

Author:

Xiaodong Duan duanxiaodong@chinamobile.com
Shuaiyu Wang wangshuaiyu@chinamobile.com

Change controller:

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

4. Mapping MIME Parameters into SDP

The information carried in the MIME media type specification has a specific mapping to fields in the Session Description Protocol (SDP) [3], which is commonly used to describe RTP sessions. When SDP is used to specify sessions employing the compact bundled format for GSM half-rate speech, the mapping is as follows:
The MIME type ("audio") goes in SDP "m=" as the media name.

The MIME subtype ("GSM-HR") goes in SDP "a=rtpmap" as the encoding name and the sampling rate for the GSM-HR codec is 8 KHz.

The optional parameters "ptime" goes in the SDP "a=ptime" attributes, respectively.

The payload type payload type value for GSM-HR is created dynamically and is used in the PT field of the RTP data header.

5. Security Considerations

RTP packets using the GSM-HR payload format are subject to the security considerations discussed in the RTP specification [5].

A potential denial-of-service threat exists for data encodings using compression techniques that have non-uniform receiver-end computational load. The attacker can inject pathological datagrams into the stream which are complex to decode and cause the receiver to be overloaded. However, this encoding does not exhibit any significant non-uniformity.

As with any IP-based protocol, in some circumstances a receiver may be overloaded simply by the receipt of too many packets, either desired or undesired. Network-layer authentication MAY be used to discard packets from undesired sources, but the processing cost of the authentication itself may be too high.

6. IANA Considerations

It is requested that one new media subtype (audio/GSM-HR) is registered by IANA. For details, see Section 2.
7. References

7.1. Normative References


[6] ETS 300 969: "Digital cellular telecommunications system (Phase 2+); Half rate speech; Half rate speech transcoding (GSM 06.20)".

[7] ETS 300 971: "Digital cellular telecommunications system; Half rate speech; Comfort noise aspects for the half rate speech traffic channels (GSM 06.22)"

Author’s Addresses

Xiaodong Duan  
China Mobile Communications Corporation  
53A, Xibianmennei Ave., Xuanwu District  
Beijing, 100053 P.R. China  
Email: <duanxiaodong@chinamobile.com>

Shuaiyu Wang  
China Mobile Communications Corporation  
53A, Xibianmennei Ave., Xuanwu District  
Beijing, 100053 P.R. China  
Email: <wangshuaiyu@chinamobile.com>
Rocky Wang
Huawei Technologies
Xibianmennei AveHuawei Industrial Base, Bantian Longgang
Shenzhen, Guangdong Province 518129, P.R.China

Email: <rocky.wang@huawei.com>

Ying Zhang
Huawei Technologies
Huawei Industrial Base, Bantian Longgang
Shenzhen, Guangdong Province 518129, P.R.China

Email: <zhangying67324@huawei.com>

Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in BCP 78 and BCP 79.

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF