Internet Engineering Task Force   Peter Barany, Nortel Networks
Audio Video Transport WG   William Navarro, Nortel Networks
INTERNET-DRAFT
November 14, 2001
Expires: May 14, 2002

RTP payload format for EFR speech codec
<draft-barany-avt-efr-00.txt>

Status of this Memo

Abstract

   This document specifies a Real-Time Transport Protocol (RTP) payload
   format for the Global System for Mobile communications (GSM) Enhanced
   Full Rate (EFR) speech codec. The EFR speech codec RTP payload format
   specified in this document closely resembles the EFR speech codec RTP
   payload format defined in TS 101 318 "Using GSM Speech Codecs Within
   ITU-T Recommendation H.323". It is designed specifically to optimally
   interoperate with existing (i.e., legacy) GSM circuit-switched
   transceiver equipment in the sense that it supports the following
   EFR speech codec circuit-switched domain functionality in the packet-
   switched domain: error concealment of lost speech frames and SIlence
   Descriptor (SID) frames. The EFR speech codec RTP payload format
   defined in TS 101 318 does not support this functionality. A MIME
   type registration for the EFR speech codec is also included.

Revision history

    -00: Document created for specification of an RTP payload format for
         the EFR speech codec.

Conventions used in this document

    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
    [ref-RFC-2119].

Table of contents

1. Introduction

    This document specifies a Real-Time Transport Protocol (RTP) payload
    format for the Global System for Mobile communications (GSM) Enhanced
    Full Rate (EFR) speech codec. The EFR speech codec RTP payload format
    specified in this document closely resembles the EFR speech codec RTP
    payload format defined in [ref-EFR-RTP]. It is designed specifically
    to optimally interoperate with existing (i.e., legacy) GSM circuit-
    switched transceiver equipment in the sense that it supports the
    following EFR speech codec circuit-switched domain functionality in
    the packet-switched domain: error concealment of lost speech frames
    and SIlence Descriptor (SID) frames [ref-EFR-ERR].

1.1. EFR speech codec

    The Enhanced Full Rate (EFR) speech codec [ref-EFR-COD] was developed

by the European Telecommunications Standards Institute (ETSI). The
EFR speech codec is standardized for the Global System for Mobile
communications (GSM).

The EFR speech codec is a single-mode speech codec with a bit rate of
12.2 kbps (i.e., 244 speech bits per 20 ms speech frame). The
sampling frequency is 8,000 Hz, consequently there are 160 samples
per 20 ms speech frame.

In the circuit-switched domain, the EFR speech codec supports the
following functionality:

(1) DTX operation [ref-EFR-DTX]; and

(2) error concealment of lost speech frames and SID frames [ref-EFR-
ERR]

This functionality is important because it makes it possible to
achieve optimum Mean Opinion Scores (MOS) for GSM circuit-switched
voice service using the EFR speech codec.

1.2. Existing RTP payload format for EFR speech codec

An existing RTP payload format for the EFR speech codec is defined
in [ref-EFR-RTP] which is referenced in [ref-RTP-PROF]. A MIME
registration for this RTP payload format is defined in [ref-RTP-
MIME].

While this EFR speech codec RTP payload format can be used to
interoperate with existing (i.e., legacy) GSM circuit-switched
transceiver equipment, the functionality will be suboptimal in the
sense that it does not support the following EFR speech codec
circuit-switched domain functionality in the packet-switched domain:
error concealment of lost speech frames and SID frames [ref-EFR-ERR].

Error concealment of lost speech frames and SID frames is not
possible because the RTP payload format does not incorporate a
payload quality indicator.

1.3. Legacy transceiver interoperability

The GSM/EDGE Radio Access Network (GERAN) (where EDGE stands for
Enhanced Data Rates for Global Evolution) is described in [ref-
GERAN]. GERAN is an evolution of:

(1) GSM circuit-switched voice and data radio access networks; and

(2) General Packet Radio Service (GPRS) and Enhanced GPRS (EGPRS)

packet-switched radio access networks.

GERAN provides an interface between these radio access networks and the Universal Mobile Telecommunications System (UMTS) core network.

Currently, there are a great deal of legacy GSM circuit-switched transceivers deployed in the field by service providers that implement a standardized scheme for channel coding/decoding, interleaving/deinterleaving, CRC, modulation/demodulation, etc. [ref-EFR-CH] for EFR speech codec based GSM circuit-switched voice service.

GERAN defines a service known as the "optimized speech bearer" [ref-GERAN] that makes it possible for a service provider to reuse these legacy GSM circuit-switched transceivers for EFR speech codec based GERAN packet-switched voice service. For the optimized speech bearer service, network level and transport level headers (i.e., IP/UDP/RTP) are not transmitted over the air interface (i.e., Uu interface). The receiving entity (i.e., terminal or radio network controller) can regenerate the headers based upon (1) information submitted during call setup and (2) information derived from lower layers (i.e., link and physical layers). Note that the regenerated headers may not always be semantically identical to the original headers.

Figure 1 illustrates a likely EFR speech codec based GERAN optimized speech bearer scenario where the EFR speech codec is used as a packet-switched application in a GERAN system with existing (i.e., legacy) GSM circuit-switched transceiver equipment.

```
       Uu interface                        Iu-ps interface
+----------+     +-------------+     +------------+     +-----------+
|          |---->|   LEGACY    |---->|   RADIO    |---->|           |
| TERMINAL |     |    BASE     |     |  NETWORK   |     |  GATEWAY  |
|          |<----|   STATION   |<----| CONTROLLER |<----|           |
|          |     | TRANSCEIVER |     |            |     |           |
+----------+     +-------------+     +------------+     +-----------+
```
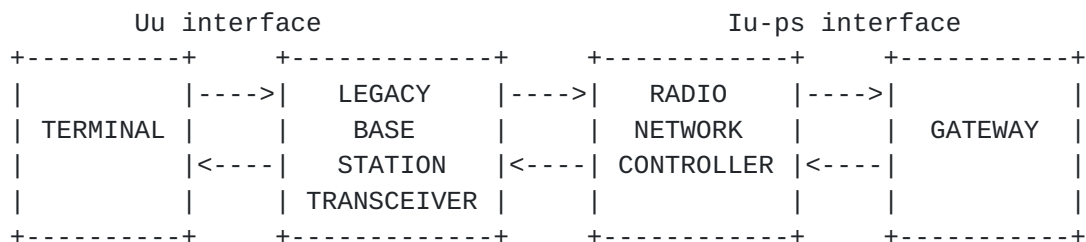
Figure 1. Terminal to gateway scenario.

1.4 EFR speech codec and AMR speech codec comparison

As mentioned in Section 1.1 of this document, the EFR speech codec is a single-mode speech codec with a bit rate of 12.2 kbps (i.e., 244 speech bits per 20 ms speech frame). The sampling frequency is 8,000 Hz, consequently there are 160 samples per 20 ms speech frame. The

original order of the 244 speech bits for the EFR speech codec as
delivered from the speech encoder is defined in Table 5 in [ref-EFR-
COD]. The 244 speech bits pass through a preliminary channel encoder
which produces 260 bits corresponding to 244 input speech bits and 16
redundancy bits [ref-EFR-CH]. The 260 bits are then reordered in
descending bit error sensitivity order according to Table 6 in [ref-
EFR-CH]. This enables the use of Unequal Error Detection (UED) and
Unequal Error Protection (UEP). There are a total of 182 Class 1 bits
(protected) and 78 Class 2 bits (unprotected). The Class 1 bits are
further divided into Class 1a (the 50 most important bits) and Class
1b bits (the 132 next most important bits). The Class 1a bits are
protected by a cyclic code and a convolutional code whereas the Class
1b bits are protected by the convolutional code only.

The 12.2 kbps speech mode is one of the eight Adaptive Multi-Rate
(AMR) speech codec speech modes [ref-AMR-COD]. The original order of
the 244 speech bits for the 12.2 kbps speech mode of the AMR speech
codec as delivered from the speech encoder is defined in Table 9a in
[ref-AMR-COD]. This is the same as that defined for the EFR speech
codec. However, for the AMR speech codec, the 244 speech bits do not
pass through a preliminary channel coder and 16 redundancy bits are
not added. Also, the 244 bits are reordered in descending bit error
sensitivity order in a different manner than that done for the EFR
speech codec (see Table 7 in [ref-EFR-CH]), with the bits being
classified as Class A bits (the 81 most important bits), Class B bits
(the 103 next most important bits), and Class C bits (the least
important 60 bits). See Table 2 in [ref-AMR-FRM].

Another significant difference between the two speech codecs is in
regards to DTX operation. The SID frames are different. The SID frame
for the EFR speech codec is defined in [ref-EFR-CN]. The SID frame
for the AMR speech codec is defined in [ref-AMR-CN, ref-AMR-FRM].
Also, the AMR speech codec has a SID_FIRST and SID_UPDATE frame (in
addition to the SID frame) while the EFR speech codec does not.

In light of these differences, the upshot of all this is that the
EFR speech codec RTP payload format specified in this document is
not based upon the AMR speech codec RTP payload format defined in
[ref-AMR-RTP]. Instead, the EFR speech codec specified in this
document closely resembles the EFR speech codec RTP payload format
defined in [ref-EFR-RTP].

2. Payload format

As mentioned throughout this document, The EFR speech codec RTP
payload format specified in this document closely resembles the EFR
speech codec RTP payload format defined in [ref-EFR-RTP].

The only difference is that the 4 bit signature (0xC, binary 1100)
at the beginning of every buffer for the EFR speech codec RTP payload
format defined in [ref-EFR-RTP] MUST be replaced by a 1 bit payload
quality indicator Q followed by 3 reserved bits R. The payload
quality indicator, if not set, indicates that the payload is severely
damaged and the receiver should set the Bad Frame Indicator (BFI),
see [ref-EFR-DTX], to either "Unusable frame" (for speech frames) or
"Invalid SID frame" (for SID frames). The 3 reserved bits MUST be set
to zero. All R bits MUST be ignored by the receiver.

As is the case for the EFR speech RTP payload format defined in [ref-
EFR-RTP], the bits in the buffer are numbered in the big-endian
manner, starting from r1 (the MSB of the first octet) and finishing
to r248 (the least significant bit of the last octet). Therefore, for
the EFR speech codec RTP payload format specified in this document,
the first octet in the buffer contains QRRR in its 4 MSBs as opposed
to 1100 for the EFR speech codec RTP payload format defined in [ref-
EFR-RTP].

3. IANA considerations

One new MIME sub-type as described in this section is to be
registered.

The MIME-name for the EFR speech codec is allocated from the IETF
tree since the EFR speech codec may be a widely used speech codec for
for GERAN packet-switched voice service using existing (i.e., legacy)
GSM circuit-switched transceiver equipment.

4. Security considerations

RTP packets using the payload format defined in this specification
are subject to the security considerations discussed in the RTP
specification [ref-RTP], and any appropriate profile. This implies
that confidentiality of the media streams is achieved by encryption.
Because the data encoding used with this payload format is applied
end-to-end, encryption may be performed after encoding so there is no
conflict between the two operations.

A potential denial-of-service threat exists for data encodings using
receiver side decoding. The attacker can inject pathological
datagrams into the stream, which are complex to decode and cause the
receiver to be overloaded. The decoder software should consider this
possibility and take the necessary precautions.

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.

5. MIME type registration

Media Type name: audio

Media subtype name: GERAN-EFR

Required parameters: none

Optional parameters: none

Encoding considerations: See Section 2 of this document.

Security considerations: See Section 4 of this document.

Intended usage: COMMON

5.1. Mapping to SDP parameters

Example of usage of EFR speech codec in SDP [ref-SDP], possible GERAN
"optimized voice bearer" service that utilizes existing (i.e.,
legacy) GSM circuit-switched transceiver equipment:

```
m=audio 49120 RTP/AVP 97
a=rtpmap:97 GERAN-EFR/8000
```

6. References

[ref-RFC-2119] RFC 2119 "Key Words for Use in RFCs to Indicate
               Requirement Levels".

[ref-EFR-RTP] TS 101 318 "Using GSM Speech Codecs Within ITU-T
               Recommendation H.323".

[ref-EFR-ERR] 3GPP TS 46.061 "Substitution and muting of lost frames
               for Enhanced Full Rate (EFR) Speech Traffic Channels".

[ref-EFR-COD] 3GPP TS 46.060 "Enhanced Full Rate (EFR) Speech
               Transcoding".

[ref-EFR-DTX] 3GPP TS 46.081 "Discontinuous Transmission (DTX) for
               Enhanced Full Rate (EFR) Speech Traffic Channels".

[ref-RTP-PROF] draft-ietf-avt-profile-new-11.txt "RTP Profile for
               Audio and Video Conferences with Minimal Control".

   [ref-RTP-MIME] draft-ietf-avt-rtp-mime-05.txt "MIME Type Registration
                   of RTP Payload Formats".

   [ref-GERAN]   3GPP TS 43.051 "GSM/EDGE Radio Access Network (GERAN);
                   Overall Description-Stage 2".

   [ref-EFR-CH]   3GPP TS 45.003 "Channel Coding".

   [ref-AMR-COD]   3GPP TS 26.090 "AMR Speech Codec; Transcoding
                   Functions".

   [ref-AMR-FRM]   3GPP TS 26.101 "AMR Speech Codec Frame Structure".

   [ref-EFR-CN]   3GPP TS 46.062 "Comfort noise aspects for Enhanced
                   Full Rate (EFR) Speech Traffic Channels".

   [ref-AMR-CN]   3GPP TS 26.092 "AMR Speech Codec; Comfort Noise
                   Aspects".

   [ref-AMR-RTP]   draft-ietf-avt-rtp-amr-10.txt "RTP Payload Format and
                   File Storage Format for AMR and AMR-WB Audio".

   [ref-RTP]   draft-ietf-avt-rtp-new-10.txt " RTP: A Transport
                   Protocol for Real-Time Applications".

   [ref-SDP]   draft-ietf-mmusic-sdp-new-03.txt " SDP: Session
                   Description Protocol".

7. Authors' Addresses

   Peter Barany   Tel: +1 972 685 2471
   Nortel Networks   EMail: pbarany@nortelnetworks.com
   2201 Lakeside Boulevard
   Richardson, Texas 75083
   United States of America

   William Navarro   Tel: +33 1 39 44 57 56
   Nortel Networks   EMail: navarro@nortelnetworks.com
   19, Avenue du Centre
   Montigny-le-Bretonneaux - PC CT111
   78928 Yvelines Cedex 9
   France