

vMRF Overview

Virtual Multimedia Resource Function

Function Survey

Copyright

© Ericsson AB 2016, 2017. All rights reserved. No part of this document may be reproduced in any form without the written permission of the copyright owner.

Disclaimer

The contents of this document are subject to revision without notice due to continued progress in methodology, design and manufacturing. Ericsson shall have no liability for any error or damage of any kind resulting from the use of this document.



Contents

1	Function Overview	1
1.1	vMRF Architecture	2
1.2	Virtual Machine Functions in vMRF	2
1.3	Roaming SC in vMRF	3
2	Audio Codec Selection Principles	5
3	Audio Conferencing	6
4	Audio Announcements	7
5	Detection of Dual Tone Multi Frequency (DTMF) Tones	8
6	Forwarding of DTMF Tones through Audio Conference	9
7	Playing Tones	10
8	Emergency and Priority Call Handling	11
9	Overload Protection	12
10	Shared Storage	13
11	Address Resolution and Fetching the Next Hop Address	15
12	Media Stream Processing Services	16
12.1	Adaptive Multi-Rate (AMR) Speech Coder Service	16
12.2	Adaptive Multi-Rate Wideband (AMR-WB) Speech Coder Service	16
12.3	Announcement Service	17
12.4	Audio Mixing Service	17
12.5	DTMF Receiver Service (DTMF-R)	17
12.6	DTMF Sender Service (DTMF-S)	17
12.7	Enhanced Voice Services (EVS)	18
12.8	G.722 Service (G.722)	18
12.9	G.729 Service (G.729)	18
12.10	Jitter Handling Service	19
12.11	Pulse Code Modulation Service (PCM)	19
12.12	Real-Time Transport Protocol and Real-Time Transport Control Protocol Service (RTP/RTCP)	19



12.13	Tone Sender Service	19
12.14	User Plane Frame Handler Service (UP FH)	20
Reference List		21



1 Function Overview

The Virtual Multimedia Resource Function (vMRF) is used to mix and process media streams. vMRF is controlled by the Multimedia Telephony Application Server (MTAS), using H.248 control protocol. Figure 1 illustrates signaling and payload traffic related to vMRF.

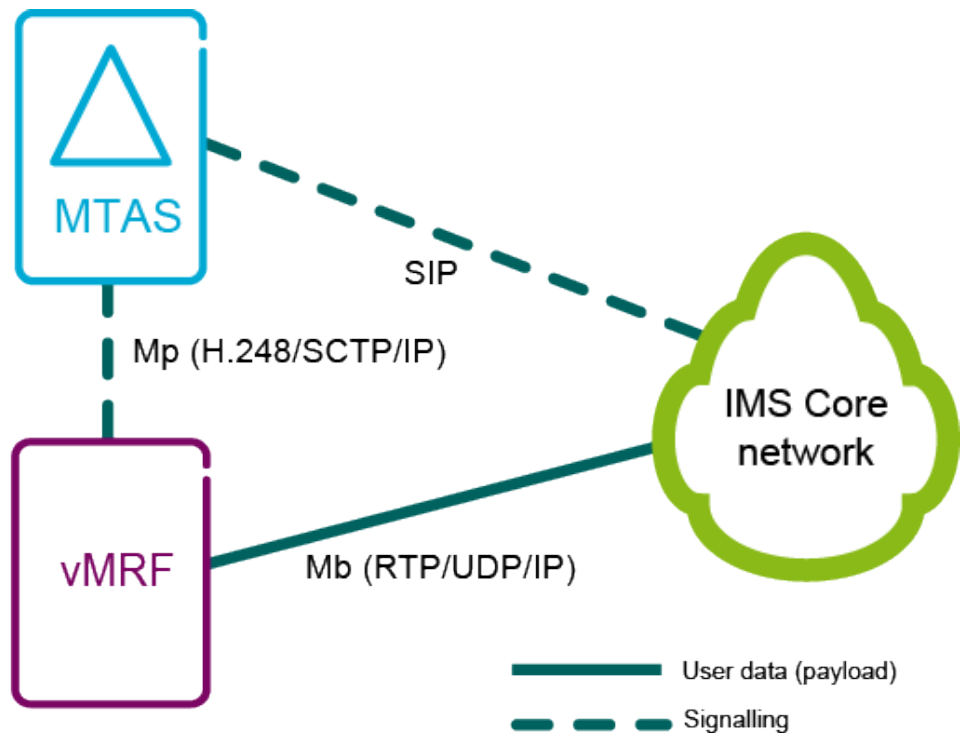


Figure 1 vMRF Overview

vMRF supports the following functions:

- Playing audio announcements
- Audio conferencing
- Detection of DTMF tones
- DTMF tone forwarding for audio conference
- Playing tones

vMRF supports the following audio codecs: AMR-NB, AMR-WB, EVS, G.711, G.722, and G.729.



Note: The Consumer Communication capacity license is needed for vMRF functions.

The use of EVS requires the Enhanced Voice Services capacity license and connection to a Network License Server (NeLS).

For more information on licensing, refer to *License Management*.

1.1 vMRF Architecture

vMRF can contain one or more Virtualized Network Functions (VNFs). A single VNF contains multiple VMs. See [Figure 2](#) for an example overview of the vMRF with two VNFs.

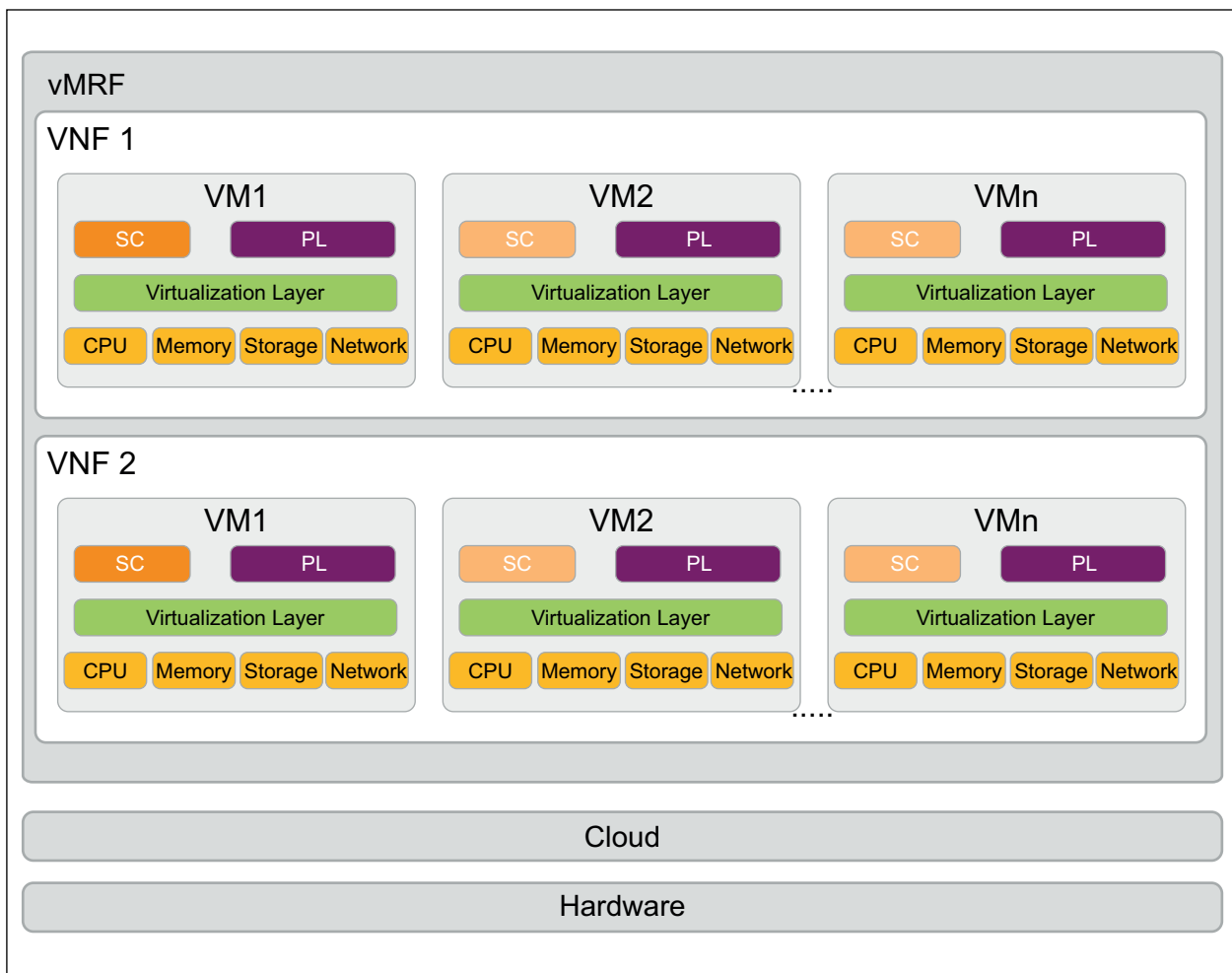


Figure 2 vMRF Architecture

1.2 Virtual Machine Functions in vMRF

In the vMRF VNF, each VM provides the following functions:



- **Payload (PL) function:** The PL function processes user plane traffic and H. 248 signaling traffic. Each VM in a VNF provides the same PL functionality and has the same configuration (except local IP addresses).
- **System Controller (SC) function:** The SC function processes O&M traffic and is responsible for the VNF internal clustering function that is needed for scaling out and scaling in. All VMs in the VNF can act as the SC, but the SC function is active in only one VM at a time. The active SC VM is selected by the Roaming SC function, see [Roaming SC in vMRF](#) on page 3.

See [unresolved external reference] for an overview of the vMRF architecture.

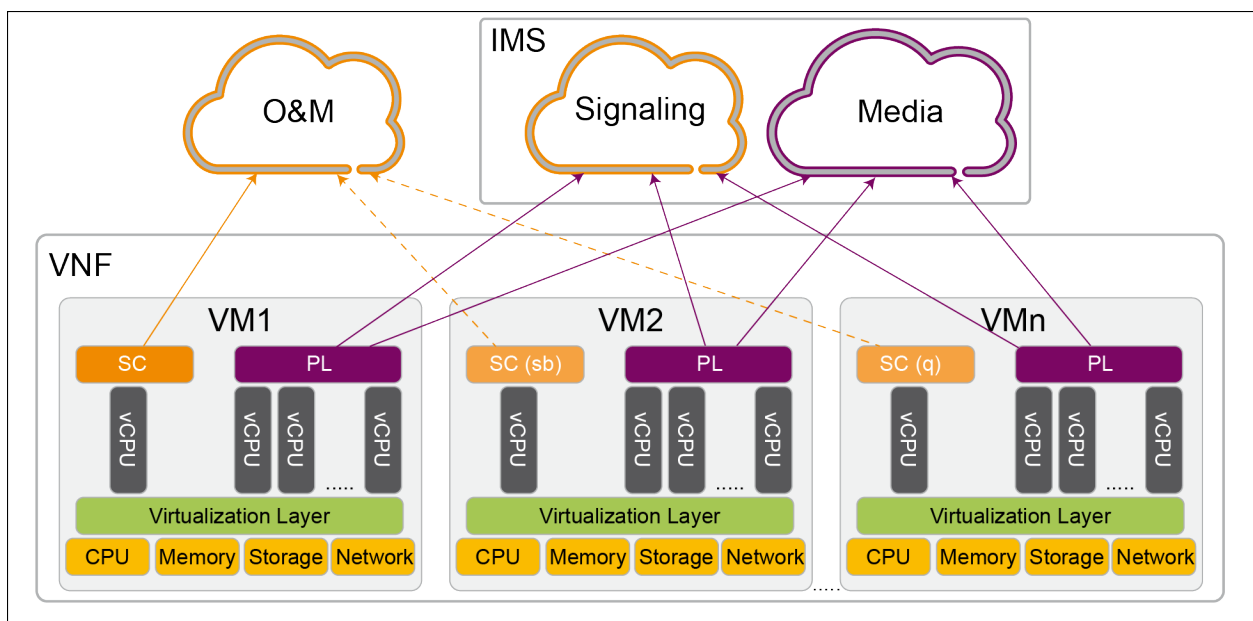


Figure 3 vMRF Architecture

1.3 Roaming SC in vMRF

In vMRF, O&M functions are secured with a Redundancy scheme called Roaming SC. This means that all virtual machines can process payload and also act as a system controller when necessary. If the SC fails, any virtual machine in the cluster can take over the SC role. This provides the robustness required by operating the VNF in a cloud environment. This also allows the VNF to operate autonomously, eliminating the need for manual clustering-related configuration.

With the Roaming SC scheme, all SCs have one of the following states:

- **ACTIVE:** The **ACTIVE** SC is processing O&M traffic for the VNF. The SC function runs on a separate core from the PL functions.
- **STANDBY (sb):** The **STANDBY** SC functions only as a PL VM, but it becomes the **ACTIVE** SC in case of the failure of the current **ACTIVE** SC.



- **QUIESCED (q):** All VMs that are not in **STANDBY** or **ACTIVE** SC state are in **QUIESCED** state. They function as PL VMs, and are ready to take either the **ACTIVE** or **STANDBY** SC state when needed.

The SC state is indicated in the `scState` attribute of the *MrflInstance* MO as follows:

SC_ACTIVE	This VM is the active SC.
SC_STANDBY	This VM is the standby SC.
SC QUIESCED	This VM is in QUIESCED state.

Assignment of the different SC states is done according to the following:

- **ACTIVE:** After a cluster restart or the simultaneous failure of the **ACTIVE** and **STANDBY** SC VMs, the **ACTIVE** SC is selected with a leader election algorithm.
- **STANDBY:** When the **STANDBY** SC fails or becomes **ACTIVE**, a new **STANDBY** SC is selected randomly from the **QUIESCED** SC VMs in the VNF.
- **QUIESCED:** All SCs that are not **ACTIVE** or **STANDBY** are in **QUIESCED** state.

When a VM takes the **ACTIVE** SC role, it automatically inherits the O&M IP address of the former **ACTIVE** SC. This ensures a seamless and transparent switchover of the O&M end-point.

To prevent two SCs operating at the same time, the new SC VM starts processing O&M traffic only after a delay period: about 5 seconds after cluster startup, or 6.5 seconds after the failure of the **ACTIVE** SC VM.



2 Audio Codec Selection Principles

For the incoming SDP offer the highest priority audio codec received in the SDP offer that is supported by the vMRF is selected for the media stream. In addition, if the SDP offer included also a telephone-event corresponding to the RTP clock rate of the selected audio codec, then the telephone-event is also selected for the media stream. The selected audio codec, and possibly the selected telephone-event, is sent back in the SDP answer.

For outgoing SDP offer the following audio codecs are sent in the SDP offer in this priority order: EVS, AMR-WB, G.722, AMR, G.729, PCMA, and PCMU. The highest priority audio codec, and possibly the telephone-event corresponding to the RTP clock rate of that audio codec, received in the SDP answer is used for the media stream.



3 Audio Conferencing

The audio conferencing feature provides a service that allows the connection of as many as 30 participants to a conference call. For each participant, it can be specified whether the connection is unidirectional (listening only) or bidirectional. The audio conferencing feature can dynamically mix audio from several participants. The audio from a participant is mixed to the audio conference based on active speaker logic.

Each client can have its own codec in use, and the Audio Conferencing feature provides transcoding between them. Audio mixing is performed at 16 kHz sampling rate. This guarantees wideband conferencing experience for wideband-capable terminals.

The following audio codecs are supported by vMRF: PCM, EVS, G.722, G.729, AMR-NB, AMR-WB.



4 Audio Announcements

Audio announcements feature is used to provide audio announcements and to facilitate interactive services for operators and network providers. vMRF supports audio announcements through streaming of audio content from the announcement files to subscribers. Audio announcements are supported in HD quality. The sample rate of the audio announcements is 16 kHz and the audio mode is Mono. Audio announcements have to be encoded in 16 bit 16 kHz PCM and stored in Waveform Audio File Format (WAV). vMRF plays an announcement towards a subscriber when a play request is received from the controlling node over H.248. Announcement files are stored on a storage server, and referenced by an announcement ID or variable announcement type. The recently played announcement files are cached to the local disk of a vMRF VM to support faster playing of announcements for future playing requests.

vMRF supports fixed announcements and segmented announcements. A fixed announcement consists of a single basic announcement. A segmented announcement is a compilation of individual announcements referred as announcement segments. In vMRF, an announcement segment can be a basic announcement or a stand-alone variable announcement. A basic announcement is a standalone announcement file that contains audio data of a spoken word or phrase. A variable announcement is an announcement whose spoken information heard by the user depends on input data received in a play request. vMRF supports the following variable announcement types: digits, date, time, number, and money. The variable announcement has a variable announcement logic file associated to each variable announcement type.

vMRF supports announcements with multiple languages. Multi-language announcements present the operator with an option to play announcements toward end users in their preferred language. Multi-language announcements are supported for both fixed announcements and segmented announcements. Multi-language announcements are identified by a single announcement ID or variable announcement type, and different languages are identified by using language tags. This enables playing of announcements toward end users in their preferred language, and simplifies the announcement configuration in vMRF and in the controlling node.

For configuration information on audio announcements refer to *vMRF Configuration Management*.

5 Detection of Dual Tone Multi Frequency (DTMF) Tones

DTMF digits can be detected either as inband signals from an audio stream or as separate telephone events. In case of a telephone event, the DTMF digits are detected from the RTP packets that use *telephone-event* payload format with dynamically established RTP payload type number. DTMF detection is done either for inband signals or for telephone event, not for both types simultaneously. If telephone event payload type has been negotiated for the audio stream, then vMRF detects DTMF telephone events, otherwise audio inband signals.

DTMF digits are specified in a dialing plan, called a digit map. The digit map resides in the vMRF and it is used to detect and report certain DTMF patterns received from the user plane. A pattern is a set of rules which express how the user is expected to dial DTMF digits, including, for example, digit names and timer values for the detection process. Up to 16 different DTMF digits can be used, corresponding to the following symbols: 0, 1, 2...9, A, B, C, D, *, #. The digit map is preloaded dynamically for each session by the controlling node over H.248. vMRF detects as many DTMF digits as it is contained in the digit map. If all digits match, vMRF notifies the controlling node that DTMF digits were collected successfully. The notification contains the digits that were collected and number of attempts that were made to detect the digits.



6 Forwarding of DTMF Tones through Audio Conference

vMRF supports forwarding of DTMF digits through audio conference from one participant to all other participants.

The function supports detection of either telephone-event DTMF or audio DTMF from the participant, based on SDP negotiation. While detecting audio DTMF digits, the received audio DTMF digits are filtered out from the stream before mixing audio by the audio mixer. The detected DTMF digits are sent to all other participants either as telephone-event DTMF or audio DTMF, based on SDP negotiation with each participant. The context topology and termination stream mode information is checked and obeyed before forwarding DTMF digits from one participant to another participant. The function can be enabled by configuration.



7 Playing Tones

vMRF supports playing of predefined tones to the user plane based on request from the MTAS. For supported tones, see [Tone Sender Service](#) on page 19.



8 Emergency and Priority Call Handling

vMRF supports emergency and priority call handling for the following call types:

- Emergency calls
- Priority calls
- International Emergency Preference Scheme (IEPS) priority calls

A call is considered an emergency or priority call if it is indicated so in the call setup. Emergency and priority calls are treated the same way, and they are prioritized over non-emergency and non-priority calls in the case of high load on the vMRF.

A certain amount of processing capacity and internal resources are reserved by the vMRF for emergency and priority calls. This reserved fraction is called a priority pool. Normal calls are not allowed to use the priority pool. Emergency and priority calls are allowed to use resources outside of the priority pool.

The size of the priority pool can be configured in the vMRF. For more information on priority pool, refer to *vMRF Configuration Management*.

For emergency and priority call handling the following rules apply:

- Emergency and priority calls are never rejected because of network level admission control or high CPU load
- Emergency and priority calls can be rejected because of lack of internal resources

9 Overload Protection

To detect an overload situation in the network caused by high network traffic, the vMRF monitors the following:

- vMRF Application processor load (H.248 signaling, control of media stream processing)
- vMRF Media Stream Processing load (computationally demanding media processing, for example, transcoding)
- vMRF IP Pipeline load (simple media processing, for example, IP packet Header processing)
- vMRF IP Pipeline vSwitch load (outgoing packets are dropped if the vSwitch cannot process more packets)

In case of an overload situation, the following alarms are raised based on available processing capacity:

- When 80% of the available processing capacity is used in the vMRF functions, the *MRF Instance 80% Capacity Limit Exceeded* alarm is raised. It indicates that the vMRF is approaching an overload situation and new virtual machines must be created.
- When 100% of the available processing capacity is used in the vMRF functions, the *MRF Instance Overloaded* alarm is raised. In this case only emergency and priority calls are accepted. New virtual machines must be created to increase processing capacity. When a call is rejected because of overload, the H.248 ErrorCode 510 "Insufficient Resources" is reported.



10 Shared Storage

Shared storage is an optional feature of the vMRF, allowing the storage of log files and crash dump files on a remote server. The remote server and the exported files are managed by the operator.

The following files are stored:

- authentication log files
- system log files
- vMRF log files
- crash dump files

vMRF connects to the server with SSHFS (Secure Shell Filesystem), mounts a specified directory path, and creates a subfolder for the cluster, and subfolders for the files for each VM in the cluster. For authentication, an SSH key pair has to be created. This key pair can be cluster-specific, or common for all clusters. The prerequisites for the remote server are the following:

- ssh connection support with public and private keys
- authentication of the ssh connections from the vMRF virtual machines
- defined username for the SSHFS mount
- ssh public key for the SSHFS mount
- the public key is appended to the authorized keys file using the following command:

```
cat .ssh/<public_key_file_name> >> .ssh/  
<authorized_keys_file_name>
```

- path for the files

Configuration of the shared storage in the vMRF is performed during instantiation by filling in the relevant HOT template parameters. The following parameters must be defined:

- username for the shared storage SSHFS mount
- ssh private key for the SSHFS mount
- IP address of the shared storage server



- server port of the shared storage server
- mount directory path
- remote server ssh host key fingerprint

For more information on the HOT template configuration, refer to *Deployment Guide for OpenStack*.



11 Address Resolution and Fetching the Next Hop Address

IPv4 addresses are 32 bits long. In written format, the address is divided into four 8-bit (1 byte) long sections, using dots as separators. For example, 172.16.254.1 (decimal digits).

IPv6 addresses are 128 bits long. In written format, the address is divided into eight 16-bit long sections (pairs of bytes), using colons as separators. For example, 2001:db8:ffff:1:201:2ff:fe03:405 (hexadecimal digits).

For alternative writing formats of IPv6 addresses, refer to [RFC 4291](#).

Local IP addresses for all networks are fetched from a DHCP server.

IPv4 and IPv6 use the following different methods for address resolution and next hop supervision:

- Address Resolution Protocol (ARP): It maps an IPv4 address to the corresponding MAC address.
- Neighbor Discovery (ND): It maps an IPv6 address to the corresponding MAC address.

The following subfunctions are supported by both the ARP and the ND:

- Next hop MAC address fetching
- Address collision detection

A DHCP server is used for fetching the next hop addresses for the following:

- IPv4 media networks
- O&M networks
- Signaling networks

Internet Control Message Protocol for IPv6 (ICMPv6) is used for fetching the next hop addresses.

Note: Load sharing towards multiple next hops is not supported. Only one next hop per static route must be configured.

12 Media Stream Processing Services

12.1 Adaptive Multi-Rate (AMR) Speech Coder Service

The Adaptive Multi-Rate (AMR) Speech Coder is a high-quality narrowband codec standardized by 3GPP. AMR is standardized as the mandatory speech coder for the LTE radio network; it is also used in WCDMA and GSM networks, and in Voice over IP solutions. The speech coder handles the conversion between PCM coded speech and AMR coded speech. Discontinuous Transmission (DTX) is supported in both uplink and downlink. DTX allows the radio transmitter to be switched off most of the time during speech pauses. This saves power in the UE and reduces interference over the air interface.

The AMR speech coder operates in different modes, representing source rates from 4.75 kbps to 12.2 kbps. The selected rate is a trade-off between speech quality and robustness. AMR complies with [3GPP TS 26.071](#).

12.2 Adaptive Multi-Rate Wideband (AMR-WB) Speech Coder Service

The Adaptive Multi-Rate Wideband (AMR-WB) codec is used to provide better speech quality than the existing Adaptive Multi-Rate Narrowband codec. AMR-WB provides better speech quality because its analog spectrum range is 100–7000 Hz instead of the 300–3400 Hz of narrowband codecs. The sampling rate for AMR-WB is 16 kHz instead of the 8 kHz of narrowband codecs. The prerequisite for better speech quality is that the speech is carried end-to-end in AMR-WB coded form between the two UEs. HD speech quality is lost if AMR-WB speech is converted to a narrowband codec. This is the case, for example, in calls between an AMR-WB mobile and a PSTN phone.

The AMR-WB speech codec consists of the following components:

- Multi-rate speech codec
- Source-controlled rate scheme, including voice activity detector
- Comfort noise generation system
- Error concealment mechanism (to combat the effects of transmission errors and lost packets)

The AMR-WB codec has the following nine source rates: 6.60, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05, and 23.85. These AMR-WB rates correspond to SDP mode-set values from 0 to 8 (no mode-set on SDP means



the same as all rates supported). The codec is able to switch its bit rate every 20 ms. Refer to [3GPP TS 26.171](#).

12.3 Announcement Service

The Announcement Service plays an announcement towards a subscriber. The announcement is read from the announcement cache, where the announcement was fetched from the announcement storage when a play request was received. The Announcement Service supports playing of announcements that are encoded in 16 bit 16 kHz PCM and stored in Waveform Audio File Format (WAV).

12.4 Audio Mixing Service

The Audio Mixing service is responsible for audio mixing during audio conference.

The Audio Mixing service has an active speaker logic that determines which clients are included in the audio mix. This is based on voice activity detection. The detection algorithms calculate frame power, detect pitch, and estimate tonal activity to determine whether a client is speaking or not. Active speaker logic also ensures that the client has a long enough speech period before it is included in the audio mix, and a hangover period before the included client is dropped from the audio mix.

Audio mixing is performed at 16 kHz sampling rate.

12.5 DTMF Receiver Service (DTMF-R)

The DTMF-R detects DTMF digits from the user plane. Both Audio DTMF and Telephone-Event DTMF are supported. The DTMF-R can detect up to 16 different DTMF digits, corresponding to 0–9, A, B, C, D, *, and #.

For Audio DTMF, the DTMF-R detects each signaling tone, validating a correct tone pair and checking the timing. In the presence of speech, additional tests are performed to avoid interpreting a speech waveform as a valid signal tone. DTMF-R can be configured to detect overdriven DTMF tones. Without this setting, these tones would be dropped, because the harmonics produced by the low-frequency DTMF tones that fall within the passband of the high-frequency DTMF tones can be considered an indication of speech. For Telephone-Event DTMF, the digits are interpreted from RTP packets according to [RFC 4733](#).

12.6 DTMF Sender Service (DTMF-S)

The DTMF Sender (DTMF-S) sends DTMF digits to the user plane.

Both Audio DTMF and Telephone-Event DTMF are supported.

The DTMF-S can generate up to 16 different DTMF signals, corresponding to 0–9, A, B, C, D, *, and #. Audio DTMF consists of two simultaneously played frequencies. Telephone-Event DTMF is encoded to RTP according to [RFC 4733](#).

12.7 Enhanced Voice Services (EVS)

Enhanced Voice Services (EVS) is a multi-rate audio codec that operates at 8 kHz, 16 kHz, 32 kHz, and 48 kHz sampling rates, and offers full audio bandwidth ranging from 20 Hz up to 20 kHz. EVS supports bit rates from 5.9 kbps to 128 kbps. EVS supports comfort noise generation and error concealment.

The media can be encoded with EVS or with EVS AMR-WB IO mode. The frame size is 20 ms and the RTP timestamp is always generated with 16kHz clock. 20 ms and 40 ms packetization times are supported for EVS.

Note: The use of the EVS codec requires the Enhanced Voice Services capacity license and connection to a Network License Server (NeLS).

12.8 G.722 Service (G.722)

The G.722 is an audio codec that operates at a sampling rate of 16 kHz and offers audio bandwidth ranging from 50 Hz up to 7 kHz. The coding system uses Sub-Band Adaptive Differential Pulse Code Modulation (SB-ADPCM) within a bit rate of 64 kbps. In the SB-ADPCM technique used, the frequency band is split into two sub-bands (higher and lower) and the signals in each sub-band are encoded using ADPCM. The audio coder complies with [ITU-T G.722](#). The G.722 codec supports Packet Loss Concealment (PLC).

12.9 G.729 Service (G.729)

The G.729 high compression codec enables packet voice transmission using G.729 towards external IP-based networks. vMRF supports the G.729 speech codec according to [ITU-T G.729](#), with the following variants:

- G.729A, reduced complexity 8 kbps Conjugate Structure Algebraic Code Excited Linear Prediction (CS-ACELP)
- G.729AB, meaning G.729A with built-in silence suppression

vMRF supports G.729A with packet loss concealment. A packet loss concealment mechanism is supported in order to minimize the distortion in output speech caused by lost or excessively late speech packets. The G.729A is bit stream interoperable with the full version of G.729.

vMRF supports G.729AB, which adds on silence suppression including voice activity detection, generation and reception of SID (Silence Insertion)



Descriptor) packets, as well as comfort noise generation. These algorithms are used to reduce the transmission rate during silence periods of speech. The G.729AB (defining reduced complexity version of the G.729 speech codec with built-in silence suppression) is bit stream interoperable with the full version of G.729 with silence suppression.

12.10 Jitter Handling Service

Jitter (delay variation between data packets) is removed or decreased using jitter compensation performed by the Jitter Handling service. The Jitter Handling service supports adaptive jitter service.

In the beginning of the call the jitter buffer size is always the configured initial jitter buffer size, but during the call the jitter buffer size adapts to the measured jitter.

12.11 Pulse Code Modulation Service (PCM)

The Pulse Code Modulation (PCM) service handles coding and decoding of PCM ([ITU-T G.711](#)) samples in vMRF. Both A-law and μ -law formats are supported. PLC is supported for the PCM service.

12.12 Real-Time Transport Protocol and Real-Time Transport Control Protocol Service (RTP/RTCP)

The RTP/RTCP service represents the RTP and RTCP protocol termination for IP user plane traffic.

RTCP can be used to monitor the aliveness of the session as well as quality of service during an ongoing RTP session, according to [RFC 3550](#).

12.13 Tone Sender Service

The Tone Sender service supports the generation of the following tones:

- H.248.1 E.7 Call Progress Tone Generator Package
 - Special Information Tone
 - Ringing Tone
- H.248.27 Conferencing Tones Generation Package
 - Conference Enter Tone
 - Conference Exit Tone



The tone can be added on top of an established connection or it can temporarily interrupt the speech.

12.14 User Plane Frame Handler Service (UP FH)

The User Plane Frame Handler (UP FH) service provides RTP payload framing according to the corresponding RTP profile.



Reference List

- [1] *RTP Payload for DTMF Digits, Telephony Tones, and Telephony Signals*, [IETF RFC 4733](#)
- [2] *RTP: A Transport Protocol for Real-Time Applications*, [IETF RFC 3550](#)
- [3] *G.729: Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)*, [ITU-T G.729](#)
- [4] *G.711: Pulse code modulation (PCM) of voice frequencies*, [ITU-T G.711](#)
- [5] *7 kHz audio-coding within 64 kbit/s*, [ITU-T G.722](#)