5G Voice over New Radio (VoNR)

White Paper | Version 01.00 | Reiner Stuhlfauth

ROHDE&SCHWARZ

Make ideas real

CONTENTS

1 INTRODUCTION

The main driver behind the development of 5G networks has always been improved data services: higher mobile data rates and reduced latency, etc. However, the legacy services of voice and video communications remain key elements for mobile services, and subscribers continue to demand them.

It is this demand for voice services from mobile subscribers that keeps it part of services providers' packages and business models. However, voice over 5G networks is not just about keeping customers happy; voice also has a role to play in the new data services provided by 5G. According to analysts, the number of voice subscriptions worldwide will double until 2025 [Ref. 21].

While it may seem simple to just keep the voice services that you already have in place, it actually presents some thorny technological challenges to accomplish this feat. In fact, a single technical solution is not possible because whatever solution is adopted needs to adapt to the existing network deployment.

Prior to this latest migration of voice over 5G networks (5G Voice over New Radio (5G VoNR)), the migration of voice services from 3G and circuit switched mobile networks to the 4G Long-Term Evolution (LTE) network was initially slowed by technical challenges. Part of the ability of Voice over LTE (VoLTE) to overcome its early technical challenges was the fact that it was based on the IP Multimedia Subsystem (IMS) architecture. In a sense, you can track the adoption of IMS by service providers to the rise of VoLTE.

This IMS architecture will play an increasingly important role in 5G VoNR. Like 4G LTE networks, 5G voice calls are implemented as end-to-end Voice over IP (VoIP) connections managed by the IMS core. Voice and video communication services in these networks ride on top of the IP data connection. Unlike voice services provided by external applications (i.e. so-called OTT speech services), voice over IMS supports Quality of Service (QoS) management across the entire 5G System (5GS).

While IMS can provide voice services for any type of access (fixed, cable, and 2G/3G) as well as for any 5G-deployment model, 5G is not as flexible and absolutely has to have an IMS network to handle voice services no matter the type of deployment. Just as we could track the adoption of IMS to the rise of VoLTE, the introduction of 5G serves as a catalyst to accelerate voice core modernization to IMS from older technologies in networks.

As we said previously, a single technical approach to 5G VoNR is not possible. This white paper will go into detail for all the approaches that can be taken. In general, however, VoNR has to adapt to the existing deployment modes: Non-Standalone (NSA) or Standalone (SA).

The NSA deployment mode, known as option 3, involves LTE plus NR with an Evolved Packet Core (EPC). In contrast, the SA deployment involves NR with 5G Core (5GC). This deployment mode is known as option 2.

Once it has been determined whether the RAT is 5G NR or EUTRA, it's necessary to know if these are co-existing SA networks or in a dual connectivity scenario: E-UTRAN New Radio–Dual Connectivity (EN-DC) or NR-E-UTRA Dual Connectivity (NE-DC). Of course, you also need to know whether the RAT even supports voice services.

Most recent 5G deployments have employed what's known as the option 3. This approach essentially means that the network providers have an existing 4G LTE network and they have deployed a 5G network along side it. In this way, the 5G NR serves as a secondary cell, and the core technology remains Evolved Packet Core (EPC).

In the operation of option 3, the UE registers to the IMS over Evolved Packet System (EPS). When the 5G UE launches (or receives) a voice call, this follows typical VoLTE procedures over the EPS system.

When a service provider can choose option 2, they deploy the 5G network as a SA network, without relying on any other network. In this option, the IMS core provides voice as a 5G-application service. Voice services on this kind of network are referred to as Voice over New Radio (VoNR).

Even this option 2 presents challenges. In these early years of 5G, geographic coverage for 5G will be incomplete. As a result, when a mobile device moves out of a 5G NR coverage area, a VoNR voice call in progress would need to handover to use the VoLTE in a 4G network. Meticulous network planning and coverage considerations is needed.

In the previous VoLTE systems, circuit-switched fallback (CSFB) led to delays causing an interruption to the call. A procedure that will be explained in detail in this whitepaper is the "EPS fallback for IMS voice". This procedure avoids these dropped calls by directing the UE to fall back to the EPS at the start of any voice call as soon as it is launched. Basically, the EPS fallback performs the handover when the call is first setup instead of during the call and thereby does not impact the user's in-call experience. With EPS fallback, the UE will camp on the higher prioritized 5G RAT. There is no need for full support of 5GC yet, thus EPS is a kind of interim step speeding up the time to market approach for voice services.

The benefit of 5G VoNR for voice-only calls is surely the quality and ultra-high definition of the calls. But, as we pointed out earlier, and perhaps more importantly, 5G VoNR also has an important role to play in the new data services provided by 5G.

5G VoNR provides a point of integration with applications and content such as announcements, music, conferencing and more. It will also provide enhanced support for real-time communications including Rich Communications Services (RCS) integration. An example of this RCS integration, we can envision the 5G VoNR will enable interactive capabilities, such as real-time language translations. Many of the more advanced functions will work only in a 5GNR environment with the support of 5GC infrastructure.

This White Paper will provide in-depth guidance in navigating 5G VoNR from the network deployment and the connectivity options aimed at not only supporting voice over NR, but also ensuring the data services that VoNR should enable. At the end of the White Paper, a brief summary is provided on the signaling parameters for both the network and UE side. A general introduction into 5G fundamentals, procedures, system and service aspects and technology drivers can be found in Ref. 20.

2 VOICE CALL ASPECTS IN 5G NR

Although data services in the context of enhanced mobile broadband (eMBB), ultra-reliable and low latency communications (URLLC) and massive machine-type communications (mMTC) are the pivotal drivers behind the 5G evolution, legacy services like voice and video communications still represent important services that operators want to offer to their subscribers. As part of the technology evolution, we have seen a major change from circuit-switched 2G networks with an initial focus on telephony to fully packet-switched 4G networks focused on Internet data communications.

This white paper elaborates the technological details on how a 5G network may support voice services. Unfortunately, a single technical solution in the 5G system will not be offered for voice services including the radio access technology, infrastructure deployment and protocol layers involved. The objective here is to describe those technical evolution steps intended to support voice services to guarantee that those services will not be curtailed due to the introduction of 5G.

From a high-level perspective, voice over NR needs to adapt to the existing deployment, whether it is NSA or SA mode and whether the EPC is the core network or the 5GC. The type of voice support in 5G depends on the available Radio Access Technology (RAT) in the form of 5G NR or EUTRA and whether both are used either as coexisting standalone networks or in a dual connectivity scenario EN-DC or NE-DC, and obviously which RAT supports voice services. The second question influencing the type of voice service is the availability of a core network (either EPC or 5GC) and again, obviously, which core network supports voice services.

The objective of this white paper is to present the various voice services in 5G in greater detail along with some evolution paths illustrating how the offered voice services may change with the evolution of the 5G System (5GS), 5G Access Network (5G-AN) and User Equipment (UE) deployment, and lastly, some supplementary services that are supported such as the emergency service, SMS or the automotive emergency service eCall.

From a high-level perspective, voice over NR is voice over IP using the IP Multimedia Subsystem (IMS) infrastructure. The advantage of using IMS is the ability to set up and guarantee a Quality of Service (QoS) for each application. The task for IMS is to establish, control and maintain a Protocol Data Unit (PDU) session, including all relevant data bearers with corresponding QoS flow for best end-user quality experience. One difference compared to a data PDU session in 5G is that with the Non-Access Stratum (NAS) signaling procedure PDU session establishment request, the UE requests a PDU session for IMS signalling. Like in Voice over Long-Term Evolution (VoLTE), voice over IMS in 5GS also supports QoS as a major difference compared to voice services offered by external applications, e.g. so-called over-the-top (OTT) speech services.

Thus, one question arises concerning how to connect IMS to the next generation network, 5GC. The evolutionary paths describe whether in an NSA connection, voice will be supported by EUTRA only and if the simultaneous NR data connection can be sustained or will be suspended. This option is referred to as **voice over LTE (VoLTE) in EN-DC** setup.

The **Evolved Packet System (EPS) fallback** use case describes the situation where 5GC does not offer voice services; if necessary, the connection will be handed over to an EPS connection (VoLTE). Another fallback mode is **RAT fallback**, assuming the current core network supports voice service but the current RAT, presumably NR, does not. In that case, the connection is transferred from NR to EUTRA only.

Voice over NR (VoNR) indicates the case where the NR network does support voice services and the 5GC offers a connection to IMS. In general, VoNR is independent of the dual connectivity and thus, it would work with EN-DC, but the focus is on the NR SA where 5GC connects to IMS supporting voice services. Since LTE is considered here to operate in parallel, inter-system handovers are mandatory to guarantee UE mobility and avoid call drops and leverage high-quality Key Performance Indicators (KPIs). Note that in 3GPP Release 15, there is no handover defined between 5G and 3G/2G. Thus, no Circuit-Switched Fallback (CSFB) scenario is possible either. A handover to circuit-switched 2G/3G is only possible in two steps via an interim connection to 4G. In 3GPP Release 16, a Single-Radio Voice Call Continuity (SRVCC) is introduced where a VoNR connection can be transferred to 3G. The objective is to avoid call drops when coverage of 5G services is getting weaker and LTE coverage is not available. See Fig. 1 for relevant setups.

Besides the terminal capabilities, support for voice services in 5G NR needs to consider the various network deployment options. Critical questions are e.g. which RAT to use, EUTRA or NR, which core network is available, EPC or 5GC, and whether the Evolved NodeB (eNB) is an next generation evolved NodeB (NG-eNB) or just a legacy eNB. Consequently, we may speak about EPS fallback, RAT fallback, voice over NG-eNB (NGEN-DC) or standalone VoNR. For sure, the frequency band allocation accompanies such voice call deployment options, i.e. network operators plan to re-farm legacy frequency bands in the lower frequencies from LTE to 5G. With such enhanced coverage, services like VoNR also become feasible.

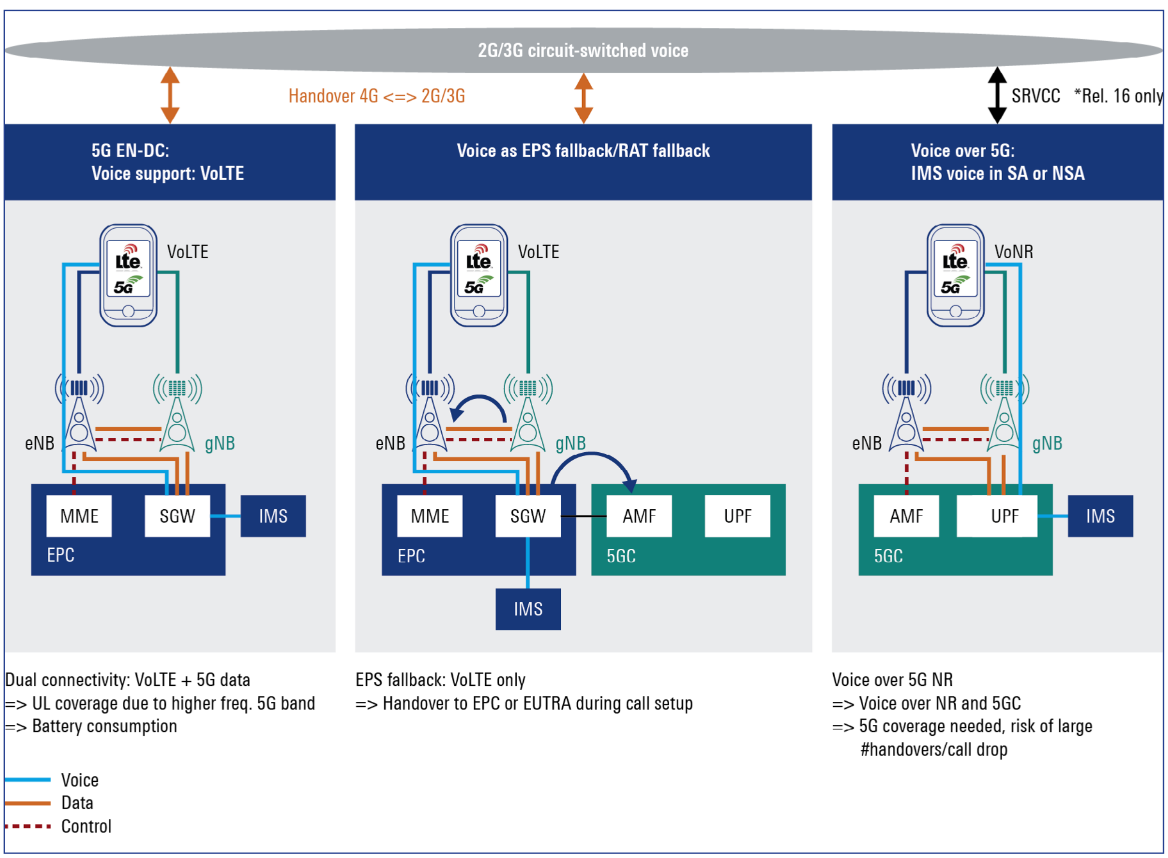

Figure 1: Deployment scenarios supporting voice in 5G.

Besides the technology evolution from 4G to 5G on both RAT and the core network, especially for voice services we see an evolution towards the introduction of more and more sophisticated and high-quality voice and video services. One term that accompanies VoNR is the enhanced voice service (EVS) with wider audio bandwidth, higher sampling ratio, better quantization and higher resolution. The EVS speech coder has been introduced with LTE in several networks already, but 5G voice services rely on this advanced speech coding algorithm more extensively. A concise introduction to EVS with the corresponding speech codec principles is given as well.

Due to the N3IWF interworking, there is even an option to establish a voice call over IMS on a non-3GPP access network (e.g. WLAN) and handover to VoNR. For the sake of brevity, the details of this procedure are omitted here, but they can be found in TS 23.502.

A popular derivate of emergency services in the automotive sector known as eCall may operate on LTE as a next-generation eCall (NGeCall), but there is also ongoing research into how to enable future-proof emergency call services over 5G NR.

2.1 Voice services in an NSA dual connectivity setup

Recent deployments of 5G reveal widespread usage of option 3 as one option in the connectivity scenario background in the infrastructure evolution. As a result, EN-DC, where there are two radio links, EUTRA and 5G are offered. To accelerate the introduction of voice services, the major aspect of this method is that NR does not support voice services at all since this will be covered by LTE. The criterion in EN-DC mode is that NR cannot exist standalone; thus, LTE coverage is the prerequisite. Operators who have deployed VoLTE for 4G can continue to use IMS services over EUTRAN (partially or fully upgraded to EN-DC). There is no impact on IMS. In fact, the IMS system is not even aware of the EUTRAN upgrade to EN-DC. All location information exposed to IMS is based on LTE, as before.

All VoLTE principles remain applicable. A PDN connection to the IMS (Access Point Name (APN) is used for Session Initiation Protocol (SIP) signaling (on the default bearer with QoS Class Identifier (QCI)=5) and for voice media (on a dedicated bearer with QCI=1. [Ref. 1]

The prerequisite for voice services is support for IMS, which is already the case in EN-DC since the EPC is connected to IMS already. A possible recommendation includes setup of an IMS Packet Data Network (PDN) connection for Session Initiation Protocol (SIP) signaling messages inheriting the LTE QoS profile QCI = 5 and a parallel PDN connection for the voice content inheriting QCI = 1 as this is already recommended for VoLTE. Since the setup supports dual connectivity, a third PDN default bearer with suggested QCI = 9 can now operate as a split bearer to provide additional packet data over EUTRA and NR radio interfaces. It would be up to the network implementation and UE maturity whether the NR link is suspended or maintained simultaneously to the EUTRA bearer. This proposal presents an early implementation and support for voice services when deploying 5G since it reuses investments operators made into their LTE network. Bit rate and latency requirements are well satisfied over LTE. In a longer-term rollout perspective, EN-DC option 3 may present an interim stage that extends the existing EPS by a concurrent 5GC deployment and then allows the support for VoNR.

The challenge as well as a minor drawback involves providing reliable voice services on early standalone NR deployments since a common scenario uses middle bands for 5G and lower bands for LTE, which may induce some coverage limitations. Another negative aspect is the higher energy or battery consumption since the UE needs to operate on two radio connections simultaneously. [Ref. 1] suggests RAN features such as RLC unacknowledged mode, Transmission Time Interval

(TTI) bundling, Robust Header Compression (RoHC) enabled, etc. that should increase the service KPI with optimized quality of experience.

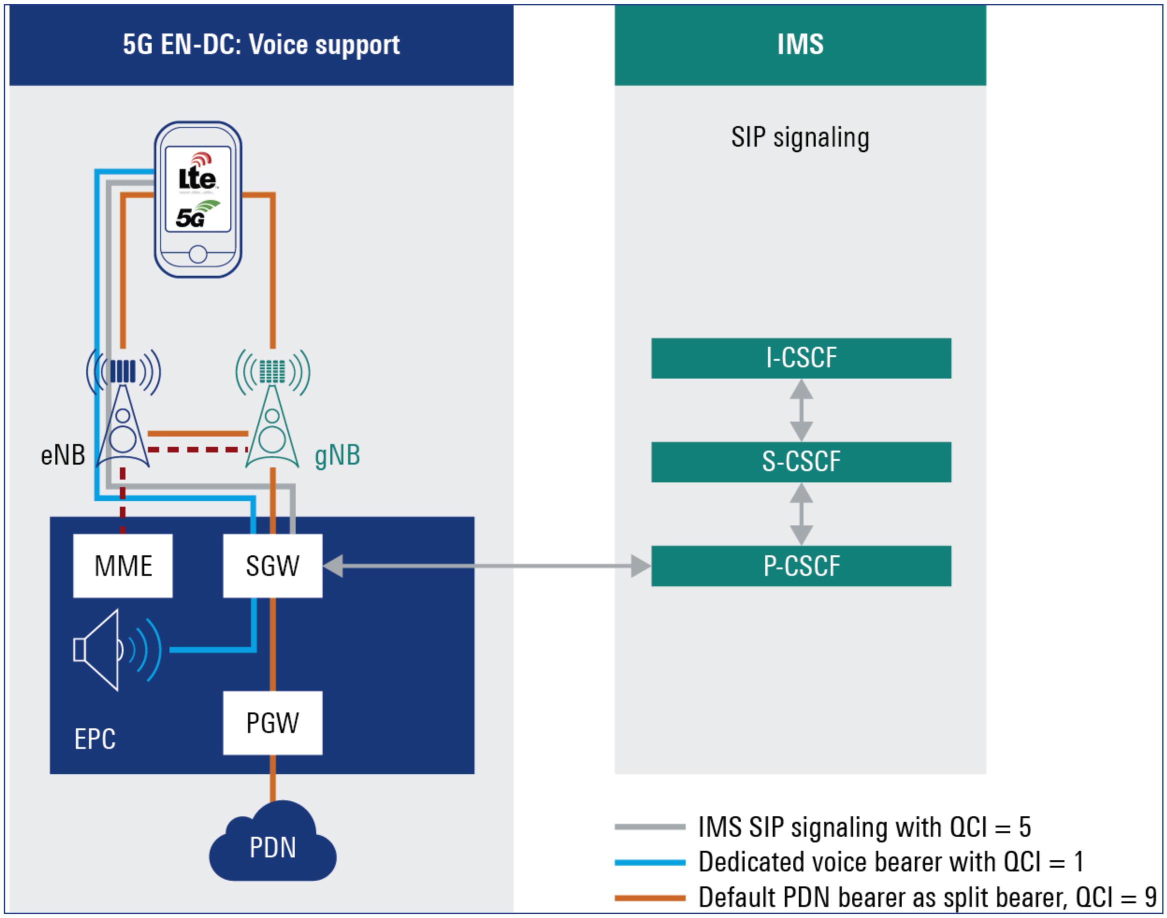

Figure 2: Deployment scenario EN-DC supporting VoLTE and Internet split bearer NR + LTE.

2.2 Voice services using EPS fallback

2.2.1 Network interworking requirements supporting voice services

Regardless of whether 5G deployment is based on NSA or SA mode, the lesson has been learned from legacy network introductions that initial deployments will not have full coverage. Consequently, 5GC needs to be tightly coupled to EPC, especially to an existing IMS VoLTE supporting infrastructure to provide seamless voice services across the entire network with acceptable Quality of Experience (QoE) as KPIs. The objective is to register the UE primarily in the 5G network, even if voice services are not supported and a handover to LTE would be required. This method is RAT agnostic and the IMS support for voice services is available via the EPC. A UE camping on NR will then be redirected to EPC as core and the serving RAT may change from NR to EUTRA as well. Note that a fallback procedure where only the RAT is changed from NR to EUTRA is called voice services with RAT fallback. This is further described in section 2.2.3. To leverage this motivation of tight coupling between EPC and 5GC, the objective is first to introduce an additional interface between the core network entities and functions. Note that their implementation depends on the deployment strategy of operators and infrastructure vendors. A few of these new connections include:

- N6 interface to connect the 5GC function User Plane Function (UPF) to the IMS. TS 23.501 defines N6 generally as the reference point between the UPF and a data network. For voice support, this data network is now represented by the IMS signaling system. A PDU connection can be established with the required QoS flow for voice services. N6 is also the prerequisite for full VoNR support

- S5 interface for control and user plane connection between Session Management Function (SMF) / UPF representing the 5GC entities and the Serving Gateway (SGW) representing the EPC entity. From the EPC point of view, the S5-U interface replaces the Public Data Network Gateway (PGW) for voice services as they are now logically established via the UPF

- N26 interworking interface between Mobility Management Entity (MME) and Access & Mobility Management Function (AMF) to enable context transfer and network-controlled mobility scenarios like handover between LTE and 5G. N26 represents an optional network deployment, but as the main advantage, TS 23.501 indicates that interworking procedures with N26 provide IP address continuity on inter-system mobility to UEs that support 5GC Non-Access Stratum (NAS) and EPC NAS and that operate in single registration mode. Without the N26 interface and assuming a UE in single registration mode, the control coordination between MME and AMF needs to be routed via SMF and PGW

- Home Subscriber Server (HSS) interworking with Unified Data Management (UDM), providing a connection to IMS and supporting the querying of the serving AMF when requested by IMS (see TS 29.563 for further details)

- Policy Control Function (PCF) with Rx/N5 PCF: When an indication for a session arrives over the Rx/N5 interface and the UE does not have priority for the signaling QoS flow, the PCF derives the Allocation and Retention Priority (ARP) and 5G QoS Identifier (5QI) parameters, plus the associated QoS characteristics as appropriate, of the QoS flow for signaling as per service provider policy (see TS 23.503 for further details)

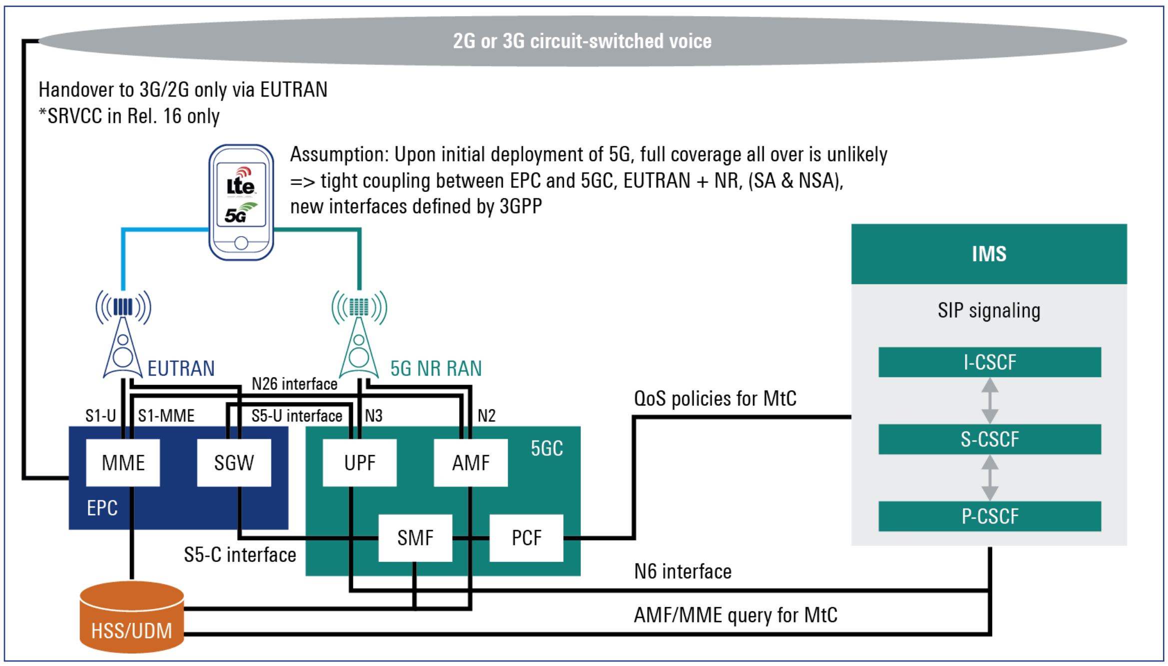

Figure 3: Reference points between EPC, 5GC and IMS to guarantee tight interworking for voice support.

Thanks to these interworking and reference points, one can guarantee an IP flow controlled via UPF and SMF independent of whether the UE is camping on LTE or NR. In addition, mobile terminated connections initiated by IMS can be routed properly and even in a single registration situation, due to the optional N26 interworking it is possible to maintain the IP address allocated to the UE.

2.2.2 Registration procedure – aspects for voice support

The first important procedure in a possible offering of voice services in 5GS is the registration procedure in which the UE and network exchange the intentions and capabilities of both entities. Further details on the registration procedure in the NAS message flow are available in TS 24.501. With the registration request message, the UE reveals its usage setting to the network. Usage settings can be data-centric or voice-centric and if the UE intends to use voice, the UE usage settingsshould be set on voice-centric mode. Another important flag in this control message is the S1 mode flag contained in the 5G Mobility Management (5GMM) capability element. With this flag, the UE indicates whether a possible EPS fallback procedure for IMS voice from 5GS to EPC is feasible or not. Via the capability information exchange, the UE discloses its IMS-related parameters. Common IMS parameters the UE may support are, for example, the indication of voice over EUTRA support, voice over Secondary Cell Group (SCG) bearer support and an indication of voice fallback into EPS. The parameter voiceOverNR indicates the UE support for voice over NR. If the UE does not support voice over NR but only EPS fallback, it is recommended to set the parameter flag as *voiceOverNR = False and voiceFallbackIndicationEPS-r16 = True*. Note that the UE capability is a procedure between UE and NG Radio Access Network (NG-RAN) (TS 38.331). The 5GC receives the registration request and needs to provide a proper response depending on the network capabilities or offerings. Obviously, one of them is the support for voice services by the network. With the registration accept message, the network does not only confirm the successful transfer into 5GMM_REGISTERED state but also confirms the support for voice call related features. It is worth mentioning in greater detail the information element of the registration accept message 5GS network feature support. This Information Element (IE) contains flags such as the support for "IMS voice over Packet Switched (PS) session supported over 3GPP access", "non-3GPP voice support", "emergency call services" or "emergency call fallback support" and whether the network interface N26 does or does not exist. In the following figure, the registration procedure is depicted with a focus on voice-specific control information only. Further details on the NAS registration procedure are available in TS 24.501.

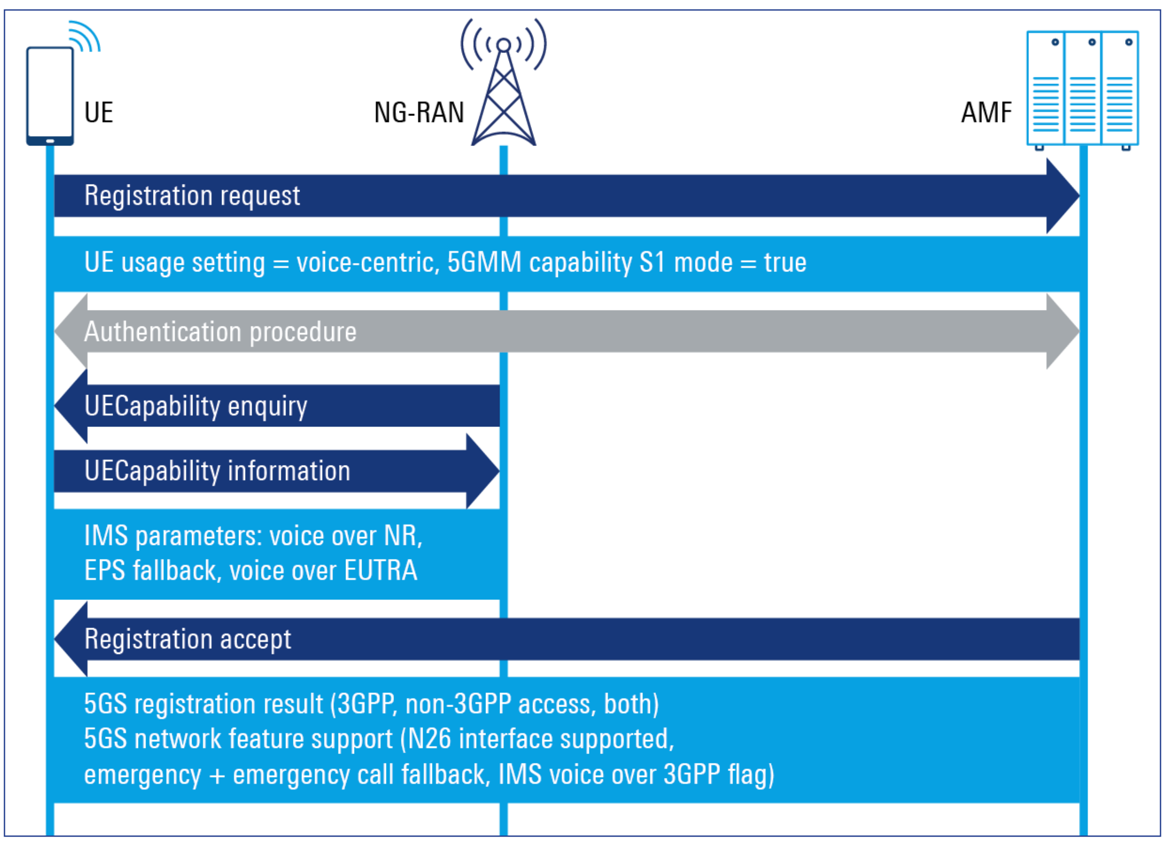

Figure 4: Registration Procedure with voice aspects.

TS 23.501 defines certain interworking scenarios between 5GS and EPC and provides more information on interfaces such as N26. S1 mode indicates a successful EPS attach and N1 mode indicates a successful 5GC attach. The definition in TS 24.501 is that in N1 mode, the UE has access to the 5G core network via the 5G access network. Single or dual registration indicates simultaneous handling of mobility states. In single registration mode, there is only one active mobility state at any given time. The UE stays in either 5GC NAS mode or EPC NAS mode. Concerning the UE identifiers (see Fig. 3-13), the UE maps the EPC Globally Unique Temporary Identify (EPC-GUTI) to 5G-GUTI during mobility between EPC and 5GC and as stated, if the network supports the N26 interface, the UE keeps 5G context such as IP address allocations for re-use when moving from 5GC to EPC. In order to interwork with EUTRAN connected to EPC, the UE supporting both S1 mode and N1 mode can operate in single-registration mode or dual-registration mode. The first mode (single-registration mode) is mandatory for UEs supporting both S1 mode and N1 mode. Dual-registration mode requires the UE to handle 5GMM and EMM context independently and simultaneously. In this mode, the UE maintains the identifiers 5G-GUTI and EPC-GUTI independently. The UE can perform 5GC or EPC re-registration/TAU using the corresponding GUTIs.

A UE operating in the dual-registration mode may register to N1 mode only, S1 mode only, or to both N1 mode and S1 mode.

During the EPS attach procedure or initial registration procedure, the mode for interworking is selected if the UE supports both S1 mode and N1 mode and the network supports interworking. Depending on the N26 interface support, the UE has different options to operate in single- or dual-registration mode and certain submodes. If both 5GMM and EPS Mobility Management (EMM) are enabled, a UE operating in single-registration mode should maintain one common registration for 5GMM and EMM. Coordination between 5GMM and EMM for a UE, which is capable of N1 mode and S1 mode and operates in dual-registration mode is not needed. TS 24.501 defines the details of the coordination between 5GMM and EMM in single-registration mode depending on whether the N26 interface is supported or not.

2.2.3 5G IMS support

The IP multimedia subsystem (IMS) represents connection management for voice services in 5G, like in legacy LTE networks. 3GPP Release 5 introduced IMS to evolve Universal Mobile Telecommunications System (UMTS) networks to deliver IP-based multimedia to mobile users. Readers not familiar with basic IMS aspects may consult [Ref. 6] for further information. IMS has become a core component within the 3G, cable TV and 4G networks, and refinements of technology aspects were realized by subsequent releases. Nevertheless, IMS is considered as being agnostic and independent from 5GS.

Initially, IMS was an all-IP system designed to help mobile operators to deliver next generation interactive and interoperable services, cost-effectively, over an architecture providing the flexibility of the Internet.

The session initiation protocol (SIP) was selected as the signaling mechanism for IMS, thereby allowing voice, text and multimedia services to traverse all connected networks. The 3GPP works closely with experts in the IETF to ensure maximum re-usability of Internet standards, preventing fragmentation of IMS standards. For further information on general IMS aspects, see TS 22.228 and TS 23.228.

Since voice services in 5G are not mandatory for the UE and network, the 3GPP agreed to a common implementation and indication policy to guarantee proper functioning of voice services when offered. The Global Supplier Association (GSMA) organization published a permanent reference document [Ref. 8] defining a profile that identifies a minimum mandatory set of features which are defined in 3GPP and GSMA specifications that a wireless device and network are required to implement in order to guarantee interoperable, high-quality IMS-based communications services for voice, video and messaging over a next generation (NG) radio access connected to 5GC.

A network offering voice services in 5GS needs to support IMS with the following functionality:

- Indication toward the UE if IMS voice over PS session is supported
- Capability to transport the Proxy Call Session Control Function (P-CSCF) address(es) to UE
- Paging policy differentiation for IMS
- IMS emergency service
- Domain selection for UE originating sessions
- Terminating domain selection for IMS voice
- Support for P-CSCF restoration procedure
- Network Repository Function (NRF)-based P-CSCF discovery
- NRF- or SCP-based HSS discovery

The serving Public Land Mobile Network (PLMN) AMF should send an indication toward the UE during the registration procedure over 3GPP access to indicate if an IMS voice over PS session is supported or not supported in 3GPP access and non-3GPP access. The UE's usage setting applies to voice-capable UEs in 5GS and indicates whether the UE has preference for voice services or data services. When a UE selects the usage setting as "voice-centric", this includes IMS voice and when the UE sets the usage setting as "data-centric", data services include any kind of user data transfer without a voice media component (TS 24.501).

A mobile device offering voice services in 5GS needs to support IMS capabilities that are required over the Gm and Ut reference points. Reference point Gm supports communications between the UE and IMS related to registration and session control. Reference point Ut facilitates the management of subscriber information related to services and settings [Ref. 9]. To guarantee proper operation of IMS-based services, 3GPP and GSMA recommend a set of eleven general IMS functions that should be supported; further details can be found in [Ref. 8]:

1.) SIP registration

The UE is required to perform a registration to IMS via the SIP protocol and certain aspects of such a registration procedure should be supported and fulfilled:

► Single SIP registration to the known IMS Access Point Name/Data Network Name (APN/DNN) for all IMS services. This is mandatory for all UEs and agnostic with respect to the home operator or roaming condition

► Two separate IMS registrations using different APNs/DNNs, each supporting a subset of the IMS services. An application could be an IMS registration to the IMS APN/DNN in concurrent registration to a Home Operator Service (HOS) APN/DNN to combine voice services with Rich Communications Services (RCS) services. This depends on the RCS VoLTE single registration parameter [Ref. 10] and is applicable if at least one of the RCS messaging services or Message Session Relay Protocol (MSRP)-based enriched calling services is enabled. If this parameter is set to zero, the UE handles two separate IMS registrations. If it is set to one, the UE uses single registration and if it is set to two, the UE uses single IMS registration if it is registered to the home network; otherwise it handles two IMS registrations

- P-CSCF discovery prior to initial IMS registration using protocol configuration option settings during the PDU session establishment

- Authentication for SIP registration depending on the related APN/DNN. Two authentication methods are used: IMS-Authentication & Key Agreement (IMS-AKA) and SIP digest [Ref. 8]

- IMS user and device identifiers in SIP registration. Depending on the APN/DNN registration and authentication, the UE allows the derivation of public and private IMS identities. Examples are the well-known International Mobile Subscriber Identity (IMSI) or International Mobile Equipment Identity (IMEI) and the IMS private user identity (IMPI) as part of the IP Multimedia Services Identity Module (ISIM) application on the Universal Integrated Circuit Card (UICC)

- Registration of IMS services. The UE must register multimedia telephony voice services (MMTel voice), short message services over IP (SMSoIP), video services (MMTel conversational video) and MMTEL call composer, always via the IMS well-known APN. Optionally via a second concurrent registration, the UE facilitates home network based complementary services like RCS messaging services and MSRP-based enriched calling services

- SIP registration procedure. The UE and IMS network must follow the SIP registration procedure defined in TS 24.229. To protect privacy, the UE must include a user part in the URI of the contact address such that the user part is globally unique and does not reveal any private information. Note that the UE can execute two separate IMS registration procedures to the default DNN for all IMS services and concurrently to a HOS DNN for RCS-specific services

- Registration to IMS with service-specific aspects. During the registration procedure, the UE reveals service-related information such as the IMS communications service identifier (ICSI) to indicate IMS multimedia telephony and media feature tags like “audio” or “video” if the UE registers for MMTEL services. Similar information disclosure is defined for services like SMS, call composer or RCS (see [Ref. 8])

2.) Authentication
To guarantee secure network access, IMS also requests device and subscriber authentication. The GSMA requirement in [Ref. 8] involves support for two authentication mechanisms (IMS-AKA as defined in TS 33.203 or SIP authentication via the Digest method as defined by GSMA [Ref. 10]) and provides further details and requirements for these authentication procedures. The latter authentication method includes the mandatory UE and optional network support for HTTP content server authentication. Integrity protection is mandatory for the network and UE, while confidentiality protection for SIP signaling is optional for the network depending on whether radio link layer security is enabled or not.

3.) Addressing
Proper addressing of user identities is an essential prerequisite for successful communications. The UE and IMS core network must support public user identities based on structures defined in TS 23.003 like alphanumeric Session Initiation Protocol-Uniform Resource Identifiers (SIP-URIs) and Mobile Station International Subscriber Directory Number (MSISDN) either represented as digit telephone URI or SIP URI. As a kind of simplified and user-friendly operation on the Graphic User Interface (GUI), TS 24.229 defines the possibility of local numbers used as geo-local numbers, home operator local numbers or other private local numbers. The UE and IMS must support local numbers and P-called-party-ID-header in the SIP header while global routable user agent URIs are not required [Ref. 8].

4.) Call establishment and termination
Concerning IMS call handling parameters, 5G requires more or less the same policies as in LTE; TS 24.229 defines the exact details. Optionally, the home operator can configure the UE with timers for round trip time estimation (T1), maximum retransmit time for INVITE responses (T2) or maximum message duration within the network (T4) (see TS 24.167). One slight difference is that the UE includes a reason information as cause when terminating a connection via CANCEL or BYE message and if available, the UE should insert the P-access-network-info [Ref. 8]. To leverage subscriber usability, some connection extension and modification services are required as well. For example, the UE and network should be able to add a video call to a voice session during session establishment via SDP message transfer and if available, the UE should support the services provided by MMTEL call composer and RCS messaging services as defined by GSMA [Ref. 10].

5.) SIP precondition considerations
TS 24.229 specifies a precondition mechanism for MMTEL voice conversational services that should be supported by the UE as requested by [Ref. 8]. The network may disable the use of preconditions via explicit signaling. For RCS services, SIP preconditions must not be used.

6.) Early media and announcements
The UE must support reception of voice and video media associated with one early dialog, e.g. when a SIP 180 message follows the SIP INVITE and the UE must support the P-early-media header field. TS 24.628 provides further details on how the UE renders locally generated communications progress information.

7.) Forking
By way of reminder, SIP forking represents a mechanism to split a SIP call into multiple clones for multiple endpoints. This leverages usability, and we may imagine an incoming call that can ring many endpoints at the same time. With SIP forking, your desk phone can ring at the same time as your softphone or a SIP phone on your mobile.

Network forking is at the discretion of the operator. For interoperability and forward-compatibility reasons, the UE must be ready to receive responses generated due to a forked request and behave according to the procedures specified in TS 24.229. Furthermore, the UE should be able to maintain at least forty parallel early dialogs until receiving the final response on one of them and the UE must support receiving media on one of these early dialogs [Ref. 8].

8.) Signaling compression
[Ref. 8] indicates that the UE must not use signaling compression methods like Signaling Compression (SIGCOMP) when the initial IMS registration is performed in 5G RAN.

9.) SIP timer operation
To enable a connection flow and potentially handle unexpected waiting and expiry times, the UE must handle session expiry timers during the INVITE process as further defined in RFC 4028 [Ref. 15].

10.) RCS feature capability discovery
The capability or service discovery mechanism is a process which enhances service usability by allowing a user to understand the subset of RCS services available to access and/or communicate with their contacts at certain points in time [Ref. 10]. If this feature is supported by the UE and the network, a configuration parameter CAPABILITY DISCOVERY MECHANISM can enable this capability exchange procedure. For the dual-registration case mentioned above, the capability exchange

can happen on either of the two registrations. In the current version of 5G IMS support, capability discovery is applicable for the following services: MMTEL conversational voice and video, MMTEL call composer, RCS messaging services and MSRP-based enriched calling services.

11.) User agent

User agent and server headers are used to indicate the release version and product information of the instant messaging (IM) clients and IM servers. The UE should include the user agent header in all SIP request messages and server headers in all SIP response messages. In addition, the user agent header should be included in HTTP requests for XCAP. [Ref. 10] provides further details on the compilation format of these headers.

As in legacy networks, the higher layers describe the voice services and other applications as multimedia telephone services (MTSI) or as MMTEL. They are defined in TS 26.980. Voice or video is the application layer itself that is transferred via the real time protocol (RTP) over User Datagram Protocol (UDP)/IP protocols [Ref. 13]. In this protocol layer view, the RAT layer can either be EUTRAN in VoLTE or 5G NR in VoNR. Beside the transport of audio-visual content over RTP, the real time control protocol (RTCP) adds additional information about the corresponding RTP stream, e.g. calculated jitter values. IMS uses RTP as a radio layer agnostic media transfer protocol. The definition of RTP and RTCP is provided in IETF RFC 3550 [Ref. 13]. The main purpose of RTP is to allow the receiver to play received media at the proper pace since IP networks introduce packet delay and loss as well as jitter. For example, assuming two IP packets are sent to the same destination with 10 ms delay between both packets, nothing guarantees that these two packets also arrive at the destination with a 10 ms delay. IP packet #2 may arrive right after packet #1, or much later or even before it. The RTP time stamps are used to recover the right timing relationship between the IP packets.

Since voice or video can run using a streaming service, the Internet Engineering Task Force (IETF) defined the real time streaming protocol (RTSP) in RFC 2326 [Ref. 19] as a network control protocol designed for use in entertainment and communications systems to control streaming media servers. The transmission of streaming data itself is not a task of RTSP. Most RTSP servers use RTP in conjunction with RTP Control Protocol (RTCP) for media stream delivery. On the control plane, there are two major protocols, i.e. the SIP and session description protocol (SDP). Since some SIP messages may be transferred over an IPsec tunnel, an optional IP key exchange (IKE) may be executed additionally.

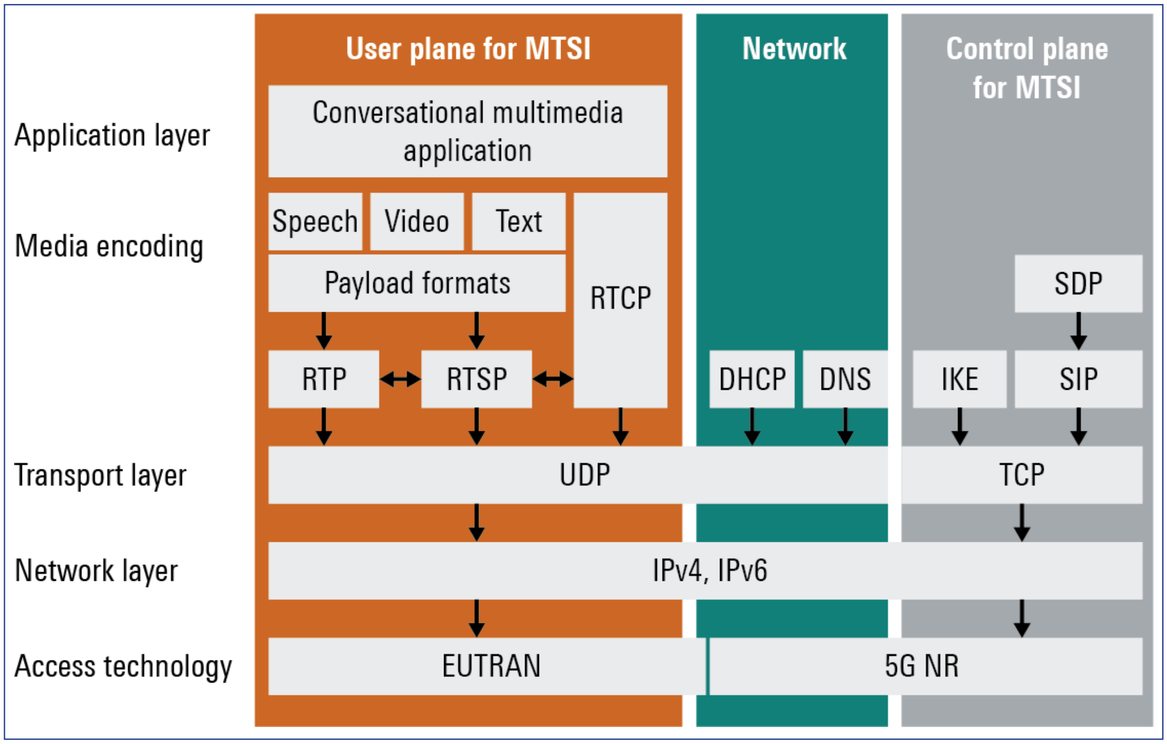

Figure 5: Protocol layer for MTSI.

The objective in this section is to describe the details of the PDU session relevant for IMS connection setup. First, we assume as a prerequisite that the UE is registered and a 5GMM context is established. The PDU session establishment message is conveyed to the AMF via the Uplink Non-Access Stratum (UL NAS) transport container message. Here, the parameter message type indicates the NAS message PDU session establishment request as initial request and the DNN is set on the name of the corresponding IMS network. It is the responsibility of the SMF to connect to the IMS, and thus the PDU session establishment request contains parameters indicating the IMS connection request (see TS 24.501). For example, the SSC mode should be set by the UE to SSC mode 1 request, the PDU session type IE indicates whether the UE prefers an IPv4 or IPv6 address or either one, and in the 5GSM capability, the UE should indicate the support for the S1 mode and reflective QoS. With the help of the extended protocol configuration option (PCO), the UE may signal that an IMS session should be established (TS 24.008). If the UE is establishing a PDU session for IMS and the UE is configured to discover the P-CSCF address during connectivity establishment, the UE should include an indicator that it is requesting a P-CSCF IP address within the Session Management (SM) container IE (TS 23.502). In case of successful PDU session establishment, the network responds with a possible setting of the P-CSCF address in the session establishment accept message and the network activates a PDU session with QoS flow 5QI = 5 inheriting the IMS signaling QoS profile. IMS support is obviously a prerequisite for VoNR as discussed in further detail in section 2.2.6.

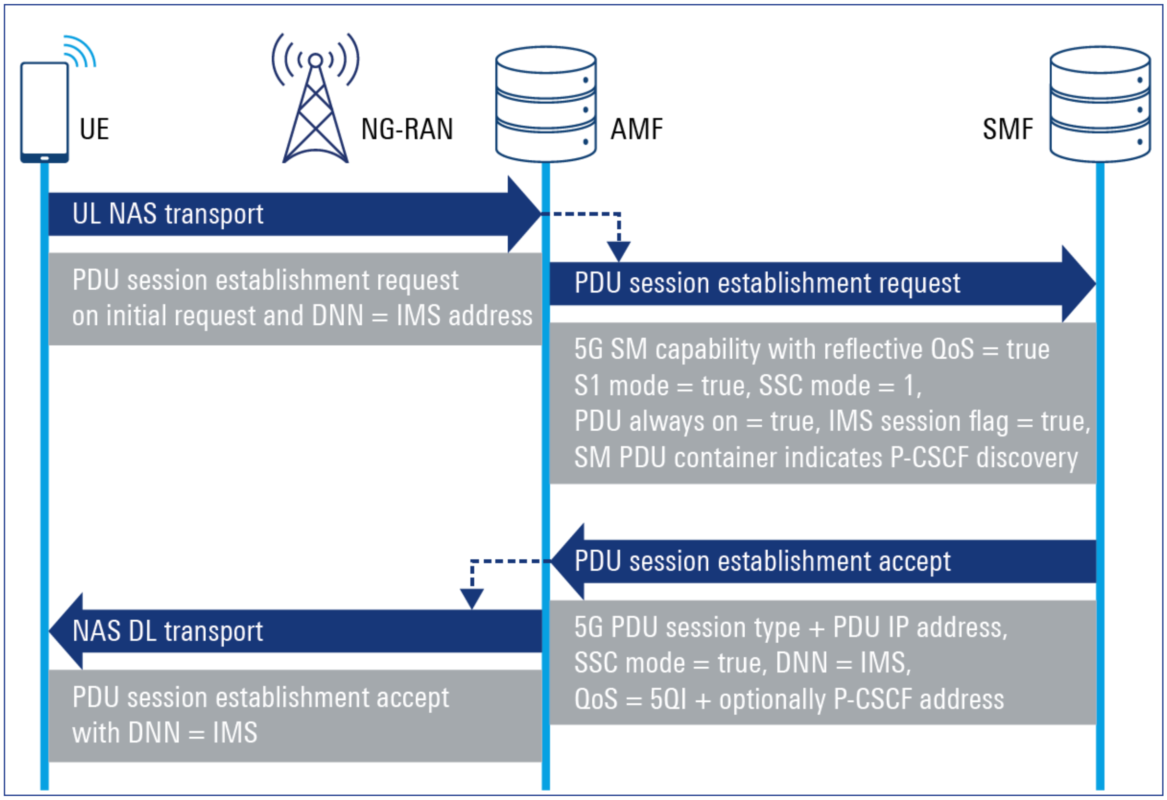

Figure 6: PDU session establishment for IMS voice.

2.2.4 EPS fallback

From a high-level perspective, EPS fallback represents a situation where a voice call is guided from NR to LTE already during the connection setup procedure to keep the connection setup time minimized. Thus, the UE may fully use the offered 5GS services and the operator maintains the expected KPI with this additional mobility trigger to EPC. For example, the advertised 5G QoS flow parameters will be mapped to the corresponding EPC bearer during this handover procedure. The trigger for EPS fallback can be UE-based, e.g. when the UE signals limited voice support only for EPS and thus indirectly requests an EPS fallback or it can be network-based. In detail, the network may use two standardized mobility procedures to trigger the EPS fallback: connection release with redirect to EUTRAN (here, the N26 interface is very beneficial to reduce the call setup time) or the inter-system handover command.

EPS fallback with the updated infrastructure EPC-5GC interconnection facilitates the transition to a VoNR deployment. Assuming an operator has initially deployed support for EPS fallback without supporting voice over NR, then the migration to VoNR is by changing the RAN configuration to not trigger fallback and letting the QoS flow for voice be established on NR connected to 5GC instead. This transition then provides a legacy-compatible and future-proof path to introduce VoNR services from a network perspective only without taking the risk of curtailing the services to existing UEs. The reason for the network commanding the UE to VoLTE can be e.g. a temporary lack of radio resources in NR for voice and moving the connection for capacity reasons to LTE or a general non-provisioning of voice services in 5G at all. The latter can be caused by an operator with a 5G focus on data services like eMBB or URLLC only. Intentionally, the operator does not yet implement IMS services connected to 5GC. It is also possible for a UE to signal via its capability information that only voice over EPS services are supported. Hence, the UE indirectly requests redirection to EPS.

An EPS fallback scenario from the protocol layer perspective is depicted in Fig. 6. First, a UE camps on 5G NR and a 5G connection is established. The 5GC at least provides a SIP control connection to recognize that a voice connection is requested by the UE. Either via redirection or handover procedures, the connection is moved from 5G NR to EUTRA networks. The advantage of EPS fallback is that the UE or gNodeB (gNB) only needs to support the IMS signaling channel (SIP over NR, low real-time requirements), and does not need to support the IMS voice/video communications channel (RTP or RTCP over NR, high real-time requirements). RTP or RTCP over NR requires lower latency and better 5G NR radio coverage [Ref. 2]. Thus, it can be considered as an intermediate step to the provisioning of VoNR, depending on device and network maturity.

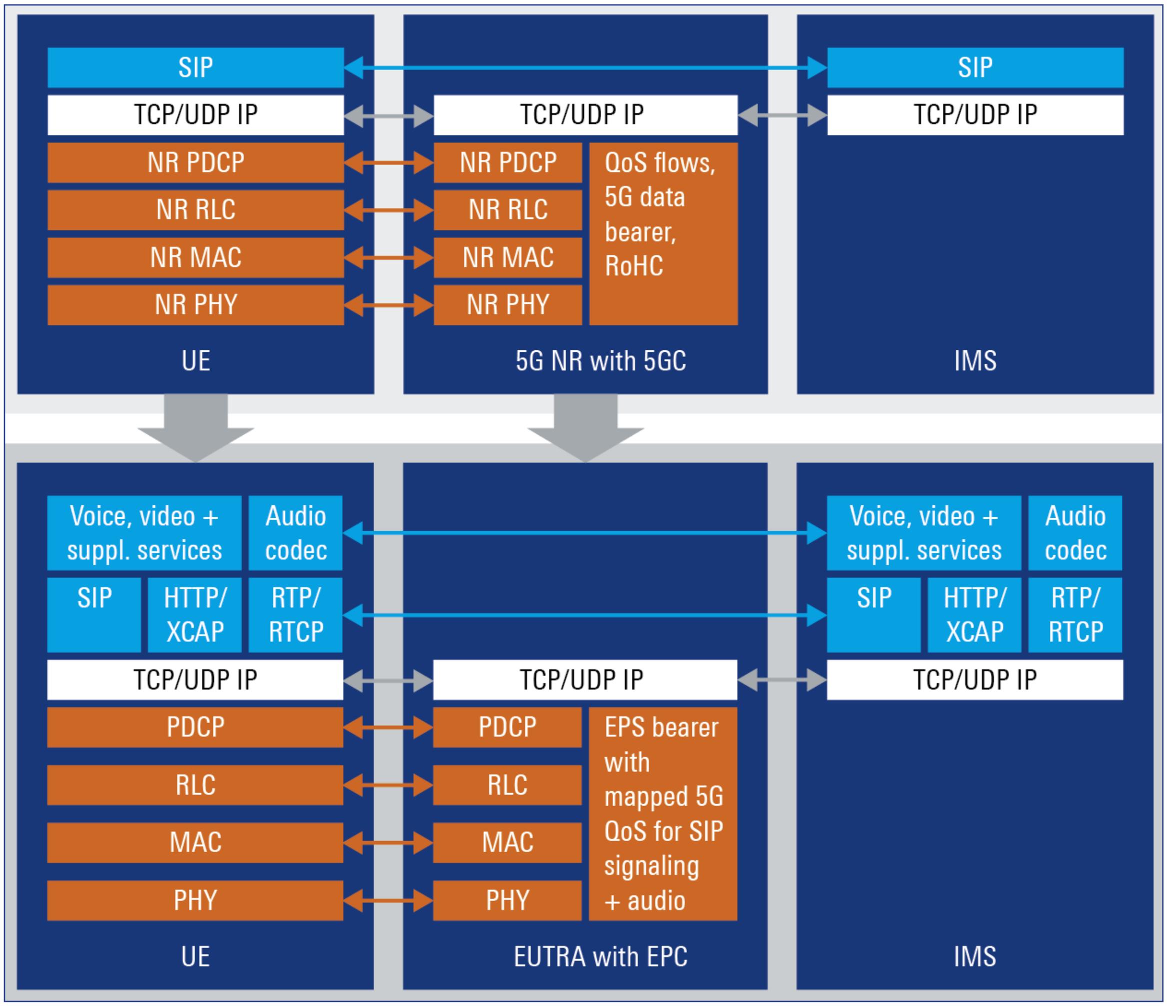

Figure 7: EPS fallback scenario.

The migration to VoNR can be accomplished once all required voice capabilities are in place in the network. Devices introduced before this step will still be in the field when voice over NR is introduced. The capabilities of these devices will determine if the devices will use voice over NR or continue to rely on EPS fallback. Hence, the network will support voice over NR including EPS fallback [Ref. 1] for a long period of time.

Upon the attempt to establish the QoS flow for the voice media over NR during call setup, the NG-RAN rejects the QoS flow setup towards the SMF with an indication that mobility is in progress. The NG-RAN initiates transfer of all PDU sessions from 5GS to EPS using one of the two standardized procedures:

- Release with redirect
- Inter-system handover

TS 23.501 and TS 23.502 define general policies of interworking between 5GC and EPC as well as registration aspects. In brief, the UE performs a registration procedure when moving from EPC to 5GC and it either executes a tracking area update or initial attach procedure when moving from 5GC to EPC.

The UE performs a tracking area update procedure if it has at least one PDU session for which session continuity is supported during interworking, i.e. the UE has EPS bearer ID and mapped EPS QoS parameters received.

The UE performs an initial attach procedure if it is registered without PDU session in 5GC or the UE is registered only with PDU session for which session continuity is not supported during interworking to EPC, and either the UE or the EPC does not support attach without PDN connectivity.

Details of the tracking area update: (TS 23.501) "When a UE is CM-CONNECTED in 5GC and a handover to EPS occur, the AMF selects the target MME based on the source AMF Region ID, AMF Set ID and target location information. The AMF forwards the UE context to the selected MME over the N26 Interface. In the UE context, the AMF also includes the UE Usage type, if it is received as part of subscription data. When the Handover procedure completes successfully the UE performs a Tracking Area Update. This completes the UE registration in the target EPS."

When the UE is served by the 5GC, the UE has one or more ongoing PDU sessions each including one or more QoS flows. The serving PLMN AMF has sent an indication towards the UE during the registration procedure that IMS voice over PS session is supported inducing an IMS registration. During this registration procedure, the network indicates whether the N26 interface is supported or not.

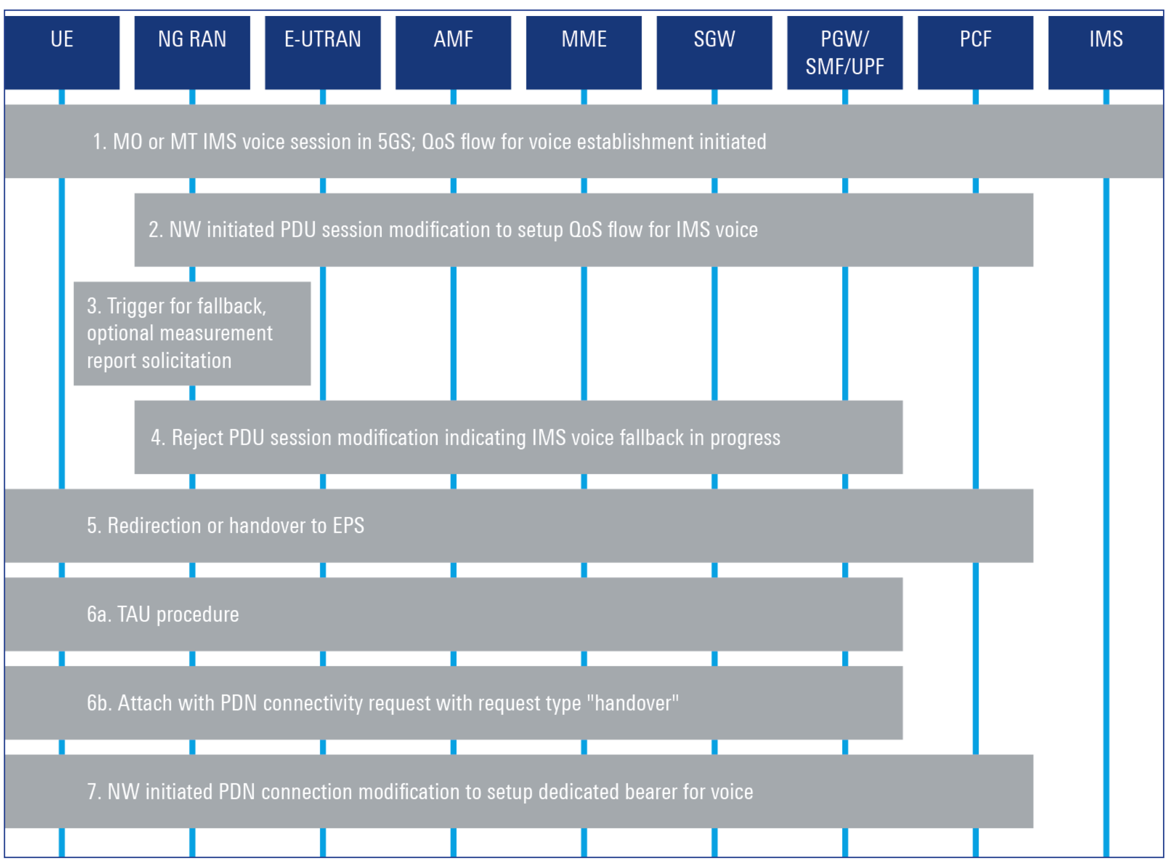

Figure 8: EPS fallback singling message flow [TS 23.502].

The signaling flow for EPS fallback as shown in Figure 8 (TS 23.502):

1. The UE camps on NG-RAN and a voice session establishment has been initiated, either IMS mobile terminated or originated (see Fig. 6)

2. Network initiated PDU session modification via N2 interface to set up QoS flow with 5QI = 1 for voice reaches the NG-RAN

3. NG-RAN is configured to support EPS fallback for IMS voice and decides to trigger fallback to EPS, considering the UE capabilities. During the initial context setup (TS 38.413) procedure, the AMF indicates that "Redirection for EPS fallback for voice is possible" plus provides network configuration (e.g. N26 availability) and radio conditions.
The NG-RAN may initiate measurement report solicitation from the UE including EUTRAN as target

4. NG-RAN responds by indicating the rejection of the PDU session modification to set up QoS flow for IMS voice received in step 2 by PDU Session Modification Response message towards the PGW-C+SMF via AMF with an indication that mobility due to fallback for IMS voice is ongoing

5. NG-RAN initiates either handover or AN Release via inter-system redirection to EPS, considering the UE capabilities

6a. In the case of 5GS to EPS handover and in the case of inter-system redirection to EPS with N26 interface, the UE initiates the tracking area update (TAU) procedure

6b. In the case of inter-system redirection to EPS without N26 interface, the UE initiates Attach with PDN connectivity request with request type "handover"

7. After completion of the mobility procedure to EPS or as part of the 5GS to EPS handover procedure, the SMF/PGW re-initiates the setup of the dedicated bearer(s) for the maintained PCC rule(s) including the dedicated bearer for IMS voice, mapping the 5G QoS to EPC QoS parameters. The PGW-C+SMF reports about successful resource allocation and access network information. The IMS signaling related to IMS voice call establishment continues after step 1 as in legacy VoLTE. At least for the duration of the voice call in EPS, the EUTRAN is configured to not trigger any handover to 5GS.

A message flow with further IMS signaling information is depicted in Fig. 9. In this example, we use a redirection command to EUTRA and assume a mobile originated call and N26 interface supported by the network:

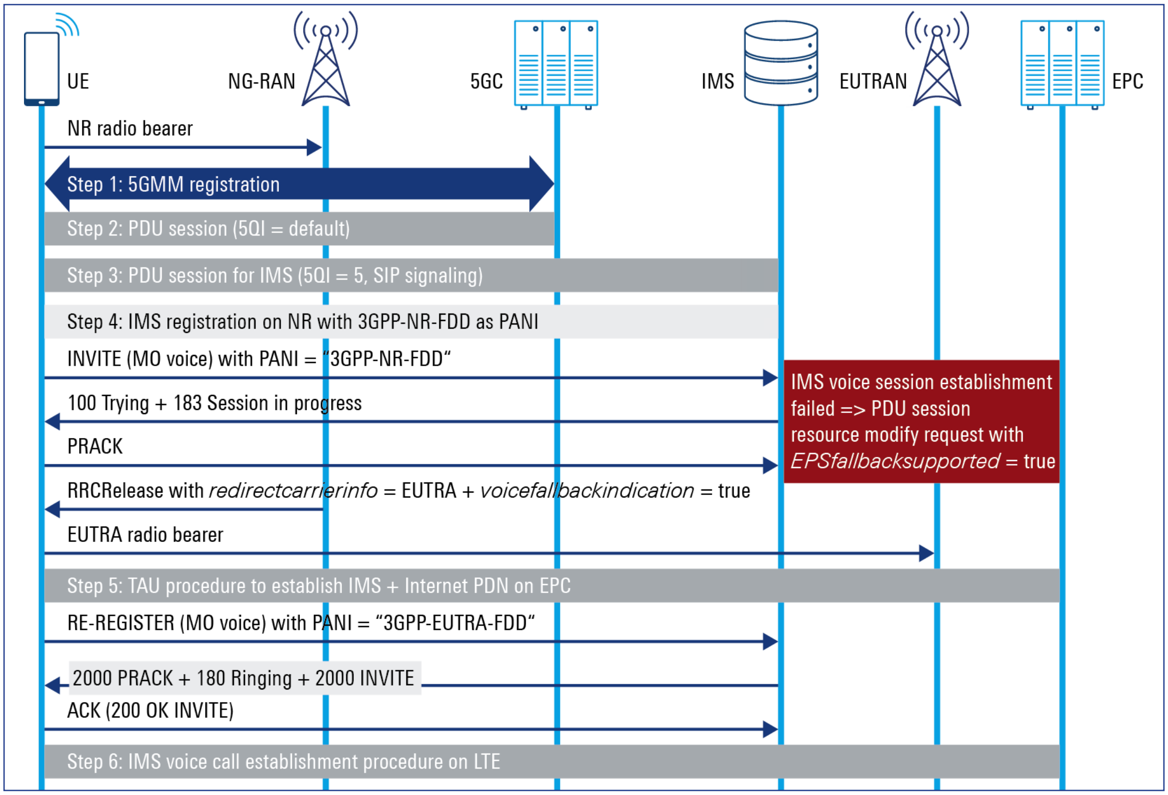

Figure 9: EPS fallback signaling procedure.

With the step 1 procedure as further explained in Fig. 4, the UE registers to 5GMM. The assumption is that the UE settings are voice-centric, S1 mode is preferred and the network supports the N26 interface plus voice services. During the registration procedure, the UE and network exchange their capabilities. EPS fallback can be indirectly triggered by the UE. If the UE capability flag *UECapabiltyInformation-NR* sets the indicator VoiceOverNR flag as FALSE, it can be understood as an indirect solicitation of EPS fallback and NG-RAN informs the AMF about the IMSVoiceSupportIndicator setting to FALSE in the UERadioCapabiltyCheckResponse message. In response, the AMF commands the NG-RAN to set the RedirectionEPSFallbackIndicator to TRUE in the IntialContextSetupReq message. When triggered by higher layers, the UE starts with steps 2 and 3 to establish a 5G NR PDU session for IMS as shown in Fig. 6 and a general Internet DNN PDU session in parallel.

The IMS registration procedure in step 4 provides some voice-specific details. Like in general IMS sessions, the network assigns an IP address to the UE and informs about the P-CSCF address. The SIP signaling messages REGISTER and SUBSCRIBE contain the primary-access-network-info (PANI) [Ref. 16] as "3GPP-NR-FDD" or "3GPP-NR-TDD" and the SIP definition of a cellular network with parameters such as Mobile Country Code (MCC), Mobile Network Code (MNC), Tracking Area Code (TAC) and NR Cell Identity (NCI) (TS 24.229). In the above example, the UE registers via 5G to IMS; thus, the TAC has a length of 6 hexadecimal digits (as length 4 in the LTE case) and NCI is present as well. Via the SDP, the UE requests the establishment of a voice call based on the Enhanced Voice Service (EVS) audio codec. To indicate a pending response, the IMS sends the 183 session in progress message. During the dedicated QoS flow establishment and resource reservation procedures, 5GC decides upon EPS fallback for the voice call because VoNR cannot be established. This decision is either based on previous UE solicitation of EPS fallback or triggered due to missing VoNR support for 5GC. As a consequence, the NR-RAN initiates the Radio Resource Control (RRC) connection release (either via the *RRCRelease* or *HandoverRequest* command) with redirection to EUTRAN info to fallback into EPS. As in our example, we are assuming there is support for the N26 interface between 5GC and EPC. This allows the procedure for a TAU and the continuation of the DNN and IMS PDU sessions; thus, the UE and P-CSCF IP addresses are preserved. In the TAU message, several parameters are set to specific values: The field containing the old GUTI is set with the "5G-GUTI" value while the EPS bearer status is set with "internet and IMS PDU" as the description and "active" as the status information. With the SIP message RE-REGISTER, the UE sets a new PANI equal to "3GPP-EUTRA-FDD" as the new network access identifier. The final step 6 in this call setup procedure is the establishment of a voice bearer inheriting QCI = 1 for voice.

Just to provide some examples of real message logs – the following figure contains two extracts from a signaling procedure executing an EPS fallback scenario with the R&S® CMX500. First, the messages on the right indicate the establishment of a PDU session. The extract given here indicates in fact two PDU sessions: one with default 5QI and the second with IMS signaling 5QI details. Via RRC messages (not shown in this extract), the 5G NR network triggers an RRCRelease message with redirection indication to EUTRA. The latter message is shown as an extract on the left side of this figure.

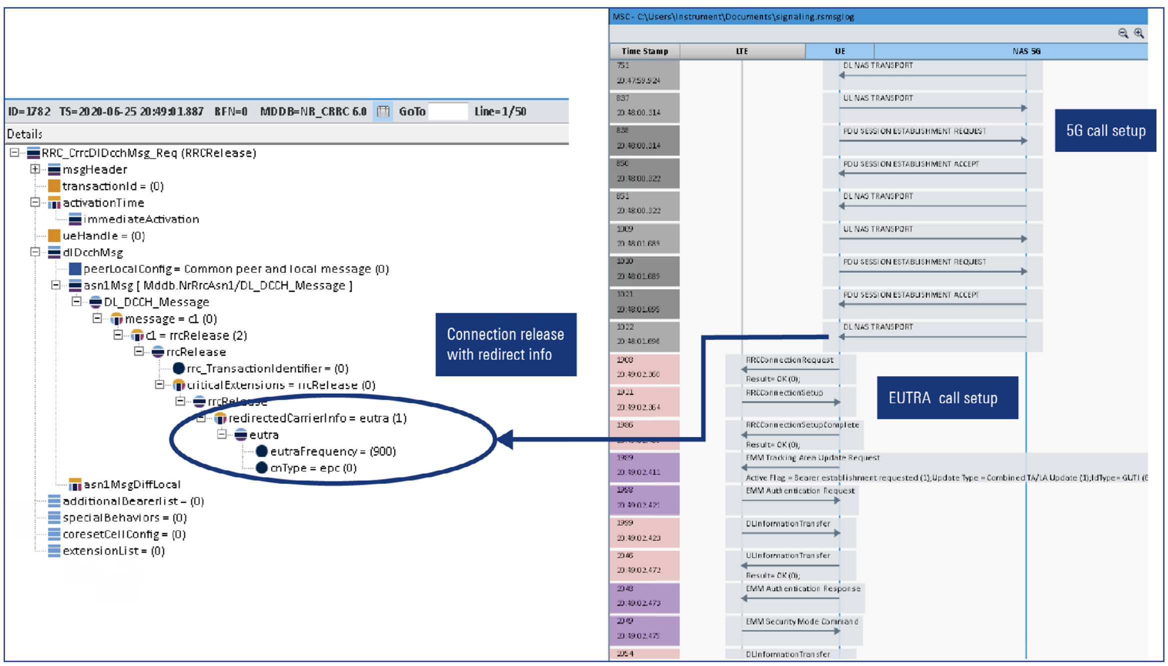

Figure 10: EPS fallback signaling scenario extract from R&S®CMX500.

2.3 Voice services using NG-eNB

Since 5G allows various deployment methods, multiple connectivity options are the consequence. For example, option 4 and option 7 introduce the 5GC as the core network and update the LTE eNB to a Next Generation NodeB (NG-eNB). The objective is to offer core network services and application layer bearers with 5G QoS profiles in an early 5G deployment stage. This connectivity option likely includes the provision of voice services, and consequently the 5GC requires a connection to IMS to support such services. One aspect in this deployment scenario is that the UE is camping on the EUTRA carrier in advance; thus, no handover is needed. The benefit is the shorter call setup time. In the protocol structure, the difference shown in the following figure is the update within the LTE protocol stack such that the Packet Data Convergence Protocol (PDCP) protocol layer is upgraded from LTE PDCP to NR PDCP. The advantage is full control of the services – like in the current situation, the voice services by 5GC. This voice scenario does not necessarily require dual connectivity as in option 4 or option 7 of the connectivity scenarios; thus, it can represent a feasible deployment for upgrading the existing LTE network with connectivity to 5GC and provide legacy services like voice over NG-eNB if no VoNR is possible due to lack of coverage. Obviously, the UE should support the enhancement with the NR PDCP implementation, so it is not compatible to legacy networks, in which case the NG-eNB still needs to provide VoLTE functionality in parallel.

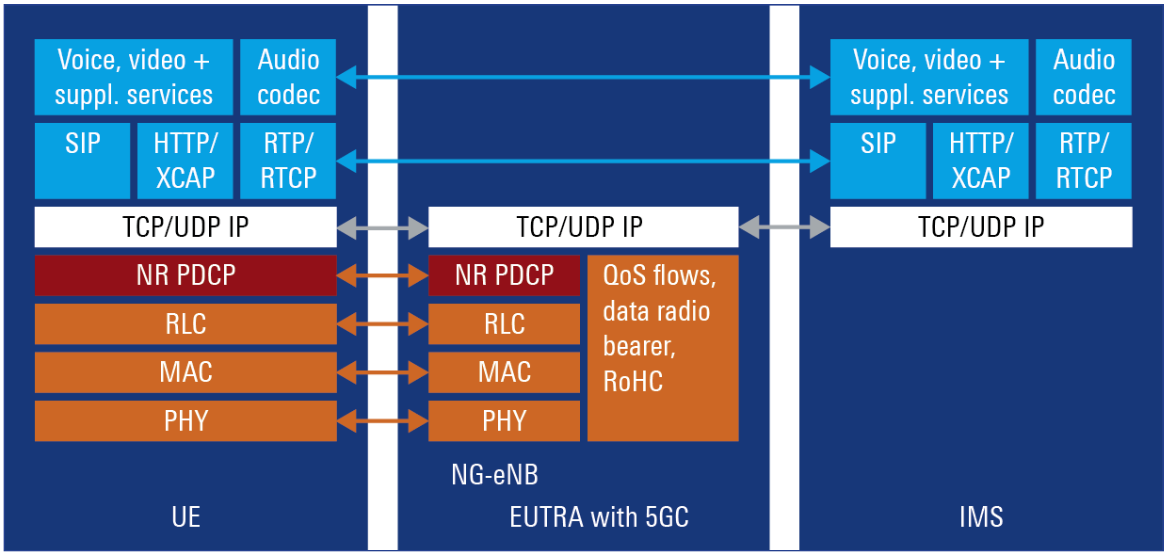

Figure 11: Voice services using NB-eNB.

2.4 Voice services with RAT fallback

RAT fallback is a similar scenario to EPS fallback. The objective is related to the fact that 5G NR may not provide full voice service coverage. The major difference is that EPS fallback is a kind of double handover, first containing a RAT change from NR to LTE and second a core network handover from 5GC to EPC as shown in Fig.11. The intention of RAT fallback is to support the enhanced eNB (NG-eNB) and the provisioning of 5GC as a core similar to the voice services offered over NG-eNB and 5GC, but with the difference that the UE has camped on NR previously. Thus, only the RAT is changed when the UE and network set up a voice connection. The objective is for the 5GC to define the QoS flow parameters for voice services at an early stage, e.g. even during the 5G NR radio bearer setup and maintain it as long as possible, even after the RAT change. TS 23.502 defines this procedure as inter-RAT fallback in 5GC for IMS voice because in a holistic view, the procedure includes any RAT as target. Thus, it may be used for handover from one NR RAT to another NR RAT, assuming operators run services on multiple NR RANs in parallel, or it can involve redirection to a target RAN as legacy EUTRA, which appears to be the most likely case. Like EPS fallback, the transition is either initiated by a direct handover command or via a redirection procedure. The advantage is that this mode does not necessarily require dual connectivity and 5G NR does not need to provide full voice services coverage including RTP and RTCP; only a basic SIP signaling connection is needed. The drawback is similar to EPS fallback, i.e. a longer call setup time affecting the user QoE.

Figure 12: RAT fallback supporting voice services.

The RAT fallback procedure can generally be used to trigger a RAT change (not necessarily EUTRA since NR to NR RAT changes are also possible). The prerequisite is similar to the EPS fallback scenario; the UE is served by the 5GC and has one or more ongoing PDU sessions each including one or more QoS flows. The serving PLMN AMF indicated during the registration procedure that IMS voice over PS session is supported, triggering an IMS registration.

The following figure describes the message flow:

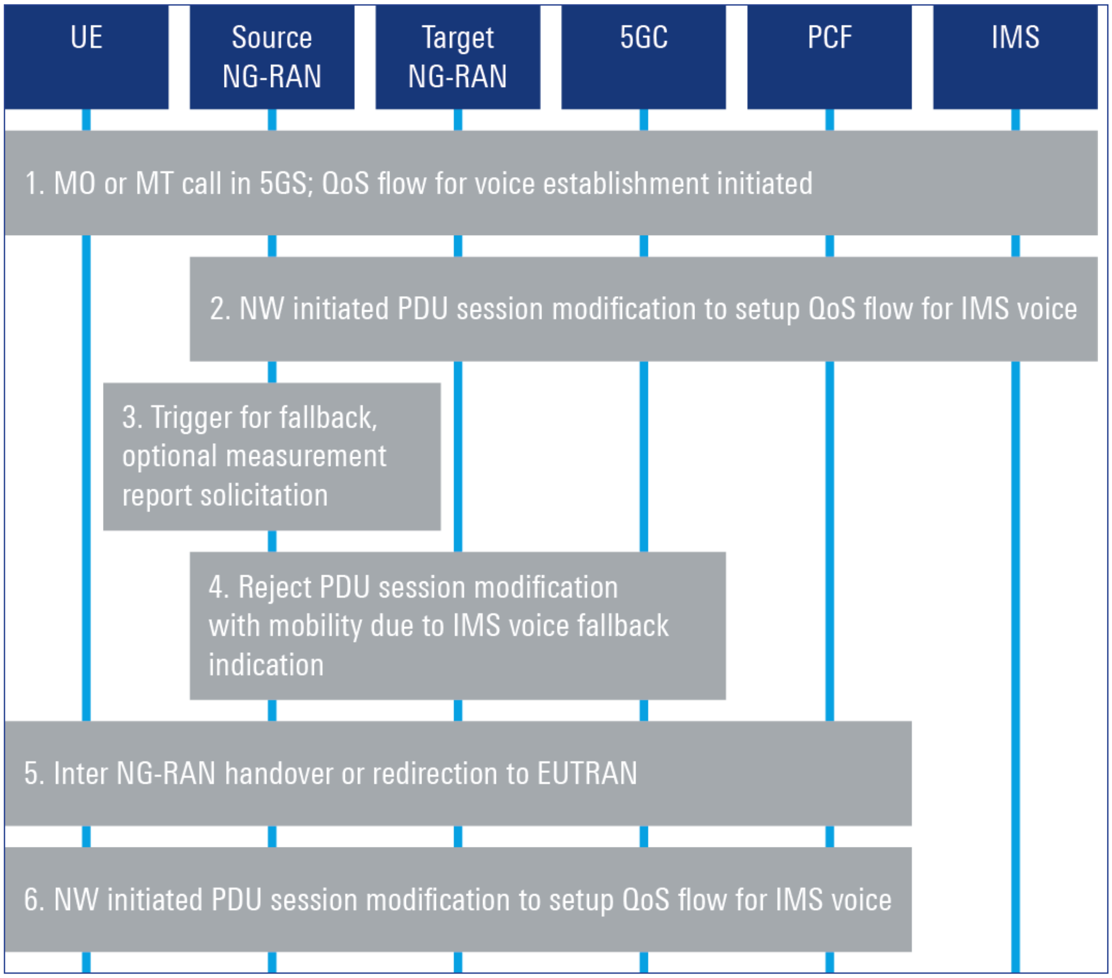

Figure 13: RAT fallback for IMS voice.

1. The UE camps on source NG-RAN in the 5GS and MO or MT IMS voice session establishment has been initiated

2. Network initiated PDU session modification to set up QoS flow for IMS voice reaches the source NG-RAN via the N2 interface

3. If source NG-RAN is configured to support RAT fallback for IMS voice, source NG-RAN decides to trigger RAT fallback, considering the UE capabilities, network configuration and radio conditions. To leverage a reliable mobility scenario, the source NG-RAN may initiate measurement report solicitation from the UE including target NG-RAN

4. Source NG-RAN responds indicating rejection of the PDU session modification to set up QoS flow for IMS voice received in step 2 by *PDU Session Response* message towards the SMF via AMF with an indication that mobility due to fallback for IMS voice is ongoing

5. Source NG-RAN initiates Xn-based inter NG-RAN handover or N2-based inter NG-RAN handover or redirection to EUTRA connected to 5GC

6. After completion of the inter NG-RAN (inter-RAT) handover or redirection to EUTRA connected to 5GC, the SMF re-initiates the PDU session modification to set up QoS flow for IMS voice. The SMF reports about successful resource allocation and access network information

The IMS signaling related to IMS voice call establishment continues after step 1 as in legacy VoLTE connections. At least for the duration of the IMS voice call, the target NG-RAN is configured to not trigger inter NG-RAN handover back to the source NG-RAN.

2.5 Voice over NR (VoNR)

Voice over NR represents the voice services provided by the 5G RAN, 5GC and IMS. Besides assumed support for voice within all the involved entities, IMS is integrated into the deployment scenario. The advantage is that IMS manages (like in VoLTE) the PDU session establishment with the relevant QoS flow for optimized voice quality. The prerequisites for VoNR are support for IMS voice over PS by the UE, signalled support for IMS voice and the emergency service support indicator by the network. The assumption is also that the 5GC is available as the core network and most likely the deployment option 2 representing 5G standalone is the underlying infrastructure. There is the possibility of supporting VoNR in option 3, i.e. NSA mode as well, but most likely operators deploying option 3 will go for voice connections via LTE (see section 2.2.4). Due to the proliferation of voice services, VoNR can be considered as a prerequisite for successful 5G deployment in standalone operation, e.g. option 2 deployments. The VoNR protocol stack includes the 5G NR stack on the radio link layers. On the application layer, there is full support for IMS and voice services containing the relevant speech codecs, i.e. EVS as demanded by the 3GPP. Not directly related to VoNR support, but strongly recommended is thus support for emergency services by 5GC and NR; otherwise, several UEs would likely refrain from camping on NR cells.

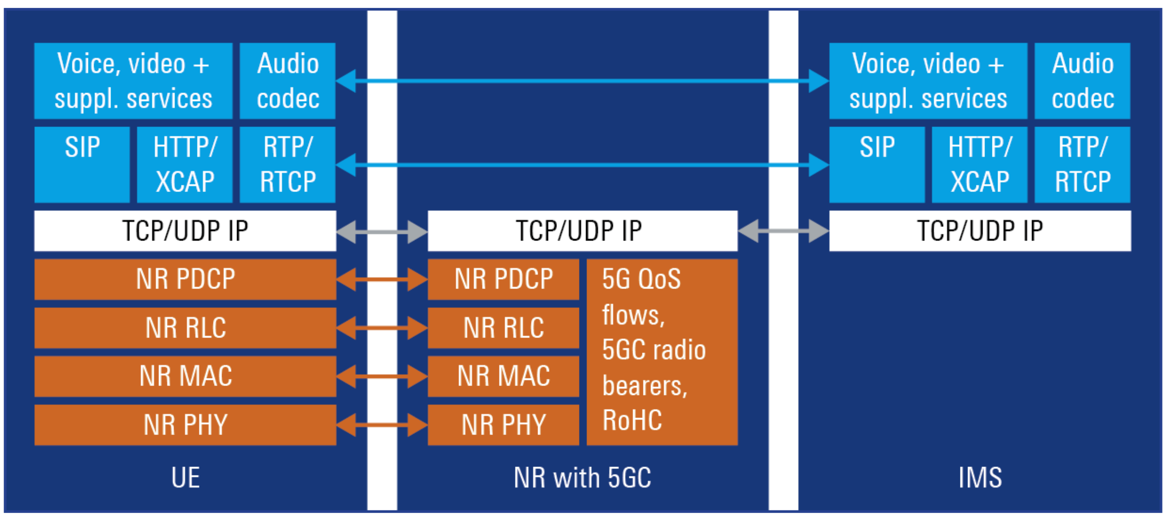

Figure 14: Voice over NR (VoNR).

Handling of voice requires that both NR and UE support the QoS flow for voice over the radio access, i.e. QoS flow establishment is supported by the gNB. The VoNR QoS consideration supports QoS flows for conversational voice and video, for IMS signaling and for non-GBR voice or video for MSRP services. It is worth mentioning that a UE depending on RCS support parameter settings may have two IMS registrations (see section 2.2.3). To fulfil the voice gap KPIs and avoid dropped calls, a certain relationship and interconnection between 5GS and EPS are needed to support mobility scenarios, which are a traditional inter-system handover, VoNR to VoLTE.

Although not a mandatory requirement but still strongly recommended, in a deployment with support for voice in 5GS, both SMF and UPF are interconnected via the S5 interface and tight interworking between AMF and MME is guaranteed by the existence of the N26 reference point. Complying with this recommendation increases the chances of providing seamless inter-system mobility with good voice characteristics and short handover interruption times.

Such an inter-system handover initiates transfer of all PDU sessions from 5GS to EPS and all QoS flows in the IMS PDU session / IMS PDN are transferred between AMF and MME, using inter-system handover signaling over N26. The SMF and UPF ensure IP address preservation and QoS mapping between 5QI and traffic flow templates (TFT) for the QoS flows in the IMS PDU session. Handover of EPS to 5GS follows the same procedures. In a real deployment, one may find a kind of phased introduction of handover capabilities, starting with the one-way handover (NR in 5GS to LTE in EPS). The two-way handover or reverse handover from EPS to 5GS may be supported at a later deployment stage. In the case of the one-way handover, the phone stays on the underlying LTE network during the rest of the call, even if it moves back into NR coverage.

The message flow for VoNR follows the connection setup message flow in a VoLTE situation [Ref. 6] except that instead of EUTRA and EPS, the 5G RAN and 5GS are involved. First, the serving PLMN AMF sends an indication towards the UE during the registration procedure to indicate whether an IMS voice over PS session is supported in the 3GPP access network. A UE with "IMS voice over PS" voice capability over 3GPP access takes this indication into account when performing the core network domain selection. The UE and network establish a PDU session for IMS DNN with QoS flow 5QI = 5 and SSC mode as 1 and always ON flag set to TRUE. Within this procedure setup, the UE includes extended protocol configuration options (ePCO) in the PDU session establishment request to the AMF. In this signaling container, the UE includes the P-CSCF IPv4/IPv6 request information element to indicate an IMS DNN target [Ref. 7]. Furthermore, the AMF forwards these ePCO options towards the SMF. The SMF fetches the P-CSCF addresses based on the DNN profile, which maintains IMS-related data and includes the P-CSCF IPv4 or IPv6 address in an N1 message towards the AMF. The SMF does not include the P-CSCF address if the UE does not set the P-CSCF container options in the ePCO field. Optionally, the UE can request a PDU session modification in case it wishes to change QoS rule. On the basis of the selected media or the voice or video codec, the network sends the session modification complete with negotiated QoS parameters that will be confirmed by the UE with the modification complete message. After these QoS and media setting agreements, the UE and the UPF start the user plane data session using the agreed audio or video codec on the N3 interface.

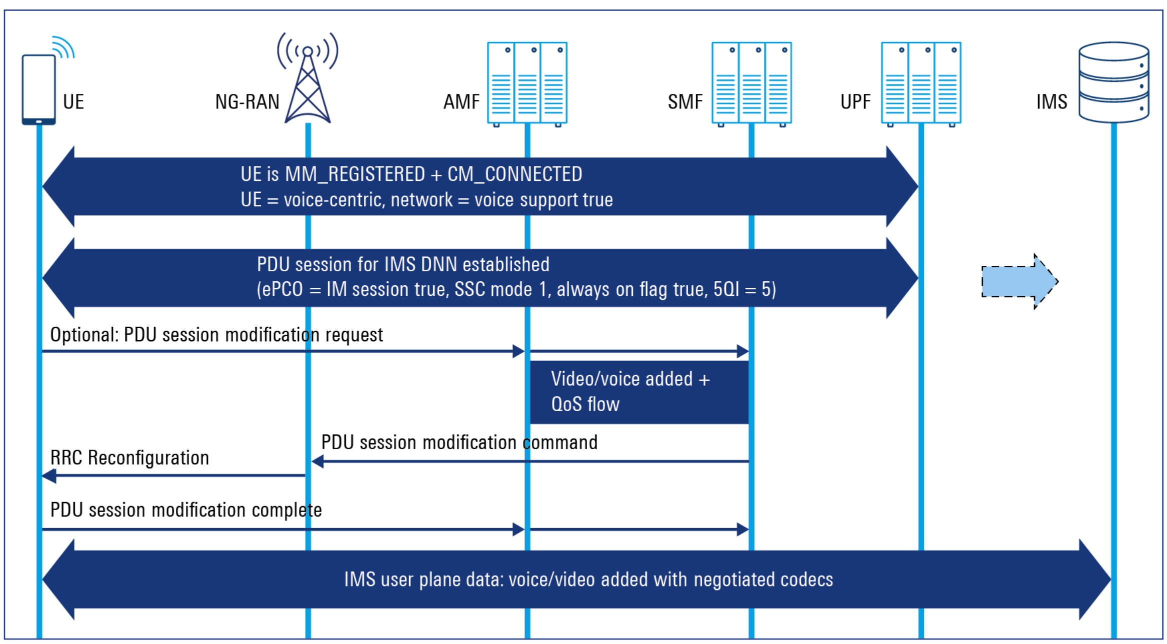

Figure 15: Voice over NR message flow.

2.5.1 VoNR radio parameter support recommendation

Support for voice services mainly involves integration and support for IMS connected to 5GC, but nevertheless to guarantee the best QoS for voice and video, a recommendation paper has been published [Ref. 8] that also describes certain lower layer functionalities. The network and UE should support these radio layer features. To support voice services with sufficient QoS, these protocol layers should support the following features:

PDCP layer – Robust header compression methods. One of the functionalities provided by the PDCP layer is RoHC as defined in TS 38.323. At a minimum, the UE and network must support the "RTP/UDP/IP" profile (0x0001) to compress RTP packets and the "UDP/IP" profile (0x0002) to compress RTCP packets. The UE and network must support these profiles for both IPv4 and IPv6 address allocations.

Radio bearers – To support 5G voice and video services, the UE and network establish data and signaling radio bearers with a certain QoS profile. The RLC layer offers transport services in acknowledged mode (AM) and unacknowledged mode (UM). For IMS signaling transfer, the UE and network establish an AM Data Radio Bearer (DRB) with 5QI = 5. An additional AM DRB with 5QI as one of the values 6/7/8/9 is likely to be established concurrently for non-Guaranteed Bit Rate (non-GBR) services. Depending on the service being either voice or video as suggested by the 3GPP, the UE and network set up an UM DRB with 5QI = 1 for voice and/or 5QI = 2 for video. One suggestion to avoid timeouts is that the NG-RAN should set the discard timer (*discardTimer*) for the IMS carrying DRB to a value of "infinity".

Discontinuous Reception (DRX) mode – To maximize the lifetime and efficiently manage UE battery consumption, [Ref. 8] suggests that the UE and network should support MAC layer DRX functionality.

PHY layer settings – There are no direct recommendations concerning VoNR support affecting the physical layer in documents such as the 3GPP specifications or [Ref. 8]. 5G NR does not specify a feature like TTI bundling known from EUTRA. Since Release 16, it has been possible to schedule via Downlink Control Indicator (DCI) certain repetitions for the Physical Uplink Shared Channel (PUSCH), but there is no automatic bundling for the downlink direction. The RRC layer can schedule radio resources in semi-persistent mode (SPS). Since voice and video are characterized by a kind of synchronous data flow, SPS could represent a scheduling mechanism to reduce the signaling overhead in VoNR connections. However, this is more of a recommendation for physical layer scheduling instead of an actual requirement.

One of the characteristics of voice is its relative non-delay tolerance and its requirement for a certain data rate. For this reason, a GBR DRB is recommended for voice services. [Ref. 8] classifies non-GBR bearers as not suitable for IMS-based voice services.

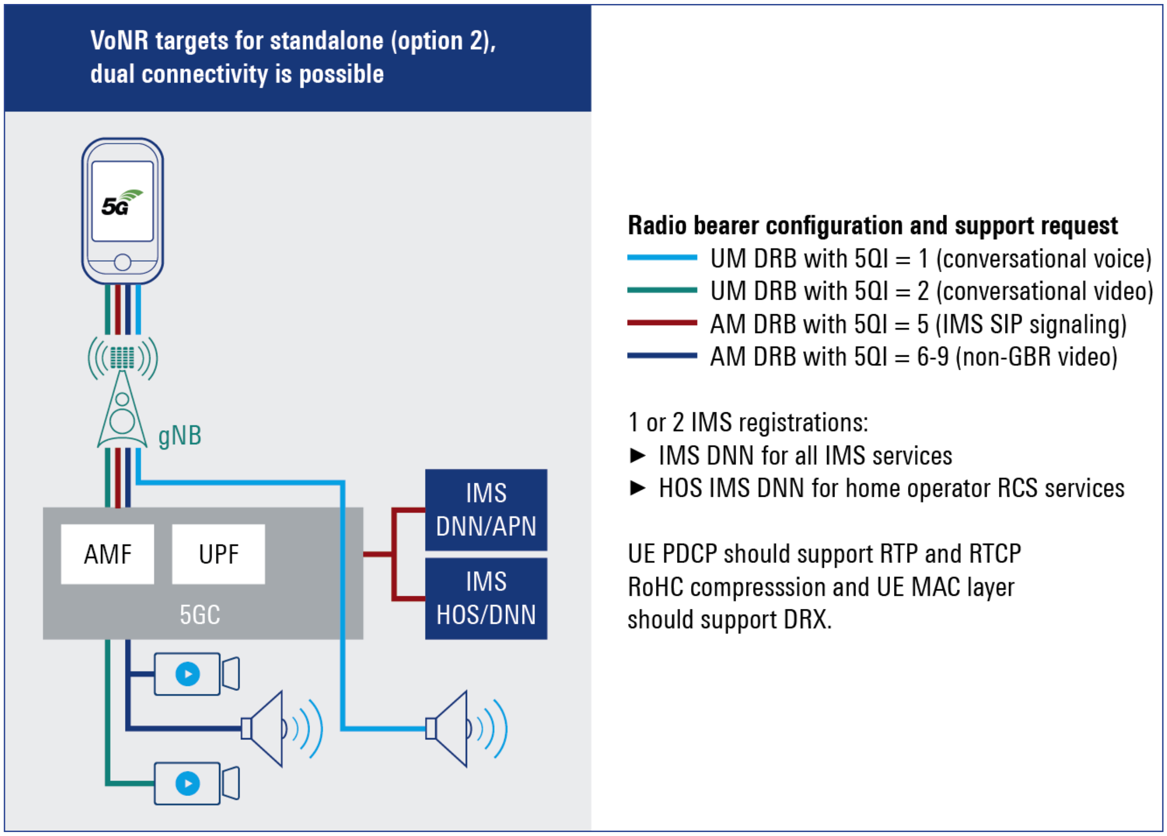

Figure 16: Voice over NR radio layer requirements.

VoNR IMS signaling and QoS considerations
As a summary of IMS support in VoNR, this section briefly discusses the signaling flow aspects and QoS considerations. There can be two IMS registrations: one is the default IMS APN/DNN and the second is an optional HOS IMS DNN. By default, the UE registers to the well-known IMS and obtains the single network slice selection assistance information (S NSSAI), an IPv4 or IPv6 address and discovers the P-CSCF address. The IMS signaling uses 5CQI = 5 and is solely reserved for IMS signaling. The UE should prevent non-IMS applications from using this QoS flow.

Optionally for RCS service support, the UE may register to a home operator (HOS) IMS DNN concurrently. This second QoS flow using the same 5QI may be used for home operator specific HTTP signaling messages with XCAP or HTTP content. Depending on the service, [Ref. 8] suggests 5QI = 1 for voice, 5QI = 2 for video and 5QI = 6-9 for non-GBR voice/video. Note that emergency services share the same QoS flow as conversational voice. A detailed description of further IMS signaling parameters and features for voice services can be found in [Ref. 8].

2.6 Network deployment and connectivity options supporting voice
The objective of this section is to provide an overview of the network deployment and connectivity options with a focus on how they support the provision of voice services. Previously, several voice scenarios were described with a focus on the protocol and bearer concept and the underlying architectural aspects. Now, these same scenarios are discussed again from the perspective of the underlying deployment options.

Option 3 supporting voice services
Option 3 EN-DC is the NSA deployment in 5G. Thus, the LTE connection is mandatory. With this option 3, the voice services are supported by EUTRA and the EPC core network. According to TS 37.340, the Master Cell Group (MCG) (or master bearer) either supports the LTE PDCP protocol

layer or the NR PDCP. Consequently, there are two possible ways to implement voice services in network option 3 mode: legacy VoLTE or enhanced VoLTE using the NR PDCP. The following figure simplifies these implementations of voice services; only the logical flow of voice is depicted, summarizing coded speech as user plane data and SIP signaling. The dual connectivity between LTE and 5G is only depicted symbolically. The advantage of this voice implementation is that it does not require any upgrade for the existing voice supporting infrastructure – except for the minor transition from PDCP to NR PDCP.

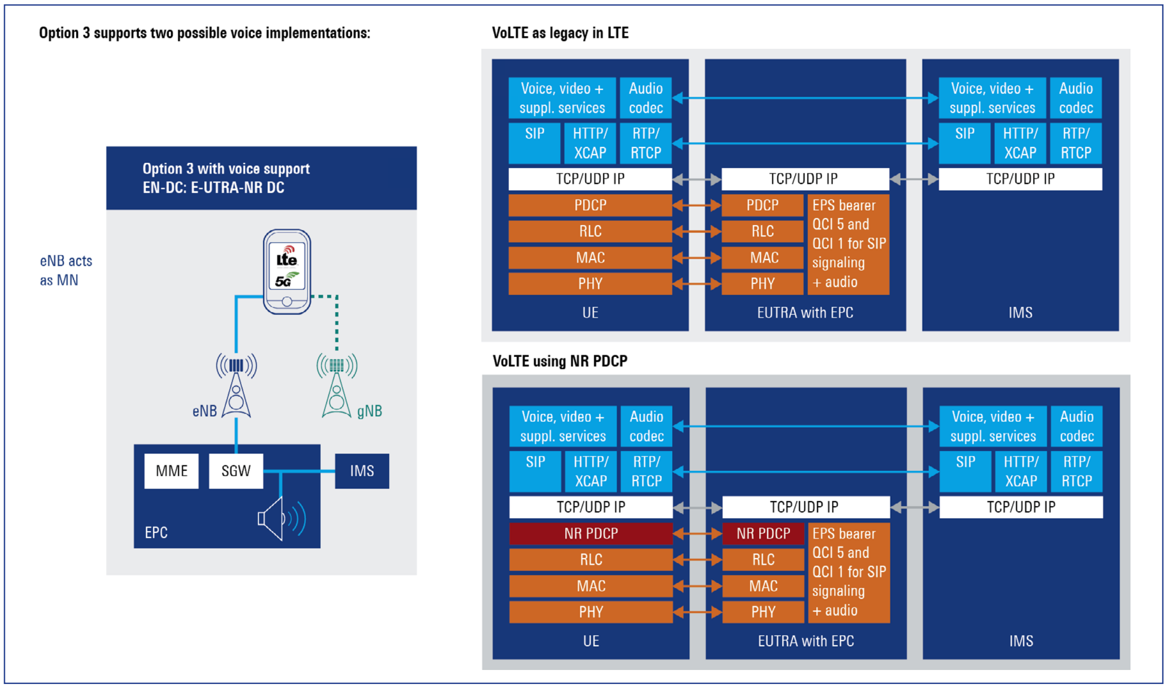

Figure 17: Network option 3 for NSA (EN-DC) supporting voice services.

Option 7 supporting voice services

Option 7 can be seen in the background of non-standalone access (NSA) schemes. It represents the NGEN-DC mode including the connection to 5GC and an upgraded eNB to NG-eNB but still with LTE as the master cell RAT. It provides an LTE MCG bearer and an NR SCG bearer in a dual connectivity scenario. This deployment offers the flexibility to either run voice services as enhanced VoLTE over NG-eNB including 5GC or additionally support to run VoNR over the secondary cell bearer directly as VoNR and then provide a smooth transition from VoLTE to VoNR.

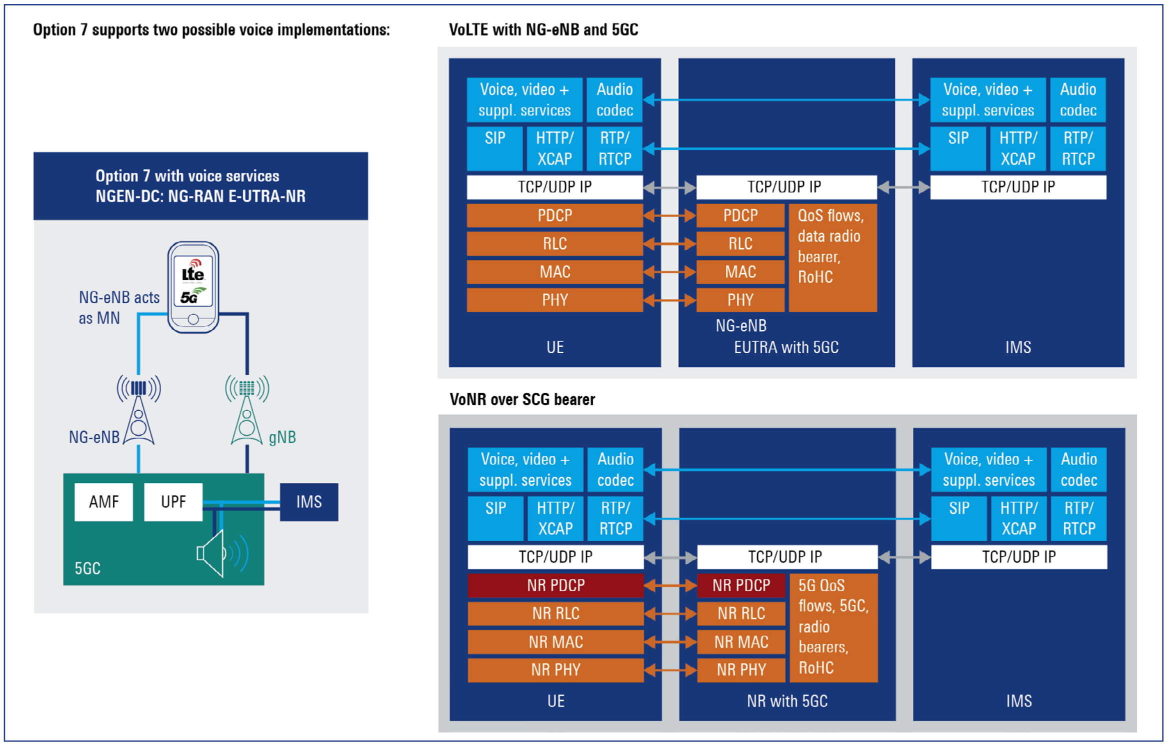

Figure 18: Network option 7 for NSA (NGEN-DC) supporting voice services.

Option 4 supporting voice services

Option 4 can also be considered as a kind of non-standalone (NSA) mode. In detail, it represents the NE-DC mode including the connection to 5GC and an upgraded eNB to NG-eNB but compared to option 7, the NR gNB is the MCG. It provides an NR MCG bearer and an LTE SCG bearer in a dual connectivity scenario. Similar to the option 7 connectivity scenario, this deployment offers the flexibility to either run voice services as enhanced VoLTE over NG-eNB including 5GC or additionally support to run VoNR over the secondary cell bearer directly as VoNR and then provide a smooth transition from VoLTE to VoNR.

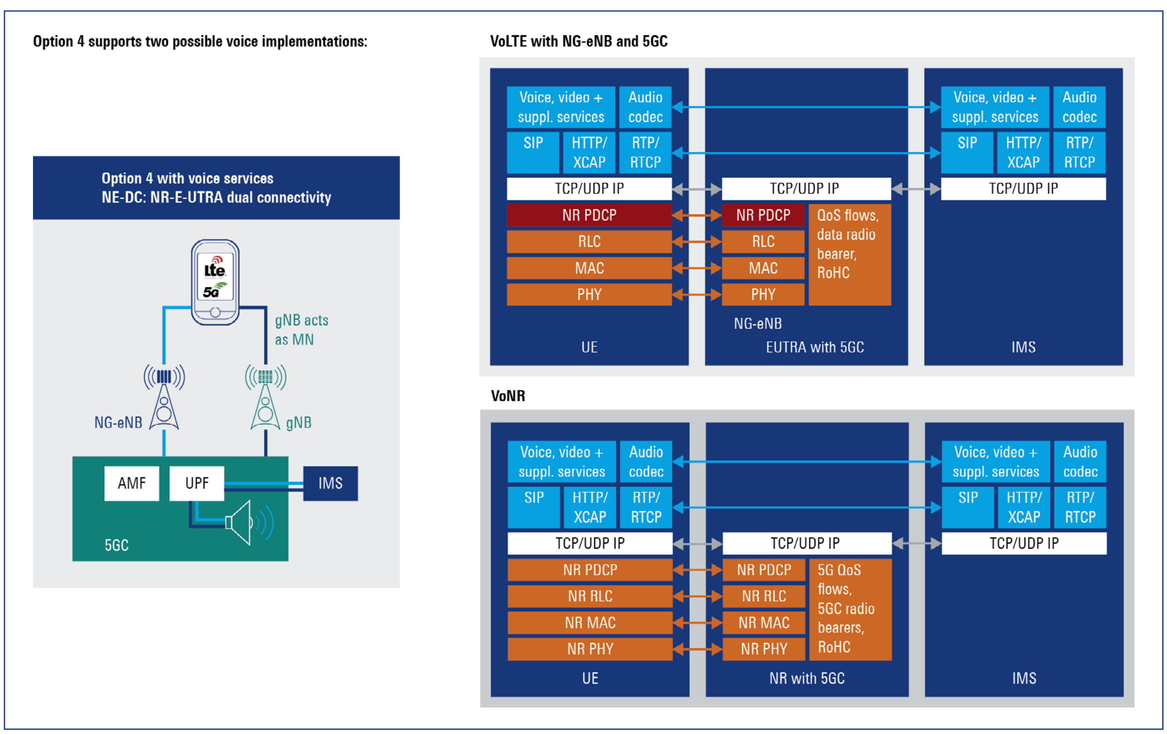

Figure 19: Network option 4 for NSa (NE-DC) supporting voice services.

Option 2 standalone (SA) supporting voice services

Option 2 represents the standalone (SA) deployment of 5G NR. If voice services should be supported in the 5G system, it is completely up to the operator to make sure that IMS and the speech codecs are fully supported. The voice services are offered based on the technology methodologies described as voice over NR (VoNR). Just for completeness and as is also known from legacy technology deployments (e.g. LTE with CSFB), in the option 2 network deployment it is very likely that an LTE carrier provides radio coverage in addition to NR radio coverage. Thus, many of the previously described voice scenarios are possible as well, i.e. EPS fallback providing a handover during connection setup from NR to LTE, or a RAT fallback if the LTE NG-eNB is connected to the 5GC. Further details on these setups can be found in [Ref. 2].

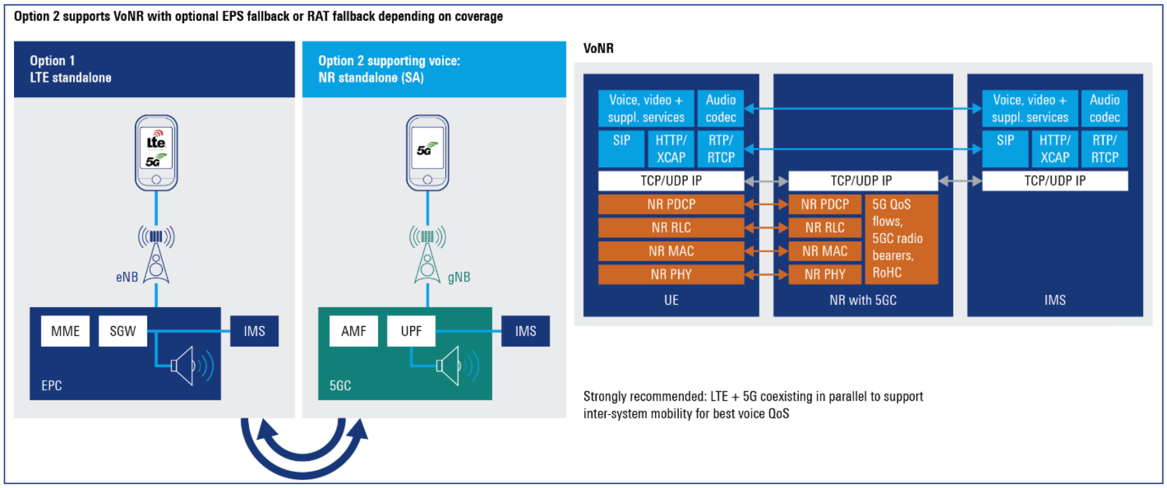

Figure 20: Network option 2 for SA supporting voice services.

Fig. 21 below attempts to depict the various voice services in 5G related to the network deployment options. Of course, this is a simplified view. Real implementations are naturally more complex and allow mixed modes depending on the UE capability support and network maturity. The RAT fallback mode is not directly depicted here since it can be understood as a similar scenario to voice over NG-eNB with 5GC support. The only difference is the assumption of whether the UE camps before call setup on the EUTRA carrier or on the NR carrier and a subsequent handover or redirection procedure is needed.

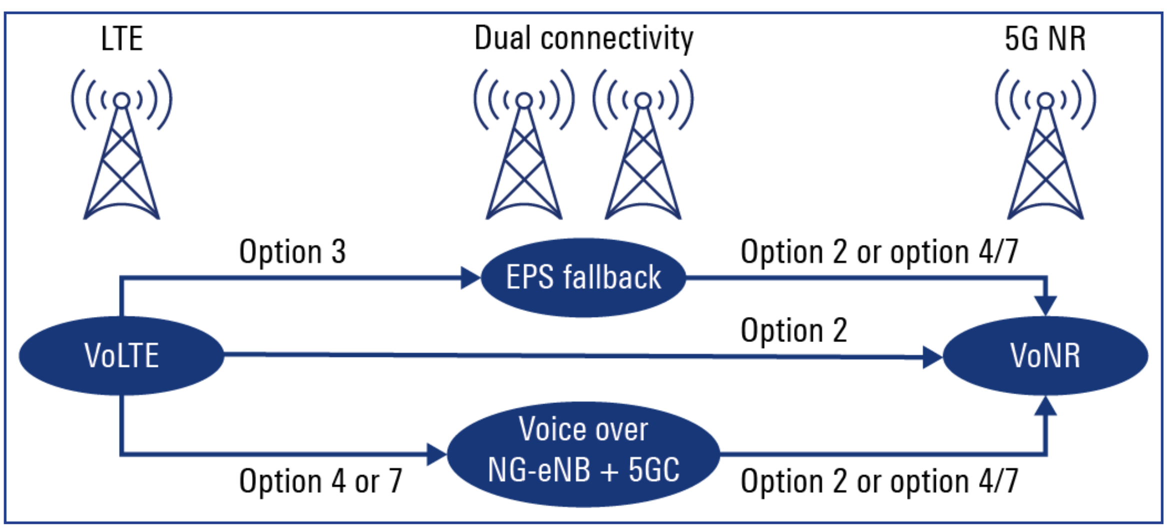

Figure 21: Possible transition scenarios from VoLTE to VoNR related to network deployment scenarios [Ref. 2].

2.7 Speech codec aspects and EVS

It is not our intention to go back into history to the very beginnings of telephone services intro duced by Johann Philipp Reiss and Alexander Graham Bell as the fathers of modern communications. Voice transmission over a channel suffers the limitation of data rate versus the requirement to provide audible services. In the early days of 2G (GSM), the limitation to a maximum throughput of 9.6 kbps led to the introduction of speech coding principles. Thus, the audio signal itself is not transferred directly as PCM digitized samples like in corresponding fixed-line telephone services such as ISDN, but instead a "digital voice coder model" of the audio signal is transferred. The basis for this speech codec approach is nevertheless a first step towards audio bandwidth limitation and a digital sample with a certain quantization ratio. The first GSM full-rate speech codec restricted the audio bandwidth to a 300 Hz lower boundary and a 3.4 kHz upper boundary. With a sampling ratio of 8 kHz, the limitation to a maximum throughput of 9.6 kbps could be fulfilled. There have even been notions of using a half-rate speech codec for greater network capacity. Facing competition with fixed-line networks and the voice quality offered here, the trend in wireless communications has obviously moved in the direction of better overall speech quality. This is first achieved through higher audio bandwidth and a higher sampling ratio for lower granularity in the quantization. The bandwidth evolution ranges from narrowband (NB) via wideband (WB) to super wideband (SWB) and ultimately to fullband (FB), covering the audible frequency range of a typical human ear from 20 Hz to 20 kHz.

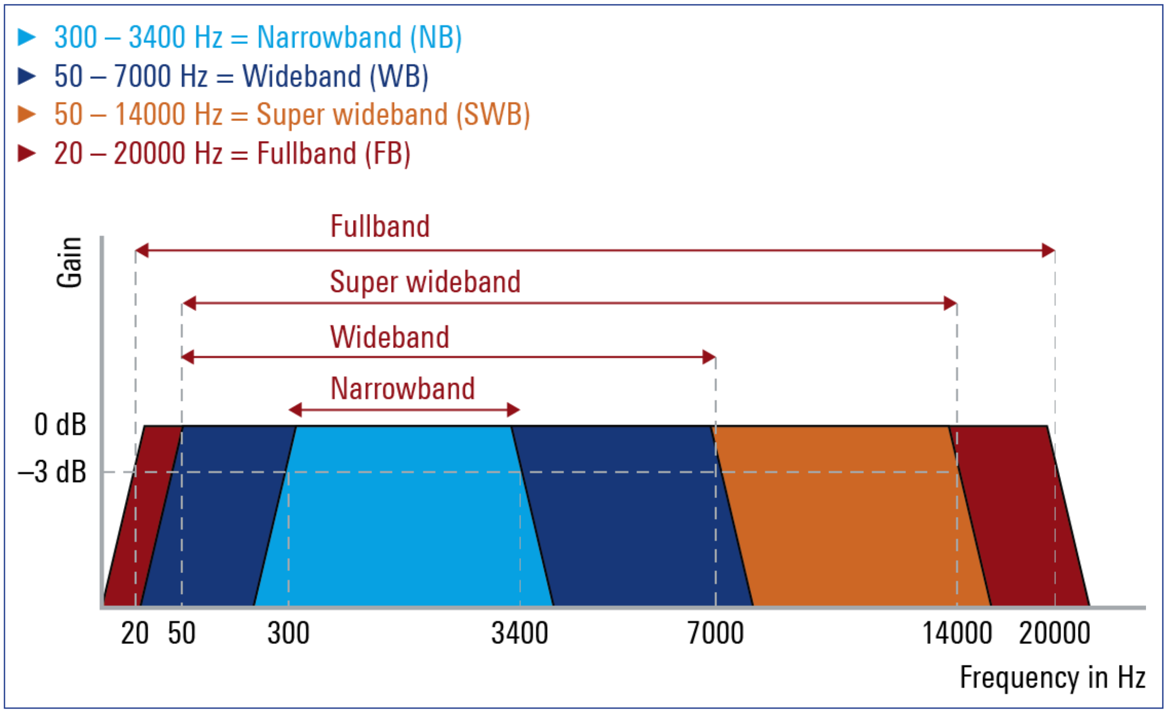

Figure 22: Evolution of audio bandwidth.

In later releases, the concept of the adaptive multi-rate speech codec (AMR) was introduced to cope with the fluctuating rate due to changing conditions on the radio interface. The speech coder provides multiple speech codec rates and depending on the current radio link quality or channel rate, the most appropriate speech rate is selected. Lower speech codec rates allow more overhead in the channel coding and ensure a more robust voice call if radio conditions deteriorate. The speed can be switched every 20 ms, which is the duration of one speech frame. Tandem free operation (TFO) in the context of AMR describes the agreement that in a bidirectional communications scheme, the same codec is used in either direction. To enable interoperability and future longevity of VoLTE and VoNR, GSMA has published a requirement document [Ref. 8]. Besides a list of required speech codecs that should be supported by the UE, a detailed description of how the SDP should indicate and select the available audio codecs is also provided in this document.

The motivation for introducing an EVS is to allow a combination of high-end audio such as music and voice [Ref. 3]. In commercial networks, for example, there is a product known as Full-HD-Voice representing an EVS speech codec that is already deployed in several LTE-Advanced networks. Initial speech codecs developed in GSM focused on voice services only while newer codecs focus on high-end multimedia services. To guarantee interworking with legacy voice codec technologies, EVS supports the nine codec rates of AMR-WB. This operation mode is known as interoperable mode (AMR-WB IO).

The audio encoder takes its input and can produce output at the decoder in the form of a 16-bit uniform pulse code modulated (PCM) signal at sampling frequencies of 8 kHz, 16 kHz, 32 kHz or 48 kHz. The audio may originate from and terminate within the audio part of the UE or from the network side or from the public switched telephone network (PSTN) via a narrowband 13-bit A-law or µ-law (8 kHz) PCM conversion or wideband (16 kHz) 14-bit uniform PCM conversion.

The source codec offers a total of eleven bit rates with support for various audio bandwidths as the EVS primary mode. For legacy interworking, nine different speech codec bit rates (AMR-WB) are supported. In detail, the source codec maps an audio sample (20 ms) as input into a 16-bit uniform PCM format resulting in encoded speech blocks. The EVS codec employs a hybrid coding scheme combining linear predictive (LP) coding techniques based upon algebraic code excited linear prediction (ACELP), predominantly for speech signals, with a transform coding method, for generic content. Like legacy codecs, voice activity detection (VAD) may signal discontinuous transmission (DTX) instead of non-existing audio and the receiving entity will play comfort noise generation (CNG) audio signals to feign link maintenance and save bandwidth. The EVS codec is capable of switching between these different coding modes without artefacts. (TS 26.445)

EVS primary mode ← EVS → EVS AMR-WB IO mode

Source codec bit rate (kbit/s)	Audio bandwidth supported	Source controlled operation available
5.9 (SC-VBR)	NB, WB	(Yes, always on)
7.2	NB, WB	Yes
8.0	NB, WB	Yes
9.6	NB, WB, SWB	Yes
13.2	NB, WB, SWB	Yes
13.2 (channel aware)	WB, SWB	Yes
16.4	NB, WB, SWB, FB	Yes
24.4	NB, WB, SWB, FB	Yes
32	WB, SWB, FB	Yes
48	WB, SWB, FB	Yes
64	WB, SWB, FB	Yes
96	WB, SWB, FB	Yes
128	WB, SWB, FB	Yes

Source codec bit rate (kbit/s)
6.6
8.85
12.65
14.25
15.85
18.25
19.85
23.05
23.85

AMR-WB interoperable mode for legacy codec compliance

Figure 23: EVS modes, bit rates and audio bandwidths.

Since our objective is to provide a concise introduction to the speech codec evolution in order to better understand voice aspects in 5G, EVS are illustrated in Fig. 24 in an exemplary manner.

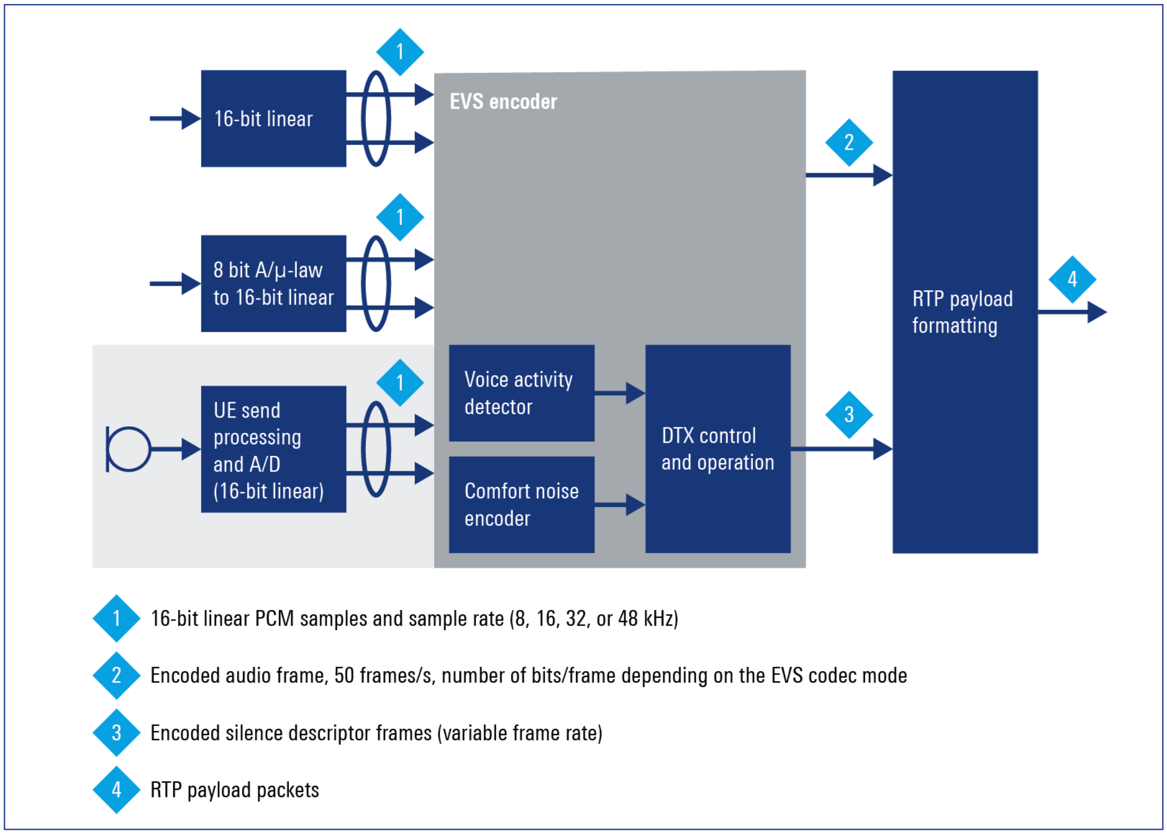

Figure 24: Audio processing functions of EVS code, TX side (TS26.441).

The enhanced voice service speech coder consists of the multi-rate audio coder optimized for operation with voice and music mixed content signals, a source-controlled rate scheme, a voice or sound activity detector and a comfort noise generation system together with an error concealment mechanism to combat the effects of transmission errors and lost speech packets (TS 26.441). Since speech packets are presumably transferred over a packet-switched network, jitter buffer management ensures time-aligned output of the audio signal without excessive delays.

The focus of this section has been on codecs for voice. Other modern and popular cellular communications systems incorporate video communications. This requires support for and definition of video codecs. In TS 26.114, the 3GPP stipulated that a network and UE supporting video over NR (ViNR) services must support ITU-T Recommendation H.264 as a video codec. There exist various video codecs and due to their evolution, the GSMA has published a list of requirements. To guarantee interoperability, these video codecs should be supported by the UE [Ref. 8]:

- ITU-T Recommendation H.264 Constrained baseline profile (CBP)
- ITU-T Recommendation H.264 Advanced video coding (AVC) or constrained high profile (CHP
- ITU-T Recommendation H.265 High efficiency video coding (HEVC) main profile

2.8 Supplementary speech services, emergency calls + SMS

2.8.1 Supplementary speech services

Along with organizations such as the GSMA and Open Mobile Alliance (OMA), TS 24.173 defines and promotes certain supplementary services for multimedia telephone calls (MMTEL). Although they are not directly related to 5G, [Ref. 8] recommends supporting these services in the interest of forward-looking service support and longevity of use cases. Besides direct IMS configuration parameters, the UE should also support a wide range of supplementary voice services. Among these voice services are well-known candidates such as calling party presentation, waiting indication, barring of incoming or outgoing calls, barring of international calls, ad-hoc multi-party conferences. etc. See [Ref. 8] for the full list of supplementary services that are requested to be supported by the UE.

Supplementary MMTEL services include protocols like the XML configuration access protocol (XCAP) [Ref. 14], which is an application layer protocol allowing a client to read, write and modify application configuration data stored on a server. Multi-party conferences can be set up on the fly via a three-way session creation process. The IMS core network must support a certain list of conference attributes such as conference info, maximum user amount, user display, media support as voice and/or video and a status indication as connected or disconnected. As is familiar from legacy networks, 5G supplementary speech services should also support mechanisms for message and call waiting indications, originating or terminating ID restrictions and communications diversion. Triggered by security concerns such as spamming and stalking or just for avoidance of high costs, 5G IMS supplementary services also support a wide range of communications barring mechanisms.

2.8.2 SMS in 5G

The short message service (SMS) concept introduced in the early days of 2G remains an important use case in 5G. There are two methods of transporting SMS messages over the 5G access (similar to LTE). The first method is SMS over IP (SMSoIP) using IMS as a management and coordination network to guarantee proper data transfer. The second method is SMS over 5G NAS (SMSoNAS), i.e. an approach to encapsulate the SMS data container into a 5G control message. In SMSoNAS, originated and terminated SMS messages are transported between the UE and AMF via NAS messages. According to [Ref. 8], both SMS transfer methods are mandatory for a UE claiming to be SMS-capable, while the network may select which options are offered. For the sake of completeness, we should mention that a short message can be transferred via RCS messaging services as well, but this is not the focus of this section. A short message service function (SMSF) has been defined in 5GS for handling these originating and terminating SMS using legacy transport protocols MAP [Ref. 11] or Diameter [Ref. 12] for SMS exchange between the UE and SMS service center (SMS-SC). In SMSoIP, an SMS service gateway for IP-based SMS (IP-SM-GW) must be present for originating and terminating SMS. This IP-SM-GW may also perform MT domain selection to other accesses (4G, 3G, 2G) while SMS transport between the UE and IP-SM-GW is via IP.

Both SMS over NAS and SMS over IP must be supported by a 5GS UE supporting voice, whereas each operator has the freedom to decide whether to support SMS over NAS, SMS over IP or both. These SMS solutions can be combined and can also be used in case EPS fallback is used for voice [Ref. 1]. During the registration procedure in 5GS, the UE includes an “SMS supported” indication in the Registration Request indicating the UE’s capability for SMS over NAS transport. TS 23.502 provides further details on that registration procedure as well as the distinct SMS connectivity scenarios involving mobile originated or terminated SMS messages when the UE is in either the CM_IDLE or CM_CONNECTED state.

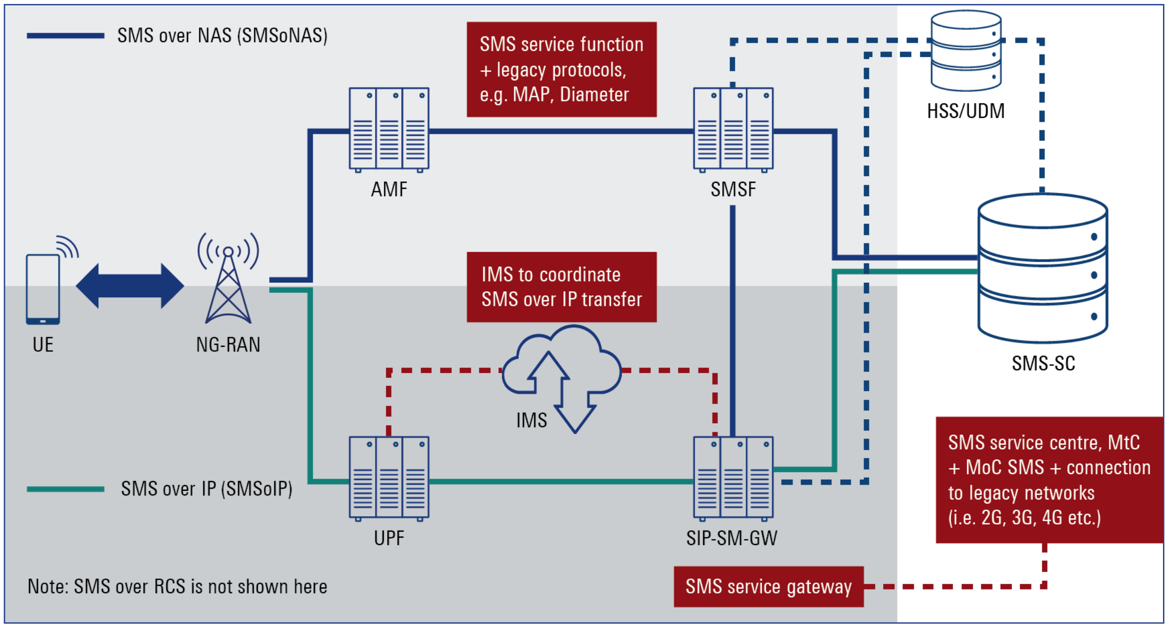

Figure 25: SMS in 5G [Ref. 1].

By way of example, the message flow for a mobile originating SMS over NAS in the CM_IDLE state is illustrated in Fig. 26. The main function applied to the 5G system is the short message service function (SMSF) representing the 5G function that provides the ability to deal with SMS and interconnect to the legacy SMS interworking MSC (SMSIWMSC). The SMSIWMSC guarantees delivery of the SMS to the endpoint regardless of the RAT of the destination UE.

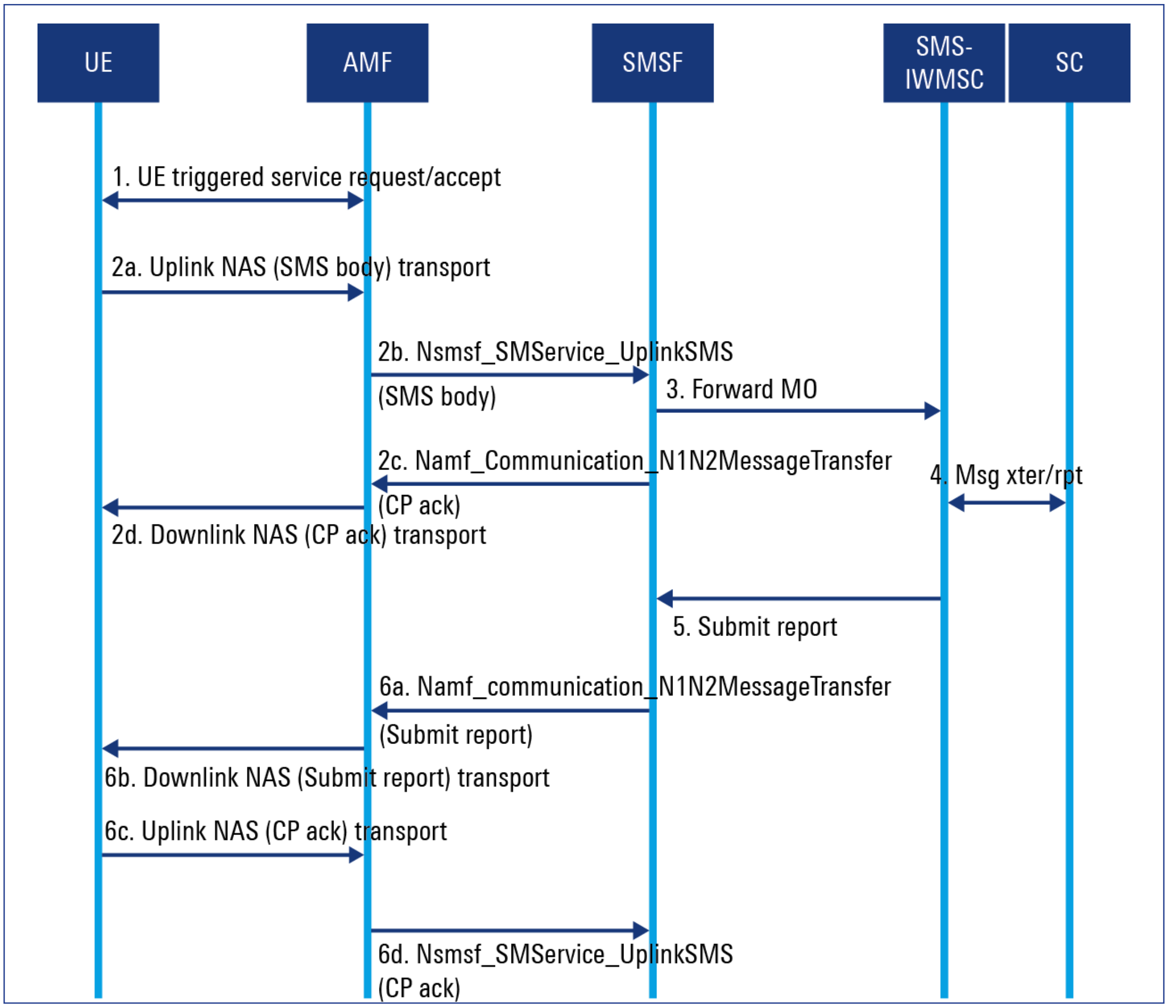

Figure 26: SMS transfer in 5G, message flow [TS 23.502].

2.8.3 Emergency services in 5G

Due to the proliferation of wireless communications technologies worldwide, cellular technologies are the number one access technology for emergency calls today according to first responder statistics. 5GS may offer emergency services directly or via fallback to EUTRA. Support for emergency services is indicated via system information broadcast. Emergency services in 5G provide the possibility to set up an emergency call via IMS-based emergency sessions, transferring emergency information such as the UE location and voice over an IP data network. Within the 5G system several possibilities exist, ranging from full support for emergency services within the 5G system through handover to legacy networks. Emergency services in 5GS behave like a VoNR connection, while during the IMS and RRC signaling flow, the RRC and SIP control messages indicate emergency as the use case. The PDN bearers, IMS registration and IMS connection setup behave according to the details described below. There is also the possibility that 5GS does not directly support emergency services. However, 5GS offers the possibility to hand over or redirect an emergency session to EPC. This procedure is defined as the emergency fallback procedure. Two derivates for 5G voice services exist, EPC and RAT fallback, and the same also applies to emergency services. With 3GPP Release 16, voice call handover to UMTS is also possible and is known as SRVCC. For emergency services, it is possible for the 5G RAN to trigger a SRVCC handover to UTRAN. Last but not least, depending on the current circumstances, the UE may also set up an emergency session via non-3GPP radio access, i.e. via WLAN. Concerning the registration status details and service offering, one may differentiate between a UE with normal 5GMM context after the registration procedure and a UE with limited registration. For example, the latter case could be an emergency call permitted for a UE without previous USIM registration to the network. This NAS agnostic permission is broadcast via SIB1 in a 5G NR RAN. A UE that is 5GMM registered obtains permission to start emergency services per RAT and per registration area during the NAS registration process. TS 23.501 describes further details for these circumstances and also whether the UE needs to be fully registered, i.e. the GMM context is established or whether the UE is allowed to set up an emergency call in emergency registration mode, i.e. possible USIM free network access. The latter mode requires a legal indication by the national regulator whether a valid network subscription is a prerequisite or not. Depending on local regulations and the operator's policy, the network may allow or reject a registration request for emergency services from UEs that are identified to be in a limited service state. If the 5GS supports emergency services, the support is indicated to the UE via the registration accept message and this permission is valid per registration area and per RAT. This indication also includes the possibility of an emergency fallback. If the 5GS supports emergency service fallback, it indicates the support to the UE via the registration accept message and this indicator flag is a complement to the emergency service support flag. This indication means that the 5GS does not directly support emergency services and if needed, they will be handed over to legacy technologies. Thus, it prevents the UE from camping on NR. TS 23.501 defines the details of when the network indicates support for emergency services, i.e. the network supports emergency calls directly or the network supports fallback or SRVCC to the legacy networks EUTRAN or UMTS. Support for emergency services is also signaled within the network from 5GC to the 5G RAN as it influences the possible setup of dual connectivity connections and the decisions about which RAT will be the primary cell RAT. A UE camping on a cell in a limited service state and not fully registered with valid 5GMM context may execute an emergency registration procedure to obtain information about whether the network supports emergency services (TS 23.501).

On the access stratum or RAT, two signaling aspects are worth mentioning in the context of emergency service signaling. At first glance, the procedures may appear similar to legacy technologies but there are some minor differences. The RAN signals support for emergency services in SIB1 with a flag, *ims-EmergencySupport*, but unlike legacy technologies, this flag indicates support for emergency services to UEs in limited service only, e.g. when a UE is not 5GMM registered or even without USIM. A UE that has a valid 5GMM context receives emergency service support during the registration procedure with important details as mentioned and this permission is valid per RAT and per registration area. In the uplink direction, the process is business as usual, i.e. the UE indicates with an RRC establishment cause set to "emergency" when it requests an RRC connection in relation to an emergency session. One can see the priority of the emergency situation since the UE signals an emergency situation with the first full and plainly visible UL message.

5G includes the concept of standalone, non-public networks (SNPN). TS 38.300 states that such SNPNs do not support emergency services.

2.8.3.1 Barring methods avoiding network congestion using access category and access identity

In the DL direction, the network has the option to restrict access to the network via sophisticated barring mechanisms known as uniform access control (UAC). Under high network load conditions, the network may protect itself against overloading by using the unified access control functionality for 3GPP access to limit access attempts by UEs. Some of the barring mechanisms are RAN-specific, such as Random Access Channel (RACH) backoff, RRC connection reject, RRC connection release and UE-based access barring mechanisms. Other mechanisms are NAS-specific in terms of where the definition of access categories and access identities happens; this is defined as UAC. Obviously, there is an interaction between the UAC on a higher layer and the barring info signaling on the AS layer. Depending on the network configuration, the network may determine whether certain access attempts should be allowed or blocked based on categorized criteria. Such criteria are access categories and access identities. The NG-RAN may broadcast barring control information associated with these access categories and access identities.

Such mechanisms allow congestion control even in emergencies, e.g. in a catastrophic situation, first responders typically apply plans in order to prioritize and cluster emergency messages. Another common solution could involve restriction of cell access to network operator business only for maintenance work or restriction of access to a certain kind of user or device group. 5GS defines access categories and access identities. When the UE needs to initiate an access attempt, the UE should determine one or more access identities from the set of standardized access identities and one access category from the set of standardized access categories. One important aspect here is that both the access categories and access identities are valid per access attempt; thus, they are not to be understood as static identifiers. To complete this procedure and allow certain operator-specific policies as well, there are operator-defined access categories to be associated with a given access attempt. TS 24.501 provides further details on these standardized access identities and categories along with what access events will trigger an access check before the attempt. Access identities could be priority settings such as a UE belonging to a mission-critical access group of devices, or a UE that has multimedia priority service (MPS) or belongs to one of the access classes 11 – 14. As access categories, TS 24.501 defines situations such as MT or MO voice or video calls, IMS signaling, access attempt for delay tolerant traffic, etc. A UE requesting emergency services uses access category 2 or access category 8 if there is a resume procedure for an emergency call. In the downlink direction, SIB1 indicates restriction parameters via the IE *uac-BarringInfo*. To allow higher flexibility, barring information can be configured to be valid for common access attempts via the IE *uac-BarringForCommon*, to be valid per PLMN via the IE *uac-BarringPerPLMN* or valid per access category and access identity via the *IE uac-BarringInfoSetList*. 5G NR introduces a more elaborate barring mechanism where a certain access category is combined with barring information. Furthermore, it is possible to provide access policies to a certain access identity within an access category. TS 38.331 defines the details of such barring information. The network can configure multiple barring sets of this kind and combine them with the access category for which this information becomes valid via an index parameter. Furthermore, the parameter *uac-BarringForAccessIdentity* allows greater barring granularity since it supports barring policy definition per access category and per access identity. The barring method itself describes a uac-BarringFactor as a probability factor that the access attempt would be allowed during that access barring check. If such an access attempt is barred, the UE waits a *uac-BarringTime* (range 4 – 512 seconds) until a new access attempt is started. Since emergency service support depends on the RAT itself, there is a possible restriction in mobility procedures like inter-RAT handover or dual connectivity in case an emergency call is ongoing.

Beside emergency services and uniform access control, TS 22.153 specifies the service requirements for multimedia priority services (MPS). MPS allows priority access to 5GS to service users such as government-authorized personnel or emergency management officials.

2.8.3.2 Emergency services in 5GS

5GS may support emergency services directly, which requires 5GC to be connected to IMS since the connection will be an IMS emergency call and the 5G RAN supports emergency bearer services. If 5GS offers emergency services, it will indicate the permission to UEs in limited registration via the *ims-EmergencySupport* flag in SIB1 and via the registration accept message, valid per RAT and registration area, to UEs that are properly 5GMM registered. 5GS emergency service support behaves similar to VoNR; IMS is the control backend coordinating the SIP message flow and providing a PDU connection with the required QoS. RRC protocol messages indicate an emergency situation to establish an early PDU connection to an IP-CAN or data network and IP is the transfer protocol for emergency-related information like voice and optional assistance information like the UE location, etc. A high-level signaling procedure for an IMS emergency session is illustrated in the following figure:

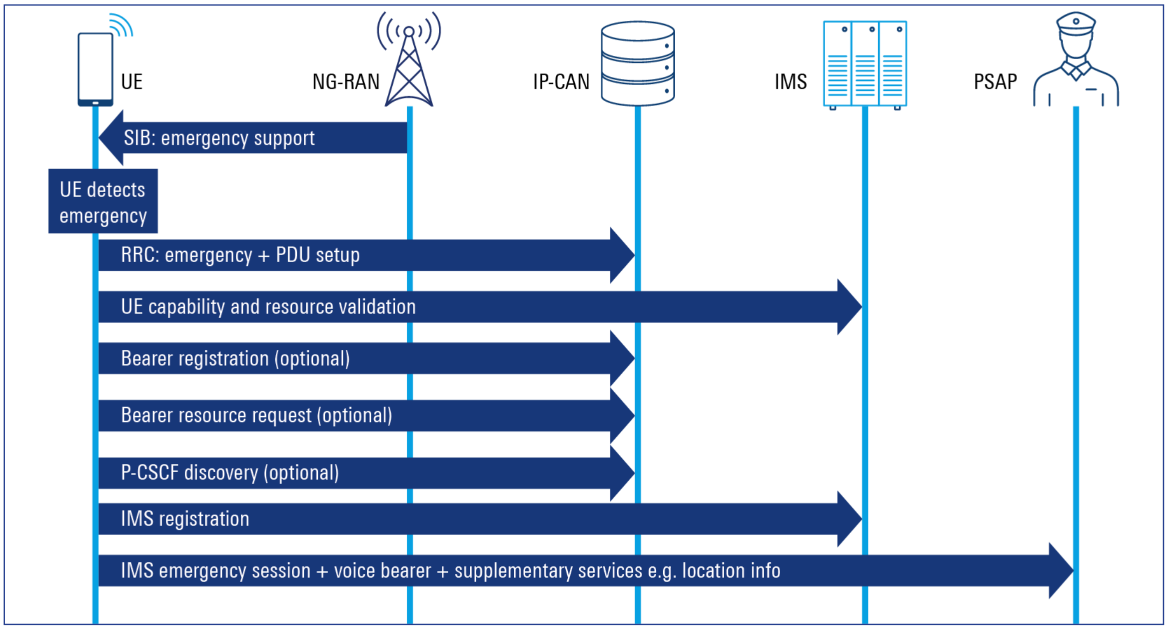

Figure 27: IMS emergency services in 5GS - high-level message flow.

As a prerequisite, we assume the proper emergency support indication from the network, either via SIB broadcast for UEs in a limited registration state or via registration accept to UEs in a fully 5GMM registered state. It is the responsibility of the UE to recognize and detect the emergency situation based on the dialed number, i.e. via dialing the known emergency numbers 112, 999, 911 or via explicit GUI selection. Next, the UE starts an SRB and DRB connection. The difference compared to a standard PDU connection is that here, there is RRC signaling of an emergency to obtain higher priority and optionally, some security procedures like UE authentication and ciphering enabling are skipped to reduce the call setup time. Unlikely but as an optional step, a UE that recognizes that it has insufficient resources or capabilities to establish an emergency call should terminate ongoing communications and suspend the radio connection. The objective is to set up a PDU connection to a so-called IP-CAN, i.e. an IP-based access network. The IP-CAN should support emergency bearers; otherwise, TS 24.229 defines alternative behavior for the UE in this unlikely situation. Most commonly, this emergency bearer is a 5G PDU bearer connection with default settings and optionally, the UE may need to register, i.e. obtain an IP address and a QoS flow and secondly (also optional), the UE may negotiate the resources with the IP-CAN. The latter case is typically the QoS flow, but in 5GS, emergency services use a default QoS flow. The IP-CAN should be aware of the emergency call QoS and reserve as well as guarantee the required bearer resources. Like in an IMS connection, the first signaling procedure is the P-CSCF discovery. Either the UE obtains the IP address of the P-CSCF via the 5GS bearer setup (most likely in emergency situations) or the UE has to start the P-CSCF query as in the standard IMS connection setup.

If the UE has sufficient credentials to authenticate with the IMS network, it should initiate an IMS emergency registration by providing its allocated IP address to the P-CSCF. The IMS registration request should include an emergency indication. TS 24.229 describes the details for when the UE needs to perform an IMS emergency registration or not; both possibilities exist. The P-CSCF establishes a SIP connection to S-CSCF like in a standard IMS message flow. Note that for emergency services, the S-CSCF is also called the E-CSCF as the emergency service function. It is up to the implementation whether there is a difference between S-CSCF and E-CSCF, or they are a common entity and the use case distinguishes between the roles. The last step is that the E-CSCF establishes an IP voice connection to a public safety answering point (PSAP). It is up to local regulations to decide who behaves as a public answering point for emergency sessions. Most common are local police stations or ambulances / firefighters, but also specific help desk functions such as an emergency service help desk with a facility running their own SNPN are possible. Thus, as the final step, the UE should initiate IMS emergency session establishment using the IMS session establishment procedures containing an emergency service indication and optionally any registered public user identifier in case the UE has performed a previous emergency registration. Note that this describes a high-level message flow for IMS emergency call setup. Some of the message signaling steps are optional.

Further signaling details for an IMS emergency connection are discussed in the following section. There is the option for an emergency registration procedure that is further described in TS 24.501. A UE triggering such an emergency registration indicates within the registration request message the "emergency" purpose in the registration type IE. The AMF may then skip a UE identity request and authentication procedure, depending on local regulations. When contemplating an emergency connection, the registration procedure, especially the network-oriented registration accept message, provides details on what kind of emergency service is supported by the network. With the 5GS network feature support information element contained in the registration accept message container, the network informs the UE about support for specific features, such as IMS voice over PS session, location services (5G-LCS), emergency services or emergency services fallback. When initiating an emergency call, the UE upper layers take the IMS voice over PS session indicator, the emergency service support indicator and the emergency service fallback indicator into account for access domain selection. First, the IMS voice over PS session over 3GPP access indicator flag signals whether IMS voice services are supported in 5GS directly or not. Second, with the emergency service flags EMC and EMF, the network may describe details of emergency service support within 5GS. The emergency service support indicator for 3GPP access (EMC) indicates the network status as one of four options:

- 5GS does not support emergency services
- Emergency services are possible when NR is connected to 5GCN
- Emergency services are possible when EUTRA is connected to 5GCN
- Emergency services are possible when EUTRA and NR are connected to 5GCN

The emergency services fallback indicator for 3GPP access (EMF) indicates the network status as one of four options:

- 5GS does not support emergency services fallback
- Emergency services fallback is possible when NR is connected to 5GCN
- Emergency services fallback is possible when EUTRA is connected to 5GCN
- Emergency services fallback is possible when EUTRA and NR are connected to 5GCN

5GS divides support for emergency services into four different behaviors, similar to IMS emergency service support in EPC (TS 23.401). Thus, emergency services are offered to UEs in valid registration only, to UEs with valid IMSI authentication only, to UEs that provide their IMSI identity, or to all UEs (even anonymous access).

Within the 5GS entities, provision of emergency services requires storage of configuration data within the various functions. The first access entity is the AMF which needs to contain emergency configuration data such as the slice selector identity S-NSSAI and the address of the emergency DNN or IP-CAN in Fig. 27 above which is used to derive an SMF. The latter can also be statically configured. Optionally, the AMF may also store statically configured UPF information. Typically, the emergency service support permission parameters and emergency service configuration parameters (e.g. PSAP info and optionally the address of a location resource function LRF if the network needs to provide the UE location) in the AMF are handled according to local regulations. One common scenario is that emergency services are offered within the entire PLMN. In such a case, every AMF and at least one SMF should contain the emergency service configuration data.
There are no specific architecture and protocol stack requirements for emergency service provision.

There is no direct mechanism to control QoS in an emergency service since the requirements range from unauthorized UE access through UEs with subscription to only emergency services. Therefore, the initial QoS parameters used for establishing emergency services are configured in the V-SMF (local network) in the SMF emergency configuration data. The QoS flows of a PDU session associated with the emergency DNN are dedicated to the IMS emergency session only.

In 5GS, the flow and initiation of an IMS emergency session follow the same principles and concept previously defined in LTE. TS 24.229 defines details such as different connection criteria when the UE is aware of being in its home network, meaning it may skip certain security-related credentials or use default bearers and address configurations. As a radio technology agnostic management system, IMS provides supplementary services enhancing the information flow in combination with emergency services. It is possible for the RAN (as the AS) to recognize the emergency situation within an ongoing connection and trigger generation of emergency signaling towards the IMS. As requested by local authorities, location information such as GNSS positioning may be included in the SIP control messages.

Following the high-level message flow in Fig. 27, a more detailed signaling message flow is shown in Fig. 27. Due to the complexity, we omit the establishment of a 5G PDU connection and emphasize the IMS signaling aspects with a focus on emergency services only. As the first step, it is the responsibility of the UE to recognize the dialed number (e.g. 112, 119, 911 known as E911, 999) as an emergency number. Some operators even recognize a kind of "panic" number such as 9999 or 911111 as an emergency. Consequently, the UE sets the top-level services "SOS" within the universal resource name (URN) signaled to the IMS. The P-CSCF is the IMS entity responsible for recognizing emergency calls and prioritizing them over normal calls. To check whether an initial request is part of an emergency call, the P-CSCF compares the Request URI to the emergency dial string (e.g. 112, 911), the *Emergency Universal Resource Name* (URN) and known PSAP SIP or TEL URIs [Ref. 4]. Beforehand, the UE may execute an initial emergency registration via the REGISTER message to the IMS. The emergency registration follows the general registration procedure in IMS as defined in TS 24.229 with these specific settings: In the contact header field of the SIP URI parameter, the UE includes the term "sos" and in the "to" and "from" fields within the REGISTER message, the UE includes the public user identity (either the default value, a temporary value or a value obtained in a previous normal registration). To start emergency session setup via the INVITE message, the UE sets the URN parameter to high-level service type "sos" with optionally detailed service URNs, e.g. "urn.service.sos.ambulance" if the UE would be aware that it is a medical emergency. TS 24.229 provides more details on UE behavior regarding emergency connection setup without previous registration. For example, the UE uses an anonymous ID in the INVITE, sets the URN to emergency as described in the registration procedure and also includes its own IP address obtained from the IP-CAN. If available, the UE may include geolocation information in the INVITE message. A similar behavior is defined in cases where the UE performs an emergency registration before the connection setup. The only difference is the placing of the public user ID used during the emergency registration in the INVITE message. There is also the possibility to start an emergency connection setup within a non-emergency registration. The difference here is that the UE includes its identity such as a telephone number instead of a public identity. Further details on timer settings, abnormal emergency connection release or network-initiated deregistration, etc. are provided in TS 24.229.

An example of an emergency reject would look as follows: If the P-CSCF recognizes an emergency number or urn:service:sos, it responds back with 380 Alternative Service and an XML body which contains the <ims-3gpp> element with the <alternative-service>.

A successful IMS emergency flow is shown in the figure below. First, the UE identifies the dialed number as an emergency and sets urn:service:sos as the emergency identity signaled to the P-CSCF [Ref. 17]. The P-CSCF handles the registration requests with an emergency public user identifier, detects and prioritizes the emergency session and selects an emergency-CSCF (E-CSCF) in the same network to handle the emergency session request. The emergency-CSCF (E-CSCF) receives an emergency session establishment request from a P-CSCF and optionally requests the location resource function (LRF) to retrieve the UE location information. It is the responsibility of the E-CSCF to route emergency session establishment requests to an appropriate destination such as a PSAP like a local police station or an ambulance, including anonymous session establishment requests.

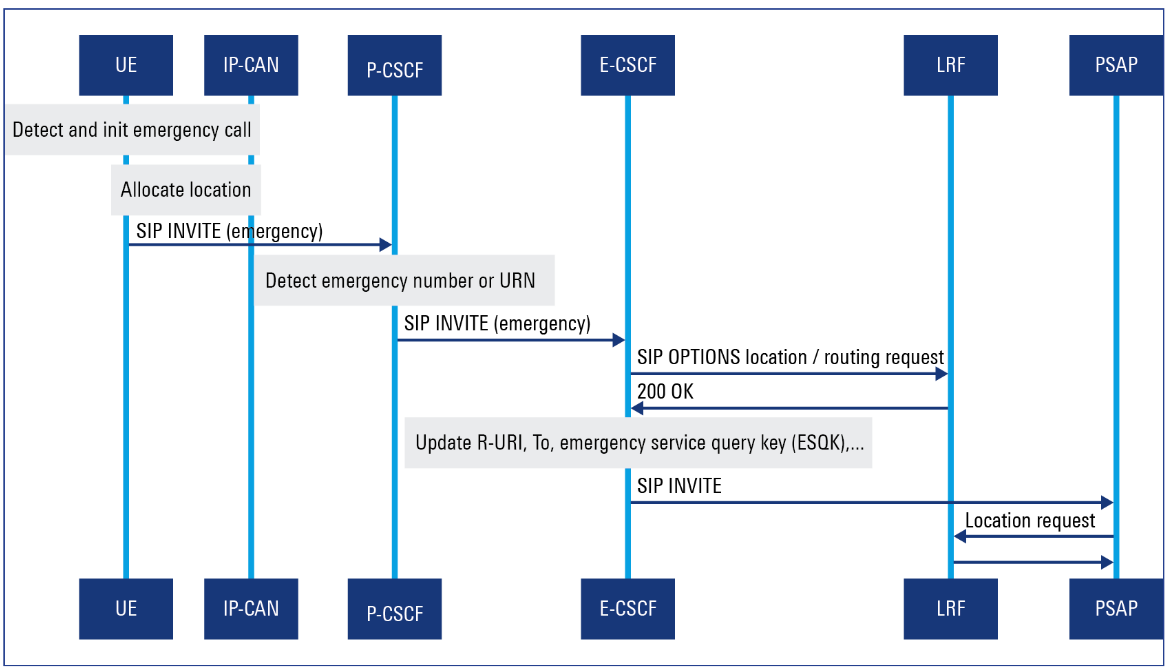

Figure 28: IMS emergency call sequence.

2.8.3.3 Emergency services fallback (ES-FB)

EPS emergency services fallback is an early handover of an emergency connection to EUTRAN. This solution is new in 5GS, i.e. there is no equivalent in 4G. Similar to the fallback procedures for normal voice and depending on the core network and RAT support, the fallback can either be a RAT fallback so only the RAT changes from NR to EUTRA, or it can be a system fallback where the fallback goes from 5GC to EPC. During the registration accept procedure, the network indicates the support for emergency services fallback to the UE. The UE indicates the service request "emergency" towards the AMF if support for emergency services has been indicated by the AMF in the registration accept message to the UE. Upon receiving the service request for emergency, the AMF interacts with gNB to perform EPS fallback. In this solution, there is no need for NR and 5GC to support emergency features other than what is needed for the service request for emergency handling. However, voice over NR (possibly including EPS fallback) has to be supported on NR and 5GC; otherwise, the UE would not even camp on NR.

If the 5GC has indicated emergency services fallback support for the routing area and RAT where the UE is currently camping and if the UE supports emergency services fallback, the UE should initiate the signaling message flow shown in Fig. 29. The architectural concept of EPS emergency fallback is depicted in this same figure:

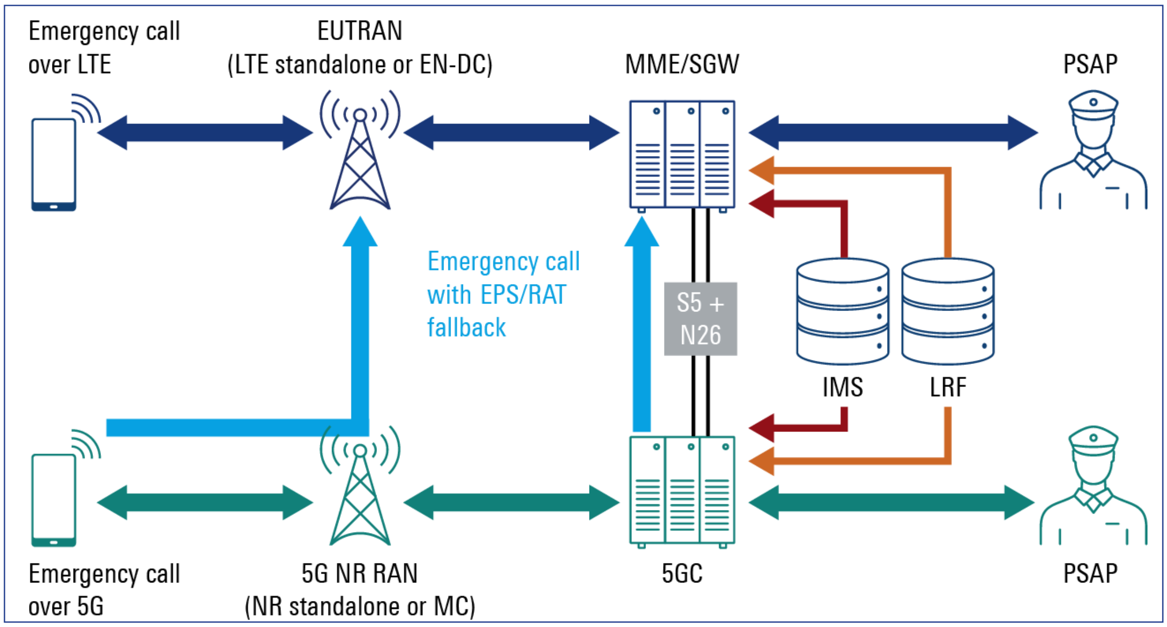

Figure 29: Emergency services fallback to EPS architecture [Ref. 1].

The architectural background and circumstances decide which scenario is applied, i.e. the first issue would be to check which core network exists (5GC or EPC) and whether emergency services are supported, and additionally which radio access technology supports emergency services. The objective is to support various deployment scenarios for emergency services. The first situation is that the UE camps on 5G NR and the NR RAT does not support emergency services. This will trigger a RAT fallback as described in section 2.4, but only for emergency services instead of standard voice; thus, the IMS signaling is according to emergency services. The second situation can be a deployment including 5GC, but according to our assumption, 5GC does not support emergency services and a fallback to EPS is initiated as described in section 2.2.4.

Upon the QoS flow establishment request for emergency services, the network triggers either the EPS fallback procedure shown in Fig. 8 or the RAT fallback procedure shown in Fig.13.

The emergency services fallback procedure follows the signaling flow shown in Fig.15. The assumption is based on prior signaling of emergency services fallback support to the UE to allow it to camp on the supporting RAT (either EUTRA or NR). It is the responsibility of the UE to recognize the emergency situation (as described previously) based on the dialed number. With the service request message to the AMF, the UE signals the request for emergency services fallback. 5GC triggers the emergency fallback by executing an NG-AP procedure in which it indicates to NG-RAN that this is a fallback for emergency services. Depending on the core network availability and its support for emergency services, the AMF indicates the target core network to the RAN node. This distinguishes whether the handover is inter-RAT fallback or inter-system fallback. The AMF may forward the security context for UEs that are already successfully authenticated and registered to the corresponding RAN. Based on the target core network being either 5GC or EPC, the NG-RAN initiates one of the following procedures (TS 23.502):

- Handover or redirection to EUTRAN cell connected to 5GC if the UE is currently camping on NR
- Handover or redirection to EUTRAN connected to EPS. NG-RAN uses the security context provided by the AMF to secure the redirection procedure

If the redirection procedure is used, the network signals the target core network type (EPC or 5GC) in order to trigger the appropriate NAS procedures (S1 or N1 mode).
The final step in the signaling message flow after handover or redirection to the target cell is that the UE establishes a PDU session for IMS emergency services (see Fig. 29).

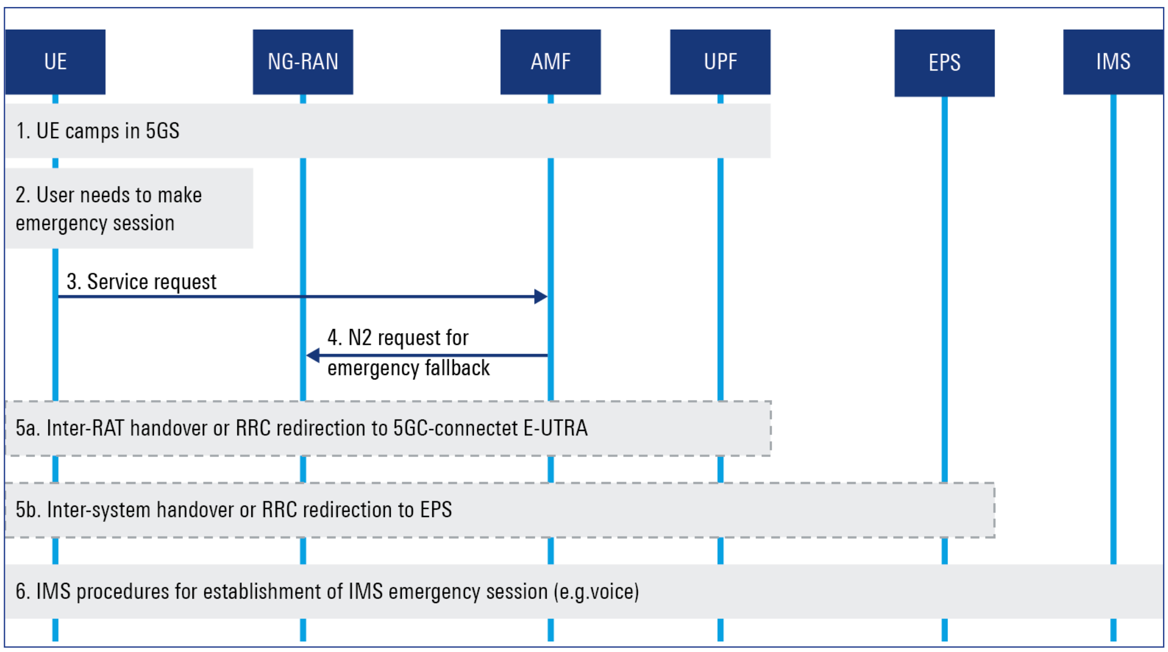

Figure 30: Emergency services fallback to EPS - message flow.

2.8.4 Automotive eCall in 5G

In May 2010, the UN General Assembly proclaimed the period of 2011 to 2020 as the decade of action for road safety. The European Union's contribution to this effort was to design eCall, an automatic emergency call system. There are several European standards by the ETSI summarized under the term eCall. Just to name one of them, CEN EN 16072 defines the Pan-European eCall operating requirements [Ref. 4].

As eCall is based on the GSM radio technology and the lifecycle of a new car is in the range of > 10 years, 3GPP specified in later releases the eCall communications over UMTS. To ensure a future-proof technology over a much longer period, eCall evolves with 3GPP Release 14 into the next generation eCall (NGeCall) [Ref. 5]. The general working principle stays untouched but as a major difference, the radio access technology used to convey the emergency minimum set of data (MSD) is LTE instead of GSM. Thus, the voice call moves from circuit-switched to packet-switched and from single voice call to IP data.

NGeCall requires voice services offered by LTE (VoLTE) connected to IMS. With the IMS system, the minimum set of data (MSD) is sent in a SIP invite message during call setup. As 5G enters the communications world with its new radio system and to guarantee the longevity of the eCall automotive emergency communications system, 5GS must support the eCall type of emergency services. The SIB1 contains a network indicator flag *eCallOverIMS-Support* to indicate that the network supports eCall over IMS. Note that this flag is defined since 3GPP Release 14 and would also be included in the system information provided by EUTRAN.

The current standard mandatorily requires the existence of a USIM in the UE; anonymous access is not allowed. The same applies for eCall over a non-3GPP radio access network. There is the possibility of a device supporting eCall-only mode, e.g. a vehicle-based emergency telematic control unit (TCU) modem. Such a UE configured for eCall-only mode should remain in the RM-DEREGISTERED state, should camp on a network cell when available but should refrain from any registration management, connection management or other signaling with the network [TS 23.501]. The UE may instigate registration management and connection management procedures in order to establish, maintain and release an eCall over IMS session.

The dispatching of the eCall connection on either 5G NR RAT or EUTRA RAT depends on the support of both radio access technologies, the support of EPC and 5GC and the interconnection among them. Depending on these circumstances and the eCall over IMS support flag, an eCall emergency session can be routed via 5G NR and 5GC to the IMS or there can be a fallback to EUTRA with routing via either 5GC or EPC to IMS. There is no difference between the general message flow of an LTE-based NGeCall or a 5G NR-based NGeCall. Assuming the prerequisites are fulfilled (5G NR supporting IMS), Fig. 31 shows a general IMS SIP message flow for such an emergency call.

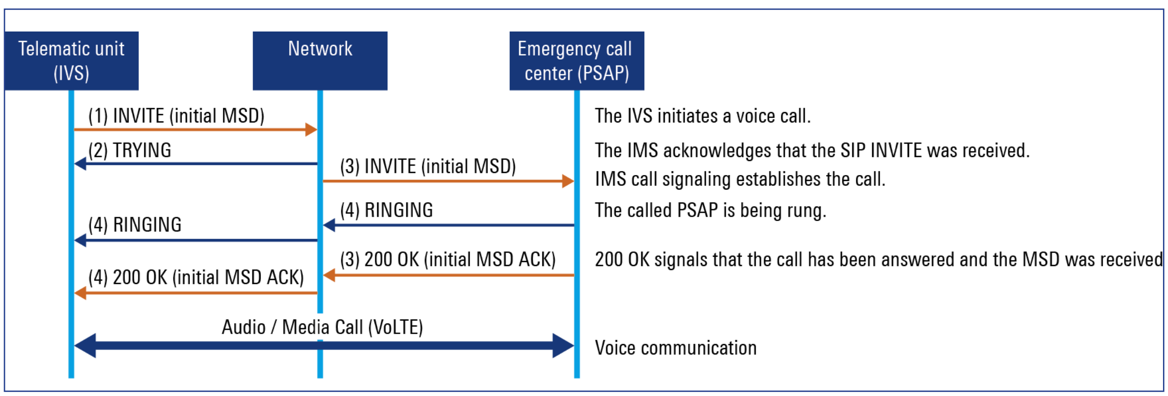

Figure 31: SIP message flow for NGeCall.

Details within the message flow between an emergency call triggering telematic control unit and the emergency call center (e.g. PSAP) show the RAT-agnostic behavior of the SIP message content. One optional SIP feature as a difference can be the signaled URN; here, the UE can set the parameter "urn:service:sos.ecall.automatic" if this is an automatically generated eCall or "urn:service:sos.ecall.manual" if this is a manually triggered eCall. Once the emergency session is established and the minimum set of data (MSD) is exchanged, the PSAP operator has the option to establish a voice channel that is answered by the telematics control unit automatically. NGeCall extends this feature so the PSAP can request transmission of the MSD once again without interruption of the voice communications channel. In a major difference between eCall, as defined in the first setup based on GSM, and NGeCall, the MSD is no longer sent as inband circuit-switched modem data but is included in the SIP message directly. The 140 bytes of information, including the position information for the caller UE, is transferred via SIP [Ref. 18]. The following screenshot, taken from the R&S®CMW500 wideband radio communication tester, provides more signaling message details for the SIP INVITE message establishing an NGeCall emergency connection.

```
INVITE urn:service:sos.ecall.manual SIP/2.0                                   → INVITE with uniform resource name (urn)
From: <tel:1220>;tag=2883044717                                               → Involved participants in the call "from/to"
To: <urn:service:sos.ecall.manual>
CSeq: 735561065 INVITE
Call-ID: 2883044713_3440660140@fc01:abab:cdcd:efe2::1
Via: SIP/2.0/TCP [fc01:abab:cdcd:efe2::1]:40589;branch=z9hG4bK1697007878
Max-Forwards: 70
Contact: <sip:425a0f9f-61c1-4c6b-aaf2-ee62552e17e0@[fc01:abab:cdcd:efe2::1]:40589;user=phone>;
+sip.instance="<urn:gsma:imei:000000000-000000-0>";+g.3gpp.icsi-ref="urn%3Aurn-7%3A3gpp-service.ims.icsi.mmtel";audio;+g.3gpp.nw-init-ussi;
+g.3gpp.mid-call;+g.3gpp.srvcc-alerting;+g.3gpp.ps2cs-srvcc-orig-pre-alerting
Route: <sip:[fc01:cafe::1]:5102;lr>
P-Access-Network-Info: 3GPP-E-UTRAN-FDD; utran-cell-id-3gpp=0010100010000100   → NG-eCall over EUTRAN
Security-Verify: ipsec-3gpp;alg=hmac-sha-1-96;q=0.900;prot=esp;mod=trans;ealg=aes-cbc;spi-c=3872527343;spi-s=2777027566;port-c=5103;port-s=5102
Proxy-Require: sec-agree; Require: sec-agree; P-Preferred-Identity: <tel:1220>
Allow: INVITE,ACK,CANCEL,BYE,UPDATE,PRACK,MESSAGE,REFER,NOTIFY,INFO,OPTIONS
Content-Type: multipart/mixed;boundary=boundary1
Accept: application/sdp,application/EmergencyCallData.control+xml
Recv-Info: EmergencyCallData.eCall.MSD
P-Preferred-Service: urn:urn-7:3gpp-service.ims.icsi.mmtel
Accept-Contact: *;+g.3gpp.icsi-ref="urn%3Aurn-7%3A3gpp-service.ims.icsi.mmtel";audio
Supported: timer,100rel,replaces,precondition,from-change,histinfo,tdialog
P-Early-Media: supported
Call-Info: <cid:user1@test.3gpp.com>;purpose=EmergencyCallData.eCall.MSD      → Message body part and body ID
Content-Length: 1151
0123456789012345678901234567890123456789012345678 90...
Session-Expires: 1800;refresher=uac                                            MSD content
Content-Type: application/sdp  --boundary 1  …session parameters … -boundary 1 → … Session description
Content-Type: application/EmergencyCallData.eCall.MSD
Content-ID: <user1@test.3gpp.com>
Content-Disposition: by-reference;handling=optional
```

Figure 32: SIP INVITE message containing details for NgeCall setup.

2.8.5 Emergency services over non-3GPP access

Since 5G in general supports radio access via non-3GPP networks, the support for emergency services is signaled to the UE during non-3GPP registration via the 5GC. There is a priority policy within 3GPP specifying that the UE should try to use 3GPP RAT for an emergency first. A UE may only attempt to use emergency services over untrusted non-3GPP access if it is unable to use emergency services over 3GPP access. TS 23.167 defines further details for the criteria to support non-3GPP emergency services. Trusted and untrusted non-3GPP access to 5GC for emergency sessions is supported.

The UE may establish an emergency session over non-3GPP access to 5GC only when 3GPP access for emergency calling is not possible or available, e.g. no coverage. The emergency service support for non-3GPP access indicator (EMCN3) within the registration accept message tells the UEs whether emergency calls over non-3GPP access technologies such as WLAN are possible.

The procedure is similar to the described procedure for emergency calling over 5G NR or EUTRA. First, the UE needs to identify and recognize the dialed number as an emergency number. For the specific case where the UE has selected to make an emergency call and the radio access network is non-3GPP access to 5GC, TS 23.167 defines some action items the UE has to perform:

- ► The UE should establish an emergency PDU session over non-3GPP and should perform an IMS emergency registration before sending an IMS emergency session request
- ► The UE should include any available location information in the IMS emergency session request
- ► For the media supported during IMS emergency sessions, standardized media codecs and formats should be applied

A UE should not establish an emergency PDU session over non-3GPP access if the UE has initiated and successfully established an emergency session via 3GPP RAT.

In a high-level protocol procedure, the UE should signal IMS emergency as establishment cause over the non-3GPP RAT and the IMS should reject any connection request without such an establishment cause from the UE. Further details are provided in TS 23.167 and TS 23.502.

2.9 Signaling aspects for voice support in 5G

The objective of this brief section is to summarize and provide an overview of the signaling parameters on the network side and UE side in AS and NAS that are relevant in relation to voice support. The network signals the support for voice and emergency calls via system information broadcast and via NAS signaling during the registration procedure. This split between AS and NAS signaling has the advantage that services like voice or emergency can be indicated in a UE-specific manner, valid per radio access technology and per registration area. The objective is to react flexibly to various UE types and reduce the signaling impact.

System information broadcast parameters relevant for voice aspects

The system information block SIB1, being part of the initial BWP, occurs at the highest repetition rate and signals the emergency call support parameters ims-EmergencySupport, *eCallOver-IMS-Suppport* and *UAC-Barring* information. The support for emergency and cell barring information is indicated to the UE at the earliest possible moment. Since the network may offer emergency calling to UEs in a limited registration state, the SIB broadcast is the way forward to provide such information to the UE, and cell barring is needed at that stage to allow network congestion control. In addition, SIB1 contains UAC access category assistance information enabling the UE to determine whether it belongs to access category 1, representing delay tolerant traffic and optionally deferring the UE radio access attempts. The barring policies and the corresponding information parameters within the system information are explained in section 2.8.3.1 and in TS 38.331. While LTE SIB2 allows transmission of barring information related to voice calls, 5G NR uses a slightly different barring concept. On the radio interface access, attempts can be barred with information parameters valid for common access, valid per PLMN or valid per access category or access identity. To allow higher flexibility, some access policies are decoupled from the AS layer and shifted to the NAS layer. Thus, support indication such as for voice services is provided during the UE registration.

UE capability indicator

Obviously, the UE needs to indicate its capability to the network during the attach procedure and this has further impact on the UE and network behavior concerning service and domain selection. In general, the UE capability transfers represent a very complex aspect in 5G, e.g. some of the UE features are clustered within feature groups and there are some dependency requirements across the different features. Due to the complexity, this section only presents the UE capabilities relevant for voice. The parameters UE usage setting as voice- or data-centric, the availability of IMS support and the UE operating in single- or dual-registration influence the behavior of the UE for domain selection. In the present situation, we understand the signaled condition “the UE supports IMS voice over 3GPP access” if one of the following is fulfilled (TS 24.501):

1. The UE supports IMS voice over NR connected to 5GCN
2. The UE supports IMS voice over E-UTRA connected to 5GCN, or
3. The UE supports IMS voice in EPS

From among the many UE capability parameters defined in TS 38.306 and signaled via the RRC message *UECapabilityInformation*, the IMS-parameters field is one of the most important in relation to UE voice service support.

- *voiceOverEUTRA-5GC* indicates whether the UE supports IMS voice over E-UTRA via 5GC. This support also includes IMS voice over EUTRA connected to EPC, i.e. VoLTE
- voiceOverSCG-BearerEUTRA-5GC indicates whether the UE supports IMS voice over SCG bearer of NE-DC. This is a future-oriented parameter; in the current release IMS voice over split bearer is not supported, neither NR-DC nor NE-DC
- voiceFallbackIndicationEPS indicates the support for EPS fallback. If this field is included, the UE should support IMS voice over NR and IMS voice over E-UTRA via EPC
- voiceOverNR indicates whether the UE supports IMS voice over NR. It is mandatory for the IMS voice capable UE in NR, and otherwise optional. This parameter represents a backward-compatible capability; VoNR support includes support for VoLTE as long the UE supports EUTRA

Although not directly an AS-related UE capability, TS 38.306 requests for proper VoNR support that the UE PDCP capabilities support the ROHC profiles 0x0000, 0x0001 and 0x0002 to perform proper compression of PDCP SDUs containing voice to correspond to the correct voice codec. As stated, one conditionally mandatory UE capability feature is support for *IMS emergency calls* in case the UE supports voice services in NR.

RRC signaling control messages for voice

While this section cannot delve into the RRC signaling details in 5G NR, we wish to provide a concise presentation of the signaling messages that are relevant for voice. In this context, the first and most important RRC message is the *RRCSetupRequest* message containing the information element establishment cause. In the current release, the UE can signal the following reasons why a connection is needed to the network: emergency, high priority access, MPS or MCS priority access, mobile terminated access and mobile originated signaling, data transfer, SMS or voice. Similar reasons can be signaled by the *RRCResume* message in a failure case when the connection should be resumed. The EPS fallback procedure explained in section 2.2.4 requires the release of the NR radio link via the *RRCRelease* message and the signaling of its cause. To trigger a redirection to EUTRA, the 5G RAN signals the *voiceFallbackIndication* as the release cause. As an alternative to connection release and redirection to EUTRA, the 5G RAN can directly send the RRC *MobilityFromNRCommand* with the *voiceFallbackIndication*.

NAS relevant network features

To allow a higher degree of flexibility, 5GS provides an indication of the voice service support not only in the AS layer but also in the NAS during the registration procedure. The objective is to provide higher granularity in feature support indication and in access restrictions. At least from an abstract point of view, the support for voice services can be realized on a per UE basis. The access restriction described previously is valid per access attempt, but the UAC mechanisms are intended to avoid network congestion while the feature signaling is intended to indicate the supported services. Using the 5GS network feature support information element, the network indicates support for certain features to the UE. It is contained in the registration accept message and the IE structure is given as (TS 24.501):

<table>
<tr><td colspan="8">5GS network feature support IEI</td><td>octet 1</td></tr>
<tr><td colspan="8">length of 5GS network feature support contents</td><td>octet 2</td></tr>
<tr><td>MPSI</td><td>IWK N26</td><td colspan="2">EMF</td><td colspan="2">EMC</td><td>IMS-VoPS-N3GPP</td><td>IMS-VoPS-3GPP</td><td>octet 3</td></tr>
<tr><td>5G-UP CIoT</td><td>5G-HC-CP CIoT</td><td>N3 data</td><td>5G-IPHC-CP-CIoT</td><td colspan="2">Restrict EC</td><td>MCSI</td><td>EMCN3</td><td>octet 4</td></tr>
<tr><td>0 spare</td><td>0 spare</td><td>0 spare</td><td>0 spare</td><td>0 spare</td><td>5G-EHC-CP CIoT</td><td>ATS-IND</td><td>5G-LCS</td><td>octet 5</td></tr>
</table>

Figure 33: 5GS network feature support for voice aspects.

Due to space and complexity constraints, only network features that have a certain relevance to voice and emergency services are mentioned here:

- MPSI indicates whether the network supports multimedia priority services (MPS) or not. The content influences whether the UE may use access identity = 1 or not
- IWKN26 is not a direct voice related indicator but it indicates whether the network supports interworking without the N26 interface between EPC and 5GC. It has an effect on network selection and camping of the UE and how signaling message parameters should be set during call establishment
- EMF as the emergency services fallback indicator to indicate emergency fallback support
- EMC as the emergency support indicator
- IMS-VoPS-3GPP is a very important flag to indicate support for voice over IMS services in the 5G network. Note that this also includes support for EPS or RAT fallback voice support, but not necessarily full VoNR support
- IMS-VoPS-N3GPP indicates that the network supports voice over non-3GPP
- MCSI is the indicator bit whether a UE can use access identity 2 representing mission critical services

Information on additional signaled parameters within the network feature support can be found in TS 24.501.

3 SUMMARY

Voice services are not just some legacy burden that sits on the back of 5G and prevents it from realizing its full potential. Voice in 5G networks will be an active part in delivering a host of key data services well into the future.

Of course, mobile subscribers will continue to demand voice as part of their services. But because of the integral role voice will play in the emerging generation of data services, voice services will play a vital role in the packages offered by service providers.

While standalone (SA) 5G deployments are increasing, these deployments are often launched alongside non-standalone (NSA) networks, indicating that there remains some hesitancy among operators to fully commit to SA 5G networks. This will continue to offer some complications to the deployment of voice services over these 5G networks. However, there are some key circumstances in place that will ease the deployment of voice over 5G.

VoLTE is a Critical Enabler of 5G VoNR

Still today LTE is the most widely used mobile radio technology, connecting more than half of all mobile users across the globe [Ref. 21]. Circuit switched fallback (CSFB) has provided a first solution for most service providers to prioritize 4G services using LTE and remaining the offer of voice services by incorporating a handover to 3G circuit switched voice only for this particular application. It was considered an interim step before VoLTE became more mature. However, service providers are increasingly adopting voice over LTE (VoLTE) in their networks, using the IP multimedia subsystem (IMS) as the management and orchestration core. This use of IMS has provided a key support for voice over 5G.

While VoLTE continues to be invested in and deployed for voice services, investment in 5G is accelerating. This expansion of 5G deployments foretells big changes for many voice service providers going forward. This is because service providers will no longer have the option of running LTE that's capable of running voice either with or without IMS as the core. 5G VoNR is only possible to operate by migrating to an IMS-based core network.

IMS Adoption is Growing

Voice over 5G uses the same IMS-based architecture as VoLTE. IMS enables the convergence of, and access to, real time services (voice, video), streaming, messaging, data and web-based technologies for the wireless user. The IMS system enables an easy-to-maintain and highly scalable IP multimedia communication network. IMS is access independent and provides a peer-to-peer addressing architecture for IP-based voice sessions, end-to-end Quality of Service (QoS), charging, authentication and security (TS 22.228).

Preparing the VoLTE network for 5G

It is neither technically nor economically feasible to launch a 5G network with nationwide full coverage from day one, capable of carrying voice over it. It has always been part of the plans for 5G's rollout that it would leverage existing networks. The IMS-based architecture of VoLTE provides a good starting point for 5G. It continues the reuse of existing infrastructure and methodologies but allows the future-proof introduction of new technology aspects.

In large part, this evolution stems from the fact the optimization of a network and its constituent parts need to be done in coordination rather than isolation, or it risks be ineffective. A coordinated optimization of everything from the EPC to the IMS is needed to ensure a good customer experience.

The Evolution of IMS in VoNR

In the initial Release 15 (Rel-15) of 5G there was an attempt to minimize the impact on IMS in order to assist deployment. In fact, from the perspective of IMS, Rel-15 looked essentially like a 4G EPC. The IMS used for VoLTE lays the groundwork for 5G VoNR. However, by Release 16 (Rel-16) the IMS becomes part of the 5GC and Service Based Architecture (SBA).

Options for deploying voice over 5G

This white paper has gone into extraordinary detail on all the factors and options in deploying voice over 5G networks. However, in general terms, it's possible to break down these deployment scenarios into option 3 and option 2.

Option 3 is probably the most widely used deployment scenario since it involves using existing 4G VoLTE networks (and in some cases 3G networks). In this case, the actual 5G connection is only used for data.

Option 2 can only be used where a 5G core is already deployed. In this deployment option voice services are provided in two ways. The first method is the EPS fallback in which a mobile phone will be forced to fallback from 5G to LTE as the call is first set up. The voice and data traffic will be carried over the LTE network during the entire call. The second method uses VoNR in which the voice is handled entirely by 5G, making possible simultaneous voice and high-speed 5G data services.

No matter which deployment option is used they both rely on IMS architecture as the basis for providing voice and video communication over a 5G network. Even in deployment options 4, 5 and 7, the same voice solutions found in option 2 still apply. All the architectural approaches that use the 5G core will look like option 2.

Emergency Services

In option 3, emergency service calls will continue to be handled using VoLTE. In option 2, it becomes possible to handle these emergency calls with either VoNR or EPS fallback. There is also a third approach, dubbed the Emergency Call Fallback, in which the fallback to LTE occurs before the IMS call setup.

In this Emergency Call Fallback, the UE operates using 5G NR. When the user initiates the call, the UE detects it and recognizes that the emergency service is not supported by the radio-core combination being used. In this case, the UE sends a special request to the Access and mobility Management Function (AMF). The AMF then forwards this request to the gNodeB (gNB) that signals it to turn it over to EPS and LTE. Not until the UE falls back to LTE is the emergency call initiated. This all means that the infrastructure for IMS emergency calls from LTE can be sued with option 2 deployments.

Conclusion

Option 3 deployments have enabled voice services to continue as either an IMS/VoLTE network or a Circuit-Switched Fallback (CSFB). With option 2, voice services for 5G UE must be an IMS-based network that supports VoLTE. This also means that a VoLTE solution is a fundamental requirement for the introduction of a 5G core.

4 REFERENCES

Ref. 1 Ericsson whitepaper, 5G voice – network evolution aspects. Paper 1-5, 2019
Ref. 2 Huawei, Vo5G technical white paper, July 2018
Ref. 3 Evolved HD voice for LTE, white paper, Ericsson, UEN 284 23-3160 Rev B | October 2014
Ref. 4 eCall conformance and performance testing (1MA241), Rohde & Schwarz application note, B. Schulz, 2016, www.rohde-schwarz.com/appnote/1MA241
Ref. 5 Next generation eCall conformance testing (GF-M312), Rohde & Schwarz application note, 2019, www.rohde-schwarz.com/appnote/GFM312
Ref. 6 Voice and SMS in LTE, white paper (1MA197), Rohde&Schwarz, C. Gessner, O. Gerlach, 2011
Ref. 7 UCC 5G SMF configuration and administration guide, Cisco, 2020, https://www.cisco.com/c/en/us/td/docs/wireless/ucc/smf/b_SMF.html
Ref. 8 GSMA NG.114 IMS profile for voice, video and messaging over 5GS, August 2020
Ref. 9 ETSI TS 185 003 V2.0.0 (2008-03), Telecommunications and internet converged services and protocols for advanced networking (TISPAN); Customer network gateway (CNG) architecture and reference points
Ref. 10 GSMA PRD RCC.07- Rich communications suite 9.0, advanced communications services and client specification
Ref. 11 Common channel signaling system no. 7 (i.e., SS7 or C7). ITU recommendation Q.700, 1993
Ref. 12 RFC 3589 - Diameter command codes for 3GPP Release 5.
Ref. 13 IETF RFC 3550 – Real time transfer (RTP): A transport protocol for real-time applications
Ref. 14 IETF RFC 4825 – The eXtensible Markup Language (XML)
Ref. 15 IETF RFC 4028 Session timers in the session initiation protocol (SIP), 2005
Ref. 16 IETF RFC 7315 Private header (P-Header) extensions to the session initiation protocol (SIP) for the 3GPP, 2014
Ref. 17 IETF RFC 5031 A uniform resource name (URN) for emergency and other well-known services, 2008
Ref. 18 IETF RFC 8148 Next-generation vehicle-initiated emergency calls, 2017
Ref. 19 IETF RFC 2326 Real time streaming protocol (RTSP), 1998
Ref. 20 Kottkamp, Meik, et al. 5G New Radio: Fundamentals, Procedures, Testing Aspects. Rohde & Schwarz, 2019. Access digital copy from this page.
Ref. 21 Bugel, Jim, et al. Frederik Jejdling, Nov. 2020, pp. 1–36, Ericsson Mobility Report.
TS 22.153 Multimedia priority service
TS 22.228 Service requirements for the Internet Protocol (IP) multimedia core network subsystem (IMS); Stage 1
TS 23.228 IMS architecture, Stage 2
TS 23.167 IMS emergency call aspects
TS 23.401 General packet radio service (GPRS) enhancements for evolved universal terrestrial radio access network (E-UTRAN) access
TS 24.008 Mobile radio interface Layer 3 specification; Core network protocols; Stage 3
TS 24.167 3GPP IMS management object (MO); Stage 3
TS 24.173 IMS Multimedia telephony communication service and supplementary services; Stage 3
TS 24.229 IP multimedia call control protocol based on SIP and SDP; Stage 3
TS 24.628 Common basic communication procedures using IP multimedia (IM) core network (CN) subsystem; Protocol specification
TS 26.114 IP multimedia subsystem (IMS); Multimedia telephony; Media handling and interaction
TS 26.441 EVS general overview
TS 26.445 EVS speech coder details
TS 26.980 Multimedia telephony over IP multimedia subsystem (IMS); Media handling aspects of multi-stream multiparty conferencing for multimedia telephony service for IMS (MTSI)
TS 29.563 5G system; Home subscriber server (HSS) services for interworking with unified data management (UDM); Stage 3
TS 33.203 3G security; Access security for IP-based services
TS 23.503 Policy and charging control framework for the 5G system (5GS); Stage 2

5 ABBREVIATIONS

5GC	5G core
5GS	5G system
5QI	5G QoS indicator
AMF	Access and mobility management function
AMR	Adaptive multi-rate
AMR-WB	AMR wideband
AS	Access stratum
CSFB	Circuit-switched fallback
CSFB	Circuit-switched fallback
DNN	Data network name
DRB	Data radio bearer
DRVCC	Dual radio voice call continuity
EPC	Evolved packet core (LTE core)
ePDG	Evolved packet data gateway
EPS	Evolved packet system
EUTRAN	Evolved UMTS terrestrial radio network
EVS	Enhanced voice services
GBR	Guaranteed bit rate
GSM	Global system for mobile communications
GSMA	GSM association – www.gsacom.com
GUI	Graphical user interface
HOS	Home operator services
IETF	Internet engineering task force
IMS	IP multimedia subsystem
IP-CAN	IP connectivity access network
IVS	In-vehicle system
KPI	Key performance indicator
LRF	Location resource function
LTE	Long term evolution
MAC	Medium access control
MCS	Mission critical services
MM	Mobility management
MME	Mobility management entity
MPS	Multimedia priority services
MSRP	Message session relay protocol
MTSI	Multimedia telephony services for IMS
NAS	Non-access stratum
NSA	Non-standalone
OMA	Open mobile alliance
OTT	Over-the-top
PANI	P-access-network-info SIP header
PCC	Policy and charging control
PCF	Policy control function
PCO	Protocol configuration option
P-CSCF	Proxy call session control function
PDCP	Packet data convergence protocol
PDN	Packet data network

PDU	Packet data unit
PSAP	Public safety answering point (emergency)
QoE	Quality of experience
QoS	Quality of service
RAN	Radio access network
RAT	Radio access technology
RCS	Rich communications services
RLC	Radio link control
RRC	Radio resource control
RTCP	RTP control protocol
RTP	Real-time protocol
SA	Standalone
SDAP	Service data adaptation protocol
SDP	Session description protocol
SIP	Session initiation protocol
SM	Session management
SMF	Session management function
SNPN	Standalone non-public network
SPS	Semi-persistent scheduling
SRB	Signaling radio bearer
SRVCC	Single radio voice call continuity
UAC	Unified access control
UE	User equipment
UPF	User plane function
URN	Universal resource name
VoLTE	Voice over LTE
VoNR	Voice over new radio
ViNR	Video over new radio
XCAP	XML configuration access protocol

Regional contact

Europe, Africa, Middle East
+49 89 4129 12345
customersupport@rohde-schwarz.com

North America
1 888 TEST RSA (1 888 837 87 72)
customer.support@rsa.rohde-schwarz.com

Latin America
+1 410 910 79 88
customersupport.la@rohde-schwarz.com

Asia Pacific
+65 65 13 04 88
customersupport.asia@rohde-schwarz.com

China
+86 800 810 82 28 |+86 400 650 58 96
customersupport.china@rohde-schwarz.com

Rohde & Schwarz

The Rohde & Schwarz electronics group offers innovative solutions in the following business fields: test and measurement, broadcast and media, secure communications, cybersecurity, monitoring and network testing. Founded more than 80 years ago, the independent company which is headquartered in Munich, Germany, has an extensive sales and service network with locations in more than 70 countries.

www.rohde-schwarz.com

Rohde & Schwarz customer support
www.rohde-schwarz.com/support

Version 01.00 | March 2021
White Paper | 5G Voice over New Radio (VoNR)