TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Tarmo Tormilind 000565IATM

VOICE SERVICES IN MOBILE PACKET CORE NETWORKS

Master's thesis

Supervisor: Marika Kulmar
Master of Science

Co-supervisor: Avo Ots
Master of Science

Tallinn 2020

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Tarmo Tormilind 000565IATM

KÕNESIDE TEENUSED PAKETTSIDE PÕHISTES MOBIILSIDE TUUMVÕRKUDES

Magistritöö

Juhendaja: Marika Kulmar
Tehnikateaduste magister

Kaasjuhendaja: Avo Ots
Tehnikateaduste magister

Tallinn 2020

Author's declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Tarmo Tormilind

04.05.2020

Abstract

This Master thesis is based on theoretical materials and standards of telecommunications industry standardisation body 3GPP and Estonian mobile operator live network implementation. The study focuses on packet switched mobile core network EPC and voice services deployment options in all-IP EPC network. Traditionally voice calls have been served via circuit switched core networks in 2G and 3G. Introduction of packet-only core networks in 4G has caused fundamental change in the way voice call services are provided to end users by mobile operators. The shift has been technically and commercially challenging to mobile operators.

In the first logical part the thesis provides theoretical basis of the voice call services solutions in EPC network like Circuit Switched Fallback, Single Radio Voice Call Continuity and Voice over LTE. In the second half the thesis goes into details of the configuration aspects of EPC core network elements. The aim of the thesis is to compare three deployment options and to find the most optimum implementation path for voice services in EPC network. The thesis also contains the analysis of current configuration of the mobile operator network elements and provides suggestions to improve and optimise the configuration to achieve better end-to-end quality of the voice call.

The conclusion of the study is that VoLTE call solution is the most advanced and offers the best quality and call experience for end users. At the same time introducing VoLTE is technically complex and needs a lot of effort. Current configuration in the mobile operator network is close to optimum, is well-balanced and allows offering high-quality VoLTE calls while there is also some room for improvements.

This thesis is written in English and is 69 pages long, including 10 chapters, 24 figures and 2 tables.

Annotatsioon

Kõneside teenused pakettside põhistes mobiilside tuumvõrkudes

Käesolev magistritöö käsitleb kõneside pakkumise erinevaid võimalusi pakettsidele orienteeritud mobiiliside tuumvõrkudes. Töös on tuginetud Eesti mobiilside operaatori võrgulahendustele ja mobiilside seadmete ja tarkvara tootja Ericsson toodetele. Mobiilside operaatori lahendused, mis võimaldavad kõneteenuseid pakkuda pakettside põhistes võrkudes on välja töötatud magistritöö autori poolt või tihedas koostöös operaatori ja töö autori kui Ericssoni esindaja vahel. Traditsiooniliselt on kõneteenust pakutud 2G ja 3G võrkudes ahelkommutatsiooni meetodi abil, kuid alates 4G võrkude rajamisest pole selline meetod enam võimalik, sest 4G tuumvõrgu EPC arhitektuur on standardiseeritult ainult pakettkommutatsioonil põhinev.

Autor keskendub töös erinevatele lahendustele, mille abil on võimalik kõneteenust pakkuda EPC võrkudes. Lahendusvariandid on väljatöötatud ja standardiseeritud telekommunikatsiooni standardimise katusorganisatsiooni 3GPP poolt. Töö esimene pool annab ülevaate EPC tuumvõrgu teoreetilistest aspektidest nagu EPC võrgu üldine arhitektuur, võrgulülid, liidesed ja kasutatavad protokollid. Ülevaade tugineb peamiselt 3GPP standarditel ja kahel teemat käsitleval raamatul. Kuna lahenduste väljatöötamise ja tarkvaraarenduse protsessid on läinud väga kiireks, siis temaatilisi raamatuid tänapäeval enam peaaegu välja ei anta, sest raamatu ilmumise ajaks oleks info juba aegunud. Seetõttu tugineb ka autor paljuski 3GPP standarditele, veebimaterjalidele ja Ericssoni sisemisele dokumentatsioonile.

Autor võrdleb kolme peamist lahendust ja pakub kokkuvõttes välja parima variandi. Erinevate lahenduste võrdluses selgub, et eelistatuim on meetod, kus kõnesidet pakutakse IP-protokolli põhiselt üle LTE ja EPC võrkude, ehk meetod VoLTE. Selline meetod võimaldab kõige paremat häälkõne helikvaliteeti ja kiireimat kõne algatamist. Autor tegi ka endapoolse testiseeria operaatori võrgus, tuginedes Ericssoni tarkvarale mõõtmaks kõne loomise aega. Test kinnitas teooriat, et VoLTE kõnealgatus on oluliselt kiirem kui teised vanema põlvkonna lahendused. Samas on VoLTE rakendamine mobiiliside võrgus

keerukas, kulukas ja aeganõudev protsess. Üldjuhul kestavad sellised projektid ca poolteist aastat.

Töö teises pooles on vaatluse alla võetud mobiilside operaatori poolt reaalses võrgus kasutatav lahendus ja konfiguratsioonid võrgulülides. Autor püüab enda töökogemusele ja Ericssoni materjalidele tuginedes leida optimaalse konfiguratsiooni, mis tagaks lõpptarbijale kvaliteetseima ja stabiilseima häälkõne ja kiireima kõnealgatuse. Üldjuhul oli konfiguratsioon vägagi optimaalne, kuid siiski on võimalik konfiguratsiooni veelgi optimeerida. Tarkvaratootjad toovad uuendusi välja vägagi lühiajaliste ajaintervallide tagant; Ericssoni näitel igakuiselt. Seetõttu on vajalik pidevalt jälgida võimalikke uuendusi ja täiustusi tarkvaras, mis võimaldavad veelgi stabiilsemat ja kvaliteetsemat kõnesidet. Autor pakkus ka enda poolt võimalusi konfiguratsiooni uuendamiseks ja täiendamiseks. Mobiiliside operaator võttis VoLTE lahenduse kasutusele 2019 aasta juunikuus. Alates sellest on tarkvaratootja toonud välja palju uuendusi ja parendusi, millest mõnedki võiks operaatori võrgus rakendada.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 69 leheküljel, 10 peatükki, 24 joonist, 2 tabelit.

List of abbreviations and terms

2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
5GC	5G Core
5GS	5G System
AAA	Authentication, Authorization and Accounting
APN	Access Point Name
CN	Core Network
CS	Circuit-Switched
CSCF	Call Session Control Function
CSFB	Circuit Switched Fallback
DECOR	Dedicated Core Network
DNS	Domain Name System
DRB	Data Radio Bearer
EDGE	Enhanced Data rates for GSM Evolution
EIR	Equipment Identity Register
ECM	EPS Connection Management
EMM	EPS Mobility Management
eNB	E-UTRAN NodeB
ePDG	evolved Packet Data Gateway
EPC	Evolved Packet Core
EPS	Evolved Packet System
ETSI	European Telecommunications Standard Institute
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
GERAN	GSM EDGE Radio Access Network
GGSN	Gateway GPRS Support Node

GPRS	General Packet Radio Service
GSM	Global System for Mobile communications
GSMA	GSM Association
GTP	GPRS Tunnelling Protocol
GTP-C	GPRS Tunnelling Protocol for Control Plane
GTP-U	GPRS Tunnelling Protocol for User Plane
GUTI	Globally Unique Temporary Identifier
HLR	Home Location Register
HSDPA	High Speed Downlink Packet Access
HSPA	High Speed Packet Access
HSS	Home Subscriber Server
HSUPA	High Speed Uplink Packet Access
IETF	Internet Engineering Task Force
IMEI	International Mobile Equipment Identity
IMS	IP Multimedia Subsystem
IMSI	International Mobile Subscriber Identity
IPX	Internetwork Packet Exchange
IRAT	Inter-Radio Access Technology
ITU	International Telecommunication Union
KPI	Key Performance Indicator
LTE	Long Term Evolution
MCC	Mobile Country Code
MCPTT	Mission-critical push-to-talk
MME	Mobility Management Entity
MMTel	Multimedia Telephony
MNC	Mobile Network Code
MO	Mobile Originated
MSC	Mobile Switching Centre
MT	Mobile Terminated
NAS	Non-Access Stratum
NB-IoT	Narrowband Internet of Things
PCC	Policy and Charging Control
PCRF	Policy and Charging Rules Function
P-CSCF	Proxy-CSCF

PDN	Packet Data Network
PDP Context	Packet Data Protocol Context
PGW	PDN Gateway
PLMN	Public Land Mobile Network
QoS	Quality of Service
RADIUS	Remote Authentication Dial In User Service
RAN	Radio Access Network
rSRVCC	Return SRVCC
RNC	Radio Network Controller
RRC	Radio Release Control
RTP	Real-time Transport Protocol
SAE	System Architecture Evolution
S-CSCF	Serving-CSCF
SCTP	Stream Control Transmission Protocol
SDM	Subscriber Data Management
SGSN	Serving GPRS Support Node
SGW	Serving Gateway
SIP	Session Initiation Protocol
SRVCC	Single Radio Voice Call Continuity
TA	Tracking Area
TAC	Tracking Area Code
TAU	Tracking Area Update
TEID	Tunnel Endpoint Identifier
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
UTRAN	Universal Terrestrial Radio Access Network
VoLTE	Voice Over LTE
VoPS	Voice over Packet Switched
WCDMA	Wideband Code Division Multiple Access
WLAN	Wireless Local Area Network

Table of contents

List of figures

List of tables

1 Introduction

The current thesis focuses on topic providing voice services in modern packet core networks. The study is done based on two EPC books, 3GPP (Third Generation Partnership Project) specifications, equipment vendor internal materials and public web materials. The study uses Estonian mobile operator solution descriptions and implementation, real configuration examples. Thesis author has been partly responsible for technical implementation and integration of aforementioned services.

Thesis is divided into three main parts. First part gives theoretical overview of the EPC network and its architecture, mostly based on 3GPP specs. Second part gives insight of the three different voice service deployment solutions from technical perspective, trying to find the best and most optimum solution. Third part summarizes findings in previous chapters, analysis the mobile operator configuration and gives recommendations for optimisation.

The aim of current thesis is to compare different voice service solutions and suggest the best deployment option of them from technical perspective. Main solutions for providing voice services in EPC (Evolved Packet Core) are Circuit Switched Fallback (CSFB), Single Radio Voice Call Continuity (SRVCC) and VoLTE (Voice over LTE). Commercial factors are left out of scope. Thesis will bring out the advantages and disadvantages of three solutions, describing the related features up to the detailed level of parameters configuration. Study is done based on mobile equipment vendor Ericsson network technology and real Estonian mobile operator.

In GSM/GPRS (2G) and WCDMA/HSPA (3G), telecoms networks were mainly defined by voice services and designed respectively, including mobile core networks. It became clear that for managing very rapid increase in data services, there is a need for enhanced core network which would be able to handle exponential data traffic growth in mobile networks. Hence industry started working with evolved packet core network which would mainly fulfil the requirements of data services focused network. But it does not fulfil the requirements of providing quality voice services so well. Finding the best solution for operators offering voice services via packet core networks is the main focus of thesis.

The Estonian mobile operator has deployed modern EPC network and all three solutions that are analysed in the thesis, into the real production network. All the services have been made available commercially by year 2019. EPC domain is from equipment vendor Ericsson and neighbouring domains – IMS and CS – from different vendor.

While EPC is the network designed for data services it has limitations and shortcomings of supporting traditional voice services. Mainly because EPC does not include CS domain as its predecessors. Providing high quality voice services in LTE/EPC networks has been a challenge for Mobile Network Operators (MNO). There are different solutions and options designed by 3GPP for providing voice services which requires the co-operation of different network domains – Circuit Switched, Packet Switched and IP Multimedia Subsystem (IMS). This is one of the key factors making voice services deployment in EPC challenging. Deployment of voice services has been complex and technically and commercially demanding journey for MNOs. The good proof point is that Voice over LTE (VoLTE) services has been launched just in recent years and yet to be launched by many operators in 2019. That is many years after commercial introduction of EPC. While roaming services between different operators has been vital part of voice services offering for many years then VoLTE roaming has started to gain coverage just in very recent years.

EPC provides support for different high-speed access network technologies, including non-3GPP technologies. An example for non-3GPP access is WLAN and accordingly voice service as Wi-fi calling or Voice over Wi-fi (VoWi-fi). Current thesis does not cover non-3GPP access-based services and does not intend to go deep into packet core neighbouring domains IMS and CS. Those two domains are described on high level only.

2 EPS main terminology

In current thesis are used many telecommunications industry and standardisation body 3GPP defined terms and acronyms. To elaborate the meaning of main terms and context in which they are used the current chapter offers brief description of relevant terms.

EPC is the new Packet Core architecture itself. Defined so starting from 3GPP Release 8.

E-UTRAN is term for marking the radio access network that implements LTE radio interface technology.

LTE is the name given to 3GPP standardisation project. Outcome of this project is a set of standards defining the new radio access network – E-UTRAN. In daily talk LTE is often used more commonly instead of E-UTRAN itself. Same approach will be used in this thesis as well.

EPS is 3GPP term that refers to end-to-end system and comprises of UE, E-UTRAN and EPC. EPS is often referred to as combination of EPC and E-UTRAN. As a matter of fact, any IP-access network connected to EPC can be referred to as the Evolved Packet System. (EPS). EPS is broader concept than LTE connected to EPC.

UTRAN is the radio access network for WCDMA/HSPA.

GERAN is the radio access network for GSM.

WCDMA is the air interface technology used for 3G standard. Quite commonly referred to the whole 3G RAN which actual official term is UTRAN.

HSPA is the term that covers the enhancements in WCDMA standards allowing more high-speed data transmission in downlink and uplink directions – HSDPA and HSUPA respectively.

GPRS is the common term for packet core networks in 2G/3G networks

3 EPC architecture

New generation of core network, called Evolved Packet Core (EPC) is developed for high bandwidth services in mind right from day one and designed to enable mobile broadband services and applications.

EPC is part of EPS system – the core part of it. EPS also comprises of terminal (frequently called UE as User Equipment) and radio access network (RAN). While EPC is mostly defined for reaching by UE via E-UTRAN access it does also support different access networks according to 3GPP. RAN domains that are supported in addition to E-UTRAN (4G access) are GERAN (2G access), UTRAN (3G access) and non-3GPP access. Nevertheless, it is important to understand that EPC is in its core elements and features still designed to provide support for packet-switched services over LTE and non-3GPP access networks which do not have relation to Circuit Core. Interworking with some specific Circuit Core services, mostly voice, does exist and is described in later chapters.

Standardisation body for EPC is Third Generation Partnership Project (3GPP) who is responsible for developing all the specifications describing EPC architecture, signalling, protocols etc. EPC is part or Evolved Packet System (EPS) which comprises of radio access, core network (EPC) and mobile terminals that form the complete end-to-end mobile system. Focusing to Packet Switching results in higher bit rates than in previous 3GPP core architectures, faster throughput and lower latency. All the communication within the EPC network is IP-based. There is “always-on“ IP-connectivity provided between end user equipment and Packet Data Network (PDN). PDN can be the Internet, a corporate network, or a dedicated service network. Hosts and terminals connected to the PDN are accessible to hosts and terminals connected to the radio network. The EPC supports mobility within the LTE RAN, enabling UE to be handed over between eNodeBs, also mobility between LTE RAN and GERAN/UTRAN. That kind of mobility is referred to as Inter-Radio Access Technology (IRAT). In 5G era EPC can simultaneously support Dual Connectivity with LTE and NR (New Radio). New Radio is 5G generation radio access network. [1]

Large and essential domain related to EPC is Subscriber Data Management (SDM). This domain is handling the data related to the subscribers. Formally, by 3GPP specifications, SDM is not a separate domain as such from EPC; it can be seen as EPC subdomain. SDM functions are embedded part of Packet Core but also Circuit Core and IMS, interacting with subscriber databases defined by 3GPP. As SDM is broad domain, the insight into it in the thesis will be not very deep but still touched as it is necessary to understand SDM main functions to understand EPC itself.

Policy and Charging Control (PCC) is an additional concept within EPC which is not mandatory but in reality very widely used. The PCC is designed to enable flow-based charging, including policy control, which includes support for service authorization and QoS management. EPC also provides functions for management and enforcement of service-, radio- and data level policies such as QoS. It can be said that PCC is a key enabler for online charging and hence implicitly makes possible of registering prepaid subscribers mobile networks. Yet another subdomain in EPC is just briefly mentioned charging and billing which consists of Online Charging and Offline Charging. [2][1]

3.1 Standardisation

Understanding standardisation in telecommunication industry is vital to understand the whole EPC framework on broader level, and also in details thus there is a need to give overview of 3GPP standardisation principles.

Main standardisation body for EPS system, including EPC, is Third Generation Partnership Project (3GPP) which “owns“ the specifications. 3GPP leads all the standardisation work throughout all the standardisation phases and different releases. Where needed and applicable 3GPP is referring to Internet Engineering Task Force (IETF) and occasionally Open Mobile Alliance (OMA) specifications.

3GPP covers cellular telecommunications technologies, including radio access, core network and service capabilities which provide a complete system description for mobile telecommunications. The major focus for all 3GPP Releases is to make the system backwards and forwards compatible where possible, to ensure that the operation of user equipment is uninterrupted. The 3GPP specifications and studies are contribution-driven which means all the decisions are made jointly by members. [3]

Specification work is done on Technical Specification Group and Workgroup level. The three Technical Specification Groups are:

- Radio Access Networks
- Services and System aspects
- Core Network and Terminals

3GPP follows a three-stage methodology as defined in ITU-T (International Telecommunication Union) Recommendation I.130

- stage 1 specifications define the service requirements from the user point of view.
- stage 2 specifications define an architecture to support the service requirements.
- stage 3 specifications define an implementation of the architecture by specifying protocols in detail. [5]

EPC was first introduced by 3GPP in Release 8 and work was completed in December 2008. 3GPP uses a system of releases which provide developers and network equipment vendors a stable platform and a set of features for implementation at certain point of time. Addition of new functionalities is always possible to be added in subsequent releases. [3]

While EPC was introduced as a whole solution in Release 8 then Release 9 was mostly about leftovers of Release 8, including e.g. MBMS (Multimedia Broadcast Multicast System), SON (Self-Organised Networks) and EPS emergency bearers. Lot of enhancements were made in RAN side, also minor improvements in EPC side. Release 9 was officially completed in December 2009. Releases 10 and 11 new studies were mostly about LTE enhancements while Release 11 also introduced machine-to-machine communication and advanced IPX (Internetwork Packet Exchange) in core side. Release 11 was completed in Q3 2012. Release 13 introduced major new features in core side like NB-IoT (Narrowband IoT) and DECOR (Dedicated Core Network) and MCPTT Mission-critical push-to-talk) and work was completed Q1 2016. Latest completed release is Release 15, which focuses mostly on 5G System (5GS). All 5GS related information, including voice services in 5GS is beyond of the scope of the thesis.

The evolution of the releases and main components of each of it is illustrated in Figure 1

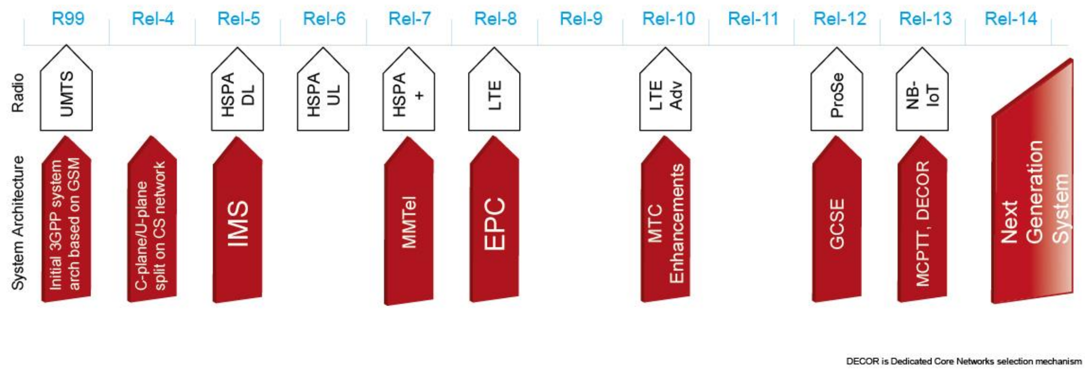


Figure 1. 3GPP Release history and key features. [4]

3.2 Architecture overview

EPC consists of nodes and functions that provide provide support for packet-switched services, primarily IP connectivity, over different RAN accesses. EPC main components, network functions, are seen in Figure 2. These are Mobility Management Entity (MME) which is purely control plain function. Serving Gateway (SGW) and PDN Gateway (PGW) which are mostly user plane functions but do have signalling component as well. And last Home Subscriber Server (HSS) which is holding the information of all the subscriber base of mobile operator and is interfaced to MME. [2]. These four core elements are the mandatory nodes but the overall architecture in reality includes many more network functions. We will look into the functions and responsibilities of these, interfaces and termination points of each node more deeply in Chapter 4. Main architecture of EPC/LTE and it’s components can be seen in Figure 2.

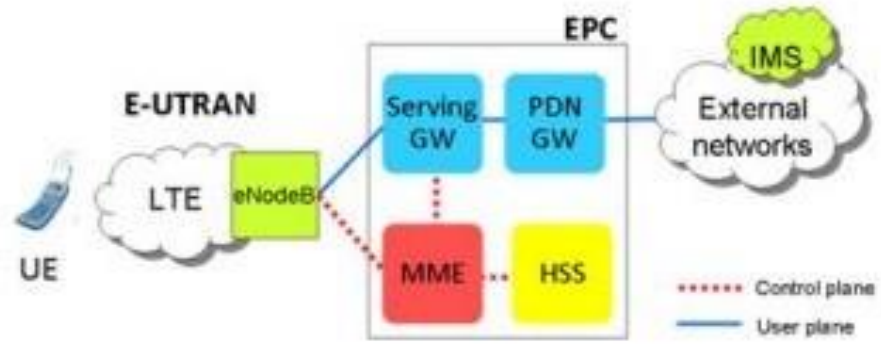


Figure 2. EPC main architecture. [1]

What is not visible in the picture is IP infrastructure supporting the logical nodes as physical components of the actual EPC network. IP infrastructure is considered as part of transport network enabling the IP connectivity and routing between EPC entities. Transport and IP infrastructure is out of scope of the thesis. Also relevant function not visible in the picture is DNS which is mainly used for selection and discovery of different network elements, e.g. SGW and PGW. While DNS as a supporting function is in reality used by most mobile operators, it is still not part of EPC architecture and hence does not get much attention.

All nodes and interfaces described in the chapter are logical nodes and interfaces which means in real network implementation these functions may be located in the same physical entity equipment. Best example is SGW and PGW as separate EPC logical functions which frequently are co-located in one physical entity. It is up to equipment vendor product solution and operator architecture if certain functions are co-located or not.

The most important key factors of EPC architecture are :

- flat architecture
- all-IP interfaces
- control plane (CP) and user plane (UP) separation

Idea behind flat architecture is that very few nodes are involved in handling data traffic, also called payload. Flat architecture is driven by the will of handling of user data as optimized way as possible.

For control plane and user plane separation is decided that user data and signalling is separated in regards of node functions and interfaces making the network scaling and dimensioning independent and flexible. Separate handling of control signalling and user data makes possible of scaling control plane and user plane nodes based on the actual traffic mix in the network. Control signalling tends to scale with the number of users (subscribers) while user data volume scales depending of services and applications introduced in the network or outside of MNO network. Simply said, growth of the number of subscribers in the network would generate the need for additional capacity in control plane nodes while new services would generate the need for additional capacity in user plane nodes. As currently in developed countries oprators' networks number of subscribers is not increasing rapidly anymore only the nodes that are associated with carrying end-user data traffic need to be scaled, for supporting high-bandwidth traffic.

Splitting control and user plane allows also different geographical network deployment options. Most commonly used scenario would be to use centralized deployment of the equipment handling control signalling and more distributed deployment for user data handling functions. This would make possible for decreased latency for heavy user data, like video streaming, gaming etc.

3GPP specifies support for multiple access technologies for EPC. Multiple access technologies can be E-UTRAN, UTRAN, GERAN and non-3GPP technologies. Standards also specify the mobility between these access technologies. The idea was to bring convergence using a unique core network providing various IP-based services over multiple access technologies. [1]

As stated EPS also allows to connect the UE to EPC via non-3GPP access. Non-3GPP means these technologies are not specified in 3GPP. Examples of such technologies can be WiMAX, CDMA2000, WLAN or also fixed networks. Non-3GPP accesses can be devided in two categories: „trusted“ and „untrusted“. Trusted non-3GPP accesses can interact directly with the EPC. Untrusted non-3GPP accesses interact with the EPC via network entity called the ePDG (evolved Packet Data Gateway). 3GPP does not specify which non-3GPP accesses should be considered trusted or untrusted. The decision is put as operators own vision and responsibility. The most wide spread example of non-3GPP untrusted access is WLAN, enabling Wi-Fi calling in mobile operators' networks. Deep insight into non-3GPP access is beyond of the scope of current thesis.

Detailed EPC architecture is illustrated in Figure 3. Following chapter explains much of the details of the architecture that is visualised in the figure but defenately not all as the scope would grow to be too large.

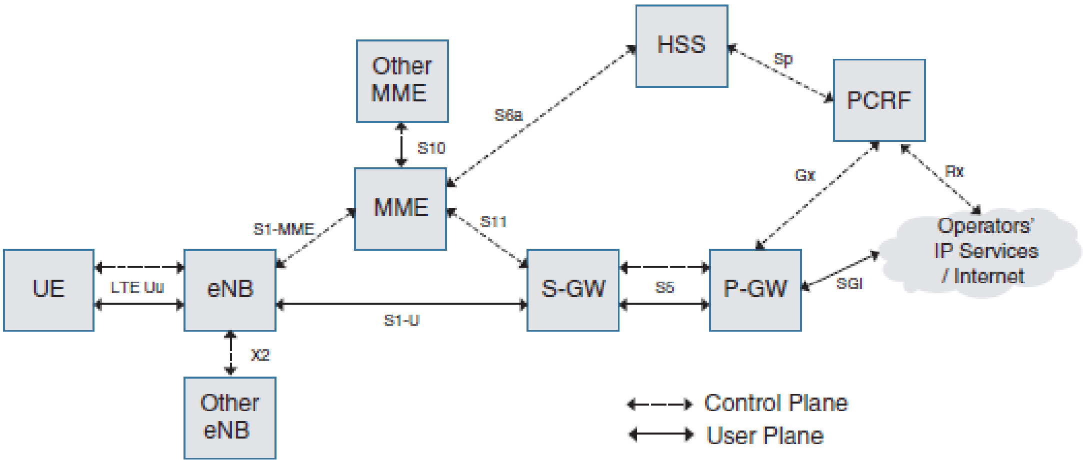


Figure 3. EPC detailed level architecture [7]

4 EPC Network entities

4.1 Network functions

The MME (Mobility Management Entity) is the key control-node for the LTE access network. MME is responsible of handling all control plane signalling, including mobility and security functions for the terminals. It is involved in the bearer activation/deactivation process and is also responsible for choosing the SGW for the UE at the initial attach. It is responsible for authenticating the user (in conjunction with the HSS). For each registered UE, the MME stores a list of parameters that is called UE context. The context information contains the assigned UE identities, UE capabilities, the UE's current location and currently used security keys. This part of UE context is called Mobility Management context. The UE context also contains a description of the UE's active PDN connections. The UE context is, within one MME, uniquely identified by the UE's IMSI number. The MME also manages all the terminals that are in idle mode. The MME is responsible for tracking and paging of the UE in idle mode. It is the termination point of the Non-Access Stratum (NAS). All eNodeBs are connected to at least one MME over the S1-MME logical interface. [6]

MME pool

MME pool is a specific concept in EPC/LTE mobility management which results radically decreased number of signalling in the network. MME pool allows for load distribution over multiple nodes leading to an optimized capacity. [2] MME pool is also very beneficial for situations where one MME as a pool member goes out of service or is manually taken out of service, e.g. is going to be upgraded, resulting node restart. In case of MMEs are in pool this does not result of any outage or disruption in service for end users while.

An MME pool area is defined as an area where the UE is served without having to change the serving MME. The MME pool area is served by one or more MMEs working in parallel, that is, one eNodeB is connected to several MMEs. An MME pool area consists of one or several Tracking Areas (TA). As long as the UE remains within the pool service area, it is attached to a specific MME. If the MME is unavailable, the eNodeB reroutes the signalling for the attached UE to another MME in the pool. MME pool example is shown in Figure 4.

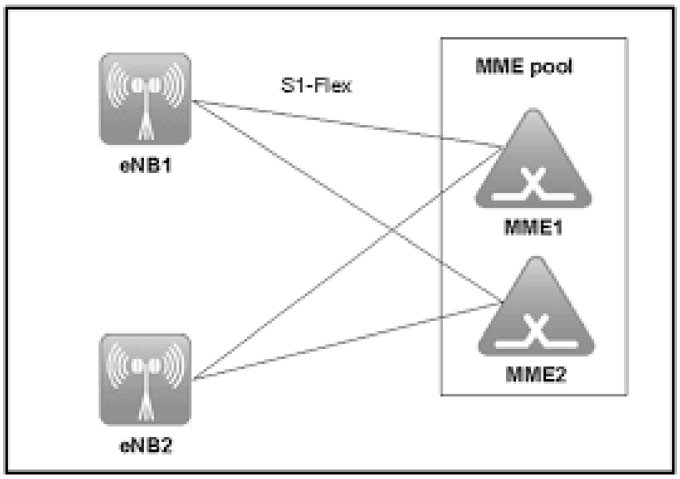


Figure 4. MME pool. [9]

The Serving GW is primarily User Plane node and terminates S1-U user plane interface towards the eNodeBs and acts as anchor point for LTE mobility. Its prime responsibility is routing and forwarding of user IP-packets. The SGW also buffers downlink IP packets sent for terminals that are at the moment in idle mode. It is important EPC architectural factor that for roaming users SGW always resides in the visited network. In roaming scenario SGW always resides in visited PLMN (Public Land Mobile Network). [6]

The PDN GW is the EPC interconnection point to external IP networks which is usually public internet but does not have to be necessarily. PDN has functionality for IP address allocation, charging, packet filtering and policy enforcement role. In addition PDN GW has a key role in supporting QoS for end-user IP services. [1, lk 23]. The PDN GW is often physically co-located with SGW in network deployments and connects to SGW via S5 interface and to external packet data networks via the SGi interface. In roaming scenario PGW always resides in home PLMN.

HSS is basically a database that contains user-related and subscriber-related information. It also provides support functions in mobility management, call and session setup, user authentication and access authorization. The Home Subscriber Server holds subscription profiles and security related parameters, just as HLR does in GPRS case.

The main functions for each EPC node mentioned above is presented below in detailed and compact form.

- MME
    - Authentication
    - GW Selection
    - Session and Mobility Management
    - Tracking Area handling
        - Paging
        - Handover 2G/3G <-> LTE
- Serving-GW
    - Session Management
    - Payload handling
    - LTE Mobility
    - Lawful Interception (LI)
- PDN-GW
    - IP adress allocation
    - Connectivity to IP Networks (ISP, PDN etc)
    - Lawful Interception (LI)
- HSS
    - Maintain and provide subscription data
    - User identification handling
    - Access Authentication
    - Provide Keys for Authentication and Encryption
    - User Registration Management
    - Maintain Knowledge of used PDN GW
- PCRF (Policy and Charging Rules Function)
    - Set QoS for each SDF (Service Data Flow)
    - Provide Service Data Flow Gating
    - Define Charging for each SDF
    - Enables QoS Control

4.2 Interfaces and reference points

In 3GPP specifications the term "reference point" is used to describe an association between two logical network entities. In the thesis term "interface" is used which is more commonly used. There is a slight difference in the formal definition of a reference point and an interface, but for the purpose of this thesis the difference has no practical value. [2] In EPS most interfaces start with the letter "S" with some exceptions. All the most relevant interfaces are illustrated in Figure 5. The key interfaces used in case of each UE attachment are S1-MME, S1-U, S6a, S11, S5 and SGi. The interfaces with dashed line are Control Plane interfaces while interfaces of User Plane are marked with solid line. Current chapter and Figure 5 does not describe the interfaces between EPC and CS domains, like SGs and Sv between MME and MSC (Mobile Switching Center). These are described in later chapters.

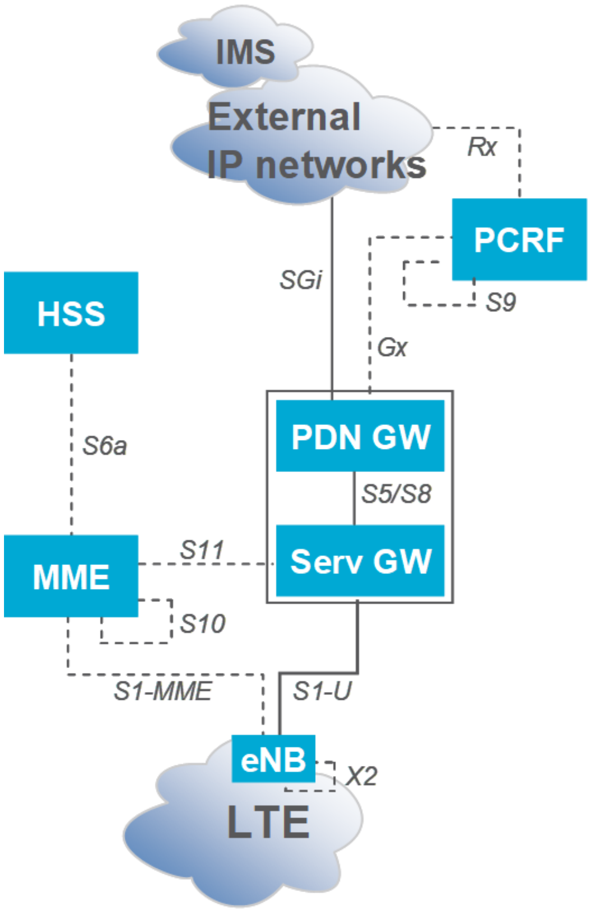


Figure 5. EPC main control plane and user plane interfaces. [7]

The S1-MME interface connects the MME to eNodeBs. S1-MME transports S1-AP protocol (S1 Application Protocol) messages over SCTP protocol and NAS (Non-access stratum) messages over S1-AP. S1-MME interface support for functionalities like paging, handover, UE Context management, E-RAB (E-UTRAN Radio Access Bearer) management and transparent transport of messages between MME and UE.

The S1-U is user plane interface. It is between E-UTRAN and the SGW for user plane tunnelling. The transport protocol over this interface is GTPv1-U (GPRS Tunnelling Protocol-User Plane). The same protocol was used in GPRS network so there is no change in this matter. GTPv1-U is defined in 3GPP TS 29.281. [7]

The S11 interface is between MME and SGW and meant for supporting mobility and bearer management. For example, to create, modify or delete EPS bearers. The interface uses GPTv2-C protocol which is defined specifically for EPS and later explained in chapter 4.3. The S11 interface activities are always triggered by events via NAS level signalling from the terminal (like device attaching to the EPC network) or they may be triggered during network-initiated procedures like PDN GW initiated bearer modification procedures. S11 interface keeps the control- and user plane procedures in sync for a terminal during the period the terminal is attached in the EPS. During the handover, the S11 interface is used to relocate the SGW when appropriate. The S11 is one of the key interfaces in EPC. Figure 5 illustrates the S11 interface protocol stack which is brought here as an illustration of one GTP based interface. GTP protocol is the cornerstone of EPC architecture

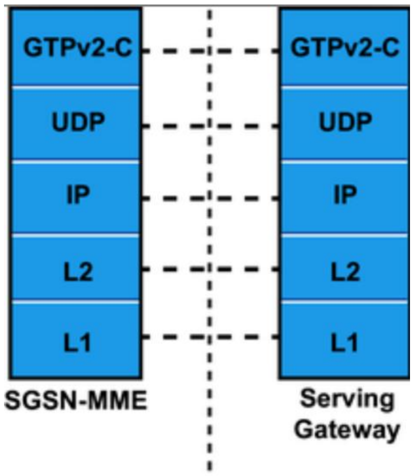


Figure 6. S11 interface protocol stack. [7]

The S6a interface is defined between MME and HSS and is used for multiple purposes. Protocol that is used is Diameter which is thoroughly described in chapter 4.3. The main purpose of the interface is that enables the transfer of subscription and authentication data for authenticating/authorizing user access. The interface has many more purposes that are described below:

- Exchanging location information. Location in this context is meant as the MME identity. The MME that is serving UE notifies HSS about the MME identity. In some cases, for example if the UE attaches to a new MME, the MME downloads information from HSS about the MME that previously served the UE.
- Authorizing a user to access the EPS. The HSS holds the whole subscription data of the subscriber, e.g. allowed APNs (Access Point Names) and other information related to the user's authorized services. The subscription profile is downloaded to the MME and used when granting a user access to the EPS. [1 lk 388]
- Exchanging authentication information. The HSS provides authentication data to the MME based on what user is being authenticated by the MME. It is relevant that MME is responsible for authenticating UE.
- Download and handle changes in the subscriber data stored in HSS. When subscriber data is modified, the updated subscription data is pushed to MME. For example, at the situation when subscription is withdrawn the HSS notifies MME about that and MME detaches UE. After the update of subscription data MME can also modify ongoing session not just detach it.

The S5 interface has both Control Plane and User Plane components. In CP case it is used for tunnel management between SGW and PGW. All the functionalities associated with creation, deletion, modification of the bearers for EPS connected users are handled over S5. In UP case user plane tunnelling is handled. The S5 is used in non-roaming scenario where SGW and PGW are both located in home network. S8 is the roaming scenario of S5 where SGW is located in the visiting network and PGW in home network. Exception could be solution called "Local Breakout" in which the SGW and PGW both are located in visited network.

The Gx interface is between the PCRF and the PGW, allowing the PCRF direct control over the policy enforcement functions of the PGW. The protocol used on the Gx interface is Diameter. [8]

4.3 Protocols

EPC architecture is designed around the tunnelling protocol named GTP (GPRS Tunnelling Protocol) and Diameter which is evolved from earlier RADIUS protocol. GTP was initially developed in ETSI and later continued within 3GPP after its creation. GTP is a fundamental part of EPC network running over IP between core entities MME (Mobility Management Entity) and SGW (Serving Gateway). [8]

IETF generated protocols play also essential role in EPS. 3GPP developed EPC surrounding systems IMS and PCC (Policy and Charging Control) in which all the protocols are built on top of initially IETF developed base protocols and later enhanced according to 3GPP needs. An example of such protocol is Diameter which base protocol is defined by IETF and defined in RFC 6733.

4.3.1 GTP

GTP protocol has two main components: the control-plane part (GTP-C) and the user-plane part (GTP-U). GTP-C is used to control and manage tunnels for individual terminals that are attached to EPC. The GTP-U uses a tunnel mechanism to carry the user data traffic. There are three different versions for GTP-C that exist: GTPv0, GTPv1 and GTPv2. The one that is designed for EPC specifically and strictly suggested to use is GTPv2. For GTP-U there are two versions: GTPv0 and GTPv1. Latter is recommended to use. In the EPS, all interfaces between SGSN and MME, between MMEs, between MME and SGW and between SGW and PGW use GTPv2-C. These are interfaces S5, S8, S10, S11 which were described previously in this chapter. Interfaces S1-U, S5 and S8 use GTPv1-U. So it is clear S5 and S8 are using both versions of GTP – GTP-C and GTP-U. The key functions of GTPv2-C are described below.

1. Mobility Management. The set of messages within this subset include functions for managing terminal’s identification and maintaining presence in the network. Also coordinating handling of data transfer between entities during handover, relocation and so on.

2. Tunnel Management. That involves creation and deletion of session; creation, modification and deletion of bearers established during the end-user data session and time user is connected. The messages exchanged during tunnel management

keep the user's different service requirements maintained while the user moves around within the network or between PLMNs.

3. Service-specific functions. One example of such service is MBMS (Mobile Broadcast Multicast Service). GTP-v2 also supports messages for services CSFB (Circuit Switched Fallback) and SRVCC (Single Radio Voice Call Continuity) procedures. These services are main focus areas of thesis. [8]

A GTP tunnel is uniquely identified in a given node by combination of IP-address, UDP port number and allocated TEID (Tunnel Endpoint ID). For the control plane for each endpoint of a GTP-C tunnel there is a control plane TEID-C as tunnel is bidirectional. The scope of the tunnel and TEID-C depends on the interface and its functions. The TEID-C is unique per PDN connection. There is only one pair of TEID-Cs per UE over "S" interfaces. The same tunnel is shared for control messages relating to the same UE operations. GTP protocols are defined in 3GPP specs TS 29.274 and TS 29.060.

4.3.2 Diameter

The Diameter protocol is Authentication, Authorization and Accounting (AAA) protocol and is the successor of RADIUS protocol. Diameter is widely used in the IMS architecture, allowing IMS nodes to exchange AAA-related information. In EPS, the Diameter protocol is used towards HSS for location management and subscriber data management and towards EIR (Equipment Identity Register) to check whether a given terminal is stolen. The Diameter protocol is also used for retrieval and provisioning of charging and QoS related information from the PCRF and Online and Offline charging systems.

Diameter is made up of two parts, the mandatory Diameter base protocol and the optional Diameter extensions. Base protocol is specified in IETF RFC 6733. The Diameter base protocol defines a set of general messages and rules that apply to all messages that are exchanged between Diameter nodes. Diameter base protocol extensions are called applications. A Diameter application does not imply a program, but a protocol that is based on the underlying Diameter protocol. The applications benefit from the general capabilities of the Diameter base protocol. Several diameter applications have been defined by IETF, like NAS application and Credit Control Application, but it is also possible to define “vendor-specific” applications. Vendor in this context can be an organization or company. For example, 3GPP as an organization has defined applications S6a and Gx which are widely used in EPC. [6], [10]

Diameter is designed as a peer-to-peer architecture, meaning that every host who implements the Diameter protocol can act as either a client or a server depending on network deployment. Diameter node can refer to a Diameter client, a Diameter server or a Diameter agent. In the EPS case MME acts as a Diameter client while the HSS acts as a Diameter server. In EPS, Diameter runs over Stream Control Transmission Protocol (SCTP). A Diameter connection is a physical link between two diameter nodes. A Diameter session is a logical association between two diameter nodes and can cross multiple connections, with the help of Diameter agent who forward the messages to appropriate destination.

4.3.3 SCTP

SCTP is important protocol in EPC and therefore it is significant to give short overview of it. SCTP is a transport layer protocol that provides reliable, in-sequence transport of messages with congestion control. SCTP is also a connection-oriented protocol that maintains a relationship between the end points of an SCTP association during the message transmission. SCTP is designed to transport telecommunications signalling messages over IP Networks. The SCTP association must be established between two endpoints before any data transfer can take place. SCTP association set up is done by using a four-way handshake unlike TCP three-way handshake.

The important “feature” of SCTP is multihoming which is implemented to increase redundancy of connectivity. A node is defined as multihomed if it can be reached through more than one IP address. The idea is to set up more than one path to a destination and supervise the path availability. If one path fails, the SCTP traffic can continue over one of the redundant paths. Multihoming is illustrated in Figure 7. [7]

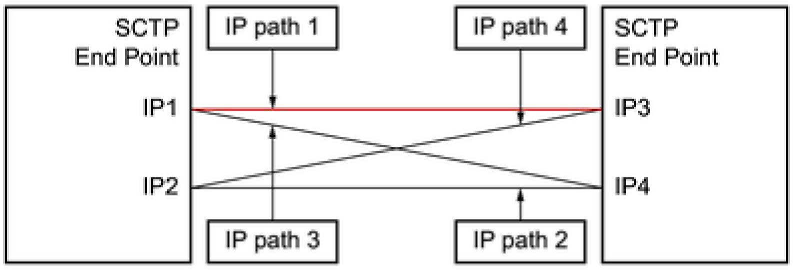


Figure 7. SCTP multihomed association with four IP paths. [7]

In figure 7, there are two multihomed SCTP end points. "End point 1" on the local side has IP1 and IP2 addresses. "End point 2" on the remote side with IP3 and IP4 addresses. Association is established and all IP paths are active. The primary path "IP path 1", assigned on the local side is IP1-IP3 (marked with the red bold line). If the transmission through the primary path (IP1-IP3) fails and some traffic is present, SCTP uses the other path (IP2-IP4) for traffic. [7]

5 Deployment options

There are different deployment options for introducing EPC into mobile operators' networks. One option is that initially EPC equipment is deployed as new, separate network nodes not to affect existing GSM/WCDMA infrastructure. Proper dimensioning of the new EPC network and individual nodes needs to be done, as well as in-depth planning how the new nodes to be integrated into operator existing IP infrastructure. Important domain needs to be taken care of is DNS and it's related configuration. EPC requires completely new DNS server entries to serve LTE access compared to DNS in GPRS. [2].

Secondly, important decision to be made by operators is how to introduce PGW functionality. It is possible that there is no change in the way GSM/WCDMA radio networks attach to SGSN and GGSN and in parallel LTE RAN and its capable UEs attach to EPC. Which in reality means RAN must be aware which terminal is LTE capable and which not and forward the terminal respectively to whether legacy core network or EPC. The solution offered by many vendors is that new PGW incorporate GGSN functionalities and contact to the SGSN over Gn interface. So the existing subscriber base, which was so far served by GGSN is now served by PGW and GGSN as a logical function is eliminated from network. In such a case SGSN "thinks" it is actually connected to GGSN over Gn interface while it is actually connected to PGW. What enables this kind of manipulation is GTPv1 user plane protocol, used in both GPRS and EPC architecture. To achieve this, existing GGSNs may be upgraded to support PGW functionality or to replace GGSN with PGW nodes. The latter is the widely used scenario by operators.

Thirdly, the decision must be made if and to what extent to use **combined nodes**. SGW and PGW can be physically split or used as a combined SGW/PGW, meaning to have it as one physical entity. In most use cases it does make sense to use combined SGW/PGW as it simplifies management and minimizes the amount of hardware. Most of the operators have gone the way using co-located gateways. It would make sense to use separate nodes for very large deployments where operator's network is geographically heavily distributed, to keep the SGW as LTE anchor point closer to RAN to keep the latency for payload as minimal as possible.

6 EPC/LTE mobility key elements

6.1 EPC/LTE terminal registration procedure

The initial attach procedure is the basis of understanding the VoLTE related call and EPC and IMS registration. This chapter gives very brief overview of UE registration procedure to EPC core network. Attach signalling flow is illustrated in Figure 8. Solid lines represent the signalling that is mandatory while dashed lines mark the optional signalling, depending on the exact preceding conditions. Below is briefly described each step of the signalling flow.

The UE initiates the Attach procedure sending an *Attach Request* message, which includes the IMSI or GUTI and usually IMEI, also information about its last Tracking Area if that is available (message nr 1 in the Figure 8).

Relevant part is that UE can ask *attach type* as *Combined* which means UE will register both to PS and CS domain within one current registration. MME checks if there is any restrictions to the UE, like roaming restrictions etc. If UE attaches to a new MME, the MME initiation *Identification* procedure during which the old SGSN/MME sends all the UE context information to new MME (messages nr 2 and 3).

If UE is not known to either of MMEs, the serving MME sends *Identity Request* message to UE to request IMSI and gets *Identity Response* from UE which consists IMSI (messages nr 4 and 5).

After successful security related activities (message flows nr 7 and 8) MME can send *ESM Information Request* to UE, if the UE had initially in the Attach Request set specific Information Element flag to 1. Reason for that is *Attach Request* is not secure procedure and sensitive information is not sent in this message. The UE sends an *ESM Information Response,* containing the UE requested APN (messages nr 8 and 9).

If there are any active bearers for the UE in the new MME, the new MME asks from PGW do delete active bearers, with *Delete Session Request* (message nr 10).

Next step is to send *Update Location Request* to the HSS. That is optional procedure in few cases, for instance if UE provided IMSI is to be informed or there is no valid subscription data for the UE in the MME. The HSS sends *Update Location Answer* over Diameter protocol, including all the subscription data (messages nr 12 and 13).

After that MME will send *Create Session Request* to SGW who forwards it to the PGW. The PGW creates entries about requested bearers into its bearer context table and responds with *Create Session Response* to SGW which forwards the message to MME. Now MME sends updated UE related info to MSC in CS domain in *Location Update* procedure. Illustrated by messages nr 14-17.

After previous core internal signalling is completed MME sends *Initial Context Setup Request* to the eNB. That message contains *Attach Accept* message which is forwarded to UE by eNB. The MME also includes *Activate Default Bearer Request* message in the *Attach Accept* message. The purpose of this is to have "always-on" PDN connection at the time UE is attached to EPC. The UE sends the *Attach Complete* message containing the identity of the bearers through the eNodeB to the MME. These are messages nr 18-20.

For a Combined Attach Request, the MME sends a *TMSI Reallocation Complete* message to the MSC/VLR (message nr 22). The MME sends a *Modify Bearer Request* message to the SGW, containing the TEID of the eNodeB and the IP address of the eNodeB, SGW acknowledges the request with *Modify Bearer Response* (messages nr 23, 24) and with that attach procedure is completed.

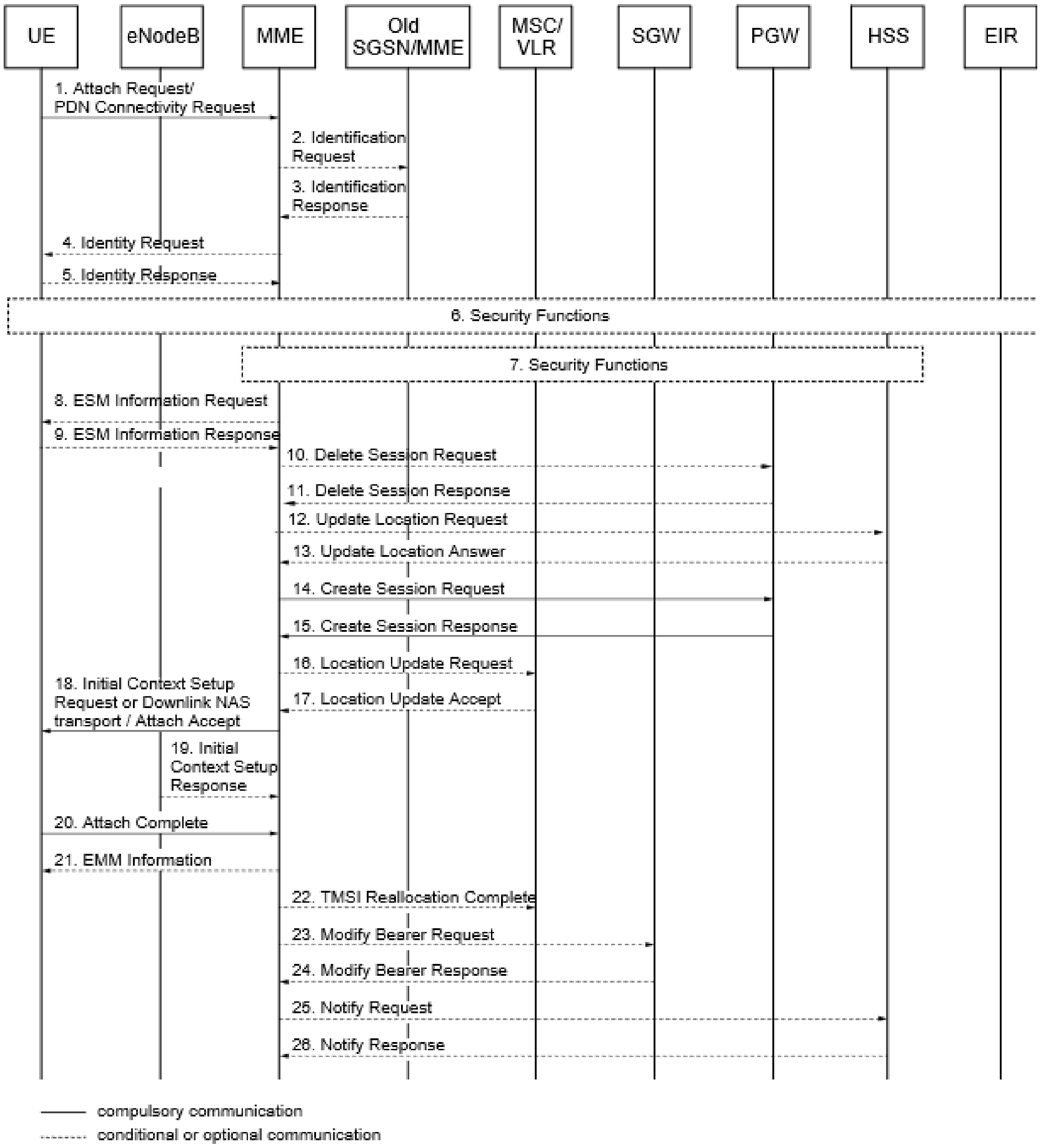


Figure 8. Terminal initial attach procedure in EPC network. [7]

6.2 EPC and 2G/3G core interworking

The ability to allow for continuous service coverage through interworking with other radio networks is a key feature in any mobile network architecture. For LTE deployment, interworking with existing access networks supporting IP connectivity becomes crucial because during introduction of LTE its coverage is much more limited compared to older RAN technologies. Hence UE can drop out of LTE coverage and register in legacy RAN. Smooth interworking in such case is necessary. From core perspective 3GPP has defined two different solutions to achieve interworking. The two options make difference how SGSN (Serving GPRS Support Node) is handled and implemented – there is Gn-interface based SGSN and S4 interface based SGSN. SGSN is a legacy node in 2G/3G packet core (GPRS). As Gn-SGSN solution is a lot more widely deployed interworking based on that implementation is described.

In the GPRS architecture, an SGSN connects to a GGSN (Gateway GPRS Support Node), which acts as the point of interconnect to external IP networks for all data sessions over GPRS. The logical SGSN has a key role to play also for LTE/EPC while that is not the case for logical GGSN node. Existing SGSNs and GGSNs can keep serving non-LTE users as before, but the SGSNs are also utilized by multi-RAT (Radio Access Type) LTE devices when out of LTE coverage. [2] The interworking solution includes the MME and the PDN GW acting towards the SGSN as another SGSN and GGSN respectively. The MME and PGW are replicating the signalling needed for movements between 2G/3G and LTE. The MMEs and PGWs act towards the SGSN as SGSNs and GGSNs respectively. All above means that both the MME and the PGW interface SGSN over the standard packet core Gn interface, which defines the name of the whole solution. Historically SGSN is interfaced to logical node HLR (Home Location Register) and MME to HSS. While moving between networks there must not be incoherent information in the network, for example to what RAN UE is currently connected. This means HLR and HSS need to ensure consistency of information between them through close interaction between each other or as an alternative share a single set of data. Which means same user database is in use for HLR and HSS and these functions behave like frontends to the remaining EPC.

7 Voice services in EPC

Current chapter covers main functionalities of different voice service options in LTE/EPC networks. Detailed level overview is given in chapters 8-10.

There are two fundamentally different ways that voice services can be implemented in EPC networks for LTE users. These two options are Circuit Switched Fallback and VoLTE/MMTel based services which relies on IP Multimedia Subsystem technology. From the beginning of EPC standardisation these two approaches have been the main options for all mobile operators to offer voice services in their LTE and IP-based core networks. Simply said, CSFB solution utilises the legacy network infrastructure in which voice calls were conducted over CS core in GSM and WCDMA. In VoLTE scenario voice calls rely on IMS and MMTel (Multimedia Telephony) applications and packet-switched infrastructure, hence there is no need for legacy core. Both options have its pros and cons which are investigated deeper in subsequent chapters [2].

Circuit Switched Fallback solution does not involve IMS architecture and in fact voice calls are never served over LTE at all. CSFB relies on a temporary inter-system change that moves the UE from LTE to a system with 2G or 3G radio access and where circuit-switched voice calls can be served. As briefly described in Chapter 6.1, CSFB relies on the LTE terminals that are registered not only in EPC but also in the circuit-switched domain when powered up and attached to LTE. This kind of registration mode is called “combined attach”. All the voice capable terminals, also called voice-centric terminals, are obliged to register also in CS domain. Most of the modern smartphones are voice-centric. If such an option to register to CS domain is not available, then registration of the terminal fails also in PS domain. The dual-domain registration is handled by the network through an interconnection between MME and MSC Server. The latter is purely CS domain network entity. The interface between these two entities is called “SGs” and plays vital role in CSFB solution enabling. There are two main use cases for CSFB – voice calls initiated by the mobile user or voice calls received by the mobile user. These use cases are called Mobile Originated CSFB and Mobile Terminated CSFB respectively. From core perspective these two are very different setups to deal with and serve.

If the user is to make a voice call, the terminal switches from LTE to a system with circuit-switched capabilities. Any packet-based services are handed over to and continue to run in the new system or they are suspended until the time the voice call is terminated and terminal moves back to LTE. Which of these case applies is depending on the capabilities of the system the call is switched to.

If the user is attached to LTE and there is an incoming voice call to that user the MSC requests paging in LTE over SGs interface. Paging is managed by MME. The terminal receives the paging in LTE and then immediately switches to circuit-switched capable network where the voice call is received. After the voice call is terminated the UE moves back to LTE automatically. Packet-based services are handled the same way as in MO CSFB case.

MMTel is the IMS-based service offering for voice calls, standardized by 3GPP. MMTel is a natural choice for offering voice services in LTE coverage as EPS is designed to efficiently carry IP flows between IP hosts [2]. IMS basic architecture is covered in Chapter 7.2 and deeper insight to MMTel in Chapter 7.1

Single-Radio Voice Call Continuity (SRVCC) is designed to allow handover of a voice call between a system that supports the IMS/MMTel voice service and where there is not sufficient radio access support for carrying the MMTel service. Such an example could be, due to insufficient bandwidth for IP services, or insufficient QoS support in the network. SRVCC defines a solution for how an IMS-based voice call in one system is handed over to another system, which serves the voice call using circuit-switched mechanism. The term “single radio” very simply said, means that terminal is capable of handling voice call either in GSM/WCDMA or in any IMS-based service, like e.g. Wi-Fi. If one option turns impossible to handle the radio can do handovers between the systems. The dual-radio would mean, the terminal is simultaneously working in both, GSM and Wi-Fi system at the same time. That would make the radio of the terminal very complex and that is why “single-radio” concept is used. [11]

3GPP has specified the following combinations of SRVCC (handovers between systems):

- LTE to GSM
- LTE to WCDMA
- WCDMA to GSM

- WCDMA to WCDMA
- GSM to LTE
- WCDMA to LTE

The last two options, where the call is started as a CS call in GSM or WCDMA and then transferred to MMTel Service on LTE, was defined by 3GPP in Release 11, and the feature is known as “return SRVCC” or rSRVCC [2]. It is also important to note here that IMS-based voice calls is something that is not strictly related to LTE network. Even though in most realisations the IMS-based call is propagated over LTE networks, subscriber can have IMS-based call also in WCDMA. The option “WCDMA to WCDMA” in the list above means handing over the subscriber call in the WCDMA that is IMS-based to WCDMA that is not IMS-based or vice versa.

7.1 MMTel

MMTel is a standard for providing multimedia services over IMS. MMTel standard is described in 3GPP TS 24.173 and was initially a joint project by 3GPP and ETSI/TISPAN (European Telecommunications Standard Institute/Telecoms and Internet Converged Services and Protocols for Advanced Networks). A principal characteristic of the MMTel standard is that the mobile access is based on Internet Protocol (IP). This makes the standard future proof. MMTel is designed with the aim of replacing fixed and mobile circuit-switched telephony. Characterized by quality, interoperability and reliability, the MMTel standard is a telco-grade service. [7]

Voice over LTE (VoLTE) is most widely used implementation of MMTel. It is important to realise that VoLTE is just one possible realisation of MMTel but is erroneously used as a synonym for MMTel. VoLTE is described and defined in GSMA profile IR.92, based on MMTel. GSMA profile IR.92 covers every layer of the network, including IMS features, media requirement, bearer management, LTE radio requirements and common functions, such as IP version. It includes a subset of general IMS and MMTel service features, selected to provide an IP telephony service on at least the same quality level as current CS based WCDMA/GSMA networks. In reality operators are using different enhancements in addition to IR.92 profile, mainlu to raise voice quality. VoLTE profile represents the minimum set of features that can be a starting point for operators who plan implementing Voice over LTE. [2]

MMTel is standardized service offering for voice calls and is built on the IMS and offers more possibilities than traditional circuit-switched voice calls. The aim is to enhance the communication experience for end-users and enables to use real-time multimedia services, for instance adding video, text, chat, instance messaging and conference calls to the basic voice component. Users can easily change the service by adding and dropping media streams and calling parties during an ongoing session. They can also easily switch between sessions, devices and fixed or mobile connections, or start a new chat session, upgrade the session to a voice or video call, or add a new participant. It also allows multimedia communication between two or more end points hence the conversation can be between more stakeholders than two. An end point is typically located in a UE, but can also be located in a network entity. [12]

MMTel allows for interoperable services between operators and towards legacy networks. The standardized interfaces mean that operators can use multiple vendors within a network and integrate with Internet services. It is important to note that MMTel also complies with regulatory requirements associated with voice services, in contrast to the so called Over The Top (OTT) VoIP services. Example of latter are Skype, Viber, WhatsApp. These services do not have standardized specification and they rely heavily on application. While MMTel solutions can be integrated into the mobile terminal modem. There are many benefits of this like by saving battery life by not using power-hungry application processers. The intention of implementing MMTel is to eventually phase out circuit-switched technologies, replacing them with an all-IP solution.

7.2 IMS architecture

IMS is a global, access-independent and standard-based IP connectivity and service control architecture, that enables various types of multimedia services to end users using common Internet based applications. [13]

MMTel as an example is a service set that relies on IMS architecture. IMS is very broad and complex topic and most of it is beyond the scope of current thesis. Yet it is relevant to understand the basics of IMS architecture. IMS is a standardized architecture for controlling and delivering multimedia services that employ IP for transport Session Initiation Protocol (SIP) for service signalling. “Standardized” refers only to the nodes, protocols, interfaces and not to the services delivered on top of it. The fact that standardized architecture does not include any standardized services, has been the main factor that IMS has not been widely spread before VoLTE has become relevant service offering for mobile operators.

The main drivers for deploying IMS architecture has been following:

- Voice over LTE
- PSTN modernization - multimedia telephony over IP
- Fixed Mobile Convergence

PSTN (Public Switched Telephone Network) is global interconnected voice-oriented public telephone network which relies traditionally on circuit-switched technology. The driver for IMS would be moving away from CS based telephony service to all-IP – voice over IP. Fixed Mobile Convergence in this context means that subscribers attached to IMS can be wireless and also fixed because IMS is access network agnostic and standardized for both accesses. An illustration of how IMS as a system is interfaced to EPC is visible in Figure 9. It must be noted that IMS related architecture on this figure is very simplifed but still gives high level overview of EPC and IMS interconnections.

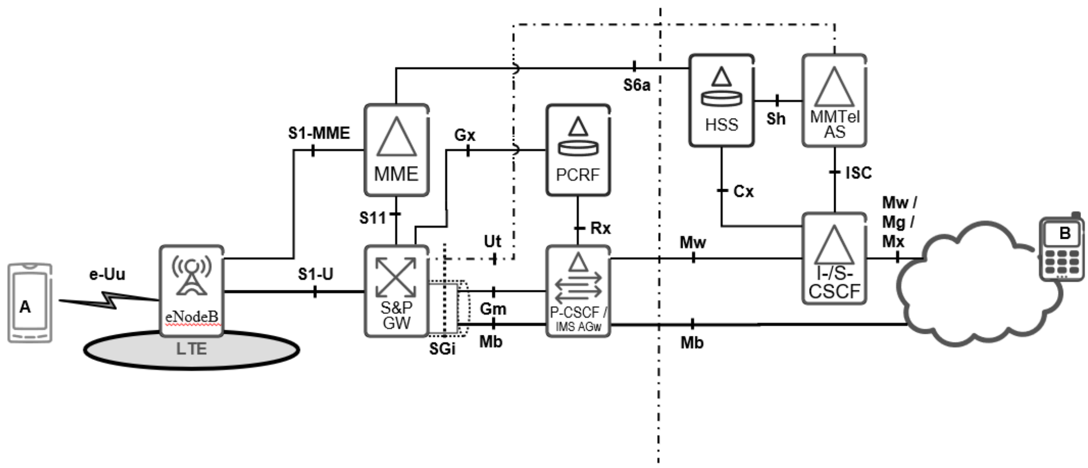


Figure 9. IMS architecture interfaced to EPC. [7]

IMS is defined as a subsystem within the mobile core network architecture. It consists of a number of logical entities interconnected via standardized interfaces. Note that there are many logical entities in IMS that vendors of IMS infrastructure equipment may combine into one single physical or virtual product.

IMS end-to-end framework consists of four main layers which are all visible and distinguishable in Figure 9:

- User Equipment. The UE must be compatible for IMS services and must be prepared by UE vendor for specific mobile operator needs and adjusted based on the characteristics of IMS network of the mobile operator.
- Access Network. The network through which UE is able connecting to EPC and IMS. Originally IMS was designed for GSM/WCDMA but later enhanced for LTE and non-3GPP accesses. eNB represents the access network.
- Core Network. The layer which assures the authentication and authorization of the user and connecting sessions to the relevant application servers. This is the role of EPC.
- Application Layer. Consists of numerous application servers and web servers enabling subscriber to use the services.

At the core of the IMS subsystem is the Call Session Control Function (CSCF) which on high level is responsible for session management and routing. CSCF is the node handling SIP signalling, invoking applications and controlling the media path. The CSCF is logically separated into three different entities:

- The Proxy CSCF (P-CSCF)
- The Serving CSCF (S-CSCF)
- The Interrogating CSCF (I-CSCF)

These three entities may reside as a different software features in the same product as also visible in Figure 9, where S-CSCF and I-CSCF are co-located.

P-CSCF is an entry point to IMS from any access network. The primary role of it is a SIP proxy function. It is in the signalling path between the terminal and the S-CSCF and can inspect every SIP message that is flowing between two endpoints. The P-CSCF manages quality of service and authorizes the usage of specific bearer services in relation to IMS-based services. The P-CSCF also maintains a security association with the terminal and may also optionally support SIP message compression for efficient use of radio resources. [2] The P-CSCF is always located in the same network as the PGW is located. Therefore, both the PGW and the P-CSCF are located either in the visited PLMN or the home PLMN. Note that in roaming scenarios the MME is always located in the visited PLMN.

The S-CSCF is the central node of the IMS archtiecture. It manages the SIP sessions and interacts with the HSS server for subscriber data managent The S-CSCF also interacts with the Appliation Servers. S-CSCF is a stateful SIP server providing session control, acts as a SIP registrar. It is always located in the home network and is a central point for control of operator provided services.

The primary role of I-CSCF is to be the contact point for SIP requests from external networks. Basically it acts as a SIP proxy at the edge of the network. It interacts with the HSS to assign the S-CSCF that handles the SIP sessions for a user.

The rest of the main logical components of IMS architecture are listed below:

- **Multimedia Resource Function** (MRF)
  MRF is split into two functional parts: Multimedia Resource Function Controller (MRFC) and Multimedia Resource Function Processor (MRFP).
- MRFP is a media plane node that can be invoked to process media streams. Examples of the use cases where the media data are routed via MRFP are conference calls.
- MRFC interacts with the CSCF and controls the actions taken by MRFP. It acts as a SIP User Agent and manages the features of the MRFP.
- The **Breakout Gateway Control Function** (BGCF) handles routing decisions for outgoing calls to circuit-switched networks. It normally routes the sessions to a Media Gateway Control Function (MGCF)
- The **MGCF** provides the logic for IMS interworking with external circuit-switched networks. It controls the SIP signalling towards the S-CSCF.
- The **Session Border Controller** (SBC) is an IP gateway between the IMS domain and an external IP network. It manages IMS sessions and provides support for controlling security and quality of sessions. [13]

- **Multimedia Telephony Application Server** (MMTel AS) is the application server for IMS based multimedia telephony (MMTel). MMTel AS supports both basic calls and various telephony supplementary services in a multimedia context. MMTel AS supports the MMTel Supplementary Services as defined in IR.92. MMtel AS has an integrated MRFC but does also support an external MRFC. MMtel AS controls MRFP via MRFC (internal or external) for announcements and voice/video conferencing.

IMS detailed architecture is visible in Figure 10.

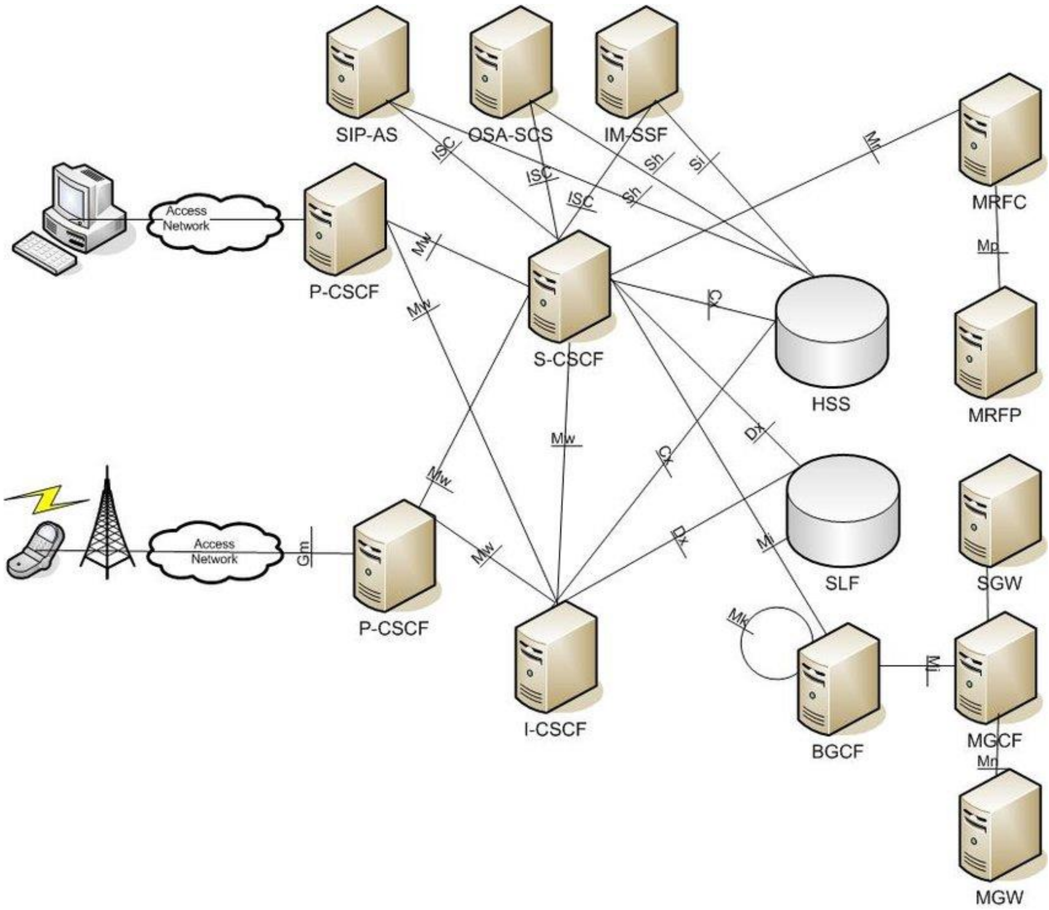


Figure 10. IMS core detailed architecture[14]

There are entities that are not part of IMS architecture, but which are vital in the end-to-end session management and handling. The Application Server (AS) implements a specific service and interacts with the CSCF in order to deliver it to end-users.

The Home Subscriber Server (HSS) is the main data storage for all subscriber and service-related data. It consists of following functionalities: IMS functionality, subset of Home Location Register and Authentication Centre (HLR/AUC) functionality required by the PS and CS domain. HLR functionality is required to provide support to PS domain entities, such as MME. In similar fashion the HLR provides support for CS domain entities, like MSC/MSC servers. The AUC stores a secret key for each mobile subscriber, which is used to generate dynamic security data for each mobile subscriber.

The Policy and Charging Rule Function (PCRF) is responsible for making policy and charging control decisions based on session and media-related information obtained from the application function such as P-CSCF and IMS environment. The PCRF generates charging rules and authorises the IP flows of the chosen media components. Based on available information in the PCRF it makes an authorisation decision which will be enforced in the access network function, e.g. eNB. [13]

8 Voice services setup

Current chapter will give an in-depth overview of voice services implementation in EPC. It will provide the high-level requirements for each solution and provide configuration examples from EPC perspective, not going deeper in IMS nor CS domain. The configuration examples are based on Estonian mobile operator where all the solution options are implemented by the time this thesis is written. Much of the configuration has been done by the author of the thesis or with assistance with author. The trace examples and configuration examples presented below are taken from mobile operator network and based on the thesis author's personal subscriptions.

8.1 Circuit Switched Fallback

CSFB is basically a mandatory solution for all mobile operators who launch LTE radio access network. It is considered as an interim step in moving towards all-IMS based voice. In most cases operators do not have IMS and VoLTE capability at the time LTE is being launched. Even if IMS and VoLTE capability is there, LTE coverage is implemented step by step and there will be geographical areas where subscriber will move from LTE coverage area to the area where is 2G/3G coverage only. In such a scenario CSFB capability is the only option to guarantee continuous voice services. Moreover, even if the LTE coverage is hundred percent in PLMN, as it is quite commonly so, there are still situations where LTE coverage might be lost by UE because the radio quality goes very bad and UE is switched backed to 2G/3G coverage. An example of such case would be the subscriber located indoors in densely populated areas or modern office buildings which are covered with materials shielding radio waves effectively. Relevant factor here is also the frequency band used in LTE. The higher is the frequency, the poorer are the radio waves propagation capabilities.

The CSFB feature allows reusing of CS infrastructure when the UE is served by E-UTRAN. CSFB is only possible in areas where E-UTRAN coverage overlays with GERAN and/or UTRAN. CSFB allows a UE to be reachable for CS voice service even when attached to LTE, after fallback call is handed over to overlapping CS domain.

Before establishing voice call the UE moves to 2G/3G access. After voice call completion UE will move back to LTE access. During CSFB network pages UE over LTE to indicate an incoming call on CS. Paging initiator is MSC which is interfaced to EPC over SGs interface. The SGs interface and combined procedure are the key enablers for CSFB. Combined procedure enables a supporting UE to connect to both Packet Switched (EPC) and Circuit Switched core services through the EPS network. UE requests the MME for combined procedures in *Attach Request* or *Tracking Area Update* (TAU) *Request* message. The MME establishes initial registration of UE and maintains UE location information updated with the MSC/VLR (Visitor Location Register). The MME and the MSC/VLR serving a UE can only communicate if there exist an SGs association for the UE.

VLR is Visitor Location Register which contains the exact location of all mobile subscribers currently present in the service area of the MSC. This means also that MSC/VLR must be informed about the location of UE if it is attached in LTE. This information is necessary to route a call to the right base station – basically to succeed in establishing Mobile Terminating CSFB. In some occasions the Ericsson MME can be configured to prevent from specific UEs from connecting to the CS service through SGs interface in combined procedures. Such an occasion would be if CS services are not available for some UEs. Most of the modern UEs – smartphones, request always combined procedure during attach and are even not registering to PS services when they get response from core that CS service is not available. The reason is to prevent the situation where it is not possible to set up a voice call over CS core when the voice over PS core is not available. For instance, LTE coverage is lost, or Voice over LTE is not present. In such a scenario there would be not possible to even set up emergency call and for that purpose UE rejects the registration procedure while CS domain not available. For dongles which do not use voice services normally, the MME can be configured to restrict CS service to SMS only, that is accept the combined registration with SMS only in the Location Area. The trace example of EPS/IMSI attach with combined procedure Information Element set to “1” by UE in Attach Request is visible in Figure 11. Trace taken in Estonian mobile operator MME and IMSI is author’s personal one. End of the IMSI is hidden for security reasons. Attach Request is inside *Initial UE message,* in NAS (Non-Access Stratum) PDU (Packet Data Unit) message unit. Mobile identity is IMSI but could be as well GUTI (Globally Unique Temporary Identifier).

```
NAS EPS Mobility Management Message Type: Attach request (0x41)
0... .... = Type of security context flag (TSC): Native security context (for KSIasme)
.111 .... = NAS key set identifier: No key is available (7)
.... 0... = Spare bit(s): 0x00
.... .010 = EPS attach type: Combined EPS/IMSI attach (2)
EPS mobile identity
   Length: 8
   .... 1... = Odd/even indication: Odd number of identity digits
   .... .001 = Type of identity: IMSI (1)
   IMSI: 2480222
   [Association IMSI: 2480222
      Mobile Country Code (MCC): Estonia (248)
      Mobile Network Code (MNC): RLE (02)
```

Figure 11. UE combined attach in LTE

The SGs association is created during the combined Attach procedure, or during a combined TAU procedure. In regard to SGs, MME establishes initial registration of UE towards CS network and maintains UE location information updated with the MSC/VLR. The SGs interface between PS and CS core networks is illustrated in Figure 12. Transport protocol for SGs is SCTP and application layer protocol SGs Application Part (SGsAP).

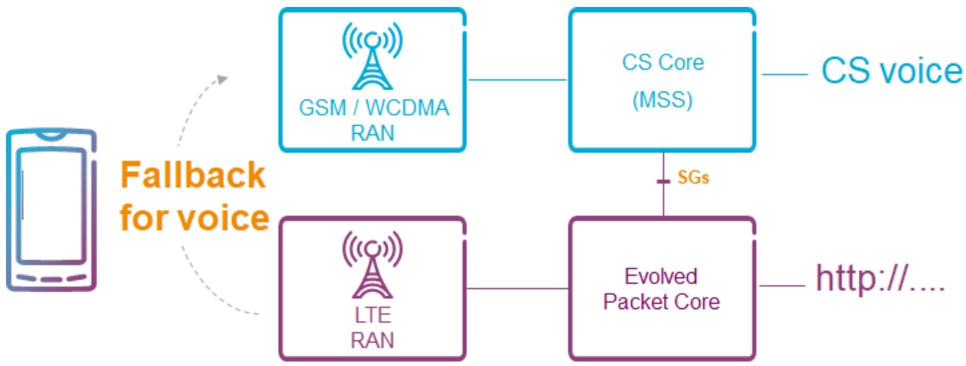


Figure 12. The SGs interface. [7]

To guarantee successful and smooth CSFB there is a need to configure TA-LA (Tracking Area – Location Area) mapping in the MME. Main purpose of this is that MME must be able to notify MSC about UE location in granularity of LA. In the CS domain the mobile network radio coverage area is divided into LAs, supporting GSM/WCDMA. For EPS the LTE coverage is divided into different TAs. The SGSN-MME keeps the Mobile services MSC/VLR updated with location information for the UE. The SGSN-MME knows in which Tracking Area (TA) the UE is, but since the MSC/VLR needs to be updated with the correct Location Area (LA), each TA served by the SGSN-MME needs to be configured to map to an LA in order to send correct LA information to the MSC/VLR. The LA and the TAs are mapped by connecting them to the same geographical area. The LAs and TAs often cover the same geographical area which means the UE can access to any radio if it has relevant support. The MSC/VLR does not have any information about TAs, so the MME sends location information based on the LAs in which the UE is located. That is why the configuration of TA-LA mapping is required in MME. [7]

It is common that an LA is greater than a TA. In this case several TAs are needed to cover the same geographical area as on LA is covering. It is possible that the area covered by a TA is covered by several LAs, which means a TA can map to more than one LA. An example of TA-LA mapping is shown in Figure 13.

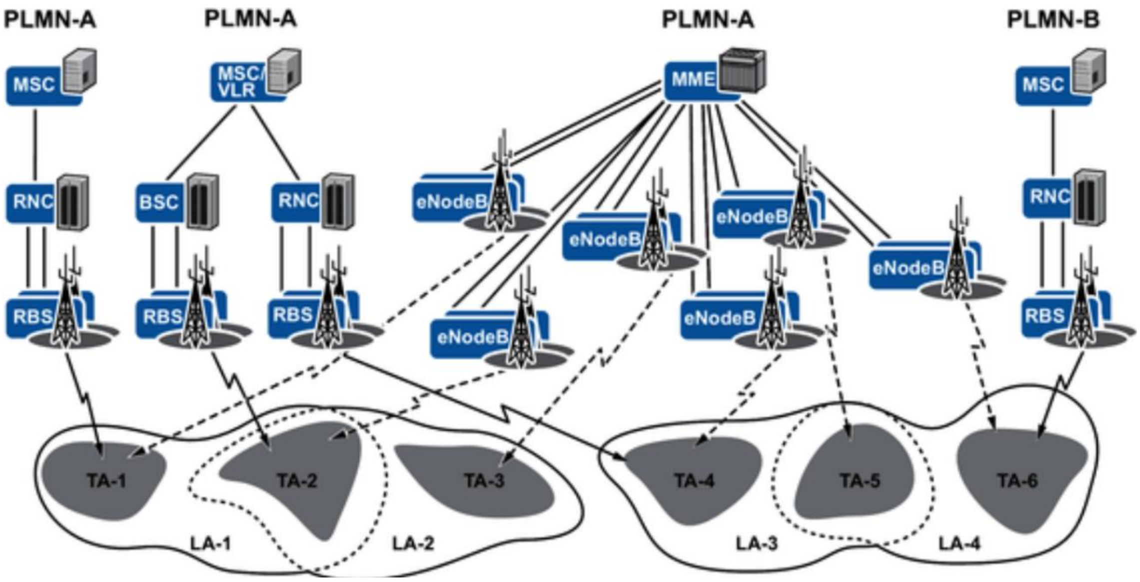


Figure 13. TA-LA mapping example [7]

In this example TA-1, TA-3, TA-4 and TA-6 ach map to only one LA. TA-2 maps to LA-1 and LA-2. TA-5 maps to LA-3 and LA-4. In this example LA-3 and LA-4 belong to different PLMNs. In this kind of LA mapping CSFB to GSM/WCDMA works fine but it is still recommended to avoid situation where e.g. TA-2 partly belongs to LA-1 and LA-2.

A faulty or bad TA-LA mapping causes extra signalling and even dropped calls. Assume that the TA-LA mapping indicates that the UE belongs to an LA defined in one MSC/VLR, but the UE is in an LA belonging to another MSC/VLR. The Mobile Terminated (MT) call is routed through the wrong MSC/VLR. The handover of the call between the MSC/VLRs is necessary as the UE tries to connect to the GSM or WCDMA network. A handover between MSC/VLRs takes time and results in extra Home Location Register updates. The delay can also cause the originating party to hang up, before the receiver gets a chance to answer the call.

There are two main options for CS fallback realization:

- RRC (Radio Release Control) Release with Redirect. In this case the source network releases the RRC connection indicating the UE to switch to UTRAN/GERAN.
- Packet Switched Handover (PSHO). In this case upon fallback PSHO is performed to fallback RAT – UTRAN/GERAN.

If during fallback time there is ongoing active data transfer in PS domain, for example video stream, then the impact to the session is quite different in the two scenarios. In the Redirection option some packets are lost during the radio access change from source to target network. In the PSHO case PS bearer is prepared first in the target cell before CSFB takes place and hence the packets will be not lost. There can be just a minimal delay in data transfer which usually is not noticeable to the service user. The signalling flow for these two options is quite different during the CSFB. In the following chapters 8.1 and 8.2 the both options are covered.

As described above CSFB can be made to WCDMA and GSM radio networks. Therefore, a strategy is needed to define which RAT and which frequency the UE should be released to when CSFB takes place. The eNB takes the decision based on the RAT priorities defined in LTE. WCDMA and GSM should have different priorities and only if UE does not support the target RAT with first priority the second RAT would be chosen.

A feature called Dual Transfer Mode (DTM) should be supported by UE and target systems to achieve good quality fallback. In case of DTM is supported PS services may continue in the target system during CS Fallback. If DTM is not supported PS services cannot be transferred to target system and will be suspended. Differences of supported DTM and non-supported DTM is described in the chapter describing SRVCC signalling flow, chapter 8.2. After the CS call is finished, the UE may switch back to LTE or remain in GSM/WCDMA but enhanced CSFB solutions are designed to move back to LTE and as fast as possible. In case of non-DTM the PS service resumes.

Ericsson enhanced CSFB solution includes features Deferred Measurement Control (DMCR) and RAN Information Management (RIM) transfer. Both features must be supported by EPC in SGSN and MME. Both features are implemented to reduce the call setup time. For example, during system change the UE reads target system information before it is able to switch to target RAT. Without DMCR support the maximum allowed reading time is 1280 ms, while for DMCR supported case it is 640 ms in worst case. The RIM feature provides means to request and forward radio-related system information between the GERAN/UTRAN and the eNodeB through the SGSN-MME. The system information, referred to as CSFB information, is included in containers as a part of the RIM messages. The content of the RIM messages is transparent to the SGSN-MME. [7]

The call set-up time for a call using CSFB depends of the implemented CSFB variant. At CSFB to WCDMA, RRC Connection Release with Redirect using Deferred Measurement Control Reading (DMCR) gives approximately the same CS call setup time as if the system information can be provided to the UE already in LTE, the RIM case. DMCR has less impact on the existing WCDMA network in comparison of the introduction of RIM. Therefore, DMCR is the preferred solution.

Table 1 illustrates the measured performance figures (CS call setup times) for the different CSFB alternatives. These measurements are done under lab conditions and show average values of a series of test calls. As can be seen the difference for setting up RIM compared to DMCR is just 0.3 seconds in average. [15]

Table 1. CSFB Call Setup time characteristics. [15]

	Call setup times by CSFB (s)	**Additional call setup delay compared to CS call setup**
RRC Connection Release with Redirect to UTRAN, <u>Sysinfo read in UTRAN</u>	MO 5.4 MT 6.7	MO 2.7 MT 2.7
RRC Connection Release with Redirect to UTRAN, <u>DMCR UTRAN support</u>	MO 3.7 MT 5.0	MO 1.0 MT 1.0
RRC Connection Release with Redirect to UTRAN, <u>RIM transfer</u>	MO 3.4 MT 4.7	MO 0.6 MT 0.6
Release with Redirect to GERAN, no DTM	MO 5.5	2.1
Release with Redirect to GERAN, <u>RIM transfer</u>	MO 4.3	0.9

8.1.1 CS Fallback Mobile Originated

This chapter gives an overview of CSFB signalling in mobile originating scenario with PS Handover and without PS Handover. Figure 14 illustrates the call flow without PSHO and Figure 15 with PSHO.

CS Fallback Mobile Originated **without PS Handover** procedure signalling is described as follows. For CSFB MO first step is to perform combined procedures between UE and MME (message nr 1 in figure 14). Next the UE sends *Extended Service Request* message to the MME indicating that it needs to exit E-UTRAN and redirect to WCDMA/GSM (message nr 2). UE registration state (ECM state) at that time can be either active or idle, ECM-CONNECTED or ECM-IDLE respectively. ECM is EPS Connection Management state. When the MME receives *Extended Service Request* it sends *SGsAP-MO-CSFB-INDICATION* message to the MSC/VLR with service type inside that message set “MO CSFB” or “MO CSFB emergency call” (message nr 3). Depending on UE ECM state MME sends *Initial Context Setup Request* (ECM-IDLE) or *UE Context Modification Request* to eNB and receives response from eNB (messages 4-5). The eNodeB triggers a *Radio Resource Control* (RRC) connection release message to the UE, to redirect the UE to WCDMA/GSM (message 6). The eNodeB sends a *UE Context Release Request* message with a cause code indicating CS fallback, to the MME (message nr 7). If the ECM state was ECM-CONNECTED, MME needs to delete bearer from SGW and thus sends *Release Access Bearer Request* message to SGW and gets response from SGW (messages 8 and 10). In ECM-IDLE state that procedure is not applicable. To release S1 connection MME sends *UE Context Release Command* to eNB and gets response to that from eNB (messages 9 and 11). If target radio network does not support DTM, the MME sends a *Suspend Notification* to SGW. The SGW marks the bearers as suspended status and forwards the message to PGW. The PGW responds to SGW with *Suspend Acknowledge* and SGW forwards message to MME (messages 12,13). At that point UE is transferred completely to WCDMA/GSM network and deleted from EPS. [7]

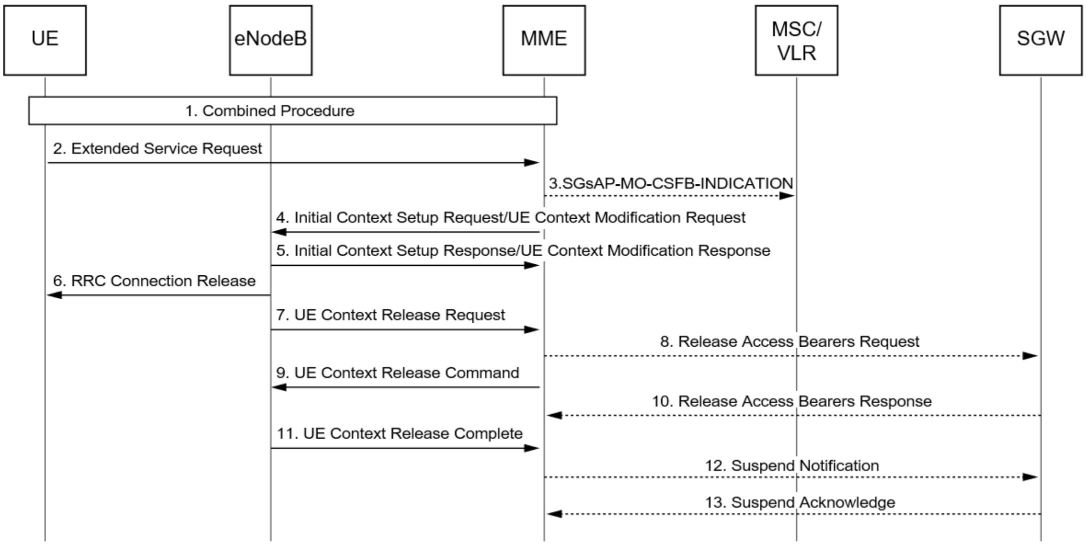


Figure 14. CSFB Mobile Originated without PSHO [7]

CS Fallback Mobile Originated **with PS Handover** procedure signalling is described as follows. Messages from one to five are identical as in without handover case. After that IRAT (Inter Radio Access Type) HO procedure to WCDMA over Gn interface is performed. When a UE in the LTE radio network enters a WCDMA/GSM routing area which is served by the SGSN, it initiates a RAU (Routing Area Update) to SGSN procedure. The Gn interface connects the MME to the SGSN. The MME converts EPS bearer context information to PDP context (Packet Data Protocol Contexts) and sends it in a GTPv1 format over the Gn interface that can be interpreted by the SGSN. The RAU to SGSN and SGSN Context exchange can be seen in Figure 15 as messages nr 1-5. Next the SGSN asks GGSN to update PDP context (messages 6-7). PDP context is a data structure that allows the device to transmit. data using Internet Protocol. It includes the device's IP address, IMSI and additional. parameters to properly route data to and from the network. After these steps, communication between SGSN and HLR takes place. SGSN informs HLR that it does not have subscription data for UE and sends *Update Location Request* message to HLR. The HLR acknowledges the renewal of UE location and responds to SGSN with *Update Location Answer* (messages 8-11). After that the SGSN validates the UE presence in the new routing area and sends a *RAU accept* message to UE. The UE acknowledges *RAU accept* and responds with *RAU complete* (messages 12-13).

Like in the case without PSHO, there are subsequent activities in LTE domain which is not visible in Figure 15. The MME removes the bearers and sends a *UE Context Release*

Request to eNB and also MME deletes the context in the SGW for PDN connections and bearers that have been transferred to the SGSN previously. Hence MME sends *Delete Session Request* to SGW and SGW responds with *Delete Session Response*. As a final step for PSHO, the MME deletes EPS bearers internally and states UE state to EMM-DEREGISTERED. EMM is EPS Mobility Management state. The state means UE is detached in MME after set to EMM-DEREGISTERED.

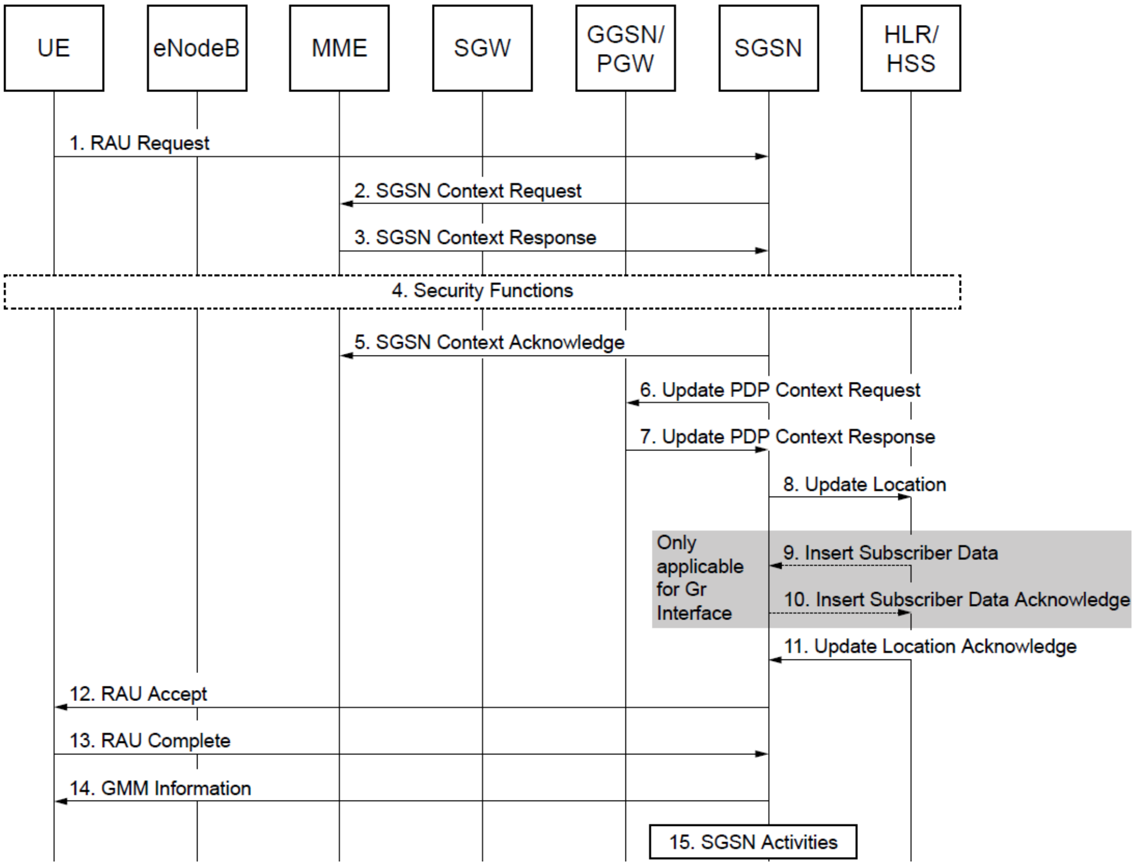


Figure 15. CSFB Mobile Originated with PSHO. [7]

8.1.2 CS Fallback Mobile Terminated

CSFB MT flow is quite similar like the signalling for the Mobile Originating case. The beginning of signalling differs a bit and the focus is set on that in this chapter. The whole CSFB MT signalling is visualized in Figure 16. The example is taken from Mobile Operator MME with thesis author’s IMSI, which has been hidden, using UE tracing software feature.

The whole procedure is started by MSC who sends *SGsAP Paging Request* to MME over SGs interface. MME sends paging to eNB and eNB forwards it to UE. The Paging by MME includes “CSFB indicator”. As the UE has been in ECM-IDLE state it sends *Initial*

UE message together with *Extended Service Request* to MME, indicating acceptance of the MT call. The MME sends *SGsAP Service Request* to MSC, indicating acceptance of the MT call. Since now the messaging is identical to Mobile Originated case, exchanging *Initial Context Setup* and *UE Context Setup* messages between eNB and MME. And so on until the end of procedure.

Node 1		Node 2	Interface	Protocol	Message
MME	<-	MSC	SGs	SGsAP	SGsAP paging request
eNodeB	<-	MME	S1	S1AP	Paging
eNodeB	<-	MME	S1	S1AP	Paging
eNodeB	<-	MME	S1	S1AP	Paging
eNodeB	->	MME	S1	S1AP	Initial UE message
UE	->	MME	S1	NAS	Extended service request
MME	->	MSC	SGs	SGsAP	SGsAP service request
eNodeB	<-	MME	S1	S1AP	Initial context setup request
eNodeB	->	MME	S1	S1AP	Initial context setup response
eNodeB	->	MME	S1	S1AP	UE context release request
eNodeB	<-	MME	S1	S1AP	UE context release command
eNodeB	->	MME	S1	S1AP	UE context release complete
MME	<-	SGSN	Gn_Gp	GTP	SGSN context request
MME	->	SGSN	Gn_Gp	GTP	SGSN context response
SGSN	<-	SGSN_MME	Gn_Gp	GTP	SGSN context response
SGSN	->	RNC	Iu	RANAP	Security mode command
SGSN	<-	RNC	Iu	RANAP	Security mode complete
SGSN	->	SGSN_MME	Gn_Gp	GTP	SGSN context acknowledge
MME	<-	SGSN	Gn_Gp	GTP	SGSN context acknowledge
SGSN	->	GGSN	Gn_Gp	GTP	Update PDP context request
SGSN	<-	GGSN	Gn_Gp	GTP	Update PDP context response
SGSN	->	HLR	Gr	MAP	Update location request
SGSN	<-	HLR	Gr	MAP	Insert subscriber data request
SGSN	->	HLR	Gr	MAP	Insert subscriber data answer
SGSN	<-	HLR	Gr	MAP	Update location answer
UE	<-	SGSN	Iu	MM	Routing area update accept
SGSN	->	RNC	Iu	RANAP	Direct transfer
SGSN	<-	RNC	Iu	RANAP	Direct transfer
UE	->	SGSN	Iu	MM	Routing area update complete

Figure 16. CS Fallback Mobile Terminated signalling in MME

8.1.3 Mobile Operator configuration

All the configuration for enabling CS Fallback in EPC is done in MME. SGW, PGW and PCRF have no information about combined procedures. The configuration is made quite simple and straightforward by equipment vendor. There are three features to be turned to enabled status to enable full SGs functionality. These features are *„sms over sgs“*, *„csfb to wg“* and *„MSC pool SGs*. MSC pool concept is similar to MME pool. In this case all the MMEs in the MME pool are connected to all the MSCs in the MSC pool.

SMS over SGs must be turned on because it will activate listening SGsAP protocol in MME. The SMS over SGs feature provides a transparent transfer of SMS between the

UE and the MSC/VLR through signalling relay by the MME. The benefits of the feature is that allows the UE to remain in the existing LTE network and enables the UE to use SMS concurrently with other ongoing services in EPS. CSFB to WCDMA with PS HO also requires that the "*Packet Handover"* feature is activated.

In addition to enabling features, on the high level the following configuration steps are done:

- IP-based interface configuration
- SCTP-based interface configuration
- TA-LA mapping
- SGsAP configuration

It is out of the scope of the thesis to cover interface connectivity based configuration, like IP addresses and SCTP parameters. For TA-LA mapping, the Geographical Area (GA) is configured. Within a Geographical Area the Tracking Area ranges are configured. The ranges contain the TAs mapping to one or more LAs. After that is created LA connected to the desired Geographical Area. LA as identifier itself consists of MCC (Mobile Country Code), MNC (Mobile Network Code), Location Area Code and also optionally access type, whether WCDMA or GSM. When a TA is configured to map to more than one LA then many restrictive rules apply for LA configuration.

Figure 17 gives an overview of the operator configuration about TA-LA mapping. It is not full configuration but a portion of it. The TA-LA mapping is done by connecting the TA and the LA to the same Geographical Area (GA), using the Geographical Area Name (GAN). In the figure can be seen there is created four location areas and four different Geographical Areas – marked with parameter "-gan". Each GA involves number of Tracking Areas. For example, GA "RNC201LA1021" does involve four Tracking Areas, which is marked with "first" and "last" in the configuration. Location Area with Location Area Code 1021 (lower block of the figure) is mapped with GA "RNC201LA1021". Like this, all the LAs are mapped to TAs through the Geographical Area.

```
ps Class          Identifiers                           | mcc  mnc  first  last
--------------------------------------------------------------------------------
A  ga_ta_range   -gan  RNC201LA1021  -tan  2021         | 248  02   2021   2021
A  ga_ta_range   -gan  RNC201LA1021  -tan  2921         | 248  02   2921   2921
A  ga_ta_range   -gan  RNC201LA1021  -tan  3021         | 248  02   3021   3021
A  ga_ta_range   -gan  RNC201LA1031  -tan  2031         | 248  02   2031   2031
A  ga_ta_range   -gan  RNC201LA1031  -tan  2931         | 248  02   2931   2931
A  ga_ta_range   -gan  RNC201LA1031  -tan  3031         | 248  02   3031   3031
A  ga_ta_range   -gan  RNC201LA1041  -tan  2041         | 248  02   2041   2041
A  ga_ta_range   -gan  RNC201LA1041  -tan  2941         | 248  02   2941   2941
A  ga_ta_range   -gan  RNC201LA1041  -tan  3041         | 248  02   3041   3041
A  ga_ta_range   -gan  RNC201LA1051  -tan  2051         | 248  02   2051   2051
A  ga_ta_range   -gan  RNC201LA1051  -tan  2951         | 248  02   2951   2951
A  ga_ta_range   -gan  RNC201LA1051  -tan  3051         | 248  02   3051   3051

ps Class Identifiers                            | at     cssr  defsvmsc   gan           defsgsmsc  defmsc
-------------------------------------------------------------------------------------------------------------
A  la    -mcc  248  -mnc  02  -lac  1021      | NULL  NULL  undefined  RNC201LA1021  EMSS4      EMSS4
A  la    -mcc  248  -mnc  02  -lac  1031      | NULL  NULL  undefined  RNC201LA1031  EMSS4      EMSS4
A  la    -mcc  248  -mnc  02  -lac  1041      | NULL  NULL  undefined  RNC201LA1041  EMSS4      EMSS4
A  la    -mcc  248  -mnc  02  -lac  1051      | NULL  NULL  undefined  RNC201LA1051  EMSS4      EMSS4
```

Figure 17. TA-LA configuration example.

For SGs related configuration there is a need to configure basic entities and number of different parameters depending on the wanted solution. Example of the basic configuration is creating MSCs in the MME configuration and connecting created MSC to a Location Area. When a combined registration procedure is performed, the MME selects an MSC/VLR configured for the indicated or selected LA. If some of the LAs is associated to more than one MSC, which is quite common case, then default MSC per LA must be configured. In the configuration example in Figure 17, the default MSC is "EMSS4 ", as can be seen from LA configuration.

When a TA is configured to map to more than one LA, the node level default LA selection profile must be configured which is complex configuration and not looked deeper. If MME is interfaced to MSC Pool instead of many sole MSCs then MSC Pool support feature will be configured in MME. When TA-LA mapping is performed, an LA can only belong to one MSC/VLR if MSC pools are not used. If MSC pools are used, an LA can belong to several MSC/VLR in the pool. Several LAs can belong to the same MSC/VLR. The Mobile Operator is using MSC Pool which makes the CS services more redundant and flexible. For example, if one MSC needs to be taken into maintenance it does not cause possible impact to end users as other MSCs in pool will carry all the traffic.

For achieving SGs connectivity to MSC SGs peer must be configured. One MSC can theoretically own more than one SGs peer. There are many parameters that can be configured to drive the SGsAP protocol behaviour. One relevant parameter is *SGsRecoveryMscUnreachable* which specifies method the MME uses to reselect an MSC to recover the SGs association when the MME detects that the MSC the UE has registered is unreachable. The operator has set it to "none" which means MME does not take any action. This value protects the MME processor load from going to high load if adjacent MSC should restart but is not the most desired option for end user as it can cause disturbance during CSFB call set up time. If the *SGsRecoveryMscUnreachable* parameter is set to “signalling”, the MME triggers MSC reselection based on UE signalling. Looking through all the SGsAP parameters is beyond the scope of the thesis.

It is worth to note that there is an option to configure Dummy CS service in MME. That is needed for the terminals that need CS service availability through combined procedures to get registered to packet services, but CS service is not available through PS network. That way the terminal is still able to register to EPC even if CS domain is not available.

8.2 Single Radio Voice Call Continuity

As briefly described in Chapter 7, the SRVCC is a solution that offers a mechanism where the UE performs a coordinated radio level handover in combination with a change from IMS VoIP to circuit switched voice using IMS procedures for service continuity. [2]. For SRVCC, single radio means that the UE transmits and receives signals on only one radio access at a given time (LTE or WCDMA) in order to minimize power consumption and radio emission. A single radio UE scans and measures the signal quality of the LTE radio access and sends the measurements to the serving radio network. If the signal quality of the LTE radio is below a certain threshold the eNB will request the MME to make a handover to the UTRAN network.

There are different variants of SRVCC and these are introduced in Table 2.

Table 2. SRVCC versions.

Common Name	3GPP Release	Description
Basic SRVCC	Rel 8	Call Continuity from E-UTRAN to UTRAN/GERAN
aSRVCC	Rel 10	PS to CS call transfer during Alerting Phase
eSRVCC	Rel 10	Enhanced SRVCC. Support for shortened call setup delay time
bSRVCC	Rel 10	SRVCC at Pre-Alerting Phase
vSRVCC	Rel 11	Video SRVCC
rSRVCC	Rel 11	SRVCC from UTRAN/GERAN to E-UTRAN

SRVCC enables transfer of voice calls between PS access and CS access when UE is not capable of transmitting/receiving in both access networks simultaneously (hence "single radio"). The SRVCC feature requires explicit enhanced support in many entities: in the UE, the eNB, the MME, the MSC, the HSS and the PCRF. It is also required that the voice call has been set up as an IMS telephony session and that the call remains anchored in IMS after the transfer. The anchoring point in IMS is logical entity called Service Centralization and Continuity Application Server (SCC-AS). The SCC-AS coordinates access transfer when an ongoing VoLTE session needs to change access to the Circuit Switched network of either WCDMA or GSM.

The UE must be able to handle SRVCC, which is described in 3GPP TS 36.300 for interaction between UE and LTE and in 3GPP TS 25.331 for the interaction between UE and UTRAN. UE must be explicitly configured for IMS speech service support by mobile operator so that UE indicates to the core network it is SRVCC capable. In reality it means UE vendor together with mobile operator must create the specific operator-based SW and enable that in the terminal. Once the SW is enabled for one model of the UE there will be usually support for all the coming next models.

The MME is doing checking of UE SRVCC capability and performs the PS bearer splitting by separating the voice PS bearer from the non-voice PS bearers. For instance, for vSRVCC, the MME identifies the vSRVCC marked video PS bearer in addition to the voice PS bearer and handles the non-voice PS bearer handover with the target cell. The MME initiates and coordinates the handover preparation and execution for the voice bearers during SRVCC procedures. The MME also releases E-UTRAN resources after handover completion.

In addition to the standard MSC procedures, an MSC supporting SRVCC must also support the relocation preparation procedure, which is requested by the MME for voice components of a call. The MSC must also coordinate the CS handover session transfer procedure. The MSC must also negotiate with the SCC-AS to determine whether it should perform an SRVCC or vSRVCC procedure. In order to support rSRVCC, the MSC informs the RAN about the possibility of performing CS to PS SRVCC by sending a "CS to PS SRVCC operation possible" to the UTRAN/GERAN. The possibility of performing such a handover is based on the capability of the UE and the registration of the UE within the IMS.

The HSS stores the SRVCC related flags for subscribers. During the E-UTRAN attach procedure, the MME downloads the Session Transfer Number for SRVCC (STN-SR) and SRVCC flag from the HSS. STN-SR is s routing number used to identify an SCC-AS as part of SRVCC handover procedure. It is a unique number that is generated by UE and stored in HSS. For rSRVCC sessions, the MSC downloads the “CS to PS SRVCC allowed” from the HSS during the attach procedure.

The eNB selects a target cell during the SRVCC handover and must be able to send an indication to the MME that SRVCC is required for handover. The PCRF enforces specific QoS (Quality of Service) principles while IMS session is anchored in the SCC-AS. Going in detail to QoS is beyond of scope of the thesis. [2]

For CS and PS interworking scenarios, like SRVCC, UE capabilities and configuration play very important role for which services and domains are used by the subscriber and this principle is called Voice Domain Preference. It is UE’s mode of operation that has decisive role how the UE relates to CS and PS domain. The UE’s mode of operation reflects two things: the domain it is attached to and whether it prefers voice or data. “Voice” in this context can mean Circuit Switched Telephony or Voice over IP. A UE in PS mode is registered to EPS services only, while a UE in CS/PS mode is registered in both, PS and CS domain. The UE mode of operation is decided during initial attach and it informs the network about its preferences by including the “Voice domain preference and UE usage setting” Information Element (IE) in the *Attach Request*. The attach type in *Attach Request* can be “Combined EPS/IMSI attach” or “EPS attach”.

It is upon the end user (or application) configuration whether UE prioritises either voice or data, called “voice centric” or “data centric” respectively. The UE can be configured to use only CS voice, only IMS PS voice or both. Configuration here does not mean any setting in the terminal menu but firmware configuration. There are four different modes that apply for CSFB and IMS/SRVCC capable UE to make an appropriate domain selection:

- *CS Voice only.* The UE will not attempt to initiate voice sessions over IMS. The UE will attempt combined EPS/IMSI attached.

- *CS Voice preferred, IMS PS Voice as secondary*. The UE will try to use the CS domain for voice call. The UE will attempt combined EPS/IMSI attach and if combined attach fails for the CS domain, the UE attempt voice over IMS.
- *IMS Voice preferred, CS Voice as secondary*. The UE will try to use IMS to originate and terminate voice sessions. If the UE fails to use IMS for voice, then the services are provided using the CS domain. The UE can either perform combined EPS/IMSI attach or EPS attach when attaching to E-UTRAN.
- *IMS PS Voice only*. The UE will not attempt combined EPS/IMSI attach and performs IMS registration indicating support for voice. [6]

The before mentioned Information Element "Voice domain preference and UE usage setting" is visible in Figure 18 in *Attach Request* message which shows the typical VoLTE capable UE during combined attach. As can be seen the UE is voice centric and preferred voice is IMS PS, while CS voice is secondary.

```
    .... .010 = EPS attach type: Combined EPS/IMSI attach (2)
  > EPS mobile identity
  > UE network capability
  > ESM message container
  > Tracking area identity - Last visited registered TAI
  > DRX Parameter
  > MS Network Capability
  > Location area identification - Old location area identification
  > Mobile station classmark 2
  > Mobile station classmark 3
  > Supported Codec List - Supported Codecs
  v Voice Domain Preference and UE's Usage Setting
      Element ID: 0x5d
      Length: 1
      0000 0... = Spare bit(s): 0
      .... .0.. = UE's usage setting: Voice centric
      .... ..11 = Voice domain preference for E-UTRAN: IMS PS voice preferred, CS Voice as secondary (3)
```

Figure 18. UE Voice Domain Preference in EPS.

The key enable for SRVCC procedure is Sv interface that connects the MME with an MSC server which is enhanced for SRVCC. The sole purpose of Sv is to support handover of E-UTRAN IMS VoIP calls to the UTRAN/GERAN CS domain. SGs is another interface between MME and SGs but in addition to CSFB it has more functions, e.g. SMS over SGs. The Sv interface is based on the GTPv2-C protocol, like many other EPC related control plane interfaces but makes use of an extension to the GTPv2-C protocol.

It should be noted that SRVCC is defined as a service. It is possible that subscription data in the HSS dictate that the service is not allowed under certain circumstances, such as when the UE is in a visited network. The SRVCC feature transfers an ongoing call between the PS and CS domain, which is in contrast to the CS Fallback feature, which transfers a call even before it has been established. The transfer of a call from the EPS domain to the CS domain with SRVCC is similar to a normal handover.

On the high level the main steps for SRVCC service are following. UE detects the LTE frequency and attaches via LTE to EPC network, sets up the IMS APN and registers with IMS APN to the IMS Core. Usually the name of the APN, or more specifically the APN-NI (APN Network Identifier) is “IMS”. During attach UE also indicates SRVCC support to the core network. UE makes or receives the call over IMS. LTE radio network triggers the SRVCC procedure. UE tunes to WCDMA/GSM and resumes the voice call in CS domain. Any PS bearers are either released, suspended or moved. The latest SRVCC enhancements starting from 3GPP Release 10 enable to transfer PS bearers without significant interruption using PS HO procedure. The aim for voice call is to secure speech interruption time below 300 milliseconds.

The detailed signalling is illustrated in Figure 19 and Figure 20. SRVCC procedure with PS HO is divided into two phases: handover preparation, and session transfer and handover completion. Figure 19 shows the preparation phase and Figure 20 the completion part. Note that SRVCC target RAN can be UTRAN or GERAN but in this example we look only into transferring call to UTRAN which is more commonly used scenario. SRVCC signalling is quite different in case of using PS HO compared to SRVCC without PS HO, using Dual Transfer Mode only.

The handover preparation is started by UE who sends measurement reports to the source E-UTRAN, eNB (msg nr 1). Based on UE measurement reports, the source E-UTRAN decides to trigger an SRVCC access transfer to the CS domain (activity nr 2 in figure). E-UTRAN sends a *Handover Required* message including the “SRVCC Handover Indication” IE to the source MME. The “SRVCC Handover Indication” indicates to the MME that a CS/PS handover is required. The “SRVCC HO” indicator indicates to the MME that this is an SRVCC access transfer operation for the voice part towards the CS domain (msg nr 3). For next step, the MME splits the voice bearer from the non-voice bearers and initiates, for the voice bearer, the relocation toward the MSC and the SGSN, respectively (activities nr 4 and 5,6). The bearer splitting is relevant step during SRVCC HO and it is done based on QoS Class Identifier (QCI) and “SRVCC Handover

Indication”. QCI is the criteria for telling whether a UE has active voice bearers. If the UE have at least one bearer with the value of QCI=1 it has a voice bearer and MME takes actions based on that. QCI meaning in EPS is described in chapter 8.3.

Then MME initiates the PS-to-CS Handover procedure for the voice bearer by sending an *SRVCC PS to CS Request* message to the MSC server. Address of the MSC Server is configured in the MME. The MME decides about the MSC where to send PS to CS Request based on LA identity derived from target RNC ID received from E-UTRAN (msg 5). The message includes information relevant to the CS domain only, including parameters like Target RNC ID, MME Tunnel ID, IMSI, Sv flags, supported codecs, CS keys, etc. Then the MME sends a *Forward Relocation Request* message to the target SGSN and this message contains Packet Data Protocol Contexts (PDP Contexts) which is a 3G term for data bearer (msg nr 6). Relocation procedure requests resource allocation in target RAN (msg nr 7,8). Next the SGSN and MSC send response messages to MME for previous request messages (msg nr 9 and 10). The source MME synchronizes the two prepared relocations and sends a *Handover Command* message to the source E-UTRAN. The E-UTRAN forwards *Handover Command* to UE and the UE tunes to the target UTRAN cell and with that HO preparation is ended.

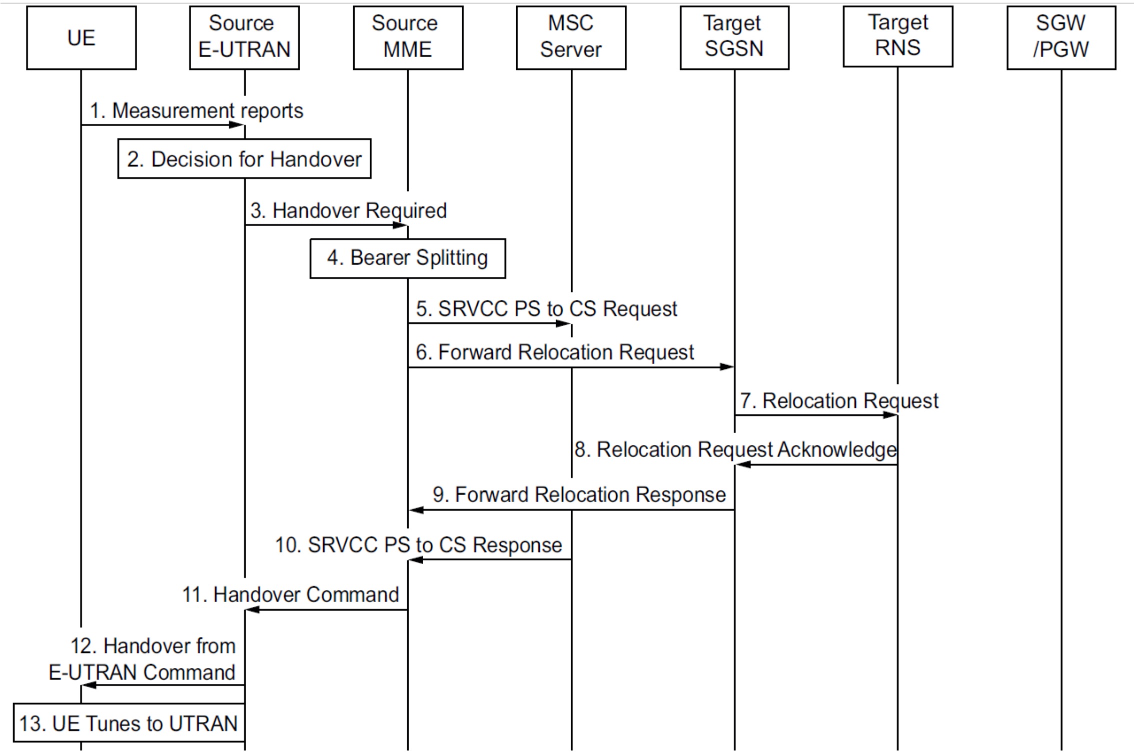


Figure 19. SRVCC Handover preparation. [7]

Handover completion, visible in Figure 20, starts with *Handover Detection* occurrence at the target UTRAN. The MSC sends *an SRVCC PS to CS Complete Notification* message to the source MME (msg nr 15) informing it that the UE has arrived on the target side and MME acknowledges that (msg nr 16). The acknowledgement finishes the PS to CS Handover negotiation procedure between MME and MSC which was started in message nr 5. Messages nr 17-19 are about deleting the so far existing voice bearer in MME and SGW/PGW. The MME asks SGW/PGW to delete the bearer by *Delete Bearer Command* and sets the "PS to CS handover indicator" in the message. The SGW/PGW sends a *Delete Bearer Request* to source MME and MME responds with *Delete Bearer Response.* The target RNC sends a *Relocation Complete* message (msg nr 20) to the target SGSN ending the Relocation procedure between RNC and SGSN, started by *Relocation Request* in message nr 7. For next, the target SGSN sends a *Forward Relocation Complete Notification* message to the source MME and MME acknowledges that (msg nr 21,22). Also, timer is started in the MME to supervise when resources in the eNB and PGW can be released. When the timer expires, the MME sends a *Delete Session Request* to SGW/PGW and *Release Resources* to eNodeB. By that all the resources are released in EPS system – in eNB, MME and SGW and PGW. [7]

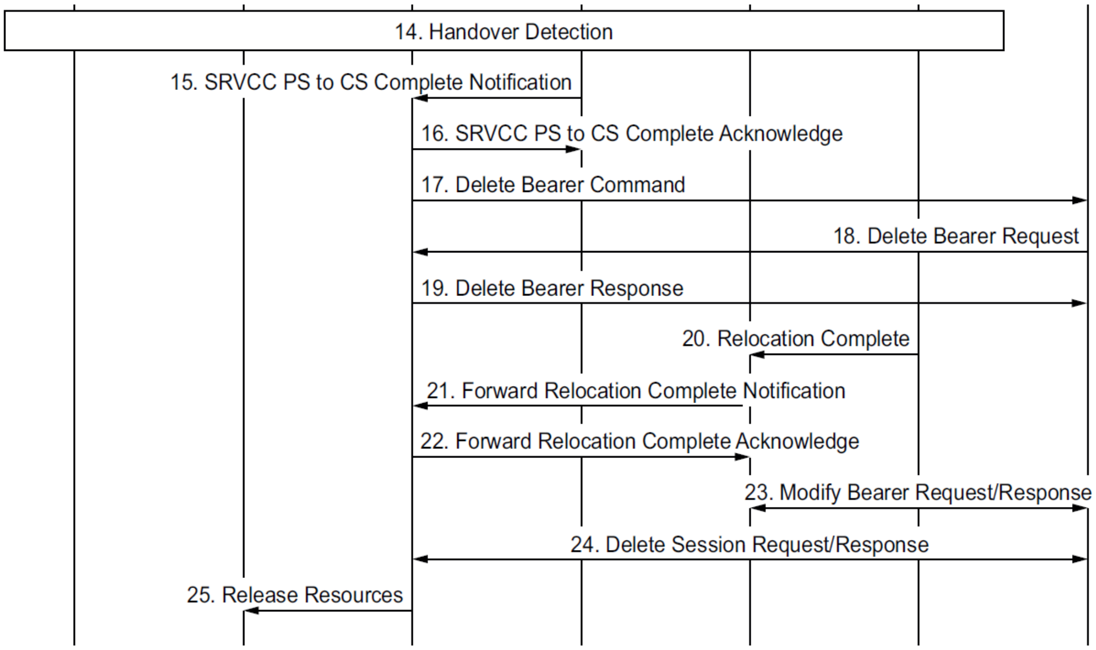


Figure 20. SRVCC Handover completion. [7]

It is relevant to highlight that SRVCC makes possible the SRVCC capability exchange in the signalling. UE makes its SRVCC capability visible in the *Attach Request* and MME does the same including "SRVCC operation possible" indication to the eNB in *Initial Context Setup Request* or *Handover Request* message.

8.2.1 Mobile Operator Configuration

First thing in SRVCC configuration procedure is to configure Sv interface IP connectivity in MME towards MSC server(s). While for SGs case there was a need to configure SGs peer(s) with its IPs, then for Sv the target MSC IPs are configured with the same config command than creating MSC itself. If network is planned to support both CSFB and SRVCC, it is recommended the MSC used for SGs interaction is also configured for Sv interaction as long as the MSC sever is enhanced for SRVCC. Such configuration can reduce the complexity of network topology and help to reduce network signalling between MME and MSC server for involved combined procedure. [16]

Next step could be to decide the method based on what the MSC selection is done by MME. There are two options: using DNS (Domain Name System) query for choosing MME during SRVCC handover or do the selection based on manual configuration in MME. The *SvMscDnsQueryEnabled* parameter specifies whether the MME discovers and selects the MSC by DNS query for the SRVCC handover from LTE to WCDMA.

The operator has set the parameter value to “false” which means the MME selects an MSC supporting Sv (not all MSCs necessarily must support Sv), that is configured in the MME for the SRVCC handover. The MSC servers must be created manually in MME configuration. For each MSC the following parameters must be configured: Ordinal Number, capacity, SGs and Sv support, Sv IP address, MCC, MNC. The “Ordinal Number” parameter defines in which order the value ranges are assigned to the MSC/VLRs in the MSC pool. The “Capacity” parameter specifies the relative capacity of an MSC/VLR, that is, the number of UEs the MSC is capable of handling in relation to the other MSCs in the MSC pool. Each value range corresponds to a set of IMSI numbers At least two MSCs enhanced for SRVCC should be configured to achieve redundancy in Sv. Operator has configured two MSCs for Sv and enabled Sv support for both MSCs. The Location Areas has to be created and an MSC connected to the LAs. If an LA is connected to more than one MSC over the Sv interface and the MSC pool enabling feature “MSC pool Sv” is not enabled, a default MSC for the LA must be configured. If the “MSC pool for Sv” feature is enabled, then default MSC configuration per LA is not mandatory. Unlike CSFB which requires TA and LA mapping, there is no demand for planning the mapping relationship between TA and LA for SRVCC function.

Relevant parameter to be configured is *VoiceBearerInHandoverCommand* so that MME and eNB configurations would be aligned. The parameter specifies if the voice bearer is sent in the *Handover Command* message for SRVCC with or without PS Handover procedures. If the voice bearer is sent in the *Handover Command* message, it is included in the *E-RAB to be released list* IE. Based on that setting eNB knows to release the voice bearer during handover procedure.

There are several timers in GTPv2 level that can be adjusted to achieve the optimized signalling result. Just as an example of one of such timers is *T3ResponseSrvccPsToCs* which specifies the maximum waiting time for a response to an SRVCC PS to CS Request message. These are messages number five and ten in Figure 19. Value range of parameter is one to fourteen seconds. The operator has set the value to seven seconds. If timer expires the Request message is repeated the nr of attempts specified with *N3RequestSrvccPsToCs* timer. This value is configured to two.

As a last step of SRVCC configuration specific features for SRVCC and PS HO must be turned on. If MSC pool concept is used for Sv, that feature must be turned on as well.

8.3 Voice over LTE

Voice over LTE is similar to any VoIP (voice over IP) service but it is based on the IMS network, with specific profiles for control plane and media planes of voice service defined by GSMA in IR.92. In VoLTE case voice service (control and media plane) is delivered as data flows over LTE data bearers, also called all-IP voice. Hence the VoLTE is not dependent on traditional circuit switched telecommunications networks. VoLTE system architecture can be divided into three main domains: access, EPC core and control. Access domain contain only eNB, EPC is common Evolved Packet Core with all mandatory elements and control domain consists of three main parts: IMS Core, PCRF and HSS. In addition, terminal must be VoLTE capable with support in its firmware. [13]

Different domains must support different elements to achieve end-to-end call with good quality. For instance, the UE must support the Robust Header Compression (RoHC) to compress the large size VoLTE packets, being able to handle GBR (Guaranteed Bit Rate) bearers and indicate proper Voice Domain Preference to the core network, support relevant audio codecs etc. The EPC must guarantee the configuration with proper IMS APN, is responsible for P-CSCF discovery and is doing all the data bearer management related activities, including QoS relate bearer handling. The IMS must have support for IPSec (IP Security) protocol to protect the signalling, must support SIP URI (Session Initiation Protocol Uniform Resource Identifier) and Tel-URI schemes and have support for high quality audio codecs. The SIP URI and Tel-URI are addressing schemes to bridge the gap between VoIP and PSTN.

In EPS IP connectivity between UE and external PDN (Packet Data Network) is established with the help of EPS bearer. An EPS bearer is established when the UE connects to a PDN and the bearer remains established throughout the lifetime of the connection to provide the UE always-on IP connectivity. The initial EPS bearer is referred to as default bearer. An additional EPS bearer that is established for the same IP connection is referred to as dedicated bearer. For IMS-based services like VoLTE, there must be always a connection with default bearer established for the IMS APN before IMS registration. During VoLTE call setup towards IMS, a dedicated EPS bearer is created for voice media transfer. The dedicated EPS bearer for voice is temporary one as it lasts only during the voice media session which is different from the default EPS bearer. The default EPS bearer is persistent until the UE is detached from the LTE. Default bearer establishing is always initiated by UE while dedicated bearer setup is initiated by the network. The creation of dedicated EPS bearer is triggered by the P-CSCF sending service data information to the PCRF. Once the dedicated EPS bearer is created for voice, there comes two EPS bearers created between the UE and the IMS APN, e.g. the default EPS bearer for IMS related signalling and a dedicated EPS bearer for voice media. A dedicated bearer is for QoS differentiation purposes.

An EPS bearer is always characterized by the QoS parameters bound to the bearer. Different sets of QoS values are defined by parameter named QCI. The QCI values 1-4 are reserved for resource type Guaranteed Bit Rate (GBR) and QCI 5-9 for non-GBR. Default bearers are always non-GBR bearers while dedicated bearers can be GBR and non-GBR bearers. For GBR bearers dedicated network resources associated with the EPS bearer are permanently allocated at bearer establishment. During VoLTE call setup dedicated bearer with QCI 1 is used for real time conversational voice. IMS related signalling use QCI 5. The illustration of established EPS bearers in EPC during VoLTE call can be seen in Figure 21. Dedicated bearer with QCI 1 is used for voice media over RTP (Real-time Transport Protocol) protocol, default bearer with QCI 5 is used for IMS signalling over SIP protocol and default bearer with QCI 9 for first default bearer for Internet APN.

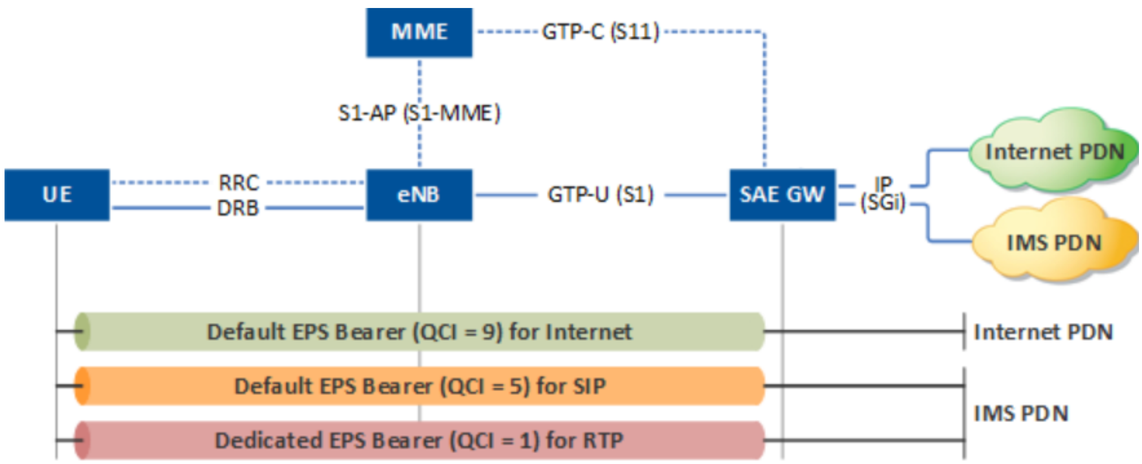


Figure 21. EPS bearers in VoLTE call. [7]

The steps to set up end-to-end VoLTE call are following. UE to detect available LTE network, attach to LTE network and check by UE that network is voice capable, set up IP connection with IMS APN and find P-CSFC, register and authenticate in IMS and finally place a call itself.

It is stated in GSMA profiles IR.88 (LTE Roaming Guidelines) and IR.92 (IMS Profile for Voice and SMS) that IMS application must use the IMS well-known APN. The APN name must be “IMS”, which is also APN Network Identifier part of the APN. Any other application must not use that APN. [11]

VoLTE signalling call flow is long and complex. The extensive call flow is presented in Appendix 1. In total three EPS bearers must be created to set up the VoLTE call. The bearers can be seen also in Figure 21. On high level the main steps are following:

1. Authenticate UE by MME and negotiate voice preferences between UE and EPC
2. UE initiates establishing first Default EPS bearer for general data connectivity, with “internet” APN and QCI 9
3. UE initiates establishing additional Default EPS bearer for IMS signalling purpose over SIP protocol, with IMS APN and QCI 5
4. UE registers to IMS network via SIP protocol
5. Network triggers Dedicated bearer establishment for actual VoLTE call

UE attach to EPS network and EPS bearer set up has been described in previous chapters and hence it is left out from the VoLTE call signalling flow presented in Figure 22. The figure shows the VoLTE call setup originating side and is focusing to IMS signalling part and interaction between different systems. Service Request procedure, radio and initial EPS data bearer, SIP registration, are mandatory to be in place prior to sending SIP INVITE message to ensure data bearer is up and running. UE initiates SIP signalling towards IMS core and “SIP INVITE” is sent from UE all the way to the terminating side. “SIP 100 Trying” from IMS Core is basically the acknowledgment to INVITE message. “SIP 183 Session Progress” means that target party is ready for negotiation about resources. Instantly after “SIP 183 Session Progress” sent IMS contacts PCRF over Rx interface to reserve resources for the call. PCRF will check what type of media is requested and will create Dedicated Bearer setup towards PGW in EPC, with QCI 1.

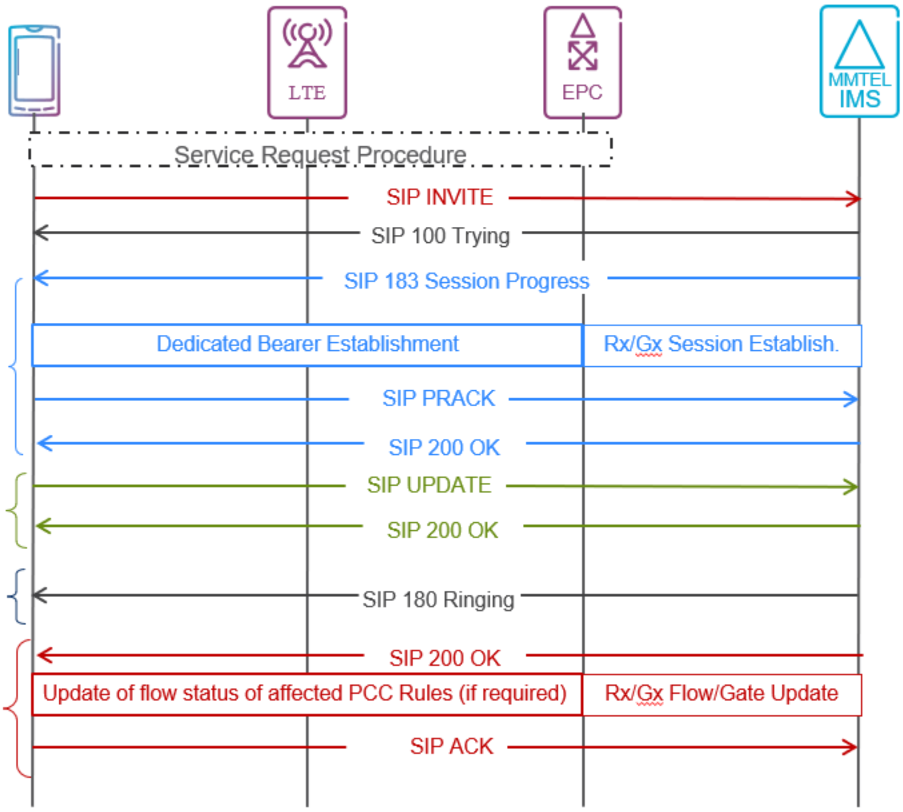


Figure 22. VoLTE call originating call flow. [7]

Terminal acknowledges “SIP 183” with “SIP PRACK” and terminating side responds with “SIP 200 OK”. After successful dedicated bearer setup in originating UE side “SIP UPDATE” will be sent to notify terminating side that originating party has been allocated resources and call can begin. Terminating party answers with “SIP 200 OK” and also sends “SIP 180 Ringing” indicating that the voice call setup request is being notified to the call recipient. After the recipient has answered the call the “SIP 200 OK” is sent to the UE. Upon receiving this message, the UE allocates the media resources and media can flow both directions. Last message sent is “SIP ACK” once call is established.

8.3.1 Mobile operator configuration

Current chapter describes how to configure the Multimedia Telephony Service (MMTel) for IMS voice service in EPC and also briefly the same for IMS Emergency service. In EPC the VoLTE related configuration must be done in MME and PGW. PCRF and Gx, Rx related configuration is out of scope of the thesis as it is broad and complex separate topic. Rx is an interface between PCRF and Application Function. There is a gray zone for which aspects can be assumed already present in an existing network before VoLTE and which aspects should be considered VoLTE specific configurations. The intention is to cover only explicitly specific VoLTE configuration..

In MME the „mmtel“ feature must be activated as a first step. The feature allows Packet Switched access for IMS-based voice, IMS Emergency and other multimedia services that are anchored in the IMS. The feature supports IMS Emergency Service for authenticated UE, unauthenticated UE with SIM cards, and unauthenticated UE without SIM cards. This feature is also enhanced by Service Priority Based Paging, which is used to prioritize the paging for specific services, such as MT VoLTE calls. This enhancement reduces the time needed for VoLTE call setup as makes paging procedure faster.

The “mmtel” feature requires also activation of the “Network-Initiated Dedicated Bearers” feature which enables establishment of multiple, dedicated bearers, in addition to the default bearers. Next step would be to enable IMS voice service for configured PLMN. It is possible to enable IMS voice with or without emergency service, to allow emergency services only for authenticated users, only for UEs with SIM card (both authenticated and unauthenticated) and also allow IMS voice and emergency service for all UEs, with and without SIM card. The mobile operator has decided to allow emergency services and to allow it for all users, configuring the relevant parameter “VoIPServices” to *“voip with emergency all”*. As emergency service has been enabled there is also a need to configure emergency number list and emergency profile name under PLMN. The emergency number list contains emergency numbers used by the serving network. Once an emergency number list has been created, it can be populated with emergency numbers, adding ISDN number of emergency service. Operator has defined “112” for all emergency categories like police, ambulance etc. As in Estonia only one common emergency number is used. The emergency profile is a configuration profile containing data used during IMS Emergency Service. The emergency profile applies to the entire PLMN. It contains parameters like emergency APN name, QoS parameters (including QCI), optionally PGW identities, configuration to allow or prevent the MME from sending an emergency PDN connection to a UTRAN and many more. For emergency APN it is recommended to configure a unique APN name and only Network Identity part; in current operator case it is configures as “sos”. QCI is set to five, PGW identities are not configured thus EPC is relying on DNS server for GW selection, sending emergency PDN to UTRAN is not allowed.

Important node function value to consider is “allow CSFB during voice call”. This parameter is used to control if *SGsAP-PAGING-REQUEST* messages for MT CS services and *Extended Service Request* messages for MO CS services are accepted by the MME during an ongoing IMS voice call. Even though such events would be exceptional, it is recommended to allow CSFB. Otherwise, for example for ongoing emergency VoLTE call the CSFB would always be rejected.

There are many parameters to configure VoLTE behavior not only on node level but also with specific IMSI range or Geographical Area granularity. If LTE coverage in some areas is poor, causing extra signalling because SRVCC procedure is often triggered, it is possible to define GA where VoLTE calls are disabled. Another relevant parameter to be set with IMSI range granularity is whether VoLTE roaming for inbound roaming subscribers is allowed or not. If VoLTE roaming is allowed it has to be allowed for each IMSI series separately. The function is called Home-Routed VoLTE. For mobile operator currently there is approximately ten partners for which VoLTE roaming is allowed. There is also a parameter “SrvccPossibleS8” which specifies if SRVCC is possible for visiting subscribers with home-bound IMS voice connection. When this parameter is not defined, then SRVCC is supported. To disable the parameter, it makes possible avoiding SRVCC attempts that would eventually fail. For example, the subscribers visiting a network where the MSC is not connected to the home IMS system of the subscriber. Although disabling the parameter would lead to improved KPIs (Key Performance Indicatior) and reduced signalling, it is not configured by the operator. It is also possible to restrict IMS emergency services based on IMSI range, for example to disable IMS emergency service for visited subscribers, regardless if IMS emergency services are allowed on PLMN level for home users. The operator has chosen the least restrictive value for visitors, set “IMS Emergency Service Supported “parameter to value which allows IMS emergency for all UEs with SIM card.

There are two node functions that needs to be considered in MME VoLTE configuration: “vops based on IMS APN” and “vops based on UE SRVCC capability”. The first parameter controls whether the MME checks the subscribed APNs in the subscription data of UE when setting the IMS VoPS bit in the *EPS network feature support* IE during Attach. The mobile operator has set value to “off” which means the MME does not check the IMS APN configuration profile in the subscription data. Hence, whether the UE has subscribed IMS APN configuration profile or not has no impact on the setting of the IMS

VoPS bit. If the parameter was set to “on” the MME checks all the subscribed APNs. If there is not subscriber APN “ims” (case insensitive parameter) the MME sets the IMS VoPS bit to 0, indicating IMS voice over PS session in S1 mode is not supported. [7]

The “vops based on UE SRVCC capability” parameter controls whether the MME checks the SRVCC capability of UE when setting the IMS VoPS bit in the *EPS network feature support* IE during Attach. When the parameter is set to “off”, the MME does not check the SRVCC capability of the UE. Hence, whether the UE has SRVCC capability or not has no impact on the setting of the IMS VoPS bit. When the parameter is set to “on”, but the UE has no SRVCC capability, the MME sets the IMS VoPS bit to 0, indicating IMS voice over PS session in S1 mode not supported. This configuration would ensure that UE which is not SRVCC capable is not able to register to IMS and set up VoLTE call which would be unwanted result.

In MME “TimerBasedGbrBearerDeactivation” (tgbd) parameter is highly recommended to configure to significantly improve VoLTE call retainability. If the mentioned parameter is set to “true” the deactivation of GBR bearers is delayed when the MME receives a *UE Context Release Request* message with an S1 release reason that is defined by another MME parameter “S1RelCauseDelayGbrBearerDeact”. The “S1TimerGbrBearerDeactivation” parameter specifies the maximum time that the MME preserves the GBR bearers after receiving a *UE Context Release Request* message with an S1 release reason that is defined by the “S1RelCauseDelayGbrBearerDeact” parameter. The mobile operator has set the “tgbd” value to “true” and the timer value for “S1RelCauseDelayGbrBearerDeact” is configured to four seconds. That means the Dedicated Bearer set up for VoLTE call is preserved for four seconds before bearer deletion is started in MME. This behaviour saves the signalling time because the dedicated bearer is retained and if the radio conditions should improve during four seconds, there is no need to create the new bearer.

For PGW configuration Gx interface towards PCRF is configured and is a prerequisite for following configuration. The IMS APN is configured with dedicated bearers enabled since this is crucial for VoLTE functionality. Traffic on the IMS APN default bearer is restricted to allow only IMS signalling towards the available P-CSCF. Traffic filtering is enabled for the APN to ensure that the APN is not used for any other traffic exept traffic related to VoLTE services. UE can request IP address type (PDP type) for both, IP version

4 (IPv4) and IP version 6 (IPv6), which is called DualStack, but it in PGW it is recommended to configure PDP type as single stack type and preferrably for IPv6. In this case PGW allocates IP address to UE as IPv6 only regardless what UE is asking. In addition to IMS APN also the APN for emergency service must be configured. From the MME config part we know it is „sos“ and APN is configured as logical APN. The logical APN is bound to the real APN and is using the same APN settings as real APN.

To disallow any type of traffic not related to IMS the rulespace is configured. Rulespace is associated with Service Set which is configured to use VoLTE service data flow. The fragment of the config example of VoLTE APN in PGW is shown in Figure 23.

```
Apn=ims
   allowRuleSpace="rs-ims"
   routingInstance="GmMb"
   AccessRestrictions=1
      selectionMode=public
      up
   Gtp=1
      disableWlanDedicatedBearer
      wlan
      up
   PCscf=1
      primaryPCscfPool="ims_default_primary_pcscf_pool"
      up
   PdpContext=1
      addressAllocation=shared-ip-pool
      creation=unblocked
      enableDedicatedBearer
      ipv6AddressAllocation=shared-ip-pool
      pdpType=ipv4v6
      sharedIpPool="ip-pool-ims"
      Policing=1
         maximumBandwidthDownlink=4000000
         maximumBandwidthUplink=100000
         up
      SessionControl=1
         up
      Signaling=1
         ggsnDeletesPerSecond=250
         up
```

Figure 23. IMS APN configuration in PGW.

9 Optimum deployment solution

For mobile operators who started deploying EPS networks, radio and core part, and who have legacy 2G/3G network in place, implementing CSFB as a voice service is more or less mandatory. CSFB is seen as an interim step towards all-IMS voice calls but CSFB ensures the voice call capability during the LTE coverage building which is done usually step by step until the coverage of LTE is 100% of the PLMN. Theoretically the LTE network could be built up to the coverage 100% and then enabling VoLTE calls together with SRVCC and skipping CSFB phase but this approach is complex technically and commercially and in reality not used. Also operators usually started marketing 4G in their network as soon as possible to gain publicity and marketing benefits. There are also operators on the market who did not have legacy CS network and came to the market starting building LTE and EPC as entry level networks. In this case VoLTE calls were enabled since day one. It is still worth to note most of such operators provide data services only, like Fixed Wireless Access, for example and do not provide LTE mobility and voice services at all.

The Estonian mobile operator based on which the thesis is written is typical operator with legacy 2G/3G networks. It started deploying LTE step by step which means CSFB was planned and implemented as a first step to offer voice service for subscribers who were attached to EPS to consume data services. VoLTE services were implemented during years 2018-2019 for smartphones from Sony, Samsung and Huawei. The VoLTE is not yet activated to all the customer base, for example not to prepaid SIM cards, as well as not all the terminals do not support VoLTE service in operator network.

VoLTE solution has technically many advantages over the voice calls that do not use VoLTE technology like CSFB:

- VoLTE allows for simultaneous usage of high-speed packet services and voice calls. It is quite common that using demanding and high-quality packet services (e.g. videoconference) within LTE the subscriber receives terminated CSFB call and packet services are moved to 2G/3G network which are not capable of handling the data stream and initial packet service is halted. It can easily happen even with the most enhanced radio access type switching technology, like seamless PS Handover. Furthermore, often the coverage of 2G radio is better than 3G and subscriber data services are moved to 2G because of that. Swithcing the radio back to 4G after CSFB call has ended is also quite complex procedure in packet core and can take quite some time.

- VoLTE offers better voice quality than legacy networks. The key driver here is HD (High Definition) Voice which is a suite of services enabled by VoLTE. HD Voice provides clearer, more natural sounding audio, when both calling parties are using HD voice-enabled phones that are connected to LTE network. In addition to producing clearer, richer, and more natural sounding audio it also helps to eliminate background noise. Most of the new smartphones sold in 2020 do support HD voice.
- VoLTE offers better battery life for a terminal because the UE does not need to switch between different radio types during a voice call which is a procedure with high energy consumption.
- VoLTE offers more efficient radio spectrum usage as the newer and more efficient modulation technologies are used to save radio resources.
- VoLTE enables shorter call setup times compared to CSFB.
- In CSFB scenario, the call decision (which results in moving UE from PS to CS) has to be made before call initiation by core network. For VoLTE scenario (IMS registered UE) the moving from PS to CS can be done during the VoLTE call and call remains anchored in IMS after PS to CS transfer. That is VoLTE/SRVCC procedure advantage over CSFB.

During the writing of the thesis at April 2020 the author conducted a set of tests to measure call setup times for two different scenarios. First scenario is where both UEs – calling and called party – reside in LTE coverage and are IMS registered. For second scenario the IMS registration was turned off for both UEs and they needed to perform CSFB procedure. The test was done with two Sony identical mobile terminals and using Ericsson proprietary Android application in the mobile operator cell with ECGI (E-UTRAN Cell Global Identifier) 248-02-256320534. Measurement can vary significantly in different cells because most of the call setup delay comes from radio access not from core related signalling, For both scenarios one hundred call repetitions were done and average result calculated. It was assured before every call attempt that both UEs went to ECM-IDLE state in LTE network.

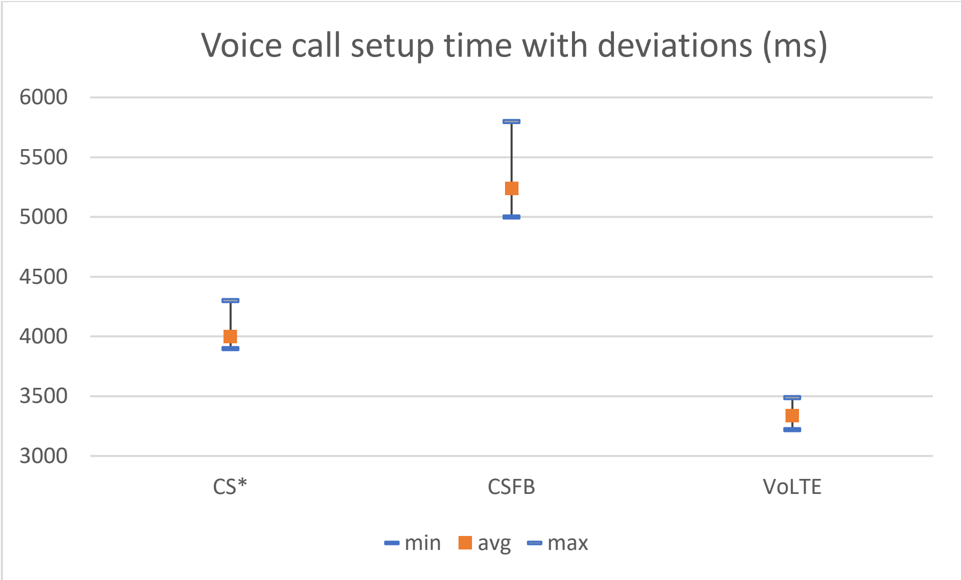


Figure 24. Voice call setup delay times. CS* is estimated value.

For the scenario with VoLTE to VoLTE call setup time was calculated with following formula for which the messages in the formula are described in Figure 22:

$$\mathit{Call\ Setup\ Time\ [ms]} = T_{180\ Ringing} - T_{INVITE} \tag{1}$$

The average call setup time was 3340 ms (3,34 seconds) and maximum deviation from the average 150 ms.

For the scenario with CSFB call to CSFB call the average call time was 5240 ms (5,24 seconds) and the largest deviation from the average 560 ms. The time interval was measured between the *Extended Service Request* message by the calling party UE and ringing tone sent by MSC in the called party. The measurements prove that voice call setup time in VoLTE case is significantly shorter than for calls without VoLTE and smaller deviation from average call setup time proves the stability and reliability of VoLTE calls.

According to the row 3 in Table 1 on page 61 can be considered that extra delay for CSFB call compared to ordinary CS call is 0.6 seconds in mobile originated and mobile terminated end. As a result of this consideration the CS call setup would be 4 seconds. The author had not an opportunity to test CS call in practice, but the result would fit well with theoretical results in Table 1. VoLTE call setup delay time would be 3.3 seconds, CS to CS call 4 seconds and CSFB 5.2 seconds. The Figure 24 illustrates the voice call setup times for different voice solutions. On the figure can be seen average setup time and minimum and maximum values. It is clearly seen that voice call setup time is shortest for VoLTE as well as deviations from the average are the smallest which makes the VoLTE a preferred solution.

The main strengths of CSFB solution include:

- No need to rely on deployment of IMS infrastructure and services before offering voice as a service to LTE users
- Same feature and service set offered for voice services when in LTE access as when in a system supporting circuit switched voice calls. The circuit switched core network infrastructure can be utilized for LTE users.

It should be noted that both solutions can be simultaneously supported in the same network and it can be assumed operators initially deploying CSFB will over time migrate towards VoLTE/IMS solutions. [2]

10 Summary

Current thesis contains an overview of theoretical aspects of standardised packet switching based core network EPC, its architecture, interfaces and protocols. The main focus is on voice services and different deployment options in the IP protocol-based telecommunications core networks. Traditionally voice calls have been served in 2G and 3G networks by Circuit Switched core network. After introducing EPC, which is packet-only network, mobile operators have been forced to introduce the solution for providing voice services in Packet Switched networks which is technically not very easy task.

First half of the thesis gave an in-depth overview of technical details of EPC and the different voice service deployment options which are Circuit Switched Fallback, Single Radio Voice Call Continuity and Voice over LTE. Effort has been put to go into details in the call establishment signalling flow for all three different deployment options. Understanding the call setup signalling allows to compare the complexity of different options and to draw conclusions on the strengths and weaknesses of each option.

Second half of the thesis described the actual configuration needed to enable voice services in EPC. The configuration examples are given based on mobile equipment vendor Ericsson products and Estonian mobile operator who has introduced all three mentioned options step by step in their network. The author of the thesis has done much of the configuration that is described in the last chapter of the thesis or has done consulting for the operator for configuration details. The author has worked closely with the operator to achieve the optimum configuration of live core network.

While CSFB is implemented by the mobile operators to guarantee voice calls in spotty LTE coverage, it is still seen as an interim step towards VoLTE. Voice over LTE is the most advanced option for voice service in EPC and has several advantages over legacy voice solutions. It enables much clearer and more natural voice quality than provided in legacy CS networks. Moreover, the call setup time is significantly shorter for VoLTE than for CSFB. During the writing of the thesis the author made a series of tests to measure voice call setup time for different options. The results prove that call setup time using VoLTE call solution is noticeably shorter than using the call forwarding to legacy CS network.

Last part of the thesis looks into the currently existing configuration in live production core network of the Estonian operator and also gives a few recommendations for changes to achieve even more optimum configuration. The findings have been shared with the operator. As vendors launch the new software frequently (monthly per Ericsson) there are constant improvements and enhancements in the software which need to be considered to run and implement the newest possibilities of the software.

The main goal of the thesis was to find the best voice service option and within that option to find the optimum configuration. It became clear that technically VoLTE call is undoubtedly the most advanced and best solution. The biggest obstacle for launching VoLTE is the complexity of IMS subsystem which the VoLTE relies on. It requires lot of effort from solution architects and integration engineers with high competence and long experience to deploy VoLTE. The project takes usually much time (1-1.5 years, estimated) and effort and is also commercially expensive. The abovementioned operator launched VoLTE commercially in June 2019, and at the time this thesis was written service has not been enabled to all of the subscribers because of technical considerations.

References

[1] 3GPP, “The Evolved Packet Core”, [Web article]. Available: https://www.3gpp.org/technologies/keywords-acronyms/100-the-evolved-packet-core

[2] Magnus Olsson, Shabnam Sultana, Stefan Rommer, Lars Frid, Catherine Mulligan, “EPC and 4G Packet Networks: Driving the Mobile Broadband Revolution”, Second Edition, USA, Academic Press, 2013

[3] Magnus Olsson, Shabnam Sultana, Stefan Rommer, Lars Frid, Catherine Mulligan, “SAE and the Evolved Packet Core”, First Edition, UK, Academic Press, 2009

[4] 3GPP, “Releases”, [Web article]. Available: https://www.3gpp.org/specifications/67-releases

[5] Wikipedia, “3GPP”, [Web article]. Available: https://en.wikipedia.org/wiki/3GPP

[6] IP-Solutions AB, “EPC Protocols and Signalling”, [Training material]

[7] Ericsson, “EPC architecture”, [Internal company material] http://cpistore.internal.ericsson.com/alex

[8] GSMA, “Volte Service description and implementation”, [Web document]. Available: https://www.gsma.com/futurenetworks/wp-content/uploads/2014/05/FCM.01-v1.1.pdf

[9] LTE blog, “Long Term Evolution”, [Web material]. Available: http://ltespace.blogspot.com/p/mme-pooling.html

[10] IETF, “Diameter Base Protocol”, [Web Document]. Available: https://tools.ietf.org/html/rfc6733

[11] GSMA, “IMS Profile for Voice and SMS”, 2019, [Web Document]. Available: https://www.gsma.com/newsroom/wp-content/uploads//IR.92-v13.0-1.pdf

[12] 3GPP, “TS 24.173 IMS multimedia telephony communication service and supplementary services”, [Web Document]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=1049

[13] Miikka Poikselkä, Harri Holma, Jukka Hongisto, Juha Kallio, Antti Toskala, “Voice over LTE”, Nokia Siemens Networks, Finland, John Wiley & Sons Ltd, 2012

[14] Researchgate.net, “Overview-of-the-3GPP-IMS-architecture”, [Web material]. Available: https://www.researchgate.net/figure/1-Overview-of-the-3GPP-IMS-architecture_fig19_266491904

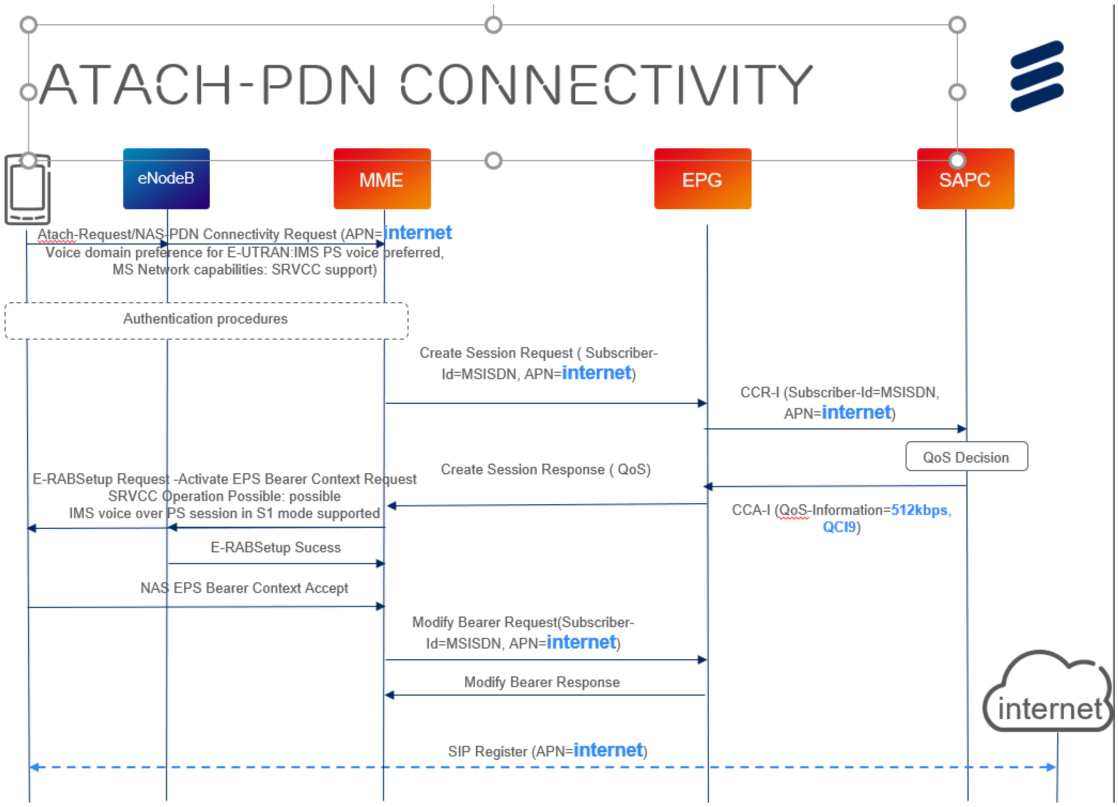
[15] Ericsson, “CSFB and SMS over SGS solution guideline”, [Internal company material], Available: http://gask2web.ericsson.se/service/get?DocNo=22112FGD101073&Lang=EN&Rev=B&Format=PDFV1R2

[16] Ericsson, “Single Radio Voice Call Continuity SRVCC Method of Procedure”, [Internal company material],

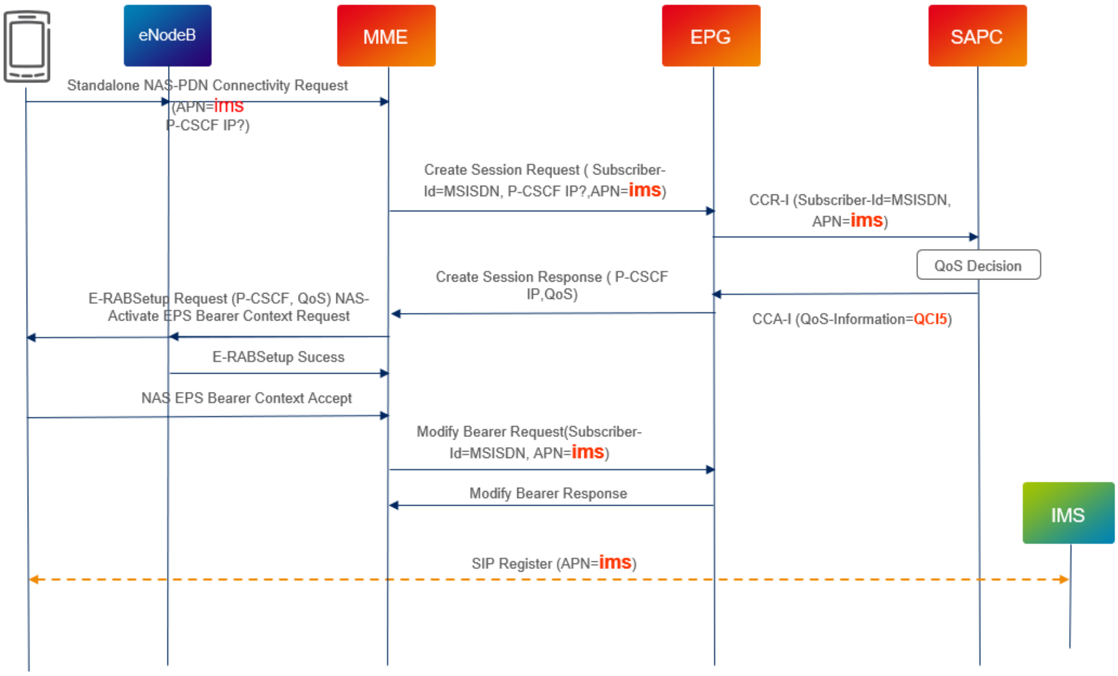
Available:
http://gask2web.ericsson.se/pub/get?DocNo=122/1020...02483&Lang=EN&Rev=A&Format=MSWORD2003

[17] Ericsson, “Voice over LTE Method of Procedure”, [Internal company material], Available:
http://gask2web.ericsson.se/pub/get?DocNo=1/1020FGD101073&HighestFree=N&HighestNotPRev=N

Appendix 1 – VoLTE call establishment



2ND PDN CONNECTIVITY



DEDICATED BEARER