3GPP TR 38.801 V14.0.0 (2017-03)

Technical Report

3rd Generation Partnership Project;
Technical Specification Group Radio Access Network;
Study on new radio access technology:
Radio access architecture and interfaces
(Release 14)

Keywords
<keyword[, keyword]>

3GPP

Postal address

3GPP support office address
650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet
http://www.3gpp.org

Contents

Foreword

This Technical Report has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

x the first digit:

1 presented to TSG for information;

2 presented to TSG for approval;

3 or greater indicates TSG approved document under change control.

y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.

z the third digit is incremented when editorial only changes have been incorporated in the document.

Introduction

Work has started in ITU and 3GPP to develop requirements and specifications for new radio (NR) systems, as in the Recommendation ITU-R M.2083 "Framework and overall objectives of the future development of IMT for 2020 and beyond", as well as 3GPP SA1 study item New Services and Markets Technology Enablers (SMARTER) and SA2 study item Architecture for NR System. 3GPP has to identify and develop the technology components needed for successfully standardizing the NR system timely satisfying both the urgent market needs, and the more long-term requirements set forth by the ITU-R IMT-2020 process. In order to achieve this, evolutions of the radio interface as well as radio network architecture are considered in the study item "New Radio Access Technology" [1].

1 Scope

The present document covers the Radio Access Architecture and Interface aspects of the study item “New Radio Access Technology” [1]. The purpose of this TR is to record the discussion and agreements that arise in the specification of the “New Radio Access Technology” from an Access Architecture and Interface specification point of view.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP RP-160671: "New SID Proposal: Study on New Radio Access Technology".

[2] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[3] 3GPP TS 36.401: "Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Architecture description".

[4] 3GPP TR 22.862: "FS_SMARTER - Critical Communications".

[5] 3GPP TR 38.913: "Study on Scenarios and Requirements for Next Generation Access Technologies".

[6] 3GPP TR 23.799: "Study on Architecture for Next Generation System".

[7] RP-161266, "5G Architecture Options- Full Set". Joint RAN/SA Meeting, Busan, S. Korea, June 2016.

[8] R3-161646, "Migration path towards NR with EPC", NTT DOCOMO, INC.

[9] R3-161772, "AT&T Views on 5G Architecture Evolution: Early Phase 1 To Mature Phase 2 Deployment", AT&T.

[10] R3-161809, "Analysis of migration paths towards RAN for new RAT", CMCC.

[11] R3-161813, "Transport requirement for CU&DU functional splits options", CMCC.

[12] 3GPP TS 36.300: "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2".

[13] 3GPP TS 36.423: "Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 application protocol (X2AP)".

[14] R3-162102, " CU-DU split: Refinement for Annex A (Transport network and RAN internal functional split)", NTT DOCOMO, INC.

[15] R3-162174, "TIM view on 5G Migration Path", Telecom Italia.

[16] R3-170939, "Views on 5G Migration Paths", China Telecom.

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in 3GPP TR 21.905 [2] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [2].

eLTE eNB: The eLTE eNB is the evolution of eNB that supports connectivity to EPC and NGC.

gNB: A node which supports the NR as well as connectivity to NGC.

NOTE : The definitions of gNB and eLTE eNB may require revision in the normative phase.

New RAN: A Radio Access Network which supports either NR or E-UTRA or both, interfacing with the NGC (see TR 23.799 [6]).

New Radio: A new radio access technology which is studied under [1].

Network Slice: A Network Slice is a network created by the operator customized to provide an optimized solution for a specific market scenario which demands specific requirements with end to end scope as described in TR 23.799 [6].

Network Function: A Network Function is a logical node within a network infrastructure that has well-defined external interfaces and well-defined functional behaviour.

NG-C: A control plane interface used on the NG2 reference points (defined in TR 23.799 [6]) between New RAN and NGC.

NG-U: A user plane interface used on the NG3 reference points (defined in TR 23.799 [6]) between New RAN and NGC.

Non-standalone NR: A deployment configuration where the gNB requires an LTE eNB as anchor for control plane connectivity to EPC, or an eLTE eNB as anchor for control plane connectivity to NGC.

Non-standalone E-UTRA: A deployment configuration where the eLTE eNB requires a gNB as anchor for control plane connectivity to NGC.

User Plane Gateway: Termination point of the NG-U interface.

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [2] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [2].

CCNF	Common Control Network Function
NF	Network Function
NGC	Next Generation Core
NG-C	NG Control Plane interface
NG-U	NG User Plane interface
NR	New Radio
UPGW	User Plane Gateway
TRP	Transmission Reception Point

4 Objectives and requirements

4.1 Support for high reliability low latency services

The NR shall be able to support the highly reliable (i.e. with low ratio of erroneous packets), and low latency services.

Support for high reliability services is subject to the following requirements and assumptions:

- High Reliability:

  High reliability is about providing high likelihood of delivering error free packets through the 3GPP system within a bounded latency. A performance metric for high reliability is the ratio of successfully delivered error free packets within a delay bound over the total number of packets. The required ratio and latency bound may be different for different URLLC use cases.

- High Availability:

  High availability is related to a communication path through the 3GPP system providing reliable services. This communication path between the communication end points is made up of radio links as well as transport links and different HW and SW functions. The NR should provide high availability, in addition to deploy redundant components and links for these radio, transport and HW/SW.

- Low Latency:

  According to TR 38.913, the NR should support latencies down to 0.5 ms UL/DL for URLLC.

Whether a new function is needed at the NR to support high reliability low latency services was not addressed during study phase.

Whether and how end to end delays will be considered was not addressed during study phase.

5 Deployment scenarios

5.1 General

The following example scenarios should be considered for support by the RAN architecture.

Although it is not always explicitly specified, it should be assumed that an inter BS interface may be supported between an gNB and other gNBs or (e)LTE eNBs.

The Heterogeneous deployment comprises in the same geographical area two or more deployments as defined in sections 5.2 to 5.5 of the present document.

5.2 Non-centralised deployment

In this scenario, the full protocol stack is supported at the gNB e.g. in a macro deployment or indoor hotspot environment (could be public or enterprise). The gNB can be connected to “any” transport. It is assumed that the gNB is able to connect to other gNBs or (e)LTE eNBs via a RAN interface.

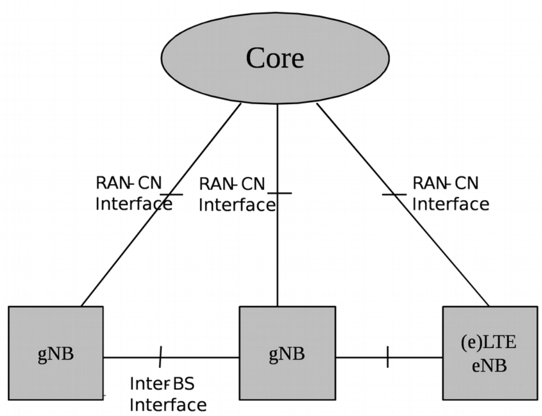


Figure 5.2-1: Non-centralised deployment

5.3 Co-sited deployment with E-UTRA

In this scenario the NR functionality is co-sited with E-UTRA functionality either as part of the same base station or as multiple base stations at the same site. Co-sited deployment can be applicable in all NR deployment scenarios e.g. Urban Macro. In this scenario it is desirable to fully utilise all spectrum resources assigned to both RATs by means of load balancing or connectivity via multiple RATs (e.g. utilising lower frequencies as coverage layer for users on cell edge).

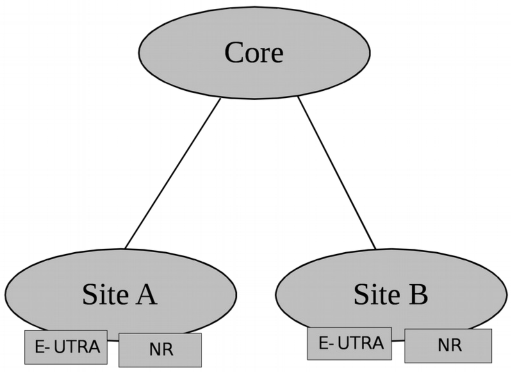


Figure 5.3-1: Co-sited deployment with E-UTRA

5.4 Centralized deployment

NR should support centralization of the upper layers of the NR radio stacks.

Different protocol split options between Central Unit and lower layers of gNB nodes may be possible. The functional split between the Central Unit and lower layers of gNB nodes may depend on the transport layer.

High performance transport between the Central Unit and lower layers of gNB nodes, e.g. optical networks, can enable advanced CoMP schemes and scheduling optimization, which could be useful in high capacity scenarios, or scenarios where cross cell coordination is beneficial.

Low performance transport between the Central Unit and lower layers of gNB nodes can enable the higher protocol layers of the NR radio stacks to be supported in the Central Unit, since the higher protocol layers have lower performance requirements on the transport layer in terms of bandwidth, delay, synchronization and jitter.

Both non co-sited deployment and co-sited deployment with E-UTRA can be considered for this scenario.

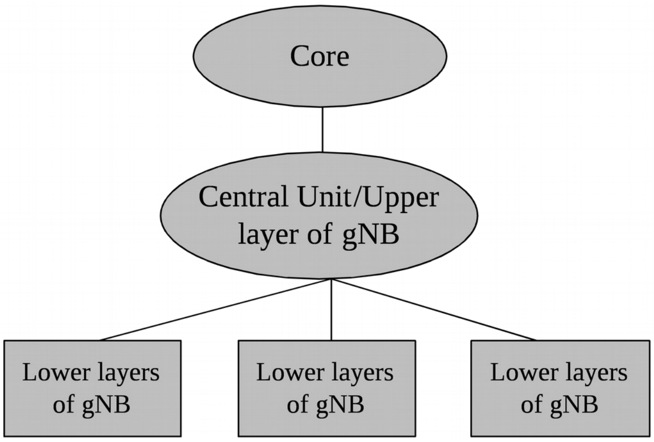


Figure 5.4-1: Centralized deployment

5.5 Shared RAN deployment

NR should support shared RAN deployments, supporting multiple hosted Core Operators. The Shared RAN could cover large geographical areas, as in the case of national or regional network sharing. The Shared RAN coverage could also be heterogeneous, i.e. limited to few or many smaller areas, for example in the case of Shared in-building RANs. A shared RAN should be able to efficiently interoperate with a non-shared RAN.

Each Core Operator may have their own non-shared RAN serving areas adjacent to the Shared RAN. Mobility between the non-shared RAN and the Shared RAN shall be supported in a way at least as good as for LTE.

The Shared RAN may (as for the case of LTE) operate either on shared spectrum or on the spectrum of each hosted Operator.

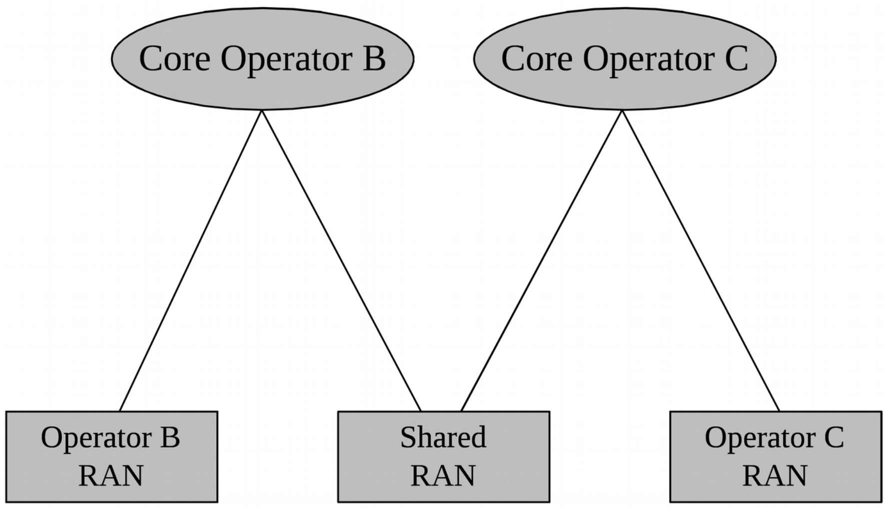


Figure 5.5-1: Shared RAN deployment

6 Overall New RAN architecture

6.1 RAN-CN functional split

The general aspects of RAN-CN functional split are captured in TR 23.799 [6]. The topics still open in TR 23.799 [6] for RAN-CN functional split need to be addressed during normative phase.

6.2 RAN functions description

NOTE 1: The further detail of the functionalities will be decided in normative work.

Functions similar to E-UTRAN as listed in TS 36.401 [3]

- Transfer of user data
- Radio channel ciphering and deciphering
- Integrity protection
- Header compression
- Mobility control functions:
    - Handover

- Inter-cell interference coordination
- Connection setup and release
- Load balancing
- Distribution function for NAS messages
- NAS node selection function
- Synchronization
- Radio access network sharing
- Paging
- Positioning

Functions specific for New RAN:

- Network Slice support
  - This function provides the capability for New RAN to support network slicing
- Tight Interworking with E-UTRA
  - This function enables tight interworking between NR and E-UTRA by means of data flow aggregation. This function includes at least dual connectivity. Interworking with E-UTRA is supported for collocated and non-collocated site deployments.
- E-UTRA-NR handover through a New RAN interface
  - This function provides means for E-UTRA-NR handover via the direct interface between an eLTE eNB and a gNB.
- E-UTRA - NR handover via CN (both intra-system and inter-system, i.e. EPC-NGCN, handovers)
  - This function provides means for E-UTRA - NR handover via CN.

NOTE 2: When discussing solutions to support this function, RAN3 needs to consider factors such as adaptation of the source RAN/CN to the target RAN/CN.

- Session Management
  - This function provides means for the NGC to create/modify/release a context and related resources in the New RAN associated with a particular PDU session of a UE and the corresponding tunnel between the New RAN node and the UPGW.
- Contacting UEs in inactive mode
  - In the inactive mode, the UE context is stored in RAN and UP data is buffered in RAN.

NOTE 3: Work on detailed description of the function will be done in the work item phase.

Functions specific for New RAN which are postponed:

- Direct services support (further study related with D2D, coordinate with RAN1, RAN2)
  - This function provides communication whereby UEs can communicate with each other directly
- Interworking with non-3GPP systems
- This function provides interworking between NR and Non-3GPP RAT (e.g. WLAN).

7 RAN Architecture and Interfaces

7.1 New RAN Architecture

The New RAN consists of the following logical nodes:

- gNBs providing the NR U-plane and C-plane protocol terminations towards the UE; and/or
- eLTE eNBs providing the E-UTRA U-plane and C-plane protocol terminations towards the UE.

NOTE : Whether to define a New RAN logical node that is able to provide both NR and E-UTRA U-plane and C-plane protocol terminations towards the UE will be determined in the normative phase.

The logical nodes in New RAN are interconnected with each other by means of the Xn interface.

The logical nodes in New RAN are connected to the NGC by means of the NG interface. The NG interface supports a many-to-many relation between NG-CP/UPGWs and the logical nodes in New RAN.

The New RAN architecture is illustrated in Figure 7.1-1.

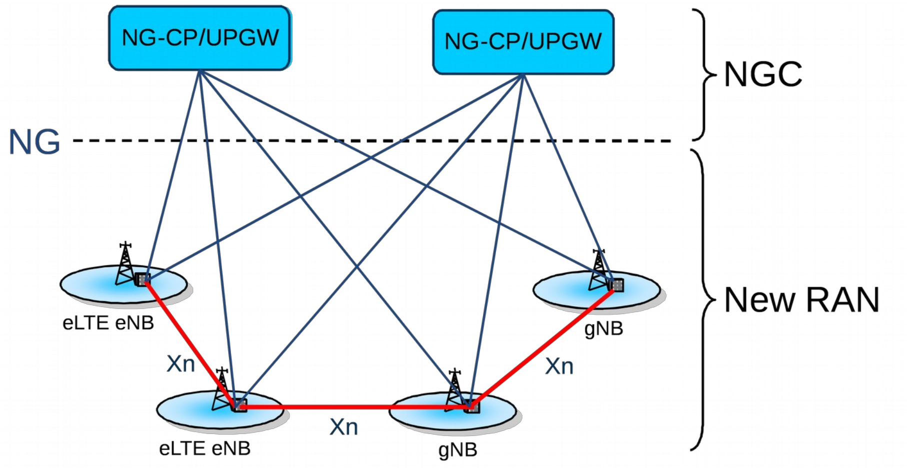


Figure 7.1-1: New RAN architecture

7.2 5G Architecture Options

The following options of [7] for providing NR access to suitably capable UEs should be considered in discussions on the RAN-CN interface, and the interface between E-UTRA and NR RAT.

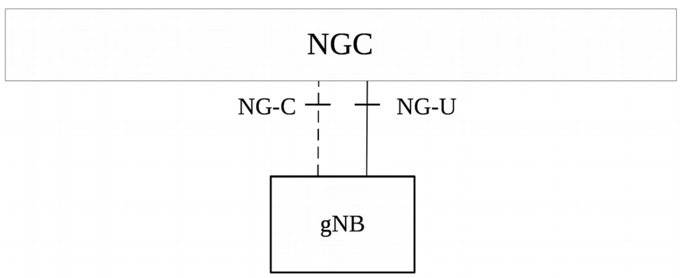


Figure 7.2-1: Option 2

In Option 2, the gNB is connected to the NGC.

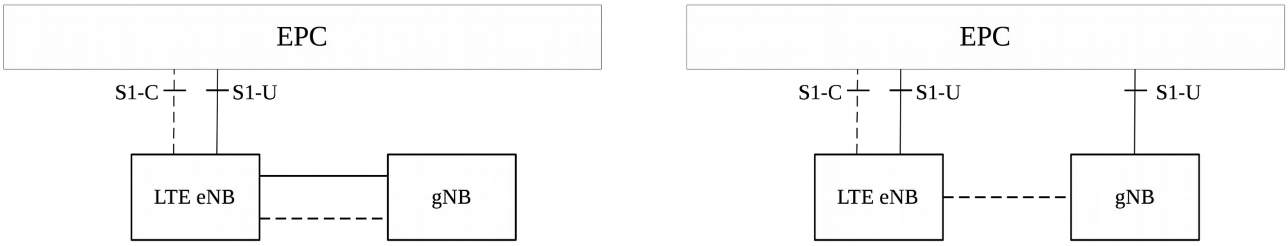


Figure 7.2-2: Options 3 and 3A

In Option 3/3A, the LTE eNB is connected to the EPC with Non-standalone NR. The NR user plane connection to the EPC goes via the LTE eNB (Option 3) or directly (Option 3A).

NOTE: Terminology related to Option 3/3A can be further discussed in normative phase, if needed.

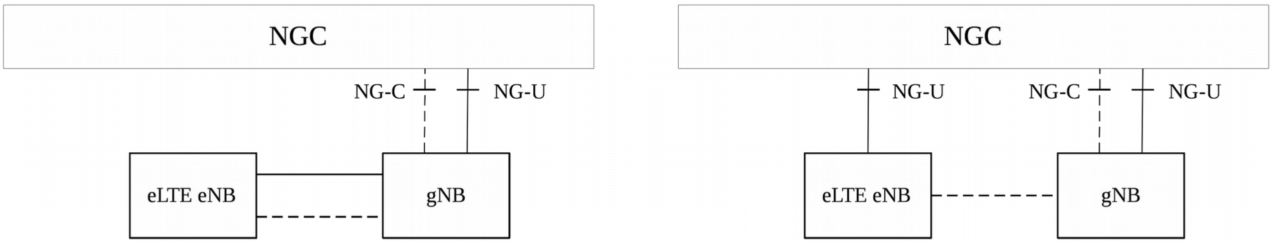


Figure 7.2-3: Options 4 and 4A

In Option 4/4A, the gNB is connected to the NGC with Non-standalone E-UTRA. The E-UTRA user plane connection to the NGC goes via the gNB (Option 4) or directly (Option 4A).

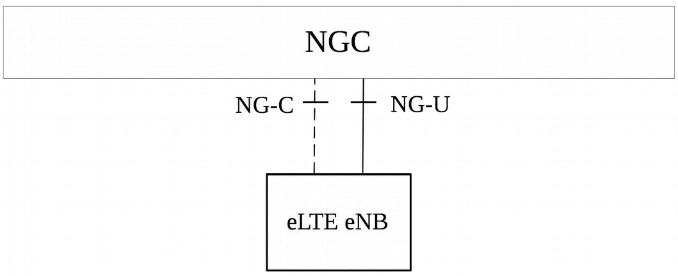


Figure 7.2-4: Option 5

In Option 5, the eLTE eNB is connected to the NGC.

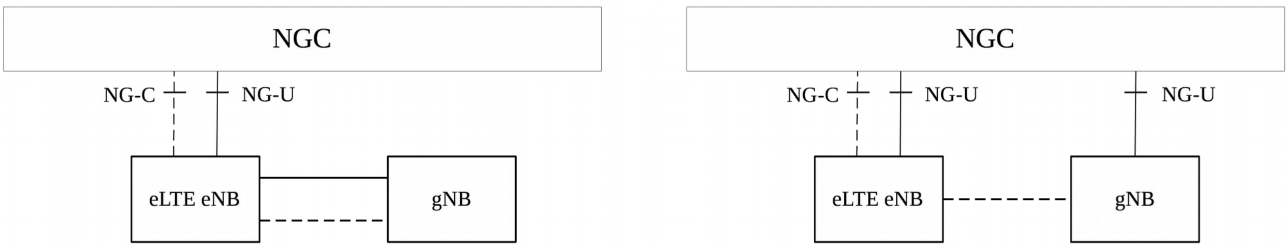


Figure 7.2-5: Options 7 and 7A

In Option 7/7A, the eLTE eNB is connected to the NGC with Non-standalone NR. The NR user plane connection to the NGC goes via the eLTE eNB (Option 7) or directly (Option 7A).

7.3 RAN-CN interface

7.3.1 RAN-CN interface connectivity scenarios

In order to support the options described in section 7.2, the following scenarios for connectivity between RAN consisting of E-UTRA and NR, and a CN consisting of an NGC and an EPC should be considered in the discussions on RAN-CN interface definition . The connectivity scenario in figure 7.3.1-1 includes support for Option 3/3A, while figure 7.3.1-2 includes support for Option 2, Option 4/4A, Option 5, and Option 7/7A.

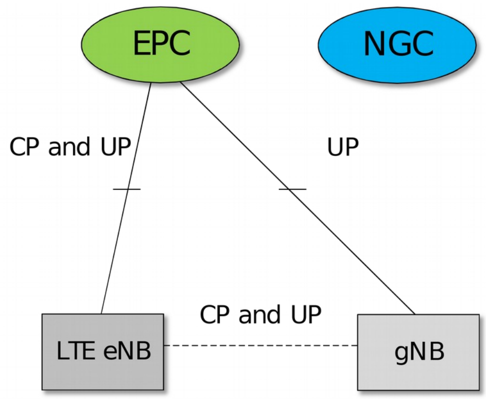


Figure 7.3.1-1: E-UTRA and NR connected to the EPC.

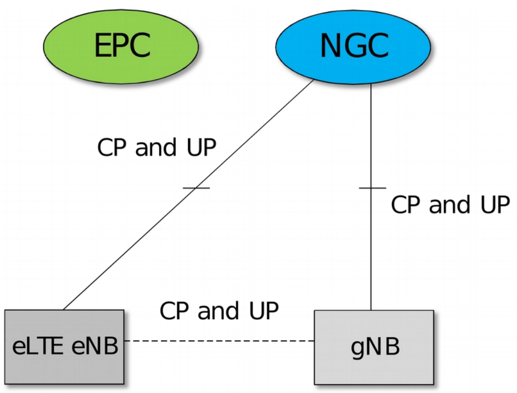


Figure 7.3.1-2: E-UTRA and NR connected to the NGC (Note: In this scenario, eLTE eNB and gNB can be collocated.).

7.3.2 General principles

The general principles for the specification of the NG interface are as follows:

- the NG interface shall be open;
- the NG interface shall support the exchange of signalling information between the New RAN and NGC;
- from a logical standpoint, the NG is a point-to-point interface between a New RAN node and an NGC node. A point-to-point logical interface shall be feasible even in the absence of a physical direct connection between the New RAN and NGC;
- the NG interface shall support control plane and user plane separation;
- the NG interface shall separate Radio Network Layer and Transport Network Layer;

- the NG interface shall be future proof to fulfil different new requirements and support of new services and new functions;
- the NG interface shall be decoupled with the possible New RAN deployment variants.
- the NG Application Protocol shall support modular procedures design and use a syntax allowing optimized encoding /decoding efficiency.

NOTE: Whether and how to document the application protocol for NG with regards to S1AP will be decided in the normative phase.

7.3.3 NG Interface Functions

NG-C interface supports following functions:

- Interface management: The functionality to manage the NG-C interface;
- UE context management: The functionality to manage the UE context between the New RAN and CN;

NOTE 1: The UE context information may include roaming and access restriction and security information.

NOTE 2: The UE context information may include the information related with network slicing.

- UE mobility management: The functionality to manage the UE mobility for connected mode between the New RAN and CN;
- Transport of NAS messages: procedures to transfer NAS messages between the CN and UE;
- Paging: The functionality to enable the CN to generate Paging messages sent to the New RAN and to allow the New RAN to page the UE in RRC_IDLE state;
- PDU Session Management: The functionality to establish, manage and remove PDU sessions and respective New RAN resources that are made of data flows carrying UP traffic.
- Configuration Transfer: the functionality to transfer the New RAN configuration information (e.g. transport layer addresses for establishment of Xn interface) between two New RAN nodes via the NGC.

NOTE 3: It can be discussed on whether congestion and overload control function is needed or not in normative phase.

7.3.4 NG Interface Procedures

7.3.4.1 General

To support the functions listed in Section 7.3.3 the NG interface should support the following procedures. The procedures are classified in different categories.

Interface management procedures:

- NG Setup: To establish an NG interface between the New RAN and NGC nodes
- Configuration Updates from New RAN and NGC nodes: To update configuration of the interface from the New RAN or NGC - NG Reset: To reset the NG interface
- Error Indication: To report detected errors in one incoming message

UE Context Management Procedures:

- Initial Context Setup: To establish the initial UE context both in New RAN and in NGC
- UE Context Release: To remove UE context information
- UE Context Modification: To modify an already established UE context

UE Mobility Management Procedures:

- Handover Preparation: Needed to prepare resource allocation from the source RAN to the NGC for a UE handing over
- Handover Resource Allocation: Needed to perform admission control and reserve resources at the target RAN for a UE handing over
- Path Switch Request: Needed to request NGC to switch a UP connection to a different new RAN node UP termination point

NOTE 1: the Path Switch over NG-C procedure is taken as baseline. Other Path Switch procedures may be considered during normative work, if justified.

- Handover Cancel: Needed to cancel an ongoing handover preparation or an already prepared handover
- Handover Notification: Needed to indicate to the NGC that handover has been successfully completed

Transport of NAS Messages Procedures:

- Initial UE Message: To transport the first NAS PDU when no UE signaling connection is established from the UE to the NGC

NOTE 2: It should be discussed in normative phase whether the merge of Initial UE Message and UL NAS Transport is possible or not.

- DL NAS Transport: To transport NAS PDUs in DL
- UL NAS Transport: To transport NAS PDUs in UL
- Non Delivery NAS Indication: To indicate that the New RAN node did not deliver a NAS PDU to the UE
- NAS Rerouting: In order to route a NAS request to a different CN node

NOTE 3: The need of this procedure will be assessed during normative phase.

Paging Procedures:

- Paging: To provide a paging instruction from the NGC to the New RAN

PDU Session Management Procedures:

- PDU Session Setup: Needed to establish a PDU Session and assign resources to it according to the QoS and other configuration parameters assigned to the flow(s) configured within the session
- PDU Session Modify: Needed to modify a previously configured PDU session, related QoS flows and respective New RAN resources
- PDU Session Release: Needed to release a previously configured PDU Session and related New RAN resources
- PDU Session Modification Indication: Needed to switch a data path from one termination point at the New RAN to a different termination point.

Configuration Transfer Procedures:

- RAN Configuration Transfer: to request and/or transfer New RAN configuration information via the NGC.
- NGC Configuration Transfer: to request and/or transfer New RAN configuration information to the New RAN.

In order to support IMS multimedia services, including emergency calls, according to SA2 requirements, it may be necessary to introduce additional procedures for UE radio capability handling.

7.3.4.2 NG Interface Procedures Descriptions

7.3.4.2.1 General

This section describes NG procedures. The NG interface can terminate in an eLTE eNB or in a gNB. Therefore, NG procedures apply to both an eLTE eNB and a gNB.

7.3.4.2.2 Interface Management Procedures

NG Setup Procedure

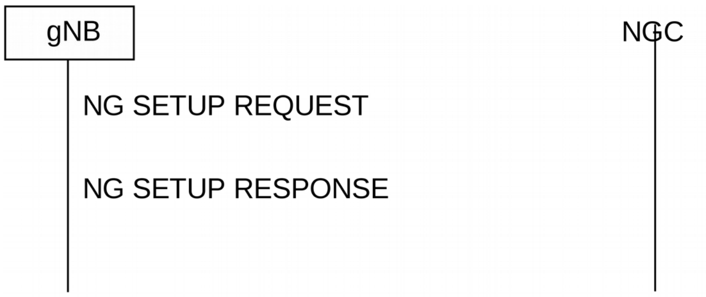


Figure 7.3.4.2.2-1: Example of NG Setup procedure

This procedure is used to setup the NG interface. The procedure enables exchange of configuration parameters between the New RAN and the NGC.

NG Reset Procedure

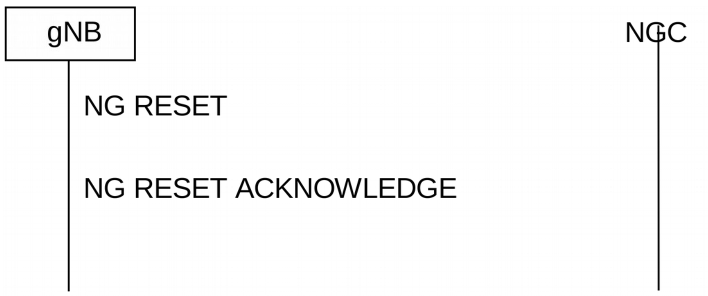


Figure 7.3.4.2.2-2: Example of NG Reset procedure initiated from NGC

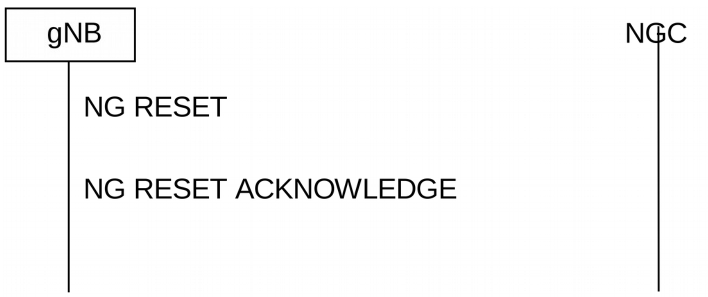


Figure 7.3.4.2.2-3: Example of NG Reset procedure initiated from gNB

This procedure is used to initialize or re-initialize the peer entity or part of the peer entity in the event of a failure in the NGC or vice versa. It can be initiated by either NGC or New RAN.

Error Indication Procedure

Figure 7.3.4.2.2-4: Example of Error Indication procedure initiated from NGC

Figure 7.3.4.2.2-5: Example of Error Indication procedure initiated from gNB

The Error Indication procedure is initiated by a node in order to report detected errors in one incoming message over NG interface, provided they cannot be reported by an appropriate failure message. It can be initiated by either NGC or New RAN.

7.3.4.2.3 UE Context Management Procedures

Initial Context Setup

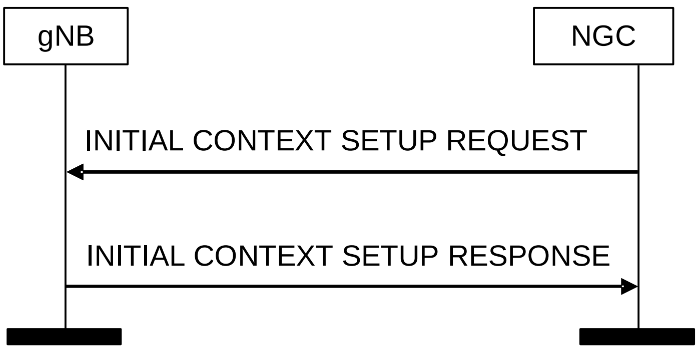


Figure 7.3.4.2.3-1: Example of Initial Context Setup procedure

This procedure is used to establish a context at the New RAN and to establish PDU Sessions between the New RAN and the NGC for a given UE. In this procedure, UE context related information held in the NGC is signalled to the New RAN. When needed, the NGC signals to the RAN information concerning the PDU Sessions for which resources need to be allocated by the serving New RAN node.

UE Context Release

Figure 7.3.4.2.3-2: Example of UE Context Release Request procedure

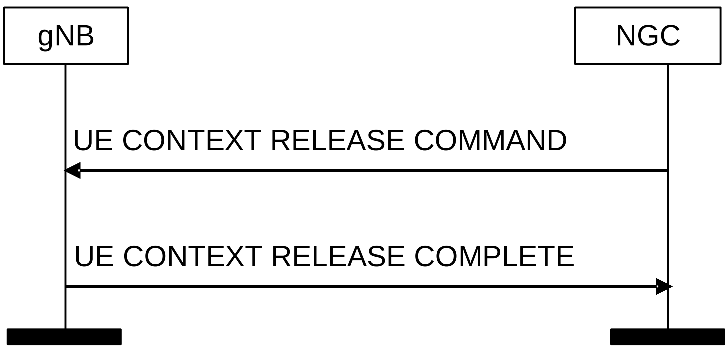


Figure 7.3.4.2.3-3: Example of UE Context Release procedure

This procedure consists of two sequences of messages: one initiated by the New RAN and requesting the NGC to initiate a UE context release; the other initiated directly by the NGC.

This procedure is used to remove an already established UE context and the PDU Sessions associated to it, if any, at any time.

7.3.4.2.4 Transport of NAS Messages Procedures

UL NAS Transport and DL NAS Transport

Figure 7.3.4.2.4-1: Example of Uplink NAS Transport procedure

Figure 7.3.4.2.4-2: Example of Downlink NAS Transport procedure

These two procedures are needed in order to allow an immediate and independent transmission of NAS PDUs between New RAN and NGC.

7.3.4.2.5 UE Mobility Management Procedures

Handover Preparation

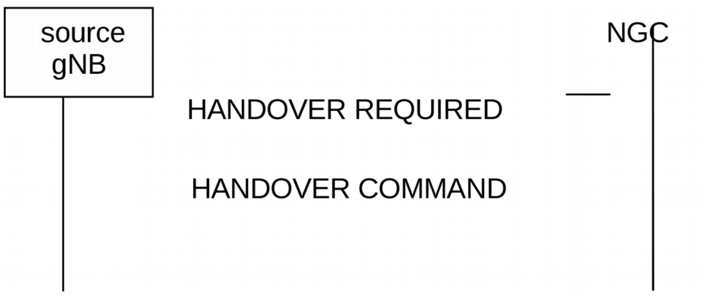


Figure 7.3.4.2.5-1: Example of Handover Preparation procedure

This procedure enables initiation of a handover involving the NGC. In this procedure the source RAN signals to the NGC the intention to handover a UE to a given target New RAN. The source RAN produces and signals information that should be passed transparently to the target New RAN.

In return, the source RAN receives from the NGC information about the resources that can be handed over and those cannot be handed over to the target RAN and that should be released. Also, the source RAN receives information that was sent by the target RAN transparently to the NGC.

Handover Resource Allocation

Figure 7.3.4.2.5-2: Example of Handover Preparation procedure

This procedure is mainly needed to communicate to the target New RAN on resources to be allocated as part of the handover. The target New RAN needs the information to run admission control and to decide whether resources can be allocated. The other fundamental role of this procedure is to create a UE context at the target RAN for the UE that is handed over. Therefore, the NGC should signal to the target RAN information that are relevant to establish a UE context.

In return, the target New RAN signals to the NGC information about the resources that were admitted and those that failed to be admitted. Additionally, the target RAN signals to the NGC information that should be transparently conveyed to the source New RAN.

Path Switch Request

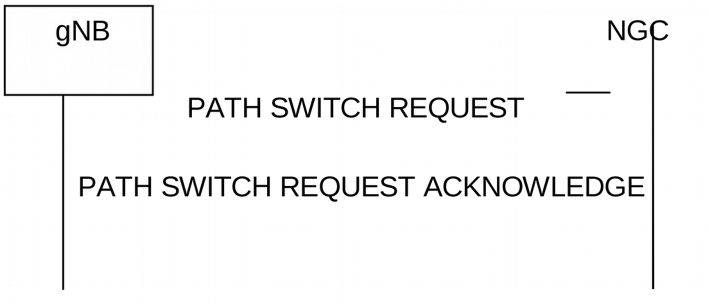


Figure 7.3.4.2.5-3: Example of Path Switch Request procedure

The purpose of the Path Switch Request procedure is to request the switch of a downlink GTP tunnel towards a new GTP tunnel endpoint.

Handover Cancellation

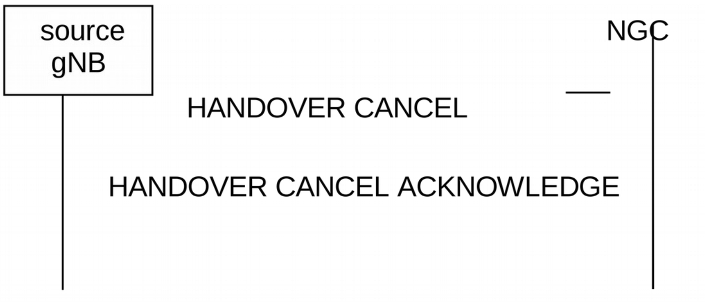


Figure 7.3.4.2.5-4: Example of Handover Cancel

The purpose of the Handover Cancel procedure is to enable a source New RAN node to cancel an ongoing handover preparation or an already prepared handover.

Handover Notification

Figure 7.3.4.2.5-5: Example of Handover Notification procedure

The purpose of the Handover Notify procedure is to enable a target New RAN node to notify NGC about the successfully completed handover.

7.3.4.2.6 Paging Procedure

Figure 7.3.4.2.6-1: Example of Paging procedure

This procedure is used to enable the NGC to page a UE in the specific gNB. At the reception of the PAGING message, the gNB performs paging of the UE.

7.3.4.3 Potential NG Enhancements

Paging

When the NGC is to page a full tracking area (or set of tracking areas), it may send the paging message via IP multicast. One way to support the IP multicast may be:

- each tracking area is configured with an IP multicast address
- all the gNBs of the tracking area join the IP multicast group.
- the paging to the tracking area is sent by NGC to the gNBs by IP multicast using UDP protocol.

7.3.5 NG interface architecture

The NG interface supports a one-to-many relation between New RAN nodes and NGC nodes.

7.3.6 NG Control Plane

7.3.6.1 General

The NG control plane interface (NG-C) is defined between the NR gNB/eLTE eNB and NG-Core entity. The control plane protocol stack of the NG interface is shown on Figure 7.3.6-1. The transport network layer is built on IP transport. For the reliable transport of signalling messages, SCTP is added on top of IP. The application layer signalling protocol is referred to as NG-AP (NG Application Protocol).

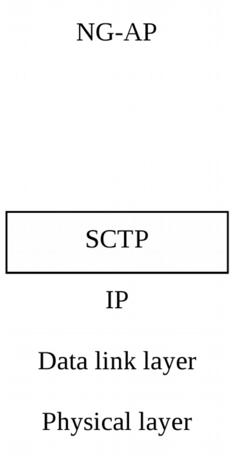


Figure 7.3.6-1: NG Interface Control Plane

The SCTP layer provides the guaranteed delivery of application layer messages.

In the transport IP layer point-to-point transmission is used to deliver the signalling PDUs.

SCTP/IP is the agreed transport protocol for NG-C.

7.3.6.2 List of Potential Requirements/Issues with usage of SCTP

Availability

Problem Statement

NG-C is likely to be terminated in the selected CCNF in an intermediate independent front end function in order to not expose the CCNF internal processing structure to the gNB. With a single SCTP termination point per gNB/CCNF pair a failure affecting the SCTP termination point may require recovery action such as re-initialisation of SCTP associations before service between the eNB and MME can be re-established.

Scalability

Problem Statement

Scalability of a CCNF may require the ability to add or remove both SCTP termination points without interrupting service.

7.3.7 NG User Plane

The NG user plane (NG-U) interface is defined between the gNB/eLTE eNB and the UPGW. The NG-U interface provides non guaranteed delivery of user plane PDUs between the gNB/eLTE eNB and the UPGW.

The NG-U interface shall support per PDU Session tunneling.

NOTE 1: Support of other type tunneling e.g. per node tunneling may be decided at WI phase.

The protocol stack for NG-U is shown in Figure 7.3.7-1.

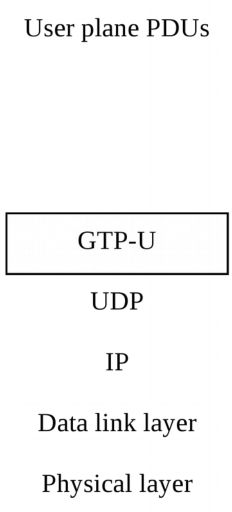


Figure 7.3.7-1: NG-U protocol structure

NOTE 2: GTP-U is a baseline for NG-U, however, later some enhancements or introduction of a new alternative protocol is still possible if justified.

7.4 RAN internal interface

7.4.1 Xn Interface

7.4.1.1 General

The interface allowing to interconnect two gNBs or one gNB and one eLTE eNB with each other is referred to as the Xn interface. The interface Xn is also applicable for the connection between two eLTE eNBs.

7.4.1.2 General principles

The general principles for the specification of the Xn interface are as follows:

- the Xn interface shall be open;
- the Xn interface shall support the exchange of signalling information between the endpoints, in addition the interface shall support data forwarding to the respective endpoints;
- from a logical standpoint, the Xn is a point-to-point interface between the endpoints. A point-to-point logical interface should be feasible even in the absence of a physical direct connection between the endpoints.
- the Xn interface shall support control plane and user plane separation;
- the Xn interface shall separate Radio Network Layer and Transport Network Layer;
- the Xn interface shall be future proof to fulfil different new requirements, support new services and new functions.

NOTE: Whether and how to document the application protocol for Xn with regards to X2AP will be decided in the normative phase.

7.4.1.3 Xn Interface Functions

The Xn-C interface supports the following functions:

- Xn interface management and error handling function to manage the Xn-C interface;
  - Error indication;

  - Setting up the Xn;
  - Resetting the Xn;
  - Updating the Xn configuration data;
  - Xn removal.
- UE connected mode mobility management: function to manage the UE mobility for connected mode between nodes in the New RAN;
  - Handover preparation;
  - Handover cancellation.
- UE context retrieval: function to retrieve UE context from another node in the New RAN;
- Dual connectivity: function to enable usage of additional resources in a secondary node in the New RAN;

NOTE: Dual connectivity between two gNBs can be discussed in normative phase..

- Interference coordination: function to manage inter-cell interference;
- Self-optimization: function to autonomously adapt radio parameters.

The Xn-U interface supports the following functions:

- Data forwarding
- Flow control

7.4.1.4 Xn Interface Procedures

To support the functions listed in Section 7.4.1.2, the Xn interface should support the following procedures. The procedures are classified in different categories.

Xn Interface Management and Error Handling Procedures:

- Xn Setup: to establish an Xn interface between the New RAN nodes
- Xn Reset: to reset the Xn interface
- Error Indication: to report detected errors in one incoming message
- New RAN Configuration Update: to update the configuration for the New RAN nodes over Xn interface
- Xn Removal: to remove the signaling connection between two New RAN nodes in a controlled manner

UE connected mode Mobility Management Procedures:

- Handover Preparation: to establish necessary resources in a New RAN node for an incoming handover
- Handover Cancellation: to cancel an ongoing handover preparation or an already prepared Xn handover
- UE Context Release: to indicate to the source new RAN that radio and control plane resources for the associated UE context are allowed to be released

Dual Connectivity Procedures:

- Secondary Node Addition
- Secondary Node Modification (Master node initiated)
- Secondary Node Modification (Secondary node initiated)
- Secondary Node Release (Master node initiated)
- Secondary Node Release (Secondary node initiated)

NOTE 1: Work in other RAN WGs on inter-RAT DC for New RAN, on 5G intra-system mobility and NR specifics needs to be revisited in order to understand whether DC signalling schemes as specified for E-UTRA can be re-used for 5G. It is however expected, that role definitions from E-UTRAN (master node, secondary node) can still be applied for options 4/4a and 7/7a.

Interference coordination procedures:

NOTE 2: Interference coordination procedures will be addressed in the work item phase, and are pending input from RAN1 on NR-ICIC schemes. LTE supported a framework for inter-node coordination that enabled a number of different ICIC schemes, and a similar framework may be adapted for NR.

7.4.1.5 Xn Control Plane

The Xn control plane interface (Xn-C) is defined between two neighbour New RAN nodes. The control plane protocol stack of the Xn interface is shown on Figure 7.4.1.5-1 below. The transport network layer is built on SCTP on top of IP. The application layer signalling protocol is referred to as Xn-AP (Xn Application Protocol).

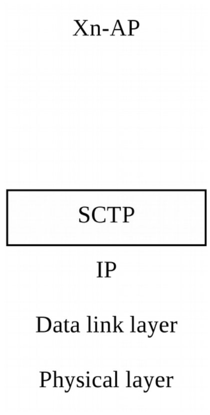


Figure 7.4.1.5-1: Xn Interface Control Plane

7.4.1.6 Xn User Plane

The transport layer for data streams over Xn is an IP based Transport. The following figure shows the transport protocol stacks over Xn.

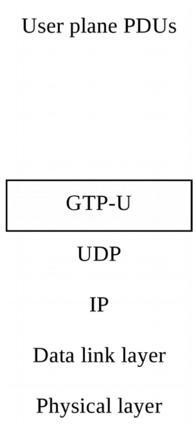


Figure 7.4.1.6-1: Xn Interface User Plane

NOTE: GTP-U is a baseline for Xn-U, however, later some enhancements or introduction of a new alternative protocol is still possible if justified.

8 Realization of Network Slicing

8.1 Key principles for support of Network Slicing in RAN

Network Slicing is a new concept to allow differentiated treatment depending on each customer requirements. With slicing, it is now possible for Mobile Network Operators (MNO) to consider customers as belonging to different tenant types with each having different service requirements that govern in terms of what slice types each tenant is eligible to use based on Service Level Agreement (SLA) and subscriptions.

NSSAI (Network Slice Selection Assistance Information) includes one or more S-NSSAIs (Single NSSAI). Each network slice is uniquely identified by a S-NSSAI, as defined in TR 23.799 [6]. The UE may store a Configured and/or Accepted NSSAI per PLMN. The NSSAI can have standard values or PLMN specific values.

NOTE 1: For signaling between RAN and CN a Slice ID is represented by an NSSAI or S-NSSAI. For the air interface, it is up to RAN groups to decide how to carry/define NSSAI information in RRC (the term “slice ID” is used in the following to refer to this).

The following key principles apply for support of Network Slicing in RAN. The RAN throughout the whole section refers to the new RAN, including both gNB and eLTE eNB.

RAN awareness of slices

- RAN shall support a differentiated handling of traffic for different network slices which have been pre-configured. How RAN supports the slice enabling in terms of RAN functions (i.e. the set of network functions that comprise each slice) is implementation dependent.

Selection of RAN part of the network slice

- RAN shall support the selection of the RAN part of the network slice, by one or more slice ID(s) provided by the UE or the CN which unambiguously identifies one or more of the pre-configured network slices in the PLMN. The Accepted NSSAI is sent by CN to UE and RAN after network slice selection

Resource management between slices

- RAN shall support policy enforcement between slices as per service level agreements. It should be possible for a single RAN node to support multiple slices. The RAN should be free to apply the best RRM policy for the SLA in place to each supported slice.

Support of QoS

- RAN shall support QoS differentiation within a slice.

RAN selection of CN entity

- For initial attach, the UE may provide one or more slice ID(s). If available, RAN uses the slice ID(s) for routing the initial NAS to an NGC CP function. If the UE does not provide any slice ID(s) the RAN sends the NAS signalling to a default NGC CP function.

- For subsequent accesses, the UE provides a Temp ID, which is assigned to the UE by the NGC, to enable the RAN to route the NAS message to the appropriate NGC CP function as long as the Temp ID is valid (RAN is aware of and can reach the NGC CP function which is associated with the Temp ID). Otherwise, the methods for initial attach applies.

NOTE 2: the definition of the Slice ID for use over the air interface is subject to further discussions

Resource isolation between slices

- RAN shall support resource isolation between slices. RAN resource isolation may be achieved by means of RRM policies and protection mechanisms that should avoid that shortage of shared resources in one slice breaks the service level agreement for another slice. It should be possible to fully dedicate RAN resources to a certain slice. How RAN supports resource isolation is implementation dependent.

Slice Availability

- Some slices may be available only in part of the network. Awareness in a gNB of the slices supported in the cells of its neighbouring gNBs may be beneficial for inter-frequency mobility in connected mode. If such awareness is also beneficial for intra-frequency mobility may be discussed in the normative phase. It is assumed that the slice configuration does not change within the UE's registration area.

- The RAN and the CN are responsible to handle a service request for a slice that may or may not be available in a given area. Admission or rejection of access to a slice may depend by factors such as support for the slice, availability of resources, support of the requested service by other slices.

Possible solutions for how slice availability may be handled during mobility may be discussed in the normative phase e.g.:

- Neighbours may exchange slice availability on the interface connecting two nodes, e.g. Xn interface between gNBs.

- The core network could provide the RAN a mobility restriction list. This list may include those TAs which support or do not support the slices for the UE.

- The slices supported at the source node may be mapped, if possible, to other slices at target node. Examples of possible mapping mechanisms that can be studied in normative phase are:

  - Mapping by the CN, when there is naturally a signalling interaction between RAN and CN and performance is thus not impacted;

  - Mapping by the RAN as action following prior negotiation with the CN during UE connection setup;

  - Mapping by the RAN autonomously, when involving the CN would not be a practical solution and if prior configuration of mapping policies took place at RAN;

Support for UE associating with multiple network slices simultaneously

- In case a UE is associated with multiple slices simultaneously, only one signalling connection shall be maintained.

8.2 Solutions for selection of Network slice and CN entity by gNB

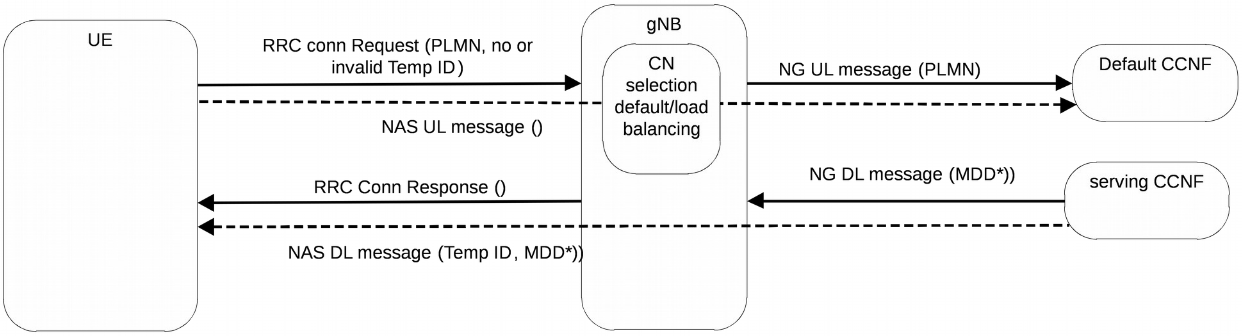
8.2.1 Solution 1

NOTE : this section presents a RAN solution for network slice selection and CN entity selection which is based on the solution described in section 6.1.2 of TR 23.799 [6]. This solution enables support of multiple slices simultaneously supported by a UE.

In this solution the Network Slice and the CN entity are selected based on a Multi-Dimensional Descriptor called MDD as described in TR 23.799 [6]. In this case the CN entity is a common control plane entity in NGC (CCNF) – common to all supported slices - as described in section 6.1.2 of TR 23.799 [6].

The RAN selection of the Network slice and the selection of the appropriate CCNF can be partitioned into the following five cases.

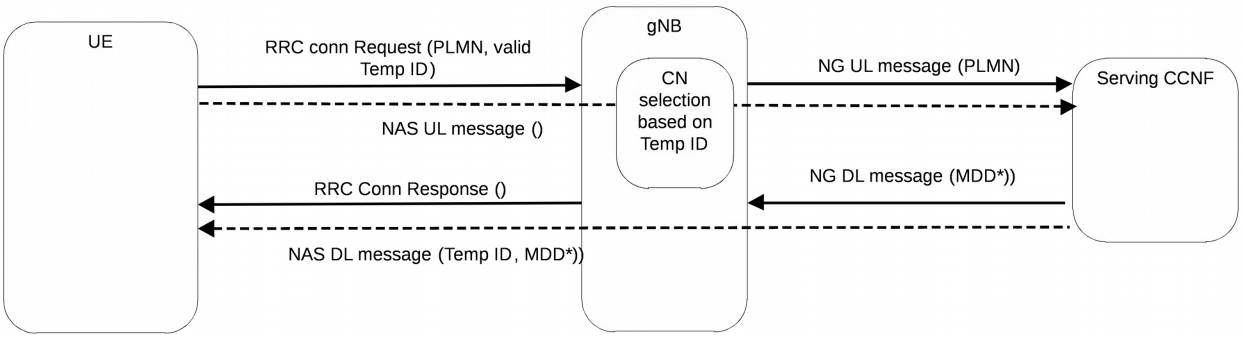
<u>Selection case 1: no MDD, no or invalid Temp ID</u>



Valid temp ID means valid for this PLMN.

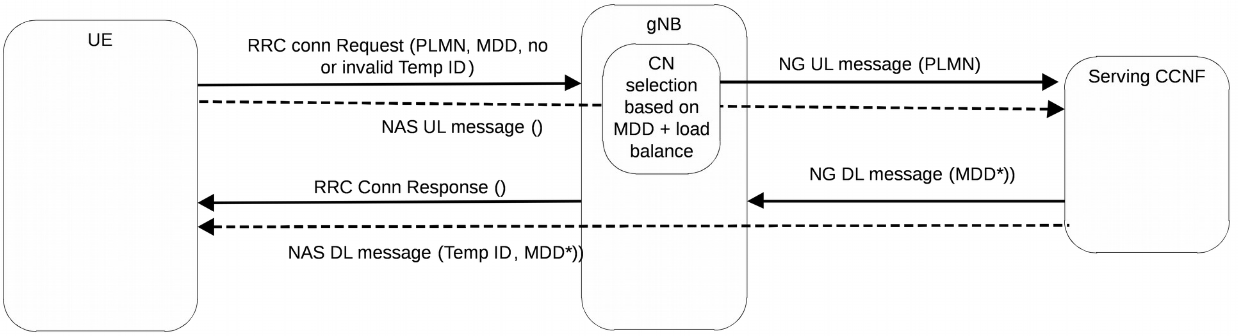
In absence of Temp ID information valid for the indicated PLMN in the RRC connection request, the gNB needs to route to a default configured CCNF. After validation steps, the Default CCNF, depending on UE capabilities and subscription information, will select a suitable serving CCNF. This serving CCNF could be the default CCNF itself. A Temp ID corresponding to serving CCNF is sent back to the UE at NAS level. In case an MDD applies for the selected PLMN, this MDD is sent back to the UE together with the Temp ID at NAS level. The MDD is then also indicated in the NG DL message to gNB.

<u>Selection case 2: no MDD, valid Temp ID</u>



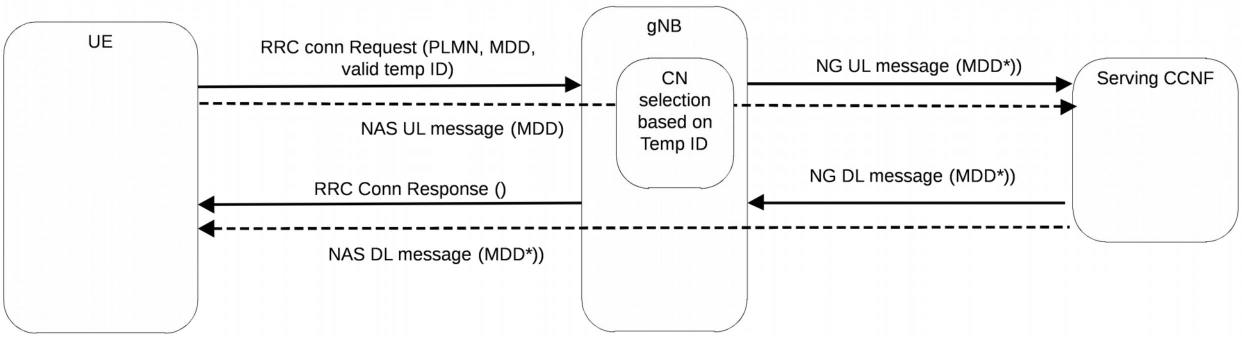
If a Temp ID information valid for the indicated PLMN is included in RRC Connection Request, the gNB routes based on Temp ID to the serving CCNF. In case an MDD applies for the selected PLMN, this MDD is sent back to the UE at NAS level. The MDD is then also indicated in the NG DL message to gNB.

<u>Selection case 3: MDD, no or invalid Temp ID</u>



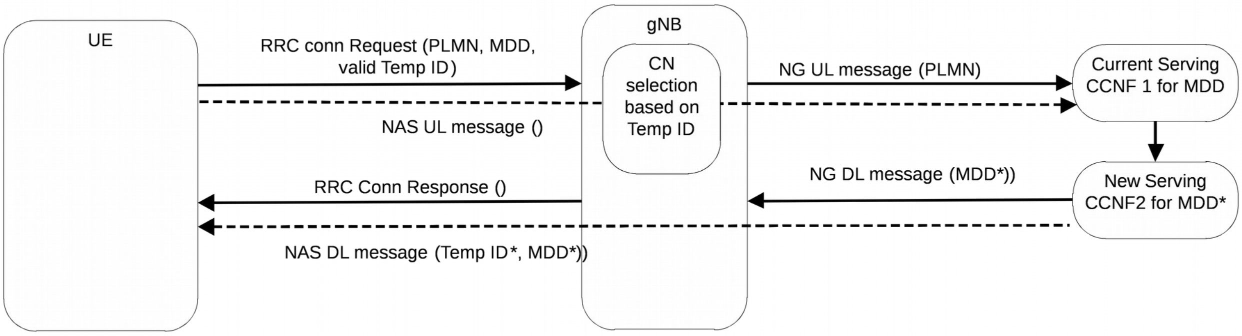
n absence of Temp ID information valid for the indicated PLMN in the RRC connection request, the gNB selects a serving CCNF based on the MDD when present. If multiple CCNF are available and suitable for the MDD, gNB can apply load balancing criteria. RAN also uses the included MDD for RAN slice awareness (slice specific behavior, congestion control, resource isolation). After validation steps, a Temp ID corresponding to serving CCNF is sent back to the UE at NAS level. The MDD is sent back to the UE together with the Temp ID at NAS level to either reconfirm the initially requested MDD or indicates a new one if policy has changed. The MDD is also indicated in the NG DL message to gNB for confirmation.

Selection case 4: MDD, valid Temp ID



If a Temp ID information valid for the indicated PLMN is included in RRC Connection Request, the gNB routes based on Temp ID to the serving CCNF regardless of MDD presence. RAN uses the included MDD for RAN slice awareness (slice specific behavior, congestion control, resource isolation). The MDD is sent back to the UE with the Temp ID at NAS level to either reconfirm the initially requested MDD or indicates a new one if policy has changed . The MDD is then also indicated in the NG DL message to gNB for confirmation.

Selection case 5: Modification of MDD, valid Temp ID



If a Temp ID information valid for the indicated PLMN is included in RRC Connection Request, the gNB routes based on Temp ID to the serving CCNF regardless of MDD presence. RAN uses the included MDD for RAN slice awareness (congestion control, resource isolation). If the current MDD needs to be updated into MDD* because some of the MDD vectors are no more supported (i.e. operator policy has changed for the UE due to UE capabilities e.g. the subscriber may use different types of UEs by swapping the UICC) and if CCNF1 is not compatible with the new MDD*, the NGC may select a new CCNF2 compatible with MDD*. In this case the MDD* is sent back from CCNF2 to the UE together with a new Temp ID* corresponding to CCNF2 at NAS level. The new MDD* is then also indicated to gNB over NG DL message.

8.2.2 Solution 2

Network Slice selection may be realised in at least two mechanisms. One possible way is to configure the UE with a list of slice IDs to which it is allowed to access. Another way is to let the CN notify the UE of the Slice IDs it can access and let the UE rely on such information onwards.

The first case of UE configured Slice ID is made of the following general steps:

- The UE will be configured with Slice IDs and with criteria to present the right Slice ID to the RAN.
- When the UE requests to access the RAN it will present a Slice ID as per criteria previously configured.
- By receiving the Slice ID, the RAN shall be able to
  - Identify the policies that apply to the selected slice and assign RAN resources accordingly.
  - Identify the CN node that supports the slice ID presented by the UE.

In order to achieve the CN node selection the RAN would need to be informed about the slice IDs supported by connected CN nodes. This can be achieved by letting RAN and CN exchange information on supported slice IDs at RAN-CN interface setup.

Figure 8.2.2-1 summarises the steps described in the case of Slice ID provided by the UE.

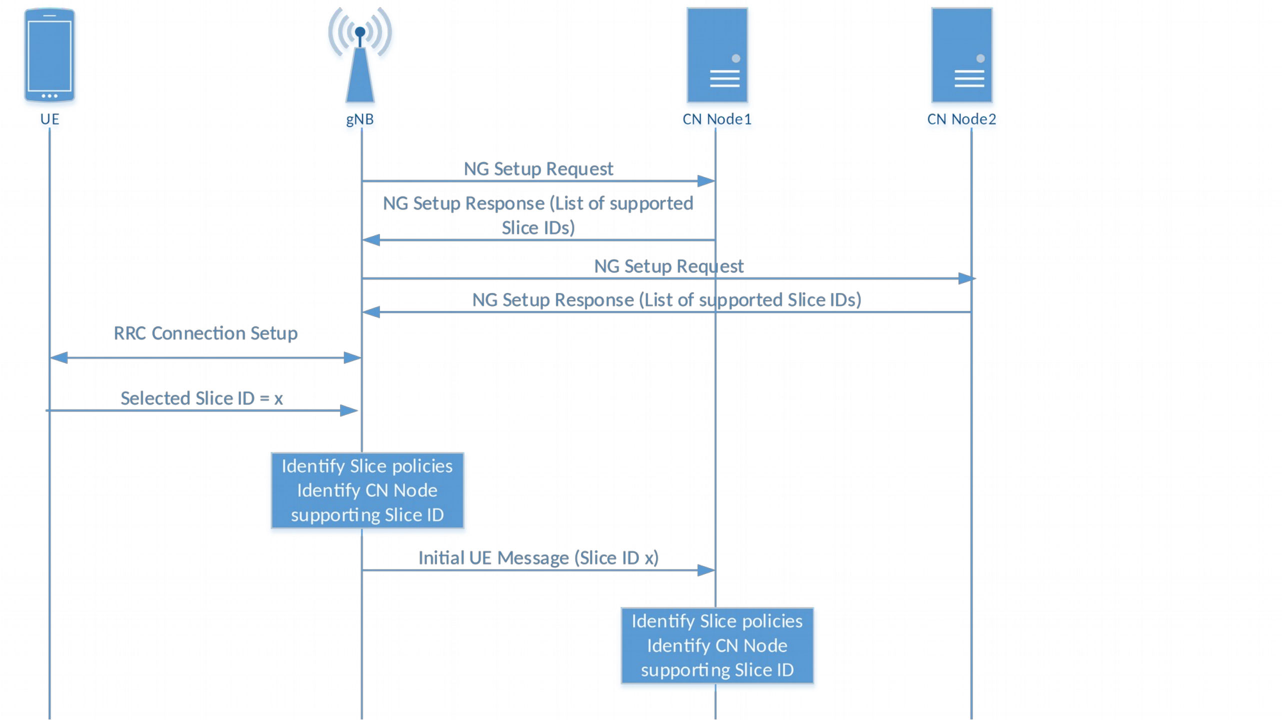


Figure 8.2.2-1: Slice selection when Slice ID is provided by the UE to RAN

The second case, where the slice ID is not initially provided by the UE and it is configured by the CN to the UE, consists of the following general steps:

- The UE presents a NAS Service Request with no specific Slice ID.
- The serving RAN routes the NAS PDU to a default CN Node.
- The default CN node analyses the NAS request and retrieves UE subscriber information. Based on the UE subscriber information and NAS Service Request the default CN Node reroutes the request to a different CN Node. The latter could be achieved by reusing procedures similar to existing S1: Reroute NAS Request.
- The new CN Node (reached via reroute) responds to the RAN with an assigned Slice ID for the UE and it responds to the UE with a NAS PDU in which usable Slice IDs are provided.
- From this moment onwards the UE will present to the RAN a slice ID that intends to access at RRC connection establishment.

Figure 8.2.2-2 summarises the steps described in the case of Slice ID configured by the CN to the UE.

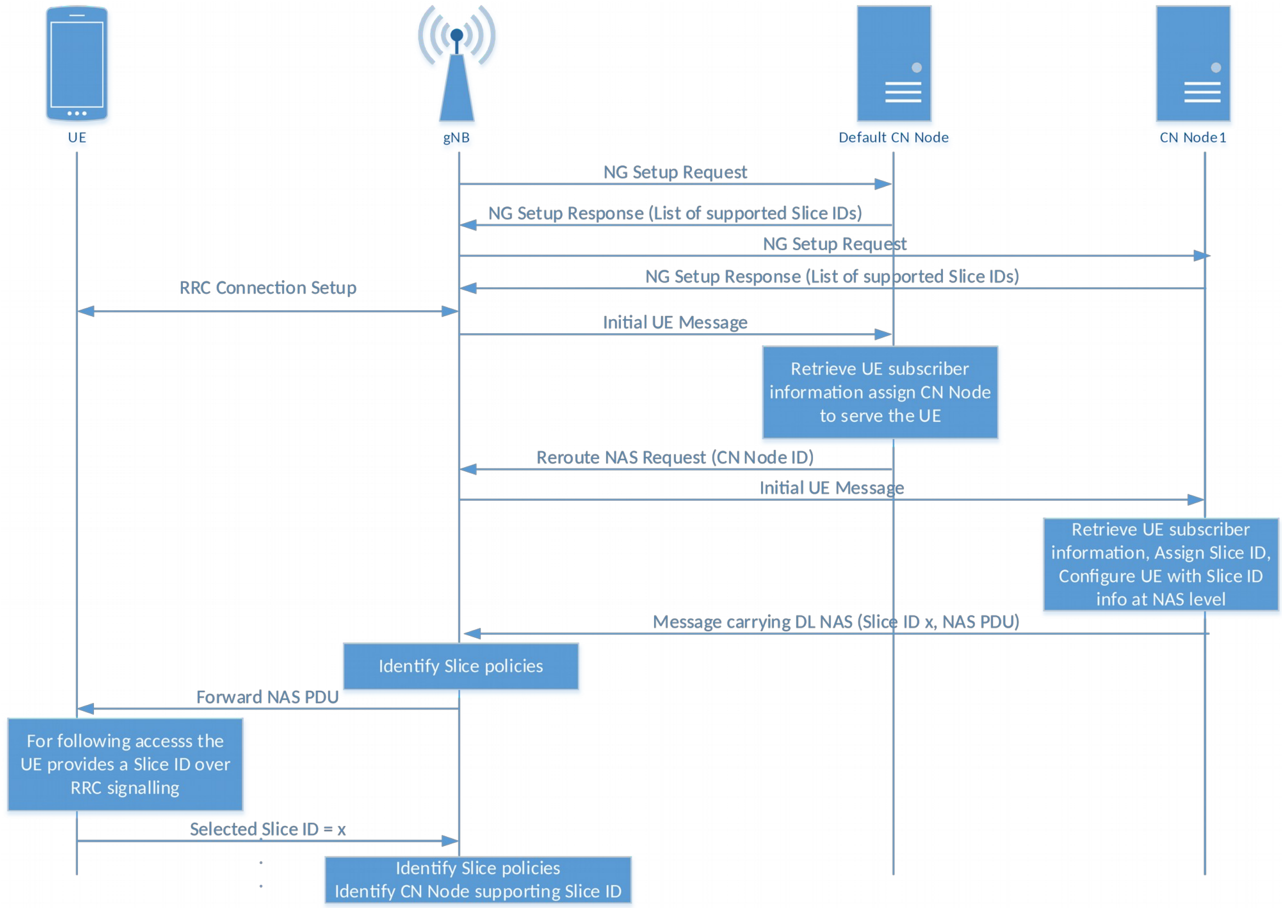


Figure 8.2.2-2: Slice selection when Slice ID is provided by the CN

In both mechanisms described above it is assumed that a UE can only access in parallel the network slices supported by the CN node serving it. Namely, a UE cannot access multiple slices served by different CN nodes at the same time.

8.2.3 Solution 3

NOTE : this section presents a RAN solution for network slice selection and CN entity selection which is based on solution described in section 6.1.1 of TR23.799 [6]. The solution does not make any assumption on any potential RAN internal slicing and any association with network instance is performed network internally. This solution also enables a UE to support multiple slices simultaneously..

The RAN selection of the Network slice and the selection of the appropriate CN entity can be partitioned into the following two cases:

Case 1: Initial Signaling Routing

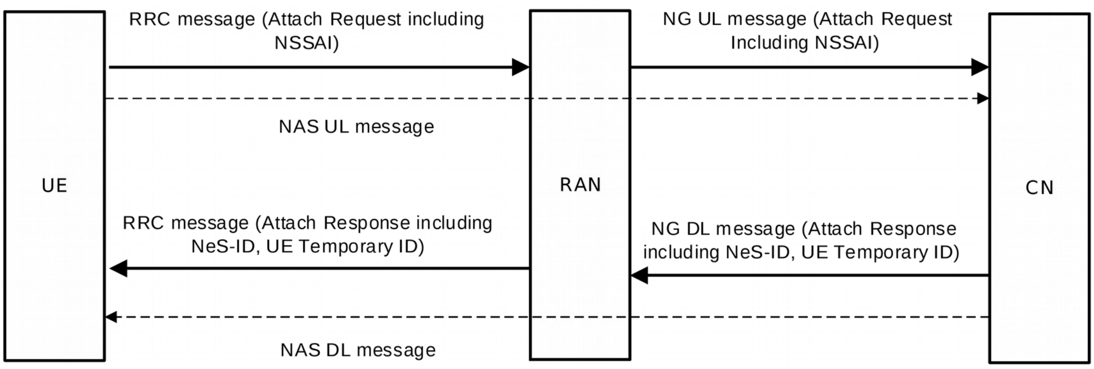


Figure 8.2.3-1: Selection of CN entity for initial signalling routing

As Figure 8.2.3-1 shows, the UE sends initial Attach Request to the RAN including network slice selection assistance information (NSSAI) including UE capability, Service type and requested service (optional) etc. The RAN forwards the Attach Request to the CN. The CN determines the slice(s) and selects the appropriate Common CP NFs as well as related NeS-ID of the slice instance. The selected Common CP NFs assigns the UE Temporary ID and the CN sends the Attach Response with NeS-ID and UE Temporary ID to the UE.

<u>Case 2: Subsequent Signaling Routing</u>

After the UE registered with a network slice instance, the UE is provisioned with two identities in terms of network slice type ID (NeS-ID) and Temporary ID which are used for RAN selection of CN entities for accessing to the network slice.

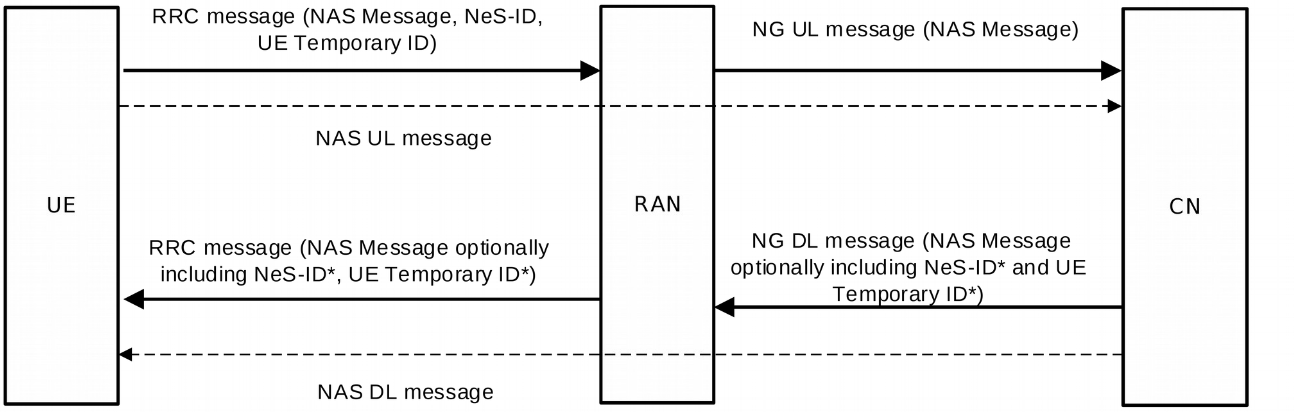


Figure 8.2.3-2: Selection of CN entity for subsequent signalling routing

As Figure 8.2.3-2 shows, for the subsequent NAS signalling routing after the UE is attached to the system, it can send a NAS message to the RAN with the NeS-ID and the UE Temporary ID carried in RRC message. If the UE Temporary ID is valid, the RAN can send the NAS message directly to the Common CP NF, based on the UE Temporary ID information. Otherwise, the RAN determines the dedicated Common CP NFs based on the NeS-ID. The dedicated Common CP NFs responds to the UE by sending to the RAN a suitable NAS message, which includes, if necessary, a new NeS-ID (optional) and new UE Temporary ID to the UE.

8.2.4 Conclusion

RAN selects CCNF based on Temp ID and slice ID(s) from RRC. Table 8.2.4-1 summarizes the CCNF selection strategy in different Temp ID and Slice ID(s) presence combinations.

Table 8.2.4-1 CCNF selection based on Temp ID and slice ID(s)

Temp ID	Slice ID(s)	CCNF selection by RAN
NA or invalid	NA	Default CCNF is selected
NA or invalid	Present	Selects CCNF which supports UE requested slice ID(s)
Valid	NA or present	Selects CCNF per CN identity information in Temp ID

RAN may use slice ID(s) from RRC for selection of the RAN part of Network Slice before final slice(s) selection is indicated by the CN.

8.3 Resource Isolation between slices

Resource isolation enables specialized customization and avoids one slice affecting another slice. Hardware/software resource isolation is up to implementation. Each slice may be assigned with either shared or dedicated radio resource up to RRM implementation and SLA.

When assigned dedicated radio resource the slice may be isolated and configured by RAN with one or more of below items:

- Time/frequency/code resources etc;
- Access channel;

NOTE : It is up to RAN1/RAN2 to decide how to partition access channel e.g. in frequency, time and preamble.

- Independent Access control, Load control, QOS etc.

Logically, slices may be isolated in terms of DRBs.

The RAN should be allowed to serve traffic for different slices via shared resources.

8.4 Resource management between slices

To enable differentiated handling of traffic for network slices with different SLA:

- RAN is configured with a set of different configurations for different network slices;
- To select the appropriate configuration for the traffic for each network slice, RAN receives a slice ID indicating which of the configurations applies for this specific network slice.

8.5 Mobility

8.5.1 General

To make mobility slice-aware in case of Network Slicing, Slice ID is introduced as part of the PDU session information that is transferred during mobility signalling. This enables slice-aware admission and congestion control.

8.5.2 Connected Mode Mobility

When a target cell is selected, handover signalling is initiated. Such procedure attempts to move PDU Session resources for all active slices of the UE from one source node to a target node. The source gNB needs to pass on slices that a UE in question is using to a target gNB as part of the HO procedure.

If a handover procedure involves a NGC, during such procedure the target AMF is responsible for aligning the set of slices supported in the new Registration Area between UE and network at NAS level. PDU Sessions that are associated with the removed slices are not admitted at target node.

An example of such call flow, for the case of active mode mobility across different Registration Areas, is shown in Figure 8.5.2-1 for the case of CN involved handover.

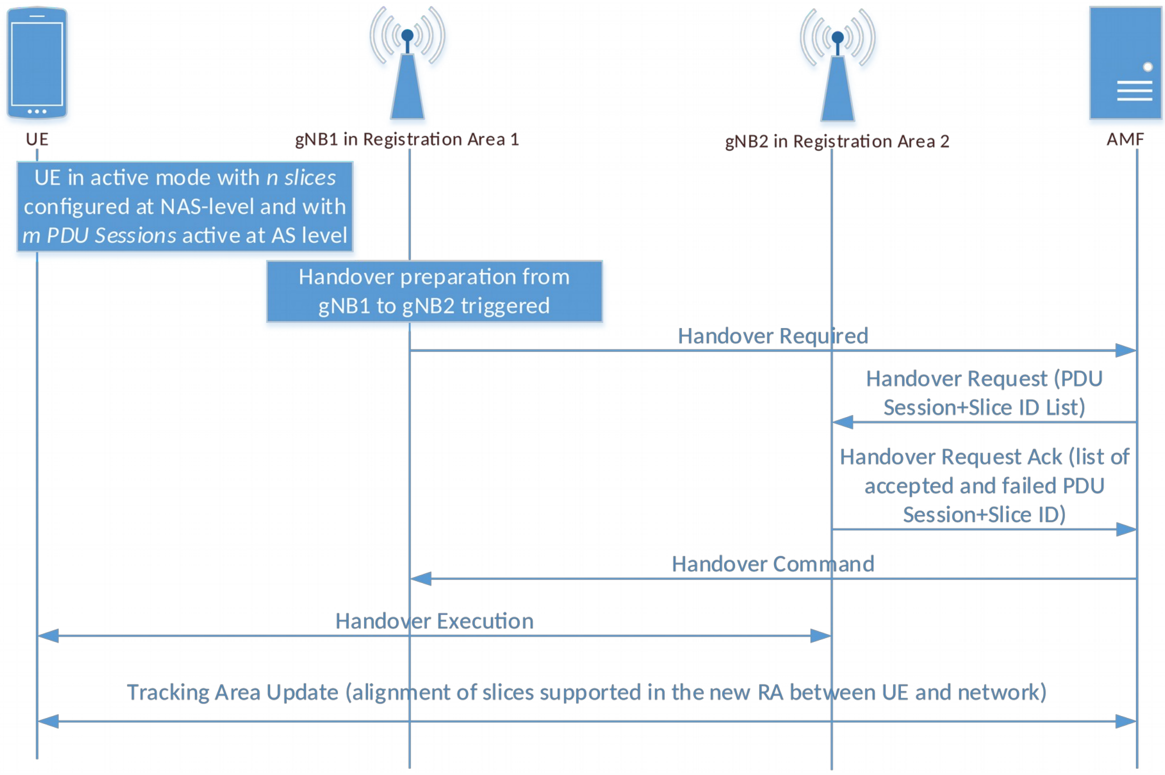


Figure 8.5.2-1: example of call flow for slice access management in Active mode CN involved mobility across different Registration Areas

8.6 Validation of the UE rights to access a NW slice

It is the responsibility of the CN to validate that the UE has the rights to access a NW slice. Prior to receiving the Initial Context Setup Request message, the RAN may be allowed to apply some provisional/local policies, based on awareness of which slice the UE is requesting access to.

It should be noted that, additionally, the CN is aware of all NW slices the UE belongs to. During the initial context setup, the RAN is informed for all NW slices for which resources are being requested.

8.7 Granularity of slice awareness

To respect the key principles for the support of Network Slicing in RAN, outlined in sub clause 8.1, slice awareness in RAN is introduced at PDU session level, by indicating the Slice ID corresponding to the PDU Session.

This implies:

- All QoS flows within a PDU session shall belong to the same NW slice;
- Within a slice, QoS differentiation is supported;
- Connection of a UE to multiple NW slices is supported, as multiple PDU sessions per UE can be established;
- As a consequence of slice awareness at PDU Session level, user data pertinent to different NW slices will not share the same NG-U tunnel;
- By adding the Slice ID information to the PDU session information, mobility signalling becomes also slice-aware and enables per-slice admission and congestion control.

8.8 Impacts on RAN signalling

8.8.1 General

Following the principles of UE rights validation and slice awareness granularity highlighted in previous sub clauses, it can be seen that NW slice information can be conveyed in RAN signalling as follows:

8.8.2 UE Context Handling

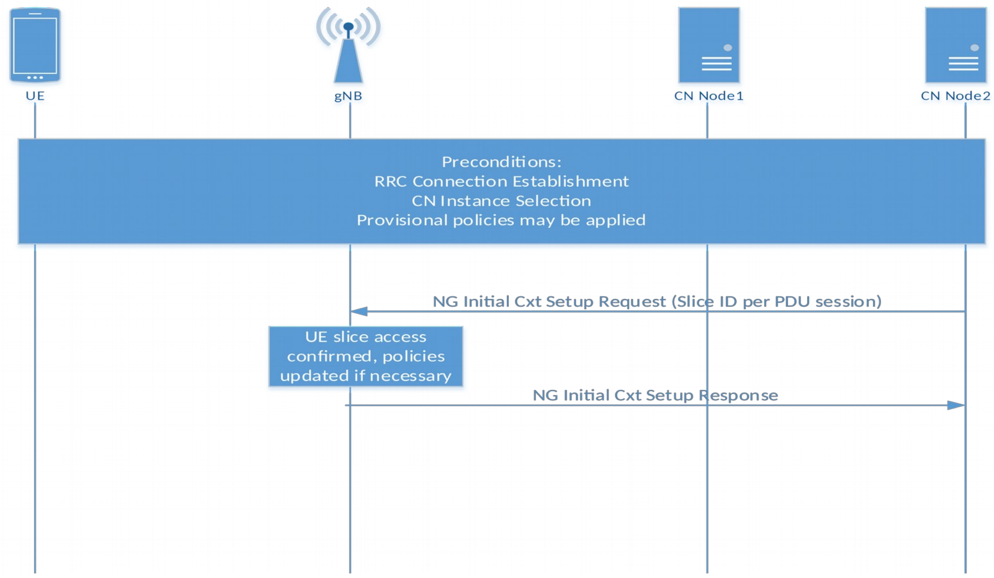


Figure 8.8.2-1: NW Slice-aware Initial Context Setup

Following the initial access, the establishment of the RRC connection and the selection of the correct CN instance, the CN establishes the complete UE context by sending the Initial Context Setup Request message to the gNB over NG-C.

The message contains the Slice ID as part of the PDU session/s resource description.

Upon successful establishment of the UE context and allocation of PDU resources to the relevant NW slice/s, the RAN responds with the Initial Context Setup Response message.

8.8.3 PDU Session Handling

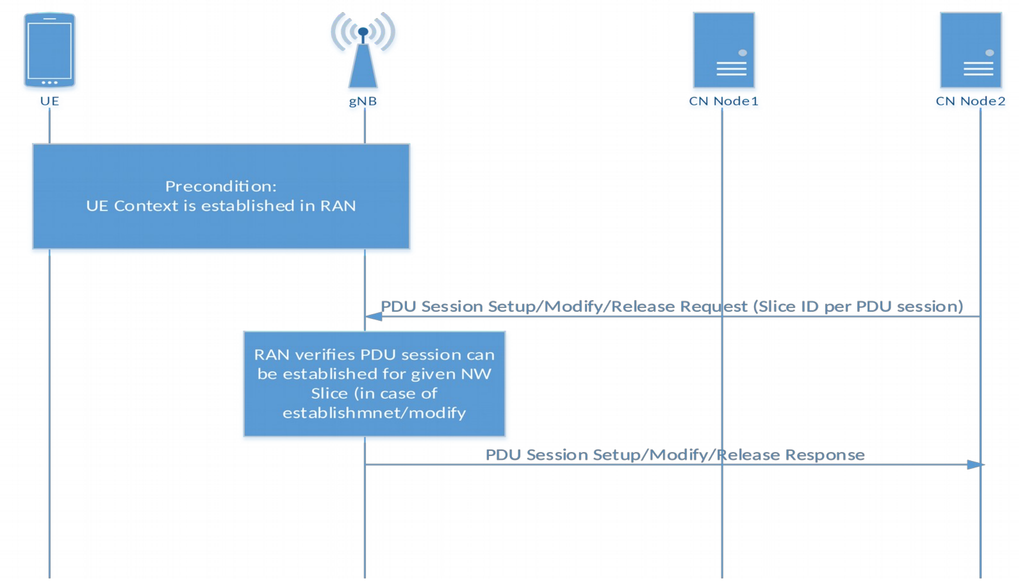


Figure 8.8.3-1: NW Slice-aware PDU Session Setup/Modify/Release

When new PDU sessions need to be established or existing ones modified or released, the CN requests the RAN to allocate/release resources relative to the relevant PDU sessions by means of the PDU Session Setup/Modify/Release procedures over NG-C. In case of network slicing, Slice ID information is added per PDU session, so RAN is enabled to apply policies at PDU session level according to the SLA represented by the NW slice, while still being able to apply (for example) differentiated QoS within the slice.

RAN confirms the establishment/modification/release of a PDU session associated to a certain NW slice by responding with the PDU Session Setup/Modify/Release Response message over the NG-C interface.

9 QoS

9.1 Key principles for QoS in RAN

The following design principles for handling PDU Sessions and QoS flows related context data and respective resources at the New RAN apply:

- New RAN receives QoS related information through NG-C PDU Session control signalling.
- A PDU Session context includes a per PDU session default QoS profile and may include per PDU session pre-authorised QoS profiles.

NOTE 1: Whether RAN needs to be aware of which NAS-level QoS profile is to be regarded as the default profile is to be further discussed in normative phase.

NOTE 2: Any impact on RAN from reflective QoS is to be discussed in the normative phase.

- During the lifetime of PDU Session context, the per PDU session QoS profile may be modified, added or deleted.
- GBR QoS flow management requires explicit signalling on NG-C.

- New RAN does not support packet filtering.
- A QoS profile is either defined as a “A-type QoS profiles” which have standardized QoS characteristics, or along a “B-type QoS profiles”which has QoS characteristics dynamically signalled over NG-C.

NOTE 3: The content of a QoS profile rules, A-type and B-type, needs to be determined in the normative phase.

- QoS profiles established for a PDU Session can be represented as a linear, indexed list of implicit (standardised) or explicit (dynamic) QoS profile descriptors.

The following design principles for handling QoS aspects of the UE context at the New RAN apply:

- A per UE UL and DL rate limit for non-GBR QoS flows, provided to the serving New RAN node, shall be obeyed.

The following design principles for handling QoS aspects on NG-U apply:

- User plane marking for QoS is carried in encapsulation header on NG-U. The QoS mark provides sufficient information for the RAN to identify the related QoS flow.
- Upon detection of a new non-GBR QoS flow on NG-U the New RAN node decides the mapping to radio resources, i.e it may decide to create new radio resources or map it to existing radio resources

Figure 9.1-1 depicts an example representation of a UE Context descriptor containing PDU Session and QoS flow related data and does not preclude any future protocol definition:

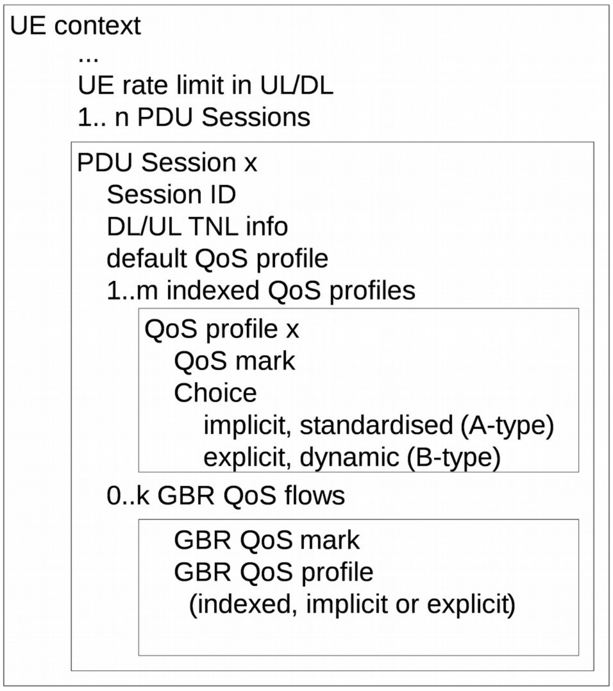


Figure 9.1-1: Example representation of a UE Context descriptor containing PDU Sessions and QoS flows

The format and content of the NG-U encapsulation header is to be discussed in normative phase.

An example of content of the respective NG-U encapsulation header is depicted as follows:

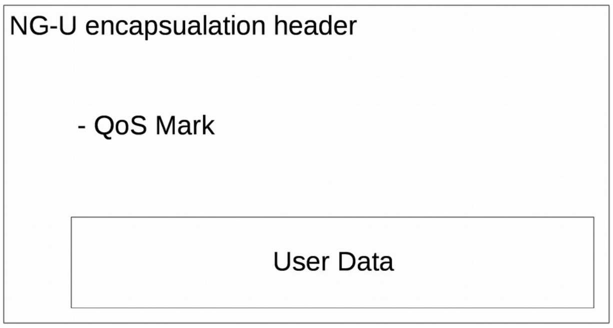


Figure 9.1-2: Example representation of the NG-U encapsulation header

NOTE 4: Whether the QoS mark is always sent over NG-U is related to NOTE 1.

The resulting information flow between the UE, New RAN and NGC can be depicted in the following way:

- The NGC provides QoS rule to the UE: NAS-level QoS profiles (A- or B-type), packet filters and precedence order.
- RAN will apply a specific QoS profile based on information received from the Core Network. RAN receives QoS profiles using NG-C PDU Session control signalling.
- User plane marking for QoS is carried in encapsulation header on NG-U.
- For downlink traffic, it is up to RAN to bind the traffic onto a corresponding DRB based on the NG-U marking and the corresponding QoS characteristics provided through NG-C signalling, also taking into account the PDU session associated with the DL packet.
- For uplink traffic, RAN determines the appropriate QoS Mark and includes it in the NG-U encapsulation header towards the CN as described in section 9.2.

Figure 9.1-3 shows the end-to-end NR QoS framework with the mapping of the QoS attributes.

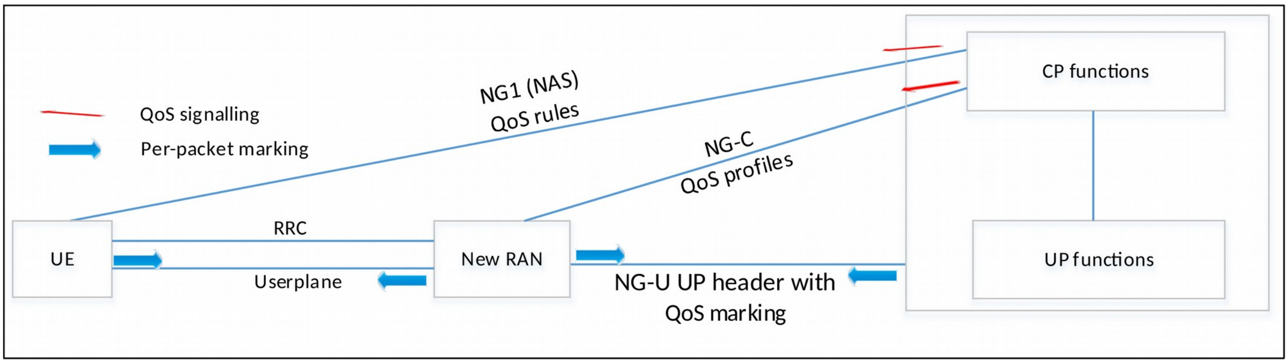


Figure 9.1-3. NR QoS Framework description

NOTE 5: NG1 (defined in TR 23.799 [6]) is a reference point for the control plane between UE and NGC.

9.2 Uplink QoS marking

The New RAN shall determine and include the NG-U QoS mark in the NG-U encapsulation header of uplink packets so that the CN can apply appropriate charging and may do some validation checks.

Depending on discussion in RAN2, the New RAN may e.g. either:

- infer the NG-U QoS mark value from the DRB over which the UL packets have been received in case there is only one UL QoS flow mapped on a DRB,

- or infer the NG-U QoS mark value from the flow marking value received over the radio in case there are multiple UL QoS flows mapped on the same DRB.

9.3 QoS impact on handover procedure

During handover, data forwarding may be performed at PDU level, RB level or QoS flow level. The possibilities of data forwarding solution will be discussed in normative phase.

10 Radio access network procedures

10.1 Dual Connectivity between NR and LTE

10.1.1 General

Option 3/3a/3x, 4/4a and 7/7a/7x of deployment scenarios can be considered as tight interworking between NR and E-UTRA.

In Option 3/3a, Dual Connectivity (DC) specified in TS 36.300 [12] and relevant stage 3 specifications (e.g., TS 36.423 [13]) should be reused as baseline considering the fact that EPC should not be impacted. Therefore, for the Xx interface between LTE eNB and gNB, the procedures and protocols would remain alike those of DC, while minor enhancements might not be ruled out. In Option 3x defined in Section 10.1.2.4, further enhancements are needed on top of LTE based DC.

In Option 4/4a, Dual Connectivity can be realized, in which the gNB (similar role as MeNB in TS 36.300 [12]) is connected to the NGC with Non-standalone E-UTRA (similar role as SeNB in TS 36.300 [12]). The E-UTRA user plane connection to the NGC goes via the gNB (Option 4) or directly (Option 4a). For the Xn interface between eLTE eNB and gNB, the procedures and protocols should be newly designed.

In Option 7/7a/7x, Dual Connectivity can also be achieved, in which the eLTE eNB (similar role as MeNB in TS 36.300 [12]) is connected to the NGC with Non-standalone NR (similar role as SeNB in TS 36.300 [12]). The NR user plane connection to the NGC goes via the eLTE eNB (Option 7) or directly (Option 7a). For the Xn interface between eLTE eNB and gNB, the procedures and protocols should also be newly designed.

NOTE: The procedures and protocols over Xn interface for Option 7/7a should be the same as Option 4/4a.

10.1.2 Option 3/3a/3x

10.1.2.1 General principles for Xx interface

The interface allowing to interconnect EPC connected gNB and LTE eNB, is referred to as the Xx interface. The general principles for the specification of the Xx interface are as follows:

- the Xx interface shall be open;
- the Xx interface shall support the exchange of signalling information between the endpoints, in addition the interface shall support data forwarding to the respective endpoints;
- from a logical standpoint, the Xx is a point-to-point interface between the endpoints. A point-to-point logical interface should be feasible even in the absence of a physical direct connection between the endpoints.
- the Xx interface shall support control plane and user plane separation;
- the Xx interface shall separate Radio Network Layer and Transport Network Layer;
- the Xx interface shall be future proof to fulfil different new requirements, support new services and new functions;
- the Xx interface shall support LTE based Dual Connectivity operation where LTE eNB is MeNB;

- the Xx interface shall support flow control functions;
- the Xx interface does not support handover preparation functions.

10.1.2.2 Architectural aspects

DC specified in TS 36.300 [12] is applied as the baseline for tight interworking between NR and E-UTRA in this option. Radio Protocol Architecture for the User Plane is defined in Figure 10.1.2.2-1 for Option 3 (split bearer) and Option 3a (SCG bearer). The Radio Protocol Architecture for option 3x (SCG split bearer) is discussed in clause 10.1.2.4. LTE eNB and gNB are assumed to have the role similar to MeNB (Master eNB) and SeNB (Secondary eNB) specified in TS 36.300 [12], respectively. NR PDCP is one of the NR sub-layers to handle SDUs of the S1-U interface into different DRBs according to the QoS information associated with the SDU.

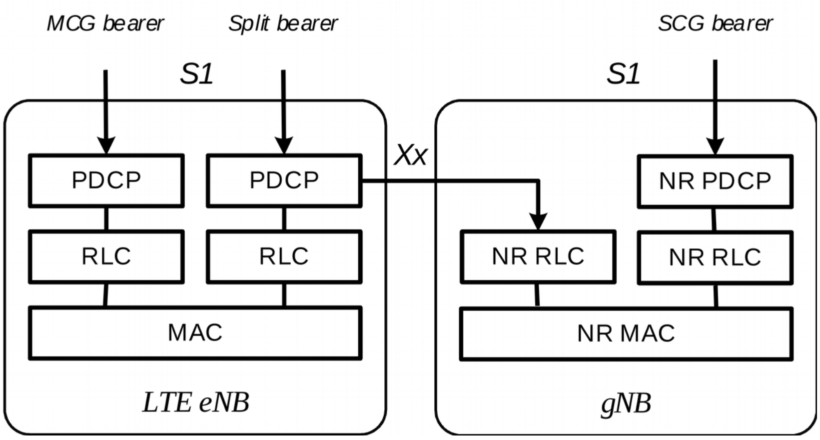


Figure 10.1.2.2-1: Radio Protocol Architecture for split bearer and SCG bearer in Option 3/3a

Network interface configurations can be defined in Figure 10.1.2.2-2 and Figure 10.1.2.2-3.

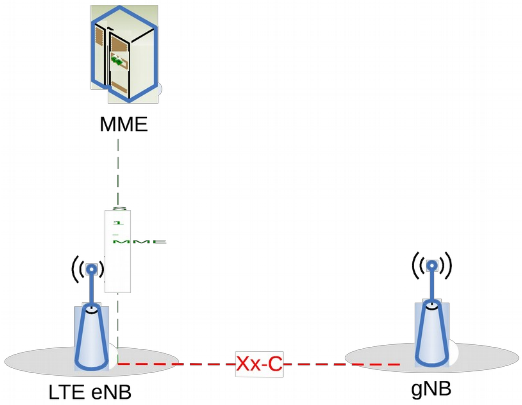


Figure 10.1.2.2-2: C-Plane connectivity for Option 3/3a/3x

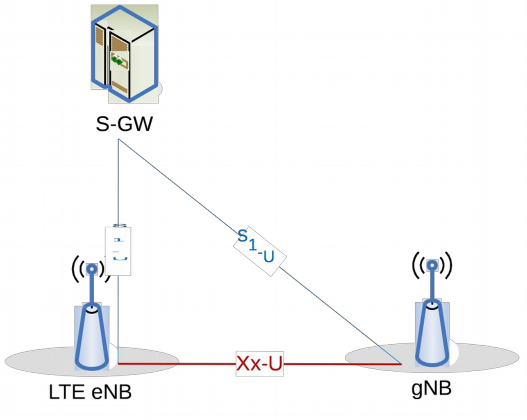


Figure 10.1.2.2-3: U-Plane connectivity for Option 3/3a/3x

10.1.2.3 Procedural aspects

The procedures defined under section 10.1.2.8 (Dual Connectivity operation) in TS 36.300 [12] listed below apply. In this list, LTE eNB and gNB connected via Xx are considered to have the role similar to MeNB (or eNB without DC) and SeNB, respectively. In this context, the Xx-U has all the functionality of X2-U for LTE DC operation. There are no impacts to S1 procedure foreseen with the support of Option 3/3a/3x. The DC procedures defined in TS 36.423 [13] are used as baseline.

- SeNB Addition
- SeNB Modification (MeNB initiated SeNB Modification)
- SeNB Modification (SeNB initiated SeNB Modification)
- Intra-MeNB handover involving SCG change
- SeNB Release (MeNB initiated SeNB Release)
- SeNB Release (SeNB initiated SeNB Release)
- Change of SeNB
- MeNB to eNB Change
- SCG change
- eNB to MeNB change
- Inter-MeNB handover without SeNB change

NOTE: Support of hybrid HeNB as the SeNB is not justified in RAN3.

Additional DC functionalities required for option 3x are discussed in clause 10.1.2.4.

SeNB Addition

- To support UE capability coordination, the MeNB would need to signal NR UE capabilities in the SENB ADDITION REQUEST message.
- In the SENB ADDITION REQUEST ACKNOWLEDGE, the SeNB indicates its choice of the NR configuration over Xx.

Existing procedure defined in TS 36.300 [12] can be reused without any procedural change, but with minor update to adapt the message information to NR.

SeNB Modification (SeNB/MeNB initiated SeNB Modification)

Some enhancements to the SENB MODIFICATION REQUEST, SENB MODIFICATION REQUEST ACKNOWLEDGE, SENB MODIFICATION REQUIRED and SENB MODIFICATION CONFIRM messages may be needed e.g. due to use cases and differences between E-UTRA and NR technologies.

Existing procedures defined in TS 36.300 [12] can be reused without any procedural change, but with minor update to adapt the message information to NR.

Intra-MeNB handover involving SCG change, Inter-MeNB handover without SeNB change

These cases involve intra-/inter-MeNB handover where the NR part may or may not change. Any procedural changes to this part should be similar as considered in SeNB Addition, SeNB Modification, Change of SeNB and SCG change procedures.

Existing procedures defined in TS 36.300 [12] can be reused without any procedural change.

MeNB/SeNB initiated SeNB Release

Either MeNB or SeNB should be able to release the connection between eNB and gNB. No procedural changes are foreseen.

Existing procedures defined in TS 36.300 [12] can be reused without any procedural change.

Change of SeNB

Since this is essentially just SeNB Release + SeNB Addition procedures, any procedural changes should be covered by the component procedures.

Existing procedures defined in TS 36.300 [12] can be reused without any procedural change.

MeNB to eNB Change, eNB to MeNB change

These procedures allow using a handover between MeNB and eNB involving SeNB Addition and SeNB Release procedures. As also these procedures are using the components of SeNB Addition and SeNB Release, no procedural changes are foreseen.

Existing procedures defined in TS 36.300 [12] can be reused without any procedural change.

SCG change

In E-UTRAN, SCG change is used whenever synchronous reconfiguration is required for the SeNB. With the tight interworking between E-UTRA and NR, similar baseline can be assumed and no procedural changes are foreseen.

Existing procedures defined in TS 36.300 [12] can be reused without any procedural change.

10.1.2.4 SCG split bearer

10.1.2.4.1 General

In order to support SCG split bearer, another deployment option needs to be supported. In option 3x shown in Figure 10.1.2.4.1-1, the solid line shown between LTE eNB and gNB is used for U-plane data transmission terminated at the gNB, i.e. S1-U data from EPC is split at the gNB.

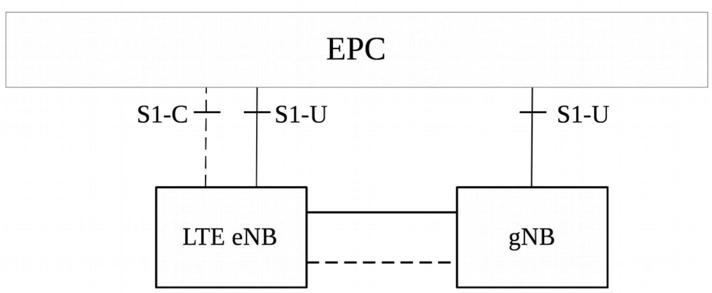


Figure 10.1.2.4.1-1: Option 3x

In Option 3x, S1-MME is still terminated at LTE eNB while U-plane is split in the gNB as described in Figure 10.1.2.2-1 and Figure 10.1.2.2-2. Radio Protocol Architecture for the User Plane can be defined in Figure 10.1.2.4.1-2 for SCG split bearer.

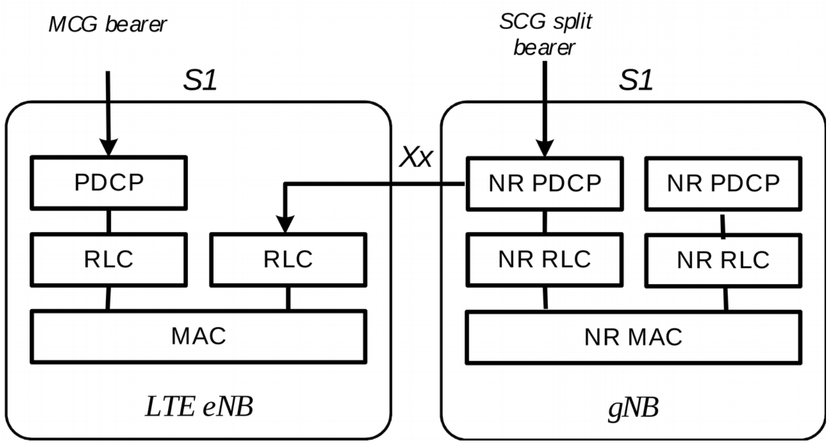


Figure 10.1.2.4.1-2: Radio Protocol Architecture for SCG split bearer in Option 3x

10.1.2.4.2 Evaluation on SCG split bearer

The following evaluation can be considered for the analysis of this bearer type:

- Signalling between Master node and Secondary node
- The backhaul of Xx interface
  - The Xx interface has to offer sufficient capacity to cope with LTE bitrates.
- Signalling to CN due to mobility in/out of Secondary node coverage

The following evaluation can be considered for the analysis of this bearer type:

Signalling between Master node and Secondary node

The following additional functions need to be supported on top of LTE Dual Connectivity specified in TS 36.300 [12].

- LTE eNB (MeNB) provides its tunnel endpoint information in SENB ADDITION REQUEST or SENB MODIFICATION REQUEST message to receive DL data transmission from the gNB;
- gNB (SeNB) provides its tunnel endpoint information in SENB ADDITION REQUEST ACKNOWLEDGE or SENB MODIFICATION REQUEST ACKNOWLEDGE message to receive UL data transmission from the LTE eNB;
- Flow control procedure needs to be applied to the other direction, i.e. LTE eNB signals DL DATA DELIVERY STATUS to the gNB, and the gNB signals DL USER DATA to the LTE eNB.

There is no difference on the number of signalling messages between SCG bearer and SCG split bearer with some enhancements to LTE DC specification.

10.1.2.5 QoS aspect

For this option, the LTE QoS framework defined in TS 36.300 [12] applies:

- E-RAB (and concerning S1-U bearer) is established between EPC and gNB for SCG bearer option and SCG split bearer option
- Xx-U is established between LTE eNB and gNB for split bearer option and SCG split bearer option

- DRB is established between gNB and UE according to SCG bearer option or split bearer option or SCG spilt bearer option

10.1.2.6 General principles for S1 interface

It is concluded that from a functional point of view the S1 interface is able to support deployment of Option 3/3a/3x.

10.1.2.7 Xx interface protocols

Since Option 3/3a/3x connects with EPC via S1, it is essential to adopt X2-C and X2-U as Xx interface if there are no obstacles.

One possible consideration is whether Global eNB ID can be reused for gNB or not. However, this can be achieved by e.g., gNB using it or converting to it. The details can be specified during normative phase. Therefore, X2 interface protocols, i.e. X2AP and X2 U-Plane protocol are baseline for Xx interface

10.1.2.8 TNL address discovery for Xx interface establishment

TNL address discovery solutions for Xx interface establishment for the deployment options 3/3a/3x shall be considered and concluded during the normative work.

10.1.3 Option 4/4a

10.1.3.1 Architectural aspects

Radio Protocol Architecture for the User Plane is shown below.

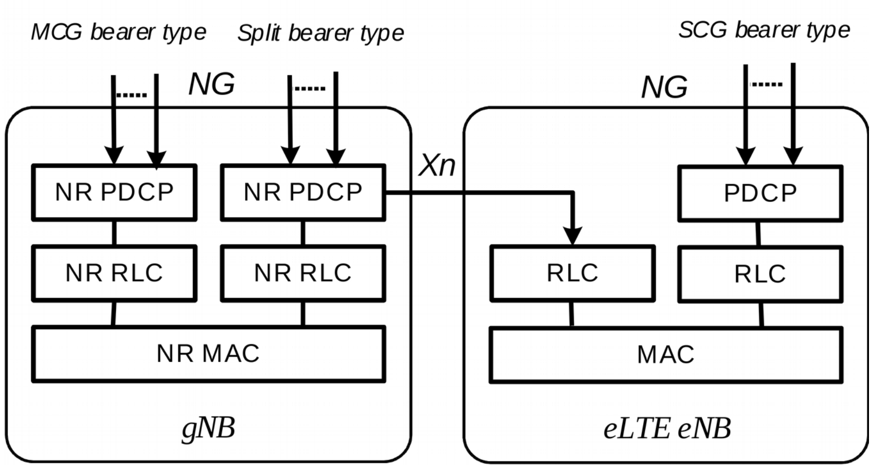


Figure 10.1.3.1-1: Radio Protocol Architecture for split bearer and SCG bearer in Option 4/4a

Network interface configurations can be defined in Figure 10.1.3.1-2 and Figure 10.1.3.1-3.

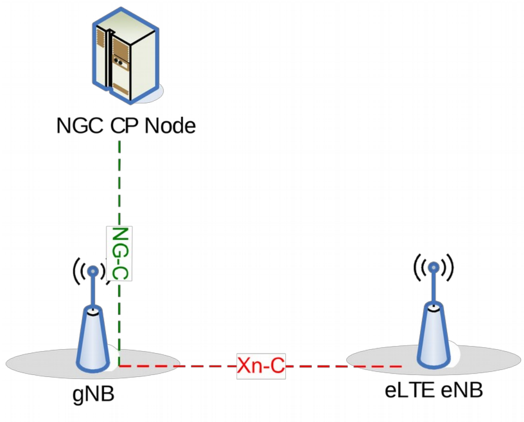


Figure 10.1.3.1-2: C-Plane connectivity for Option 4/4a

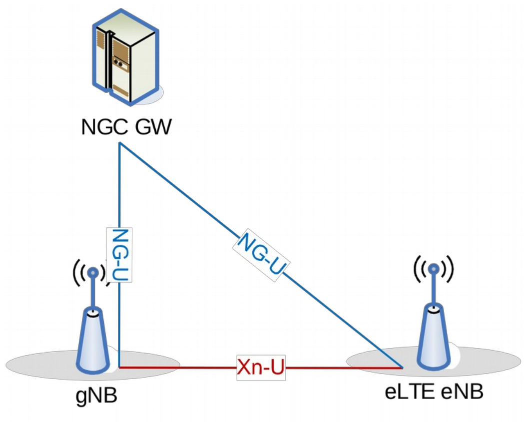


Figure 10.1.3.1-3: U-Plane connectivity for Option 4/4a

10.1.3.2 Procedural aspects

The procedures defined under section 10.1.2.8 (Dual Connectivity operation) in TS 36.300 [12] can be a reference for defining the new Xn based procedures, in which gNB and eLTE eNB connected via Xn are considered to have the role similar to MeNB and SeNB, respectively. The NR new QoS model defined in section 9 will be applied for the new Xn based procedures.

NOTE: The details on the procedures can be defined in the normative phase.

10.1.4 Option 7/7a/7x

10.1.4.1 Architectural aspects

Radio Protocol Architecture for the User Plane is shown below.

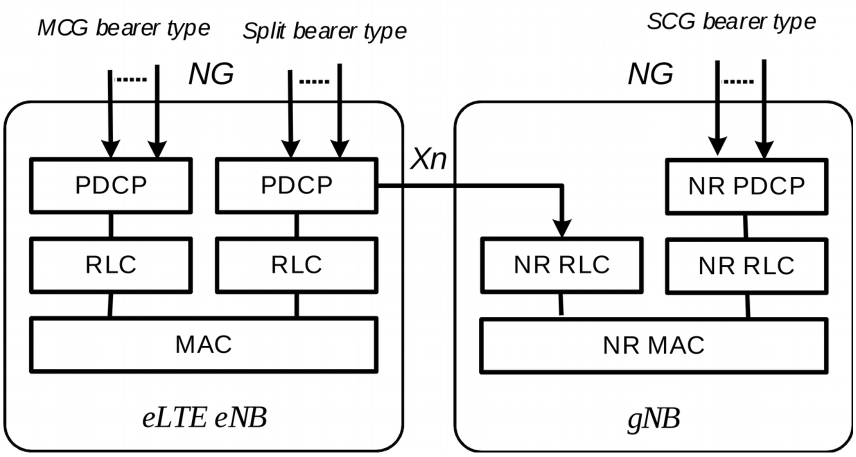


Figure 10.1.4.1-1: Radio Protocol Architecture for split bearer and SCG bearer in Option 7/7a

Network interface configurations can be defined in Figure 10.1.4.1-2 and Figure 10.1.4.1-3.

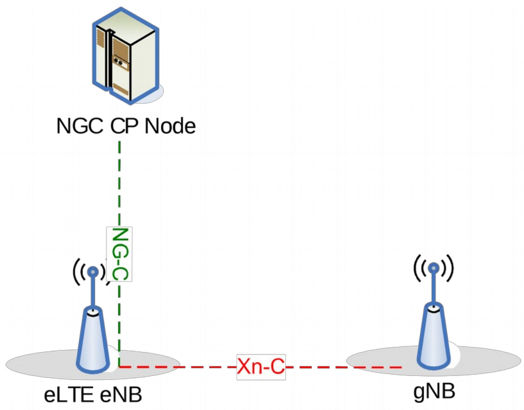


Figure 10.1.4.1-2: C-Plane connectivity for Option 7/7a/7x

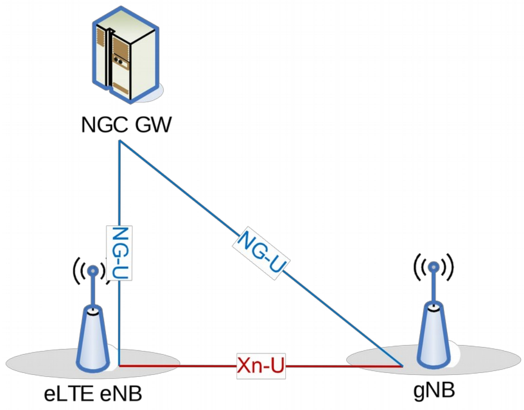


Figure 10.1.4.1-3: U-Plane connectivity for Option 7/7a/7x

10.1.4.2 Procedural aspects

The procedures defined under section 10.1.2.8 (Dual Connectivity operation) in TS 36.300 [12] can be a reference for defining the new Xn based procedures, in which eLTE eNB and gNB connected via Xn are considered to have the role similar to MeNB and SeNB, respectively. The NR new QoS model defined in section 9 will be applied for the new Xn based procedures.

NOTE: The details on the procedures can be defined in the normative phase.

10.1.4.3 SCG split bearer

In order to support SCG split bearer, another deployment option needs to be supported. In option 7x shown in Figure 10.1.4.3-1, the solid line shown between eLTE eNB and gNB is used for U-plane data transmission terminated at the gNB, i.e. NG-U data from NGC is split at the gNB.

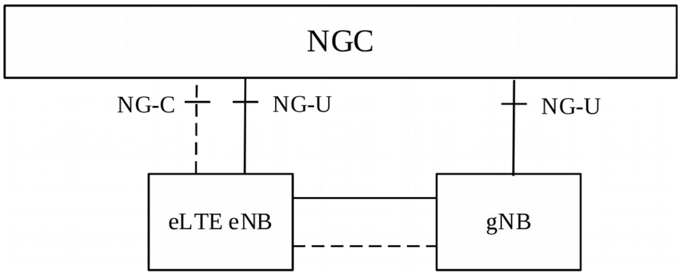


Figure 10.1.4.3-1: Option 7x

In option 7x, NGC C-plane is terminated at eLTE eNB while U-plane is split in the gNB as described in Figure 10.1.4.1-2 and Figure 10.1.4.1-3. Radio Protocol Architecture for the User Plane can be defined in Figure 10.1.4.3-2 for SCG split bearer.

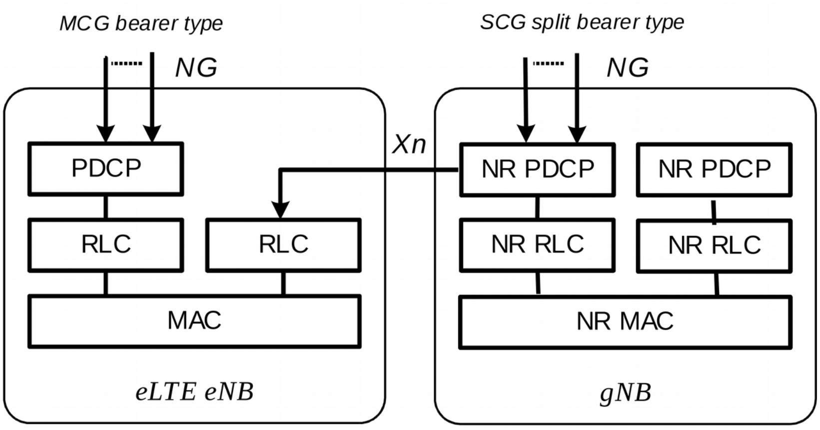


Figure 10.1.4.3-2: Radio Protocol Architecture for SCG split bearer in Option 7x

NOTE: Further details will be considered during normative work by taking final RAN2 agreement into account.

10.2 New RAN operation

10.2.1 Intra-system Mobility

10.2.1.1 General

Principle: The LTE X2 handover procedure as in TS 36.300 Figure 10.1.2.1.1-1: Intra-MME/Serving Gateway HO is taken as a baseline for intra-system mobility i.e. intra RAT (gNB <-> gNB; eLTE eNB <-> eLTE eNB) and inter-RAT (eLTE-eNB <-> gNB).

Principle: The NG-based handover shall be supported and LTE S1-based handover procedure as in TS 23.401 Figure 5.5.1.2.2-1 is taken as a baseline.

10.2.1.2 Intra-system Intra-RAT mobility

10.2.1.2.1 Scenario: Inter gNB mobility

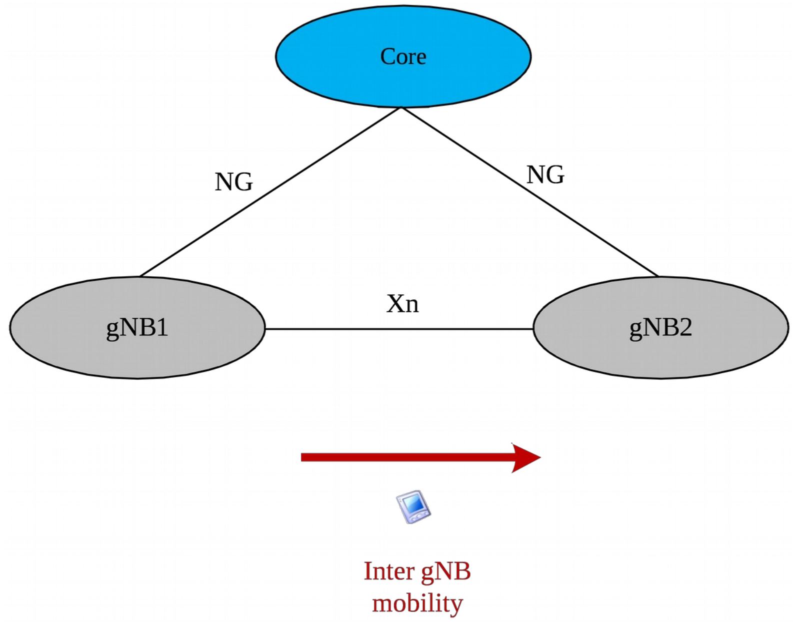


Figure 10.2.1.2.1-1: Inter gNB mobility scenario

NOTE 1: The details of intra-gNB mobility scenario can be further considered during normative phase.

NOTE 2: The intra 5G intra-RAT handover is normally based on Xn-based handover. However NG-based handover may be used if Xn interface is not available and in some cases where Xn is available.

10.2.1.2.2 Inter-gNB Handover Variant with in-band Path Switch

The following call flow presents a variant of the inter-gNB handover derived from LTE X2-handover where the path switch is triggered over the user plane for latency optimization. This variant is shown below in an example where two PDU sessions are active towards different UPGWs:

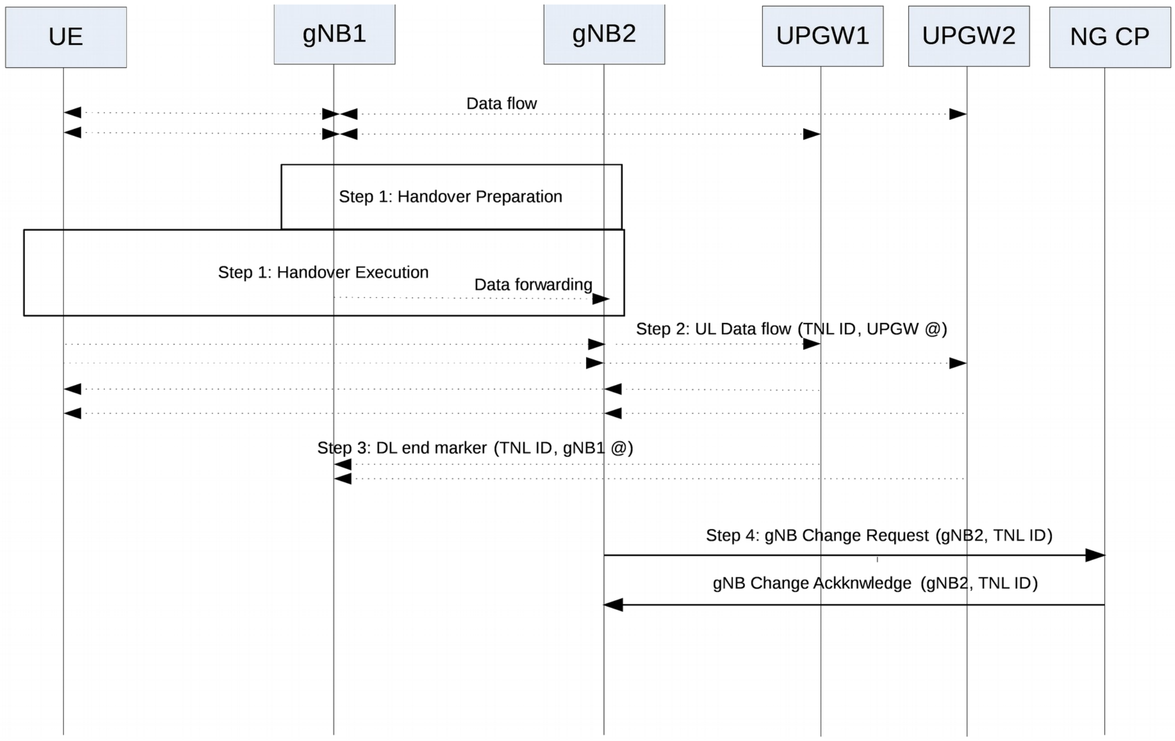


Figure 10.2.1.2.2-1: Intra-new RAN Handover using in-band Path Switch over NG-U

Starting point: The TNL ID is the identifier in the tunnel header over NG-U carrying all the packets for a given UE and a given PDU session and the UPGW @ is the address of the UPGW which has been assigned at PDU session creation.

- At step 1 during handover preparation gNB2 receives the context information for the UE and the sessions including the tunnel information (UPGW @, TNL ID) for each relocated session then performs handover execution.
- At step 2, at the end of handover execution gNB2 sends over NG-U for each session a special packet including the TNL ID to inform the UPGW that it now handles the traffic for that session. The UPGW identifies from the TNL ID in the tunnel header the UE and associated PDU session. Reception of this first packet automatically switches the DL path between the gNB2 and the UPGW for this (UE, PDU session).
- At step 3, The UPGW sends an End Marker packet to gNB1. Receiving the End Marker packet in gNB1 including TNL ID in the tunnel header automatically releases the tunnel between the gNB1 and the UPGW for this (UE, PDU session).
- At step 4 the gNB may send a gNB Change Request message to inform NG CP of the new gNB ID.

This above described procedure is applicable to both eLTE eNB and gNB.

10.2.1.3 Intra-system Inter-RAT handover with E-UTRA

There are two intra-system inter-RAT handover with E-UTRA scenarios as shown in Figure 10.2.1.3-1.

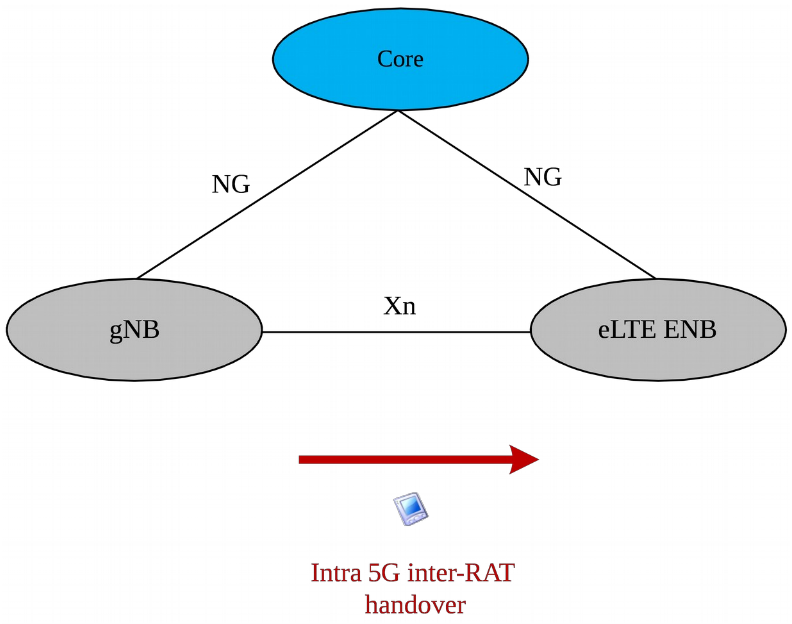


Figure 10.2.1.3-1: Scenarios for Intra-system inter-RAT handover with E-UTRA

NOTE 1: Intra 5G system inter-RAT handover with NGC relocation is pending on SA2.

NOTE 2: The Intra 5G inter RAT handover is normally based on Xn-based handover. However NG-based handover may be used if Xn interface is not available and in some cases where Xn is available.

10.2.1.4 Cell Reselection, Release and Redirection

This function provides for support of cell reselection, release and redirection between the radio accesses supported within the New RAN..

NOTE: The need and details of Cell Reselection, Release and Redirection can be further considered during normative phase.

10.2.2 Inter-system Mobility

10.2.2.1 Inter-system Inter-RAT handover with E-UTRA

Inter-system inter-RAT handover with E-UTRA scenario is shown in Figure 10.2.2.1-1.

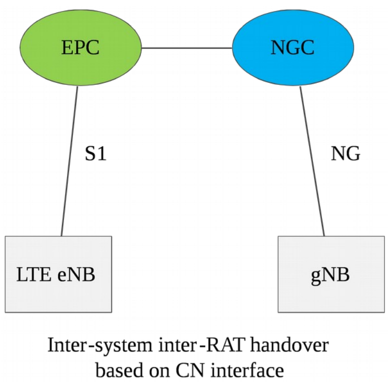


Figure 10.2.2.1-1: Inter-system inter-RAT handover with E-UTRA

NOTE: An interface between EPC and NGC, namely, NGx interface, may be available.

Figure 10.2.2.1-2 depicts a procedure for handover from the NG System to the EPS. The procedure for the opposite direction is similar and is omitted for brevity.

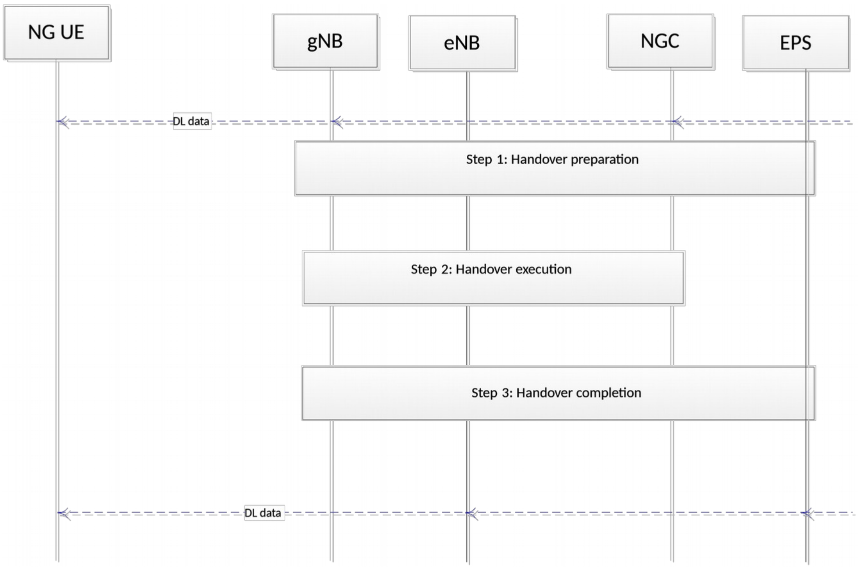


Figure 10.2.2.1-2: Procedure for handover from the NG System to the EPS

- At step 1 the gNB decides that the UE should be handed over to the E-UTRAN and triggers the handover preparation procedure to the E-UTRAN via NGC and EPC.
- At step 2 the gNB commands the UE to hand over to E-UTRAN after receiving the successful response from the target E-UTRAN through EPC and NGC.
- At step 3 the E-UTRAN notifies to the MME that the UE is handed over to the E-UTRAN. The notification message is also transmitted to the gNB through NGC in order to trigger resource release.

10.2.2.2 Inter-system intra-RAT handover

One inter-system intra-RAT handover scenario is shown in Figure 10.2.2.2-1.

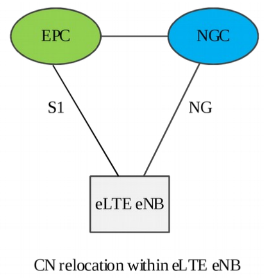


Figure 10.2.2.2-1: Inter-system Intra-RAT handover scenario

A CN relocation can be performed for the scenario of an eLTE eNB connected to both EPC and NGC. The details for this will be discussed in the normative phase.

10.2.3 PDU Session Management

10.2.3.1 Session Setup

The NG Session Setup is initiated by the NG CP over the NG-C interface following a PDU session request from the UE in order to exchange respective information between the NG Core and the gNB resulting in the creation of a context for the PDU session and related resources in the gNB.

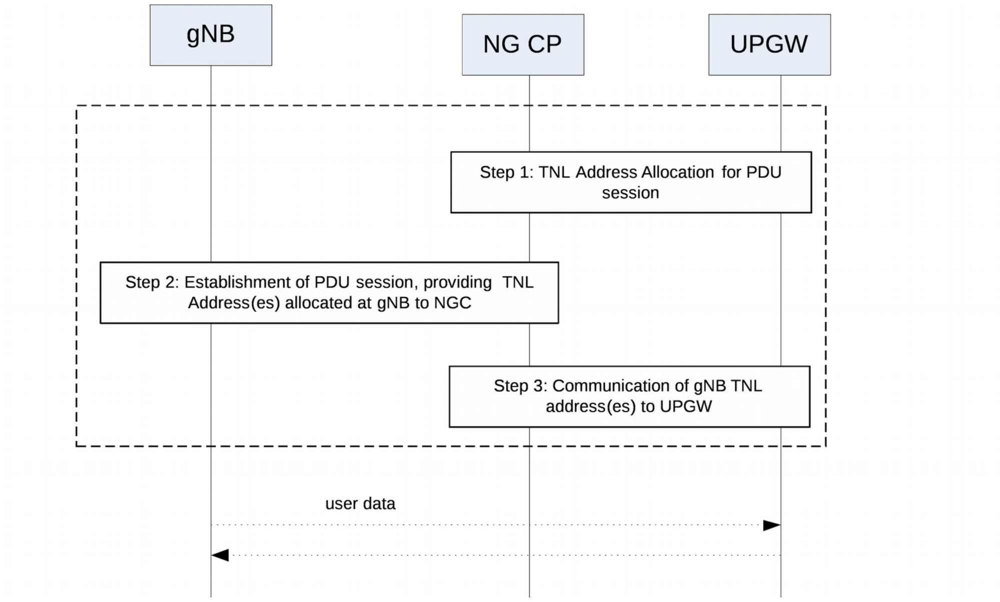


Figure 10.2.3.1-1: PDU Session Setup over NG

- At step 1 following a UE request for setting up a PDU session the NG CP determines a UPGW address and a TNL address corresponding to this UE and PDU session
- At step 2 the NG CP establishes the PDU Session providing UPGW address, TNL address and other context information e.g. QoS information.
- At step 3 the gNB TNL address, as received in step 2 is sent to the UPGW.

The TNL address(es) are then subsequently used by the gNB and the UPGW in the tunnel header of all packets exchanged over the NG-U interface.

NOTE: The above described procedure applies to both eLTE eNB and gNB.

10.2.3.2 Session Modification

The NG Session Modification is initiated by the NG CP over the NG-C interface in order to modify the PDU Session and the context for the PDU session in the gNB.

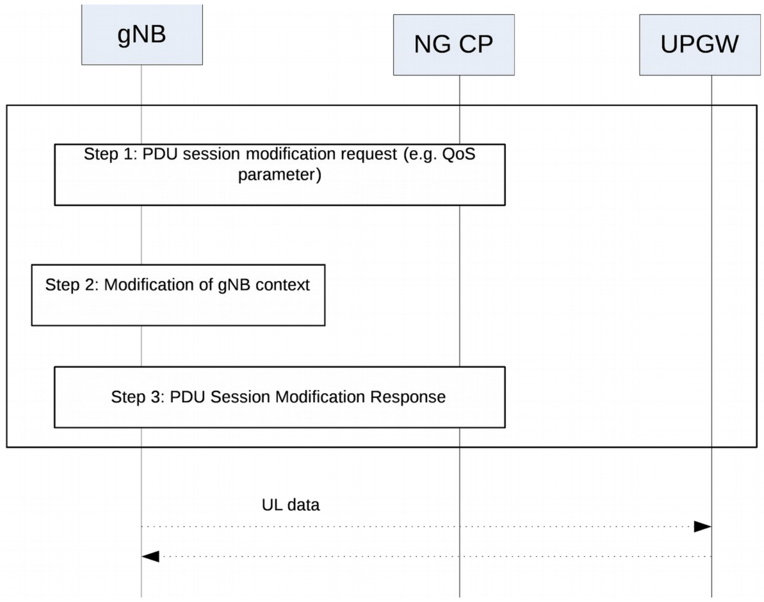


Figure 10.2.3-2: PDU Session Modification over NG

- At step 1 the NG CP initiates the PDU Session Modification procedure containing the identifier of the session to be modified and the session parameters to be modified (e.g. QoS information).

NOTE 1: The detailed parameters for Session Modification can be further considered during normative phase.

- At step 2 the gNB modifies the context and takes into account the updated parameters.

- At step 3 the gNB indicates to the NG CP the successful completion of the PDU Session Modification.

NOTE 2: The above describted procedure applies to both eLTE eNB and gNB.

10.2.3.3 Session Release

The NG Session Release is initiated by the NG CP over the NG-C interface in order to release the PDU Session and the context for the PDU session in the gNB.

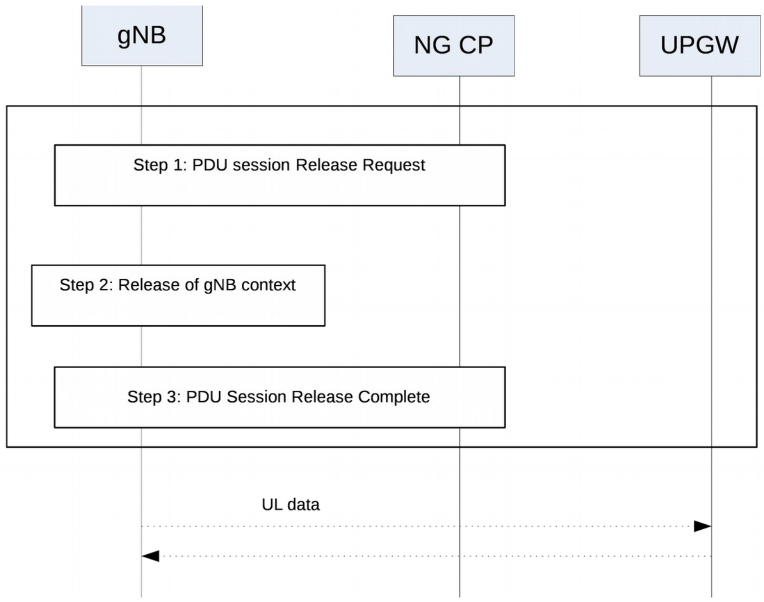


Figure 10.2.3-3: PDU Session Release over NG

- At step 1 the NG CP initiates the PDU Session Release procedure containing the identifier of the session to be released.

- At step 2 the gNB releases the context and triggers the associated reconfiguration towards the UE
- At step 3 the gNB indicates to the NG CP the successful completion of the PDU Session Release.

NOTE 1: The above describted procedure applies to both eLTE eNB and gNB.

NOTE 2: The need and details of a PDU Session Release procedure triggered from gNB (respectively eLTE eNB) can be further considered during normative phase.

10.2.4 Initial UE Access

Following a UE connection request, this is initiated by the New RAN to establish a signalling connection between New RAN and NGC in case of an Idle-to Active transition. The first message from the NGC, following the INITIAL UE MESSAGE, completes the establishment of a signalling connection. With the first reply message the NGC may initiate a UE context setup.

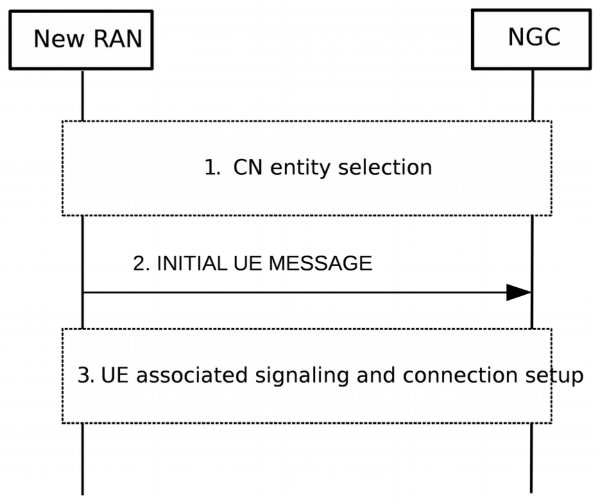


Figure 10.2.4-1: Initial UE Access over NG

- At step 1 following the reception of the first UL NAS message to be forwarded to NGC, the selection of CN entity is performed.
- At step 2, the New RAN uses the INITIAL UE MESSAGE to deliver the NAS PDU to the CN entity selected in the previous step
- At step 3, the NGC may complete the UE associated signaling connection setup. The New RAN configures the UE accordingly.

NOTE: The details of the selection of the CN entity and whether INITIAL UE MESSAGE is used in step 2 can be further considered during normative phase.

10.2.5 Intra-NR dual connectivity

Unlike NR/LTE Tight Interworking between different RATs, the Intra-NR DC is assumed to be more like legacy LTE DC from RAN3 procedure viewpoint. The main principles from legacy LTE DC can be inherited by Intra-NR DC with potential enhancement. The details shall be discussed further in normative phase.

11 RAN logical architecture for NR

11.1 Functional split between central and distributed unit

11.1.1 General description of split options

In the study item for a new radio access technology, 3GPP is expected to study different functional splits between central and distributed units. E-UTRA protocol stack is taken as a basis for further discussion, with the understanding that the conclusions may need to be revisited, once RAN2 defines the protocol stack for NR. The following functional splits between central and distributed unit are possible, as illustrated in Figure 11.1.1-1.

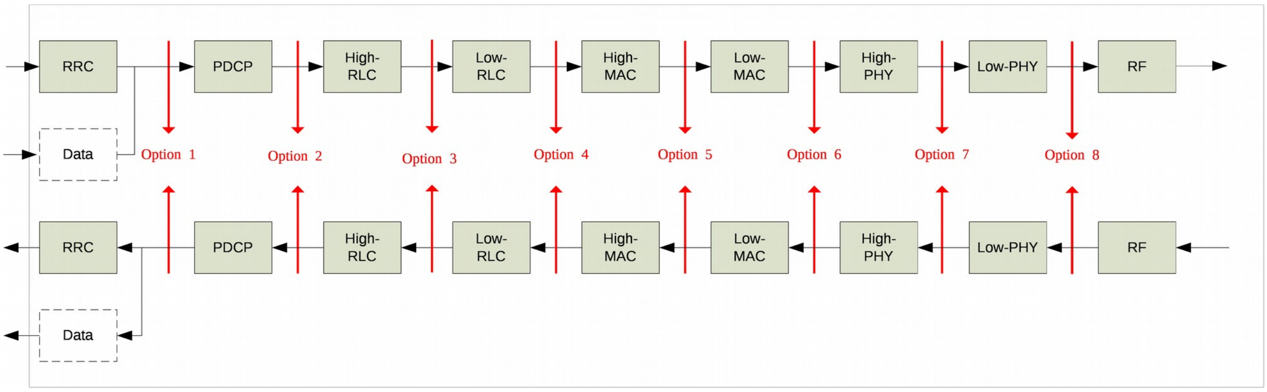


Figure 11.1.1-1: Function Split between central and distributed unit

Option 1 (1A-like split)

- The function split in this option is similar as 1A architecture in DC. RRC is in the central unit. PDCP, RLC, MAC, physical layer and RF are in the distributed unit.

Option 2 (3C-like split)

- The function split in this option is similar as 3C architecture in DC. RRC, PDCP are in the central unit. RLC, MAC, physical layer and RF are in the distributed unit.

Option 3 (intra RLC split)

- Low RLC (partial function of RLC), MAC, physical layer and RF are in distributed unit. PDCP and high RLC (the other partial function of RLC) are in the central unit.

Option 4 (RLC-MAC split)

- MAC, physical layer and RF are in distributed unit. PDCP and RLC are in the central unit.

Option 5 (intra MAC split)

- RF, physical layer and some part the MAC layer (e.g. HARQ) are in the distributed unit. Upper layer is in the central unit.

Option 6 (MAC-PHY split)

- Physical layer and RF are in the distributed unit. Upper layers are in the central unit.

Option 7 (intra PHY split)

- Part of physical layer function and RF are in the distributed unit. Upper layers are in the central unit.

Option 8 (PHY-RF split)

- RF functionality is in the distributed unit and upper layer are in the central unit.

NOTE: The options represented consist of a non-exhaustive list. The assumption on protocols and functions definition was based on what was available during the study phase.

Flexible functional split

Some of the benefits of an architecture with the deployment flexibility to split and move NR functions between central and distributed units are below:

- Flexible HW implementations allows scalable cost effective solutions
- A split architecture (between central and distributed units) allows for coordination for performance features, load management, real-time performance optimization, and enables NFV/SDN
- Configurable functional splits enables adaptation to various use cases, such as variable latency on transport

The choice of how to split NR functions in the architecture depends on some factors related to radio network deployment scenarios, constraints and intended supported services. Some examples of such factors are:

- Need to support specific QoS settings per offered services (e.g. low latency, high throughput)
- Need to support specific user density and load demand per given geographical area (which may influence the level of RAN coordination)
- Need to be able to function with transport networks with different performance levels, from ideal to non-ideal

The NR design should support the flexibility to move RAN functions between the central unit and distributed unit depending on the factors above, and should be studied.

The support of cascaded functional splits with different split options should not be precluded. A cascaded function split is a deployment with e.g. one intermediate CU and/or DU between a CU and DU pair.

11.1.2 Detailed Description of Candidate Split Options and Justification

11.1.2.1 Option 1 (RRC/PDCP, 1A-like split)

Description: In this split option, RRC is in the central unit. PDCP, RLC, MAC, physical layer and RF are in the distributed unit, thus the entire user plane is in the distributed unit.

Benefits and Justification:

- This option allows a separate U-plane while having a centralised RRC/RRM.
- It may in some circumstances provide benefits in handling some edge computing or low latency use cases where the user data needs to be located close to the transmission point.

Cons:

- Because of the separation of RRC and PDCP, securing the interface in practical deployments may or may not affect performance of this option.
- It needs to be clarified whether and how this option can support aggregation based on alternative 3C.

11.1.2.2 Option 2 (PDCP/RLC split)

Option 2 may be a base on an X2-like design due to similarity on U-plane but some functionality may be different e.g. C-plane since some new procedures may be needed.

Option 2-1 Split U-plane only (3C like split)

Description: In this split option, RRC, PDCP are in the central unit. RLC, MAC, physical layer and RF are in the distributed unit.

Benefits and Justification:

- This option will allow traffic aggregation from NR and E-UTRA transmission points to be centralized. Additionally, it can facilitate the management of traffic load between NR and E-UTRA transmission points.

- Fundamentals for achieving a PDCP-RLC split have already been standardized for LTE Dual Connectivity, alternative 3C. Therefore this split option should be the most straightforward option to standardize and the incremental effort required to standardize it should be relatively small.

NOTE 1: U-plane aspect was justified in the study phase. C-plane aspect was not addressed in the study phase.

- The alignment between LTE-NR tight interworking and functional split may be beneficial at least in user-plane, considering migration.

NOTE 2: An enhancement for the fast-centralized retransmission of lost RLC PDUs in this option was identified, but the solution details were not discussed in this study.

Option 2-2: In this split option, RRC, PDCP are in the central unit. RLC, MAC, physical layer and RF are in the distributed unit. In addition, this option can be achieved by separating the RRC and PDCP for the CP stack and the PDCP for the UP stack into different central entities.

Benefits and Justification:

- This option will allow traffic aggregation from NR and E-UTRA transmission points to be centralized. Additionally, it can facilitate the management of traffic load between NR and E-UTRA transmission points.

- This option enables centralization of the PDCP layer, which may be predominantly affected by UP process and may scale with UP traffic load.

- This option allows a separate U-plane while having a centralised RRC/RRM.

Cons:

- Coordination of security configurations between different PDCP instances for Option 2-2 needs to be ensured.

11.1.2.3 Option 3 (High RLC/Low RLC Split)

Two approaches based on Real-time/Non Real-time function split are as follows:

Option 3-1 Split based on ARQ

Description:

- Low RLC may be composed of segmentation functions;

- High RLC may be composed of ARQ and other RLCfunctions;

This option splits the RLC sublayer into High RLC and Low RLC sublayers such that for RLC Acknowledge Mode operation, all RLC functions may be performed at the High RLC sublayer residing in the central unit, while the segmentation may be performed at the Low RLC sublayer residing in the distributed unit. Here, High RLC segments RLC PDU based on the status reports while Low RLC segments RLC PDU into the available MAC PDU resources.

Benefits and Justification:

- This option will allow traffic aggregation from NR and E-UTRA transmission points to be centralized. Additionally, it can facilitate the management of traffic load between NR and E-UTRA transmission points.

- This split option may also have better flow control across the split.

- Centralization gains: ARQ located in the CU may provide centralization or pooling gains.

- The failure over transport network may also be recovered using the end-to-end ARQ mechanism at CU. This may provide protection for critical data and C-plane signaling.

- DUs without functions of RLC may handle more connected mode UEs as there is no RLC state information stored and hence no need for UE context.

- This option may provide an efficient means for implementing integrated access and backhaul to support self-backhauled NR TRPs.

NOTE: As part of the analysis with RAN2, there is no consensus on the following benefits and drawbacks from RAN2 point of view.

Benefits and Justification:

- This option may have the advantage of being more robust under non-ideal transport conditions because the ARQ and packet ordering is performed at the central unit.
- It may reduce processing and buffer requirements in DU due to absence of ARQ protocol
- Could be used over multiple radio legs of different DUs for higher reliability (U-Plane and C-Plane) [Pending to multi-connectivity]
- This option may provide an efficient way for implementing intra-gNB RAN-based mobility.

Cons:

- Comparatively, the split is more latency sensitive than the split with ARQ in DU, since re-transmissions are susceptible to transport network latency over a split transport network.

Option 3-2 Split based on TX RLC and RX RLC

Description:

- Low RLC may be composed of transmitting TM RLC entity, transmitting UM RLC entity, a transmitting side of AM and the routing function of a receiving side of AM, which are related with downlink transmission.
- High RLC may be composed of receiving TM RLC entity, receiving UM RLC entity and a receiving side of AM except the routing function and reception of RLC status report, which are related with uplink transmission.

Transmitting: Tx RLC receives RLC SDU from PDCP and transmits these packets under the format indicator of MAC.As soon as RLC receives the PDU request from MAC, RLC must assemble the MAC SDU under the format indicator of MAC and submit the MAC SDU to MAC. In order to adapt the transport network between CU and DU, it is critical that Tx RLC is placed in DU.

Receiving: Routing receives RLC PDU from MAC and judges CONTROL PDU/DATA PDU, then submits DATA PDU to Rx RLC and CONTROL PDU to Tx RLC. When PDCP/RLC reestablishment procedure is triggered, placing Rx RLC in CU is critical in order to real-timely deliver data packets to PDCP.

Benefits and Justification:

Option3-2 not only is insensitive to the transmission network latency between CU and DU, but also uses interface format inherited from the legacy interfaces of PDCP-RLC and MAC-RLC. Some benefits of Option3-2 are as follows:

- This option will allow traffic aggregation from NR and E-UTRA transmission points to be centralized. Additionally, it can facilitate the management of traffic load between NR and E-UTRA transmission points.
- Flow control is in the CU and for that a buffer in the CU is needed. The TX buffer is placed in the DU, so that the flow controlled traffic from the CU can be buffered before being transmitted. Flow control can be done depending on fronthaul conditions
- As Rx RLC is placed in CU, there is no additional transmission delay of PDCP/RLC reestablishment procedure when submitting the RLC SDUs to PDCP
- This option does not induce any transport constraint, e.g. transport network congestion. MAC submits RLC PDUs as a whole packet to RLC rather than RLC sending RLC SDUs to PDCP.

Cons:

- Compared to the case where RLC is not split, STATUS PDU of AM Rx RLC may lead to additional time delay. Because STATUS PDU must be submitted through PDCP-Tx RLC interface from CU to DU before Tx RLC in DU transmits it over air interface, which may lead to additional transport delay.
- Due to performing flow control in the CU and RLC Tx in the DU two buffers are needed for transmission, one at the CU, which allows to flow control data submission to the RLC Tx, and one at the DU in order to perform RLC TX

11.1.2.4 Option 4 (RLC-MAC split)

Description: In this split option, RRC, PDCP and RLC are in the central unit.MAC, physical layer and RF are in the distributed unit.

Benefits and Justification: In the context of the LTE protocol stack a benefit is not foreseen for option 4. This might be revised with NR protocol stack knowledge.

11.1.2.5 Option 5 (intra MAC split)

Description:

Option 5 assumes the following distribution:

- RF, physical layer and lower part of the MAC layer (Low-MAC) in the Distributed Unit
- Higher part of the MAC layer (High-MAC), RLC and PDCP in the Central Unit

Therefore by splitting the MAC layer into 2 entities (e.g. High-MAC and Low-MAC), the services and functions provided by the MAC layer will be located in the Central Unit (CU), in the Distributed Unit (DU), or in both. An example of this distribution and its justification is given below.

In High-MAC sublayer:

The centralized scheduling in the High-MAC sublayer will be in charge of the control of multiple Low-MAC sublayers. It takes high-level centralized scheduling decision.

The inter-cell interference coordination in the High-MAC sublayer will be in charge of interference coordination methods such as JP/CS CoMP.

In Low-MAC sublayer:

Time critical functions in the Low-MAC sublayer include the functions with stringent delay requirements (e.g. HARQ) or the functions where performance is proportional to latency (e.g. radio channel and signal measurements from PHY, random access control). It reduces the delay requirements on the fronthaul interface.

Radio specific functions in the Low-MAC sublayer can for perform scheduling-related information processing and reporting. It can also measure/estimate the activities on the configured operations or the served UE’s statistics and report periodically or as requested to the High-MAC sublayer.

Depending on the different implementations of the intra-MAC functional split, the following pros and cons can be defined:

Benefits and Justification:

- This option will allow traffic aggregation from NR and E-UTRA transmission points to be centralized. Additionally, it can facilitate the management of traffic load between NR and E-UTRA transmission points.
- Reduce the bandwidth needed on fronthaul, dependent on the load of RAN-CN interface;
- Reducing latency requirement on fronthaul (if HARQ processing and cell-specific MAC functionalities are performed in the DU);
- Efficient interference management across multiple cells and enhanced scheduling technologies such as CoMP, CA, etc., with multi-cell view;

Cons:

- Complexity of the interface between CU and DU;
- Difficulty in defining scheduling operations over CU and DU;
- Scheduling decision between CU and DU will be subject to fronthaul delays, which can impact performances in case of non-ideal fronthaul and short TTI;
- Limitations for some CoMP schemes (e.g. UL JR).

11.1.2.6 Option 6 (MAC-PHY split)

Description: The MAC and upper layers are in the central unit (CU). PHY layer and RF are in the DU. The interface between the CU and DUs carries data, configuration, and scheduling-related information (e.g. MCS, Layer Mapping, Beamforming, Antenna Configuration, resource block allocation, etc.) and measurements.

Benefits and Justification:

- This option will allow traffic aggregation from NR and E-UTRA transmission points to be centralized. Additionally, it can facilitate the management of traffic load between NR and E-UTRA transmission points.
- This option is expected to reduce the fronthaul requirements in terms of throughput to the baseband bitrates as the payload for Option 6 is transport block bits.
- Joint Transmission is possible with this option as MAC is in CU.
- Centralized scheduling is possible for Option 6 as MAC is in CU.
- It allows resource pooling for layers including and above MAC.

Cons:

- This split may require subframe-level timing interactions between MAC layer in CU and PHY layers in DUs. Round trip fronthaul delay may affect HARQ timing and scheduling.

11.1.2.7 Option 7 (intra PHY split)

Description: Multiple realizations of this option are possible, including asymmetrical options which allow to obtain benefits of different sub-options for UL and DL independently (e.g. Option 7-1 is used in the UL and Option 7-2 is used in the DL). A compression technique may be able to reduce the required transport bandwidth between the DU and CU.

In the UL, FFT, and CP removal reside in the DU. Two sub-variants are described below. Remaining functions reside in the CU.

In the downlink, iFFT and CP addition reside in the DU. Three sub-variants are described below. The rest of the PHY resides in the CU.

Benefits and Justification (common among Option 7-1, 7-2 and 7-3):

- This option will allow traffic aggregation from NR and E-UTRA transmission points to be centralized. Additionally, it can facilitate the management of traffic load between NR and E-UTRA transmission points.
- These options are expected to reduce the fronthaul requirements in terms of throughput.
- Centralized scheduling is possible as MAC is in CU. e.g. CoMP
- Joint processing (both transmit and receive) is possible with these options as MAC is in CU.

Cons:

- This split may require subframe-level timing interactions between part of PHY layer in CU and part of PHY layer in DUs.

Option 7-1

Description:

In the UL, FFT, CP removal and possibly PRACH filtering functions reside in the DU, the rest of PHY functions reside in the CU. The details of the meaning of PRACH filtering were not discussed in the study phase.

In the DL, iFFT and CP addition functions reside in the DU, the rest of PHY functions reside in the CU.

Benefits and Justification:

- Allows the implementation of advanced receivers

Option 7-2

Description:

In the UL, FFT, CP removal, resource de-mapping and possibly pre-filtering functions reside in the DU, the rest of PHY functions reside in the CU. The details of the meaning of pre-filtering were not discussed in the study phase.

In the DL, iFFT, CP addition, resource mapping and precoding functions reside in the DU, the rest of PHY functions reside in the CU.

It is a requirement that both options allow the optimal use of advanced receivers. Whether or not these variants meets this requirement was not discussed in the study phase.

Option 7-3 (Only for DL)

Description:

Only the encoder resides in the CU, and the rest of PHY functions reside in the DU.

Benefits and Justification

- This option is expected to reduce the fronthaul requirements in terms of throughput to the baseband bitrates as the payload for Option 7-3 is encoded data.

11.1.2.8 Option 8 (PHY-RF split)

Option 8 allows to separate the RF and the PHY layer. This split permits centralisation of processes at all protocol layer levels, resulting in very tight coordination of the RAN. This allows efficient support of functions such as CoMP, MIMO, load balancing, mobility.

Benefits and Justification:

- This option will allow traffic aggregation from NR and E-UTRA transmission points to be centralized. Additionally, it can facilitate the management of traffic load between NR and E-UTRA transmission points.
- High levels of centralization and coordination across the whole protocol stack, which may enable a more efficient resource management and radio performance
- Separation between RF and PHY enables to isolate the RF components from updates to PHY, which may improve RF/PHY scalability
- Separation of RF and PHY allows reuse of the RF components to serve PHY layers of different radio access technologies (e.g. GSM, 3G, LTE)
- Separation of RF and PHY allows pooling of PHY resources, which may enable a more cost efficient dimensioning of the PHY layer
- Separation of RF and PHY allows operators to share RF components, which may reduce system and site costs

Cons:

- High requirements on fronthaul latency, which may cause constraints on network deployments with respect to network topology and available transport options
- High requirements on fronthaul bandwidth, which may imply higher resource consumption and costs in transport dimensioning (link capacity, equipment, etc)

11.1.2.9 Summary table

Summary on characteristics of different CU-DU split options is shown in Table 11.1.2.9-1.

Table 11.1.2.9-1 Summary on characteristics of different CU-DU split option

	Opt. 1	Opt. 2	Opt. 3-2	Opt. 3-1	Opt. 5	Opt. 6	Opt. 7-3 (only for DL)	Opt. 7-2	Opt. 7-1	Opt. 8
Baseline available	No	Yes (LTE DC)	No							Yes (CPRI)
Traffic aggregation	No	Yes								
ARQ location	DU			CU May be more robust under non-ideal transport conditions						
Resource pooling in CU	Lowest	in between (higher on the right)								Highest
	RRC only	RRC + L2 (partial)				RRC + L2	RRC + L2 + PHY (partial)			RRC + L2 + PHY
Transport NW latency requirement	Loose				NOTE 7	Tight				
Transport NW Peak BW requirement	N/A	Lowest	in between (higher on the right)							Highest
	No UP req.	baseband bits						Quantized IQ (f)		Quant. IQ (t)
	-	Scales with MIMO layers							Scales with antenna ports	
Multi-cell/freq. coordination	multiple schedulers (independent per DU)				centralized scheduler (can be common per CU)					
UL Adv. Rx	NOTE 7						NA	NOTE 7	Yes	
Remarks	NOTE 4				NOTE 5/6	NOTE 5	NOTE 5	NOTE 5		

NOTE 1: This summary is based on LTE protocol stack and is to be updated if necessary based on NR protocol stack.

NOTE 2: This summary table is not to be used for evaluation of split options in its current form.

NOTE 3: The table is intended to provide a high-level summary on the characteristics of the different CU-DU split options. Therefore, the items listed are non-exhaustive (but rather limited to some of the main items), and the descriptions are abstractive (rather than being accurate but too detailed).

NOTE 4: May be beneficial for URLLC/MEC.

NOTE 5: Complexity due to separation of Scheduler & PHY processing.

NOTE 6: Complexity due to separation of Scheduler & HARQ.

NOTE 7: Was not clarified during the study phase.

11.1.3 Architectural and specification aspects

11.1.3.1 Number of split options to be specified and supported by open interface

There are transport networks with performances that vary from high transport latency to low transport latency in the real deployment. 3GPP specifications should try to cater for these types of transport networks. For transport network with higher transport latency, higher layer splits may be applicable. For transport network with lower transport latency, lower layer splits can also be applicable and preferable to realize enhanced performance (e.g. centralized scheduling). Thus, preferable option would be different between different types of transport networks (ranging from lower layer split for transport networks with lower transport latency to higher layer split for transport networks with higher transport latency). Furthermore, within lower layer split discussion, there are both demands to reduce transport bandwidth and demands to support efficient scheduling and advanced receivers.

The Option 8 has been available in today deployment based on a de facto standard from a forum other than 3GPP, 3GPP should not attempt to specify this option 8.

11.1.3.2 Implications of LTE/NR tight interworking

LTE <-> NR interworking is mainly based on Dual-Connectivity-like mechanisms. Such approach does not imply any particular functional split. The requirement that could be extrapolated by the LTE-NR tight interworking requirement is that of allowing aggregation of PDCP functionalities, in case of split bearers.

11.1.3.3 Granularity of the Functional Split

Some possible options for the granularity of the CU/DU functional split are listed below:

- Per CU: each CU has a fixed split, and DUs are configured to match this.

- Per DU: each DU can be configured with a different split. The choice of a DU split may depend on specific topology or backhaul support in a given area.

NOTE 1: For 2 cases above, one possible way on how the CU/DU decide or coordinate the split would of course be through configuration. Alternatively the split could be "negotiated" taking into account capabilities of the two units, and deployment preference e.g. based on backhaul topology.

- Per UE: different UEs may have different service levels, or belong to different categories, that may be best served in different ways by the RAN (e.g. a low rate IOT-type UE with no need for low latency does not necessarily require higher layer functions close to the RF).

- Per bearer: different bearers may have different QOS requirements that may be best supported by different functionality mapping. For example, QCI=1 type bearer requires low delay but is not SDU error sensitive, while eMBB may not be delay sensitive but has challenging requirements on throughput and SDU error rate.

- Per slice: it is expected that each slice would have at least some distinctive QOS requirements. Regardless of how exactly a slice is implemented within the RAN, different functionality mapping may be suitable for each slice.

From above, Per CU and Per DU options pertain to flexibility of network topology, and should be straightforward to support. Whether procedures are required to handle the initial configuration (or O&M is relied upon) was not discussed during the study phase. Note that in the Per DU option, one CU may need to support different split levels in different interfaces, which is not the case in the Per CU option.

Further granularity (Per UE, Per bearer, Per slice) requires analysis and justification based on QOS and latency requirements. Note that the Per UE, Per bearer and Per slice options imply that a particular instance of the interface between CU/DU would need to support simultaneously multiple granularity levels on user plane.

NOTE 2: The baseline is CU based or DU based. If there are demands to have finer granularity (e.g. Per UE,Per bearer, Per slice), justification should be made clear first.

11.1.3.4 Reconfiguration dynamicity of the functional split

Dynamicity implies that the protocol distribution and the interface between the CU and DU need to be reconfigured. If the switching only occur in CU-DU setup procedure, the interface design will not be influenced largely as the split option will not be changed during operation. If the switching occurs during operation, there will be impact on complexity of interface.

11.1.3.5 Standardization of Centralized RRM Functions

Most if not all of the defined functional splits allow for having RRM functions like Call Admission Control and Load balancing in the CU controlling multiple DUs. This allows for the potential of increased efficiency in inter-cell coordination for RRM functions like the coordination of interference management, load balancing and Call Admission Control.

However that efficiency can only be realized if the CU can have reliable and accurate understanding of the current environment at the DU which can include issues beyond just radio conditions, but can include current processing capabilities, or in the case of wireless or mesh backhauling help in determining current terrestrial capacity. We in RAN3 have been dealing with issues like this since Release 99 in UMTS and in LTE (a big example is load reporting, but there are others).

11.1.3.6 Standardization Issues with Centralized scheduling Options

Functional split Option 5, Option 6, Option 7 and Option 8 allow for scheduling of data transmission in the CU.

Having centralized scheduling can provide benefit particularly for interference management and coordinated transmission in multiple cells (like soft handover in UMTS, or CoMP in LTE). However this requires the CU to have an even better understanding of the state of the DU radio conditions than for CAC and other centralized RRM functions.

It also requires either very low latency/jitter transport or sufficiently tight coordination of timing and reception of user plane data (one solution is the window mechanism used on the UP in UMTS), but this can be challenging particularly for lower latency use cases in NR.

Centralization of RAN functions has strong potential for some benefits such as reduced cost, improved scalability, more efficient inter-cell coordination for interference management as well as improved mobility in ultra-dense deployments.

11.1.3.7 Transmission of RRC message over the CU-DU link

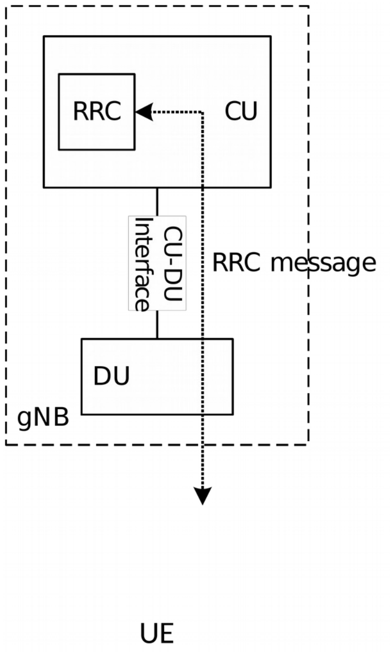


Figure 11.1.3.7-1: Transmission of RRC message between the CU and the UE via the DU

The RRC related functions should be located in the CU for all functional split options. The RRC message between the gNB and the UE should be transferred through the interface between the CU and the DU as illustrated in Figure 11.1.3.7-1. RRC messages could require a differentiated transport between CU and DU compared to data transport, e.g. in terms of robustness and delay.

NOTE: How to carry the RRC message via CU-DU interface needs further investigation.

11.1.3.8 CU-DU specification aspects

Architectural aspects

The architecture of gNB with CU and DUs is shown in Figure 11.1.3.8-1. Fs-C and Fs-U provide C-plane and U-plane over Fs interface, respectively.

In this architecture, CU and DU can be defined as follows.

Central Unit (CU): a logical node that includes the gNB functions as listed in section 6.2 excepting those functions allocated exclusively to the DU. CU controls the operation of DUs.

Distributed Unit (DU): a logical node that includes, depending on the functional split option, a subset of the gNB functions. The operation of DU is controlled by the CU.

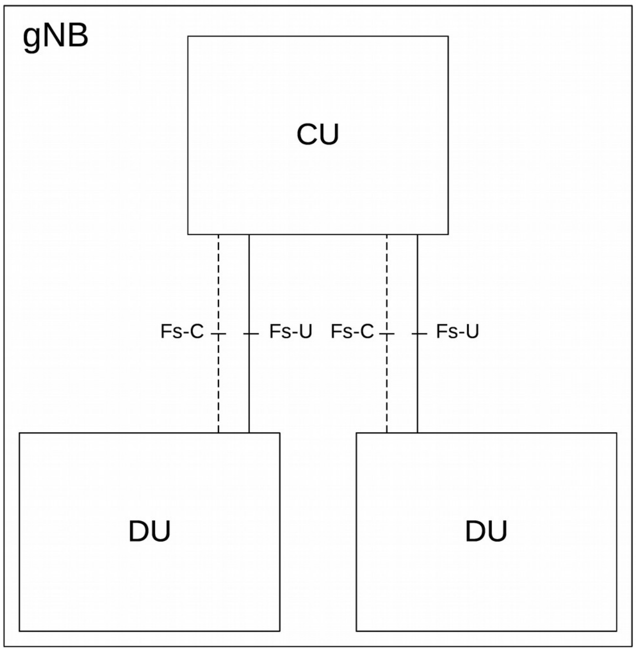


Figure 11.1.3.8-1: gNB architecture with CU and DUs

11.1.4 Transport network aspects

11.1.4.1 General

This section summarizes transport network requirements of different functional splits.

NOTE: It is understood that RAN3 has no intention to specify any transport network.

11.1.4.2 Transport network requirements for an example RAN architecture for NR

According to TR 38.913 [5], the NR shall support up to 1GHz system bandwidth, and up to 256 antennas. A calculation relative to one of several possible transport deployments applied to a possible RAN architecture example shows that transmission between base band part and radio frequency part requires a theoretical maximum bitrate over the transport network of about 614.4Mbps per 10MHz mobile system bandwidth per antenna port.

When the system bandwidth is increasing as well as the number of antenna ports, the required bitrate is linearly increasing. An example with rounded numbers is shown in the following table. Note that the figures in Table 11.1.4.2-1 are a maximisation of the needed bandwidth per number of antenna ports and frequency bandwidth.

Table 11.1.4.2-1 Examples of maximum required bitrate on a transmission link for one possible PHY/RF based RAN architecture split

Number of Antenna Ports	Frequency System Bandwidth			
	10 MHz	20 MHz	200 MHz	1GHz
2	1Gbps	2Gbps	20Gbps	100Gbps
8	4Gbps	8Gbps	80Gbps	400Gbps
64	32Gbps	64Gbps	640Gbps	3200Gbps
256	128Gbps	256Gbps	2560Gbps	12800Gbps

NOTE: Peak bitrate requirement on a transmission link = Number of BS antenna elements * Sampling frequency (proportional to System bandwidth) * bit width (per sample) + overhead. The calculation is made for sampling frequency of 30.72 Mega Sample per second for each 20MHz and for a Bit Width equal to 30.

11.1.5 Conclusions on functional split between central and distributed unit

Higher Layer Split

There shall be normative work for a single higher layer split option, i.e. Stage 2 and Stage3.

In the meantime, if other decisions cannot be made, RAN3 recommends to progress on Option 2 for high layer RAN architecture split. The contributions to the April meeting with regards to option2 against option 3-1 should be limited to address the fast centralized retransmission of lost RLC PDUs. If no agreement can be reached, a formal vote will be set-up, which will result in a down selection between Option 2 or Option 3-1, by April 2017.

Normative contributions to different options are also expected.

Lower Layer Split

The study on lower layer split RAN architectures is not completed and postponed.

Further study is required to assess on low layer splits, their feasibility, the selection of options and assess the relative technical benefits, based on NR, before a decision to go to specification phase can be made. Discussions in the Study Item, favored option 6 and 7 for future study.

11.2 UP-CP Separation

11.2.1 General

Next Generation networks will see an increasing use of very dense deployments where user terminals will be able to connect to multiple transmission points simultaneously.

It may be favourable and beneficial for the next generation RAT, to base the architecture on a separation of the CP and UP functions. This separation would imply to allocate specific CP and UP functions between different nodes.

The following points can be taken as examples for expected benefits of a separation of CP and UP:

- A centralization of CP functions, controlling different transmission points, has the potential to achieve enhanced radio performance.
- Flexibility to operate and manage complex networks, supporting different network topologies, resources and new service requirements.
- Alignment with SDN concept that would result in a functional decomposition of the radio access, based on a partial de-coupled architecture, between user and control plane entities and on network abstractions [10].
- For functions purely handling with CP or UP processes, independent scaling and realization for control and user plane functions operation.
- Support of multi-vendor interoperability.

In addition, the following points also need further considerations regarding splitting gNB to two nodes: one is specific for gNB Control plane, the other is specific for gNB user plane:

- The relationship and the co-existence with RAN protocol split architectures
- The impacts on NR/LTE tight interworking

11.2.2 UP and CP Functions Description and Grouping

As a starting point the subsequent analysis considers the LTE protocol Stack as listed by RAN3 in 3GPP TR 38.913. The functions are augmented and aligned with the RRM functionalities defined in 3GPP TS 36.300.

In current TS 36.300[2], the LTE functions are listed as in the figure below:

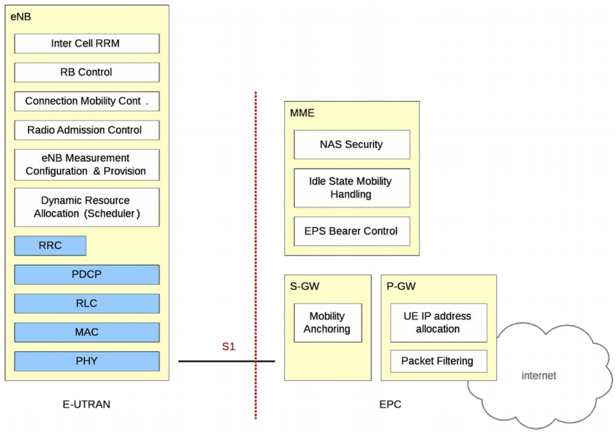


Figure 11.2.2-1: Function Split between E-UTRAN and EPC

If we start from the blue boxes (radio protocol layers) the functions in RRC layer could be considered as belonging to control plane, which is controlling all the radio resources.

The Inter-Cell RRM, RB control, Connection Mobility Continuity, Radio admission control, eNB measurement configuration and provision, Dynamic resource allocation are functions with a control plane component; while transfer of user data are user plane functions.

The functions listed above are those existing in LTE.

There were identified two basic architectures for UP-CP split, which conduct to a difference in the assignment of control functions to the central control plane:

- A flat UP-CP separation architecture, where there is a clear UP-CP function separation as LTE function separation;
- A hierarchical UP-CP separation architecture, where the UP-CP are not separated for the functions used in LTE, instead the less time-sensitive operation of these functions could be coordinated by a higher hierarchy Central Coordinator.

Since the protocol stack for the new RAT will most likely a being based on what has been introduced for LTE, the synthesis of CP and UP functions will need to take this circumstance into account.

While Inter-Cell RRM, RB control, Connection Mobility Continuity, Radio admission control, eNB measurement configuration and provision have a strong CP component, all other radio protocol layers are shared by user plane and control plane, as they are sharing the same radio resources.

All the control signalling and data are transmitted via the same protocols i.e. using the same function groups.

The mobile system especially the air interface design is quite optimized. The frequency resource is limited, and the air condition changes quickly. Therefore, the LTE protocol stack has been carefully design and strict time pattern during the interaction of the radio protocol layers for the Uu interface have been considered. The function for control plane and user plane have been considered together to maximise the performance of the air interface. The functions have strong dependency and are very difficult to separate.

Some examples of the mix of control plane/ user plane function in Uu interface:

- PHY: there are several control channels including PBCH, BCH, PDCCH, PDSCH, EPDCCH etc, and also the channel data i.e. PDSCH. However, the PDSCH can also carry the PCH which is the function for control plane. The whole system operates relying on other control signaling including reference signals and synchronization etc.
- MAC: it is designed as user plane to unify the data transmission regardless the data is for signaling or real package. However, there are still controlling functions/procedures in MAC including RACH procedure, cell activation/deactivation, TA command, PHR etc. There is only one MAC entity in E-UTRAN for a UE. The entity

is responsible for all the control signalling and data transmission. It is impossible to separate the user plane and control plane in MAC layer in network side.

- RLC/PDCP: The two layer are relatively simpler. However, there are still controlling PDU for the two layers e.g. RLC status reporting, the encryption and integration protection for the signalling and data etc.

- Scheduling: Scheduling is used to control the assignment of resource to different traffic sources, both for signalling and for user data transmission. Scheduling is the mechanism that provisions UP channels with the resources that make data traffic exchange possible. Scheduling is ruled via policies established via CP and influences future CP settings, while at the same time being the core of the UP resource provisioning mechanism. In order to provide optimal centralised scheduling a real time detailed knowledge of the UP traffic available in UL and DL is needed. For this reason, scheduling has a role both in the CP and in the UP.

- Logical channel multiplexing: this function is in charge of multiplexing data of different logical channels into the same time frequency resources. This is clearly a task performed on UP traffic to increase the efficiency of UP data transmission. However, multiplexing is controlled by CP established configurations and policies, which make it difficult to classify this function as a pure UP one.

Unless further analysis proves a difference in the outcomes above, the dependencies between some control and user plane functions, as described in the examples above, make the separation of CP and UP for all functions of an E-UTRAN based function set for NR, difficult and likely not practical.

11.2.3 RAN architecture and interfaces for UP-CP Separation

Figures 11.2.3-1 and 11.2.3-2 show some deployment scenarios below based on Higher Layer Split Option 2 as example and reusing Release 12 Dual Connectivity concepts. This CP-UP separation permits flexibility for different operational scenarios, such as:

- Move the PDCP to a Central Unit while keeping RRM in a master cell.

- Move RRM to a more central location where it has oversight over multiple cells; while allowing independent scalability of user plane and control plane.

- Central RRM with local breakout of some data connections of some UEs nearer (or at) the base station site.

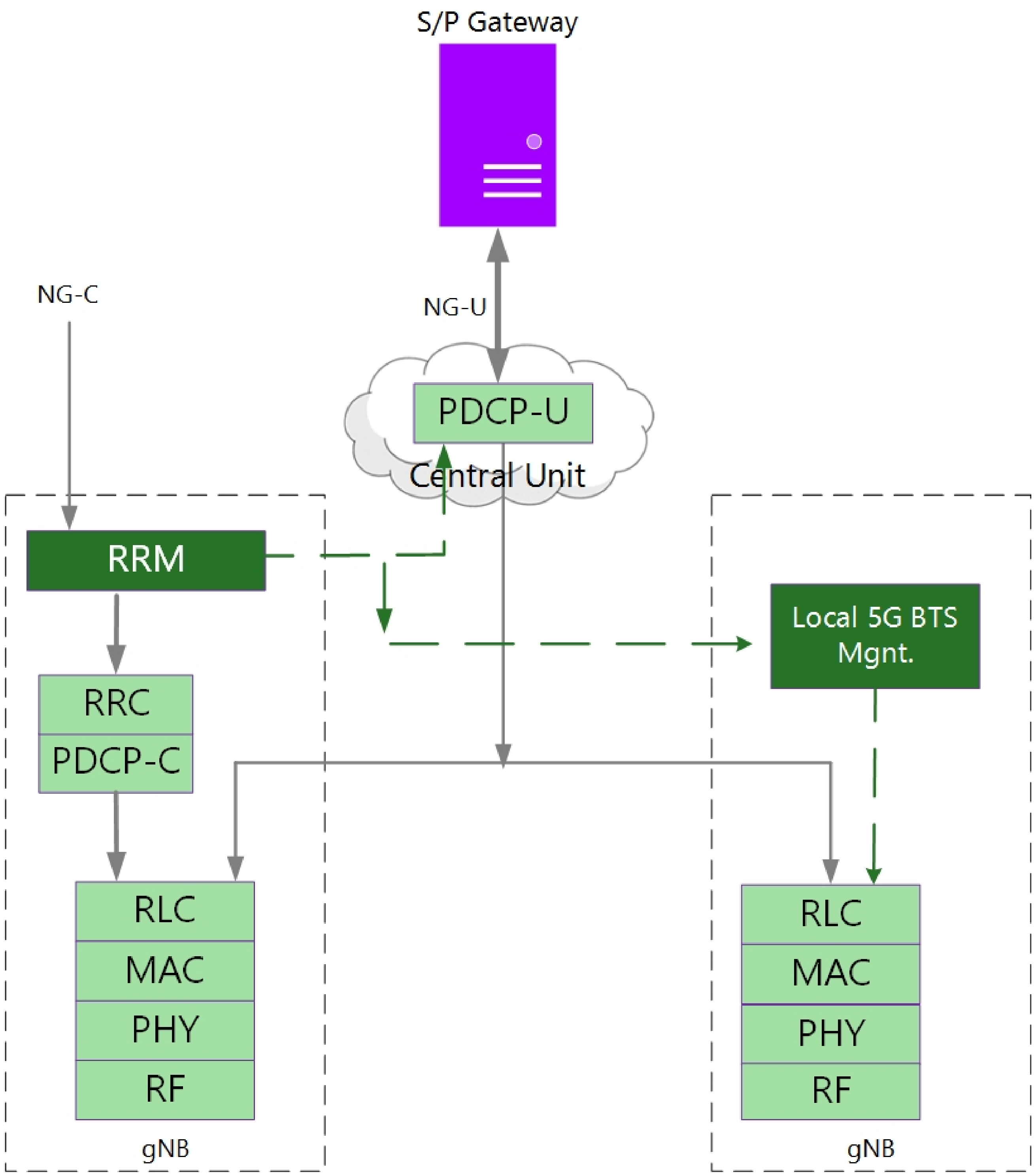


Figure 11.2.3-1: Centralised PDCP-U with Local RRM

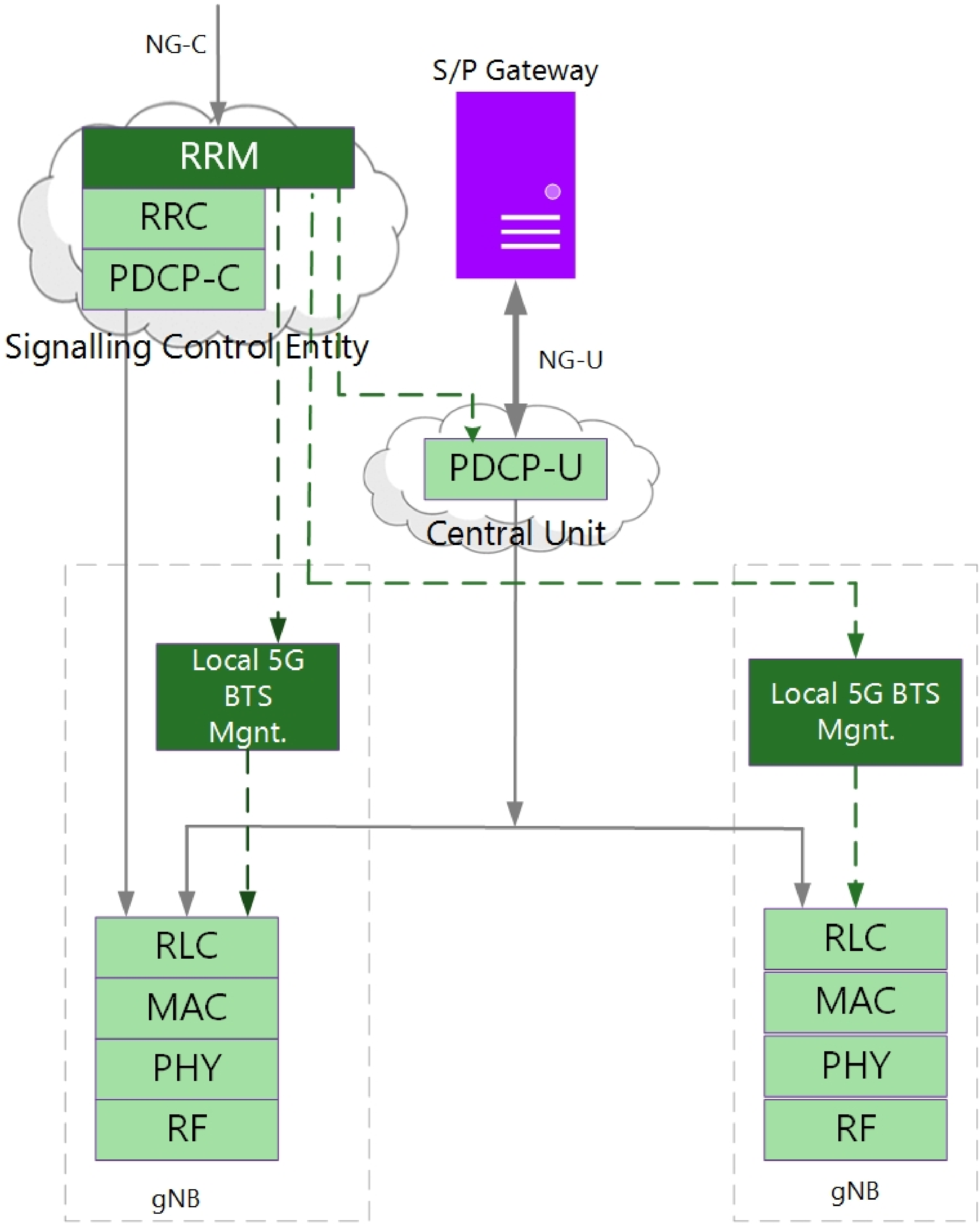


Figure 11.2.3-2: Centralised PDCP with Centralised RRM in separate platforms

The above figure illustrates how a higher layer CU/DU functional split based e.g. on Option 2 would allow for user and control plane separation. It allows for the independent scalability and evolution (e.g. addition of new encryption algorithms) of user and control planes.

11.2.4 Conclusions

Some of the benefits of Control and User Plane separation based on Release 12 Dual Connectivity were identified, but solutions details were not discussed during the study item.

11.3 Realization of RAN Network Functions

NR shall allow Centralized Unit (CU) deployment with Network Function Virtualization (NFV).

12 SON

12.1 Scope of SON for NR

Self-organizing network (SON) functions such as support for mobility robustness optimization, mobility load balancing, RACH optimization, energy savings, etc. were introduced in LTE to make deployment easier and cheaper. Equivalent functionality shall also be considered for the NR. In addition, new SON functionality may be considered, taking into account new NR features, new use cases supported by NR, and the operation in complex multi-RAT, multi-band and multi-vendor deployments.

12.2 Self configuration procedures

12.2.1 Automatic neighbour relations

Relations in 5G can be between entities, supporting different radio resource management procedures. Such procedures may include coordination of resources and identifiers, mobility, load and traffic sharing, etc. An example of such neighbour relation is relation among gNBs and eLTE eNBs interconnected with Xn.

13 Wireless relay

13.1 Scenarios

The following scenarios for wireless relay in New RAN should be considered.

- Single-hop stationary relay
    - A relay node may connect to a donor node through wireless backhaul to extend network coverage. Relay nodes could be used for network densification and coverage extension with a trade-off between deployment cost, deployment flexibility, capacity loss, and increased latency at the donor with respect to other solutions.
- Multiple-hop relay
    - Due to the limited coverage of a relay node, it may be needed to consider the support for multiple hop relay to extend network coverage. In this case, the traffic may be transmitted via one or more intermediate relay nodes, i.e. hop by hop. Multi hop relay may apply to some practical use cases (e.g. street canyon), with a trade-off between deployment cost, deployment flexibility, capacity loss, and increased latency at the donor with respect to other solutions.
- Multiple donor relay
    - To further improve the bandwidth, a relay node may connect to multiple donors. The relevance of this scenario with respect to e.g. small cell and dual connectivity use cases, should be analysed.
- Mobile relay:
    - A relay node may be deployed on a vehicle, and provides wireless connectivity service to end user inside the vehicle. The relay node's donor node may be changed, e.g. moving across the coverage of different donor node. The relevance of this scenario with respect to previous studies in LTE should be considered; furthermore, the feasibility of this scenario with respect to the physical layer should be evaluated by the appropriate WG(s).

14 Migration towards RAN for NR

14.1 Potential migration path 1 [8]

14.1.1 Step 1 deployment

When NR is launched on day 1, a likely scenario is to deploy NR on frequencies higher than those being used for LTE. In this case, the NR coverage is most likely to be much smaller than the existing LTE coverage, especially for frequencies above 6 GHz. For eMBB which can be regarded as continuous evolution of the existing cellular service, it is desirable if the existing LTE coverage can be leveraged to provide nationwide continuous coverage and mobility. In addition to that, the NR coverage enables to boost U-plane capacity in the target stop area where the traffic load is high. Option 3/3a/3xenables operators to launch the NR service as such; eNB acts as MeNB and gNB acts as SeNB. Since LTE eNB as MeNB is already connected to EPC, leveraging EPC can further drives cost effective and early launch of the NR service for eMBB. LTE-NR Dual Connectivity via EPC.

14.1.2 Step 2 deployment and on wards

After the day 1 deployment, next step in question is how NGC is introduced. As for migration from EPC, the following three roles can be considered.

- NGC is evolution of EPC.
- NGC encompasses EPC as a slice.
- NGC replaces EPC.

From operator’s viewpoints, it has to be understood whether:

a) NGC is backward compatible with EPC and so can provide the same service as EPC can.

b) NGC is not backward compatible with EPC and so is designed to provide a new service and use case which cannot be done by EPC.

Even after the day 1 NR deployment and the existing LTE coverage is replaced with NR, operators can reap the benefit if EPC can accommodate NR as a standalone RAT. Even though NGC is introduced, NGC might work as EPC as explained above. As such, NGC in relation to EPC has to be clarified before making a final decision on the RAN-CN interface scenarios to be standardised.

14.2 Potential migration path 2 [9]

One identified sequence of migration from existing E-UTRAN to the New RAN is as follows:

Step 1: Early 5G Deployment utilizing Option 3

Step 2: Migration to Option 7 – Includes Simultaneous Support for Option 3

Step 3: Migration to Option 4 and/or 2 – Plus Simultaneous Support for Option 3 & Option 7

NOTE: Options 2, 3, 4 and 7 are described in section 7.1.

- Option 3 may include 3 and/or 3a and/or 3x.
- Option 4 may include 4 and/or 4a.
- Option 7 may include 7 and/or 7a.

The applicable mobility and service continuity scenarios should be supported for each migration step.

14.3 Potential migration path 3 [10]

14.3.1 General

The migration paths from existing E-UTRAN to RAN for NR are based on the deployment architecture set options considered in [7]. The potential migration paths are as follow.

14.3.2 Potential migration path 3-1

LTE/EPC -> Option 2 + Option 5 -> Option 4/4a -> Option 2

- Step 1: Option 2 and 5 are deployed in parallel once the NG core is available. Option 7&7a can also be supported at this step.
- Step 2: Option 4/4a could be deployed
- Step 3: Only Option2 is used.

UE mode:

- Dual mode UE in step 1 with NG NAS.
- Only NR NAS in step 2
- NR UE in step 3

14.3.3 Potential migration path 3-2

LTE/EPC -> Option 2 + Option 5 -> Option 2

- Step 1: Option 2 and 5 are deployed in parallel once the NG core is available. Option 7&7a can also be supported at this step.
- Step 2: Only Option 2 is used.

UE mode:

- Dual mode UE in step 1 with NG NAS
- NR UE in step 2

14.3.4 Potential migration path 3-3

LTE/EPC -> Option 3/3a/3x -> Option 4/4a-> Option 2

- Step 1: Non-standalone NR is deployed for support of LTE-NR tight interworking, based on legacy LTE network. Option 3&3a&3x can be applied.
- Step 2: Standalone NR is deployed together with the NG core. Option 4/4a could also be deployed.
- Step 3: Only Option 2 is used.

UE mode:

- Only LTE NAS in step 1
- Only NR NAS in step 2
- NR UE in step 3

14.3.5 Potential migration path 3-4

LTE/EPC -> Option 7/7a -> Option 2

- Step 1: Option 7/7a is deployed for support of eLTE-NR tight interworking
- Step 2: Only Option 2 is used.

UE mode:

- Only NR NAS in step 1
- NR UE in step 2

14.3.6 Potential migration path 3-5

LTE/EPC -> Option 3/3a/3x -> Option 1 + Option 2 + Option 7/7a -> Option 2 + Option 5

- Step 1: Non-standalone NR is deployed for support of LTE-NR tight interworking, based on legacy LTE network. Option 3/3a/3x can be applied, in which option 3 maybe prefer in high priority.
- Step 2: eLTE-NR tight interworking, with LTE anchor connected to 5G core, standalone LTE and standalone NR should be supported.
- Step 3: Only NG core are used.

UE mode:

- Only LTE NAS in step 1.
- NR NAS and LTE NAS in step 2.
- NR UE in step 3.

14.4 Potential migration path 4 [15]

14.4.1 Deployment considerations

In order to limit dependency on legacy EPC, to reduce complexity in roaming scenarios and to enable wide business opportunities, any phased approach in the migration path that results in fragmentation of the ecosystem should be avoided, hence UEs under New RAN coverage shall be able to connect to the NGC since early phases of New RAN deployment.

<u>Initial deployment</u> shall be “future-proof”, by avoiding any legacy with respect to subsequent deployment phases. In the early phase eLTE and the NR will be deployed gradually.

- NR gNBs will be mainly deployed in Non-standalone mode through an interworking with eLTE eNBs and with a connection to the NGC. Some of the existing LTE eNBs need to be gradually upgraded to eLTE eNBs to support the NG interface towards NGC, in addition to supporting the existing S1 towards the EPC.
- All New RAN capable UEs can be steered to NGC.

<u>In a second phase</u>, NGC will support most of the UEs, whilst EPC will gradually shrink. Besides the Non-standalone NRs, also the Standalone NRs will be deployed, depending on use cases and commercial availability.

The final target architecture will be based on New RAN only, i.e. with all LTE eNBs upgraded to eLTE eNBs, and on NGC, whilst the EPC will be maintained to manage legacy UEs only, which are unable to access the New RAN.

14.4.2 Two-step migration path

The following steps are identified based on previous deployment considerations:

- Step 1: Early New RAN deployments are based on Option 5 and Option 7
- Step 2: Extension of Step 1 by introducing deployments based on Option 2

NOTE: Options 2, 5 and 7 are described in section 7.1.

- Option 7 may include 7 and/or 7a.

14.5 Potential migration path 5 [16]

14.5.1 Potential migration path 5-1

LTE/EPC -> Option 2 + Option 7/7a -> Option 2

- Step 1: eLTE-NR tight inter-working with LTE anchor connected to 5G core and standalone NR should be supported.
- Step 2: Only NG core and option2 are used.

UE mode:

- Dual mode and single mode NR NAS in step 1
- NR UE only in step 2

14.5.2 Potential migration path 5-2

LTE/EPC -> Option 2 + Option 7/7a ->Option2+Option7/7a+Option4/4a-> Option 2

- Step 1: eLTE-NR tight inter-working with LTE anchor connected to 5G core and standalone NR should be supported.
- Step 2: Option 4/4a could be deployed.
- Step 3: Only NG core and option2 are used.

UE mode:

- Dual mode and single mode NR NAS in step 1
- Only NR NAS in step 2
- NR UE only in step 3

14.6 Implications of migration paths on RAN3

In order to support the potential migration paths from existing E-UTRAN to the New RAN:

- RAN3 should focus on the following with highest priority (i.e. Rel-15): Xx interface support, NG interface support, and Xn interface support for essential functionality that is not “option-specific” (e.g. interface management, UE connected mode mobility management, etc.).
- Xn interface functionality that is “option-specific” requires input from other groups.

15 Interworking with non-3GPP systems

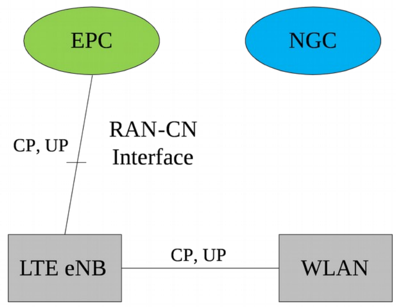


Figure 15-1: LTE connected to the EPC, WLAN interworking with LTE via inter node interface. In this scenario it is assumed that there is an interface between LTE and WLAN.

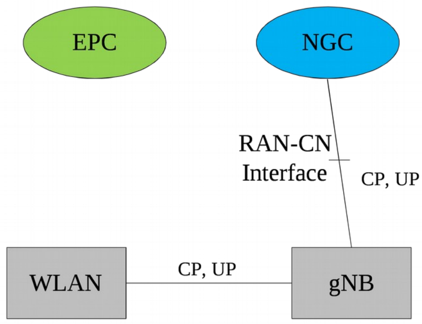


Figure 15-2: NR connected to the NGC, WLAN interworking with NR via inter node interface. In this scenario it is assumed that there is an interface between gNB and WLAN

Annex A: Transport network and RAN internal functional split

When considering functional split options, the following transport performance requirements may be expected. The values given in the table are informative and for reference. The following transport characteristics deemed to be relevant:

1) Transport latency

2) Transport bandwidth

Those transport characteristics are contributing finally to deployment costs.

On the other hand, certain features and/or use cases like Ultra Reliable and Low Latency communication (URLLC) may require a certain split to support the features and/or use cases.

The following Table A-1 is proposed to be maintained during the SI while the knowledge about the protocol stack for NR and the related requirements on the transport are evolving.

Table A-1 Requirements on the underlying transport network due to a certain functional split, as a consequence to support a certain feature/use case

Protocol Split option[1]	Required bandwidth	Max. allowed one way latency [ms]	Delay critical feature[2]	Comment
Option 1	[*DL: 4Gb/s*] [*UL: 3Gb/s*]	[*10ms*]		
Option 2	[*DL: 4016Mb/s*] [*UL:3024 Mb/s*]	[*1.5~10ms*]		[16Mbps for DL and 24Mbps for UL is assumed as signalling]
Option 3	[*lower than option 2 for UL/DL*]	[*1.5~10ms*]		
Option 4	[*DL:4000Mb/s*] [*UL:3000Mb/s*]	[*approximate 100us*]		
Option 5	[*DL: 4000Mb/s*] [*UL: 3000 Mb/s*]	[*hundreds of microseconds*]		
Option 6	[*DL: 4133Mb/s*] [*UL:5640 Mb/s*]	[*250us*]		[133Mbps for DL is assumed as scheduling/ control signalling. 2640Mbps for UL is assumed as UL-PHY response to schedule]
Option 7a	[*DL:10.1~22.2Gb/s*] [*UL:16.6~21.6Gb/s*]	[*250us*]		[713.9Mbps for DL and 120Mbps for UL is assumed as MAC information]
Option 7b	[*DL:37.8~86.1Gb/s*] [*UL:53.8~86.1 Gb/s*]	[*250us*]		[121Mbps for DL and 80Mbps for UL is assumed as MAC information]
Option 7c	[*DL:10.1~22.2Gb/s*] [*UL:53.8~86.1Gb/s*]	[*250us*]		
Option 8	[*DL:157.3Gb/s*] [*UL: 157.3Gb/s*]	[*250us*]		

Note: The values are examples provided by LTE reference, as provided in [11] and [14] (modification of required bandwidth in [11]), and are to be replaced by NR values when available. The assumptions for required bandwidth are in Table A-2.

1 Description of the split option

2 Driving feature / use-case requiring a certain split option

Table A-2 Assumptions for required bandwidth in Table A-1

Items	Assumption	Applicability
Channel Bandwidth	[100MHz(DL/UL)]	All options
Modulation	[256QAM(DL/UL)]	
Number of MIMO layer	[8(DL/UL)]	
IQ bitwidth	[2*(7~16)bit(DL), 2*(10~16)bit(UL)]	Option 7a Option 7b Option 7c
	[2*16bit(DL/UL)]	Option 8
Number of antenna port	[32(DL/UL)]	Option 7b Option 7c(UL) Option 8

Annex B: NG Interface Protocol Stacks for User Plane

B.1 Description of candidate solutions

This section presents description of possible protocol stack options for the user plane.

Solution 1: GTP-U/UDP/IP

GTP-U/UDP IP (TS29.281) is a protocol already used over S1, X2, S5/8 LTE interfaces. This is illustrated below:

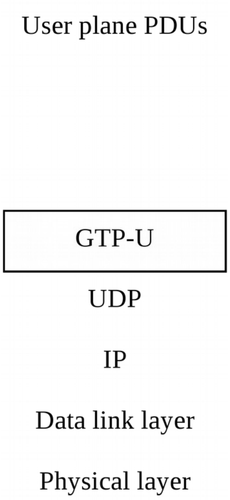


Figure B.1-1: GTP-U/UDP/IP Protocol stack

The details of the GTP-U structure is shown below:

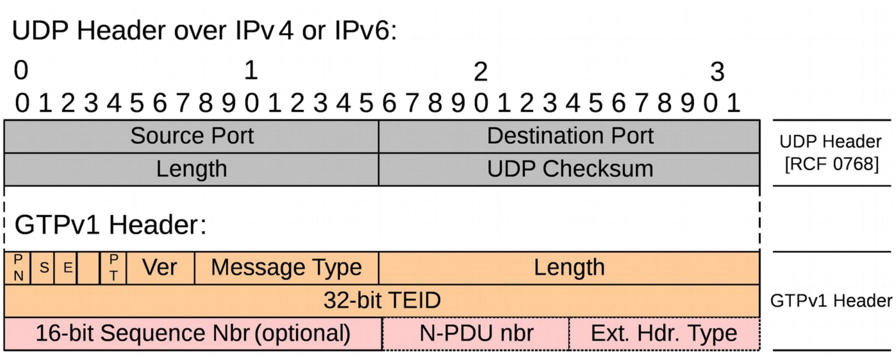


Figure B.1-2: GTPv1 Header

This protocol stack offers the following key characteristics:

Table B.1-1 Key Characteristics of GTP/UDP/IP

Feature	GTP
Message Type	Yes.
Length	16-bit payload length
Protocol multiplexer	No: all packets of a given tunnel must be of same type and this type

	needs to be configured through NGAP if it is different than IP.
payload Type	Supports any payload.
Tunnel multiplexer	Mandatory 32-bit TEID
Sequence Number	Optional 16-bit
Checksum	In UDP header
QoS transport marking	DSCP in outer IP header (TS 36.414)
5G QoS marking	Needs specific extension (encapsulation) header as per SA2 interim agreement (extension must be at least 4 octets long).
Carried over	UDP/IP port 2152
End Marker in HO	Message Type 254
Transport overhead	IP + UDP + GTP Hdr (20 bytes + IP header)
U-plane possible without tunnelling	No.
NAT Traversal	Yes.

Solution 2: GRE/IP

Generic Routing Encapsulation protocol (GRE) over IP has been specified in the IETF (RFC 2784, RFC 2890) and has been applied for e.g. PMIP based S5/S8, 3GPP LWIP, 3GPP2. This is illustrated below:

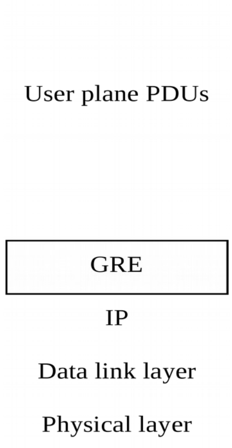


Figure B.1-3 GRE/IP Protocol stack

The details of the GRE structure is shown below:

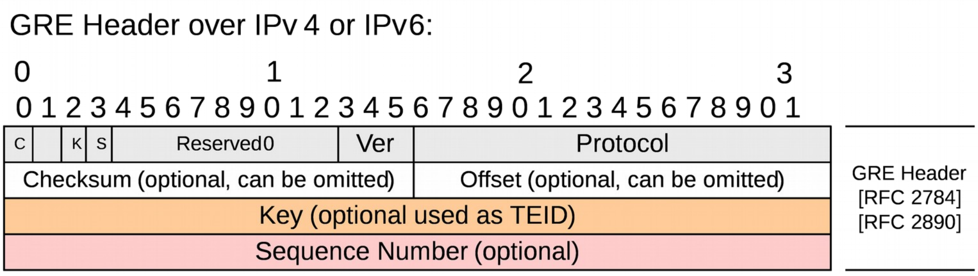


Figure B.1-4: GRE Header

This protocol stack offers the following key characteristics:

Table B.1-2 Key Characteristics of GRE/IP

Feature	GRE
Message Type	No (NOTE 1)
Length	In IP header
Protocol multiplexer	Yes: 16-bit payload identifier (Ethertype).
payload Type	Supports any payload specified as Ether Protocol Type (e.g. can support Ethernet over GRE).
Tunnel multiplexer	Optional 32-bit Key
Sequence Number	Optional 32 bit
Checksum	Optional
QoS transport marking	DSCP in outer IP header
5G QoS marking	Requires extension (NOTE 1)
Carried over	IPv4/IPv6 (IETF protocol number 47)
End Marker in HO	No (NOTE 1)
Transport overhead	IP + GRE (4-16 bytes depending on fields used + IP header).
U-plane possible without tunnelling	Yes (NOTE 2)
NAT Traversal	No (NOTE 3).

NOTE 1: Partitioning of the KEY field could be done to reserve a few bits for this feature. The interpretation of KEY field by the receiver can be specified in 3GPP specification.

NOTE 2: Whether GRE/IP/L2 can be removed from the protocol stack in some cases and then the remaining L2 can be carried directly over any underlying technology (MPLS-EVPN, VXLAN, ...) was not concluded during the study phase.

NOTE 3: It needs to be determined from requirements (SA2) if NAT traversal is required over NG interface. If needed, an alternative approach is to use GRE over IPV6.

Solution 3: Protocol Oblivious Encapsulation (PoE)

PoE is expressed without limiting the structure to specific formulations. This can be done by enhancing the configurability of the encapsulation protocol over the control plane.

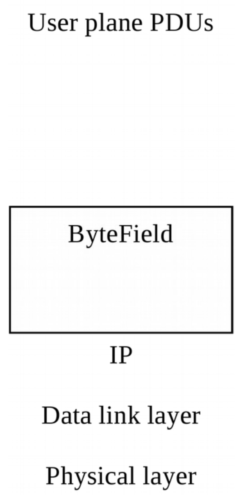
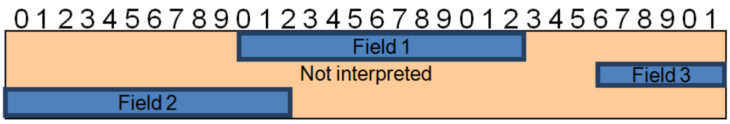


Figure B.1-5: Protocol Oblivious Protocol stack

On the DL during interface establishment, the control plane configures the locations of each needed field. This location is indicated as a pair of numbers <length, offset> from the underlying TNL. Example fields could be PDU session identifier, QoS marking etc. The location of the User plane PDUs is similarly indicated.

The details of an example PoE structure is shown below, with 3 fields identified over NG-C at different locations in an NG-U transport packet.

PoE Header definition on DL



Field	<Length, Offset>
Field 1	<12, 10>
Field 2	<12, 64>
Field 3	<6, 58>

Figure B.1-6: PoE Header in DL

On the UL as part of PDU session configuration, the NG-C provides a bytestring. The gNB prepends this bytestring to every PDU associated with this PDU session before transmission on the TNL.

PoE Header definition on UL

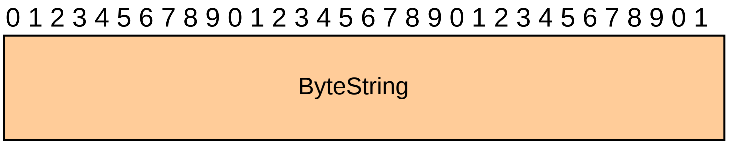


Figure B.1-7: PoE Header in UL

NOTE 1: Whether it is feasible for RAN to read or write the tunnelling header was not concluded during the study phase.

NOTE 2: The performance of the protocol e.g. on flexibility and efficiency was not concluded during the study phase.

<u>Solution 4: Configuration of NG-U protocol stack over NG2</u>

Similar to solution 3, this solution also proposes to not limit the structure used for protocol encapsulation by enhancing the configurability of the NG-U encapsulation protocol over the control plane as follows:

- The protocol stack to be used by gNB over NG-U is signalled by 5G CN over NG-C by a *NG-U Protocol* IE included in the PDU Session Setup Request message.

<u>Solution5: Geneve/UDP/IP</u>

Geneve/UDP IP is a tunnelling protocol with flexible options infrastructure defined in [xx]. It is specifically designed for future tunnelling use cases, including SDN. This is illustrated below:

User plane PDUs

Geneve

UDP

IP

Data link layer

Physical layer

Figure B.1-8: Geneve/UDP/IP Protocol stack

The details of the Genve structure is shown below:

```
Geneve Header:
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |Ver|  Opt Len  |O|C|    Rsvd.  |          Protocol Type        |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |        Virtual Network Identifier (VNI)       |    Reserved   |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                    Variable Length Options                    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Ver (2 bits)
- Opt Len (6 bits): the length of the options fields
- O(1 bit): OAM frame
- C (1 bit): Critical options present
- Rsvd. (6 bits): Reserved
- Protocol Type (16 bits): Ethertype
- Virtual Network Identifier (VNI) (24 bits): An identifier for a unique element of a virtual network (e.g. a tunnel id)

It has a flexible options structure which can be tailored for all needs and is defined as follows:

```
0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |          Option Class         |      Type     |R|R|R| Length  |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                      Variable Option Data                     |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

It has a flexible options structure which can be tailored for all needs and is defined as follows:

- Option Class (16 bits): Namespace for the 'Type' field (allocated by IANA)
- Type (8 bits): Type indicating the format of the data contained in this option
- Length (5 bits)

- Variable Option Data

This protocol stack offers the following key characteristics:

Table B.1-3 Key Characteristics of Geneve/UDP/IP

Feature	Geneve
Message Type	Yes (NOTE).
Length	In IP header
Protocol multiplexer	Yes: 16-bit payload identifier (Ethertype).
payload Type	Supports any payload.
Tunnel multiplexer	24-bit VNI field
Sequence Number	Yes (NOTE)
Checksum	In UDP header
QoS transport marking	DSCP in outer IP header (TS 36.414)
5G QoS marking	Yes (NOTE)
Carried over	UDP/IP
End Marker in HO	Yes (NOTE)
Transport overhead	IP + UDP + Geneve header + options header
U-plane possible without tunnelling	Yes
NAT Traversal	No

NOTE: As Geneve is designed to be extendable, any number of options (e.g. message type, 5G QoS, etc) can be easily added.

Annex C Change history

Change history							
Date	**Meeting**	**TDoc**	**CR**	**Rev**	**Cat**	**Subject/Comment**	**New version**
2016-04	RAN3#91 bis	R3-160947				Included Editor's notes. Reflected agreed text proposals from RAN3#91bis (R3-160989, R3-161005, R3-161006, R3-161008, R3-161010, R3-161011, R3-161013).	0.1.0
2016-05	RAN3#92	R3-161381				Editorial updates.	0.1.1
2016-05	RAN3#92	R3-161442				Editorial updates (on referencing). Moved Annex A to section .6.3.1.1 and Annex B to section 6.1.2.1. Reflected agreed text proposals from RAN3#91bis (R3-161009, R3-161012) and RAN3#92 (R3-161296, R3-161344, R3-161444, R3-161449, R3-161453, R3-161485, R3-161486, R3-161495, R3-161503, R3-161504, R3-161506, R3-161507, R3-161508, R3-161509, R3-161510).	0.2.0
2016-08	RAN3#93	R3-161687				Added the purpose of this TR in section 1.	0.3.0
2016-08	RAN3#93	R3-161962				Reflected agreed text proposals from RAN3#93 (R3-161867, R3-161972, TP from R3-161973for TR is agreed only the sentence not the new section, R3-161975, R3-162049, R3-161978, R3-161979, R3-161980, R3-162011, R3-161966, R3-162056, R3-162057, R3-162058, R3-162032, R3-162034, R3-162059, R3-162062, R3-162051, R3-162063, R3-162005, R3-161995, R3-161997, R3-161998, R3-162077, R3-162053, R3-162054, R3-162060, R3-162078).	0.4.0
2016-10	RAN3#93 bis	R3-162526				Reflected agreed text proposals from RAN3#93bis (R3-162102, R3-162546, R3-162174, R3-162355, R3-162175, R3-162528, R3-162589, R3-162590, R3-162594, R3-162624, R3-162574, R3-162575, R3-162625, R3-162577, R3-162579, R3-162597, R3-162596, R3-162598, R3-162629, R3-162628, R3-162529, R3-162531, R3-162533, R3-162534, R3-162535, R3-162541, R3-162542, R3-162544, R3-162545, R3-162631, R3-162634, R3-162549, R3-162618, R3-162635, R3-162619, R3-162592, R3-162623, R3-162626, R3-162627, R3-162630, R3-162633, R3-162647).	0.5.0
2016-10	RAN3#93 bis	R3-162527				Restructured the TR according to the new TR structure proposal which was endorsed in R3-162525.	0.6.0
2016-11	RAN3#94	R3-163034				Editorial updates (fixed numbering of figures and tables, fixed section numbering referred in texts, added missing change history comments).	0.6.1
2016-11	RAN3#94	R3-163178				Reflected agreed text proposals from RAN3#94 (R3-162686, R3-162687, R3-162865, R3-162955, R3-163106, R3-163107, R3-163108, R3-163109, R3-163110, R3-163112, R3-163113, R3-163114, R3-163115, R3-163117, R3-163118, R3-163119, R3-163120, R3-163121, R3-163168, R3-163170, R3-163173, R3-163174, R3-163176, R3-163177, R3-163179, R3-163180, R3-163214, R3-163217, R3-163221, R3-163222, R3-163248, R3-163249, R3-163250, R3-163251, R3-163252, R3-163253).	0.7.0
2016-12	RAN#74	RP-162255				Presentation to RAN for information (no change in contents compared to v0.7.0).	1.0.0
2017-1	RAN3 Adhoc	R3-170313				Some editor's note are deleted. Reflected agreed text proposals from RAN3 Adhoc (R3-170178, R3-170237, R3-170285, R3-170199, R3-170135, R3-170268, R3-170270, R3-170317, R3-170272, R3-170274, R3-170275, R3-170318, R3-170319, R3-170320, R3-170280, R3-170322, R3-170282, R3-170323, R3-170284, R3-170298, R3-170324, R3-170325 , R3-170290, R3-170327, R3-170296, R3-170299, R3-170307, R3-170308, R3-170306, R3-170302, R3-170305, R3-170332, R3-170314, R3-170321, R3-170333, R3-170328, R3-170329, R3-170297, R3-170330, R3-170334). Editorial updates(fixed numbering of figures, Font format changes…)	1.1.0
2017-2	RAN3#95	R3-170744				Reflected agreed text proposals from RAN3#95 (R3-170798, R3-170799, R3-170800, R3-170801, R3-170548, R3-170422, R3-170854, R3-170856, R3-170424, R3-170863, R3-170865, R3-170516, R3-170717, R3-170872, R3-170795, R3-170797, R3-170853, R3-170821, R3-170822, R3-170851, R3-170852, R3-170855, R3-170900, R3-170901, R3-170902, R3-170899, R3-170876, R3-170906). Editorial updates(fixed numbering of notes, fixed section numbering)	1.2.0

2017-03	RAN#75	RP-170490				Presentation to RAN for approval (Editorial updates from v1.2.0, Capturing missed update in R3-170899.).	2.0.0
2017-03	RAN#75					TR is approved by RAN plenary	14.0.0