D5.2: Preliminary Report on Multi-Connectivity and Software Defined Network Control

Revision: v.1.0

Work package	WP 5
Task	Tasks 5.1 and 5.3
Due date	31/06/2024
Submission date	01/07/2024
Deliverable lead	Benjamin Barth, Roshith Sebastian (DLR)
Version	1.0
Authors	Benjamin Barth, Roshith Sebastian (DLR), Marius Corici, Hemant Zope (FHG), Nuria Candelaria Trujillo Quijada (HSP), Fanny Parzsz (ORA)
Reviewers	Mehdi Ait Ighil (TAS-F)
Abstract	This deliverable provides a preliminary assessment of the feasibility of integrating Terrestrial and Non-Terrestrial Network for Multi-connectivity scenario using Access Traffic Steering, Switching, Splitting (ATSSS) and a Software-defined Networking (SDN) management to control the data path across the transport network.
Keywords	Multi-Connectivity, Software Defined Network Control, ATSSS, MPTCP, MPQUIC

Co-funded by the European Union 6GSNS
Grant Agreement No.: 101096573 Topic: HORIZON-JU-SNS-2022-STREAM-A-01-02
Call: HORIZON-JU-SNS-2022 Type of action: HORIZON-JU-RIA

Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.1	29/01/2024	1st edit, draft ToC	Benjamin Barth
V0.2	19/02/2024	ToC Update	Benjamin Barth, Marius Corici
V0.3	15/04/2024	Multi-Connectivity testbed initial version	Roshith Sebastian, Benjamin Barth
V0.4	14/05/2024	ATSSS contribution, Intro update, Annex added	Benjamin Barth
V0.5	23/05/2024	ATSSS architectural investigations	Benjamin Barth
V0.6	17/06/2024	Final version for review	Marius Corici, Roshith Sebastian
V0.7	22/06/2024	Internal review	Mehdi Ait Ighil
V0.8	25/06/2024	Responding to comments	Marius Corici, Roshith Sebastian
V0.9	28/06/2024	Final editing	Roshith Sebastian
V1.0.F	01/07/2024	Final review and approval for EC portal submission	Tomaso de Cola

DISCLAIMER

Co-funded by the European Union

5G-STARDUST (*Satellite and Terrestrial Access for Distributed, Ubiquitous, and Smart Telecommunications*) project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101096573.

This work has received funding from the Swiss State Secretariat for Education, Research and Innovation (SERI).

COPYRIGHT NOTICE

Project co-funded by the European Commission in the Horizon Europe Programme		
Nature of the deliverable:	**R**	
Dissemination Level		
PU	*Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page)*	✓
SEN	*Sensitive, limited under the conditions of the Grant Agreement*	
Classified R-UE/ EU-R	*EU RESTRICTED under the Commission Decision No2015/ 444*	
Classified C-UE/ EU-C	*EU CONFIDENTIAL under the Commission Decision No2015/ 444*	
Classified S-UE/ EU-S	*EU SECRET under the Commission Decision No2015/ 444*	

* R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

DATA: Data sets, microdata, etc.

DMP: Data management plan

ETHICS: Deliverables related to ethics issues.

SECURITY: Deliverables related to security issues

OTHER: Software, technical diagram, algorithms, models, etc.

EXECUTIVE SUMMARY

The deliverable will introduce the preliminary achievements with respect to multi-connectivity (NTN/TN) within an SDN-based architecture for 5G-STARDUST project, which focuses on seamless integration of Terrestrial and Non-Terrestrial networks.

Access Traffic Steering, Splitting and Switching (ATSSS) has been taken as the baseline to establish multi-connectivity to Terrestrial and Non-terrestrial Networks. Although it is possible to connect to multiple TNs or to NTNs, the main focus here is to connect User Equipment (UEs) TN and NTN simultaneously. The legacy ATSSS approach supports dual connectivity by connecting to a 3GPP access and a non 3GPP access. Hence, in this work the current definition has been extended to realise connectivity to multiple 3GPP accesses. The control plane aspects to realize this has been evaluated and the necessary enhancements have been suggested. The architectures considering direct and indirect access, via IAB-like nodes have been discussed with attention given also to the protocol stack adaptations. The advantages that the use of AI/ML can bring into the optimization process has been considered in the architectural design. The deliverable includes the initial steps in the development of a test bed to verify the suggested solutions for Multi-connectivity.

For being able to control the data path across the transport network an SDN based management is proposed which enables the transmission of the routing related information prior to their usage, with this being able to immediately adapt to topology changes.

As this deliverable is a preliminary report, the findings and solutions after the completion of the tasks will be reported in the final deliverable.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

3GPP	3rd Generation Partnership Project
5G NR	5G New Radio
5GC	5G Core
AI	Artificial Intelligence
AMF	Access and Mobility Management Function
AnLF	Analytics Logical Function
ATSSS	Access Traffic Steering, Splitting and Switching
BAP	Backhaul Adaptation Protocol
CLI	Commnad Line Interface
CN	Core Network
CoMP	Coordinated Multi-Point
COTS	Commercial off-the-shelf
CU	Central Unit
DN	Data Network
DU	Distributed Unit
DVB	Digital Video Broadcasting
FR	Frequency Range
FRR	Fast Reroute
gNB	5G Node B
GEO	Geostationary Earth Orbit
GS	Ground Stations
GTP	General Packet Radio Service (GPRS) Tunnelling Protocol
HTTP	Hypertext Transfer Protocol
IAB	Integrated Access Backhaul
IE	Information Element
IETF	Internet Engineering Task Force
IP	Internet Protocol

ISL	Inter-Satellite Link
LDP	Label Distribution Protocol
LEO	Low Earth Orbit
MA	Multi-Access
MAI	Measurement Assistance Information
MAR	Multi-Access Rule
MC	Multi-Connectivity
MP	Multi-Path
MPLS	Multiprotocol Label Switching
MPTCP	Multi-Path Transmission Control Protocol
MPQUIC	Multi-Path Quick User Datagram Protocol Internet Connection
ML	Machine Learning
MTLF	Model Training logical function
N3IWF	Non-3GPP Inter-Working Function
NAS	Non-Access Stratum
NOC	Networks on Chip
NF	Network Function
NTN	Non-Terrestrial Network
NWDAF	Network Data Analytics Function
OSPF	Open Shortest Path First
PCF	Policy Control Function
PDU	Protocol Data Unit
PLMN	Public Land Mobile Network
PM	Path Manager
PMF	Performance Measurement Function
PoC	Proof of Concept
PSA	PDU Session Anchor
QoS	Quality of Service

RAN	Radio Access Network
RFC	Request for Comments
RRC	Radio Resource Control
RTT	Round-trip time
SDN	Software Defined Networks
SIM	Subscriber Identity Module
SLA	Service Level Agreement
SMF	Session Management Function
SNMP	Simple Network Management Protocol
TCP	Transmission Control Protocol
TCO	Total Cost of Ownership
TLS	Transport Layer Security
TM/TC	Telemetry/Telecommand
TNGF	Trusted Non-3GPP Gateway Function
TN	Terrestrial Network
UDM	Unified Data Management
UDP	User Datagram Protocol
UDR	Unified Data Repository
UE	User Equipment
UPF	User Plane Function
URSP	UE route selection policy rules
UTs	User Terminals
QoS	Quality of Service

1 INTRODUCTION

This deliverable reports the preliminary outcomes of two tasks of the 5G-STARDUST-project: Task 5.1 – "Multi Connectivity Solutions" and Task 5.3 – "Software Defined Network Control". As the other task from WP5 both are dealing with network and transport layer aspects. The work presented is building on the architecture and baseline solutions elaborated in WP3 in Deliverable D3.1 and D3.2 [1], [2]. The architecture consists of a terrestrial 5G system and a 5G-Non-Terrestrial Network (NTN) on a Low Earth Orbit (LEO) constellation which are either offering connectivity in transparent or regenerative mode, i.e. hosting a gNB on-ground or on-board satellites, respectively. Both are complemented by a Geostationary Earth orbit (GEO) satellite. User Equipments (UEs) can connect either directly to terrestrial or space based base stations or indirectly using integrated access backhauling (IAB).

For multi-connectivity Access Traffic Steering, Splitting and Switching (ATSSS) as specified by 3GPP is the baseline architecture. Generally, multi-connectivity means the usage of at least two parallel connections to increase the Quality of Service (QoS). Usually the main aim is either to increase throughput or resilience. One link can be used as back-up in case the connection with the main access is lost, also called switching, or they can be used complementary. In this case, traffic can either be split over the two links (usually based on a defined threshold) or, depending on the QoS flow, steered towards a specific link full-filling the QoS demand. The benefits and selection of the setup depend on the characteristics of the available links. For instance, in a typical GEO NTN-TN parallel topology the NTN link is considered as back-up of the TN due to its higher availability and worse latency. If the links are similar in throughput, latency and jitter, they can be used in parallel increasing the throughput, while if one link has higher latency, arriving packets must be reordered slowing down the overall connection.

In this deliverable we use the Software Defined Networks (SDN) concept to develop a mechanism enabling transport level routing solution for mega-constellations to significantly cut routing convergence time and reduce processing demands on space nodes, by providing centralized computed routing information to the space nodes prior to its usage and by enabling an automatic switching of the routing tables without inter-node communication.

The document is organised as follows:

- Section 2 discusses the architectural options for multi-connectivity following the baseline scenarios of the project;
- Section 3 presents the preliminary test-bed design for multi-connectivity investigations;
- In Section 4 Software Defined Networking-based routing mechanism is presented;
- Section 5 concludes the document and provides an outlook the second iteration of this deliverable.

2 MULTI-CONNECTIVITY ARCHITECTURES

Two baseline use cases for 5G-STARDUST's proof of concept have been selected [3], one of them, connecting UEs on-board of an airplane (or ship), is used in the following to discuss the options for multi-connectivity (MC). Nevertheless, the presented architectural options and investigations can be generalized and apply also to other use cases.

In the above-mentioned use case, MC is established by connecting to TN and NTN simultaneously. There are two radio frequency bands for 5G NR namely, FR1 and FR2. FR1 ranges from 410 MHz to 7125 MHz [4], while FR2 offers a wider range from 24.25 GHz to 71 GHz [5]. FR1 and FR2 have frequencies that can be used by both TN and NTN, allowing devices to connect to either networks [6]. On connecting to both frequencies UE can use FR1 for TN connectivity and FR2 for NTN or vice versa. This will help in reducing interference making the MC solution more efficient and the system more robust.

In Deliverable D3.1 the 5G-STARDUST system architecture was introduced. UEs can connect to TN or NTN for ubiquitous coverage. NTNs are either transparent in line with 3GPP Release 17 standard or regenerative following work in progress of Release 19. Furthermore, UEs can connect directly to the terrestrial and non-terrestrial networks accessing the gNBs or indirectly, by an gNB which is connected via IAB. Several options for MC on different layers of the communication stack have been discussed in Deliverable D3.2 [2]. As baseline ATSSS was selected, which is using Multi-Path TCP (MPTCP) or MPQUIC to connect to a non-3GPP access technology and to a 3GPP access with a common 5G Core (5GC). It is basically an adaptation of MPTCP and MPQUIC, compliant with the 5GC which allows to terminate MC at the PDU Session Anchor (PSA) User Plane Function (UPF), i.e. the user serving UPF. The specification includes also an ATSSS lower layer part for Ethernet traffic. Using MPTCP and MPQUIC to directly connect to data network (DN) on top of the 5G system can be considered as baseline for multi-core setups.

For a better overview the available links are shown again in Figure 2-1 for direct access and in Figure 2-2 for indirect access, respectively. UEs or IAB[1] nodes can connect to a GEO satellite, potentially multiple LEO satellites, potentially multiple terrestrial links, or eventually only a specific sub-set of these. The UE illustrated here can also be a group of UEs. An example following the airplane scenarios is: a group of users on-board of an airplane connect either directly or indirectly using an IAB-node on-board the plane. Close to ground, terrestrial links can be used in parallel to NTN links. While after take-off in high altitudes and above the oceans, only NTNs are available. Each of these links has different characteristics in terms of QoS parameters such as throughput, latency, or jitter that need to be considered and might affect the performance.

[1] IAB nodes are taken here as reference here mostly for acting as relay nodes. However, the full protocol architecture of IAB nodes is not considered in this report, not will it be in the forthcoming phases of the 5G-STARDUST project. As such, it is about IAB-like nodes. Under this assumption, the attributes 'IAB' and 'IAB-like' will be used interchangeably throughout the entire document.

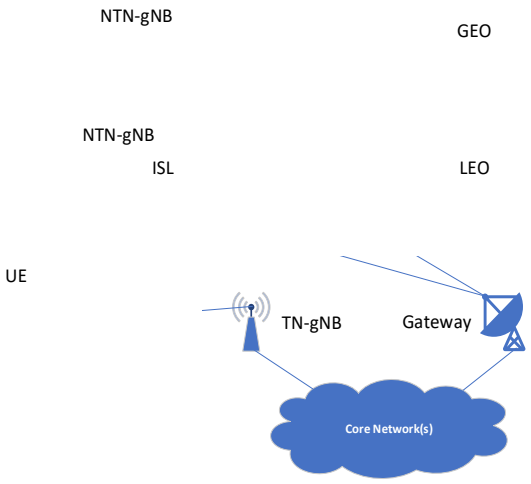

Figure 2-1: MC scenario direct access [2]

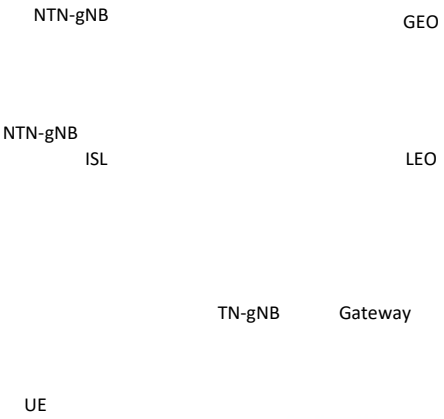

Figure 2-2: MC scenario indirect access [2]

With direct connectivity the MC can be started and controlled by the UE or even the user application, while with indirect connectivity, the UE connects to a gNB that is backhauled by multiple links (TN and NTN) which is not possible so far by standardised system components. We discuss in the following sections several options and extensions required to enable this beyond the current standard.

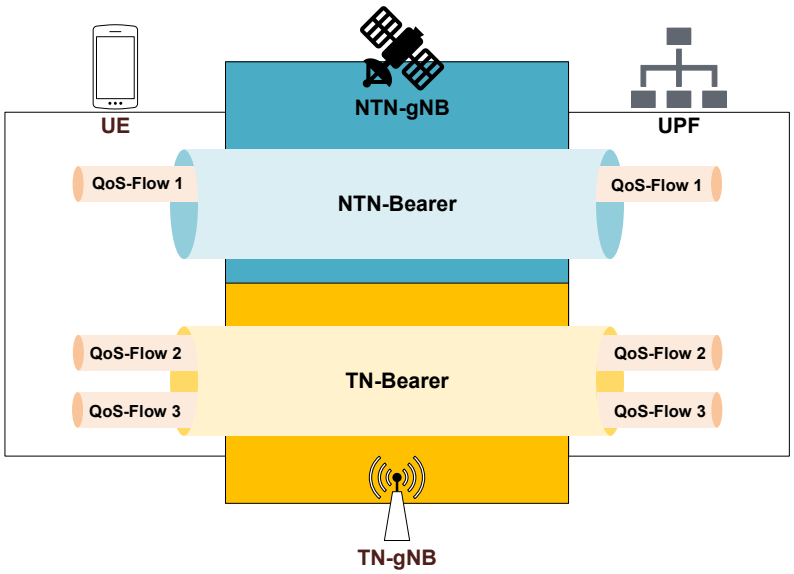

Figure 2-3: MC bearers direct connection

Figure 2-3 shows the bearers for the baseline case. The UE has access to the NTN and TN with a dedicated bearer. The UE can open for applications various QoS-Flows at each bearer and distribute the traffic accordingly using the multi-path (MP) protocol. As mentioned, the traffic can be split, steered or switched. The figure exemplary shows three QoS-Flows but it is not limited to this. The QoS flows reach via the NTN- and TN-gNBs the UPF which terminates the MC and reorders packets, if needed. With information about the network performance, QoS

flows could be scheduled according to QoS profiles from applications. Characteristics of NTN and TN bearers might be different in terms of delay, packet loss, or jitter. Additionally, a user might prefer one of the bearers due to resource consumption and costs, e.g., there might be different cost for NTN or TN services or different service level agreement (SLA) with the providers. In principle, QoS flows can be established based on application level profiles such that flows with low requirements on latency can select a GEO path while those having a more demanding requirements can use LEO or TNs.

Figure 2-4 and Figure 2-5 show two different possibilities for the indirect connection from bearer perspective. In Figure 2-4 the first option is shown, the IAB-node transparently forwards the NTN and the TN bearer to the UE. The UE can establish and control QoS-flows as in direct communication case. From implementation perspective the IAB-node must be able to handle and forward different bearers, or simply the IAB functionality is duplicated, i.e., an IAB-node for each connection supported. But this scales with the number of bearers and providers that, e.g., at the airplane use case, might be multiple ones hosted at the NTN or TN system. In the second case presented in Figure 2-5, the IAB-node provides a dedicated IAB bearer to the UE and is backhauled by the NTN and TN bearer. Two options can be considered: (i) the UE handles the MP-connection as in ATSSS, (ii) the IAB nodes handles the MP connection. In first case, UE and IAB node must coordinate the different bearers and how the QoS-flows should be mapped to them at the IAB. In the second this can be optional, since the IAB could establish MC and provide its benefits transparently to the UE.

It must be noted that the NTN/TN bearers have been selected as example to simplify the description. However, in the principle can be applied generally for other types of connections and for all link combinations, for instance two TNs or two NTNs (e.g., two LEOs or a LEO and a GEO).

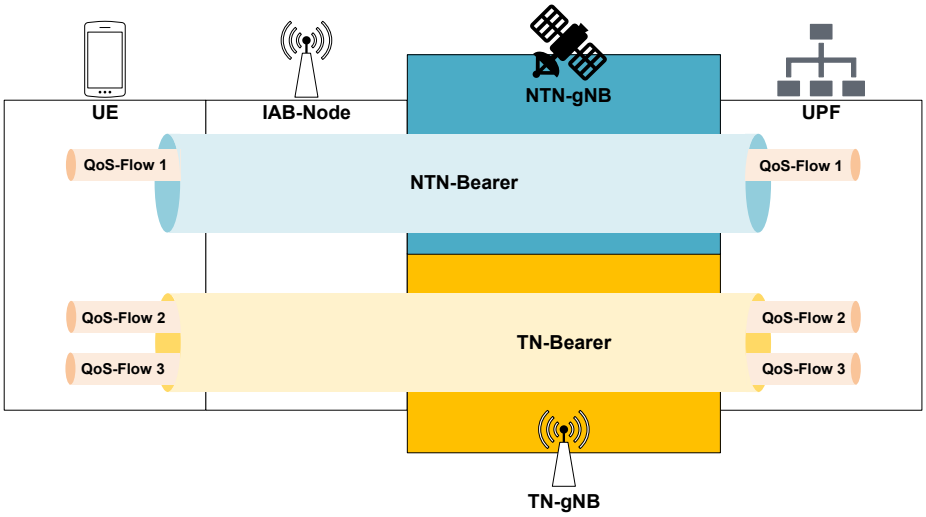

Figure 2-4: MC bearers indirect connection with transparent IAB

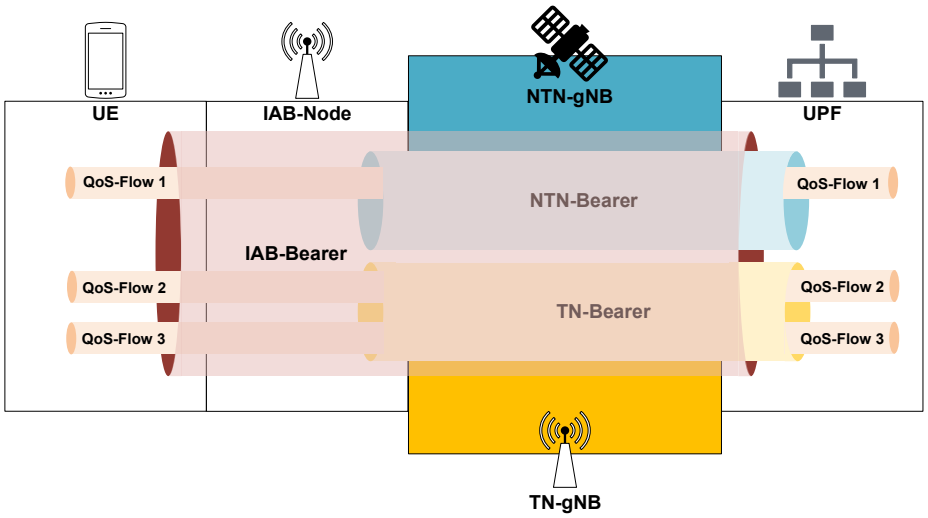

Figure 2-5: MC bearers indirect connection with IAB bearer

In all cases the core must be able to handle the different QoS flows, which are, as mentioned, parallel 5G connections. To be able to control these data bearers and to allocate the appropriate resources a comprehensive 5G control plane has to be deployed. The control plane is described in Chapter 4 of this deliverable.

2.1 ACCESS TRAFFIC STEERING, SWITCHING AND SPLITTING

ATSSS is an optional part of the 3GPP standards to connect a non 3GPP technology and a 3GPP technology to the 5GC, described in 3GPP TS 24.193 [7] and TS 23.501 [8]. The Non-3GPP Interworking Function (N3IWF) and the Trusted Non-3GPP Gateway Function (TNGF) are used for this. Since 5G-STARDUST is considering the 5G NTN in parallel to TN, the ATSSS functionality must be extended in order to allow two or even multiple 5G-bearers connected to one UPF. The functionality of N3IWF or TNGF are not needed in this case.

ATSSS establishes a Multi-Access (MA) PDU from the UE to the UPF. It supports ATSSS-Lower Layer, MPTCP or MPQUIC. MPTCP and MPQUIC steer traffic above TCP/IP and UDP/IP, respectively, while ATSSS-LL is below IP layer supporting Ethernet traffic. The UPF is including a MPTCP or MPQUIC proxy functionality. It shall be noted that MPQUIC was added to the standard in Release 18 and is referring to the current draft version of MPQUIC, since a standardised version by the IETF does not exist at the moment of writing this deliverable. On the user plane, there can be traffic on both of the connections at the same time, only on one of them, or on none of them keeping the connection open. MA-PDU session are established based on UE route selection policy rules (URSP) specified in 3GPP TS 24.526 [9]. As explained in the following section on the control plane process the setup of the connection depends on the capabilities of the UE, in principle plain MPTCP or MPQUIC are also supported.

MPTCP is standardised by the IETF and is an extension to TCP. Basically, it builds on top of it and opens several TCP sub-flows. One design goal was that it can be used without changes to legacy applications [10]. The standards comprise of IETF RFC 6182 presenting the MPTCP architecture, RFC 6356 dealing the congestion control, RFC 6897 with application considerations, and finally RFC 8684 gives the necessary TCP extensions.

Similar to MPTCP, MPQUIC is the multi-path extension to QUIC, currently only available as draft version at IETF in which it is also referred in ATSSS [11]. QUIC is a transport protocol specified by the IETF in May 2021 as RFC 8999 – RFC 9002. It is the specified transport protocol for HTTP/3. Implementations of the core QUIC specs are available and QUIC is already widely deployed to provide web content across the Internet. QUIC is connection oriented and runs on top of UDP, it handles the same functionality that is performed usually by TCP (e.g., congestion control). According to [12], QUIC provides the following features:

- Multiplexing of separate objects using streams identified by a connection ID
- Mandatory security establishment via TLS1.3 extensions embedded within QUIC, decreasing the initial hand-shake procedure to 1-RTT and 0-RTT
- Avoidance of Head-of-blocking effect by managing retransmissions
- Enforcement of congestion control, reordering and error recovery at application level (this aspect in particular allows the possibility of protocol updates independently from operating system through an update of the web clients
- Definition of different priorities among objects and pushing of objects
- Connection migration to a new network with new IP address and/or port while the connection remains up and running

2.1.1 Control Plane Process

In order to support MA PDU sessions, the AMF, the SMF, and the PCF must be extended to be able to handle the MA-requests. Its then on the AMF to inform the SMF about the two connections available, the SMF sets up the UPF, accordingly, using the N4 interface and a Multi-Access Rule (MAR). In order to introduce parallel 5G connections the specification must be extended on this regard.

The MA-PDU session establishment is a two-way hand shake: the UE is sending the MA-PDU session request, from the AMF it receives the PDU SESSION ESTABLISHMENT ACCEPT message including the ATSSS container information element of the Non-Access Stratum (NAS). The connection and the dedicated user plane resources can then be considered established. ATSSS allows the UE to register in the same Public Land Mobile Network (PLMN) or even via two different PLMNs, consequently it covers already cases in which the NTN-Provider would offer its own PLMN in parallel to the TN-Provider, or in which both share a PLMN. In case both connections use the same PLMN, the MA-PDU session request can be sent via either one of these links, which is implementation specific. If the UE is connected to two PLMNs, it shall send the MA-PDU session request sequentially; the link selected first is implementation specific. In a third, special case, if only one connection is available at the point in time the UE wants to establish the MA session, then it sends the request via the network available and via the second if it becomes available.

For transmission, it is on the UE to decide how the traffic must be distributed, but during the MA PDU session establishment it can receive ATSSS rules from the core that must be applied. In general, the decision can be based on performance measures that can vary based on the UEs capabilities. One option is that the SMF provides the UE with Measurement Assistance Information (MAI) which can be used for default-QoS rules. If the UE is capable it can also use measures available at MPTCP or MPQUIC layer applying the non-default QoS rules. Alternatively, feedback of the current link status can be provided by a performance measurement function (PMF). A protocol for message exchange between UE and UPF is

specified. It should be noted that in TS 24.193 section 8, timer values are specified for PMF measurements at UE and UPF after which the measurement might be canceled, most of the timers are set to 1s which might lead to unwanted cancellations since especially in GEO higher values can be expected, not on average, but in extremes cases. ATSSS rules and MAI can be updated by the SMF. On the other hand, the link conditions, especially in LEO, can vary a lot due to hand-overs and the change of elevation of the satellites. A frequent update might be necessary to make efficient use of the available links.

In ATSSS, steering refers to selecting for the user traffic, the best service for a flow with linked QoS-type. Switching means a handover to the second link in case of interruptions. Splitting allows to distribute the traffic between both links for load balancing. According to the standard [8], five modes are specified for the MAR and the ATSSS rule:

- active-standby: one access is defined as main access, the other as active-standby. In case of unavailability, traffic is steered towards the stand-by access.
- smallest delay: the access network with the smallest round-trip time is used.
- load balancing: the flow is steered across both the 3GPP access and the non-3GPP access, with a given percentage. It supports autonomous or UE-assisted operation.
- priority based: the UE uses the access with high priority unless it is congested or unavailable, then the traffic is split over both the access networks
- redundant: traffic is duplicated on both access networks. Optionally a primary access can be selected in which case the primary must be used for all packets, the secondary may be used to send the duplicate packets.

If the UE changes the default steering parameters to adjust its steering based on its own decision it can use:

- Autonomous load-balance indicator: the UE applies its own distribution in percentages for the load-balancing case to maximize bandwidth.
- UE-assistance indicator: the UE applies its own distribution in percentages for the load-balancing, considering its internal states (e.g., battery state).

2.1.2 Controlling Multi-Connectivity by Artificial Intelligence and Machine Learning

The ATSSS baseline allows for introducing the use of artificial intelligence (AI) and Machine Learning (ML) for traffic control in two options: first is to make use of the fact that the UEs control the uplink traffic while it is on the PSA UPF, to control the downlink. An AI/ML controller can use the congestion state available at these nodes and traffic predictions to control the end-to-end flows from and to the UE to satisfy the QoS needs. Second option is to make use of an AI/ML controller to define and update MAR and ATSSS rules directly from the core. This provides a possibility to centrally control UE behavior and traffic distribution within the network. Here the optimization goal is on general network traffic distribution since multiple UE connections can be considered. At this point it shall be noted that the use of AI/ML also for MC is investigated in Task 5.2 and preliminary results are presented in D5.3 [13]. Within this task, findings from these investigations are taken into account in the MC testbed presented in section 3. This testbed is following the first option to optimize single UE traffic distribution.

It should be noted that another option to make use of AI/ML for optimization is to place it in the UPF functionality that terminates the MP-traffic, e.g., this could be a satellite of the constellation which has at the moment a good link to the ground network, or a UPF located in the ground network.

2.1.3 Architecture Adaptations and Protocol Stacks

As mentioned already, in D3.2 [2] high-level options for the architecture have been presented and discussed preliminary based on the 5G-STARDUST architecture and use cases. Taking ATSSS as reference architecture, the technological principle is to introduce a MC-layer into the protocol stack which is enabling the benefits of MP. In the following, we separated the layer into MP-High representing MPTCP or MPQUIC and MP-Low since both use one or more TCP or QUIC connections, respectively. At 3GPP, MC functionality is not considered so far for indirect connections, i.e., backhauling a base station from different paths. Nevertheless, the presented architectural options and topologies provide different possibilities to place the MC-layer where we expect different benefits. In the Annex we present the complete protocol stacks for each of these topologies and, for a better overview again, the high-level architectures and the single connectivity cases. The stacks show only the user plane part since (as presented in the previous section) the control plane stacks stay unchanged, registrations are done sequentially. The topologies can be classified as direct and indirect connection (i.e., IAB), for both we considered the UPF either on-ground or on-board the satellite. For indirect case, three cases can be distinguished, the MC-layer can be at UE, at the IAB-node, or at an additional UPF which is included specifically for this layer.

1) Direct Connection

   a) with MC-layer at UE: this is the baseline ATSSS having the MPTCP/MPQUIC at UE and UPF (Annex 0);

   b) with MC-layer at UE, MP-termination at UPF in space: 5G-STARDUST is looking into regenerative and transparent payload and, as discussed in D3.2, if the MC-layer is implemented as in reference ATSSS UE-UPF PSA connection there is no effect of this option. On the other hand, if not only the gNB but also a UPF is provided as payload of a satellite, MP can be terminated at this layer, providing single path connectivity from there on. This only makes sense if multiple paths are opened towards the NTN, connecting to two or more LEOs. At the same time the feeder link can be off-loaded. This approach is deviating from the standard, since in principle relaying MC traffic is specified but MP traffic is terminated at the PSA-UPF, so at the end-to-end path of the 3GPP system. The corresponding protocol stack and topology can be found in Annex 0.

2) Indirect Connection

   a) with MC-layer at UE: the baseline ATSSS case considering IAB, combining both options. This case was not introduced in D3.2, for completeness we present it here. The MC-layer is implemented at UE and core side as in ATSSS case, however, the difference is that the UE is connecting to one IAB-node (as introduced in Figure 2-4). As seen in Figure 5-10 in the Annex, the UE and UPF PSA can directly communicate on top of the available infrastructure. But either the infrastructure is duplicated (two IAB-nodes), the IAB-node is extended to transparently forward multiple gNB connections (e.g., the TN and NTN-gNB), or the IAB-node is extended to communicate with the UE to configure the MP links. This is needed to efficiently organize the communication for which the UE should be aware that there are multi-paths it can make use of, and the IAB node must be aware of how to distribute the traffic among its backhaul links.

b) with MC-layer at UE and MP-termination at UPF in space: this case is similar to 1b) with the difference of the indirect connection having an IAB node in-between. From protocol stack, it can be seen as a merge of 1b) and 2a) as presented in Annex 0. As for both cases the IAB-node must be extended to enable the MP connection and for an efficient use of it to allow coordination between the UE or the core side with the IAB-node to control the traffic flows and it must be allowed that the MP traffic is terminated at a relaying UPF.

c) with MC-layer at IAB-Node: enabling the bearer split as presented in Figure 2-5. This is a special case moving the MC-layer from UE-UPF to IAB level. This approach has been selected for the 5G-STARDUST proof of concept (PoC) initially introduced in D6.1 [14]. We are describing the stack and some related considerations in the following in more detail.

d) with MC-layer at Forwarding UPF function: in this topology, an UPF is introduced right after the IAB-node. From use case perspective, this would mean that UEs, the IAB-Node and the UPF are hosted all on-board of an airplane or a ship. This is making the satellite network being part of the Transport Network, following classical backhauling setups. For the SatCom links, 5G-NR or any other access technology such as DVB-SX could be used. The actual UE 5G communication runs on top of it. Consequently, this topology is allowing the backhauling of indirect communication but is not in line with the 5G-STARDUST architecture having the NTN as part of the access network. For the sake of completeness, we are still presenting it here.

As mentioned, in the Annex, the protocol stacks for all considered Multi-Connectivity (MC) topologies are presented that have been derived from the 5G-STARDUST baseline use cases and architecture, D3.1 and D3.2. The illustration is following the color-code presented in Figure 2-6.

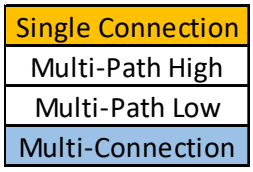

Figure 2-6: Protocol stacks, color-coding

It means that orange is single connections which is split, switched or steered by the MP-High (MPH) and MP-Low (MPL) layer colored in white. The blue connection is the MC which means that these parts of the protocol stacks must be available at least twice, e.g., on TN and NTN path. The topologies show always an NTN and a TN path, but the stacks generally apply to other combination of paths (e.g., NTN-NTN). The general illustration approach is that if moved down in layer the upper layer is encapsulated within the lower layer. Finally, at physical (PHY) layer the actual transmission is done and the stack is decapsulated up to a certain point. In order to further forward traffic it needs to be encapsulated again for the next hop which might follow another stack. Communication links are illustrated as lines between the protocol blocks. A layer can communicate with its corresponding part on sender or receiver side; it must be decapsulated properly to this point. Figure 2-7 shows exemplarily the stack for the baseline case 1a). UE connects to the gNB via the 5G New Radio (NR) stack either via TN or NTN. The blue color-coding is indicating that there can be two gNBs. The gNB is connected on the Transport side to the UPF-PSA which connects to the data network. The UE opens the MA-PDU session and opens the MPH connections which are either MPTCP or MPQUIC. From there on, its two TCP/QUIC connections building on the 5G-NR stacks below it.

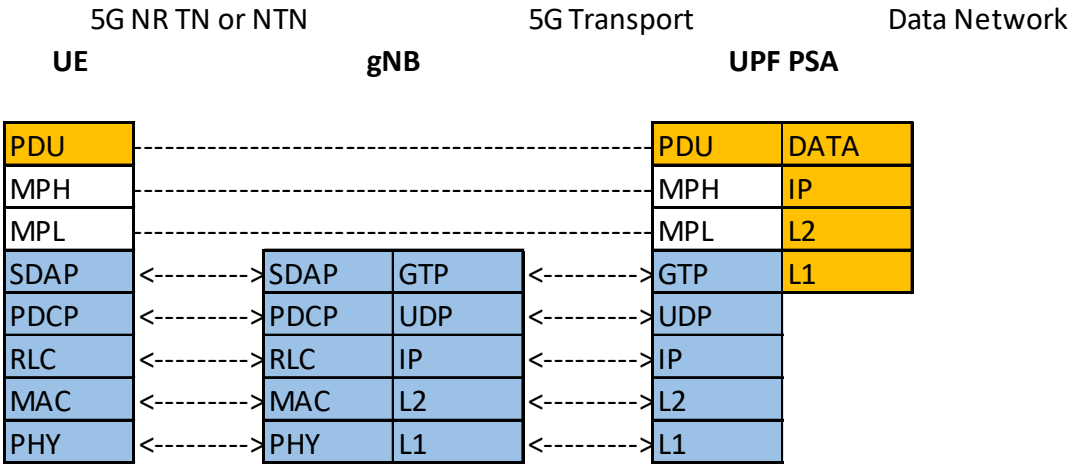

Figure 2-7: Protocol stack, direct connection with MC-layer at UE

In this example the MC-layer is at the UE. In Figure 2-8 an example for the second option, having the MC-layer directly at the IAB-node is presented. This is one possible option identified from the single connection protocol stack (cf. Annex 0, Figure 5-3) in which a UDP connection is established between the Distributed Unit (DU) of the IAB-Node and the Central Unit (CU) of the IAB-Donor. This connection is established below the GTP connection in the stack. This is a first intuitive solution aiming to minimize changes of the single connection stack to simplify potential implementation updates and keeping most part of the original stack unaffected. However, it must be pointed out that this approach is turning a connectionless link into a connection-oriented one, provided by features of TCP and QUIC which might introduce unwanted latencies due to re-transmissions and flow-control. Other options are shown in Annex 0. A way to circumvent this is to introduce extensions such as specified in IETF RFC 9221 [15] which enables unreliable data-flows of QUIC. However, with MPQUIC not specified is not clear if such extension could be included.

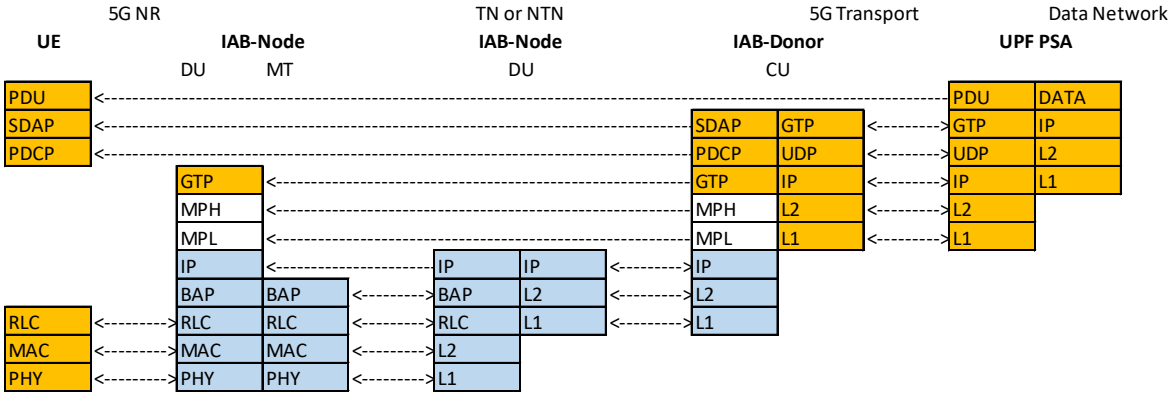

Figure 2-8: Protocol stack, Indirect Connection with MC-layer at IAB-node

The presented stack is the direct result from taking the standard IAB stack. From stack side there is no difference if the IAB-Donor is considered to be at the NTN gateway following e.g., a transparent approach or if it is part of the satellite payload in a regenerative approach. However, for actual implementation having an IAB-Donor on-board a satellite opening multi-paths via NTN and TN might introduce additional challenges, it might be preferable for NTN-NTN paths.

It is important to note that if the MC-layer is introduced in this way, at the IAB-layer, it cannot be terminated at UPF, neither in space nor on ground, since the encapsulation order will be broken. This can be seen in Figure 2-8, there is no option of introducing a MC-layer that is able to communicate with the UPF, all protocols communicate with the IAB-donor. Layers above the GTP of the IAB-node's DU are from the UE preventing an introduction. Consequently, if a MC-layer needs to be terminated at a UPF it must be initiated by the UE or another UPF as shown in the other topologies. MC-layer that has been introduced beyond the standard at IAB must be terminated at this level.

A similar approach has been followed in the 5G-STARDUST PoC design presented in D6.1 [14] and presented in Annex 0 Figure 5-15. It is on the IAB-node to establish and control the MP-connection. Also, optionally communication between the UE and the IAB node would be beneficial to optimize the use of both links. QoS parameters of traffic flows could be exchanged, but also completely switching between two links makes the end-to-end connection more robust. The stack was presented in D6.1 and is shown in the Annex in Figure 5-16. The stack includes two UPFs which are introduced due to a general problem which can be observed on introducing a MC-layer at an indirect node. IAB is basically an introduction of a gNB before a gNB, i.e., chaining base stations. If the MC-layer is introduced at the first, UE-serving gNB, it breaks the GTP connections making it necessary to introduce a second GTP layer at the UPF to ensure proper decapsulation (The Problem is illustrated in Figure 5-17 in the Annex.). This additional GTP block is breaking the standard 3GPP UPF stack, but it can be circumvented by introducing another UPF before the PSA-UPF. From use case side, this UPF could be the one from an NTN/TN provider while the UPF PSA is the UPF contracting the UE.

The protocol stack shown in Figure 2-8 introduces MPQUIC or MPTCP below GTP, i.e., encapsulating the GTP traffic. It is important to point out that MPQUIC is by default encrypting the traffic and headers making it impossible to access the GTP information before decapsulating at the MP-termination. This means that if the GTP information is needed along the path, e.g., for UPF chaining, it is better to introduce the MPH and MPL protocols before GTP. In this way, each path would have each own GTP tunnel.

2.2 CONTROL PLANE FOR MULTI-CONNECTIVITY

2.2.1 Introduction

Multi-access in systems up to 5G have been limited to support a 3GPP and a non-3GPP connection, multi-5G connectivity being implemented in the RAN through solutions like 5G Coordinated Multi-Point (CoMP) or secondary RAN. However, with the advent of multiple frequency ranges and 5G Non-Terrestrial Networks (NTN)-Terrestrial Networks (TN) convergence, multi-access across two independent accesses of the same 3GPP technology type has become a new high capacity and reliable communication opportunity where RAN solutions like CoMP cannot be realized as the two RANs may pertain to different subnetworks requiring signalling and data path aggregation at core network level. To answer to this opportunity, we present a new connectivity concept and 5G architectural advancements required to achieve this type of multi-access, with focus on TN-NTN, although the principles and architecture would apply to other scenarios briefly mentioned as well. Following a background analysis of the multi-connectivity concept in 5G system, we propose the necessary architectural additions to support multi-access across two independent 5G connections. We assess the feasibility of the solution by considering the additional functionality to be added. This study demonstrates that despite its complexity, such a solution is feasible for 5G deployments and can be easily implemented in the beyond 5G and the foreseen 6G architecture.

Enabling multi-connectivity on devices represents an effective way to improve data rates and improve communication reliability. Whether having two connections to base stations of the same 3GPP technology or by using a secondary more cost-effective but less reliable non-3GPP access, multi-connectivity enables balanced network resource management. Through multiple connections, user traffic can be distributed across various paths, optimizing resource utilization based on applications' requirements. This dual connectivity ensures that even if the secondary connection is less reliable, it can handle less critical traffic, reserving the primary, more robust connection for essential data.

Seizing these advantages, several options for multi-connectivity on various layers have been summarized in [16], while 3GPP has standardized three mechanisms for multi-connectivity. The first is deeply embedded in the 3GPP RAN systems, enabling connections to multiple base stations of the same type, where the base stations collaborate to schedule the device's data traffic – 5G Coordinated Multi-point (CoMP). Second, dual connectivity/multi-connectivity is defined at 3GPP bearer level ([17]) with one gNB used as master the other as secondary solution. The master gNB is the entity communicating with the core handling the control plane and aggregating the data bearers, looking like a single gNB towards the core network. Last, 3GPP has also standardized a multi-radio system that enables connections across a 3GPP access network and a non-3GPP one, balancing data traffic between a high-reliability link and a cost-effective one.

However, with the adoption of multiple frequency ranges, the potential use of 5G in unlicensed or privately allocated spectrum, the loosing of network reliability requirements for specific frequencies and the integration of the 5G NTN networks, the same 5G radio technology is deployed in systems with very different communication characteristics and requirements. As a result, multi-connectivity using two heterogeneous and independent 5G access networks should be also considered. These systems being separately administrated, will not enable direct interfacing between the RANs, making direct RAN collaboration like CoMP, secondary RAN or the new 6G user-centric architecture ([18]) unfeasible. Additionally, coordination may be impractical due to significantly different RAN coverage areas, ranging from square meters for high-frequency radio to hundreds of square kilometers with frequent handover for NTN.

To enable multi-3GPP connectivity without deep RAN coordination, there is a need to extend the standard architecture. We address this challenge by introducing a new concept for multi-3GPP connectivity and analyzing the architecture to provide specific enhancement details. Since interoperability is not feasible at the RAN level, our proposed multi-connectivity relies on the device and core network support. Our proposed solution is based on the core network Access Traffic Steering, Switching and Splitting (ATSSS) feature ([19]), further adapted for the specific target access networks. Although this level of convergence presumes independent RAN operations, it significantly minimizes the interactions between different administrative domains, making it effective and more likely to be deployed in the beyond 5G and 6G networks.

To evaluate the feasibility of the proposed solution, we assess the additional features required as software additions needed to be developed on top of the Fraunhofer FOKUS Open5GCore toolkit ([20]), a reference software implementation of the 3GPP 5G core network standards able to handle both 3GPP and non-3GPP connectivity. This assessment provides insight into the necessary additions and implications of adopting such a solution. Based on this evaluation, we introduce a set of considerations for the foreseen 6G network architecture, enabling an overall architectural and signaling simplification.

2.2.2 Use Cases and Requirements

The multi-3GPP connectivity addresses the use cases that cannot be covered by centralized RAN solutions or deep RAN coordination such as CoMP. These scenarios involve multiple 3GPP RAN deployments with heterogeneous characteristics, covering the same device area,

able to provide complementary communication services to user devices. Below, we present an initial list of such use cases, highlighting the importance of this feature.

1) TN-NTN Interoperability – to have the most gains from 5G NTN deployments, a seamless handover mechanism to terrestrial networks is required. As 5G NTN is highly dependent on the line of sight which can be suddenly interrupted and the handover procedure through the core network incurs significant delays due to the NTN, maintaining dual connectivity is the only effective solution. This approach ensures no interruptions or significant jitter during handovers by splitting the user traffic based on link availability.

2) Public-Private 5G Interoperability – similar to the NTN use case, private 5G deployments in enterprises, hospitals, malls, or stadiums may offer additional 5G connectivity to devices complementing the public network. However, this connectivity may be limited in capacity and reliability. Therefore, a multi-connectivity solution for these multi-administrated domains should be considered to enhance device available bandwidth and overall service reliability. For example, this feature would enable private 5G networks not to deploy emergency services, relying on the public network one, reducing their Total Cost of Ownership (TCO).

3) Sub-Networks – often referred to as “network of networks”, the dividing of an end-to-end network into smaller, logically separated segments can enhance the network security by isolating administrative domains and limiting the attack spreading. Sub-networks can have their own QoS policies and access control specific to their administrative domains, optimizing local operations. In the sub-networks environment, the devices can improve their connectivity by connecting to multiple co-located 3GPP-based sub-networks. In this situation, only a high-level coordination of the multi-connectivity would be possible.

4) Unlicensed/low reliability 5G networks – the network operators may deploy 5G networks in either unlicensed or in spectrum where only low reliability SLAs would be required, to complement the existing high reliability network with additional capacity. This approach is similar to the current 3GPP-non-3GPP multi-connectivity use case but involves two 3GPP access networks. This enables operators to offload a large amount of the user traffic within their own network, drastically reducing the TCO by deploying more cost-effective networks. Devices maintain their high reliability services over the existing 5G network while connecting to the less reliable 5G network in parallel, resulting in multi-3GPP connectivity.

To realize these use cases, the network should be able to coordinate the user identity, its access control and the resource allocation across two parallel 3GPP connections. Additionally, the system should support separately coordinated handovers within each of the 3GPP connections and handle situations where one of the connections is unavailable. Furthermore, dynamic user traffic splitting depending on the momentary established sessions and on the link conditions should be considered to optimize communication across the two links.

2.2.3 Standardisation Background

3GPP CoMP enhances transmission and reception proactively managing the interference among the users increasing the spectral efficiency and the throughput especially at the cell edges. It is based on the coordination of the different base stations through direct exchanges using a direct backhaul interface between base stations through which Channel State Information (CSI) is exchanged [21] enabling joint transmissions, coordinated scheduling and coordinated beamforming.

CoMP demands high-capacity, low-latency backhaul links between the base stations and increased synchronization, making it impractical for the scenarios proposed.

Dual connectivity is a simpler method than CoMP, the latter requiring an interface between the two base stations. It splits user traffic into two independent RAN connections that are transparent to the core network. In this setup, the master RAN communicates with the core

network, while the secondary RAN only handles the second data path connection, merged through the interface between the base stations at the master RAN.

To support multi-connectivity between 3GPP and non-3GPP accesses, a third mechanism, named ATSSS was standardized by 3GPP ([19]), illustrated in Figure 2-9. Non-3GPP technologies are converging in the core network through a Trusted Non-3GPP Gateway Function (TNGF) or a Non-3GPP Inter-Working Function (N3IWF), both adapting the specific signaling to the 3GPP stack. Based on the type of access the Access and Mobility Function (AMF) is able to determine if the connection of the device is 3GPP or non-3GPP and to permit both of them to be separately established and separately executing horizontal handovers when needed. The Session Management Function (SMF) is in charge of establishing data bearers over the two connections while the Policy Control Function (PCF) is selecting the appropriate policies to be enforced for the specific user. All operations are executed using the subscription profile in the User Data Management (UDM), the same network function where also information on the current connectivity of the User Equipment (UE) is maintained.

To be able to split the user traffic between the two the UE and the User Plane Function receive a set of ATSSS policies respectively Multi-Access Rule (MAR) transmitted from the PCF. Based on the momentary status of the two connections, the user traffic is split either at the lower layers: ATSSS-LL with an explicit split based on the data bearers or at the Multipath TCP (MPTCP) [22] or Multipath Quick UDP Internet Connections (MPQUIC) [23] where the user traffic is split at the higher levels.

The user traffic can use both connections at the same time, one connection or none while keeping them both open. The split between the two links is executed based on the link performance measurements where the SMF could signal the UE with additional Measurement Assistance Information (MAI), or the UE could interact directly at the data path level with the Performance Measurement Function (PMF) in the UPF. Specifically, for the ATSSS, the non-3GPP link is considered less reliable and more cost-effective to carry the user traffic.

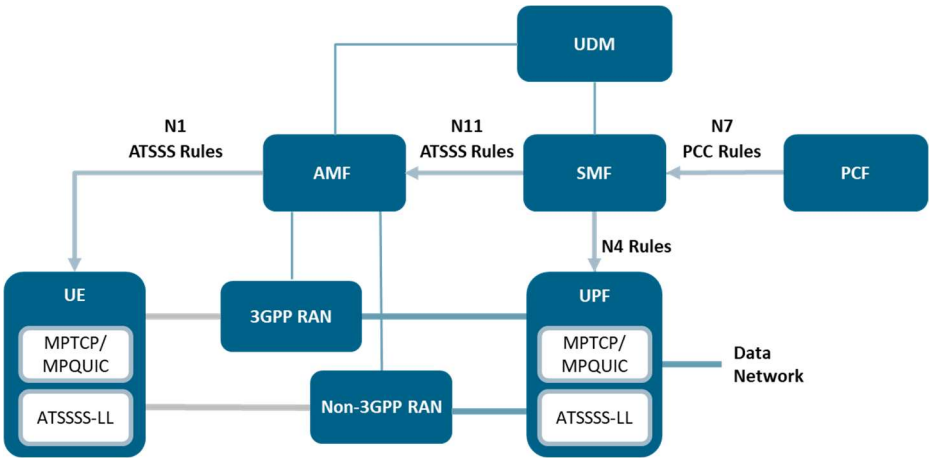

Figure 2-9 : ATSSS Architecture

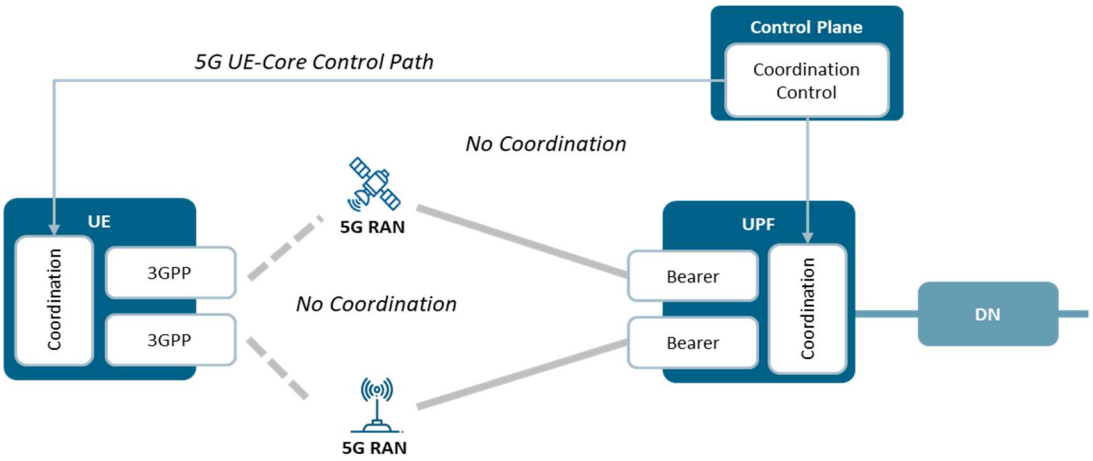

Figure 2-10: Multi-3GPP Connectivity Concept

2.2.4 Multi-3GPP Connectivity Concept

The multi-3GPP connectivity concept, as the name says, presumes that the User Equipment (UE) will be able to connect to two 3GPP accesses at the same time and establish bearers across them. Although the implementation seems straightforward, as illustrated in Figure 2-10, it is difficult due to the usage of the same identity across both links having the same subscription profile.

Considering that there is no coordination between the two RAN networks, the UE will have to register to both networks consecutively and gain access. Similarly, the sessions over the two links have to be established independently similar to the ATSSS.

In the current 3GPP situation, the initiation of a new connection over a 5G RAN will result in taking over the connectivity of the previous connection, executing a hard handover. For this not to happen in the core network, there should be a second differentiator of the established connection included in all the messages from the registration on.

Using the second differentiator, the core network is able to signal to the UE the policies how to treat each of the access networks, similar to the current ATSSS policies, however not split on technology.

However, this differentiator does not have a direct impact on the UPF where the information received is not about access networks but about data bearers. As the bearers are access independent, we assume that the same functionality as in the case of ATSSS will be used to signal the policies.

A critical element for the functioning of the multi-3GPP is the possibility to split and to gather the user traffic of the two links. This operation should be executed in the UPF where the two bearers will be associated with a single IP address towards the data network as well as in the UE. Here, the splitting of the user traffic is highly dependent on how many 3GPP stacks are implemented i.e., if the UE actually includes 2 separate UEs, one for each link, acting as single devices and an overlay split and switch layer.

2.2.5 Multi-3GPP Architecture

2.2.5.1 Required Architecture Extensions

To support a multi-3GPP connectivity it is essential that a new connection will not replace an existing connection as in the case of a handover. Specifically, a differentiator is required between the two connections to be able to split the user traffic, differentiator which has to be

known by the UE, AMF, SMF and PCF to be able to appropriately allocate the bearers according to the UE policies.

The same connection discriminator should be used for the ATSSS level in order to be able to define the appropriate user traffic split policies. However, the actual ATSSS functionality does not require other modifications and the policies distributed remains the same, the only open question being which connection is the main one and which will act as secondary for the signaling.

It should be noted that timer values specified for PMF measurements might be needed to be updated since especially in GEO higher values can be expected in extremes cases. On the other hand, the link conditions, especially in LEO, can vary a lot due to handovers and the change of elevation of the satellites. A frequent update might be necessary to make efficient use of the available links.

Furthermore, a new mechanism for handling the idle mode is required as well as for the service activations and de-activations, as the UE may use two, a single connection or none at specific moments of time.

Additionally, one of the connections may be lost quickly or for very short amounts of time, specifically for the satellite link when losing the line of sight to the currently used satellite or in case of terrestrial networks in areas with low number of base stations and using of higher frequencies such as roads through the forests. In these situations, in order to be able not to lose user traffic and to adapt fast to the situation, notifications should be sent from the link management towards the ATSSS in the user device as well as from the RAN towards the UPF to be able to appropriately redirect the user traffic for the very short duration of such interruptions.

2.2.5.2 Architectural Options

While most of the architectural requirements can be solved with an adapted parametrization of the existing ATSSS solution the critical element in implementing the architectural requirements is the capability of the user device to handle the UE functionality for two parallel 3GPP connections and the coordination between them. As such, there are two major architectural options for implementing the user device: using two different UEs coordinated between them, as illustrated in Figure 2-11 or using a single integrated UE as illustrated in Figure 2-12.

In the case of two UEs, the derivation of the keys and the establishment of the connections remains unmodified. Practically the UE is seen like a dual-SIM device connected to the same network with both of the UEs. In this situation, the problem is not the key derivation ([24]) as

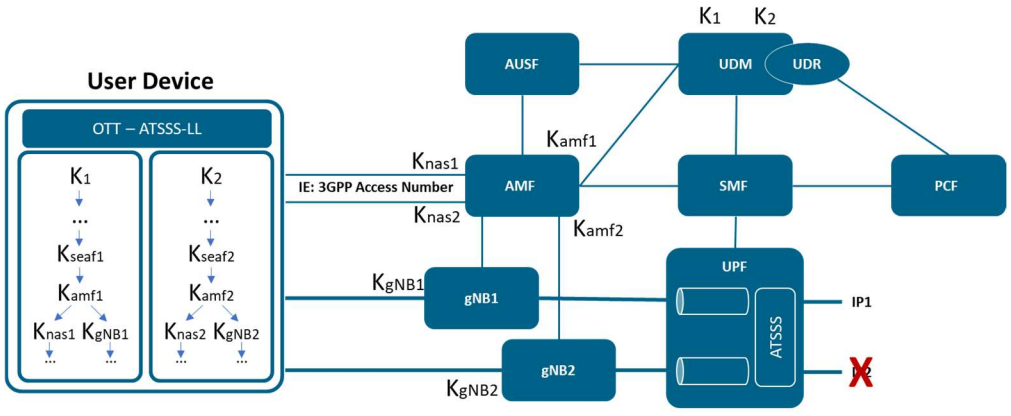

Figure 2-11: Dual UE User Device

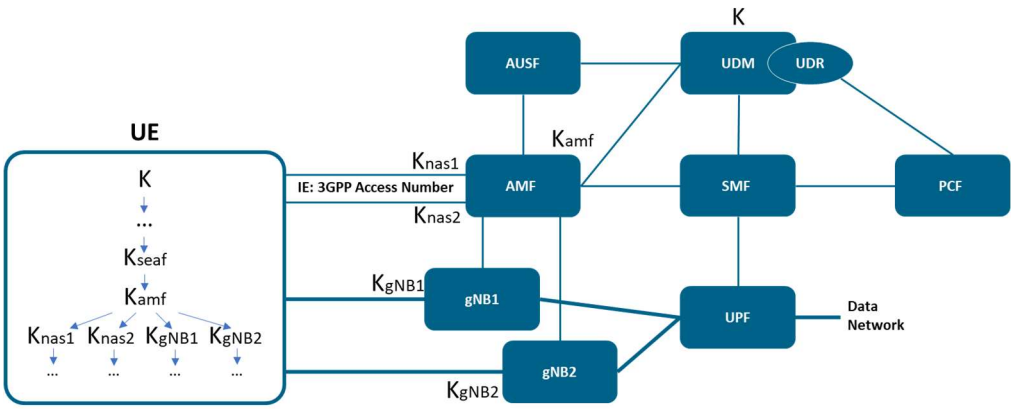

Figure 2-12: Single UE User Device

this is done independently by both of the UEs, but the matching of the user traffic as to enable the splitting across the two UEs. The key derivation follows the standard ([24]) from the main key down to K_{amf} followed by the derivations for control and data plane for each gNB (K_{gNB} and K_{nas}. In this situation, the ATSSS and the MAR policies transmitted to the UE have to consider a primary connection and a secondary one. For such policies the PCF should be aware that the two UEs are co-located into the same device, information which should be introduced within the UDR. Based on this, the PCF may generate specific ATSSS policies and the afferent MAR policies. As the ATSSS are transmitted transparently to the UE, there is no major issue. However, the MAR policies are more problematic as both of the UEs receive their own IP addresses (IP1 and IP2 in the Figure 2-11). In this situation a single IP address should be used towards the data network and the user traffic should be split across the bearers of the two UEs. For these, the same MAR policies should be used. The two UEs is simpler to implement considering the current system as it does not require major changes to the system. However, it will have independent tracking area updates and idle mode management, deemed unnecessary. And especially problematic is the dual SIM requirement for the devices which also implies the modification of the current phones.

In case of a single UE, the key derivation will remain the same down to the K_{AMF} as the authentication and authorization can freely flow between the two 3GPP accesses with having a single index increased based on the operations across both connections. This is in line with the current single UE which executes handovers between different 3GPP access networks. However, as the device is connected to 2 RANs, two distinct K_{gNB} have to be generated. In order not to take over the existing the second connection, a new Information Element (IE) has to be introduced in the NAS messages to enable the network to know this is a second connection and not a takeover of a previous one. If the new IE has more than one bits, virtually the number of 3GPP connections can be increased to more than two through this giving a large advantage to the current ATSSS solution which is strictly bound to two connections. Based on the distinct K_{gNB}s all the RRC and user plane encryption keys can be generated. In this situation, the UE should be able to maintain two K_{gNB} keys and to use them appropriately for the two connections.

However, for a solution to function, it should work also for the current UEs with no modifications. In this case, the network is taking over the functionality of determining which access it is used and to create the two bearers as to be ready in case of a handover with a secondary end-to-end bearer. This is a special case of a single UE connectivity where the network is keeping an idle mode connection over the unused 3GPP link. In this situation, the UE is not able to differentiate between the different access networks. The differentiator should come from the access itself for example NTN and TN through this making the AMF aware that the two 3GPP access networks should be run in parallel and not handed over. When the

second connection is established, through a handover, the first connection is not destroyed but kept in idle mode. In this situation, all the specific bearer information is maintained, however without any resources reserved over the wireless link. When the UE wants to make a handover to the second network after establishing the RRC connection, the user traffic can be immediately forwarded without requiring additional operations. Similarly, the downlink user traffic will be forwarded to the other link based on notifications from the access. As the notifications will not come from the PMF, the UE not having one, such notification should be transmitted through the handover commands. This way the handovers do not require a specific preparation phase being instead replaced by a fast takeover. A conditional handover mechanism can be employed to reduce the option that a handover command will not be received.

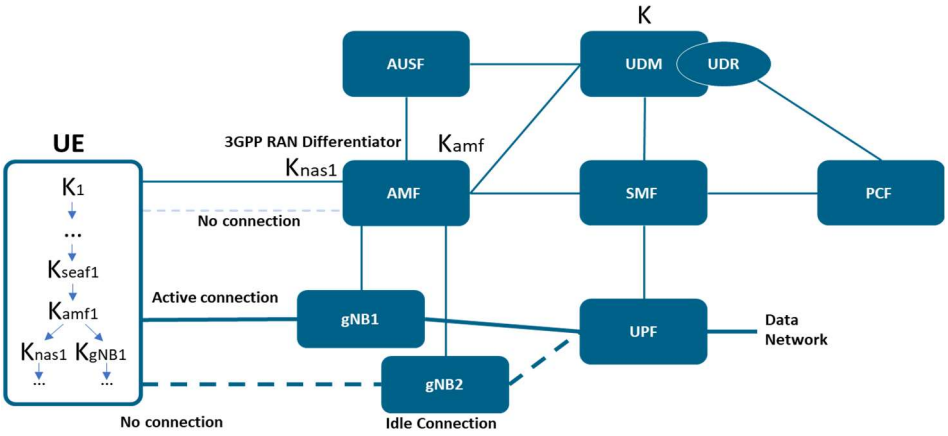

Figure 2-13: COTS UE

2.2.6 Implementation Feasibility and Complexity Assessment

In this section we assess the implementation feasibility of both the 2 UEs, 1 UE and not modified options and compare how complex it would be to become standard behavior and what are the implications on the user traffic. First, all the solutions require extensions of the core network, specifically for the policies transmitted to the UE and for the bindings between them. This is a complex feature which mostly impacts the PCF as the policy generator as well as the SMF and UPF for being able to properly treat the user traffic in the network. From this perspective, the easiest solution is to implement the single UE solution and to adapt the MAR policies that are now used for the 3GPP and non-3GPP accesses for the adaptation to the two 3GPP technologies.

However, the single UE solution proves to be highly complex to be implemented in the UE, requiring all the devices to support ATSSS as well as to be able to coordinate the authentication and cryptography across the two accesses, this being a limiting factor for the adoption of such solution.

More realistic would be to have an unmodified UE. In this situation, the communication service will be able to handover quickly between accesses as the second access will maintain an idle mode state for the UE. However, this will still require a handover time between the two accesses even though smaller than the current solution which may induce a longer delay in forwarding the downlink data due to the need to notify the UPF of the change as in the case of handovers. Also, such a solution maintains two states for the same device, and although one is idle, it still consumes more resources than a single connection with a hand over. More

evaluation and practical implementation are required to assess the opportunity of such complexity increase.

2.2.7 Conclusions and Further Work

In this chapter we have identified the main use cases and the architectural requirements for supporting two 3GPP connections at the same time for a device. This is an important feature in order to be able to support NTN-TN interoperability as well as for other systems where the 3GPP connectivity may have different reliability, security, or coverage characteristics and the 3GPP system cannot efficiently interwork such as private network-oriented setups. Starting from the ATSSS architecture, we have proposed three models on how the user device can be implemented resulting in different characteristics within the core network. These models were empirically analysed from the perspective of added complexity to the system and the viability of implementation starting from today's network architecture. A common 5G UE of today can significantly benefit from the supported mechanism as well as a device which implements ATSSS while a double-UE device would require a significant development of the UE as well as of the network.

3 PRELIMINARY MULTI-CONNECTIVITY TESTBED DESIGN

The testbed for Multi-Connectivity scenario will be developed in Mininet[2]. Mininet is a network emulator which creates a network of virtual hosts, switches, controllers, and links. In Mininet, it is possible to easily interact with the network using the Mininet CLI (and API), customize it, share it with others, or deploy it on real hardware which makes it useful for research, development, learning, prototyping, testing, debugging, and any other tasks that could benefit from having a complete experimental network on a laptop or other PC.

3.1 MININET OVERVIEW

Mininet hosts run standard Linux network software, and its switches support OpenFlow for highly flexible custom routing and Software-Defined Networking. Mininet provides a simple and inexpensive network testbed for developing OpenFlow applications. It enables complex topology testing, without the need to wire up a physical network. The CLI is topology-aware and OpenFlow-aware for debugging or running network-wide tests. It supports arbitrary custom topologies, and includes a basic set of parametrized topologies usable out of the box without programming. It also provides a straightforward and extensible Python API for network creation and experimentation.

Mininet provides an easy way to get correct system behavior (and, to the extent supported by the hardware, performance) and to experiment with topologies. Mininet networks run real code including standard Unix/Linux network applications as well as the real Linux kernel and network stack. Due to these advantages, the code developed and tested on Mininet, for an OpenFlow controller, modified switch, or host, can move to a real system with minimal changes, for real-world testing, performance evaluation, and deployment. Importantly this means that a design that works in Mininet can usually move directly to hardware switches for line-rate packet forwarding.

3.1.1 Virtualization in Mininet

Nearly every operating system virtualizes computing resources using a process abstraction. Mininet uses process-based virtualization to run many hosts and switches on a single OS kernel. Since version 2.2.26, Linux has supported network namespaces, a lightweight virtualization feature that provides individual processes with separate network interfaces, routing tables, and ARP tables. The full Linux container architecture adds chroot() jails, process and user namespaces, and CPU and memory limits to provide full OS-level virtualization, but Mininet does not require these additional features. Mininet can create kernel or user-space OpenFlow switches, controllers to control the switches, and hosts to communicate over the simulated network. Mininet connects switches and hosts using virtual ethernet pairs.

Mininet's code is almost entirely Python, except for a short C utility.

3.1.2 Advantages

Mininet combines many of the best features of emulators, hardware testbeds, and simulators.

Compared to full system virtualization-based approaches, Mininet:

[2] https://mininet.org/

- Boots faster: seconds instead of minutes
- Scales larger: hundreds of hosts and switches vs. single digits
- Provides more bandwidth: typically 2Gbps total bandwidth on modest hardware
- Easily installable

Compared to hardware testbeds, Mininet

- is inexpensive and always available
- is quickly reconfigurable and restartable

Compared to simulators, Mininet

- runs real, unmodified code including application code, OS kernel code, and control plane code (both OpenFlow controller code and Open vSwitch code)
- easily connects to real networks
- offers interactive performance

The limitation of Mininet is that it cannot exceed the CPU or bandwidth available on a single server.

3.2 THE BLOCK DIAGRAM WITH AI CONTROLLER

The block diagram for multi-connectivity setup with Machine Learning (ML) controllers is shown in Figure 3-1. The data transfer happens between node1 and node2.

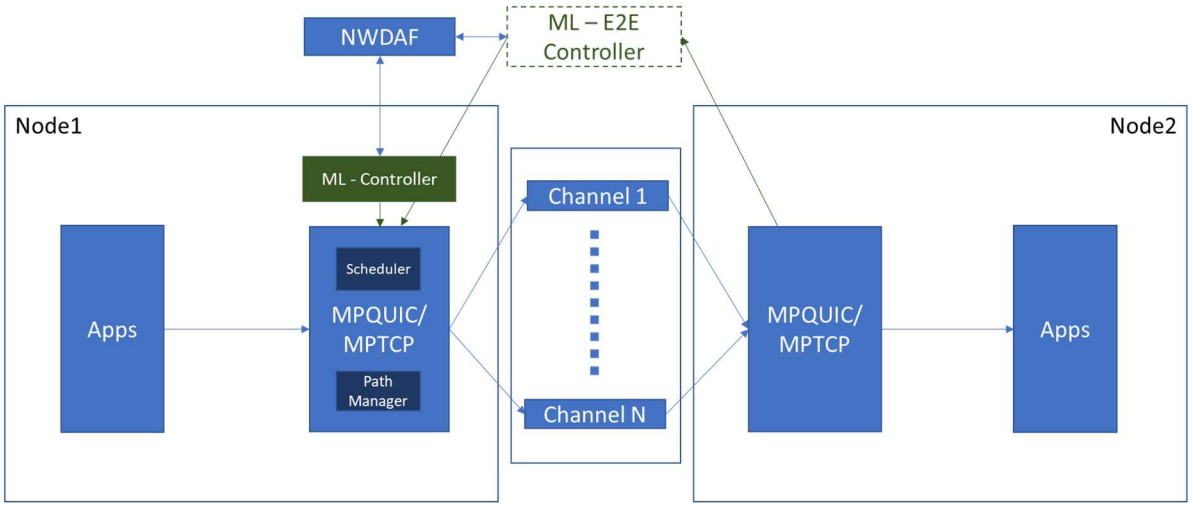

Figure 3-1 : Building Blocks in testbed design

The nodes are modelled as Ubuntu hosts with MPTCP module, which inherently supports multi-connectivity. It is possible in Mininet to realize the implementation of MPQUIC for further investigation on the Multi-Connectivity setup. After establishing a TCP connection, if it is possible to create additional sockets between these hosts a new subflow is created. For the

hosts to the use of MPTCP, the newly added TCP option field of the underlying subflow should be acknowledged by the remote host. If not, the connection will be downgraded to normal TCP and it will continue with a single path.

There are two internal components which helps in the realization of MPTCP:

- Path Manager

  The Path Manager creates new sockets in addition to the primary sockets for setting up subflows used for MC as shown in Figure 3-2. It is in charge of subflows, from creation to deletion, and also address announcements. Typically, it is the client that initiates subflows, and the server announces the additional addresses to create subflows. In our scenario, the subflows have to be established between TN and NTN. Upon establishing these connections, the apps in the nodes communicate with each other using the subflows.

  From Linux versions 5.19, there are two path managers: in kernel and in userspace[3]. If the PM in kernel is used, same rules are applied for all the connections. The application of different rules to each connection is possible only if the PM in userspace is used. It is possible to switch between these PMs by setting the net.mptcp.pm_type sysctl as “0” (kernel) or “1” (userspace).

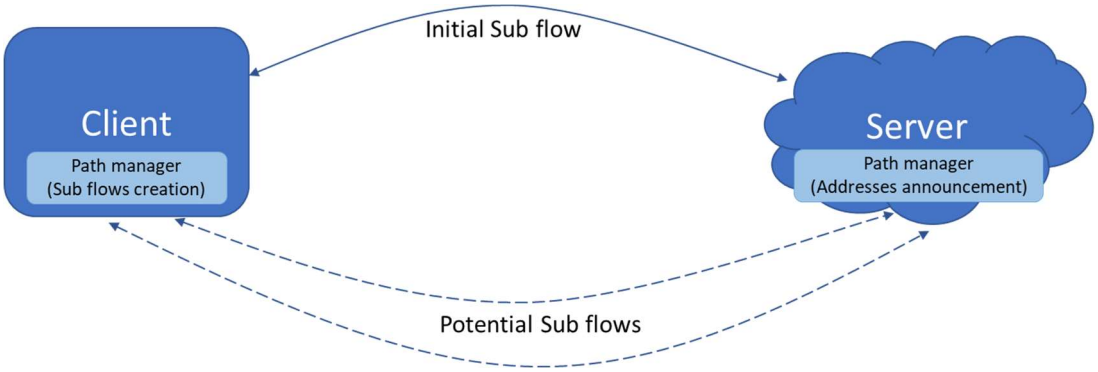

Figure 3-2 : Path manager

- Packet Scheduler

  The path scheduler decides on which available subflow to send the next available data packet as shown in Figure 3-3. To impose different rules for each subflow, it is essential to use the PM in userspace. Once it is set to use the PM in userspace, a daemon (i.e. mptcp) can be used to decide the rules for each connection. The decision can be based on available bandwidth, latency, and any other parameter defined by the scheduling policy. Since, these characteristics are different from TN to NTN, the Path Scheduler plays a vital role in the efficiency of the MC setup. The ML controller helps the Path Scheduler in this decision-making process. The overall throughput depends on the ML algorithm employed for optimization.

[3] https://www.mptcp.dev/

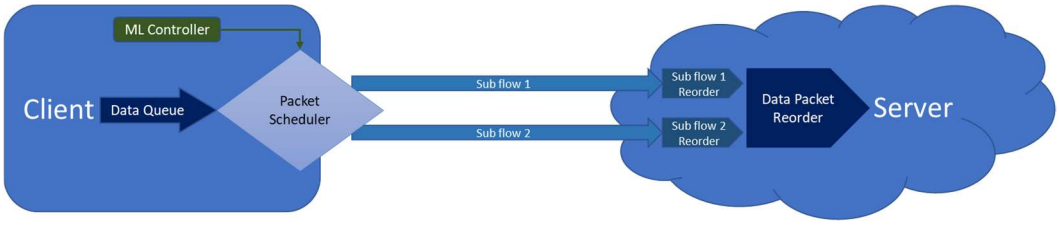

Figure 3-3 : Packet Scheduler and data packet reordering at receiver side

At the receiver end, the data packets are handled in two steps. The first step is to receive data at the subflow level, and reorder it according to the subflow sequence numbers. The second step is to reorder the data at the connection level by using the data sequence numbers, and finally deliver it to the application [25]. An end-to-end ML controller can be used to collect feedback from the receiver. This periodic feedback can also be used an input to the scheduler to decide on the subflow to send the subsequent packets.

NWDAF [26] also plays a major role in the optimization process. It is part of the 5G architecture specified in TS 23.501 [8] and has the interfaces to interact with various NFs to collect data for analytics and for ML inclusion. NWDAF contains the following logical functions:

- Analytics logical function (AnLF): Performs inference, derives analytics information (i.e. derives statistics and/or predictions based on Analytics Consumer request) and exposes analytics service.

- Model Training logical function (MTLF): Trains ML models and exposes new training services (e.g. providing trained ML model).

The ML controllers continuously interact with NWDAF to fetch the latest analytical data and ML models to improve the decision process. NWDAF also receives information from the ML controllers regarding the performance of the MC setup which can be used to update the analytics/statistics and the ML models.

4 SOFTWARE DEFINED NETWORK CONTROL

4.1 INTRODUCTION

The arrival of mega-constellations with inter-satellite links marks a major shift in space communication, evolving toward large-scale networking systems. This spurred significant interest in creating new routing mechanisms designed to tackle the unique challenges of space-based networks, characterized by predictable but high-latency and predictably changing links. In this deliverable we propose a Software Defined Networks (SDN)-based routing solution for mega-constellations, designed to significantly cut routing convergence time and reduce processing demands on space nodes. It relies on a centralized terrestrial controller that asynchronously computes routing information and deploys it to space nodes prior to its usage. These nodes are equipped with lightweight semantic routing capabilities that allows them to autonomously determine optimal routing based on their current network status. As any SDN solution, the design is flexible enough to accommodate various routing optimizations available in the literature, facilitating their practical implementation. This chapter 4 provides a detailed description of the proposed solution, assesses its feasibility using an example routing algorithm, and compares it with existing routing approaches, highlighting its strong potential for real-world commercial system application.

With the mass deployment of satellites in Low Earth Orbits (LEO), mega-constellations are revolutionizing broadband communication by offering efficient and robust connectivity service ([16]). The emergence of high-bandwidth space optical communication, new flat panel antennas, and direct-to-space connectivity for 5G phones allows space networks to compete with traditional terrestrial networks ([27]). Mega-constellations represent large-scale networks with 400 to 5,000 or even more satellites, often with low compute capacity, a large number of inter-satellite communication links, and intermittent ground connectivity to the ground ([28]).

Due to the constant movement of LEO satellites, the network topology changes frequently and predictably. Ground-space links face issues like atmospheric turbulence or heavy rain, weather-related disruptions and quality of link decrease, and limited line-of-sight for mobile-connected users, adding to the system's complexity ([29]).

Traditional routing protocols were designed to fit specific topologies, from Internet backbone and operator backhaul networks ([30]) to ad-hoc wireless networks ([31]) and Networks on Chip (NOC) ([32]). Over more than 50 years, these protocols have been refined to improve efficiency within their respective environments, continually increasing device connectivity and broadband capacity while enhancing service quality, resilience, security, and privacy. However, these same characteristics that drive their success make them unsuitable for immediate adoption in environments like mega-constellations, where the dynamic nature of these systems demands more adaptable solutions. To bridge this gap, various space routing algorithms were developed ([33], [45], [46], [47]) , but they lack deployment mechanisms for space nodes. In this respect, this chapter presents an SDN approach for deploying routing algorithms in mega-constellations, offering a new solution to manage the complexity of space-based communication networks. By using a centralized terrestrial controller, the proposed SDN framework allows routing configurations to be asynchronously and prior to their usage transmitted to space nodes, user terminals, and ground stations. This setup helps the network

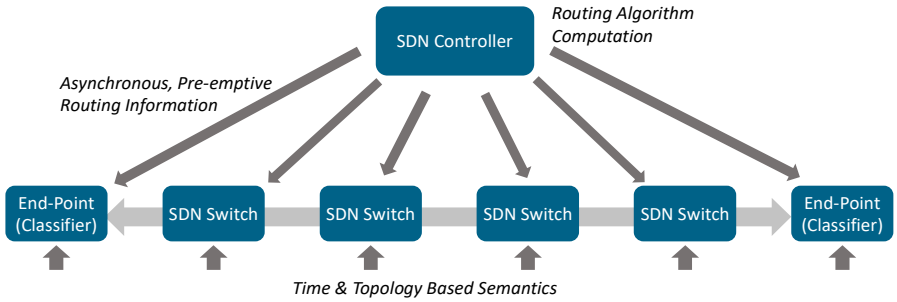

Figure 4-1 : SDN-Routing Concept for Mega-Constellations

adjust to topology changes without the need for any recalculations in non-exceptional situations, enhancing its efficiency and adaptability against increased precalculated information storage.

The SDN concept for mega-constellations we introduce here represents a generalization and adaptation of the new centralized routing decisions approach from the terrestrial networks ([34]) applying the general SDN approach to satellite networks ([35]).

Our proposed SDN concept is demonstrated using a basic Open Shortest Path First (OSPF) routing computations and a Multiprotocol Label Switching (MPLS) forwarding mechanism, showcasing its effectiveness compared to the equivalent traditional routing method ([36]).

Additionally, this chapter explores the feasibility of implementing this SDN solution, focusing on the same mega-constellation model and evaluating current software tools that can be adapted from terrestrial networks to space environments. This assessment also considers the latest developments in digital satellite payloads, underlining the alignment with ongoing compute units' developments for space.

The chapter concludes by comparing this solution to existing routing information exchange models, indicating that it holds the best potential for real-world commercial applications. The analysis demonstrates that the SDN-based approach not only streamlines the routing process using the particularities of space infrastructures, but also reduces the complexities associated with traditional routing solutions within intermediary routing nodes, making it a promising candidate for next-generation mega-constellations.

4.2 SDN ROUTING CONCEPT

Software-Defined Networking (SDN) is a modern approach to network management that decouples the control plane from the data plane, enabling centralized decisions, control, and programmability ([37]). In SDN, the control plane, which makes decisions about where data should be routed, is separated from the data plane, which forwards data to its destination. This separation allows network administrators to manage network behavior through a centralized controller, using software-based policies to control traffic flow, network configurations, and resource allocation. SDN's flexibility and programmability offer greater agility, making it easier to implement changes, optimize traffic, and deploy new services as well as to coordinate the information of the different sources within a single decision element removing potential side effects inherent to distributed decision systems.

In contrast, the traditional routing solutions rely on distributed control, where each router makes independent decisions based on routing tables and protocols such as OSPF and Label Distribution Protocol (LDP) ([36]). This approach requires complex algorithms computed at each network element to maintain consistent routing information across the network. In contrast, SDN uses a centralized controller to define and enforce routing policies, reducing the

complexity of network management. This centralization allows for quicker adaptation to network changes, simplified routing logic, and improved scalability, making SDN ideal for dynamic and evolving network environments, fact acknowledged also by the IETF and the latest terrestrial network routing protocol developments ([34]).

Adapting Software-Defined Networking (SDN) to space networks is challenging due to long distances between routing nodes and controller. Since nodes might be on the other side of the Earth from the controller, decisions propagate slowly, making real-time adaptation difficult. However, topology changes in space networks are regular and predictable, allowing for a pre-emptive approach to managing changes. Rather than using SDN to adjust the network in real time, our proposed solution uses it to distribute routing information ahead of time, before the topology changes happen.

The SDN approach here described and assumed in the rest of the project is based on the generic architecture from Figure 4-2 and involves several key operational steps. First, the SDN controller computes routing table information for all possible topologies within a mega-constellation.

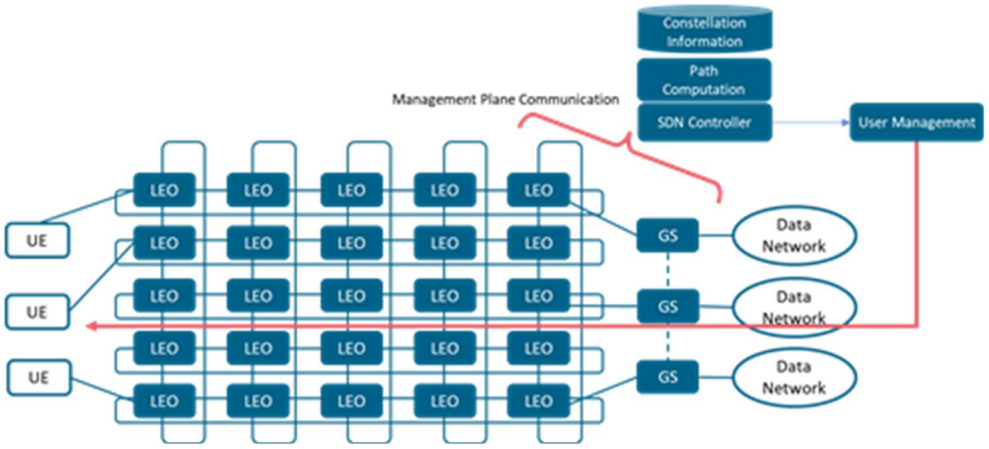

Figure 4-2 System View of the SDN Network

Since satellite orbits and their inter-satellite links are predictable, it's possible to anticipate the reached topologies. This step may take a very long time as it requires a large computation, considering that mega-constellations could have some thousands of satellites each with four or even more inter-satellite links However, when completed, it is not necessary to repeat it until new satellites are deployed.

Next, the controller uses SDN communication mechanisms to distribute this routing information to user terminals, space nodes, and ground stations, providing them with a set of potential routing tables to use when the topology changes.

When normal, non-exceptional topology changes occur, the space nodes select the appropriate routing table based on time or on the current network context i.e., when own links status changes. This decision-making is driven by the topology's status, such as which inter-satellite links are active at a given time.

Additionally, user terminals and ground stations have triggers to guide which routing table to use, based on the satellite or feeder link they are connected to. This method allows the endpoints to adapt their routing to the dynamic satellite network, following the most suitable paths.

By having all routing information preloaded, no complex algorithms are needed within the space nodes, which eliminates routing convergence time, critical especially because constellation topologies can change every two minutes, exceeding typical terrestrial routing

convergence. Moreover, this approach reduces the computational burden on space nodes, allowing them to focus on basic forwarding tasks. While the proposed solution demands substantial memory due to the large number of routing tables, this isn't a major concern, as satellite networks already include robust mass memory technologies, for many years optimized for Earth Observation applications.

Still, routing tables are not providing all the information for the nodes in many situations, when alternative routes have to be considered, for example in case of link failures or momentary link congestions. For these situations, each of the network nodes are enhanced with minimal semantic decision capabilities, enabling them to execute re-routing operations to other already installed routing rules to avoid the specific events. Also, this could be used for protecting feeder links as well as for assuring a congestion aware routing through the system, through multi-path load balancing.

4.3 SDN CONCEPT APPLICATION

In this section a mechanism to apply the previously described concept is described based on the typical terrestrial OSPF-MPLS deployment model, while comparing the two.

OSPF is a link-state routing protocol designed to manage network topology by distributing routing information among routers within an Autonomous System (AS), single administrated network similar to a mega-constellation. It uses a link-state database to maintain a map of the network and determine the shortest path for data packets, updating routes when network conditions change by executing a network wide collaborative shortest path Dijkstra algorithm. MPLS operates at a lower layer, allowing for efficient packet forwarding by assigning labels to packets, which guide them through the network without the need for complex routing table lookups. This technique reduces processing overhead and accelerates packet forwarding as it can ultimately be reduced to a matching of labels and their swapping. Within MPLS networks, the Label Distribution Protocol (LDP) distributes these labels among routers, ensuring a consistent label mapping across the network.

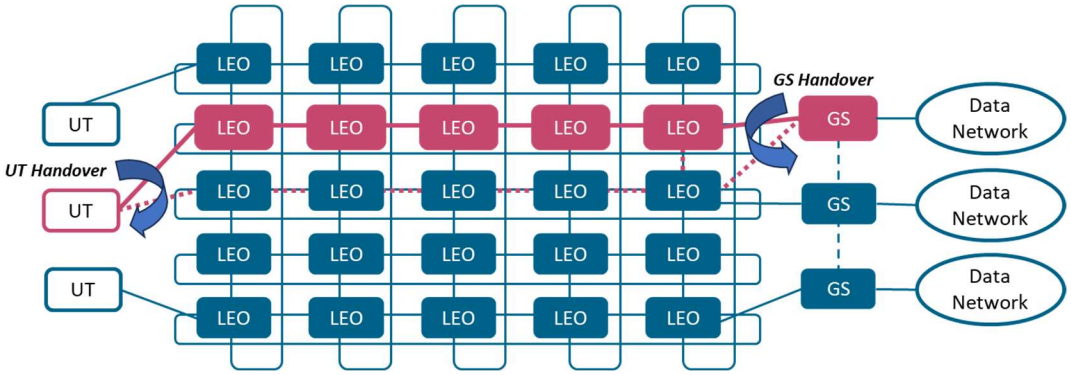

Figure 4-3 Transparent Handling of Handover procedures in the User Terminals and Ground Stations by using MPLS label swapping

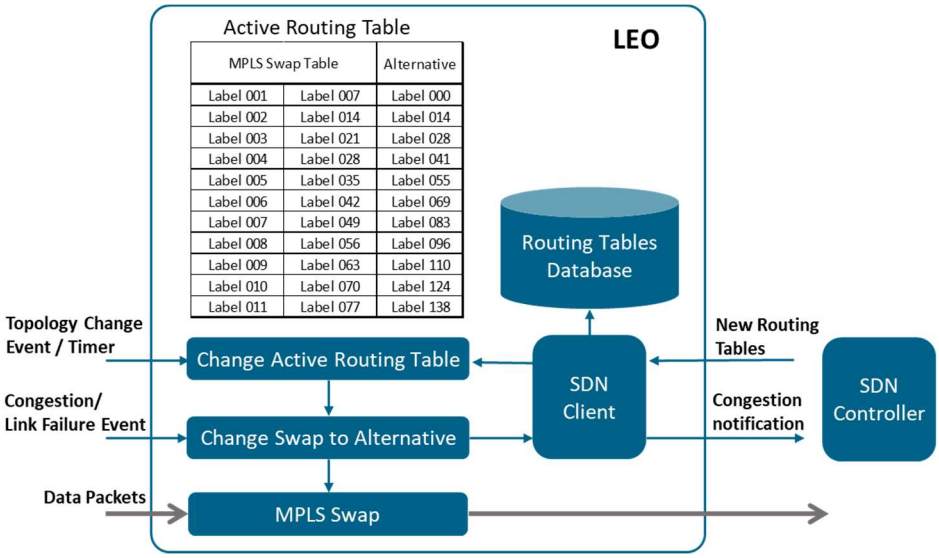

Figure 4-4 : Node View of SDN Processing

Together, these technologies create a cohesive system. OSPF provides the routing information and defines the network's logical structure, while MPLS simplifies packet forwarding based on labels. LDP handles the label distribution, ensuring that routers within the MPLS network understand how to forward packets based on the assigned labels. This combination allows networks to leverage the routing capabilities of OSPF while benefiting from the efficiency and scalability of MPLS, with LDP ensuring consistent label distribution across the network. This integrated approach is commonly used in large-scale networks to optimize routing, reduce latency, and improve overall network performance.

However, adapting OSPF to a mega-constellation poses several blocking challenges. OSPF relies on IP addresses to identify the nodes and more importantly to identify their network location. Due to the constant change of the satellites position in the network, this creates an issue: every time the network topology changes, IP addresses must be reallocated. During this reallocation process, no data can be transmitted, creating service interruptions.

Additionally, OSPF's convergence process requires the direct communication between nodes, to adapt the routing tables. This implies that after a topology change, each node must communicate with its neighbors to exchange routing information and update their routing tables accordingly. Given the grid-like structure of a mega-constellation, this propagation process can be slow due to the link delays causing a long convergence time. Even if IP addresses would be automatically allocated, the time OSPF synchronizes the routing tables will lead to misrouted data traffic.

To eliminate these issues, we propose centralizing the shortest path computation in the SDN controller, using the known constellation information to calculate optimal routing information. This method uses the Dijkstra algorithm to compute routes based on node identities and orbits, allowing for rapid execution without needing additional processing or communication between the network nodes. This centralized approach not only speeds up the process but also reduces the computational load on the space nodes, allowing them to focus solely on forwarding tasks.

However, information on the orbits and the links which will be automatically established is already available in the constellation (network) control center. We propose to use this information to compute the Dijkstra algorithm not only for the momentary topology, but for all topologies which could be reached. This requires a significant amount of computation. However, when the computation is done in advance, all the equivalent OSPF route computations would be already determined. Please note that such a computation is true for

any routing protocol chosen for the mega-constellation as the advantage of prior routing computations is drastically reducing the time needed and the computation of the nodes.

Since we assumed that nodes would not have IP addresses, we need a mechanism to embody the routing information. For this we use the MPLS label swapping as base line. To implement this, the SDN controller defines the labels for the various potential paths and allocate them to different nodes. Like in OSPF, this shortest path calculations can be done centrally in the SDN controller, significantly reducing the time required for information exchange, as all operations are conducted locally. This way, labels are assigned to all paths in the system, along with the corresponding swapping mechanisms.

This information is then transmitted to the various nodes through the management communication plane, allowing them to store the information in a routing table database.

As illustrated in Figure 4-3, the specific nodes will use this information to perform the MPLS label swapping during normal operations for all the data packets.

In an SDN network with a high latency to reach the controller as the proposed mega-constellation solution, a problem arises if a link fails. Furthermore, as the management plane shares the same links as the data plane, the SDN controller may be unreachable. To address this issue, we propose to add to the routing tables an MPLS Fast Reroute (FRR) extension ([38]), where each entry in the label swapping table has an alternative route label. This setup ensures the system can continue functioning even if a link fails, providing time for the controller to adapt the routing to accommodate the failure. This robust approach adds resilience to the system in case of unexpected link outages.

As mentioned before, all the routing tables and the backup alternatives for potential failures are pre-computed by the SDN controller and sent through the management plane before they are needed. When a topology change event occurs – whether due to ISL changes or a timer event the corresponding routing table is taken from memory and put as active routing table. This fast switch ensures that the network quickly adjusts to topology shifts, reducing the need for dedicated compute capabilities in the nodes.

Ground Stations (GSs) determine the path for data packets by adding an MPLS label to each packet. The selected label depends on the intended destination – it could be a space node, the one to which a user terminal is connected to or another ground station for backhaul traffic. To enable MPLS encapsulation we propose that each GS receives the routing labels assigned by the controller and use them to encapsulate the data packets appropriately.

User Terminals (UTs) present a different scenario, as they may only be accessible through the mega-constellation and have no role in their data path decision as they are not and should not be aware of the mega-constellation network topology. This decision-making is reserved for the network operator, similar to traditional telecom systems.

Between the UT and its anchor GS, data traffic is encapsulated in MPLS, with specific path labels to guide it. As the UT and the anchor GS do not change physical locations, however the path must be swapped through new satellite nodes at specific topology changes (e.g. to the next satellites on the same orbit), the UT receives a list of MPLS labels indicating which label to use based on the satellite node it is connected to. This setup allows for seamless handovers between satellite connections. In case the UT is a nomadic or a mobile node, and its physical location change results in a new satellite connectivity, then a new set of MPLS labels is needed to maintain the accurate routing. Please note that the UT and the GS handovers most probably will occur at different times, so there is a need for an intermediary data path when only one of the handovers has happened as illustrated in Figure 4-3. However, this can be realized transparently by the last space node as it is aware which is the path to use to reach a feeder link to the specific GS and thus it does not have to be signaled to the UT.

4.4 IMPLEMENTATION FEASIBILITY

In this section, we briefly overview notable advancements across various facets of the satellite routing practical implementation, assessing the feasibility of our proposed model, especially the impact on the space nodes.

Label matching and swapping is a straightforward operation, which can be performed on minimally complex hardware switches without extensive routing table lookups, similar to how MPLS is implemented in terrestrial networks. Additionally, due to the route pre-calculations done by the controller, the space nodes are highly simplified, not requiring having an IP protocol stack or LDP support. This reduction of functionality drastically reduces the need for compute in the space node for the operation of a routing protocol and thus the overall costs of the deployed system can be reduced. This solution allows the integration of even the smallest digital space forwarding nodes ([40]).

Furthermore, this type of switching can be done independently for each input interface, removing the need for a single centralized routing table that processes all data packets. This level of parallelization enables nodes to effectively manage multiple ground and ISLs with high efficiency.

However, the trade-off for reduced compute capacity on space nodes is the need for substantial data storage resources due to the requirement to store multiple routing tables. To make routing tables smaller, labels can be reused. This strategy takes advantage of regular network patterns and opportunities, such as simultaneous handovers of Inter-Satellite Links (ISLs) by satellites at the same latitude or the recurring visits to the same location by a satellite within a specific timeframe. Such patterns allow for the reuse of routing tables, even across different satellite topologies, reducing their size.

The SDN functionality that enables a satellite to receive routing information can be implemented over TM/TC (Telemetry/Telecommand) ([41]), but a more efficient approach is to create dedicated management communication paths for all nodes through the ISLs. These paths can be used to process locally addressed packets or relay them to the next node. Since the management- and data-path use the same physical, they are susceptible to similar link failures. To mitigate this, we proposed a mechanism like Fast Reroute (FRR) to handle potential node or link failures, ensuring resilience. Management information can be transmitted using established node management protocols like SNMP or Netconf ([42]), which are already widely used in Linux-based systems and have been thoroughly tested in terrestrial networks. It should be the same as the management protocol for the digital payload.

Implementing shortest path Dijkstra's algorithm and managing label distribution in the SDN controller is complex because it must simulate all the constellation links. Tools like Fraunhofer FOKUS OpenLanes ([43]) are developed for this task, including large-scale network emulation for both terrestrial and non-terrestrial networks. For example, OpenLanes can simulate up to 400 nodes, allowing for comprehensive system emulation. It uses cloud networking virtualization technologies to emulate connectivity between virtual payloads, adjusting link characteristics such as delay, capacity, and packet loss, main link parameters used for routing protocols. With its ability to daisy-chain virtual networks across multiple servers, OpenLanes can scale to emulate even larger constellations. The tool also facilitates automated tests across various network topologies, enabling thorough assessments of new systems in a range of scenarios, including rare or edge cases like single or multiple node failures.

The emulation capabilities of OpenLanes allow for out-of-the-box use of protocols like OSPF, MPLS, or LDP with Linux libraries. While IP addresses are required for OSPF, this can be handled statically within the emulator, reducing setup time. Using Linux-based protocol stacks is a reliable choice due to their stability and extensive long-term testing in real networks. For communication with user terminals (UT), 3GPP provides options like transmitting routing policies through the Policy and Charging Function (PCF), which is particularly useful during

handovers for nomadic and mobile terminals, enabling immediate adaptation to new network configurations ([44]).

4.5 EVALUATION

In this section, we empirically compare our proposed SDN-based routing solution with the widely used OSPF and MPLS solution, as discussed earlier, using a set of generic criteria usually used to quantify routing solutions. The comparison is graphically illustrated in Figure 4-5 and detailed in Table 1.

The primary criterion for evaluating a routing solution is the convergence duration, specifically, the time it takes for the system to stabilize after a topology change. In terrestrial networks, changes to OSPF topology occur infrequently, and result in long convergence times even when network is composed of optic fibres. In comparison, our SDN-based solution achieves rapid convergence because the necessary information is pre-computed and already available within the nodes.

Table 1: Comparison of OSPF/MPLS and SDN Routing

Criterion	**OSPF/MPLS**	**SDN Routing**
Convergence Speed	Very long duration is needed to reach convergence as the shortest path will have to be computed across long delay links and a large mesh network	Convergence is practically instant as the information is already available in the network nodes.
Complexity	Each node has to include the capability to compute the shortest path depending on the neighbours	The shortest path is computed centrally by an algorithm. It is the same algorithm as in the case of OSPF/MPLS solution
Overhead of communication	The nodes need to exchange messages to be able to compute the shortest path	The overhead is larger as the complete algorithm is centrally computed, resulting

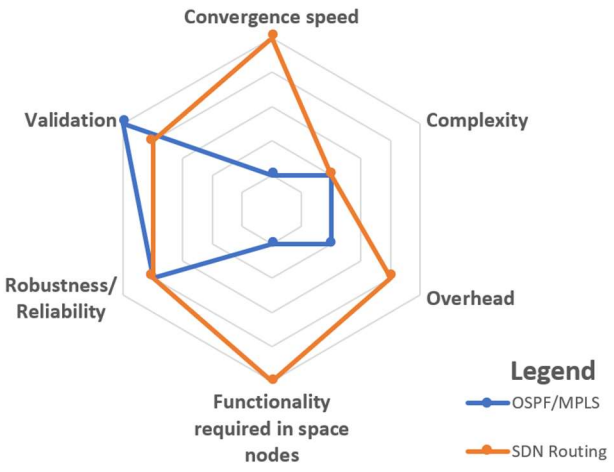

Figure 4-5 : Empirical Comparison with the classic solutions

	each time the topology changes. This includes a significantly large number of messages exchanged in space	into a large amount of information being communicated to the nodes i.e. routing tables for all the topology situations.
Functionality required in space nodes	All nodes have to maintain the OSPF and MPLS control plane	No functionality needed in the space nodes for routing control. The routing control decisions are received from the central command center.
Robustness/reliability	Mechanisms to protect paths are included in the OSPF/MPLS protocols	The same mechanisms are used for the pre-computed routing
Validation	OSPF/MPLS have a large history of being validated in real environments, however not in the dynamic, dense-link topologies of mega-constellations.	SDN approach for routing is new and no specific products are available, although also for terrestrial networks there are initial considerations.

Another factor is the overall protocol complexity. OSPF/MPLS implementations require a large number of nodes that must interact intensively during topology changes, up to where the network topology impacts the routing which is composed of all the nodes in case of a mega-constellation. Our SDN approach simplifies this by handling all operations within the emulation environment and transmitting results directly to the target nodes. Although the nodes still need to communicate within an isolated environment, the complexity of direct node-to-node interactions is significantly reduced. We assume a similar complexity as ultimately the nodes have to be emulated and communicate between them, although in an isolated environment and still information has to be communicated to the nodes.

However, the proposed SDN solution requires minimal functionality from space nodes. Most decision-making protocols reside in the controller, significantly simplifying the space nodes' requirements—they do not even need an operating system, unlike their OSPF and MPLS counterparts in terrestrial routers.

Regarding the additional functionality added by the routing solution, the SDN-based routing introduces more overhead due to the need to manage communication with the nodes, including establishing an additional SDN management communication path.

However, the proposed SDN solution requires a minimal functionality from the perspective of the space nodes. Most of the protocol stacks used to make the decisions are located in the controller. Most of the functionality does not have to be supported. As such the space nodes remain relatively simple from the perspective of the routing, not even requiring an operating system to function as in the case of OSPF and MPLS in terrestrial routers.

Both solutions implement a fast-re-routing mechanism that ensures robustness, providing the same level of node and link protection.

While the OSPF/MPLS solution has been validated in numerous commercial setups, confirming its reliability and stability, our new SDN solution has not yet been validated in a similar environment. However, since it utilizes the same protocol stacks and management protocols as those employed in terrestrial networks within the emulation environment, we anticipate that substantial validation steps have already been completed.

4.6 CONCLUSIONS AND FURTHER WORK

In this chapter, we have introduced an SDN-based approach to space-based routing that can accommodate various decision-making mechanisms from the literature. Our proposed solution employs a central SDN controller, which utilizes a large-scale network emulator and pre-existing protocol stacks to make routing decisions. This setup enables the initial validation of space designed routing protocols, such as ([45], [46], [47]), using the space network emulator before deploying them in actual satellite network environments through the designated network management protocol.

Consequently, our approach not only simplifies the deployment of space nodes but also segregates the decision-making and validation of routing protocols from their actual implementation. This separation allows for extensive testing and validation prior to deployment, enhancing system reliability.

Moving forward, we plan to address for the proposed concept a distributed controller model where the different controllers are introducing their own specific forwarding rules as well as further considerations for addressing and routing using the proposed SDN model. For this, we will prioritize minimizing memory usage by considering reoccurring patterns that take effect within constellations, that go beyond predictable topologies. We will also incorporate a minimal semantic layer to adjust routing strategies in response to significant challenges such as feeder link capacity reductions or disruptions caused by weather, as well as congestion in space links and nodes.

5 CONCLUSION AND OUTLOOK

Summarizing the main finding of the architectural considerations based on ATSSS. ATSSS is only specified for 3GPP and another single non-3GPP connection, or one 5G connection and one 4G. The standard would need to be extended to allow parallel 5G connections across independent 5G base stations beyond CoMP and Dual Connectivity. The control plane procedures specified fulfilling already all use cases identified considering several options of network operators of parallel flows, however requiring extensions on the identities and identifiers of the connections as presented in Section 2.2. Next to this, for the data path also the PMF timers have to be revised to cover the specific satellite delay.

At the moment the MP anchor is specified to be at PSA UPF, in order to make efficient use of MP also in NTN-NTN setups it could be beneficial to terminate MP-connections also at UPFs hosted as payload of satellites and forward data from there toward the PSA UPF in single connections mode. However, this represents a secondary use case to the TN-NTN interoperability which represent the main current requirement for the adoption of 5G NTN.

For indirect communication one new option introduced is to enable a MC-layer at IAB level so that the relaying node is backhauled by two or more connections that allow for steering, splitting and switching. The inclusion of MC in this implementation of indirect connection requires an update of standard. If it is introduced at IAB level, it needs to be terminated at the serving CU of the IAB-Donor, a termination at UPF is not possible to not break the stack. For indirect communication mechanism for MC handling must be developed to make UEs aware about the multiple paths available and for exchanging QoS information.

Since MPQUIC is encrypting the traffic the order placement within the stack needs to be carefully selected to not disable communication of block encapsulated within.

A comprehensive trade-off will be executed to assess the opportunity of the deployment of the additional functionality developed for the multi-connectivity architecture for load balancing, more performance end-to-end service, robustness and security against the extensive resource consumption.

Furthermore, this deliverable presented a first SDN-based mega-constellation routing approach. This provides a mechanism to transfer routing related information prior to its usage, through this drastically reducing the routing convergence time based on already available information. Going forward, to be able to prove the advantages of the SDN mechanism, at least one of the many routing algorithms available in the literature will be implemented and adapted for a mega-constellation model to prove the advantages and the limitations. Additionally, a multi-controller approach will be added in the next deliverable to enable an organizational approach towards the system routing where multiple use cases could add their own routing information without requiring the knowledge centralization.

This deliverable is a preliminary report of the ongoing activities for the tasks T5.1 (Multi-Connectivity solutions) and T5.3 (Software-defined Network Control), the results of the completed activity will be reported in the final deliverable D5.4.

REFERENCES

[1] A. Guidotti et al., 5G-STARDUST Deliverable 3.1, “System Requirements Analysis and Specifications”, v.1.3, 07.2023

[2] A. Guidotti et al.,5G-STARDUST Deliverable 3.2, “End-to-end ground and space integrated architecture”, v1.0, 02.2024

[3] N. Trujillo et al., 5G-STARDUST Deliverable 2.1, “Scenario, use cases and services”, Version: 3.0.1, 08.2023

[4] 3GPP TS 38.101-1 “User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone” V18.5.0, 03.2024

[5] 3GPP TS 38.101-2 "User Equipment (UE) radio transmission and reception; Part 2: Range 2 Standalone", V18.5.0, 03.2024

[6] 3GPP TS 38.101-5 "User Equipment (UE) radio transmission and reception; Part 5: Satellite access Radio Frequency (RF) and performance requirements", V18.5.0, 03.2024

[7] 3GPP TS 24.193, “Access Traffic Steering, Switching and Splitting (ATSSS) Stage 3”, v18.5.0, 03.2024

[8] 3GPP TS 23.501, “System architecture for the 5G System (5GS), Stage 2”, v18.5.0, 03.2024

[9] 3GPP TS 24.526, “User Equipment (UE) policies for 5G System (5GS); Stage 3”, v18.6.0, 04.2024

[10] IETF RFC 8684, “TCP Extensions for Multipath Operation with Multiple Addresses”, March 2020

[11] IETF draft-ietf-quic-multipath-07, “Multipath Extension for QUIC”, 04.2024, expiring 10.2024

[12] M. Luglio, M. Quadrini, S. P. Romano, C. Roseti and F. Zampognaro, "Enabling an efficient satellite-terrestrial hybrid transport service through a QUIC-based proxy function," 2020 International Symposium on Networks, Computers and Communications (ISNCC), Montreal, QC, Canada, 2020, pp. 1-6, doi: 10.1109/ISNCC49221.2020.9297334.

[13] 5G-STARDUST Deliverable 5.3, “Preliminary report on AI-data driven networking and QoS management”, to be submitted in June 2024

[14] T. de Cola et al., 5G-STARDUST Deliverable 6.1: “PoC functional architecture document”, v. 1.0, 04.2024

[15] IETF RFC 9221, “An Unreliable Datagram Extension to QUIC”, 03.2022

[16] C. Pupiales, D. Laselva, Q. De Coninck, A. Jain and I. Demirkol, "Multi-Connectivity in Mobile Networks: Challenges and Benefits," in IEEE Communications Magazine, vol. 59, no. 11, pp. 116-122, November 2021, doi: 10.1109/MCOM.111.2100049;

[17] 3GPP TS 37.340. “Multi-connectivity; Overall description; Stage-2”, v18.1.0., March 2024, https://www.3gpp.org/;

[18] Pan, Cunhua, et al. "User-centric C-RAN architecture for ultra-dense 5G networks: Challenges and methodologies." IEEE Communications Magazine 56.6 (2018): 14-20;

[19] 3GPP TS 24.193 Technical Specification Group Core Network and Terminals. 5G System Access Traffic Steering, Switching and Splitting (ATSSS); v.18.5.0, April 2024, https://www.3gpp.org/;

[20] Fraunhofer FOKUS Open5GCore toolkit, www.open5Gcore.org;

[21] T. Sylla, L. Mendiboure, S. Maaloul, H. Aniss, M. A. Chalouf, and S. Delbruel, “Multi-Connectivity for 5G Networks and Beyond: A Survey,” Sensors (Basel), vol. 22, no. 19, p. 7591, Oct. 2022, doi: 10.3390/s22197591;

[22] IETF RFC 8684, “TCP Extensions for Multipath Operation with Multiple Addresses”, March 2020, https://datatracker.ietf.org/doc/html/rfc8684;

[23] IETF, “Multipath Extension for QUIC”, draft-ietf-quic-multipath-08, https://datatracker.ietf.org/doc/draft-ietf-quic-multipath/, work in progress;

[24] 3GPP TS 33.501, “Security architecture and procedures for 5G System”, v18.5.0, April 2024, https://www.3gpp.org/;

[25] S. Barré, C. Paasch, and O. Bonaventure, “MultiPath TCP: From Theory to Practice,” in NETWORKING 2011, J. Domingo-Pascual, P. Manzoni, S. Palazzo, A. Pont, and C. Scoglio, Eds., Berlin, Heidelberg: Springer, 2011, pp. 444–457. doi: 10.1007/978-3-642-20757-0_35.

[26] 3GPP TS 23.288: "Architecture enhancements for 5G System (5GS) to support network data analytics services", V18.5.0, 03.2024.

[27] M. Hoyhtya, M. Corici, S. Covaci, and M. Guta, “5G and beyond for new space: Vision and research challenges,” in Advances in Communications Satellite Systems. Proceedings of the 37th International Communications Satellite Systems Conference (ICSSC–2019). IET, pp. 1-16.

[28] A. Vanelli-Coralli, A. Guidotti, T. Foggi, G. Colavolpe, and G. Montorsi, “5g and beyond 5g non-terrestrial networks: trends and research challenges,” 2020 IEEE 3rd 5G World Forum (5GWF), pp. 163—169

[29] Louis J. Ippolito, "Transmission Impairments," in Satellite Communications Systems Engineering: Atmospheric Effects, Satellite Link Design and System Performance , Wiley, 2017, pp.87-137

[30] U. D. Black and U. N, Black. IP routing protocols: RIP, OSPF, BGP, PNNI, and Cisco routing protocols. Prentice Hall Professional 2000.

[31] M. N. Alslaim, et al, “A comparative study of manet routing protocols,” in The Third International Conference on e-Technologies and Networks for Development (ICeND2014). IEEE, pp. 178-182.

[32] A. Agarwal, C. Iskander, and R. Shankar, “Survey of network on chip (noc) architectures & contributions.” Journal of Engineering, Computing and Architecture 3, no. 1, pp. 21—27.

[33] Cao, Xiaoli, et al. "Dynamic routings in satellite networks: An overview." Sensors 22, no. 12 (2022): 4552.

[34] Previdi, S., E. Aries, and D. Afanasiev, IETF "RFC 9087: Segment Routing Centralized BGP Egress Peer Engineering." (2021).

[35] Jiang, Weiwei. "Software defined satellite networks: A survey." Digital Communications and Networks 9.6 (2023): 1243-1264.

[36] Köhler, Stefan, and Andreas Binzenhöfer. "MPLS traffic engineering in OSPF networks - a combined approach." Teletraffic Science and Engineering. Vol. 5. Elsevier, 2003. 21-30.

[37] Li, Tong, Jinqiang Chen, and Hongyong Fu. "Application scenarios based on SDN: an overview." Journal of Physics: Conference Series. Vol. 1187. No. 5. IOP Publishing, 2019.

[38] Raj, Alex, et. al, "A survey of IP and multiprotocol label switching fast reroute schemes." Computer Networks 51.8 (2007): 1882-1907.

[39] Dao, N.N. et al., 2024. A review on new technologies in 3GPP standards for 5G access and beyond. Computer Networks, p.110370.

[40] Maravić, Igor, and Aleksandra Smiljanić. "MPLS implementation for the Linux kernel." 2012 IEEE 13th International Conference on High Performance Switching and Routing. IEEE, 2012.

[41] Kriezis, Anargyros, and Whitney Lohmeyer. "An Overview and Assessment of Today's US Satellite Telemetry, Tracking and Telecommand (TT&C) Spectrum Allocation Needs." Tracking and Telecommand (TT&C) Spectrum Allocation Needs (April 10, 2023)

[42] Mellia, Marco, Nur Zincir-Heywood, and Yixin Diao. "Overview of Network and Service Management." Communication Networks and Service Management in the Era of Artificial Intelligence and Machine Learning (2021): 1-18.

[43] Fraunhofer FOKUS OpenLanes: Large Nework Emulator, high level toolkit description, https://connectivity.esa.int/sites/default/files/ADS%20public%20ScyLight%202, last visited on 28.04.2023;

[44] 3GPP TS 24.526, https://www.3gpp.org/dynareport/24526.htm, last visited on 28.04.2024

[45] Roth, Manuel, Hartmut Brandt, and Hermann Bischl. "Implementation of a geographical routing scheme for low Earth orbiting satellite constellations using intersatellite links." International Journal of Satellite Communications and Networking 39.1 (2021): 92-107.

[46] Korçak, Ömer, and Fatih Alagöz. "Priority-based adaptive shortest path routing in IP over LEO satellite networks." 23rd AIAA International communications satelllite systems Conference. 2005.

[47] Z. Han, et. al, "Dynamic Routing for Software-Defined LEO Satellite Networks based on ISL Attributes," 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 2021, pp. 1-6

APPENDIX A – MULTI-CONNECTIVITY PROTOCOL STACKS

In the following the protocol stacks for all considered Multi-Connectivity (MC) topologies are presented that have been derived from the 5G-STARDUST baseline use cases and architecture, D3.1 and D3.2. The illustration is following the color-code:

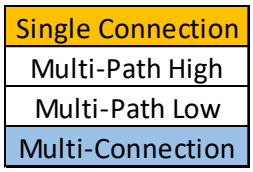

Figure 5-1: Protocol stacks, color-coding

Which means that orange is single connections which is split, switched or steered by the MP-High and MP-Low layer colored in white. The blue connection is the MC which means that these parts of the protocol stacks must be available at least twice, e.g., on TN and NTN path. The topologies show always an NTN and an TN path, but the stacks generally apply to other combination of paths (e.g., NTN-NTN).

Baseline Single Link Connection

For a better overview, we present in the following again the reference single connection stacks as presented in D3.1.

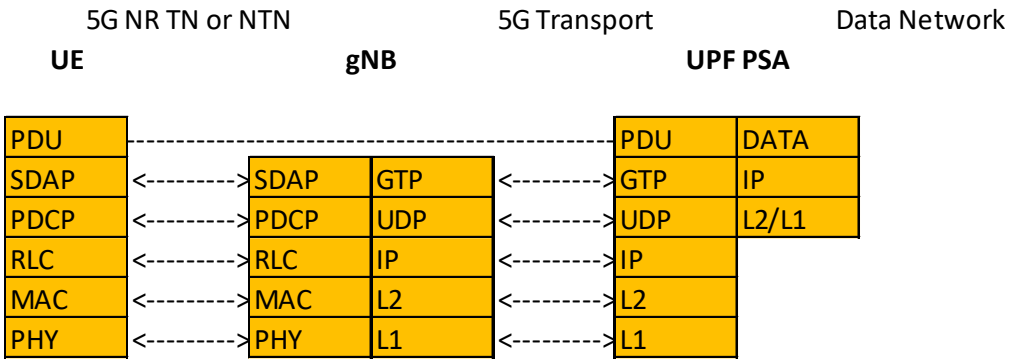

Figure 5-2: Protocol stack, baseline 5G-UP Stack single connectivity [1]

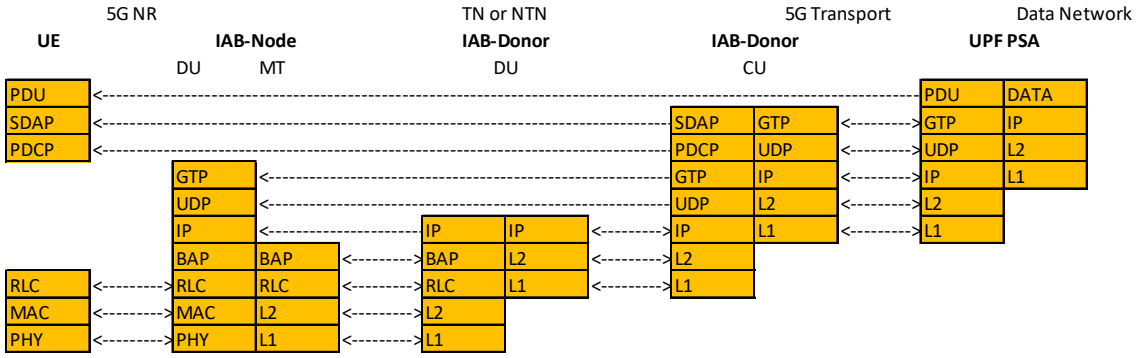

Figure 5-3: Protocol stack, baseline 5G-IAB Stack single connectivity [1]

Direct Connection with Multi-Connectivity Layer at UE

This is the baseline ATSSS case enhanced to allow parallel 5G-connections.

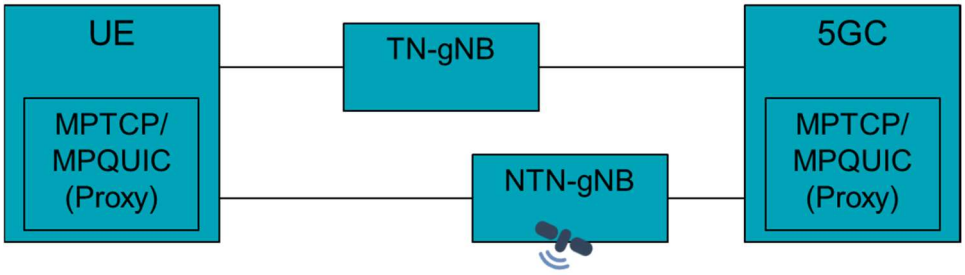

Figure 5-4: Direct connection with MC-layer at UE

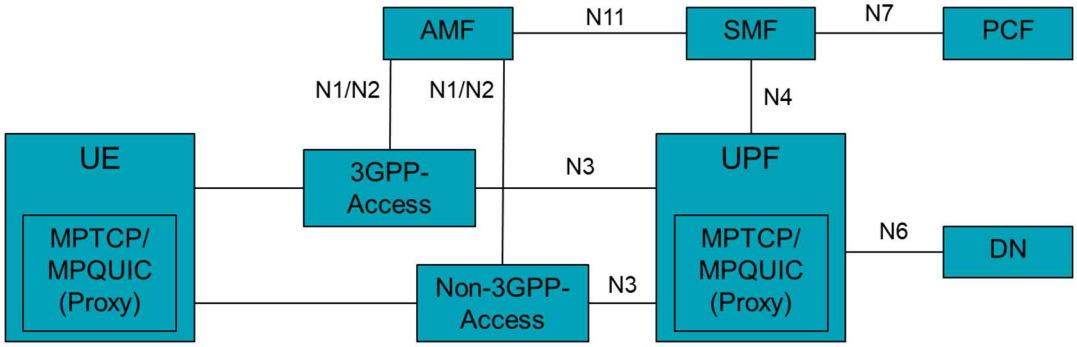

Figure 5-5: ATSSS architecture [8]

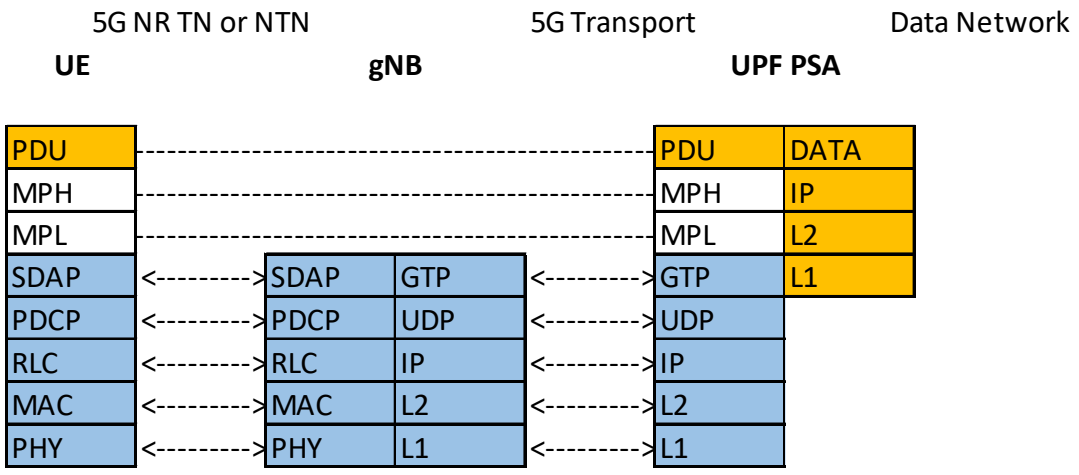

Figure 5-6: Protocol stack, direct connection with MC-layer at UE

Direct Connection with Multi-Connectivity Layer at UE, UPF in Space

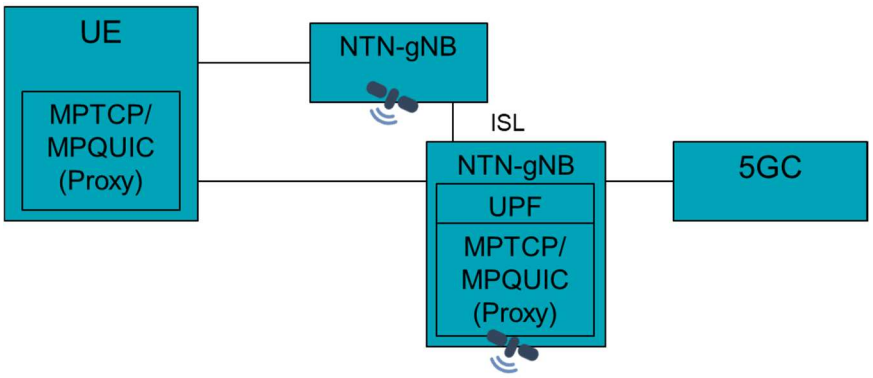

Figure 5-7: Direct Connection with MC-layer at UE, UPF in Space

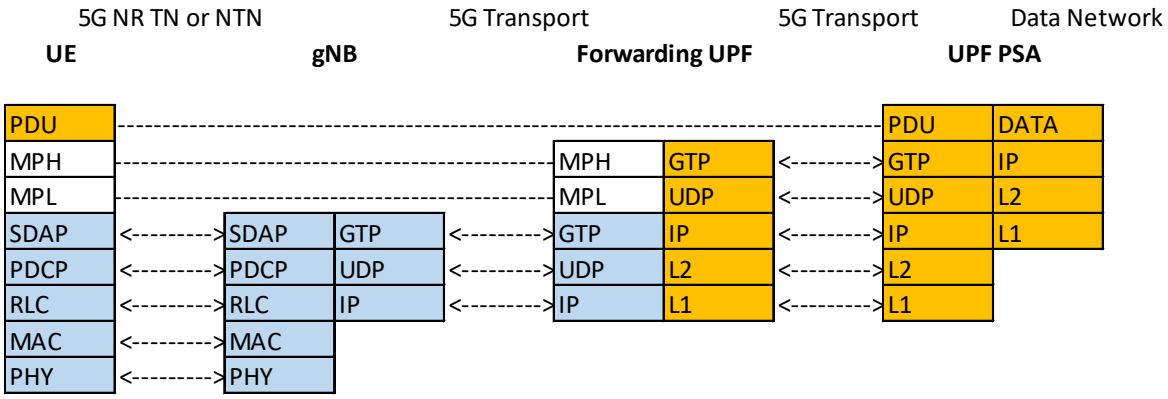

Figure 5-8: Protocol stack, direct Connection with MC-layer at UE, UPF in Space

Indirect Connection with Multi-Connectivity Layer at UE

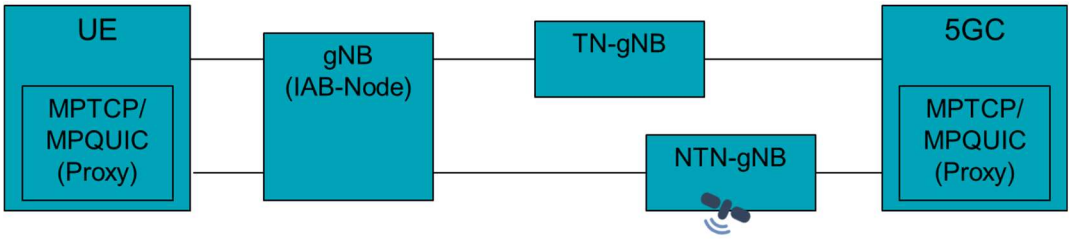

Figure 5-9: Indirect connection with MC-layer at UE

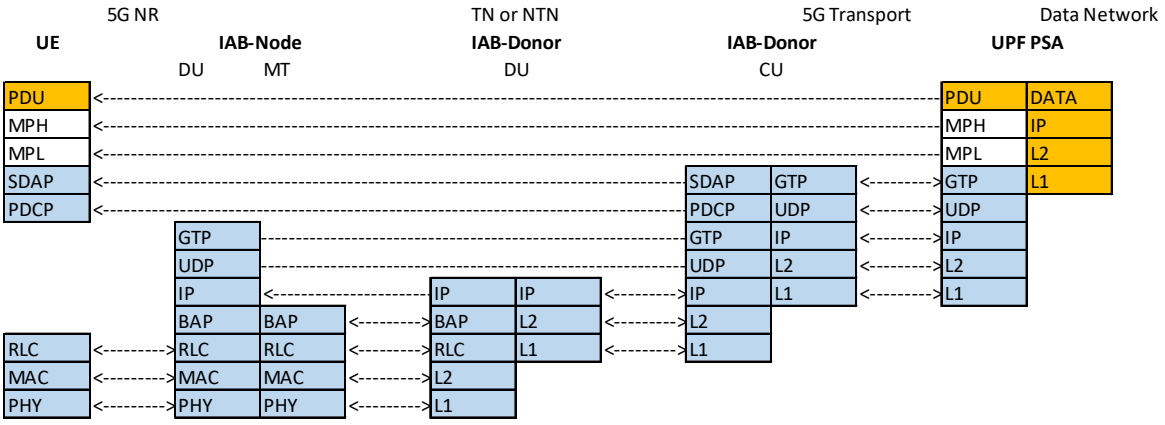

Figure 5-10: Protocol stack, indirect connection with MC Layer at UE

Indirect Connection with Multi-Connectivity Layer at UE and UPF in Space

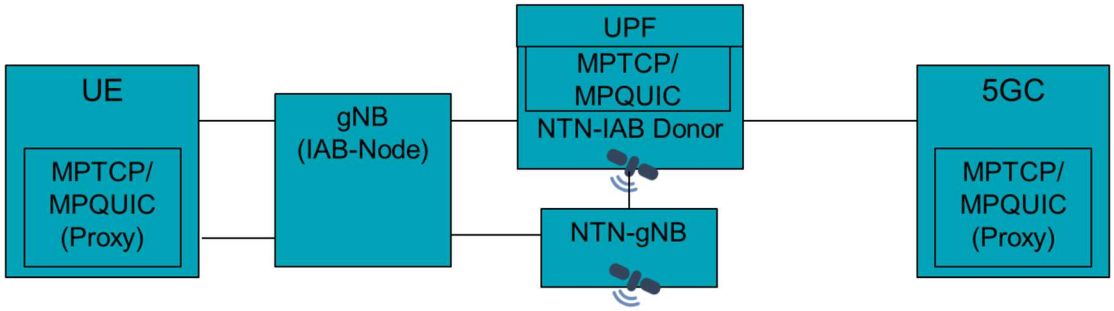

Figure 5-11: Indirect connection with MC-layer at UE, UPF in Space

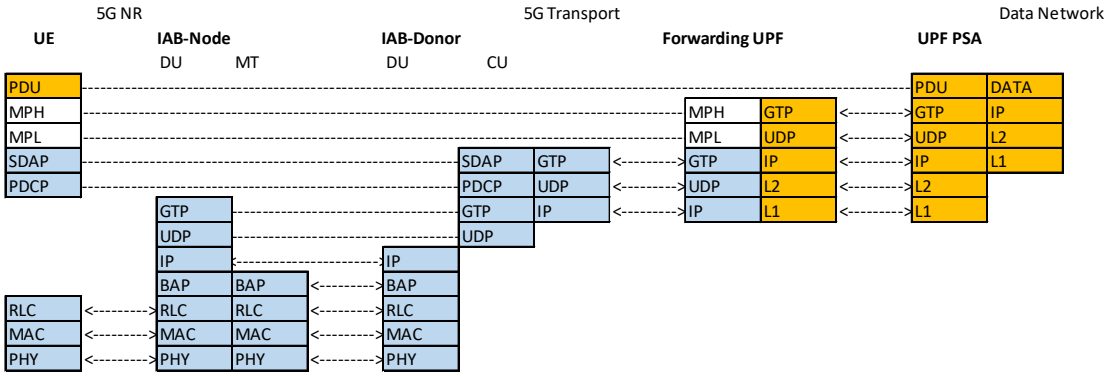

Figure 5-12: Protocol stack, indirect connection with MC-layer at UE, UPF in Space

Indirect Connection with Multi-Connectivity Layer at IAB-Node

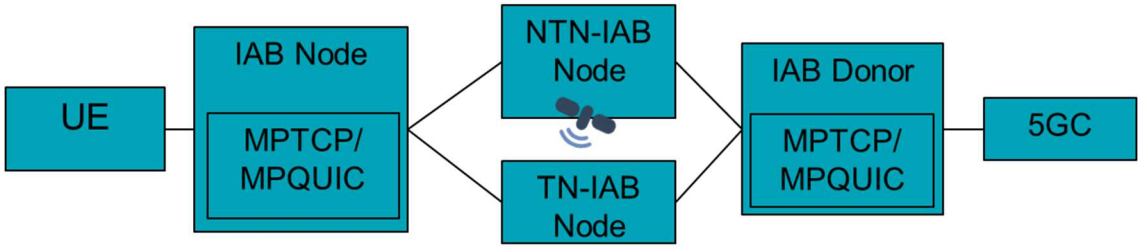

Figure 5-13: Indirect connection with MC-layer at IAB-node, donor on ground

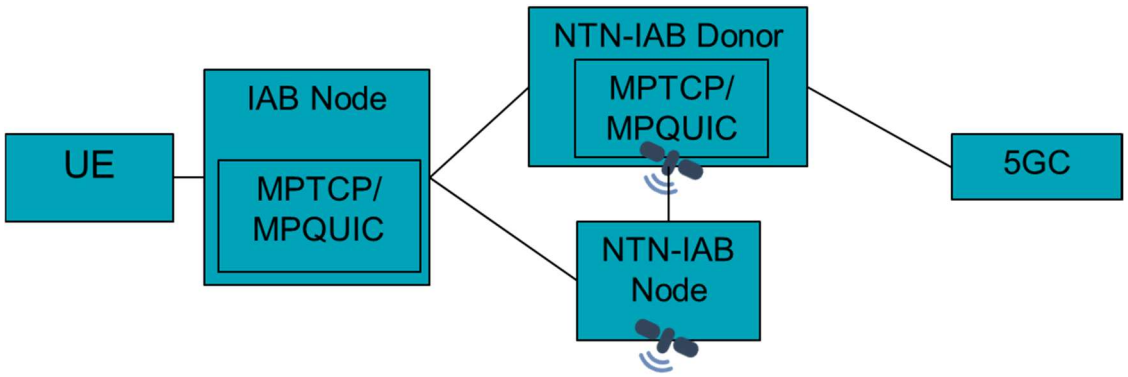

Figure 5-14: Indirect connection with MC-layer at IAB-node, donor in space

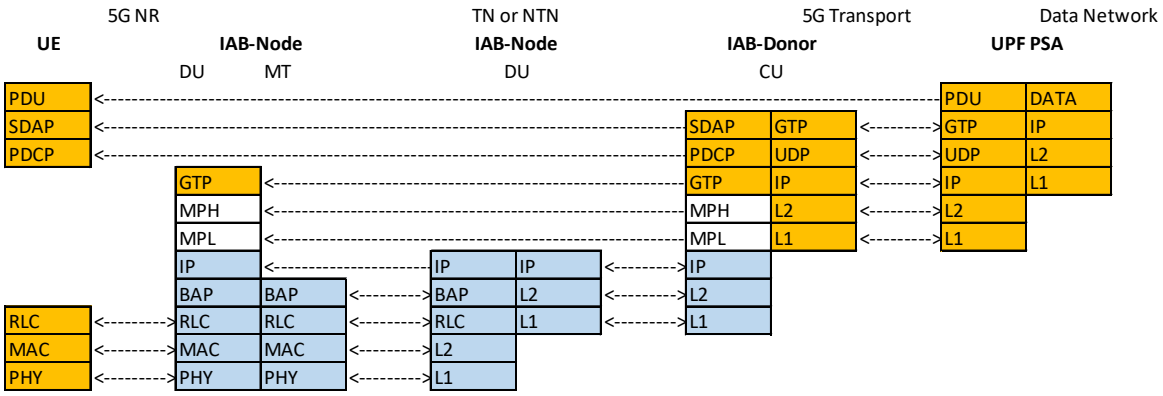

Figure 5-15: Protocol stack, indirect connection with MC-layer at IAB-node

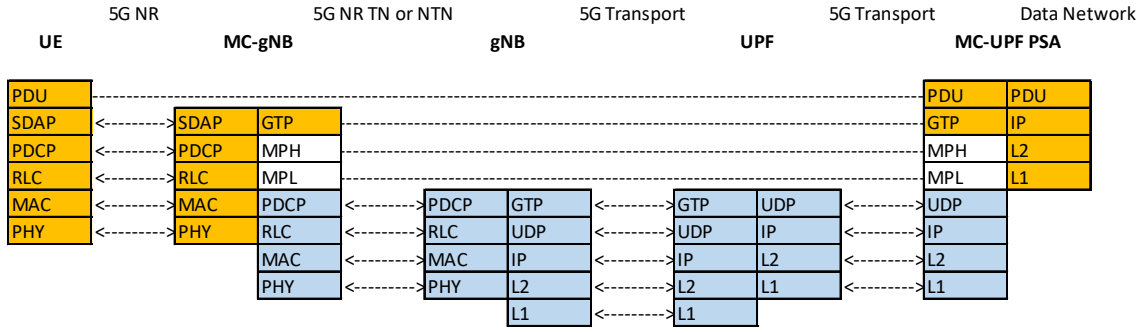

Figure 5-16: Protocol stack, indirect connection with MC-layer at IAB-node, PoC

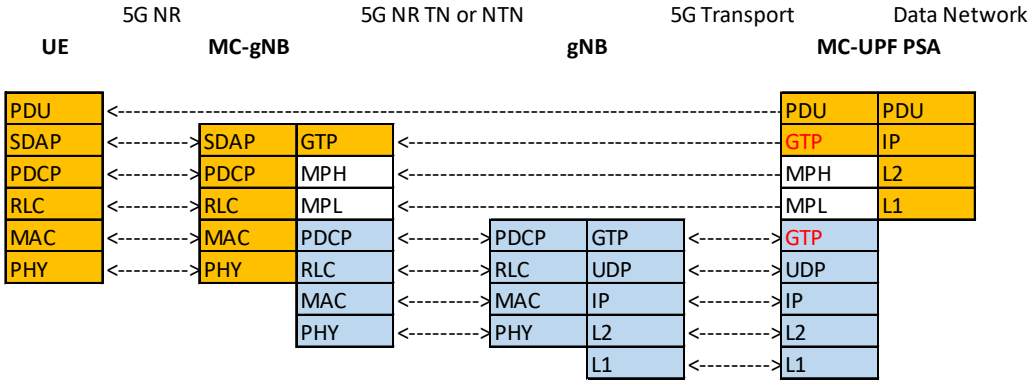

Figure 5-17: Protocol stack, indirect connection with MC-layer at IAB-node, GTP problem

Indirect Connection with Multi-Connectivity Layer at forwarding UPF

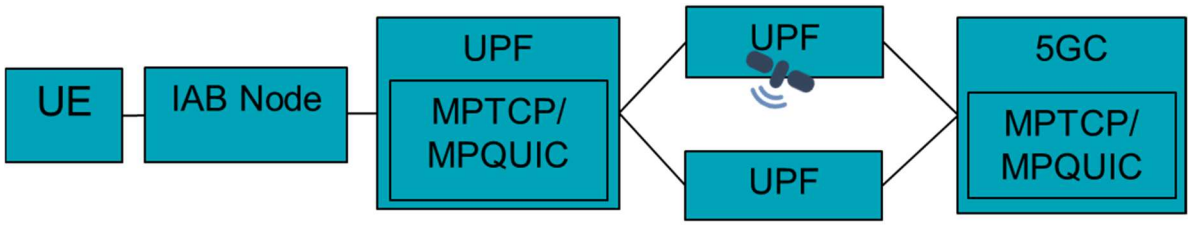

Figure 5-18 : Indirect connection with MC-layer at forwarding UPF, UPF PSA on-ground

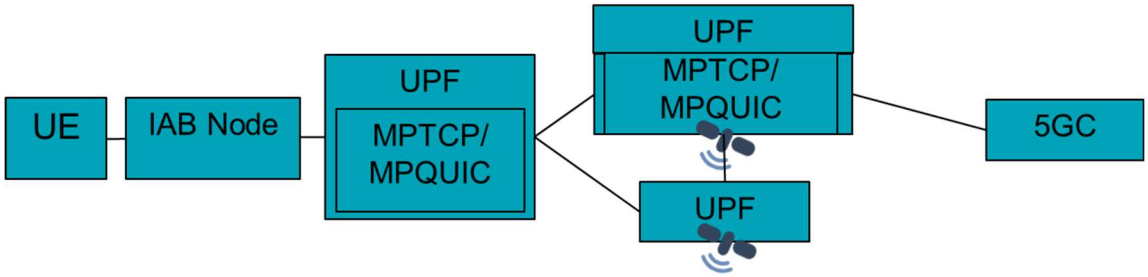

Figure 5-19 : Indirect connection with MC-layer at forwarding UPF, UPF PSA in space

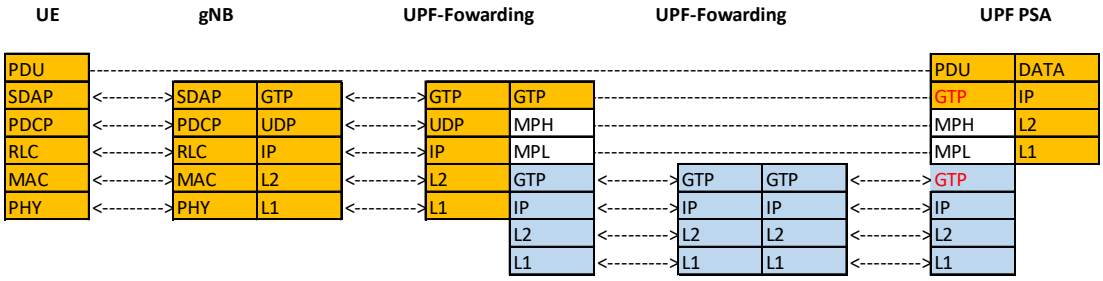

Figure 5-20 : Protocol stack, Indirect connection with MC-layer at forwarding UPF