Document Number: H2020-EUK-815323/5G-ALLSTAR/D4.1

Project Name:
5G AgiLe and fLexible integration of SaTellite And cellulaR (5G-ALLSTAR)

Deliverable D4.1

Mapping of the multi-connectivity functions onto the 5G network architecture

Date of delivery:	30/04/2019	Version:	1.0
Start date of Project:	01/07/2018	Duration:	36 months

Deliverable D4.1

Mapping of the multi-connectivity functions onto the 5G network architecture

Project Number:	H2020-EUK-815323
Project Name:	5G AgiLe and fLexible integration of SaTellite And cellulaR

Document Number:	H2020-EUK-815323/5G-ALLSTAR/D4.1
Document Title:	Mapping of the multi-connectivity functions onto the 5G network architecture
Editor(s):	**Federico Lisi (CRAT)**
Authors:	Federico Lisi, Antonio Pietrabissa, Vito Andrea Racanelli, Antonio Ornatelli, Martina Panfili, Alessandro Giuseppi, and CRAT Team (CRAT), Seok Ho Won, Sungoh Kwon, Syed Maaz Shahid. Md Mehedi Hasan (ETRI and University of Ulsan)
Dissemination Level:	PU
Contractual Date of Delivery:	30/04/2019
Security:	Public
Status:	Final
Version:	1.0
File Name:	5G-ALLSTAR Deliverable 4.1

Abstract

The document reports the preliminary design of 5G-ALLSTAR multi connectivity architecture and interfaces. It also includes a preliminary investigation of methodologies and techniques to QoE Estimation, QoE Control, and traffic flow control for enabling multi-connectivity in 5G networks.

Keywords

Multi-Connectivity, Cloud RAN, Distributed RAN, QoE, QoE Control, traffic flow control

Acknowledgements

We would like to acknowledge the following people for the valuable reviews to the deliverable: Emilio Calvanese Strinati, Nicolas Cassiau, Marjorie Thary, and Francesco Delli Priscoli

Executive Summary

This deliverable describes the goals achieved within Task 4.1 "*Multi-connectivity mapping onto the 5G network architecture*" and Task 4.2 "*Design and simulation of the multi-RAT load balancing algorithms*" in the period of July 2018 – April 2019 (M1-M10 for Task 4.2) and October 2018 - April 2019 (M4-M10 for Task 4.1).

The aim of Work Package 4 (WP4) "*Multi-Connectivity*" is to develop an architectural and functional framework to enable Multi-Connectivity in 5G networks.

The purpose of this deliverable is to introduce the architectures developed in 5G-ALLSTAR, built starting from the current standards, on which the load-balancing algorithms of T4.2 and T4.3 will be designed, and eventually, demonstrated.

A preliminary analysis of the current solutions for load balancing, and traffic control in general, is provided in Section 3.1, both from an architectural and algorithmic point of view, while Sections 3.2 and 3.3 focus on the envisaged innovations of 5G-ALLSTAR, furtherly detailed in chapters 4, 5 and 6.

Contents

List of Figures

List of Tables

List of Abbreviations

2D	2 Dimensions / 2-Dimensional
3D	3 Dimensions / 3-Dimensional
3GPP	3rd Generation Partnership Project
5GPPP	5G Public-Private Partnership
5QI	5G QoS Identifier
AHP	Analytic hierarchy process
AI	Artificial Intelligence
AIV	Air Interface Variants
AMF	Access and Mobility Management Function
BS	Base Station
CN	Core Network
CP	Control Plane
C-RAN	Central Radio Access Network
cRRM	Central Radio Resource Management
CSI	Channel State Information
CSI-RSRP	CSI Reference Signal Received Power
CSI-RSRQ	CSI Reference Signal Received Quality
CSI-SINR	CSI Signal-to-Noise and Interference Ratio
CU	Centralized Unit
DL	Downlink
D-RAN	Distributed Radio Access Network
DRB	Data Radio Bearer
dRRM	Distributed Radio Resource Management
DU	Distributed Unit
FLC	Fuzzy Logic Controller
FS	Fast Switch
GEO	Geostationary Earth Orbit
gNB-CU	next Generation Node B Central Unit
gNB-DU	next Generation Node B Distributed Unit
GRA	Grey Relational Analysis
HH	Hard Handover
KPI	Key Performance Indicator
LEO	Low Earth Orbit
LTE	Long Term Evolution
MAC	Medium Access Control
MADM	Multiple attribute decision making
MCG	Master Cell Group
MDP	Markov Decision Process
MN	Master Node
NE	Nash Equilibrium
NG-RAN	New Generation Radio Access Network
NR	New Radio
NTN	Non-Terrestrial Network
PDCP	Packet Data Convergence Protocol
PDR	Packet Detection Rule
PDU	Protocol Data Unit
PHY	Physical
QFI	QoS Flow Identifier
QoE	Quality of Experience
QoS	Quality of Service
RAT	Radio Access Technology
RLC	Radio Link Control
RRC	Radio Resource Control
RRM	Radio Resource Management
RTT	Round Trip Time
SCG	Secondary Cell Group
SDAP	Service Data Adaptation Protocol
SMF	Session Management Function
SN	Secondary Node
SNR	Signal to Noise Ratio
UE	User Equipment

UL	Uplink	**UPF**	User Plane Function
UP	User Plane	**WLAN**	Wireless Local Area Network

1 Introduction

This deliverable provides an overview of the preliminary mapping of 5G-ALLSTAR approaches in terms of Multi-Connectivity and Quality of Experience issue in the 5G networks.

1.1 Work Package and Deliverable Rationale

This document is organized to report the initial activities performed in WP4, structured as depicted in Figure 1. In particular, WP4 is divided into three main tasks, whose results are collected in three deliverables.

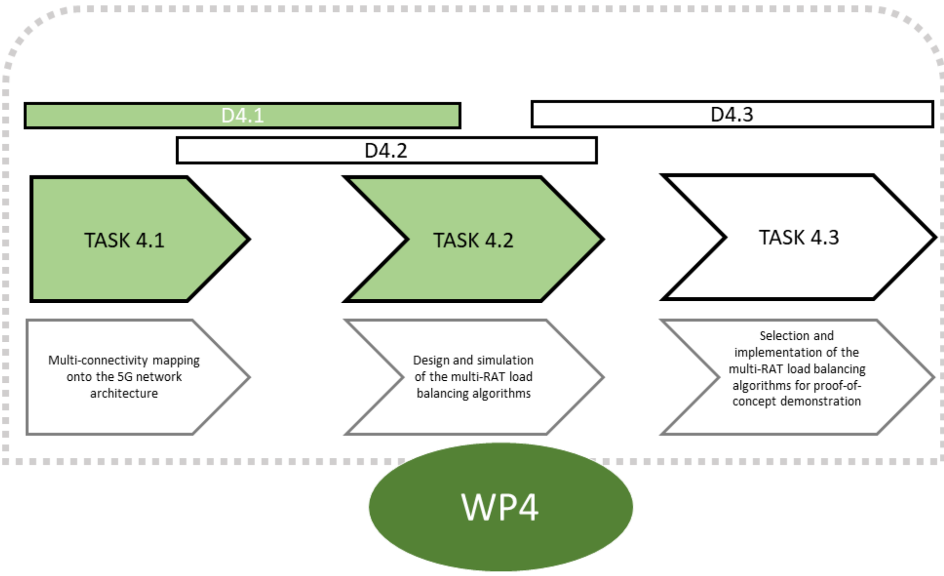

Figure 1 - Work Package 4 plan

This document undertakes the output of Task 4.1 and Task 4.2 and is organized as follows:

- Chapter 1 introduces the deliverable and reports its organization.
- Chapter 2 reports the connection of this document with the other 5G-ALLSTAR documents.
- Chapter 3 contains the investigated state-of-the-art in terms of 5G architectures and algorithms for enabling Multi-Connectivity in 5G networks. It also contains the beyond-state-of-the art brought by 5G-ALLSTAR, which will be detailed in 5G-ALLSTAR deliverable 4.2 (D4.2)
- Chapter 4 describes the Quality of Experience (QoE) methodologies introduced by 5G-ALLSTAR able to acquire User Equipment information for enabling the personalized QoE Control.
- Chapter 5 preliminarily reports the 5G-ALLSTAR ambition to cope with optimal and efficient traffic control in 5G networks. The traffic control investigated in 5G-ALLSTAR concerns the suitable solution strategies with the aim of addressing the traffic steering, switching and splitting issues.

- Chapter 6 presents the design of the inputs, outputs and interfaces among 5G-ALLSTAR components and how they are able to communicate with the standard 5G functionalities.

1.2 Summary

This deliverable is meant to report the first results attained in the scope of tasks T4.1 and T4.2.

Due to their interconnected nature, the tasks run in parallel, as T4.1 focuses on the architectural development of 5G-ALLSTAR for the problem of multi-connectivity, assuring the compliance of the tools developed in the project with the current standards, whereas T4.2 is in charge of developing the network control algorithms to enable the management of multiple RATs.

Being this deliverable the first feedback gathered by the two tasks, it reports a detailed state-of-the-art analysis for both architectural solutions and traffic control algorithms, from which the research directions currently investigated in WP4 are highlighted.

This deliverable also reports the approach that will be followed by the 5G-ALLSTAR project regarding the inclusion of the users' Quality of Experience in the multi-connectivity framework, discussing its compliance with the 5G architecture, potential benefits, and functional solutions.

To conclude the deliverable, it is reported a first version of the description of the interfaces of the various modules that WP4 will develop.

2 Relation with other work packages

WP4 has an important role in 5G-ALLSTAR since it will design and develop advanced functionality for as concern multi-connectivity and Quality of Experience control. In fact, with respect to the architecture designed in 5G-ALLSTAR deliverable 2.2, the WP4 components will impact the implementation of innovative Radio Access Network (RAN) and Core Network (CN) functionalities. Besides, the main WP4 components have a strong connection with the other results or main outputs achieved in WP2, WP3 and WP5, WP6. In detail:

- WP2 provided important inputs for WP4 during the design of requirements, Key Performance Indicator (KPI) and the high-level architecture. Such inputs have been considered in WP4 as references for the mapping between WP4 functionalities and the 5G-ALLSTAR architecture.
- With WP3, as introduced in Section 6 “Multi Connectivity interfaces” of this document (D4.1), we have preliminarily identified the interfaces to foster data exchange between WP4 and WP3 components that will be developed during the project.
- WP5 is strongly connected to WP4, in the sense that selected algorithms among the ones developed in WP4 will be also implemented in the testbed. In this regard, WP4 will enrich the 5G-ALLSTAR testbed with the main AI-based algorithms and components to foster the integration of the multi-connectivity and QoE Control in the 5G-ALLSTAR demonstration.
- WP6 is in charge of promoting the main results and achievement of the 5G-ALLSTAR project. The partners involved in WP4 disseminate such results in international scientific publication, classroom seminars.

3 5G-ALLSTAR Multi Connectivity

3.1 State of the art

This section provides the state of the art concerning 5G Networks from a twofold perspective. From one side it is dedicated to present the state of the art related to the 5G Multi-Connectivity architecture, from the other side it is dedicated to present the state of the art related to control algorithms for addressing multi-connectivity issues.

Multi-RAT access network, or heterogeneous access network, is considered to be the key enabling technology to satisfy the 5G requirements, such as high data rate, ultra-low latency and reliability. To make efficient use of all the available network resources, multi-connectivity has been proposed to simultaneously connect, and orchestrate, multiple different radio access technologies. The main advantage of the multi-connectivity approach fosters the possibility to send the user traffic in different Radio Access Technologies (RATs) that better satisfy the service requirements and the user needs. In general, multi-connectivity can be defined as the capability to configure a User Equipment (UE) to utilise resources provided by different nodes, which are characterized by different access technologies.

In [1] three multi-RAT integration methods are presented:

- *Application Layer Integration*: it consists of a higher-layer interface, providing information exchange between UEs and content provider, over multiple RATs. This solution can be easily implemented, but it is application-dependent and may not fully take into account the network state, which leads to suboptimal exploitation of resources, especially if the network state is observed to vary dynamically.
- *Core-Network-Based Integration*: this solution is proposed by 3GPP for cellular/WLAN integration based on interworking between core networks. In this case, the RAT selection is made considering operators' policy for network selection, but the overall network selection decision remains in control of the UE. The UE is then able to take its decisions considering operator policies, radio links performances and user preferences.
  It is worth remarking that typically the UE only has local knowledge about the network conditions, resulting in the suboptimal selection of decisions, degrading the overall network performances and the QoE of its user.
- *RAN-Based Integration*: this solution is proposed by 3GPP in NR/LTE dual-connectivity and allows coordination between the RATs using dedicated interfaces. The cooperation level between the different RATs is constrained by the backhaul links. Having high backhaul link capacities allows full cooperation between RATs, enabling more dynamic Radio Resource Management (RRM) mechanism and improving overall system and user performances.
  In addition, the central units may be employed as a mobility and control anchor. The benefits of this solution are the adaptation of the decisions to dynamic variations in the radio links conditions, consequently minimizing session interruptions or packet drops. Furthermore, in this configuration, appropriate feedback from UEs and operator preferences can be considered in the RATs selection.

Furthermore, in [1], multi-connectivity management approaches are presented:

- *User-Centric Approach*: with this solution the UE is continuously monitoring the radio links conditions, and, considering thresholds-based performance parameters (e.g. Signal to Noise Ratio (SNR)), the RAT selection can be performed. In advanced scenarios, the UE can consider other RATs characteristics (e.g. coverage) to better satisfy the application and user needs.
- *RAN-Assisted Approach*: User-Centric approach is limited to the local UE knowledge. For instance, the UE performs RAT selection based typically on the SNR, and in a highly dense environment the selection decision typically doesn't remain effective for long, due to the varying load of the RATs. The RAN-assisted approach employs network assistance from the RAN to the UE for RAT selection decisions. An example of assistance parameters can be network load, RAT utilization, expected resources allocation, etc.
- *RAN-Controlled Approach*: the above-mentioned schemes are user-centric by nature, resulting in suboptimal decisions from the overall system performances point of view. The RAN-controlled approach places the multi-connectivity control in the radio networks. In this approach the RAN can assign the UEs to certain RATs. Such a solution can be distributed across RATs or may utilize a central unit that manages radio resources across several cells/RATs. The UEs, in this solution, are configured to report radio measurements on their local radio environment. This solution is adopted by 3GPP for addressing dual-connectivity issues.

The *RAN-Based Integration* and *RAN-Controlled* approaches have been adopted in the Metis-II project [2] and also presented in the 5GPPP document [3]. The Metis-II project was aimed at designing an access network architecture and proving technical enablers for efficient integration between RATs. The project proposes an architecture as depicted in Figure 2 (b) composed by several Air Interface Variants (AIVs) and a central unit with AIV-agnostic functions, that based on the real-time feedback provided by the AIVs, is capable to steer the Quality of Service (QoS) Flows in a dynamic way on the different AIVs.

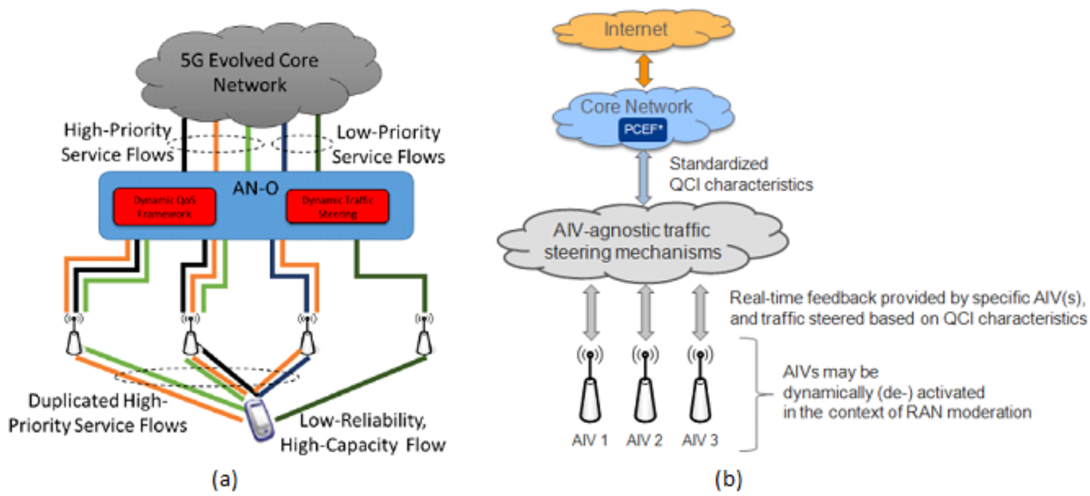

Figure 2: (a) Dynamic Traffic Steering framework in 5GPP [3]. (b) Architecture for traffic steering and RAN moderation in Metis-II [2]

Further considerations can be discussed on the *RAN-Based Integration*. In [4] three types of *RAN-Based Integration* are presented:

- *Hard Handover* (HH) enables users with poor coverage to switch to another RAT which can provide better coverage. The HH requires quite extensive Radio Resource Control (RRC) and CN signalling, as well as cell search and synchronization, which results in relatively long interruption delays. Another drawback is the low reliability since the users can be connected to only one RAT at the time.
- *Fast Switch* (FS) assumes to have common Control Plane between different technologies. In FS no signalling is required for User Plane switch, and it can respond almost instantaneously when the channel quality variations occur. Another benefit brought by the use of FS consists in the increased reliability since the users can be connected to multiple RATs at each time. Obviously, even FS has some drawbacks, these are due to the presence of greater overhead for the increased control signalling.
- *(User Plane) UP Aggregation* assumes to have both UP and CP connected to all the RATs, and that the UP data is aggregated. The benefits in this case are increased throughput, resources pooling, and the support for reliable seamless mobility. Notice that these benefits may be limited, due to the different latency and throughput of the involved RATs.

Typically, the UP Aggregation can be considered as the best choice to increase throughput and reliability, while decreasing signalling and switching time. The adoption of UP Aggregation or Fast Switch imposes the architectural constraints that have been discussed above, e.g. they need common entities to accommodate the shared functionalities.

Several works have faced with the problem of Multi-Connectivity from architectural [2], [3], [12]–[14], [4]–[11] and algorithmic [15], [16], [25], [17]–[24] perspective. The main topic of discussion is the functional split among RAN components. Following the 3GPP architecture in Figure 3, the gNB can be composed of a Central Unit (CU), also called either Central-RAN or Cloud-RAN, and of several Distributed Units (DUs). This functional split between the Central and Distributed unit has the purpose of placing the co-operative, and technology independent, RAN functionalities in a central node, so that they can benefit from the advantages of centralization (i.e., centralized decisions, high computation power available), allowing flexible RATs selection/switch, such as “fast switch”.

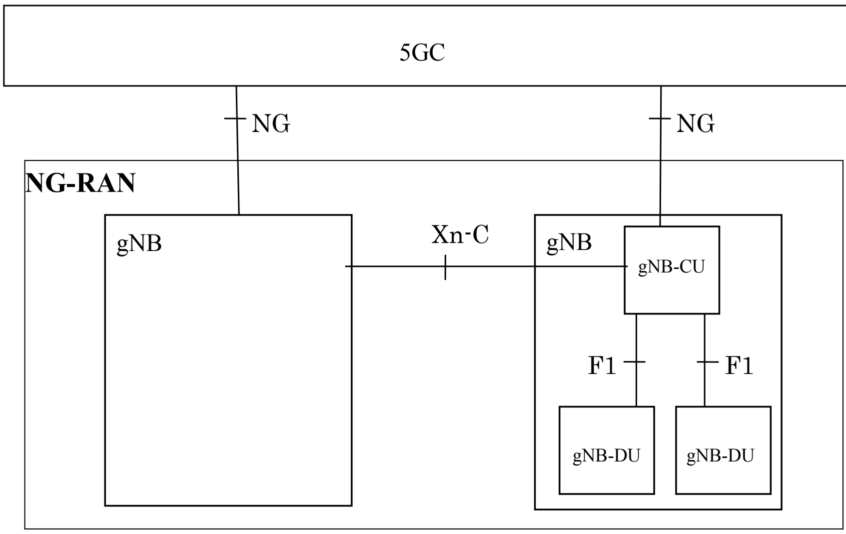

Figure 3: Overall RAN architecture [26]

The selection of the functional splits, i.e. the decision about which function should be placed in either the central or the distributed units, is a crucial point that defines the whole traffic flow control system.

From the control plane perspective, the ideal scenario would be to put the whole set of functionalities that are technology independent as well as Non-Real-Time and low bit rate, (e.g. traffic steering, spectrum sharing, etc…) in the central unit in order to have a complete view of the system, allowing optimal decision making. In this case, the distributed units have technology dependent, Real-Time and high bit rate functionalities in order to meet their requirements.

Regarding the protocol stack split, there are three main options as shown in Figure 4:

- Intra-PHY split: in this case, we face the requirements of low latency (about 1 ms one-way delay) and high throughput in the fronthaul, but there is no requirement of high computing power in the distributed units;
- PHY-MAC split: in this case the throughput is reduced compared with the intra-PHY split, but the same latency constraints are present. In this case, the amount of computing power in the distributed units is higher than in the previous case;
- Packet Data Convergence Protocol (PDCP) split: this case is the most attractive for the relaxed latency requirements (tens of ms) with a throughput comparable to the PHY-MAC, but there is a significant need for computing power in the distributed units.

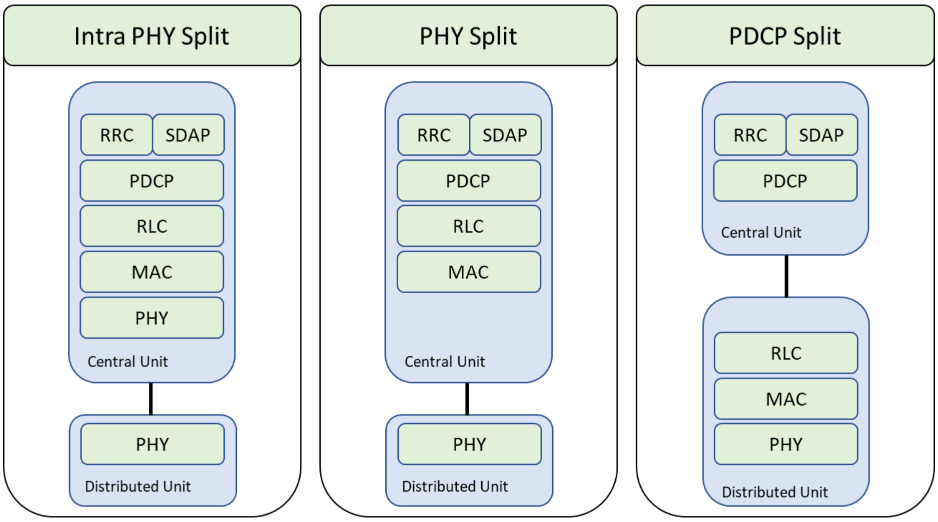

Figure 4: Functional Splits

A suitable choice is a common PDCP for the user plane and a common RRC for the control plane. In contrast to PHY, MAC and Radio Link control (RLC) functions, the PDCP functions do not have rigorous constraints in terms of synchronicity with the lower layers. Furthermore, this option will allow traffic aggregation, as it can facilitate the management of traffic load. This split has already been standardized for LTE Dual Connectivity [27].

The most recent discussions about Multi-Connectivity concern the selection of the appropriate technologies considering services and user requirements. In [28], [29], the traffic steering problem was investigated, where the traffic steering is defined as the function of distributing the traffic load optimally across different network entities and spectrum bands, considering operator and user preferences. Furthermore, the traffic steering problem needs to consider the characteristics of the different RATs, such as coverage, latency and capacity and the different traffic characteristics, defined by the QoS requirements. The goal of the traffic steering is to efficiently deliver capacity to the UEs satisfying the QoS/QoE requirements, considering the network management operations, such as energy saving, load optimization, interference management and congestion management. Traffic steering can be performed in centralized or distributed fashion. In both cases it allows the traffic steering algorithms to access information regarding load and serving capacity of all the RATs. Centralized coordination can achieve optimal performances, but this solution may introduce problems in the real-time control of the RATs. On the other hand, a distributed implementation can only perform traffic steering based on local information achieving suboptimal solutions.

The main decision for traffic steering in multi-connectivity scenarios is represented by the RAT selection issue [17], [19], [20], also known as network selection problem in heterogeneous networks. The problem consists in the selection of the most appropriate access network with characteristics able to satisfy the 5G KPI requirements. These selections can be performed by considering different network features as for instance: the mobility of the network nodes, the QoS attributes, the energy constraints, etc... .
The algorithms capable to perform the RAT selection are evaluated by considering the algorithms characteristics like computational complexity, implementation complexity,

distributed or centralized deployment with either open or closed-loop type, dynamic or static behaviour, model-based or data-driven.

RAT selection approaches, already investigated in the literature, concern the use of mathematical theories with the main characteristics detailed in Table 1. The mathematical approaches investigated will be described in the following subchapters.

3.1.1 Utility Theory

Utility theory is based on the concept of “utility” and its associated “utility functions”. In the several works [17],[30],[31] identified, the utility functions were characterized by different attributes, chosen in order to capture the networks’ and users’ characteristics, depending on the use case requirements.
Each of the monitored attributes is then associated with a function, whose type is selected so that it models correctly the behaviour of its associated characteristic (common choices for function types are linear, sigmoidal, exponential, logarithmic and polynomial).
The overall utility function is then the summation of all the attribute functions, and the goal of the network control is to maximise it.
In general, this approach requires the selection of the attributes of interest and the associated utility functions. In [31] three utility functions are designed to analyse different user attitudes to risk for economic benefits and delay preferences. The choice of attributes and their associated functions are taken to represent three different users risk profiles:

1. U_1, risk neutral users, that are the standard, in-between users.
2. U_2, risk seeking users, that pay more but experience less delay;
3. U_3, risk adverse users, that pay less but experience more delay.

The choice of the different utility functions modifies the connection experience for the users. In particular, for smaller file sizes, the price difference between U_3 and U_2 seems negligible, making U_2 the best risk profile to account for the smaller percentage of transfers exceeding the time deadline. For large files, the price difference may deter low budget users from this choice.

3.1.2 Multiple Attribute Decision Making

Multiple Attribute Decision Making (MADM) [15], [18], [19], [28] is an approach based on the selection of different alternatives, each characterized by different attributes.

The problem can be expressed in matrix form, where columns indicate attributes and rows indicate alternatives. Furthermore, the attributes shall be weighted, to represent the importance of each attribute in the choice, and normalization is needed to make the different attributes’ measurement units homogeneous. In [19] a selection network scheme is presented, based on two different MADM approaches, to select the best network in an integrated cellular/wireless LAN system, with the goal of providing the user with the best available QoS during the connection. The proposed scheme comprises two parts:

1. an analytic hierarchy process (AHP) to decide the relative weights of evaluative criteria set according to user preferences and service applications;

2. a grey relational analysis (GRA) to rank the network alternatives, because it is faster and simpler to implement than AHP.

The simulation results reveal that the proposed network selection scheme can efficiently decide the trade-off among user preference, service application, and network condition.

3.1.3 Fuzzy Logic

Fuzzy logic [18], [20], [23] solutions are based on the idea that an object or state cannot be assigned univocally to a class, but instead it is associated to a number between 0 and 1 for each of the available classes, representing the confidence of the classification. This method can be used to model the complex heterogeneous network system, and potentially can be coupled with other methods, e.g. MADM.

The work flow of this method begins with the fuzzifier, that maps the selected attributes into a fuzzy set. The second step is made by the fuzzy inference engine, that with the help of the fuzzy rule base, produces the maps that connect the input fuzzy set to the output fuzzy set. The output fuzzy set is then processed by the defuzzifier that takes the control decisions.

In [33] the authors adopted an algorithm based on a fuzzy logic controller (FLC) to evaluate fitness ranking of candidate networks. The selection is made in three phases: 1) pre-selection, to eliminate unsuitable networks; 2) discovery; 3) decision making. The discovery phase is based on FLC, where the fuzzifier maps the selected variables (network data rate, SNR and application requirement data rate) into the fuzzy set (i.e. the fuzzifier). Then, the inference engine uses these mapping functions as input to the predefined logic rule base (i.e. the inference engine). Finally, the defuzzifier obtains the overall ranking through defuzzification with a weighted average method.

3.1.4 Game Theory

In *Game theory* based approaches [17], [22], [34], [35], the problem is modelled with a set of players/agents (i.e. the decision makers), and a set of possible actions, defined as the strategy set of actions. At its fundamental level, this approach is based on the concept that each action, performed by a player, affects the actions of other players. Meanwhile, each player tries to maximize its own utility, or payoff, in an adversarial, non-cooperative, framework.

The optimal solution of this game is called Nash Equilibrium (NE), which is defined as the combination of strategies containing the “best” strategy for every player, in the sense that no player can improve its performances unilaterally. In the literature several network selection games can be found: i) games between users, where the users can be cooperative or non-cooperative; ii) games between networks, in this case the different networks shall decide the action to be taken (this is typical when each network is managed by a different operator); iii) game between users and networks, where the set of users and the set of networks are modelled by two different players. In this case, the users’ strategies are the selection of their favourable network, while the networks’ strategies are the selection of their favourable users. In this last case the NE can be reached only if the networks and the users correspondingly select each other. In [35] the network selection problem is modelled as a non-cooperative game between access networks. These access networks compete to decide the sub set of service to admit from a set of services. This selection is made maximizing a payoff that considers the users' satisfaction. Indeed, each access network admits the services that indicate the higher preferences towards the specific access network.

3.1.5 Combinatorial Optimization

Combinatorial optimization [17], [21] approaches the search for an optimum object in a finite collection of objects. These approaches are characterized by a cost function and a set of constraints and can be a good choice in case of convex formulation, which can be solved easily and, potentially, in a distributed way. In [21],optimal traffic splitting and aggregation solution are proposed. The problem is modelled as a maximization problem, whose cost function depends on the users' throughput and whose constraints imply that the user throughput is bounded by the sum of individual rate across all RATs. The peak rate of each RAT is considered known through the channel state information feedback. The algorithm was designed to be implemented in a distributed way, and the simulations showed an increasing of the throughput.

3.1.6 Markov Chains

Markov chains [16], [17], [24] are one of the most common tools for decision making, and are used for addressing several different problems. The most common approach is the Markov Decision Process (MDP) one, which is used to describe and model a stochastic dynamical system. MDP is defined by a state space, a set of available actions for each state, a transition probability function and a reward function. When the system is in a given state, the decision maker selects an action, which causes the system to evolve to the next state according to the transition probability function. The decision maker then collects the reward associated with the new state, which is an index of its performance in the new state. In [16], a Markov chain approach for network selection is used. In the formulation, each state is defined by the users' satisfaction probability for each radio technology. In each state, the possible actions are defined by the choice of radio technology to use. The proposed learning algorithm follows the Reinforcement Learning paradigm, with the reward function defined as the ratio between the sum of the weighted users' satisfaction for each technology over the total number of users. The algorithm is compared with baseline load balancing solutions, in a scenario with two different user profiles corresponding to low- and high-bit-rate services. The results demonstrate the ability of the proposed algorithm to adapt to current system conditions while achieving high user satisfaction. The proposed algorithm showed better behaviour and lower signalling requirements compared with two load balancing algorithms.

Table 1: Key Characteristics of Mathematical Theories for Network Selection

	Utility Theory	MADM	Fuzzy Logic	Game Theory	Combinatorial Optimization	Markov Chain
Objective	Utility evaluation	Combination of multiple attributes	Imprecision handling	Equilibrium between multiple entities	Allocation of applications to networks	Consecutive decisions/ rank aggregation/ priority evaluation
Decision Speed	Fast	Fast	Fast	Middle	Slow	Middle
Implementation Complexity	Simple	Simple	Simple	Complex	Complex	Middle
Precision	Middle	High	Middle	High	High	High
Model-Based/Data-Driven	Model-Based	Model-Based	Model-Based	Model-Based	Model-Based	Data-Driven or Model-Based
Open/Closed-loop	Open-loop	Open-loop	Closed-loop or Open-loop	Closed-loop or Open-loop	Open-loop or Closed-loop	Closed-loop or Open-loop
Centralized/Distributed	Centralized	Centralized	Centralized	Centralized or Distributed	Centralized or Distributed	Centralized or Distributed

In the 3GPP standard [36], a particular case of Multi-Connectivity is presented, i.e., the above-mentioned Multi-RAT Dual Connectivity, in which the multiple Tx/Rx UEs may be configured to use resources provided by two nodes: the first node provides E-UTRA access, while the second node provides NR access. As detailed in section 3.2, in the 5G ALL-STAR project, we are designing a more general multi-connectivity approach that is not only limited to NR and LTE, but which also includes satellite, 5G cellular access technologies, and other terrestrial technologies such as LTE and WI-FI access.

In 3GPP Dual-Connectivity, as reported in Figure 5, the two RATs involved in the connection are identified as Master and Secondary Nodes (MN and SN). Furthermore, three bearer types across the Uu interface are defined: i) the *MCG bearers,* only the MCG radio resources are involved; ii) the *SCG bearers,* only SCG radio resources are involved; iii) the *Split bearers,* both MCG and SCG radio resources are involved.

Each SDAP entity is placed in one RAT. The MN decides which QoS flow should be assigned to each SDAP entity. The MN or SN node, that hosts the SDAP entity, for a given QoS flow decides how to map it to DRBs.

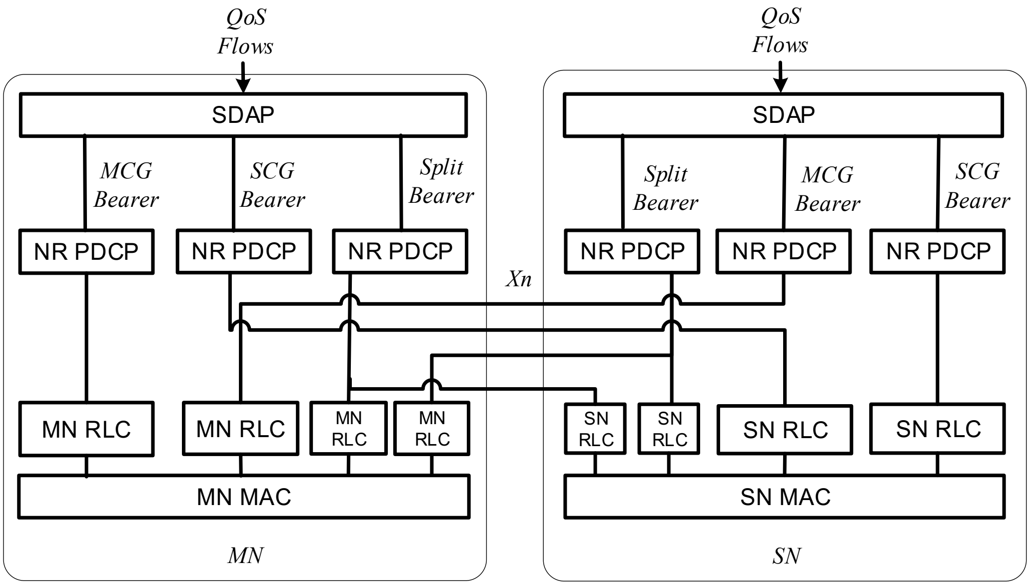

Figure 5: Network side protocol termination options for MCG, SCG and split bearers in MR-DC with 5GC, [36]

3.2 Beyond State of the art with 5G-ALLSTAR

This section provides an overview of the contribution given by the partners involved in Task 4.1 for designing the preliminary[1] 5G-ALLSTAR Multi-Connectivity with respect to the 5G Network Architecture [3]. The Multi-Connectivity architecture shown in Figure 6 is an insight into the general architecture presented in 5G-ALLSTAR Deliverable 2.2[2].

The 5G-ALLSTAR Multi-Connectivity Architecture, depicted in Figure 6, has been conceived to be in line with the current vision of distributed resource control and to maximize the whole user's experience by optimizing the radio performances. In this regard, the 5G-ALLSTAR project proposes to enrich the current entities, i.e., Core Network, Cloud Radio Access Network and Distributed Radio Access Networks, with elements (i.e., advanced functionalities) able to satisfy and control specific end-to-end Services/Applications by using both Traffic Flow Control and Quality of Experience (QoE) Control.

The QoE/QoS Management (see Figure 6) is a module, developed ad hoc in the 5G-ALLSTAR project, aimed at enhancing the "standard" QoS Control that 5G implements through the QoS Profile. Such an enhancement is performed by means of the so-called Connection Preferences which are deduced by the QoE/QoS Management taking into account personalized connection requirements aimed at satisfying even the subjective QoE of the user handling the connection in question; so, the Connection Preferences include QoE-related requirements which are additional with respect to the QoS-related requirements included in the 5G "standard" QoS Profile.

The QoE/QoS Management is logically located into the Core Networks and includes:

[1] The consolidated version of the 5G-ALLSTAR Multi-Connectivity Architecture will be provided in Deliverable 4.2
[2] 5G-ALLSTAR Deliverable 2.2 – "Preliminary architecture, API and interface"

1. the **QoE Management Repository** which includes information relevant to each past and in progress connection managed by the considered CN. Each record of the QoE Management Repository is relevant to a connection, either already terminated or in progress, which has been set-up in the considered CN. In particular, for each connection (i.e. for each record), such repository stores the following information (namely, the following three fields (a), (b), (c)):
    a) the so-called *Connection Id* which, at each connection set-up, is provided by the CN, in order to identify the key parameters of the connection; in particular, the Connection_Id includes the following subfields: (i) the Source UE Id, (ii) the Destination UE Id, (iii) the Service Type, (iv) the QoS Profile (including the 5G QoS Identifier (5QI) assigned by the CN to the connection, (v) the UE type. The CN functionalities provide the needed information to 5G-ALLSTAR by following the service APIs properly developed for this scope.
    b) the so-called *Connection Preferences* deduced, at each connection set-up, by the QoE Control (see below) by means of AI-based algorithms able to suitably analyse the (big) data included in the QoE Management Repository. The Connection Preferences are not modified for the whole connection duration;
    c) the so-called *Connection QoE History,* which includes, for each of the cells which has served the Connection (if already terminated), or is serving the Connection (if it is still in progress) the following subfields: (i) the *Cell-Id*, (ii) the *Time Duration*, i.e. the duration of the time interval in which the cell has served the Connection in question, (iii) the *Cell QoS Performance* i.e. the performance, in terms of QoS Parameters which was experienced in the Cell during the time interval in which the Cell has served the Connection in question (see the cRRM description for further information), (iv) *Implicit QoE Feedbacks* (i.e. feedbacks related to the Perceived QoE computed by a suitable *QoE Estimation* module (see below). The information included in the Connection QoE History is provided to the QoS/QoE Management module by the Traffic Flow Control module (see below) which, in turn, takes some of the information included in the Connection QoE History from the cRRM module (see below) and/or from the *QoE Estimation* module. The *Explicit QoE Feedbacks* (i.e. feedbacks related to the Perceived QoE directly provided by the users involved in the Connection) is provided by the Content Providers directly to the QoE/QoS Management which stores such information in the repository. The updates of the Connection QoE History relevant to a given cell serving a given connection are provided by the Traffic Flow Control module to the QoS/QoE Management whenever the Cell in question no longer serves the Connection in question.
2. the **Quality of Experience (QoE) Control**. The QoE Control, at each connection set-up, is in charge of deducing the Connection Preferences by analysing (by means of suitable AI-based techniques) the (big) data included in the QoE Management Repository. The rationale of the Connection Preferences is to include personalized QoE-related requirements which are additional with respect to the

QoS-related requirements which are associated with the standard 5G QoS Profile. At each connection set-up, the QoE Control (i) deduces the Connection Preferences, (ii) stores them in the QoE Management Repository, (iii) sends the Connection Preferences, related to a specific Connection and UE, to the gNB-CU and, in particular, both to the Traffic Flow Control module and to the cRRM module. At each Connection termination, the QoE Control has to inform the gNB-CU about the termination of the Connection.

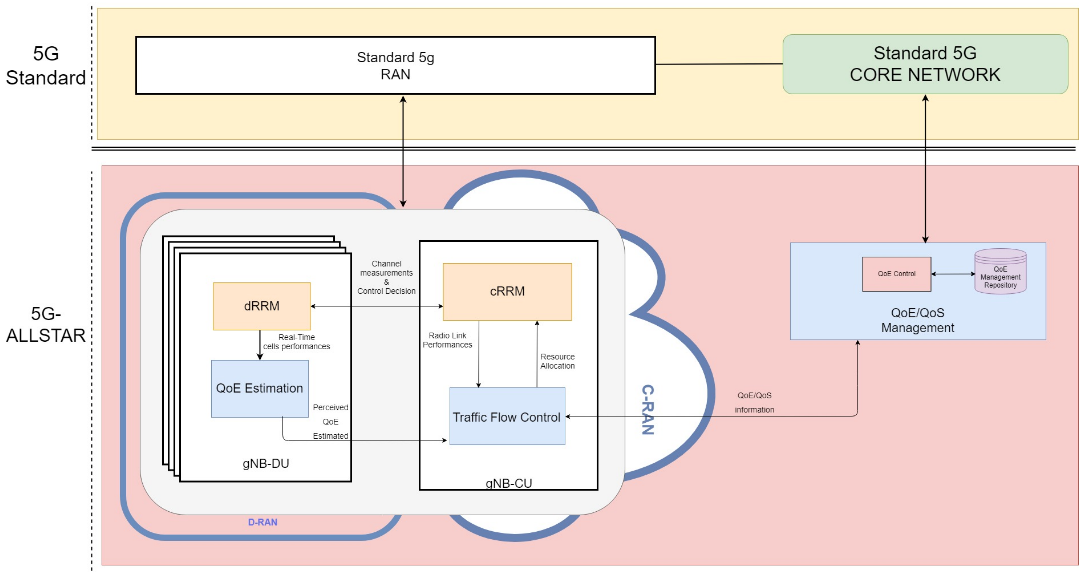

Figure 6 - 5G-ALLSTAR Multi-Connectivity Functional Architecture (Control Plane)

In the Cloud Radio Access Network, we consider a Centralized Unit, i.e., a gNB-CU, that provides Centralized control functionalities and thatis also able to manage the Distribute Units (gNB-DU) in the Distributed RAN (D-RAN) (each gNB-DU represents a RAT). Each gNB-DU manages several cells belonging to the same RATs; The area covered by the cells controlled by a given gNB-DU is referred to as *gNB-DU Area.* The area covered by all the gNB-DU managed by a given gNB-CU is referred to as *gNB-CU Area.*

The control functionalities implemented in the gNB-CU and in the gNB-DUs can be deployed in dedicated cloud-based servers. The 5G-ALLSTAR functionalities included in the gNB-CU are:

1. The **centralized Radio Resource Management** (cRRM) The cRRM manages a set of decentralized Radio Resource Management (dRRM) modules (see below). The cRRM, in cooperation with the managed dRRMs, includes algorithms and strategies able to control the multi-RAT radio bearers, i.e. able to dynamically select the appropriate radio bearers which should support the various Connections following the allocations (as explained below, referred to as *Connection-to-Cell Allocations*) decided by the Traffic Flow Control.

   In addition, the cRRM, in cooperation with the managed dRRMs, is in charge, for each of the Cell included in the *gNB-CU Area* and for each possible 5QI, of periodically computing the *Cell QoS Performance*, i.e. the performance, in terms of QoS Parameters (e.g., bit rate, latency, etc..) experienced in the considered Cell by the connections characterized by the considered 5QI; this means that the Cell

QoS Performance is expected to include a number of subfields equal to the number of different QoS Parameters; each QoS Parameter is computed as the weighted mean of the performance experienced, with respect to such parameter, by all the Connections served by the considered Cell and characterized by the considered 5QI. The information included in the *Cell QoS Performance* is periodically sent, together with the associated *5QI* and the associated *Cell_Id* from the cRRM to the Traffic Flow Control. This latter stores such information and, in turn, forwards an elaborated version of such information towards the QoS/QoE Management.

Furthermore, the cRRM, in cooperation with the managed dRRMs, is in charge, for each of the in progress Connections, to continuously monitor which are the Cells of the gNB-CU Area that, potentially, can serve the Connection in question; such cells will be referred to as *Potential Serving Cells*. For each, in progress Connection, the *Potential Serving Cells_Id* (i.e., the list of the Ids of all Potential Serving Cells of a given Connection) together with the relevant *Connection_Id* is periodically sent from the cRRM to the Traffic Flow Control which stores this important information in the In-Progress Connection repository (see below).

2. **Cell QoS Performance Repository**. This repository includes a record for each <Cell, 5QI> couple in the gNB-CU Area. In particular, for each Cell (i.e. for each record), such repository stores the following three information (namely, the following three fields (a), (b), (c))) all received from the cRRM:
    a. the *Cell_Id*
    b. the reference *Connection 5QI* which, in general, corresponds to a 5G QoS Profile. Note that the necessity to store this parameter even in this repository derives from the fact that the *Cell QoS Performance* experienced in the considered Cell by Connections characterized by different 5QI can be remarkably different since "standard" 5G QoS-assurance mechanisms (outside of 5G-ALLSTAR modules) can deal with in very different ways Connections characterized by different 5QI fields.
    c. the *Cell QoS Performance* which is currently experienced in the considered Cell (namely the Cell having the Cell_Id stored in the field (a)) by the Connections which have been assigned a given Connection 5QI (namely the one having the *Connection 5QI* stored in the field (b)).
3. **In Progress Connection Repository**. This repository includes a record for each in progress Connection. In particular, for each in progress Connection (i.e. for each record), such repository stores the following five information (namely, the following five fields (a), (b), (c), (d), (e))):
    a. the *Connection_Id* (information received from the QoE Control);
    b. the *QoE Implicit Feedback* which has been estimated by the QoE Estimator during the whole in-progress Connection.
    c. the *Actual Serving Cells* which have been selected to serve the in-progress Connection
    d. the *Connection Preferences* which have been deduced by the QoE Control (information received from the QoE Control);

   e. the current *Potential Serving Cells_Id,* including the Cell_Ids of all the Cells which, currently, potentially, can serve the in-progress Connection (information received from the cRRM);
   f. the current *Connection-to-Cell Allocations,* i.e. the current allocation, decided by the Traffic Flow Control module, of the traffic generated by the considered Connection (namely the one having the Connection_Id stored in the field (a)) to the Cells of the gNB-CU Area. Such allocation is expressed in percentages, i.e. it specifies the percentage of the traffic relevant to the considered Connection which has to be managed by each of the Cells of the gNB-CU area. Note that the necessity to store this information derives from the fact that, most likely, the *i*-th iteration of the Traffic Flow Control algorithm will provide a Connection-to-Cell Allocation expressed in terms of "Delta" with respect to the allocation decided in the (*i*-1)-th iteration. As a matter of fact, for each in progress Connection, the various iterations of the Traffic Flow Control are expected to "smoothly" move traffic amount among the serving Cells.

4. **Traffic Flow Control**. This module contains a set of strategies and algorithms (based on advanced control methodologies) able to dynamically decide, for each in progress Connection, the traffic bit rates that have to be managed by each Cell of the gNB-CU Area. So, the output of the Traffic Flow Control is the so-called *Connection-to-Cell Allocations* periodically specifying, *for all the in-progress Connections managed by the gNB-CU*, the percentages of traffic which have to be managed by each Cell. At each Traffic Flow Control algorithm iteration, the generated Connection-to-Cell Allocations are sent to the cRRM which, in cooperation with the managed dRRMs, must dynamically select the appropriate radio bearers which have to support the various Connections in the various Cells just following the received Connection-to-Cell Allocations.

   The decisions of the Traffic Flow Control are taken on the basis of the information stored in the Cell QoS Performance Repository and in the In-Progress Connection Repository. In particular, the Traffic Flow Control algorithms have to simultaneously take into account, for each in progress Connection, (i) the connection QoS requirements (implicitly, included in the Connection 5QI (field (b) of the In Progress Connection Repository)), (ii) the connection QoE requirements (included in the Connection Preferences (field (c) of the In Progress Connection Repository)), (iii) the current potential serving Cells (field (d) of the In Progress Connection Repository)), (iv) the current QoS Performance of the potential serving Cells experienced by the Connections characterized by the given Connection 5QI (information deduced from the Cell QoS Performance Repository).

The Distributed Radio Access Network (D-RAN) is designed to be equipped with advanced features able to capture real-time information from the UEs. These features are:

1. **dRRM.** The distributed Radio Resource Management, from the traffic flow control point of view, is designed to be in any gNB-DU for estimating the radio performance as closest as possible to the UE. The dRRM of a given gNB-DU provides, in real-time, the cRRM with the Cell QoS Performance and associated 5QI relevant to the Cells of its gNB-DU Area.
2. **QoE Estimation.** This module integrates, for any gNB-DU, the algorithms in charge of managing the *Implicit QoE Feedbacks* (i.e. feedbacks related to the

Perceived QoE computed by the QoE Estimation). The Implicit QoE Feedback is forwarded to the Core Network, through the Traffic Flow Control in the gNB-CU.

The communication among the various entities, i.e., the Core Network, the C-RAN and the D-RAN are performed with specific Interfaces presented in [26]. The mentioned entities exchange data information concerning the user's requirement, traffic and channel performances. The communication among entities and software modules implemented in such entities is a mandatory step for enabling the network optimization and control.

An Example: A Downlink communication procedure

In Figure 7 a downlink communication procedure in 5G-ALLSTAR system is presented. Such procedure involves four entities: i) the UE; ii) two gNB-DUs, where the gNB-DUs are characterized by different radio technologies; iii) the C-RAN, that acts as a control and user plane anchor; iv) the CN, composed by the QoE Management functionalities. The procedure represents a scenario in which a Downlink (DL) stream comes from the Data Network to the UE.

1) The Connection Preferences (that will be detailed in Chapter 4) assigned to the traffic flows is performed based on the current and past (user and service) information stored in the QoE Management Repository (UE type, service type, user's connection preferences, …);
2) traffic steering is performed by the traffic flow control based on the QoS of the traffic and the RATs information;
3) UE connection procedure to the selected RAT is performed and the DL data are sent to the UE;
4) UE send measurement report about radio link periodically during the connection;
5) Based on this information the traffic is switched from the actual RAT to another RAT by the traffic flow control;
6) at the end of the connection, each user sends feedback about its satisfaction and based on this feedback, and the other information detailed in the next sections, the QoE Control module modifies the users' Connection Preferences.

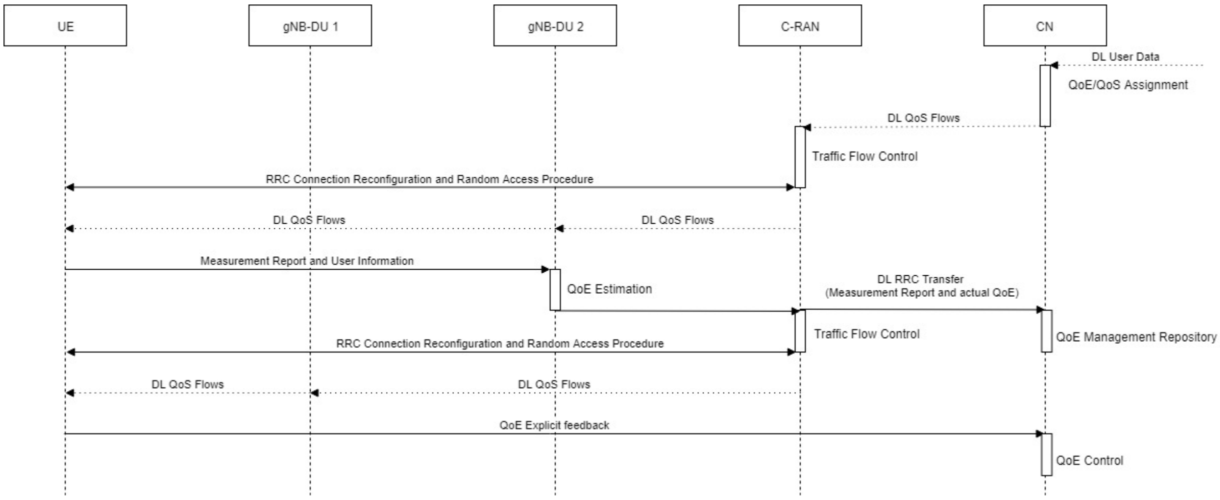

Figure 7: Downlink procedure in 5G-ALLSTAR System

3.3 Multi-connectivity functionalities

The Multi-Connectivity, as already introduced, is performed to configure a UE to be connected to multiple and heterogeneous access nodes. In a Multi-Connectivity scenario, the UE is capable of sending/receiving traffic through radio bearers established with different radio technologies, performing traffic split and switch cases. In the former case, the traffic is split and sent to simultaneously to different radio bearers. In the latter case, the traffic sent or received through a radio bearer, can be switched in one or more radio bearers.

The Multi-Connectivity is used to increase the reliability and the throughput connections by making use of above-mentioned splitting and switching functions; such functions permit to adopt the *"always-best connected"* approach with a dynamic selection of the best technologies in charge of satisfying the user requirements.

As mentioned above in the previous chapters, the 5G-ALLSTAR project deals with

1) RAN-Integration approaches [1];
2) RAN-Controlled approaches [1] ;
3) PDCP/RLC split [37]

since as already introduced they will allow adaptation to dynamic variations in the radio links conditions, avoiding session interruption or packet drops.

In the 5G-ALLSTAR EU-side design, considering the control plane functional architecture defined in Figure 6, an example of Multi-Connectivity physical architecture with three RATs and two UEs in a downlink scenario is presented in Figure 8.

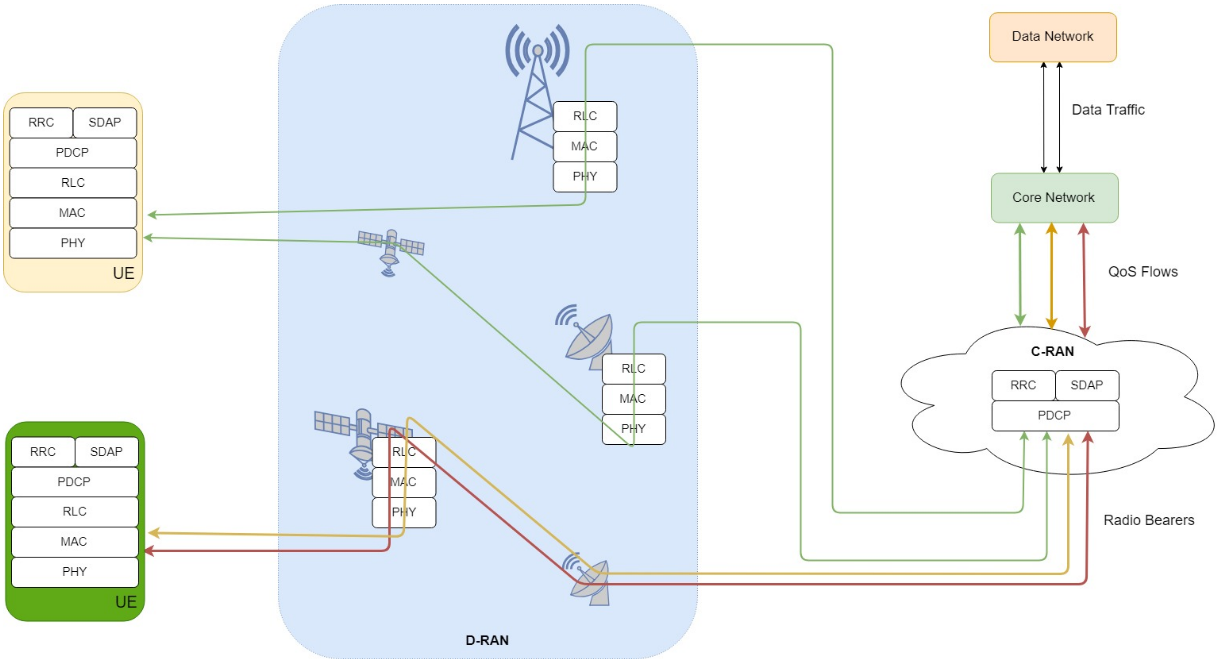

Figure 8: Multi-Connectivity Physical Architecture

Differently from Dual-Connectivity presented above as shown in Figure 6, the whole set of available RATs have common RRC and partially common UP (SDAP and PDCP layers). The RRC and PDCP common layers approach in the C-RAN bring several advantages i.e., the fast switch/UP aggregation and PDCP split.
The architecture in Figure 8 is composed of three RATs: i) a terrestrial BS; ii) a transparent satellite, and iii) a regenerative satellite. The figure presents two data flows (Data Traffic) coming from the Data Network. The Core Network divides the data flows in three QoS Flows by using the UPF functionalities (considering the Session Management Function (SMF) configuration) and, in turn, the C-RAN sends the QoS Flows to the UEs with a proper selection of radio bearers.
The C-RAN entails RRC functions capable to (i) configure the SDAP for the mapping of QoS Flows into data bearers; (ii) configure the other UP layers to establish the data bearers with the desired performances.
These functionalities are performed by the cRRM and Traffic Flow Control module as defined in the description of Figure 6.

The above-mentioned description applies to EU-side. The KR side designs different MC functionalities that are required to provide inter RAT multi-connectivity as identified in [6]. The main function is the radio access technology (RAT) selection that depends on inter RAT traffic management function. For offloading UEs data to any of the RAT, an inter RAT traffic management function is needed that operates on the network layer. This function monitors the flow of traffic and offloads UEs traffic to suitable RAN based on traffic type. The selection of suitable RAT using inter RAT traffic management function results in load balancing in the multi-RAT network limiting delays and freeing bandwidth. Optimizing inter RAT traffic management utilizes available resources in multi-RAT network to increase the capacity of the network and satisfies the QoE of users.

Traffic management can consider different characteristic of traffic, e.g., data that are delay tolerant or require a high data rate. Based on these characteristics, a suitable

RAT can be selected to satisfy the QoS for users. AI technologies can be used to analyse and manage the traffic in multi-RAT networks from network data. Machine learning (ML) and deep learning (DL) are two advanced AI methodologies, which can be used to overcome the challenges of managing 5G multi-RAT network traffic. ML-based traffic flow classification can classify delay tolerant flow from delay sensitive flow of users by analysing the data packets header. In Non-Terrestrial Network (NTN) based 5G multi-RATs networks, delay tolerant traffic can be forwarded to satellite links and delay sensitive data route to terrestrial links which increase the network performance in terms of throughput and QoS.
Other function to inter-RAT multi connectivity is the operating mode selection. This function selects the mode based on the CP and UP configuration. Two different configurations are considered for both control and data planes in the multi-connectivity physical architecture shown in Figure 8. One the configuration is known as 1A (see Figure 9) in which there is only one CP for multi connectivity in multi-RAT networks and other one is known as 3C (see Figure 9) configuration where each RAT has the CP, as depicted in Figure 9. Based on these configurations, the operating mode for both planes is selected. For instance, in 1A configuration, CP operates on diversity mode allowing signalling through single RAT, whereas 3C configuration allows the UP to operate in a reliable mode in which data flow via multiple RATs. Similarly, these configurations can be considered for the UP. As to the traffic flow of UE across multiple RATs, therefore, additional functionality for PDCP sequence number synchronization is required.

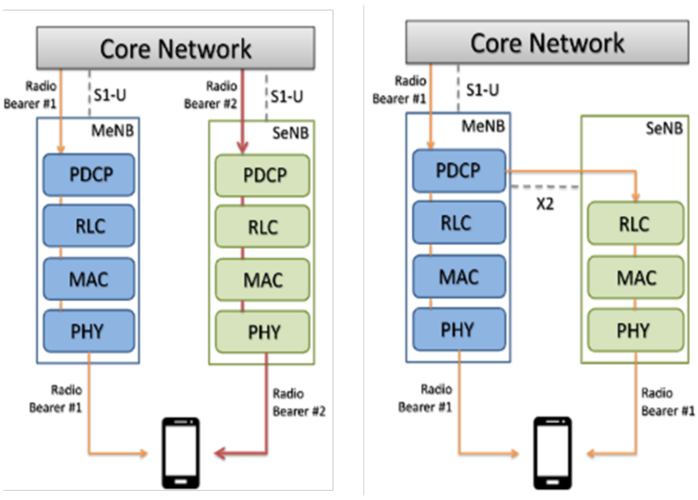

Figure 9: 1A configuration and 3C configuration for multi connectivity in multi-RAT [38]

3.4 Non-Terrestrial Systems in 5G

In the 5G-ALLSTAR project, the NTNs are considered as a key technology to reach the desired KPIs with Multi-Connectivity. The different characteristics of NTNs (e.g., coverage, reliability, latency, etc…) make such networks able to satisfy several service requirements. These networks combined with terrestrial networks and by adopting Multi-Connectivity approaches may satisfy a wide range of services requirements.
The real challenge in this case is presented by the integration between the two different technologies (terrestrial and satellite networks). The NTNs differ from typical cellular networks in terms of network functional architecture, as well as deployment scenarios

and actual coverage which may span across several countries. An appropriated mapping of NTNs physical entities onto logical NG-RAN architecture can minimize the difficulty during the network integrations.

In [39] the logical RAN architecture that was considered is depicted in Figure 3. In line with this approach three main architectures have been proposed considering bent pipe or regenerative payload. The three proposed architecture are:

1. Figure 10 shows the bent pipe payload, where the network is composed of a gNB on the ground that contains User and Control Plane functionalities and the satellite is used as a relay, with only amplification, frequency conversion and filtering functionalities (PHY-low layer). This architecture can be used when the satellite has a stringent constraint in terms of power consumption, weight, etc. On the other hand, this architecture has disadvantages in terms of Control Plane traffic (i.e., a big amount of control signals is sent along the paths) and real-time functionalities (i.e., the control information used by real-time functions has Round Trip Time (RTT) delay);

2. Figure 11 shows the regenerative payload. In this case the functionalities are split between a gNB-CU on ground and gNB-DU on the satellite. The real-time function and the lower layer protocol stack (e.g., PHY, MAC, RLC) are placed on-board, to avoid the problem on control signalling.

3. Figure 12 shows the regenerative payload, all the network functionalities are placed on-board. In this case, the integration with terrestrial networks can be actually challenging due to the difficulties of communication and synchronization between terrestrial and non-terrestrial networks functionalities.

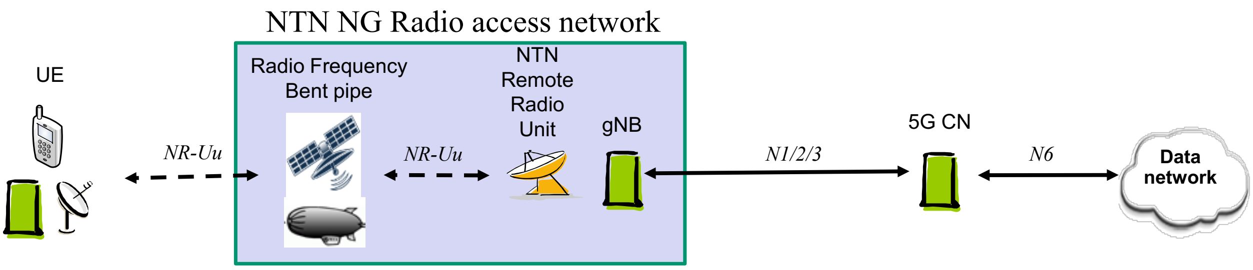

Figure 10: NG RAN architecture in Non-Terrestrial network with bent pipe payload [39]

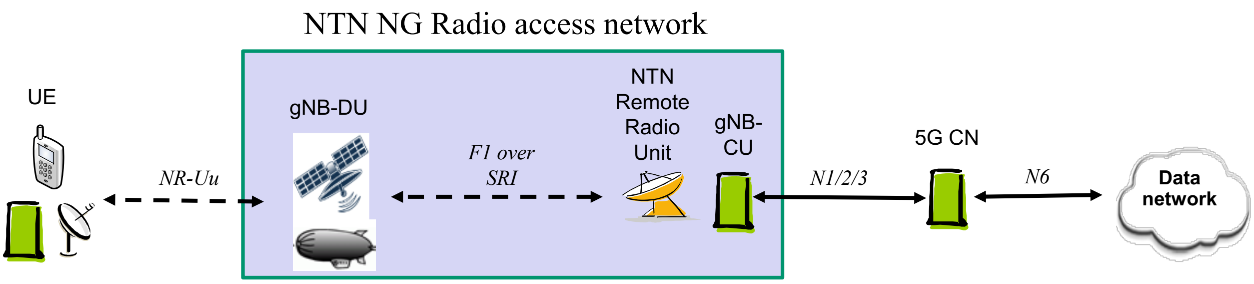

Figure 11: NG RAN architecture in Non-Terrestrial network with gNB-DU processed payload [39]

NOTE: SRI refers to Satellite Radio Interface

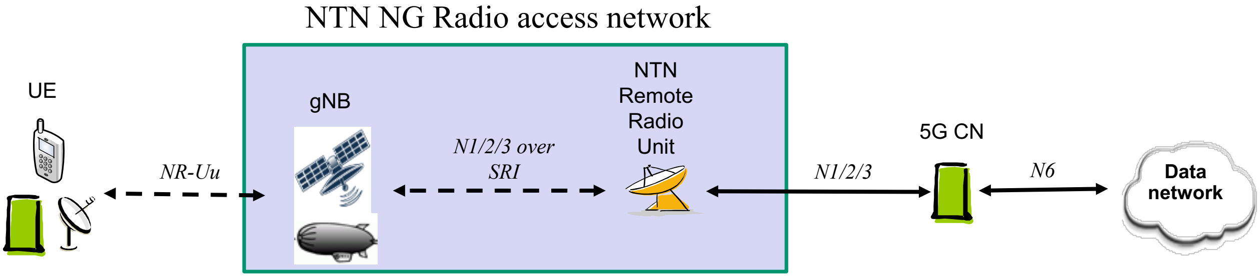

Figure 12: NG RAN architecture in Non-Terrestrial network with gNB processed payload [39]

The 5G-ALLSTAR project is investigating the adoption of the architectures presented in Figure 10 and Figure 11, thanks to their easier integration with terrestrial networks. Furthermore, these architectures allow the *RAN-Integration* and the PDCP split with the advantages defined in the previous sections.

From the KR-side perspective, in order to provide the 5G's three use cases of UMBB, URLLC, and mass-connectivity, the ITU-R M2410 document defines the KPIs such as user experience data rates of min. 100 Mbps for down-link and 50 Mbps for up-link in a dense urban area, average spectrum efficiency of 3.3 bits/sec/Hz for downlink and 1.6 bits/s/Hz for uplink, 10 ms user plane latency, and so on. In addition, some of the key drives of 5G technology are as follow.

- 100% global converge & mobility
- network consistency and security
- reduction in network energy usage
- 100 time higher data rates

5G technology has three characteristic use cases depicted in Figure 13 which are massive internet of things, enhanced mobile broadband, and mission-critical control or ultra-low reliable and latency.

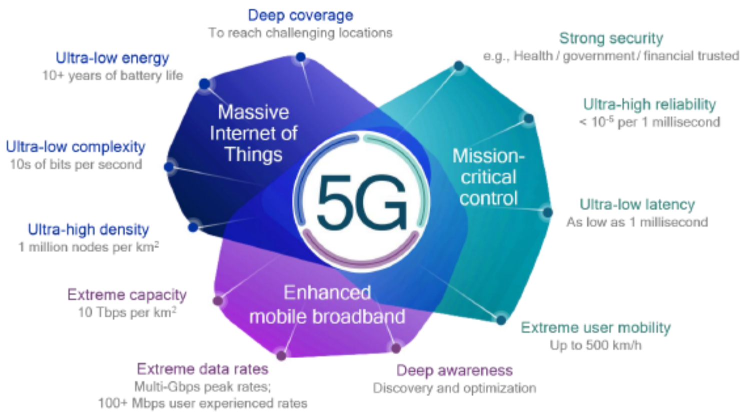

Figure 13: 5G use cases [5GUC].

3.4.1 Satellite role integrated in cellular system

From the KR-side perspective, satellite communications will be a vital portion of the 5G NTN setup to increase coverage and availability in a wide area. The integration of satellite and terrestrial systems in 5G increases spectrum availability, coverage zone, and global service ubiquity. Satellite communication systems cover an extensive service area. Therefore, NTN in 5G is capable to support multi-connectivity in a wide range coverage. Communication services through satellite are shown in Figure 14. The role and function of NTN in 5G are expected to provide services to service-deprived areas, which cannot be covered in TN. The challenges faced in NTN are as follows

- Very high propagation delay
- Continuation of service during handover between TN to NTN
- Extremely larger cell size
- The mobility of users

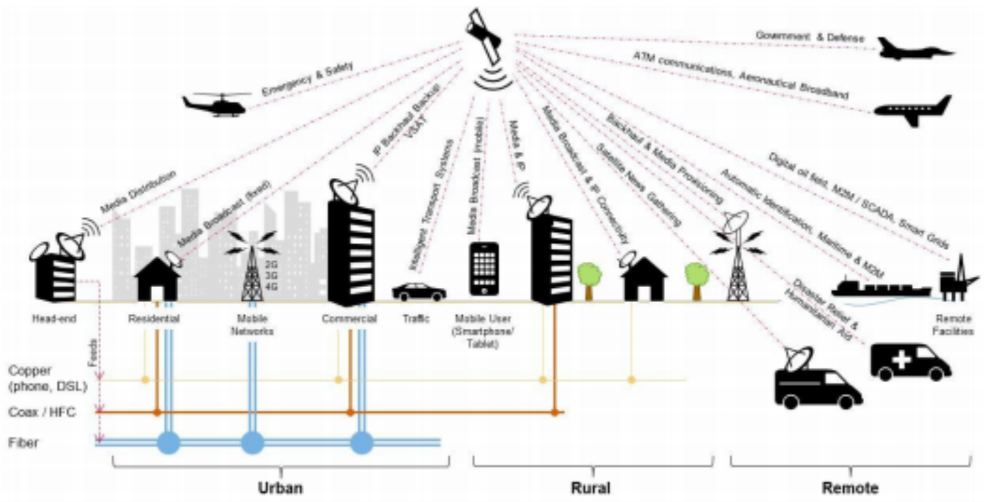

Figure 14: Satellite role integrated in cellular system [SRICS].

3.4.2 Network Architecture for NTN based 5G

From the KR-side perspective, different architecture of NTN based 5G network can be implemented based on service scenarios. Let us consider a Relay-capable UE deployed in a small office or a home in an underserved area. At cell edge, the cellular access will provide less performance than at cell center. One way to boost the throughput would be to combine the cellular access with a satellite access through the multi connectivity scheme. The corresponding architecture is depicted in Figure 15.

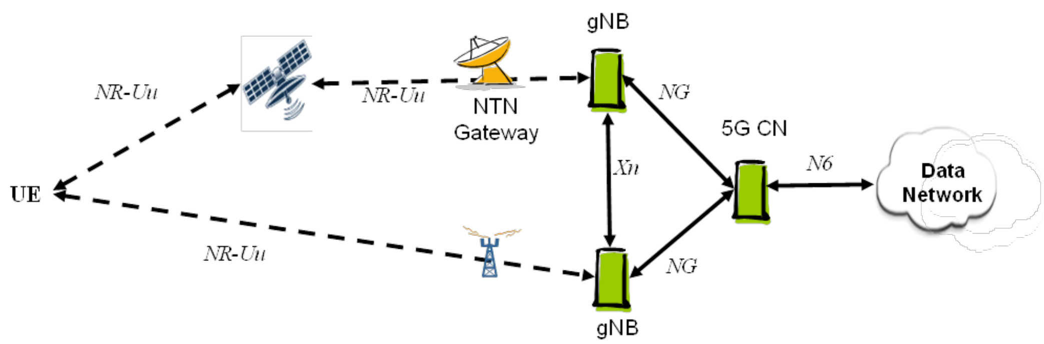

Figure 15: Multi connectivity involving satellite access (Transparent payload and NR based) and cellular access (NR or LTE based).

Another case to be considered, refers to the combination of two satellites with transparent payload either GEO or LEO or a combination of both. It is of interest to provide service to relay capable UE in a home or a small office in unserved areas. The LEO satellite access featuring relatively low latency would support the delay sensitive traffic while the GEO satellite access would provide additional bandwidth to meet the targeted throughput requirements. This is depicted in Figure 16.

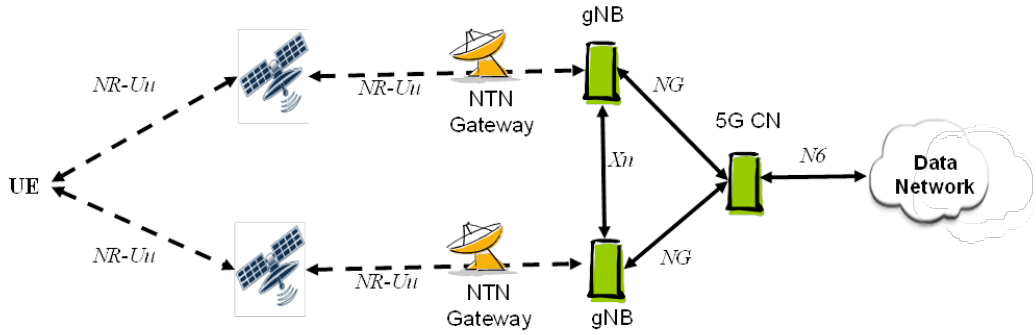

Figure 16: Multi connectivity between two NR based satellite access (Transparent payload).

The combination of two NR based satellites either GEO or LEO or a combination of both, with regenerative payload using Inter Satellite Links in between is also worth to consider providing service to relay capable UEs in a home or a small office in unserved areas. This is depicted in the figure below.

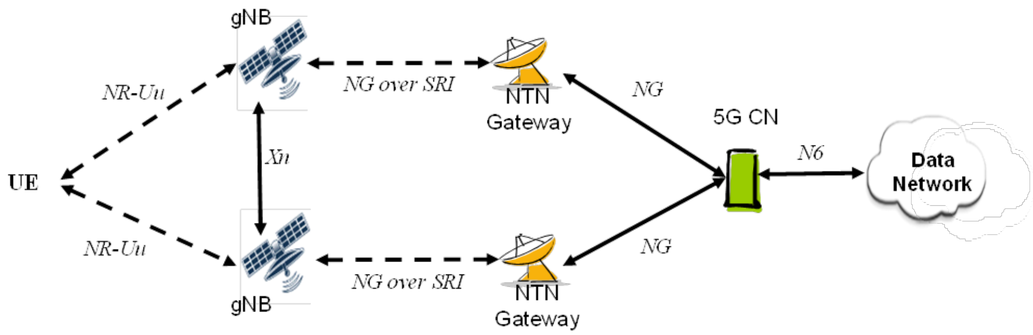

Figure 17: Multi connectivity between two NR based satellite access (both with Regenerative payload).

In addition to EU-side's, KR-side view point for the architecture includes one more multi connectivity architecture that is shown in Figure 18. This architecture provides the link between gNB and NR based satellite access for multi connectivity to edge users to boost the throughput and to reduce delay is worth considering.

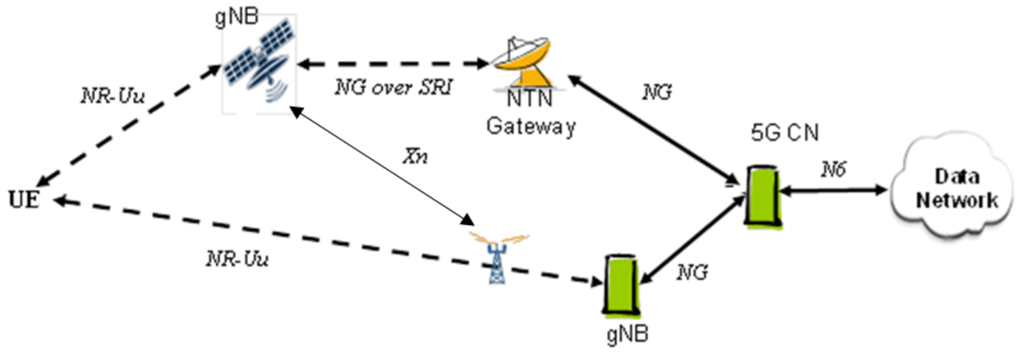

Figure 18: Multi connectivity involving satellite access (Transparent payload and NR based) and cellular access (NR or LTE based).

4 Quality of Experience Management

This section provides a draft overview of the approaches that will be investigated in a 5G-ALLSTAR project concerning the design of a whole set of Personalization System functionalities to cope with the Quality of Experience control in the Network system.

The Quality of Experience/Quality of Service Management is a fundamental part of a "Personalisation system", see Figure 19. The 5G-ALLSTAR Personalization System is aimed at providing a non-standardized Connection Preferences which enriches the inputs deployed to the Traffic Flow Control. These Connection Preferences will be considered during the Multi-Connectivity assignment tasks. The Connection Preferences contain a set of parameters deduced by the QoE Control module.

The Personalization System is also able to estimate the perceived QoE for each ongoing service/application at each UE by using the QoE Estimation module.

The Personalization System can be conceived in the 5G-ALLSTAR target system as a unique module able to expose a set of APIs (Application Programming Interface/s) for enabling the communication and integration among different modules involved in the Personalization System. The APIs are also designed for third-party usage and then to be triggered by the standardized Core Network functionalities (e.g., SMF, AMF). The Personalization System offers advanced functionalities to the Core Network.

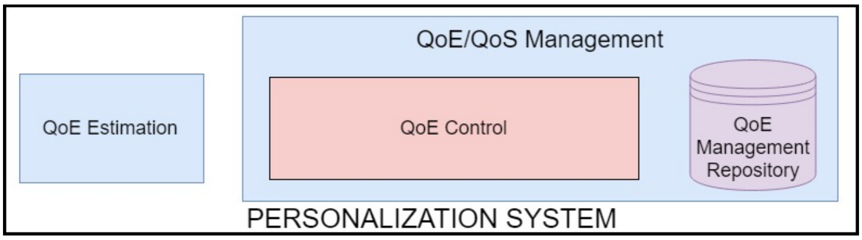

Figure 19 - Personalization System

4.1 QoE Management Repository

The QoE Management repository contains the entities as already mentioned in section 3.2. The repository is used to store individual UE information to foster the development of 5G-ALLSTAR advanced algorithms able i) to identify personalized parameters, ii) to assign such parameters to the Traffic Flow Control and iii) to cope with the personalized requirements, for each connection in each UE. In the target system, the SMF creates the new instances for each new *UE* and each new *Connection* in the 5G-ALLSTAR repository*;* it is also in charge of triggering the proper APIs for storing the *UE* and *Connection* related information.

4.2 QoE Control

The 5G networks aim at satisfying at the same time a large diversity of UE requirements. This requires the network to be flexible and adaptable to different traffic types, as defined by the 5G requirements which need a granular approach to the QoS handling. It is known that each UE can have one or more PDU sessions which may have one or more QFI (QoS Flows Identifier) at the same time. In this respect, a granular assignment of the QoS markers into the PDU session(s) for different traffic types is a fundamental feature in the design of a 5G network for both DL and Uplink (UL) cases. It also allows the Access Network to handle the data packets with different QFI, by assigning the different QoS flows to the most suitable data radio bearers.

The QoS Flow assignment system in the CN is composed of the packet classification step and the packet marking step. Packet classification is typically performed by means of a set of parameters, contained in the packets' header, such that a packet is univocally identified. Packet marking is performed after classification; typically, the packets that are classified in the same group have the same marker. In the 5G networks, these two steps are performed using Packet Detection Rules (PDRs). The PDRs are generated by control plane functions in SMF and are used in UPF to classify packets. Furthermore, the traffic classification can be performed using machine learning models, capable to classify the traffic using preselected features (e.g., packet length, port number, etc …).

The QoE Control module, in the 5G-ALLSTAR project, is developed to assign for each PDU session the enriched QoS Profile needed to cope with the related QoS Requirements. The QoE Control module relies on the information stored into the QoE Management Repository (see §3.2) and the service information provided by the data provider. The QoE Control module includes algorithms that are able to produce the Connection Preferences identified for each UE and Connection by correlating different input data.

The QoE Control is aimed at improving and guaranteeing users' degree of satisfaction while experiencing the contents and it can produce outputs for both DL and UL cases; in the DL / UL case the service provided by the QoE Control will be triggered by the SMF using the dedicated API. The mentioned approach doesn't require to have access to the packet payload and/or to use deep packet inspection in order to determine the Connection Preferences and other related information included in the QoS Profile [38]. In fact, the QoE Control by following the received QFIs which include the whole set of related information marked with these QFIs (the QFIs are computed in the CN with the standard CN functionalities) it provides only the personalized Connection Preferences. Such personalized parameters shall be assigned directly to the 5G-ALLSTAR Traffic Flow Control (see Figure 20).

The control-based strategies implemented by the QoE Control towards the personalized parameters are performed considering different input data from both (a) the data already stored in the QoE Management Repository (*Offline data*) coming from past connections and (b) the data provided in real-time (*Online data*) by the CN for each new connection stored in the QoE Management Repository.

The *Offline* data can be for instance the Implicit feedbacks, the explicit feedbacks or the selected RATs (the selected RATs input contains the set of RATs/cells used in the past connections related to the implicit and explicit feedbacks). The *Online* data can be for instance the Service Type/App and the QoS Profiles (see 4.2.2). The Service Type/App represents the applications or services requested by the UEs at each connection.

The concept of implicit feedbacks is introduced in §4.3; it is an unknown measure from the UE perspective, and it is computed by considering network performances by exploiting the QoE Estimation functionalities. In this respect, the implicit feedbacks are directly provided by the QoE Estimation for each UE and connection type (such measure is stored within the QoE Management Repository). The concept of explicit feedbacks entails the presence of a human being that evaluates her/his experience with a general content (i.e. video streaming, video conference call, mobile call, web browsing, etc…). Also the explicit feedbacks are essential [40] to provide tailored Connection Preferences and create a user-centric framework able to offer personalized services. The explicit feedbacks can be captured by the content providers by asking a simple and generic question to the end-user [40], for instance at the end of the experienced content. The explicit feedback is then stored within the 5G-ALLSTAR repository. The implicit and explicit feedbacks are processed by the QoE Control.

Whether the UE is not a user device the QoE Control evaluates the implicit feedback and the other input data to manage the Connection Preferences without considering the explicit feedbacks (it isn't released by non-user devices).

From the algorithms point of view, the QoE Control module, as it is in preliminary design, in the 5G-ALLSTAR project will be endowed by a set of functionalities by investigating techniques in the field of Machine Learning to determine, for each new connection, the optimal personalized Connection Preferences.

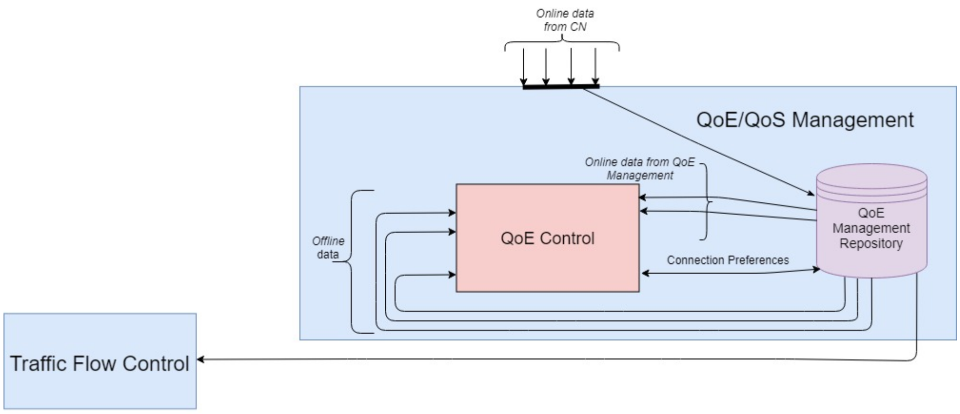

Figure 20 - QoE Control Design for Input/Output

4.2.1 QoE and QoS mapping

With respect to the above described EU-side QoE Control techniques and methodologies, the KR's design concept is as follows. 5G is expected to be a network capable of seamlessly delivering any application or service, regardless of bandwidth requirement, with perfect QoE. To meet the expectation, NTNs in 5G technology needs to satisfy several requirements: (i) consistency: the uninterrupted, seamless and invariable, but still excellent quality of the offered service; (ii) transparency: the requirement of the network to lessen its complexity and efforts on delivering excellent and seamless quality to its customers, (iii) user Personalization and Service Differentiation: comprehending the subscribers' expectations and differentiate the offered services accordingly; (iv) resource and Energy Efficient QoE-awareness: resource costs are maintained to a reasonable, minimum level.

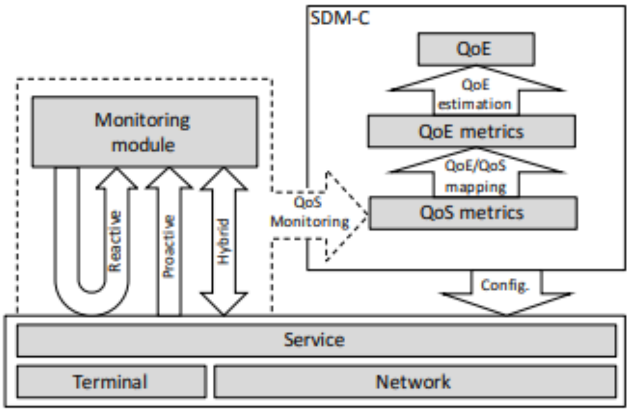

Figure 21: QoE/QoS mapping and monitoring [QOSM].

However, it is difficult to assess the QoE of the overall system due to its definition as a subjective measure of user' satisfaction. In an assessable approach, QoS is used as parameters to measure the degree of QoE, i.e., QoE/QoS mapping. Figure 21 depicts a high-level view of the QoE/QoS mapping. It also includes different monitoring mechanisms that can be used for QoS monitoring. However, monitoring QoS poses many difficulties for the 5G networks with the multiplication of the number of devices and the need for providing management processes (such as monitoring) with minimal signaling overhead. Hence, QoE enhancement remains as one of the most challenging problems for the forthcoming 5G. The relation between several QoS KPIs and the corresponding QoE presented in [41] which is shown in Table 2.

Table 2: Relationship between KPIs and QoE

KPIs	Expectation	QoE
Latency	Decreased	Increased
Connection density	Increased	Decreased
Capacity	Increased	Increased
Cost	Decreased	Increased
Battery life (D2D)	Decreased	Decreased
Mobility interruption time	Decreased	Increased

The ultimate goal of network services is to provide satisfactory, which is the quality experienced by the end user. QoE is directly related to human subjectivity, consequently, conventional QoS guarantee mechanisms do not necessarily realize QoE the end users desire [42]. QoS control, utilization of multimedia, and user interface are three methods to enhance the QoE. In order to assure QoE desired by users, KR-side thinks the design have to include some mechanism to cope with all possible combinations of user's attributes, contents, tasks, communication situations, and service prescriptions; the mechanism should be implemented into some part of the networks [42]. Another factor which impedes QoE assurance is the user's physical and mental conditions at the usage time as well as the user's attribute [42]. Hence, QoE enhancement through QoS, various models are required to control the relative parameters of QoS.

4.2.2 QoS Profile

For each QoS Flow as introduced in 3GPP [43] a QoS Profile is assigned which contains a set of different QoS parameters.
The useful QoS Parameters (according to [43]) used by both QoE Control and Traffic Flow Control in 5G-ALLSTAR will be:

- 5QI (i.e., the 5G QoS Identifier whose QoS Requirements are reported in Appendix)
- Allocation and Retention Priority
- Reflective QoS Attribute
- Guaranteed Flow Bit Rate for both UL and DL
- Maximum Flow Bit Rate for both UL and DL
- Maximum Packet Loss Rate for both UL and DL

4.3 QoE Estimation

QoE-aware resource control in the 5G network is essential to guarantee the migration from the network-centric approach to a user-centric approach. The user-centric network approach can be achieved by considering the user measures to compute the user personalized indicators, i.e., QoE measure. The QoE is a measure from the UEs' experience perspective instead of the QoS measures from network parameters [44]. The first step towards the QoE approach is the estimation of such values by considering the measured QoS. Several works have investigated potential approaches to express the QoS measures in QoE; the two promising approaches are based on the logarithmic and exponential conversion of QoS in QoE.

In the logarithmic conversion approach [45] the QoE variation is considered independent from the actual value of the QoE - the perceived level QoE for each UE in each ongoing service -, and depends only on the QoS variations:

$$\frac{\partial QoE}{\partial QoS} = -\frac{\alpha}{\gamma+\alpha QoS} \approx -\frac{1}{QoS} \tag{1}$$

By solving the differential equation, the QoE is:

$$QoE = -\ln\left(\alpha QoS + \gamma\right) \tag{2}$$

The α β and γ are constant values strictly assigned depending on the users' preferences. In the exponential conversion approach [46], i.e., the so-called IQX hypothesis (exponential interdependency of quality of experience and quality of service), the QoE variation is considered to be linked to the actual value of the QoE is:

$$\frac{\partial QoE}{\partial QoS} = -\beta(QoE - \gamma) \approx -QoE \tag{3}$$

By solving the differential equation, the QoE is then computed as:

$$QoE = \alpha e^{-\beta\, QoS} + \gamma \tag{4}$$

As already introduced, in the 5G-ALLSTAR project we are introducing a QoE Estimation module designed to be part of the Personalization System and deployed in any gNB-DU. The QoE Estimation will be implemented by following one of the two presented approaches in any distributed unit. For instance, by considering the approach related to the IQX hypothesis, the actual QoE will be computed as an exponential function of QoS_m (QoS measured at any UE level):

$$QoE_{estimate} = \alpha e^{-\beta\, QoS_m} + \gamma \tag{5}$$

The $QoE_{estimate}$ is used to compute QoE_e, that represent a QoE error, and is defined as:

$$QoE_e = QoE_{target} - QoE_{estimate} \tag{6}$$

where

$$QoE_{target} = \alpha e^{-\beta\, QoS_{target}} + \gamma \tag{7}$$

The real-time QoE Error, i.e. QoE_e, computed in the distributed units for each UE is a fundamental input data for the traffic flow control. When a generic connection *c* is established, the traffic flow *tf* related to *c* is associated to the radio resources based on RATs properties (e.g., resources, performances, etc…), Connection Preferences and QoS requirements associated to the traffic flow *tf*. As the connection *c* is on-going, the QoE Error is estimated based on equation in (6). Moreover, if the estimated QoE_e is larger than a given personalized (for each UE) threshold the Traffic Flow Control module performs proper control actions (e.g., traffic switching or splitting) trying to decrease the actual QoE Error (see equation in (5)) for the connection *c* in order to meet the UE requirements.

5 Traffic Flow Control in Multi RATs

In this section we present the preliminary solutions for multi-RAT, Multi-Connectivity functionalities identifying the relevant consideration behind the Multi-Connectivity solution in 5G networks, the relevant interfaces involved in the Multi-Connectivity solution according to the network architectures presented above (see §3.3 and §3.4), and a preliminary identification of the potential methodologies which are being investigated in the 5G-ALLSTAR project.

Multi-Connectivity consists in having multiple signal connections involving different Radio Access Technologies, either related to different radio technologies (e.g., 5G, LTE, 3G, SAT, etc…) or different cells (for both macro and small cells) belonging to the radio technologies, simultaneously used for a connection to serve a single UE (intended as multi-connectivity capable UE).

Multi-Connectivity will be performed at the C-RAN level, as already introduced in previous chapters. The implementation of the Multi-Connectivity is a fundament part of the 5G networks from, at least, three different point of views:

a) It may improve the overall data rates (throughput) to mobile UEs in 5G networks for both DL and UL.

b) It may improve the optimal exploitation of 5G network resources while meeting the 5G KPIs.

c) It can guarantee service continuity (reliability) in mobile UEs

With respect to point (a), it is clear that by exploiting simultaneous radio access technologies (or multiple cells of the same radio access technology) for transmitting the same data - splitting packets in two or more different directions - the whole data rate of a single connection of a UE will be increased (see Figure 22).

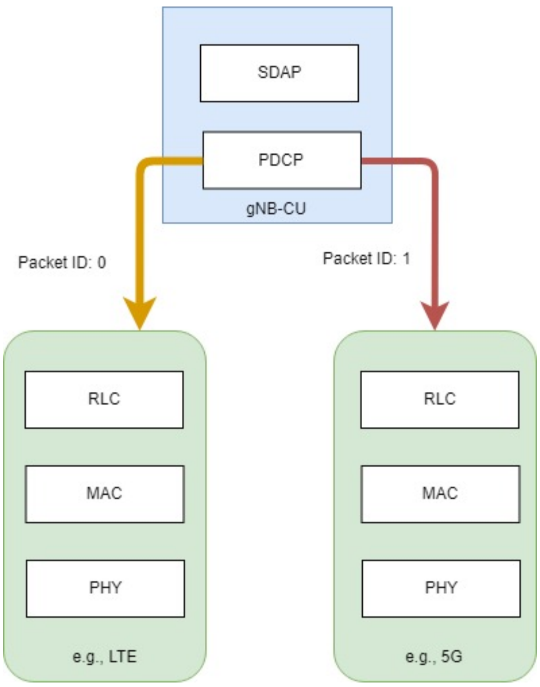

Figure 22 - MC, increase the data rate

With respect to point (b), an optimal distribution of the traffic loads, related to the multiple UEs that at the same time request for services, is enabled by the traffic separation considering the whole set of available radio access technologies in a given environment. This would be particularly suited for use cases with a simultaneous demand (e.g., people at live events, swarm of sensors, etc…).

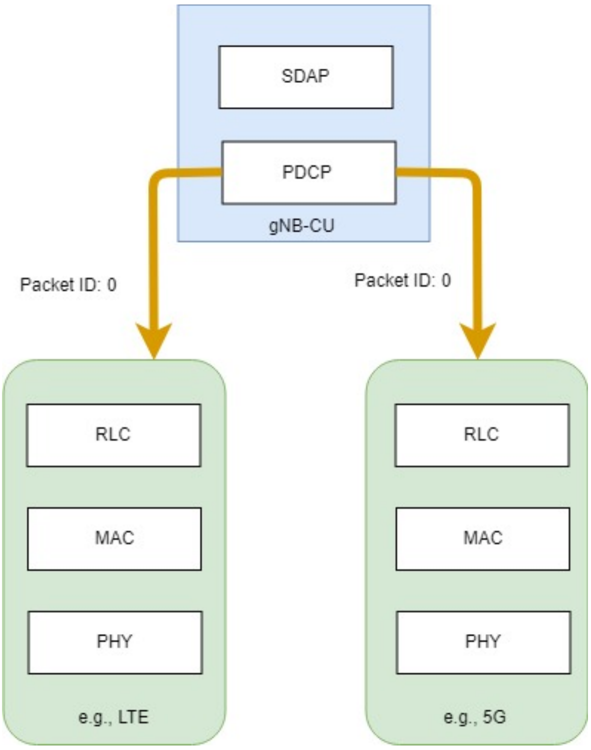

Figure 23 - MC, Service continuity

With respect to point (c), it is known that the mmWave suffers from packet loss to the intermittent channel quality and a potential solution for satisfying critical use cases is to assure the service continuity by sending the same data packets in two or more different radio access technologies (see Figure 23) to avoid packet loss.

The 5G-ALLSTAR project is aimed at implementing Multi-Connectivity solutions to satisfy both (i) data rate boosting and (ii) service continuity for reliability purposes with the implementation of innovative control methodologies leveraging the already available tendency in the cloud-based solutions for improving network performances. As already introduced in previous chapters, in 5G-ALLSTAR project the **Traffic Flow Control** module will be in charge of ensuring the 5G KPIs in terms of latency, data rate, reliability, etc, by combining different methodologies for delivering the suitable traffic steering, splitting and switching solutions in order to handle Multi-Connectivity mechanisms. The methodologies behind the Traffic Flow Control module will set up both SDAP layer and PDCP layer in the gNB-CU for enabling the Multi-Connectivity where the radio access points do not need to be divided in Master Node or Secondary node but assuming to have the Control Plane functionalities in the C-RAN and the split functions are at the PDCP level [7] (see §3.3). The Multi-Connectivity will be used by the Traffic Flow Control when needed/required taking into account the UE, network, QoS Requirements, and environmental conditions (e.g., interferences).

Traffic Flow Control Input / Output

The traffic flow control algorithm will be able to decide the dynamic association of the traffic of a given UE with one or more RATs. Indeed, these decisions will be based on the information about the traffic, the UE and the RATs conditions as shown in Figure 24.

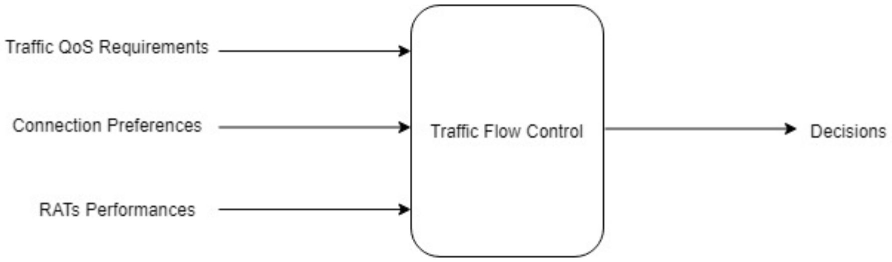

Figure 24: Traffic Flow Control Input / Output

The information that characterizes the traffic is identified by the service type. The service type contains information about the traffic requirements, i.e. the QoS Requirements. The QoS Requirements are standardized in the 3GPP with the 5QI. In [43] (see Appendix) the QoS parameters and the value for these parameters for different service types are defined. The traffic flow control algorithm on the basis of some QoS parameters such as:

- resources type (GBR, Delay critical GBR or Non-GBR);
- priority level;
- Packet Delay Budget;
- Packet Error Rate;

will be able to take into account the traffic requirements. The QoS model in the 5G network is defined in [43], the main steps are:

1. the SMF will decide for each traffic flow in a PDU session the appropriate QoS Profile,
2. SMF sent these decisions to the UPF, the UPF is are able to encapsulate the QFI (QoS Flow ID) in the headers of the traffic packets,
3. SMF sent the QoS profile to the RAN via the AMF.

The Traffic Flow Control with the QoS Profile provided by the AMF has information about the QoS Requirements of all the coming traffic flow. After all, the traffic flow control will be able to select the RATs that satisfy the QoS requirements, and the traffic flow will be recognized by the SDAP, using the QFI in the header, to carry out the decisions of the Traffic Flow Control.

The information about the UE is divided into standard UE information and personalized UE information. The standard UE information is the minimum needed to associate the UE to a set RATs, this information can be composed by the access points that are reachable by the UE, the type of RATs that the UE is able to use, and so on. The standard UE information will be provided to the traffic flow control by the AMF as specified in [37], where the AMF to AN interaction is specified with the Namf_Communication_UEContextTransfer AMF service to provide the UE context, a list of possible parameters provided by the service are:

- Mobile Equipment Identity;
- Access Type;
- Registration Area;
- Mobility Restrictions;
- User Location Information;
- Expected UE Behaviour Parameters;
- ….

The personalized UE information will be provided to the traffic flow control by the QoE/QoS Management (see §4), the provided information are composed of personalized parameters deduced on the basis of past connections performances and user feedback for each connection. In several works addressing the network selection issues (e.g., [17], [22]) the users' needs are expressed in terms of:

- Network Cost;
- Connection Quality;
- Battery Consumption;
- Mobility Needs;

These personalized parameters are used by the Traffic Flow Control in the decision process to take care of the personal needs of the users. Indeed, the selected RATs are able to satisfy the QoS requirements as well as the user needs to guarantee the best QoE during the connections.

The information about the RATs conditions is configured and collected by the RRM functionalities. The algorithm will be able to select the RATs considering the actual performances measured by the UEs. These measurements can be configured in Dual-Connectivity [36] to evaluate the performances of Master Node and Secondary Node. Furthermore, in [26] the procedure for measurement reporting in case of gNB split in gNB-CU and gNB-DU is presented. The measurement report is configured and managed by the RRC [47] and can be performed periodically or when an event occurs. The measurement can be performed considering different Channel State Information (CSI) indicators [48]:

- CSI reference signal received power (CSI-RSRP);
- CSI reference signal received quality (CSI-RSRQ);
- CSI signal-to-noise and interference ratio (CSI-SINR);
- ….

These measurements can be used by the traffic flow control to evaluate the RATs performances and to select them accordingly with the QoS and UE requirements described above.

5.1.1 Potential Approaches for Traffic Control in 5G-ALLSTAR

The starting point for the formulation of the network control algorithms that will be implemented in 5G-ALLSTAR is the identification of the most appropriate control frameworks, among the ones already explored in the literature and briefly summarised in Table 1.

For the identified inputs(?) and outputs, detailed in the previous section, a fundamental requirement that the solutions of the project should satisfy is that the proposed controller should be designed considering the network to be a dynamic system, whose evolution over time is driven by the amount, and type, of active connections.

A suitable modelling framework for the network is represented by the so-called *dynamical communication networks* [49], [50], which are network whose evolution is described by differential equations derived from the general mass conservation law. The fundamental reasoning behind such modelling framework is that any flow, be it a physical quantity or a stream of information, is conserved over the network links (or “edges”), and consequently its distribution follows simple physical-inspired laws.

Other than dynamical networks, another useful tool to capture the multi-connectivity problem characteristics (already partially introduced in the state-of-the-art section) is the concept of utility functions [30]. Utility functions are formally defined as monotonically nondecreasing functions that capture a certain performance, or quantity of interest, associated with the network state. Depending on their sign, utility functions can either be used to model performances or costs, allowing a good modelling flexibility. Common use cases are, for example, the modelling of communication channel delays (as a function of the channel-associated information flow), of connection unreliability (as a function of the number of connections), or of user satisfaction (as a function of the overall amount of network resources requested over a specific RAT).

The combination of dynamical networks with utility functions allows 5G-ALLSTAR to propose and investigate optimization algorithms for the dynamical control of traffic flows, as the utility functions may be utilised to capture several different objectives, while the differential representation of the network enables a realistic analysis of its time evolution, in terms of its congestion state and overall connection quality.

The multi-connectivity can then be seen as a multi-objective network control problem, characterized by several utility functions capturing its performances and costs, as well as physical and logical constraints, such as link capacities and user preferences.

Regarding constraints, their explicit handling by the proposed controllers will also be investigated, as, dealing with real networks that run on physical equipment, it is not reasonable to assume the number of network resources to be unlimited. *Capacitated* Networks have been extensively studied in the literature [51]–[53], and typical solution for controllers that utilise utility functions is to associate extremely negative values (i.e., high costs) to states that violate the physical/logical limitations of the controlled network. On the other hand, several control methodologies directly take into account constraints in their optimization, such as Model Predictive Control (MPC) [54] and Optimal control for Differential Algebraic Equations (DAE) Systems [55].

Other than the methodologies introduced so-far, which are related to traditional topics of Control Theory and Networked Systems, an interesting frontier that will be studied is one of the control solutions derived from the integration of Machine Learning / Artificial Intelligence solutions in standard control systems. In this regard, several works [56]–

[58] studied the joint effects of ML and Control Theory to enhance the system performance, as ML solutions are able to capture also aspects of the system that are hard to model, thanks to their ability to adapt, or learn, from the interaction with the controlled system. Several other works, on the other hand, propose controllers fully based on methodologies as Reinforcement Learning [16], [59]–[61], some of which were already applied by the authors of this deliverable in the field of communication networks and load balancing problems [62]–[65]. The last research direction that will be considered in the scope of project 5G-ALLSTAR for multi-connectivity solutions is related to game theory. Aspects coming from non-cooperative games (related to Nash/Wardrop equilibrium situations) and cooperative situations (related to operator/network optimality conditions) will be taken into account and considered in the design of the finalised control solutions. It interesting to remark that, under adequate hypotheses, game theoretic problems can be translated into an equivalent differential equation based formulation, in the field of deterministic evolutionary dynamics [66], [67], allowing once again the control solution to be based on strong, mathematically based, properties.

Overall, over the following months, Tasks T4.2 and T4.3 will investigate and develop algorithms based on:

- Standard control methods for dynamical network control,
- Reinforcement Learning based controllers,
- Game-theoretic controls and equilibria,

in compliance with the proposed architecture, and interfaces, of the previous sections.

From the KR-side perspective, the traffic flows in a RAN can be classified using the header of an incoming packet. An incoming packet consists of the header section and a data section. The header of a packet contains the information of MAC address, IP address, port number, protocol type and so on. Using the information of source IP address (SA), destination IP address (DA), source port number (SP), destination port number (DP) and protocol type information from a packet header flows can be classified. For each incoming packet, it is compared with a set of rules as shown in Figure 25 to check whether the packet is delay tolerant or not. If the packet lies in the rule set, then appropriate action is taken.
According to the traffic type, they can be offload to different RAT. Figure 25 shows a traffic Scheduling process considering the latency of traffic. In the process, if delay tolerant flow arrives in a master eNB, the eNB will offload the traffic to a secondary RAN, like Wi-Fi, satellite, etc., which is capable to handle delay tolerant traffic. However, the master eNB will serve the delay sensitive flow by itself.

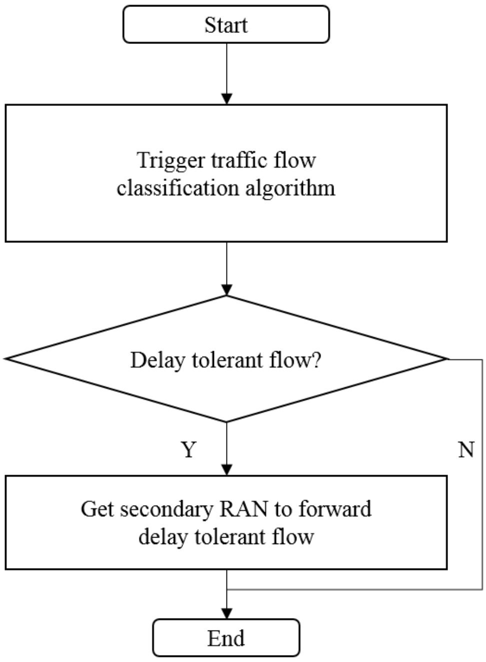

Figure 25: Traffic scheduling considering traffic latency

6 Multi Connectivity interfaces

This section introduces the preliminary design of interfaces for (i) the inner 5G-ALL-STAR components and (ii) the communication between the 5G-ALLSTAR components and the standard 5G functional modules.

The 5G-ALLSTAR services expose public APIs in order to be triggered by external and internal functions. Such APIs (WP4 APIs) will be implemented and detailed during the project to foster the integration among the services and components implemented in the 5G-ALLSTAR work packages. In the following tables, the *Interface/s Short Name* is re-ferred to the communication protocol among components as shown in Figure 26.

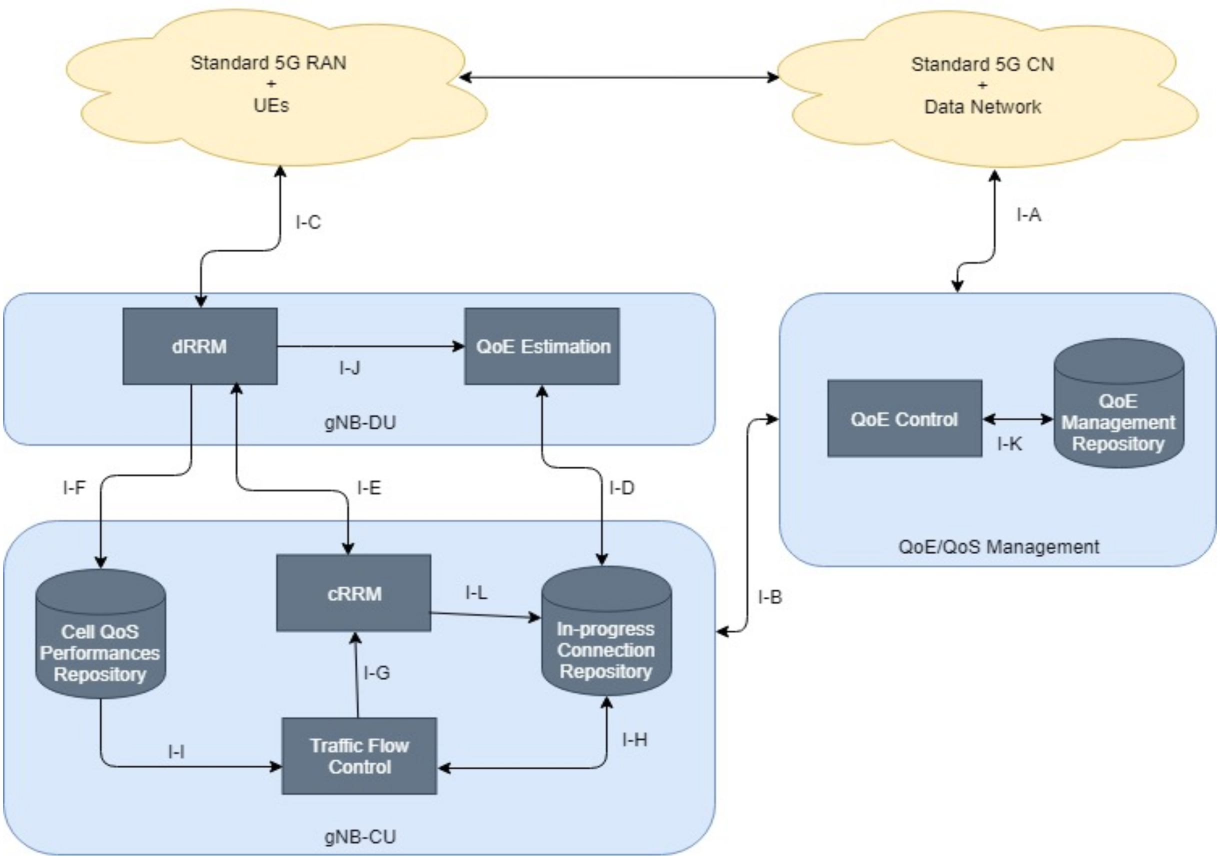

Figure 26: 5G-ALLSTAR Interfaces Diagram

Interface Short Name: **I-A**	
CN to QoE/QoS Management	{Source Id, Destination Id, Service Type, QoS Profile, UE type}, {Explicit Feedback}
QoE/QoS Management to CN	{Connection Acceptance}
Involved Module	CN functionalities (e.g. SMF) send the information about the established connections
	QoE/QoS Management stores such information

Interface Short Name: **I-B**

QoE/QoS Management to gNB-CU	{Source Id, Destination Id, Service Type, QoS Profile, UE type, Connection Preferences}
gNB-CU to QoE/QoS Management	{Cell-Id, Time Duration, Cell QoS Performance, Implicit QoE Feedbacks}
Involved Module	QoE/QoS Management sends information about the connections and user Connection Preferences.
	gNB-CU sends the information about the Connection QoE History

Interface Short Name: **I-C**	
dRRM to UEs/RAN	{Bearer Configuration, Measurement Configuration}
UEs/RAN to dRRM	{Measurement Report}
Involved Modules	dRRM considering the Connections-to-Cells Allocations performs the Radio Bearers Configuration and the Measurement Configuration
	UEs/RAN sends Measurement Reports accordingly to the Measurement Configurations

Interface Short Name: **I-D**	
QoE Estimation to In-Progress Connections Repository	{Estimated QoE Error}
In-Progress Connections Repository to QoE Estimation	{QoS Profile}
Involved Modules	QoE Estimation sends information about the actual QoE error
	In-Progress Connections Repository provides information about the target QoS, and stores the Estimated QoE error

Interface Short Name: **I-E**	
cRRM to dRRM	{Bearer Configuration}
dRRM to cRRM	{Measurement Reports}
Involved Modules	cRRM considering the Connections-to-Cells Allocations sends information about the Cells configurations (i.e. Data Radio Bearers Configuration)
	dRRM configures the Cells and periodically sends to the cRRM the CSI

Interface Short Name: **I-F**	
dRRM to Cells QoS Performances Repository	{QoS Cell Performances}
Cells QoS Performances Repository to dRRM	N.A.
Involved Modules	dRRM sends information about the serving Cells performances
	Cells QoS Performances Repository stores such information

Interface Short Name: **I-G**	
Traffic Flow Control to cRRM	{Connections-to-Cells Allocation}
cRRM to Traffic Flow Control	N.A.

Involved Modules	Traffic Flow Control sends information about the connection to cells allocations
	cRRM performs radio configurations accordingly to these allocations

Interface Short Name: **I-H**	
Traffic Flow Control to In-Progress Connection Repository	{Connection-to-Cells Allocation}
In-Progress Connection Repository to Traffic Flow Control	{Connection Preferences, Actual Serving Cells, Implicit Feedback, Potential Serving Cells}
Involved Modules	Traffic Flow Control sends information about the connection to cells allocations
	In-Progress Connection Repository sends information about the ongoing connections

Interface Short Name: **I-I**	
Cell QoS Performances Repository to Traffic Flow Control	{Cell id, 5QI, Cell QoS Performances}
Traffic Flow Control to Cell QoS Performances Repository	N.A.
Involved Modules	Cell QoS Performances Repository sends information about the cell status
	Traffic Flow Control uses the information to performs Connection-to-Cell allocation.

Interface Short Name: **I-J**	
dRRM to QoE Estimate	{UE Connection QoS Performances}
QoE Estimate to dRRM	N.A.
Involved Modules	dRRM sends information about the Cell-UE connection status
	QoE Estimation on the basis of the QoS Measurement computes the estimated actual QoE.

Interface Short Name: **I-K**	
QoE Control to QoE Management Repository	{Connection Preferences}
QoE Management Repository to QoE Control	{Source Id, Destination Id, Service Type, QoS Profile, UE type, Explicit Feedback, Cell-Id, Time Duration, Cell QoS Performance, Implicit QoE Feedbacks}
Involved Modules	QoE Control sends the Connection Preferences
	QoE Management Repository sends all the information about the connection and the Connection QoE History.

Interface Short Name: **I-L**	
cRRM to In-Progress Connections Repository	{*Potential Serving Cells_Id*}
In-Progress Connections Repository to cRRM	N.A.
Involved Modules	cRRM sends the Cell_Ids of all the Cells which, currently, potentially, can serve the in-progress Connection.
	In-Progress Connections Repository stores such information

7 Conclusions and Future work

In this document, we presented the work and the main results achieved so far in WP4 during the period M1-M10 from both EU and KR side. For as concern the activities carried out in Task 4.1, we have preliminarily identified the architectural requirements for developing AI-based algorithms that will be designed and developed in Task 4.2, and we have preliminarily identified the interfaces and the main relations with 5G-ALLSTAR layers, components. For as concern the activities carried out in Task 4.2, we have preliminarily reported the functionalities that will be implemented in 5G-ALLSTAR target system in terms of:

- Multi Connectivity: the Multi-Connectivity algorithms that will be implemented in WP4 will be used to guarantee 5G network reliability and throughput taking into account the different inputs (i.e., network status and performances, users' needs, etc..) from the other 5G-ALLSTAR components.
- QoE Management: in WP4 a set of Quality of Experience functionalities will be implemented, as detailed above, to handling and capture users' needs to definitively guarantee a satisfactory level of Quality of Experience.
- QoE Control: in WP4, AI-based Quality of Experience Control algorithms will be investigated to provide personalized parameters (also called in this document with *Connection Preferences*) in order to "drive" the Multi-Connectivity in a user-centric manner.

References

[1] S. Andreev *et al.*, “Intelligent access network selection in converged multi-radio heterogeneous networks,” *IEEE Wirel. Commun.*, 2014.

[2] S. E. El Ayoubi *et al.*, “Preliminary Views and Initial Considerations on 5G RAN Architecture and Functional Design,” *White Pap. METIS II*, 2016.

[3] P. Rost, V. Jungnicke, and 5G PPP Architecture Working Group, “View on 5G Architecture,” 2016.

[4] A. Prasad, F. S. Moya, M. Ericson, R. Fantini, and Ö. Bulakci, “Enabling RAN moderation and dynamic traffic steering in 5G,” in *IEEE Vehicular Technology Conference*, 2017.

[5] A. Checko *et al.*, “Cloud RAN for Mobile Networks - A Technology Overview,” *IEEE Commun. Surv. Tutorials*, 2015.

[6] A. Ravanshid *et al.*, “Multi-connectivity functional architectures in 5G,” in *2016 IEEE International Conference on Communications Workshops, ICC 2016*, 2016.

[7] D. S. Michalopoulos, I. Viering, and L. Du, “User-plane multi-connectivity aspects in 5G,” in *2016 23rd International Conference on Telecommunications, ICT 2016*, 2016.

[8] S. Chandrashekar, A. Maeder, C. Sartori, T. Höhne, B. Vejlgaard, and D. Chandramouli, “5G multi-RAT multi-connectivity architecture,” in *2016 IEEE International Conference on Communications Workshops, ICC 2016*, 2016.

[9] C. Pan, M. Elkashlan, J. Wang, J. Yuan, and L. Hanzo, “User-Centric C-RAN Architecture for Ultra-Dense 5G Networks: Challenges and Methodologies,” *IEEE Commun. Mag.*, 2018.

[10] Telefonica ; Ericsson, “Cloud RAN Architecture for 5G Cloud,” *White Pap.*, 2016.

[11] C. Kilinc *et al.*, “5G Multi-RAT Integration Evaluations Using a Common PDCP Layer,” in *IEEE Vehicular Technology Conference*, 2017.

[12] I. Da Silva *et al.*, “Tight integration of new 5G air interface and LTE to fulfill 5G requirements,” in *IEEE Vehicular Technology Conference*, 2015.

[13] S. Xu, M. Hou, Y. Fu, H. Bian, and C. Gao, “Improved Fast Centralized Retransmission Scheme for High-Layer Functional Split in 5G Network,” in *Journal of Physics: Conference Series*, 2018.

[14] A. Morgado, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, “A survey of 5G technologies: regulatory, standardization and industrial perspectives,” *Digit. Commun. Networks*, 2018.

[15] Z. S. Hui Leifang, Ma Wei, “A novel approach for radio resource management in multi-dimensional heterogeneous 5G networks,” *J. Commun. Inf. Networks*, vol. Vol.1, No., 2016.

[16] N. Vučević, J. Pérez-Romero, O. Sallent, and R. Agustí, “Reinforcement learning for joint radio resource management in LTE-UMTS scenarios,” *Comput. Networks*, 2011.

[17] L. Wang and G. S. G. S. Kuo, “Mathematical modeling for network selection in heterogeneous wireless networks - A tutorial,” *IEEE Commun. Surv. Tutorials*, 2013.

[18] J. O. Kim, “Feedback-based traffic splitting for wireless terminals with multi-radio devices,” *IEEE Trans. Consum. Electron.*, 2010.

[19] Q. Song and A. Jamalipour, “Network selection in an integrated wireless lan and UMTS environment using mathematical modeling and computing techniques,” *IEEE Wirel.*

Commun., 2005.

[20] M. Abolhasan, A. Al Sabbagh, and R. Braun, “A Comprehensive Survey on RAT Selection Algorithms for Heterogeneous Networks,” in *World Academy of Science, Engineering and Technology - WASET*, 2011.

[21] S. Singh, S. P. Yeh, N. Himayat, and S. Talwar, “Optimal traffic aggregation in multi-RAT heterogeneous wireless networks,” in *2016 IEEE International Conference on Communications Workshops, ICC 2016*, 2016.

[22] K. Piamrat, A. Ksentini, J.-M. Bonnin, and C. Viho, “Radio resource management in emerging heterogeneous wireless networks,” *Comput. Commun.*, 2011.

[23] and S. T. Sarabjot Singh, Mikhail Geraseminko, Shu-ping Yeh, Nageen Himayat, “Proportional Fair Traffic Splitting and Aggregation in Heterogeneous Wireless Networks,” *IEEE Communications Letters*. 2016.

[24] X. Gelabert, J. Pérez-Romero, O. Sallent, and R. Agustí, “A Markovian approach to Radio Access Technology selection in heterogeneous multiaccess/multiservice wireless networks,” *IEEE Trans. Mob. Comput.*, 2008.

[25] J. Pérez-Romero, O. Sallent, and R. Agusti, “Policy-based initial RAT selection algorithms in heterogeneous networks,” in *7th MWCN Conference, Marrakesh*, 2005.

[26] 3GPP TS 38.401, “3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NG-RAN; Architecture description (Release 15),” 2018.

[27] 3GPP TR 38.801, “3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on new radio access technology: Radio access architecture and interfaces (Release 14),” 2018.

[28] N. Zhang, S. Zhang, S. Wu, J. Ren, J. W. Mark, and X. Shen, “Beyond coexistence: Traffic steering in LTE networks with unlicensed bands,” *IEEE Wirel. Commun.*, 2016.

[29] M. Dryjanski and M. Szydelko, “A unified traffic steering framework for LTE radio access network coordination,” *IEEE Commun. Mag.*, 2016.

[30] Q. T. Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine, “On utility models for access network selection in wireless heterogeneous networks,” in *NOMS 2008 - IEEE/IFIP Network Operations and Management Symposium: Pervasive Management for Ubiquitous Networks and Services*, 2008.

[31] O. Ormond, J. Murphy, and G. M. Muntean, “Utility-based intelligent network selection in beyond 3G systems,” in *IEEE International Conference on Communications*, 2006.

[32] F. Bari and V. Leung, “Multi-attribute network selection by iterative TOPSIS for heterogeneous wireless access,” in *2007 4th Annual IEEE Consumer Communications and Networking Conference, CCNC 2007*, 2007.

[33] A. L. Wilson, A. Lenaghan, and R. Malyan, “Optimising Wireless Access Network Selection to Maintain QoS in Heterogeneous Wireless Environments,” in *WPMC 2005, 8th International Symposium on Wireless Personal Multimedia Communications*, 2005.

[34] M. Cesana, N. Gatti, and I. Malanchini, “Game Theoretic Analysis of Wireless Access Network Selection: Models, Inefficiency Bounds, and Algorithms,” in *Proceedings of the 3rd International Conference on Performance Evaluation Methodologies and Tools*, 2008.

[35] J. Antoniou and A. Pitsillides, “4G converged environment: Modeling network selection as a game,” in *2007 16th IST Mobile and Wireless Communications Summit*, 2007.

[36] 3GPP TR 37.340, “3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and NR; Multi-connectivity; Stage 2 (Release 15),” 2018.

[37] 3GPP TS 23.502, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Procedures for the 5G System; Stage 2 (Release 15)," 2018.

[38] T. T. T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," *IEEE Communications Surveys and Tutorials*. 2008.

[39] 3GPP TR 38.811, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on New Radio (NR) to support non terrestrial networks (Release 15)," 2018.

[40] H.-J. Kim, K.-H. Lee, and J. Zhang, "In-service Feedback QoE Framework," in *2010 Third International Conference on Communication Theory, Reliability, and Quality of Service*, 2010, pp. 135–138.

[41] L. Pierucci, "The quality of experience perspective toward 5G technology," *IEEE Wirel. Commun.*, vol. 22, no. 4, pp. 10–16, Aug. 2015.

[42] E. Verband der Elektrotechnik, Denshi Jōhō Tsūshin Gakkai (Japan), and Institute of Electrical and Electronics Engineers., *WTC 2014 ; World Telecommunications Congress 2014 ; Proceedings of : date, 1-3 June 2014.* .

[43] 3GPP TS 23.501, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; System Architecture for the 5G System; Stage 2 (Release 15)," 2018.

[44] E. Liotou *et al.*, "Shaping QoE in the 5G ecosystem," in *2015 7th International Workshop on Quality of Multimedia Experience, QoMEX 2015*, 2015.

[45] P. Reichl, B. Tuffin, and R. Schatz, "Logarithmic laws in service quality perception: Where microeconomics meets psychophysics and quality of experience," *Telecommun. Syst.*, 2013.

[46] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Netw.*, 2010.

[47] 3GPP TS 38.331, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; Radio Resource Control (RRC) protocol specification (Release 15)," 2018.

[48] 3GPP TS 38.215, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; Physical layer measurements (Release 15)," 2018.

[49] G. Como, K. Savla, D. A.-… on A. Control, and undefined 2013, "Robust distributed routing in dynamical networks—Part I: Locally responsive policies and weak resilience," *ieeexplore.ieee.org*.

[50] G. Como, K. Savla, D. A.-… on A. Control, and undefined 2013, "Robust distributed routing in dynamical networks–part II: Strong resilience, equilibrium selection and cascaded failures," *ieeexplore.ieee.org*.

[51] B. Gendron, T. G. Crainic, and A. Frangioni, "Multicommodity Capacitated Network Design," in *Telecommunications Network Planning*, Boston, MA: Springer US, 2011, pp. 1–19.

[52] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, "Selfish Routing in Capacitated Networks," *Math. Oper. Res.*, vol. 29, no. 4, pp. 961–976, Nov. 2004.

[53] P. Marcotte, S. Nguyen, and A. Schoeb, "A Strategic Flow Model of Traffic Assignment in Static Capacitated Networks," *Oper. Res.*, vol. 52, no. 2, pp. 191–212, Apr. 2005.

[54] J. Rawlings and D. Mayne, "Model predictive control: Theory and design," 2009.

[55] M. L. Bell and R. W. H. Sargent, “Optimal Control of Inequality Constrained DAE Systems, Part II: The Numerical Algorithm,” *Elsevier*, no. B98.21. 1998.

[56] C. Lin and C. S. G. Lee, “Neural-network-based fuzzy logic control and decision system,” *IEEE Trans. Comput.*, vol. 40, no. 12, pp. 1320–1336, 1991.

[57] T. Zhang, S. S. Ge, and C. C. Hang, “Adaptive neural network control for strict-feedback nonlinear systems using backstepping design,” *Automatica*, vol. 36, no. 12, pp. 1835–1846, 2000.

[58] F. W. Lewis, S. Jagannathan, and A. Yesildirak, *Neural Network Control Of Robot Manipulators And Non-Linear Systems*, vol. 30. 1998.

[59] R. Sutton, A. Barto, and R. Williams, “Reinforcement Learning is direct adaptive optimal control,” *ieeexplore.ieee.org*.

[60] J. Seiffertt and D. Wunsch, *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*. 2012.

[61] T. P. Lillicrap *et al.*, “Continuous control with deep reinforcement learning,” *arxiv.org*, 2015.

[62] A. Pietrabissa, F. D. Priscoli, A. Di Giorgio, A. Giuseppi, M. Panfili, and V. Suraci, “An approximate dynamic programming approach to resource management in multi-cloud scenarios,” *Int. J. Control*, vol. 90, no. 3, pp. 492–503, Mar. 2017.

[63] F. Lisi *et al.*, “Multi-agent quality of experience control,” *Int. J. Control. Autom. Syst.*, vol. 15, no. 2, pp. 892–904, Apr. 2017.

[64] J. F. Riera *et al.*, “TeNOR : Steps Towards an Orchestration Platform for Multi-PoP NFV Deployment,” *ieeexplore.ieee.org*, no. June, 2016.

[65] A. Pietrabissa *et al.*, “Resource management in multi-cloud scenarios via reinforcement learning,” in *Chinese Control Conference, CCC*, 2015, vol. 2015–Septe, pp. 9084–9089.

[66] W. H. Sandholm, *Population games and evolutionary dynamics*. .

[67] B. Wu, D. Zhou, and L. Wang, “Evolutionary dynamics on stochastic evolving networks for multiple-strategy games,” *Phys. Rev. E*, vol. 84, no. 4, p. 046111, Oct. 2011.

[5GUC] Qualcomm. "Making 5G NR a reality: leading the technology inventions for a unified, more capable 5G air interface." White paper (2016).

[SRICS] From Booz&Co, “Why satellites matter”, http://www.esoa.net/news-info-30.html

[QOSM] M. Gramaglia et al., "Flexible connectivity and QoE/QoS management for 5G Net-works: The 5G NORMA view," 2016 IEEE International Conference on Communications Workshops (ICC), Kuala Lumpur, 2016, pp. 373-379.

APPENDINX

Table 3: Standardized 5QI to QoS requirements mapping [37]

5QI Value	Resource Type	Default Priority Level	Packet Delay Budget	Packet Error Rate	Default Maximum Data Burst Volume	Default Averaging Window	Example Services
1	GBR	20	100 ms	10^{-2}	N/A	2000 ms	Conversational Voice
2		40	150 ms	10^{-3}	N/A	2000 ms	Conversational Video (Live Streaming)
3		30	50 ms	10^{-3}	N/A	2000 ms	Real Time Gaming, V2X messages Electricity distribution – medium voltage, Process automation - monitoring
4		50	300 ms	10^{-6}	N/A	2000 ms	Non-Conversational Video (Buffered Streaming)
65		7	75 ms	10^{-2}	N/A	2000 ms	Mission Critical user plane Push To Talk voice (e.g., MCPTT)
66		20	100 ms	10^{-2}	N/A	2000 ms	Non-Mission-Critical user plane Push To Talk voice
67		15	100 ms	10^{-3}	N/A	2000 ms	Mission Critical Video user plane
75		25	50 ms	10^{-2}	N/A	2000 ms	V2X messages
5	Non-GBR	10	100 ms	10^{-6}	N/A	N/A	IMS Signalling
6		60	300 ms	10^{-6}	N/A	N/A	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7		70	100 ms	10^{-3}	N/A	N/A	Voice, Video (Live Streaming) Interactive Gaming
8		80	300 ms	10^{-6}	N/A	N/A	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9		90					
69		5	60 ms	10^{-6}	N/A	N/A	Mission Critical delay sensitive signalling (e.g., MC-PTT signalling)
70		55	200 ms	10^{-6}	N/A	N/A	Mission Critical Data (e.g. example services are the same as QCI 6/8/9)
79		65	50 ms	10^{-2}	N/A	N/A	V2X messages
80		68	10 ms	10^{-6}	N/A	N/A	Low Latency eMBB applications Augmented Reality
81	Delay Critical GBR	11	5 ms	10^{-5}	160 B	2000 ms	Remote control (see TS 22.261)
82		12	10 ms	10^{-5}	320 B	2000 ms	Intelligent transport systems
83		13	20 ms	10^{-5}	640 B	2000 ms	Intelligent Transport Systems
84		19	10 ms	10^{-4}	255 B	2000 ms	Discrete Automation
85		22	10 ms	10^{-4}	1358 B	2000 ms	Discrete Automation