

# 5G ALLSTAR



Document Number: H2020-EUK-815323/5G-ALLSTAR/D4.2

Project Name:  
5G Agile and flexible integration of Satellite And cellular (5G-ALLSTAR)

## Deliverable D4.2

Design and simulation of the multi-RAT load balancing algorithms

Date of delivery: 31/12/2019  
Start date of Project: 01/07/2018

Version: 2.0  
Duration: 36 months

# Deliverable D4.2

## Design and simulation of the multi-RAT load balancing algorithms

<b>Project Number:</b>	H2020-EUK-815323
<b>Project Name:</b>	5G AgiLe and fLexible integration of SaTellite And cellular

<b>Document Number:</b>	H2020-EUK-815323/5G-ALLSTAR/D4.2
<b>Document Title:</b>	Design and simulation of the multi-RAT load balancing algorithms
<b>Editor(s):</b>	Alessandro Giuseppe (CRAT)
<b>Authors:</b>	Alessandro Giuseppe (CRAT), Antonio Pietrabissa (CRAT), Francesco Delli Priscoli (CRAT), Antonio Ornatelli (CRAT), Vincenzo Suraci (CRAT), Vito Andrea Racanelli (CRAT), Emanuele De Santis (CRAT), Alberto Isidori (CRAT), Alessandro Di Giorgio (CRAT), Martina Panfili, Federico Lisi (CRAT), Seok Ho Won (ETRI), Taesang Choi (ETRI), You Jun Choi (KATECH), Jung Ho Kim (SNETICT) Additional authors from University of Ulsan under sub-contract agreement with ETRI: Sungoh Kwon, Syed Maaz Shahid, Seyoum Yemane Teklay
<b>Dissemination Level:</b>	PU
<b>Contractual Date of Delivery:</b>	31/12/2020
<b>Security:</b>	Public
<b>Status:</b>	Final
<b>Version:</b>	2.0
<b>File Name:</b>	5G-ALLSTAR_D4.2_seconddisssue_clean.docx3

## Abstract

This document is the second issue of D4.2, which contains the improvements and answers to the comments received from the expert reviewers after the second project review.

The main changes are:

- The addition of a new section to describe at high level the multi-connectivity algorithms
- The report of the current state of development for the open source network simulator developed specifically for the project and the update of some existing simulation results on this new environment
- The addition of a more detailed discussion on the architectural implications that the different coverage of satellites and terrestrial networks have for their integration

**Abstract to D4.2:** This deliverable has been created as part of the work in the project Work Package 4 (WP4) and represents its second, and interim, output. According to the Grant Agreement, the purpose of the document is to present the suite of algorithms developed for multi-connectivity in the scope of the project, together with their preliminary testing conducted on the basis of appropriate numerical simulations. One or more of the proposed algorithms will be further developed and integrated in the project PoC during the rest of the WP4 activities, and this process will be detailed in D4.3. This deliverable collects also the updates regarding the architecture for multi-connectivity, and its software interfaces, developed in the scope of WP4.

## Keywords

*Multi-connectivity, traffic steering, network selection, network control algorithms*

## Acknowledgements

We would like to acknowledge the following people for the valuable reviews to the deliverable:

Leszek Raschkowski (FhG-HHI)

Seok Ho Won (ETRI)

Francesco Delli Priscoli (CRAT)





## Changes in the second issue

The second issue of D4.2, issued at M30, updates and improves the deliverable as suggested by the expert reviewers following their remarks after the second project review.

The document has been updated as follows:

- To answer the concerns of the experts regarding how the multi-connectivity algorithms can deal with the different coverages offered by satellite connections and terrestrial C-RANs, we added section 2.1 to better explain how our proposed resource slicing for the satellite resources could be applied in practice.
- We revised section 4 to provide a clearer overview of the algorithms' rationale and their role in the scope of the project, moving their mathematical formulation to the new section 5. We now also report the current release of the network simulator developed for the testing of the algorithms and have updated some simulation results in section 5.
- We updated the simulations of the algorithm of section 5.1 (Wardrop) with the results obtained on the new simulator that now include a more realistic setting, including interferences and network resource usage models.

## Executive Summary

The original submission of this deliverable, issued at M18, reports the activities completed within the scope of T4.2 "Design and simulation of the multi-RAT load balancing algorithms" and T4.1 "Multi-connectivity mapping onto the 5G network architecture" and represents the interim document for the whole WP4 "Multi-Connectivity".

The main objective of the deliverable is to detail the designed algorithms for the traffic flow control problem, which is the fundamental enabler of multi-connectivity. A complete mathematical formulation of such algorithms is reported, along with the results of their preliminary testing activities, conducted on the basis of numerical simulations.

At this stage, M18, the deliverable contains the final remarks regarding the multi-connectivity architecture, including a preliminary version of the specifications of the software interfaces that will be implemented in the project Proof of Concept (PoC).

The 5G-ALLSTAR consortium published in or submitted for publication to international conferences and journals part of the results and discussions of this deliverable.

## Contents

1	Introduction .....	1
1.1	Scope of the document .....	1
1.2	Relation with other work packages .....	3
2	Advancements in 5G-ALLSTAR multi-connectivity .....	4
2.1	Considerations regarding the coverage difference between satellite and ground RATs 6	
2.2	Multi-connectivity advancements offered by satellite integration.....	7
3	Multi-connectivity scenarios.....	9
3.1	Multi-connectivity simulation approach .....	9
3.2	Management and orchestration for terrestrial and satellite resources.....	12
4	Proposed algorithms for multi-connectivity and related software .....	13
4.1	Open source Network Simulator.....	13
4.1.1	5G NR and 4G LTE resource allocation description .....	14
4.1.2	Satellite resource allocation description .....	15
4.2	Overview on the 5G-ALLSTAR suite of algorithms for multi connectivity .....	16
4.2.1	Traffic steering and network selection in 5G networks based on reinforcement learning 16	
4.2.2	Wardrop Equilibrium based control .....	16
4.2.3	Friend or Foe Q-Learning.....	17
4.2.4	Maximal Load Balancing .....	17
4.2.5	Reinforcement learning resource optimization for mobile relays.....	18
4.2.6	Analytic Hierarchy Process) and Cooperative Differential Games for Multiconnectivity .....	18
4.2.7	ZeroSum games and Minimax for Multi-Connectivity.....	19
4.3	GAP analysis and comparison .....	20
4.4	Plan for the European multi-connectivity demonstrator.....	21
5	Mathematical formulation and results for the proposed algorithms .....	22
5.1	Traffic steering and network selection in 5G networks based on reinforcement learning.....	23
5.1.1	State of the art, innovations and limitations of the proposed approach.....	23
5.1.2	Preliminaries on Markov decision processes and reinforcement learning.....	24
5.1.3	Problem modelling .....	26
5.1.4	Proposed RL-based traffic steering algorithm.....	29
5.1.5	Simulations of the RL traffic steering controller .....	31
5.2	Capacity-constrained Wardrop equilibria for multi-connectivity in 5G networks .....	34
5.2.1	Multi-connectivity and traffic steering in 5G networks .....	35
5.2.2	Adversarial load balancing in 5G networks and Beckmann equilibria .....	36
5.2.3	Proposed Wardrop load balancing algorithm.....	37
5.2.4	5G traffic steering as a dynamic load-balancing problem .....	37
5.2.5	Preliminaries on Wardrop and Beckmann equilibria .....	38
5.2.6	Preliminaries on Lyapunov stability .....	39
5.2.7	Capacitated load balancing algorithm and convergence proof .....	40
5.2.8	Numerical simulation.....	46
5.2.9	Simulation setup .....	46
5.2.10	Simulation results.....	48
5.3	Network selection based on Markov games and friend-or-foe RL.....	51
5.3.1	Markov games and Nash equilibria .....	51
5.3.2	Multi-agent reinforcement learning .....	52
5.3.3	Modelling network selection as a Markov game .....	52
5.3.4	State space.....	53
5.3.5	Action space .....	53
5.3.6	Reward functions .....	53
5.3.7	$\epsilon$ -greedy policy selection.....	53

5.3.8	Maximin linear programming formulation.....	54
5.3.9	Simulations .....	55
5.3.10	Simulation one – baseline least loaded controller.....	56
5.3.11	Simulation two – no service prioritisation.....	56
5.3.12	Simulation three – different rewards .....	56
5.3.13	Simulation results.....	56
5.4	Maximal load balancing algorithm and preliminary simulation results .....	58
5.5	Reinforcement learning based resource optimization for mobile relays .....	60
5.6	AHP (Analytic Hierarchy Process) and Cooperative Differential Games for Multiconnectivity .....	63
5.6.1	AHP (Analytic Hierarchy Process).....	63
5.6.2	Cooperative Differential Game .....	65
5.6.3	Multi-Connectivity model for AHP and Cooperative Differential Game.....	66
5.7	ZeroSum Games and Minimax for Multi-Connectivity .....	69
5.7.1	Network Dynamics and Utility Function for Multi-Connectivity Modelling .....	71
6	Framework for QoE-aware Multi-Connectivity .....	73
6.1	QoE role and functional architecture in 5G-ALLSTAR .....	73
6.2	Questionnaire.....	75
6.2.1	Answers report to the first section of the questionnaire .....	76
6.2.2	Answers report to the second section of the questionnaire.....	83
6.2.3	Answers report to the third section of the questionnaire .....	84
6.2.4	Answers report to the fourth section of the questionnaire .....	87
6.2.5	Preliminary analysis of the questionnaire answers .....	89
7	Details of the multi-connectivity interfaces.....	92
7.1	UML flow chart of the multi-connectivity interfaces .....	93
8	Conclusions and future works .....	96
9	References.....	97
ANNEX	Preliminary JSON-like multi-connectivity interface description .....	102



## List of Figures

Figure 1-1: Work package 4 plan.....	1
Figure 2-1: Multi-Connectivity target physical architecture.....	4
Figure 2-2: Multi-connectivity protocol split options.....	5
Figure 2-3: Satellite 5G use cases [56].....	8
Figure 3-1: Reference architecture for the management of a satellite NR-RAT .....	12
Figure 3-2: Reference architecture for the management of a non-3GPP satellite RAN.....	12
Figure 4-1 Network Simulator GUI. Scenario with 3 ground AP and a Satellite. Grey dots: UEs. .....	13
Figure 4-2 Openbach GUI for the video server .....	21
Figure 5-1: User-perceived quality of connection for service with elastic traffic.....	26
Figure 5-2: User-perceived quality of connection for service with transmission bitrate threshold .....	27
Figure 5-3: User-perceived quality of connection for service with multi-codec traffic.....	27
Figure 5-4: Reference scenario used in the simulations .....	31
Figure 5-5: Percentage of blocked connections for the RL and the LL controllers .....	33
Figure 5-6: Dynamic Traffic Steering framework from [18].....	35
Figure 5-7 Network scenario .....	46
Figure 5-8 Maximum latency mismatch during the simulation (dotted line: tolerance $\epsilon$ ). .....	47
Figure 5-9 Network state in terms of total bitrate allocated on the various APs (solid lines: unconstrained providers, dashed line: constrained provider). .....	48
Figure 5-10: Commodity latency examples during the simulation (solid lines: unconstrained providers used by the commodity; dashed line: constrained providers; zoomed sub-plots to show the convergence within the tolerance $\epsilon = 0.5$ ). .....	48
Figure 5-11 Comparison for the three considered algorithms of the latencies over APs for commodity $i = 20$ . .....	49
Figure 5-12: Connection area covered by 3 different RATs .....	55
Figure 5-13: Number of allocated connections for the three controllers, divided by service class .....	56
Figure 5-14: Number of blocked connection for the three controllers, divided by service class .....	57
Figure 5-15: Total amount of bitrate of the blocked connections.....	57
Figure 5-16: Flow diagram of proposed MLB algorithm .....	58
Figure 5-17: Simulation environment for MLB algorithms .....	59
Figure 5-18: Simulation results with throughput enhancements.....	59
Figure 5-19: The analysis scenario for the proposed algorithm: the radio resource allocation policy with RL based interference control .....	61
Figure 5-20: AHP Hierarchy .....	63
Figure 5-21: Saaty's Scale for Pairwise comparison.....	64
Figure 5-22: Random Consistency Index for various n .....	65

Figure 5-23 Attribute example .....	66
Figure 5-24 Requirements examples.....	66
Figure 5-25 Block diagram of the AHP algorithm.....	68
Figure 5-26 Block diagram of the minimax algorithm .....	72
Figure 6-1: Personalization System from D4.1 .....	73
Figure 6-2: QoE management system functional architecture .....	74
Figure 6-3: 5G-ALLSTAR questionnaire landing page.....	75
Figure 6-4: Nationality Distribution of the Questionnaire .....	76
Figure 6-5: Age Distribution of the Questionnaire .....	76
Figure 6-6: Answers to question 1 pie chart .....	77
Figure 6-7: Answers to question 2 pie chart .....	77
Figure 6-8: Answers to question 3 pie chart .....	77
Figure 6-9: Answers to question 4 pie chart .....	78
Figure 6-10: Answers to question 5 pie chart.....	79
Figure 6-11: Answers to question 6 pie chart.....	79
Figure 6-12: Answers to questions 7 and 8 pie charts .....	80
Figure 6-13: Answers to questions 9, 10 and 11 pie charts .....	81
Figure 6-14: Answers to questions 12 and 13 pie charts .....	82
Figure 6-15: Screenshot of the first video shown during the questionnaire .....	83
Figure 6-16: Bar chart of question 1, second section.....	83
Figure 6-17: Bar chart of question 2, second section.....	84
Figure 6-18: Pie chart of question 3, second section .....	84
Figure 6-19: Screenshot of the second video shown during the questionnaire .....	85
Figure 6-20: Pie chart of question 1, third section.....	85
Figure 6-21: Bar chart of question 2, third section .....	85
Figure 6-22: Pie chart of question 3, third section.....	86
Figure 6-23: Bar chart of question 1, fourth section .....	87
Figure 6-24: Bar chart of question 2, fourth section .....	87
Figure 6-25: Pie chart of question 3, fourth section.....	88
Figure 6-26: Principal component analysis, age .....	89
Figure 6-27: Principal Component analysis, nationality .....	90
Figure 6-28: TSNE algorithm results, age labelling.....	90
Figure 6-29: TSNE algorithm results, nationality labelling.....	91
Figure 7-1: Preliminary architecture of the project EU-PoC .....	92
Figure 7-2: Interface IA flow chart.....	93
Figure 7-3: Interface IB flow chart.....	93
Figure 7-4: Interface ID flow chart .....	93

Figure 7-5: Interface IG flow chart .....	94
Figure 7-6: Interface IH flow chart .....	94
Figure 7-7: Interface II flow chart .....	94
Figure 7-8: Interface IJ flow chart .....	95
Figure 7-9: Interface IK flow chart.....	95

## List of Tables

Table 3-1: Simulation parameters.....	11
Table 4-1: GAP analysis of the proposed algorithms for traffic flow control .....	20
Table 5-1: Bitrate allocation algorithm for AP $p \in P$ in state $s \in S$ .....	32
Table 5-2 Characteristics of micro and macro cells .....	46
Table 5-3: Algorithm pseudo-code.....	55

## List of Abbreviations

<b>2D</b>	2 Dimensions / 2-Dimensional
<b>3D</b>	3 Dimensions / 3-Dimensional
<b>3GPP</b>	3 <sup>rd</sup> Generation Partnership Project
<b>5GPPP</b>	5G Public-Private Partnership
<b>5QI</b>	5G QoS Identifier
<b>ADP</b>	Approximated Dynamic Programming
<b>AHP</b>	Analytic hierarchy process
<b>AI</b>	Artificial Intelligence
<b>AIV</b>	Air Interface Variants
<b>AMF</b>	Access and Mobility Management Function
<b>AP</b>	Access Point
<b>BS</b>	Base Station
<b>CN</b>	Core Network
<b>CP</b>	Control Plane
<b>C-RAN</b>	Central Radio Access Network
<b>cRRM</b>	Central Radio Resource Management
<b>CSI</b>	Channel State Information
<b>CSI-RSRP</b>	CSI Reference Signal Received Power
<b>CSI-RSRQ</b>	CSI Reference Signal Received Quality
<b>CSI-SINR</b>	CSI Signal-to-Noise and Interference Ratio
<b>CU</b>	Centralized Unit
<b>DL</b>	Downlink
<b>D2C</b>	Device to cellular
<b>D2D</b>	Device to Device
<b>D-RAN</b>	Distributed Radio Access Network
<b>DRB</b>	Data Radio Bearer
<b>dRRM</b>	Distributed Radio Resource Management
<b>DP</b>	Dynamic Programming
<b>DU</b>	Distributed Unit
<b>FLC</b>	Fuzzy Logic Controller

<b>FS</b>	Fast Switch
<b>GEO</b>	Geostationary Earth Orbit
<b>gNB-CU</b>	next Generation Node B Central Unit
<b>gNB-DU</b>	next Generation Node B Distributed Unit
<b>GRA</b>	Grey Relational Analysis
<b>HH</b>	Hard Handover
<b>KPI</b>	Key Performance Indicator
<b>LEO</b>	Low Earth Orbit
<b>LTE</b>	Long Term Evolution
<b>MAC</b>	Medium Access Control
<b>MADM</b>	Multiple attribute decision making
<b>MCG</b>	Master Cell Group
<b>MDP</b>	Markov Decision Process
<b>MN</b>	Master Node
<b>NE</b>	Nash Equilibrium
<b>NG-RAN</b>	New Generation RAN
<b>NR</b>	New Radio
<b>NTN</b>	Non-Terrestrial Network
<b>PDCP</b>	Packet Data Convergence Protocol
<b>PDR</b>	Packet Detection Rule
<b>PDU</b>	Protocol Data Unit
<b>PHY</b>	Physical / physical layer
<b>PoC</b>	Proof of Concept
<b>PRB</b>	Physical Resource Block
<b>QFI</b>	QoS Flow Identifier
<b>QoE</b>	Quality of Experience
<b>QoS</b>	Quality of Service
<b>RAT</b>	Radio Access Technology
<b>RL</b>	Reinforcement Learning
<b>RLC</b>	Radio Link Control
<b>RRC</b>	Radio Resource Control
<b>RRM</b>	Radio Resource Management
<b>RRUR</b>	Radio Resource Usage Ratio

<b>RSRP</b>	Reference Signals Received Power
<b>RTT</b>	Round Trip Time
<b>SCG</b>	Secondary Cell Group
<b>SDAP</b>	Service Data Adaptation Protocol
<b>SMF</b>	Session Management Function

<b>SN</b>	Secondary Node
<b>SNR</b>	Signal to Noise Ratio
<b>UE</b>	User Equipment
<b>UL</b>	Uplink
<b>UP</b>	User Plane
<b>UPF</b>	User Plane Function
<b>WLAN</b>	Wireless Local Area Network

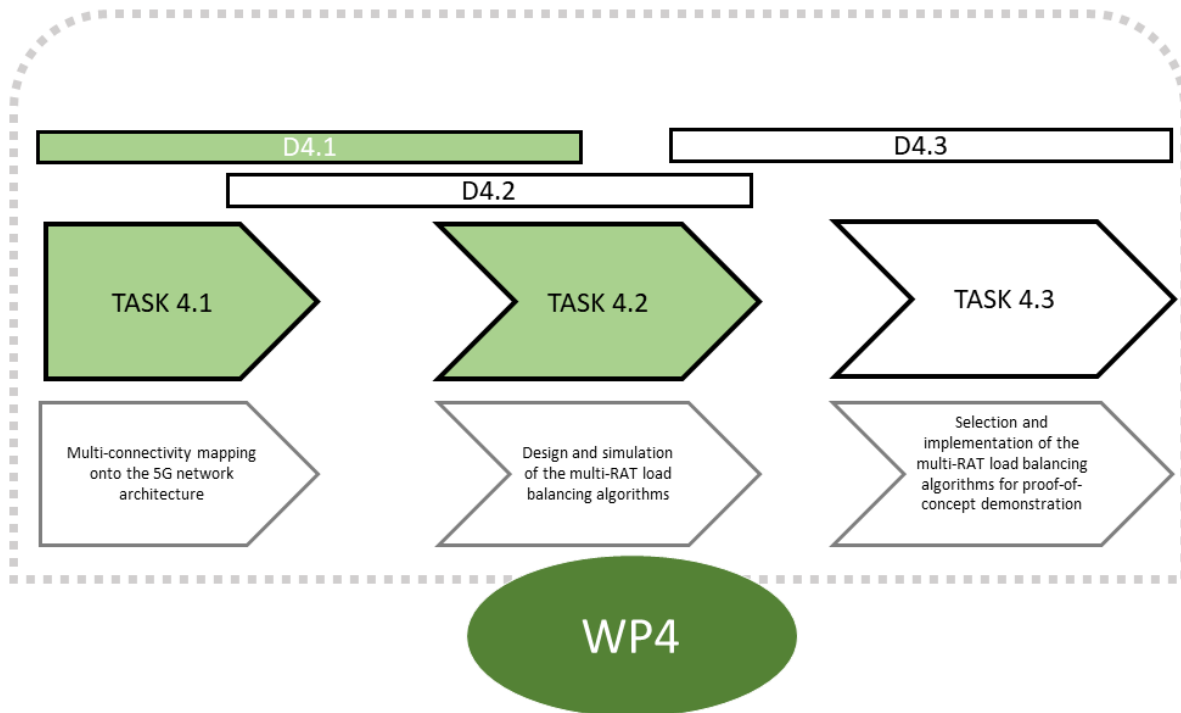
# 1 Introduction

This deliverable provides an overview of the activities conducted in both T4.1 and T4.2, detailing the traffic flow control algorithms and the framework for the Quality of Experience (QoE) management proposed by the 5G-ALLSTAR project.

## 1.1 Scope of the document

This document is organized to report the completion and refinements of the final activities performed in T4.1 “Multi-connectivity mapping onto the 5G network architecture”, as well as the first output from T4.2 “Design and simulation of the multi-RAT load balancing algorithms”. The logical relation among the deliverables of WP4 and their scopes is depicted in Figure 1-1, in compliance with the grant agreement.

We remark that WP4 is divided into three main tasks, whose results are collected in three deliverables.



**Figure 1-1: Work package 4 plan**

This document contains mainly results from Task 4.1 and Task 4.2 and is organized as follows:

- Chapter 1 serves as the introduction of the document;
- Chapter 2 reports and updates the architecture for multi-connectivity;
- Chapter 3 discusses the scenario identification process for the simulations of the developed algorithms;
- Chapter 4 details the proposed algorithms for traffic flow control;
- Chapter 5 describes the framework for QoE management in the project and reports the results of the questionnaire;
- Chapter 6 provides a first detailed analysis of the interfaces related to the activities of WP4 that will be deployed in the PoC;
- Chapter 7 draws the conclusions of the document.
- Chapter 8 contains the references of the document.
- The ANNEX includes a JSON-like description of the interfaces for multi-connectivity.

## 1.2 Relation with other work packages

WP4 is one of the three main technical work packages of the 5G-ALLSTAR project, and it concerns the design and development of a set of advanced functionalities to enable multi-connectivity in heterogeneous networks and QoE-aware traffic flow control.

In fact, with respect to the architecture designed in 5G-ALLSTAR deliverable D2.2, the WP4 components will affect 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:

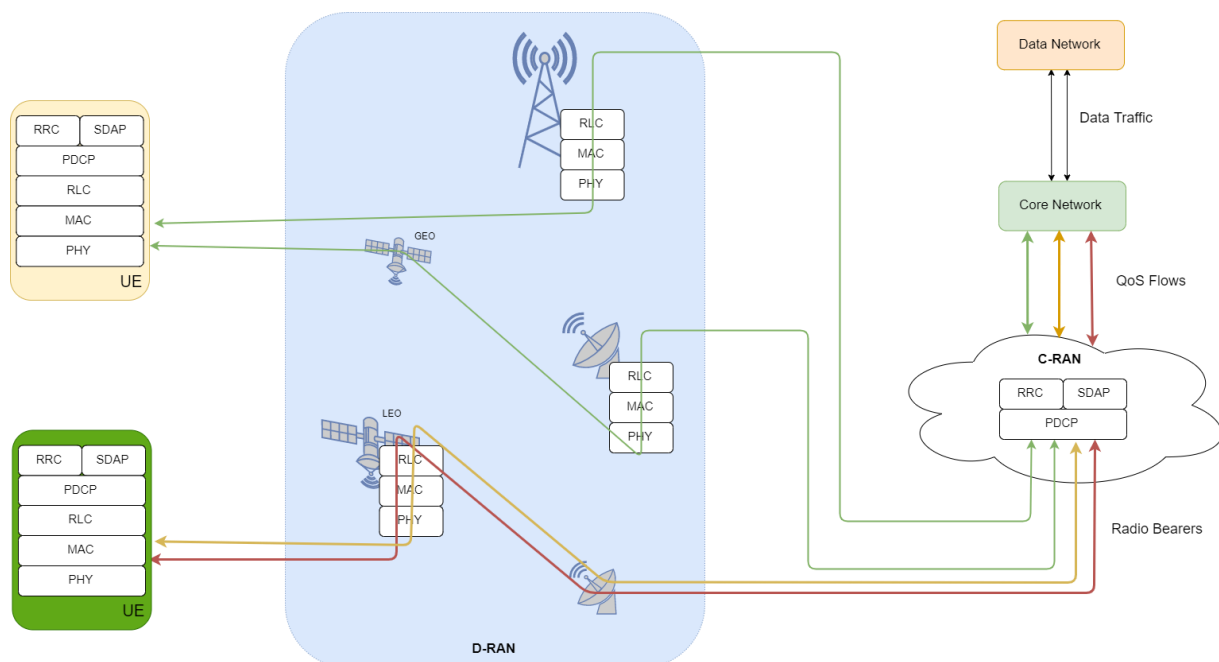
- WP2 provided important inputs for WP4 during the design of requirements, Key Performance Indicators (KPIs) and the high-level architecture. Such inputs have been considered in WP4 as references for the mapping between WP4 functionalities and the 5G-ALLSTAR architecture in the activities of T4.1, which laid the foundations for the development of the traffic flow control algorithms in T4.2.
- WP3 and WP4 activities interacted for the identification and the design of the interfaces that will foster data exchange between WP4 and WP3 components in the PoC of the project.
- WP5 represents the main user of the output of the activities of T4.2 and T4.3, as one (or more) of the algorithms presented in this deliverable will be expanded and refined so that it will be possible to demonstrate the capabilities and functionalities developed in the scope of WP4, related to both multi-connectivity and QoE management in the PoC.
- WP6 is in charge of disseminating the main results and achievements of the whole 5G-ALLSTAR project. Regarding WP4, the main dissemination will come from scientific publications detailing the traffic flow control algorithms and from graduate/general public seminars that the partners involved in the work package will organise.



## 2 Advancements in 5G-ALLSTAR multi-connectivity

This section presents a brief description of the multi-connectivity aspects that are studied in the 5G-ALLSTAR project, providing an introduction to the traffic flow control algorithms of the following sections and highlighting the rationale behind their development. The overall system architecture and its integration with multi connectivity will be discussed and detailed in the deliverable D2.3.

Figure 2-1 reports the architecture considered in the project regarding multi-connectivity. In the depicted scenario, the two pieces of User Equipment (UEs) are able to connect to several Access Points (APs) of different Radio Access Technologies (RATs) at the same time, even for the same PDU session.



**Figure 2-1: Multi-Connectivity target physical architecture**

This solution for service provision takes the name of multi-connectivity, a framework in which the traffic flows can be routed, or *steered*, over a set of heterogeneous RATs and APs to meet a variety of QoS requirements that range from connection resiliency to throughput and power saving.

Regarding multi-connectivity, three main processes for traffic flow control were identified by 3GPP for inclusion in its Release 16, namely:

- Traffic steering, which consists in the problem of optimal network selection for the connection provision;
- Traffic switching, which deals with the seamless handover between two different RATs as a dynamic response to a new network state (e.g., resource scarcity, service outage, moving UE, ...);
- Traffic splitting, which is the capability of duplicating a QoS flow over two or more different RATs to improve the resiliency of its information delivery by means of the so-called process of “Network Aggregation”.

In the scope of 5G-ALLSTAR, these processes were referred to as “traffic flow control strategies or decisions”. Traffic steering was also given the dynamical capabilities of traffic switching, so

that the assignment of (a portion of) a QoS flow may change over time and hence becomes a real-time control action that impacts the network state. Traffic splitting can also be integrated into the dynamic traffic steering process, provided that the traffic flow control accounts for the different QoS requirements of the various PDU sessions.

In this regard, the 5G-ALLSTAR project envisages also the integration of a so-called “personalisation system” able to associate to the various PDU sessions and QoS flows, a set of unstandardized user-dependant *connection preferences* that may drive the behaviour of the traffic flow control in a user-centric, or *personalised*, way.

Several control algorithms were developed in the first half of the project lifetime, to explore possible research directions and investigate the challenges associated with the multi-connectivity framework. The main objective of this deliverable is to report this investigation by detailing the designed algorithms and providing a first simulation-based analysis of their performance and differences in compliance with the grant agreement.

It is worth mentioning that the architecture of Figure 2-1 does not explicitly detail how the protocol stack was split to enable multi-connectivity. In fact, three main potential options have been identified, as reported in Figure 2-2. In the first two options, the intra-PHY and PHY splits, all the layers above the physical layer are in the Central Unit (CU), allowing an ideal scenario for a centralised control solution. Nevertheless, the shortcoming of such an option is the presence of very rigorous real-time and technology-dependent constraints for the correct functioning of the protocol stack.

Regarding the PDCP, the C-RAN is provided with a common PDCP for both the user plane and the control plane and a dedicated RRC, whereas the D-RANs are provided with PHY, MAC, and RLC layers. This configuration is advantageous since PDCP and RRC are not subject to the same constraints required for the first two solutions, allowing the usage for multi-connectivity of LEO satellites.

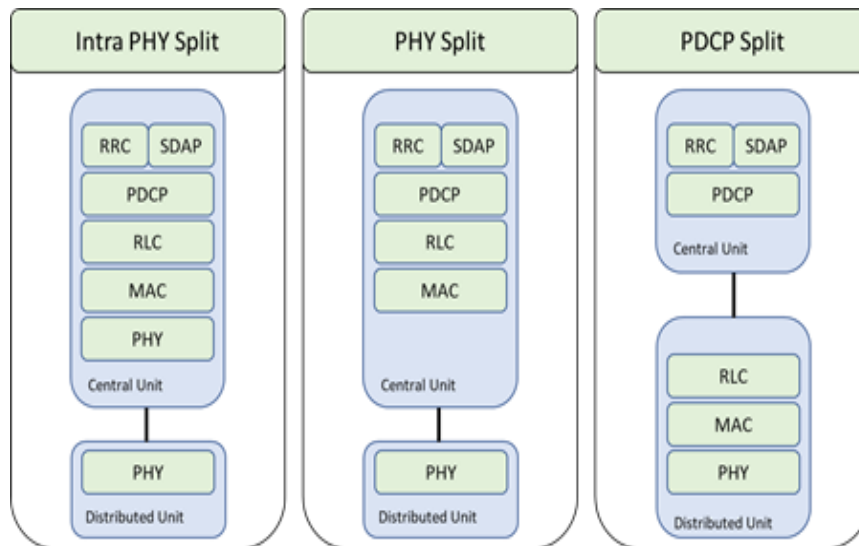


Figure 2-2: Multi-connectivity protocol split options

## 2.1 Considerations regarding the coverage difference between satellite and ground RATs

Regarding the inclusion of GEO satellites in the multi-connectivity framework of the project, we highlighted in Figure 2-1 how they act as a transparent mean of communication, as their Distributed Unit (DU) protocol stack is deployed on a dedicated ground equipment. In general, the inclusion of satellites as potential RAT for multi-connectivity poses several other challenges ranging from spectrum sharing (addressed in the scope of WP3) to optimal radio resource usage.

Satellites cover extremely wide areas, usually beyond the reasonable range of a single ground C-RAN. Thus, limiting the connectivity scope of the satellite to the area that is managed by the controller deployed in a single C-RAN would be extremely inefficient and would prevent most of the advantages that satellite connection can offer to 5G systems, e.g., their wide area availability.

Ideally, one may think to have several different C-RANs control the resources available on a single satellite, but designing a solution of this type would not be trivial and would require some awareness, as the multiple controllers may take conflicting decisions (e.g., cumulatively allocating more resources than available).

In the scope of this project, two different solutions to this issue have been identified, namely:

- The slicing of the satellite resources, so that each C-RAN that has access to the satellite manages only a portion of its resources so that no conflict on their usage may arise;
- The deployment of a wide-area C-RAN in the control plane, that supervises the functioning of multiple C-RANs of the user plane, that take their traffic flow control decisions without requiring to directly interact with the QoS flows.

Both solutions come with some associated challenges.

From a SDN point of view, by slicing its resources the satellite is virtually divided into several different satellites, each of which (virtually) covering an area comparable to the one governed by a C-RAN controller. A UE that asks for a connection to be established with one of such virtual satellites actually sends its request to the physical satellite itself, which takes the decision to accept or block the request depending on the multi-connectivity algorithms that the network controllers employ. However, according to the current protocol standards, the satellite does not receive any information regarding the position of the UE or which of the C-RAN controllers is responsible for the connection allocation. Thus, the satellite controller cannot determine which resource slice the new connection should utilise for its allocation, making the envisaged slicing impossible to be employed in practice.

For the correct integration of the slicing procedure, it is then needed either to include in the connection protocols the information regarding the C-RAN where the UE is located, or to force the UE to negotiate with the C-RAN the access to the satellite AP utilising the ground RATs. In this second solution, once the C-RAN controller enables the allocation of a certain amount of resources from its reserved slice for the connection, the UE receives connectivity from the satellite as decided by the network controllers and their multi-connectivity algorithms. In the case no terrestrial AP is available to the UE, the satellite may independently utilise an ad-hoc slice of its resources, that can be reserved for such areas, as it would be the only mean of communication and no multi-connectivity controller would be needed to decide the traffic allocation.

On the contrary, the deployment of a wide area C-RAN may be unfeasible due to the fact that the satellite coverage may span over multiple countries and different network infrastructures.

The algorithms developed in 5G-ALLSTAR assume that one of the two solutions is employed to allow the control of the network resources of the satellite as if a portion of them was dedicated to the C-RAN in which they supervise the multi-connectivity functionalities. In fact, the multi-connectivity algorithms were designed either to be distributed in the gNB-DU-CP associated to

the various APs (including the virtual APs defined following a resource slicing) or to be deployed on a centralised controller in the gNB-CU-CP that governs the resources of the entire C-RAN, including its share of resources provided by the satellite AP.

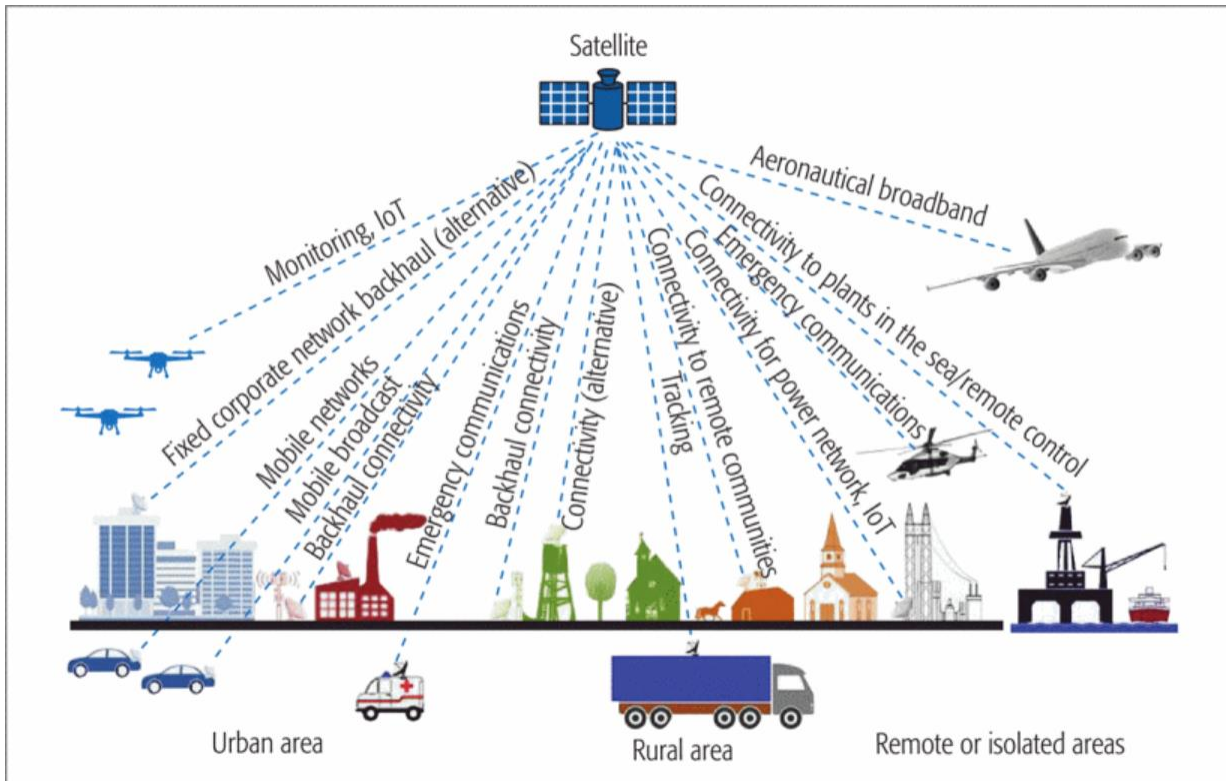
More detail on which logical entities the various algorithms should be deployed on, and on which network resources they manage, are reported in section 4.2.

## 2.2 Multi-connectivity advancements offered by satellite integration

5G-ALLSTAR multi-connectivity technology advancements are significant since satellite communications are expected to play a vital role in 5G and beyond networks in realizing ubiquitous coverage, greater resilience through diversification and network access at lower costs due to large user-base that can greatly reduce the overall total cost. Integration of satellite and terrestrial system in 5G increases spectrum availability, coverage zone and global service ubiquity. Four main use cases that indicate the role the satellite network can play when integrated with terrestrial networks are identified in [62]:

- Trunking and Head-End Feed use case – satellites provide high speed direct connectivity option to remote and hard to reach areas;
- Backhauling and Multicasting Tower Feed use case - satellites provide high speed multicast connectivity to wireless towers, access points and the cloud;
- Communications on the Move use case – satellites provide a direct and/or complementary connection for users on the move such as on planes, trains, vehicles and ships;
- Hybrid Multiplay use case - satellites are a solution for delivering content complementing or interworking with terrestrial broadband to individual homes and offices, as well as providing direct broadband connectivity in some case with the ability to multicast the same content (video, HD/UHD TV, as well as other non-video data) across a large coverage area (e.g., for local storage and consumption).

Furthermore, Figure 2-3 shows an extended list of satellite 5G use cases identified in [61] that exemplify roles of satellites in future networks.



**Figure 2-3: Satellite 5G use cases [56]**

The following section is dedicated to the identification of relevant scenarios for the simulations to test the traffic flow control algorithms, while Section 4 details the developed algorithms and their preliminary testing activities.

### 3 Multi-connectivity scenarios

This section contains the description of the process followed for the definition of the scenarios that will be utilised for the simulations of the algorithms in Section 4 and 5.

#### 3.1 Multi-connectivity simulation approach

The algorithms presented in Section 4 will be validated and evaluated using a simplified but still realistic and relevant simulation scenario. In [70], a high-level description on deployment scenarios can be found, including carrier frequency, aggregated system bandwidth, network layout, BS/UE antenna elements, UE distribution/speed and service profiles. In the remaining part of the project, the algorithms presented will be further developed and tested for the selection of the one(s) that will be demonstrated on the project PoC.

The deployment scenarios of interest are developed to cover the three main service families: eMBB (enhanced Mobile BroadBand), mMTC (massive Machine Type Communications) and URLLC (Ultra-Reliable and Low Latency Communications). The selection of the deployment of interest shall be done considering the specific algorithm objective. Some interesting deployment scenarios are:

- The **indoor hotspot** deployment scenario focuses on small coverage per site/TRxP (transmission and reception point) and high user throughput or user density in buildings. The key characteristics of this deployment scenario are high capacity, high user density and consistent user experience indoor;
- The **dense urban microcellular** deployment scenario focuses on macro TRxPs with or without micro TRxPs and high user densities and traffic loads in city centres and dense urban areas. The key characteristics of this deployment scenario are high traffic loads, outdoor and outdoor-to-indoor coverage. This scenario will be interference-limited, using macro TRxPs with or without micro TRxPs. A continuous cellular layout and the associated interference shall be assumed;
- The **rural** deployment scenario focuses on larger and continuous coverage. The key characteristics of this scenario are continuous wide area coverage supporting high-speed vehicles. This scenario will be noise-limited and/or interference-limited, using macro TRxPs;
- The **urban macro** deployment scenario focuses on large cells and continuous coverage. The key characteristics of this scenario are continuous and ubiquitous coverage in urban areas. This scenario will be interference-limited, using macro TRxPs (i.e. radio access points above rooftop level);
- The **high-speed** deployment scenario focuses on continuous coverage along the track in high-speed trains. The key characteristics of this scenario are consistent passenger user experience and critical train communication reliability with very high mobility;
- The **extreme Long Range** deployment scenario is defined to allow for the Provision of services for very large areas with a low density of users whether they are humans or machines (e.g. Low ARPU regions, wilderness, areas where only highways are located, etc). The key characteristics of this scenario are macro cells with very large area coverage supporting basic data speeds and voice services, with low to moderate user throughput and low user density;
- The **urban coverage for massive connection** scenario focuses on large cells and continuous coverage to provide mMTC. The key characteristics of this scenario are continuous and ubiquitous coverage in urban areas, with very high connection density of mMTC devices. This deployment scenario is for the evaluation of the KPI of connection density;



- The **highway** deployment scenario focuses on a scenario of vehicles placed in highways with high speeds. The main KPIs evaluated under this scenario would be reliability/availability under high speeds/mobility (and thus frequent handover operations).

Furthermore, in [70] the possible KPIs are described, a subset of these KPIs can be analysed in the simulations. The choice of the KPIs subset shall be done considering the objective of the algorithm under investigation. The possible KPIs are:

- Peak data rate;
- Peak spectral efficiency;
- Bandwidth;
- Control plane latency;
- User plane latency;
- Latency for infrequent small packets;
- Mobility interruption time;
- Inter-system mobility;
- Reliability;
- Coverage;
- UE battery life;
- UE energy efficiency;
- Cell/Transmission Point/TRxP spectral efficiency;
- Area traffic capacity;
- User experienced data rate;
- 5th percentile user spectrum efficiency;
- Connection density;
- Mobility;
- Network energy efficiency.

These KPIs are characterized in the document with evaluation methodology and target values.

After defining the deployment scenario and the KPIs, the last parameters that shall be selected are related to the channel model and the transmitter and receiver characteristics. In [71], the channel model for different frequencies and in different conditions are presented. The document presents different scenarios, antenna, path loss and fading models that can be selected considering the environment of interest. Finally, in [72] and [73] parameters for Rx and Tx dimensioning are described.

An exemplary simulation scenario, considering the references presented in this section, is summarized in Table 3-1. This scenario was used to derive some of the testing environments for the traffic flow controllers presented in Section 4 and 5.

The objective is to verify the capability of the algorithm to associate/allocate users to the three different radio access technologies (Macro, Micro and Satellite), maximizing the downlink (DL) UEs data rate and minimizing the BSs power transmission, considering a dense urban environment.

**Table 3-1: Simulation parameters**

Attributes	Macro Cells	Micro Cells	Satellite
Carrier Frequency	4 GHz	30 GHz	2 GHz (or 20 GHz)
System bandwidth	200 MHz	1 GHz	20 MHz (or 800 MHz)
Layout	Hex grid, cell radius 200 m	Random drop, 3 Micro per Macro	100 – 500 km per beam
DL Maximum Power	30 dBm	20 dBm	36 dBW/MHz (EIRP)
Maximum BTS Power	43 dBm	33 dBm	
BS antenna gain	11 dB	11 dB	
Path Loss	$128.1 + 37.6 \log_{10}(d)$ (d in m)	$24 + 45 \log(d+20)$ (d in m)	$32.45+20 \log_{10}(fc)+20\log_{10}(d)$ (fc in GHz, d in m)
UE antenna gain	0 dB	0 dB	0 dB
UE position	Random drop, 10 users per macro – 20 users per satellite beam		
Noise power	-99 dBm		
KPIs	Peak data rate, Network energy efficiency		



### 3.2 Management and orchestration for terrestrial and satellite resources

Innovative and advanced architecture concepts coupled with the adoption of autonomous network management techniques are essential for seamless integration of terrestrial and satellite networks. Towards this end, some recent studies investigated the applicability of 5G mainstream softwarisation technologies such as Software Defined Networking (SDN) and Network Function Virtualization (NFV) for integrating satellite networks with 5G terrestrial networks. Khalili et al. has surveyed the challenges and suggested solutions to integrated satellite systems into 5G terrestrial networks in [58] and recommended MANO<sup>1</sup>-like framework for rapid provisioning of network services and management and orchestration of heterogeneous terrestrial and satellite resources in [59]. Reference management and orchestration architectures for integrating satellite components into terrestrial 5G are specified by 3GPP in [60]. These reference architectures are shown in Figure 3-1 and Figure 3-2 below.

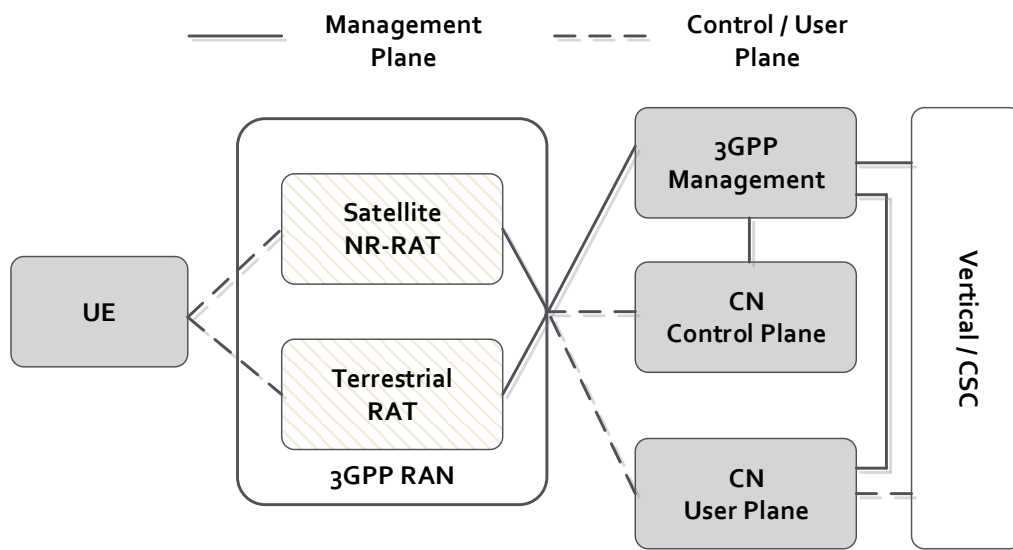


Figure 3-1: Reference architecture for the management of a satellite NR-RAT

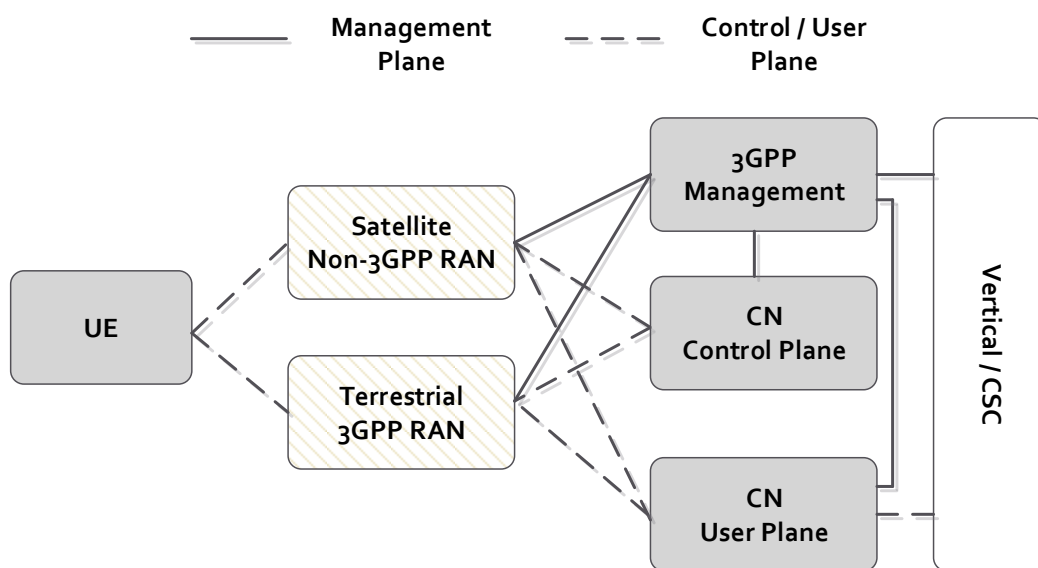


Figure 3-2: Reference architecture for the management of a non-3GPP satellite RAN

<sup>1</sup> <https://www.etsi.org/technologies/nfv/open-source-mano>

## 4 Proposed algorithms for multi-connectivity and related software

This section presents the traffic flow control algorithms designed in T4.2 and compared in T4.3. The first subsection 4.1 describes the network simulator that was developed in the scope of the project in T4.3 for the evaluation of the various algorithms proposed, while section 4.2 provides an outline of the various algorithms presented and section 4.3 reports their GAP analysis. This section is concluded with section 4.4, that provides an outline of the software that is being developed for the European demonstrator.

The goal of this section is to provide the reader with a high level understanding of the algorithms functioning and their different objectives. The entire mathematical description of the algorithms can be found in section 5.

### 4.1 Open source Network Simulator

The preliminary analysis of the proposed control algorithms conducted in T4.2 focused on their mathematical characteristics and properties, as their design required the usage of theoretical results from several research areas, spacing from machine learning to control theory.

To allow a better comparison among the performance of the algorithms in more realistic settings, aimed at the selection of the most suitable algorithms for PoC of the project, we developed in T4.3 a network simulator that we open sourced (currently available on github at: <https://github.com/trunk96/wireless-network-simulator>, but after first release it will be uploaded on zenodo to give it a DOI).

The simulator includes interference models, realistic network resource usage, the possibility to include moving APs as LEO satellites or aerial drone ones, user mobilities, different profiles for QoE depending on users and connection services. All of these functionalities are offered in a flexible way that allows the integration with a broad range of algorithms, spacing from network selection solutions to dynamic traffic steering ones, with control logics that can be deployed either in a centralised or distributed (on UEs or APs) way.

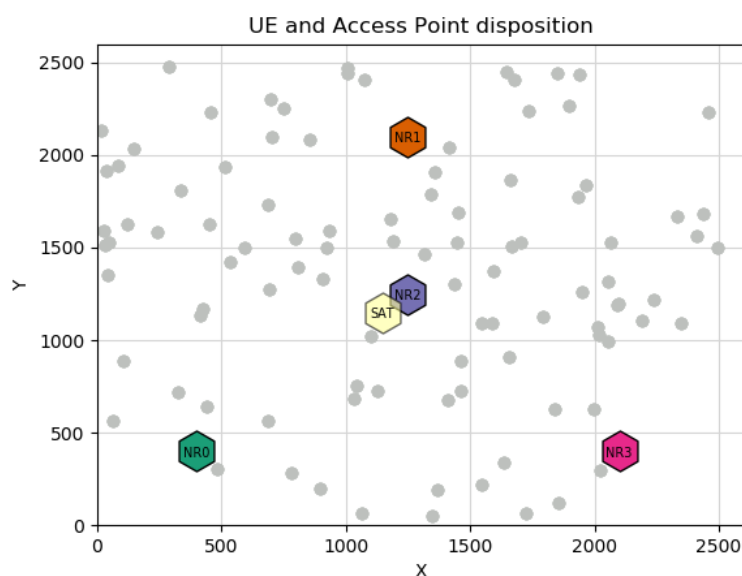


Figure 4-1 Network Simulator GUI. Scenario with 3 ground AP and a Satellite. Grey dots: UEs.

We now detail the resource allocation procedures considered in the simulator. More details on its functioning will be given, together with any update, in D4.3.

#### 4.1.1 5G NR and 4G LTE resource allocation description

As introduced, 5G New Radio (NR) APs have a limited set of resources, both in terms of frequency bandwidth and time to allocate UE requests. The goal of the algorithms designed in 5G-ALLSTAR is to optimally control such resources to enable an efficient and resilient solution for multi-connectivity, integrating also satellite related resources.

The minimum allocation unit for a 5G NR AP is the Physical Resource Block (PRB), each composed by 12 frequency subcarriers with a  $2^\mu \cdot 15$  kHz bandwidth and a time duration of  $2^{-\mu} \cdot 1$  ms, where  $\mu \in \{0,1,2,3,4\}$  is the parameter called *numerology* defined by 5G NR standards. The number of PRBs available on AP  $p$  depends on the available total bandwidth on the AP and on its numerology, as defined by 5G NR standards [74].

For 4G LTE APs the definition of PRB still stands, but the numerology parameter is constrained to  $\mu = 0$ , so there is no flexibility on using less/more subcarrier bandwidths and more/less time slot durations.

The receiving power, or RSRP,  $\mathcal{P}_i$  represents the transmission power measured by the UE  $i \in I$  between itself and the AP  $p \in P$  is computed as follows:

$$\mathcal{P}_{i,p} = \mathcal{P}_p \cdot G_p \cdot L_p \cdot L_{i,p}$$

where  $\mathcal{P}_p$  is the AP's antenna power,  $G_p$  is the AP's antenna gain,  $L_p$  is the AP's feeder losses and  $L_{i,p}$  is the path loss between UE  $i$  and AP  $p$ .

In our simulator, the path loss  $L_{i,p}$  is computed through the COST-HATA model [75,76], that is a statistical model that considers many factors as the buildings density (rural, suburban, urban), the carrier frequency used for the communications and the relative heights of UE and AP.

In order to estimate the number of resource blocks to be allocated by the AP  $p \in P$  for the communication with the UE  $i \in I$ , the *signal-over-interference-plus-noise-ratio* (SINR) has to be computed. The thermal noise part can be computed according to:

$$\mathcal{N}_p = k_b T^{env} B_p \Theta_p$$

$$\Theta_p(t) = \frac{\sum_{\tau \in (t-T, t)} \sum_{j \in I \setminus i} C_{j,p}(\tau) N_{j,p}(\tau)}{T \cdot \#R_p},$$

where  $\Theta_p(t)$  is the Resource Blocks Utilization Ratio (RBUR) of AP  $p$  at time  $t$ ,  $k_b$  is the Boltzmann constant,  $T^{env}$  is the environmental temperature,  $B_p$  is the total bandwidth for the AP  $p$ ,  $T$  is the length of the moving average,  $C_{j,p}(t)$  is equal to 1 if UE  $j$  is connected to AP  $p$  at time  $t$  and 0 otherwise and  $N_{j,p}(t)$  is the number of PRB allocated by AP  $p$  to UE  $j$  and  $\#R_p$  is the total number of resource blocks of AP  $p$ .

The interference part is computed as follows:

$$\mathcal{J}_{i,p} = \sum_{p' \neq p} F_{p,p'} \mathcal{P}_{i,p'} \cdot \Theta_{p'}(t)$$

where  $F_{p,p'}$  is 1 if AP  $p$  and  $p'$  share the same carrier frequency and 0 otherwise.

Using the equations for noise and interference it is possible to compute the SINR, and so it is possible to estimate the data-rate that can be transmitted allocating one PRB to UE  $i$  using the Shannon formula

$$r_{i,p} = 2^{-\mu} 10^{-3} B_{PRB} \log_2(1 + SINR_{i,p})$$

$B_{PRB}$  is the bandwidth of a single PRB and it can be computed as  $B_{PRB} = 12 \cdot 2^\mu 15\text{kHz}$ .

Now, given a certain bitrate request  $b_p^i$  from UE  $i$ , it is possible to compute the number of resource blocks to be allocated by AP  $p$  to satisfy the request:  $n_{i,p}^{PRB} = \lceil (b_{pk}^i / r_{i,p}) \rceil$ .

#### 4.1.2 Satellite resource allocation description

In light of what was state regarding the issues related difference coverages offered by a satellite and a C-RAN controller, we consider only a portion of resources available for the multi-connectivity of the area under the control of the considered C-RAN. Implicitly, we assume that there is no technical issue in the satellite resource slicing (e.g., the negotiation on the satellite access occurs between the C-RAN and the UE passing through the ground segment or the satellite knows which resource slice the UE may have access to) and can consider the satellite AP as a virtual AP with a different resource usage and interference models, compared with 5G NR ones.

Contrary to ground APs, the satellite APs use Time Division Multiple Access (TDMA) in order to serve multiple UEs at the same time. In this case, the minimum allocation unit is a *block of symbols* that occupies a certain time slot in the satellite time-frame.

The receiving power  $P_{i,p}$  can be still computed as in the previous section, but in this case the path loss function will be the *Free Space Path Loss*:

$$L_{i,p}^{FSP} = \left( \frac{4\pi d_{i,p} f}{c} \right)^2,$$

where  $d_{i,p}$  is the Euclidean distance between UE  $i$  and AP  $p$ ,  $f$  is the carrier frequency used and  $c$  is the speed of light.

The thermal noise can be computed as in the 5G NR case as for the interference. Using the Shannon formula (considering that this time the bandwidth is the total bandwidth of the satellite AP since the TDMA utilises all the bandwidth only for a certain amount of time), one has that the bitrate obtainable by a single block of symbols is

$$r_{i,p} = bB \log_2(1 + SINR_{i,p}),$$

where  $b$  is the ratio between the number of symbols in a single block and the total number of symbols of the satellite AP. The number of blocks to be allocated for a requested bitrate  $b_{pk}^i$  from UE  $i$  is then computed as

$$n_{i,p}^{blocks} = \lceil (b_{pk}^i / r_{i,p}) \rceil.$$

The simulator is still in development and we are now considering the addition of different communication protocols, that will impact the relation between transmitted bitrate and resource used. Any improvement will be detailed in D4.3.

## 4.2 Overview on the 5G-ALLSTAR suite of algorithms for multi connectivity

In this section we detail the main characteristics of each developed algorithm, along with its possible role in the project PoC, its level of testing (e.g., numerical validation or tested on the network simulator) and other main features.

### 4.2.1 Traffic steering and network selection in 5G networks based on reinforcement learning

**Main Characteristics:** Model free control scheme that has to be deployed in the C-RAN, as the controller requires knowledge on the entire state of the network (in terms of utilised resources).

**Control logic:** Network Selection -Every incoming request is allocated to an AP depending on the user/service profiles and the state of the network. Traffic Steering - The controller distributes the available network resources so that every connection receives a minimum amount determined by its minimum level QoS plus a share of resources that improves the QoE of the network users. Such allocation is dynamic (changes over time depending on the network state).

**Compliance with the 5G-ALLSTAR architecture:** Total, as the algorithm treats the AP as resource providers and governs how such resources shall be used. The algorithm assumes the satellite resources, as defined in section 4.1.2 can be used with no conflict with other C-RAN, thanks to an adequate slicing solution.

**Limits and cons:** being a model free solution it requires a training time, in which performances can not be assured. May be difficult to scale towards very broad C-RANs, as the training complexity increases with the number of AP (no problem for 10-20).

**Possibility to include QoE:** QoE is already included with different profiles depending on the service class (e.g., real-time, multi-codec video streaming, web browsing...) and user profile.

**Testing level:** validated on a numerical simulator and already produced a scientific publication.

**Likelihood of being integrated in the PoC:** Medium, as this algorithm is better suited for scenarios in which 5-10 APs are available to multiple UE, while in the PoC only two AP will be deployed. The algorithm is being tuned in the scope of T4.3 to evaluate its performances on scenarios similar to the PoC and also on large scale (100+ UE) systems, also in its Deep Learning variant (under development).

### 4.2.2 Wardrop Equilibrium based control

**Main Characteristics:** Model based control scheme that can be distributed either in the UEs or in the APs (requires only partial knowledge on the system state). Strong theoretically proven convergence properties.

**Control logic:** Each UE steers (migrates) dynamically the traffic associated to the various QoS-flows over the APs it can connect to according to a mathematical control law. Following this law, every UE maximises the quality of its connection routing (according to user-defined preferences, such as QoE, cost, network resource usage,...) in an unilateral fashion (i.e., without cooperation among the UE or centralisation no better equilibrium state can be found).

**Compliance with the 5G-ALLSTAR architecture:** Total, as the algorithm treats the AP as resource providers and the control logic can be distributed in the UE or in the D-RANs that govern the various APs.

**Limits and cons:** The convergence speed of the algorithm slows as the number of UE grows (tested up to 50). Even in unfavourable network conditions (e.g., huge imbalance in initial routing, very scarce network resource availability), in our testing we stressed such unfavourable situations and still obtained reasonable convergence speeds.

**Possibility to include QoE:** QoE can be included as a performance metric on which the steering can be decided. This takes the name of "latency functions" in the literature related to Wardrop Equilibria.

**Testing level:** at the first issue of the deliverable it was validated on a numerical simulator neglecting interference. The algorithm has now been tested on the new network simulator that accounts for interference models and offers several other improvements compared to the previous study.

**Likelihood of being integrated in the PoC:** High, as this algorithm requires to function a very mild mathematical assumption (boundness of the derivative of the “latency function” selected to evaluate the traffic routing) and can be used in networks constituted even by hundreds of UE. The algorithm needs to be slightly simplified to cope with the reduced setting presented in PoC but a tailored version to the PoC needs and case studies is being developed in T4.3. The algorithm is also being adapted to be implemented on current communication protocols for real adaptive video streaming.

#### 4.2.3 Friend or Foe Q-Learning

**Main Characteristics:** Model free control scheme that has to be distributed at AP level.

**Control logic:** Each AP compete with the others to allocate the most lucrative connections to optimally exploit its network resources. This adversarial setting is typical of multi-tenant networks or in scenarios in which different user tariffs are present.

**Compliance with the 5G-ALLSTAR architecture:** Good, as the algorithm is distributed in the D-RAN controllers, requiring no cooperation among them. Only issue is related to the inclusion of the satellite, as it should be able to discern which C-RAN covers the area in which the UE is located and it may not be possible without cooperation with the ground segment or changes in the communication protocol, in light of the discussion in section 2.1.

**Limits and cons:** Being a model free solution it requires a training time, in which performances can not be assured. May be difficult to train in C-RANs with more than 10 APs.

**Possibility to include QoE:** QoE can be included with a re-design of the reward function, but the adversarial nature of the algorithm and the competition of the APs is probably not the best setting for QoE inclusion.

**Testing level:** Tested on a numerical simulator, produced a scientific publication.

**Likelihood of being integrated in the PoC:** Low, the case study considered for the PoC are not characterised by an adversarial setting and a significant redesign would be needed.

#### 4.2.4 Maximal Load Balancing

**Main Characteristics:** Model based solution for the optimal control of network resources by offloading traffic from overloaded APs to unloaded ones.

**Control logic:** The offloading of traffic from overloaded APs to less utilised ones happens on a threshold-based logic and takes into account the service type (e.g., is the connection delay tolerant for satellite routing?) and potentially the user preferences (e.g. battery usage).

**Compliance with the 5G-ALLSTAR architecture:** Total, as the algorithm can be distributed in the D-RAN controllers, with some level of cooperation (state information exchange) that is compliant with the existence of the C-RAN logical units or can be deployed in the centralised controller that oversees the whole RAN.

**Limits and cons:** The optimality of the resource usage has no clear mathematical backbone but empirically the improvements are significant. May require a good optimisation solution in scenarios with many APs.

**Possibility to include QoE:** This algorithm has been designed for network performances, but QoE may be used to determine how likely is for a connection to be offloaded on a satellite AP.

**Testing level:** Tested on a numerical simulator.



**Likelihood of being integrated in the PoC:** High, as the algorithm is very simple to implement in real settings and offers immediate benefits.

#### 4.2.5 Reinforcement learning resource optimization for mobile relays

**Main Characteristics:** Model free control solution to be deployed in UEs to optimise their frequency resource block and transmit power usage.

**Control logic:** Each UE learns the best strategy to use the available frequency block groups and transmit power on its own, by interacting with the environment. The goal is to have stable connection with good QoE profiles, while reducing interferences.

**Compliance with the 5G-ALLSTAR architecture:** Total, as the algorithm manages the uplink resources of the UE and requires some partial information on the network state from the various APs.

**Limits and cons:** The training time of this algorithm increases with the number of available APs, all the control decisions are taken by UEs (low computing devices may have difficulties with the training).

**Possibility to include QoE:** QoE is included in the so-called reward function, that guides the learning process of the UEs towards its maximisation.

**Testing level:** tested on a numerical simulator.

**Likelihood of being integrated in the PoC:** Low, as the PoC use cases don't focus particularly on the setting for which this algorithm was developed for.

#### 4.2.6 Analytic Hierarchy Process) and Cooperative Differential Games for Multiconnectivity

**Main Characteristics:** Model based solution that decides an order (hierarchy) of priorities and objectives to maximise QoE.

**Control logic:** The APs of a C-RAN cooperate to maximise the QoE of their users by allocating the various connections on the available RATs depending on the (ordered) preferences of the users and their requirements in terms of QoS.

**Compliance with the 5G-ALLSTAR architecture:** Total, as the algorithm can be deployed in the C-RAN to oversee the functioning of the available APs and decide on the connection allocations.

**Limits and cons:** This algorithms requires a huge effort in identifying which are the user preferences, as the optimisation of the QoE and the satisfaction of the user requirements is the fundamental part of its optimisation. The questionnaire of the project may help in this direction but could still be too limited for a realistic application.

**Possibility to include QoE:** QoE optimisation is the core of the algorithm.

**Testing level:** preliminary tested on a numerical simulator, produced a scientific publication.

**Likelihood of being integrated in the PoC:** Low, as the QoE in the PoC will be demonstrated on a video stream by adapting its encoding and routing. Some aspects of this algorithm (e.g. connection preferences) may guide the tailoring of another algorithm to the PoC, as this algorithm is among the ones that better capture QoE related aspects.

#### 4.2.7 ZeroSum games and Minimax for Multi-Connectivity

**Main Characteristics:** Control solution for dual-connectivity (can be extended to multi-connectivity) between two AP. The AP compete to allocate the throughput requested to optimally use their radio resources.

**Control logic:** The two AP involved in the dual connectivity play a game to decide which AP routes the most traffic and consequently obtains the better return, while avoiding to overcommit their own resources.

**Compliance with the 5G-ALLSTAR architecture:** Total, as the control logic is distributed in the controllers of the two APs.

**Limits and cons:** The current formulation only takes into account dual connectivity but may be expanded to multi-connectivity. In the testing conducted so far the improvements are marginal.

**Possibility to include QoE:** QoE can be included in the optimisation process but in the current formulation the focus is on network resource usage.

**Testing level:** preliminarily tested on a numerical simulator.

**Likelihood of being integrated in the PoC:** Low, as this formulation was used as the basis to design more impactful algorithms based on Reinforcement Learning.



### 4.3 GAP analysis and comparison

The following table shows a GAP analysis of the proposed algorithms that will be used as a starting point for the activities of T4.3 regarding the selection of the algorithm(s) that will be implemented and tested in the PoC of the project and for the refinement of the algorithms themselves.

**Table 4-1: GAP analysis of the proposed algorithms for traffic flow control**

Section	Methodology	Main Characteristics	Scalability	QoE integration
4.2	RL	Model Free, Optimal Network Selection based on feedback, C-RAN based	Limited but compliant with the typical number of AP available in an AN	Explicit, directly integrated with multiple services and their different QoE functions
4.3	Control Theory and Game Theory	Dynamic (in the sense that evolves over time), adversarial environment, feedback based	The performance may degrade with a high number of commodities	Explicit, as long as the latency functions are re-defined in terms of a QoE metric
4.4	Game Theory	Model Free, Optimal Network Selection based on feedback, D-RAN based in an adversarial framework	Limited but compliant with the typical number of AP available in an AN	A redesign of the reward function is needed, as at the moment the algorithm focuses on AP performance
4.5	Linear Programming	Model based, explicit optimisation, feedback-based	High, as the linear problem is characterised by low complexity	QoE is maximised by an optimal load balancing over the APs of the network resources
4.6	RL	Model free, feedback based, D2D	Compliant with the target scenario, may benefit from ad-hoc solutions as state-space aggregation	Explicit, as it is directly included in the reward function
4.7	AHP, Game Theory	Model driven, hierarchical prioritization of objectives and performances	Limited in the preliminary formulation reported	explicit
4.8	Game Theory	Model driven, adversarial and distributed framework	Limited in the preliminary formulation reported	Explicit if included in the reward function

#### 4.4 Plan for the European multi-connectivity demonstrator

As envisaged in WP2, the EU demonstrator for multi-connectivity that will be integrated in WP5 requires the development of an ad-hoc video streaming platform to provide 8K videos over multiple radio channels at the same time.

In WP4 such software component is being developed to cope with the required flexibility needed to deploy the algorithm(s) for multi connectivity that will be selected for the final demonstrator.

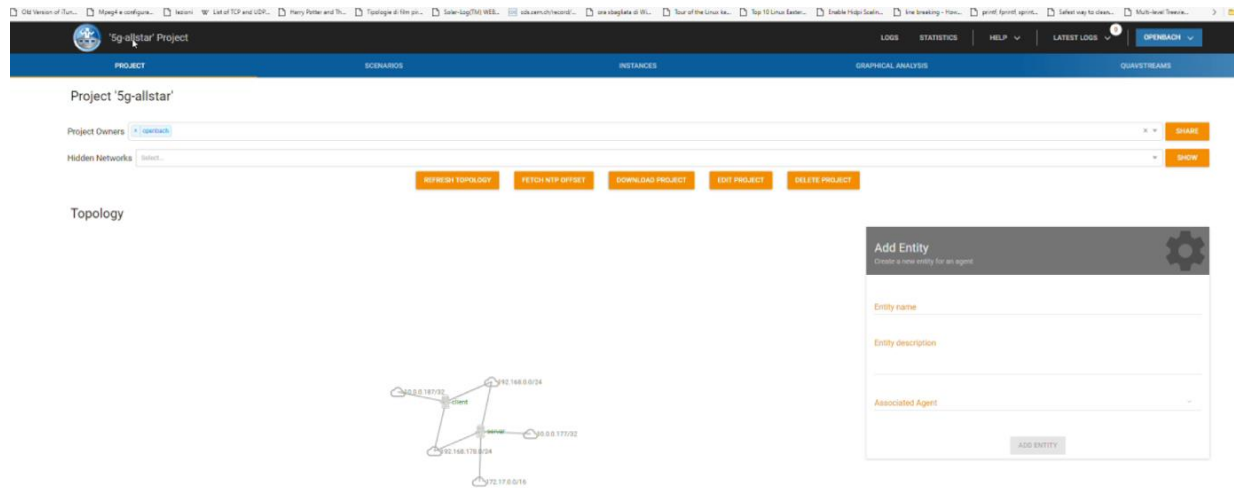


Figure 4-2 Openbach GUI for the video server

Figure 4-2 shows the Openbach Controller view for the video server. From this view it is possible to start, stop an 8K video streaming scenario, as well as to get statistics from the ongoing scenarios.

In the activities of T4.3, some testing has already been conducted with some simplified version of the proposed multi-connectivity algorithms presented in this chapter to have a feasibility assessment for their deployment. The current video streaming platform is composed by two virtual machines: the first one acts as an Openbach controller and hosts the video server and the traffic flow controller, while the second one acts as an Openbach agent and hosts the traffic flow manager and the video client. The first virtual machine has two network interfaces towards the second virtual machine, and it is possible to add a delay generator to one of these two network interfaces in order to emulate the behaviour of the satellite AP. Moreover it is possible to set dynamically and in real-time the amount of video streaming data to be sent to one link and/or to the other one.

From the Openbach Controller GUI it is also possible to show some video server and video client data about the ongoing scenarios through the well-known visualization platform *Grafana*. These data include measures of latency, streaming video quality, bitrate sent, congestion window length, buffer length at client side, etc.

This video server will be integrated with the rest of the software components of 5G-ALLSTAR and will be interfaced with the channel emulators and the cRRM to complete the PoC.

A similar setup is also being developed by the Korean partners for their specific use cases, and at the delivery month of this document no delay in the development activities is to be reported despite the ongoing pandemic.

## 5 Mathematical formulation and results for the proposed algorithms

This section contains the entire mathematical formulation for the seven algorithms developed so far in the scope of T4.2. The simulation testing conducted for D4.2 are aimed at stressing the capabilities of the algorithms and their mathematical properties in unfavorable scenarios, to assess their robustness and applicability from a scientific prospective. Further testing are undergoing in T4.3 employing also the open source simulator developed for the project, with the final aim of tailoring and deploying a subset of the presented algorithms in a real multi-connectivity scenario for the project PoC, integrating some functionalities in a real 8k video streaming platform.

Every algorithm is then described in a formal way, contrary to their higher level description provided in chapter 4, providing also the mathematical background needed to prove their properties.

In particular, the first algorithm presented in Section 5.1 presents a traffic flow controller based on Reinforcement Learning (RL) for the problem of network selection and feedback-based resource allocation for QoE maximisation. The rationale of the algorithm is to develop a controller able to optimally allocate the network resources, maximising the overall QoE of the network users, without having any knowledge on the network model and its stochastic characteristics (e.g., unknown arrival and departing rates for connections). Three different classes of services were designed, to better capture the different QoE profiles (e.g., a video stream behaves differently from web browsing in terms of QoE).

The second algorithm presented in Section 5.2 investigates a solution based on both, the frameworks of control theory and game theory that revolves around the concept of Wardrop equilibria. Wardrop equilibria are a generalisation of Nash equilibria in problems in which an infinite amount of players is competing, as in the case of telecommunication networks and data flows. The developed algorithm provides a distributed controller that dynamically steers the traffic from the various QoS flows over the available APs, exploiting multi-connectivity to the fullest. The formulation proposed, compliant with the standard modelling approach for dynamical networks typical of Wardrop-related studies, allows to define the problem in terms of an adversarial competition for network resources among services, so that a certain selected performance metric is optimised in the adversarial-Nash sense. The simulations validated the approach in a scenario in which various service providers were competing to decrease their transmission power usage in a capacitated network slice. With minor modifications to the *latency functions*, the proposed algorithm can be adapted to other performance metrics, eventually related to QoE.

The third control solution detailed in Section 5.3, is an algorithm based on game-theory and multi-agent reinforcement learning for the problem of network selection. Compared to the first algorithm, this scenario avoids the presence of a centralised network controller in favour of a distributed one (over the D-RAN as shown in Figure 2-1) attaining similar properties and results but formulating the problem as an adversarial competition between access points.

The fourth proposed controller in Section 5.4 is based on linear programming, and addresses the problem of load balancing at C-RAN level. The developed algorithm is able to enhance the throughput level in a satellite-terrestrial scenario, showing how the satellite integration in 5G scenarios may provide a significant improvement in terms of connection performance and, consequently, in the QoE of the users.

The fifth, algorithm presented in Section 5.5 a multi-agent reinforcement learning based solution for the optimisation of mobile relays, in terms of optimal resource allocation. The scenario considered for the development of this solution is compliant with the multi-connectivity framework proposed and focuses on Device-to-Device (D2D) communication and interference mitigation and other QoE-related performance indices optimisation.

The final two algorithms, presented in Section 5.6 and 5.7, are candidate approaches based on Game Theory that were explored during the life-cycle of the project that proved to be useful for

the development of the algorithms of the other sections. They are reported in a preliminary formulation, that may be finalised for D4.3 depending on the evaluations of T4.3.

The refinement, finalisation and additional testing of the algorithms will be reported also in D4.3, together with the implementation details relative to the integration with the PoC of the project.

## 5.1 Traffic steering and network selection in 5G networks based on reinforcement learning

As already introduced, traffic steering is a fundamental functionality of 5G networks for multi-connectivity and encompasses the ability of routing, or *steering*, a given traffic flow over one of the several different RATs available to the UE that initiates the connection.

Traffic steering is then heavily linked to the concept of Optimal Network Selection, to the point that the two are often used as synonyms, as the routing/steering decisions over the available APs of the considered RATs shall be driven by a feedback-based analysis of the network state and performances, potentially also taking into account user connection preferences.

In this first algorithm, the state of the network will be represented by the downlink cell allocated bitrate, while the performances will be measured in terms of connection quality, in order to capture the satisfaction level of the users. The traffic steering problem will be modelled with the Markov Decision Process (MDP) formalism, in addition Reinforcement Learning (RL) was selected for the controller design due to its ability to deal with complex scenarios without requiring an explicit model.

The main motivation behind this approach is then to characterize and show the effectiveness of such an approach through its application to a simulated 5G network scenario. For the model development, in order to account for services which can be offered with multiple bitrates and consequently with different qualities, several classes of utility functions were defined to capture the different profiles of QoE perceived by the users. Additionally, the MDP was formulated, to overcome the scalability problem, utilizing Approximate Dynamic Programming (ADP) techniques, and, in particular, state-space aggregation.

The rest of the section is organized as follows: Section 5.1.1 presents the state of the art of traffic steering in 5G networks, as well as the proposed novelties; while Section 5.1.2 provides the required background regarding MDPs; Sections 5.1.3 and 5.1.4 present respectively the problem modelling and the network selection algorithm; while Section 5.1.5 shows some simulation results to validate the proposed algorithm.

### 5.1.1 State of the art, innovations and limitations of the proposed approach

Traffic steering is the process of distributing traffic load in order to exploit the available network resources on a set of heterogeneous RATs that constitute a Radio Access Network (RAN) [1]. The enabling procedure for the process of traffic steering is the so-called *network selection* [2], in [3] a feedback-based analysis of the network aimed at identifying the best APs for the connections, also referred to as Protocol Data Unit (PDU) sessions. Such a feedback analysis is conducted based on the level of usage of the various APs, together with their characteristics (e.g., operating prices, reliability) and both, user and operator preferences.

In 5G networks, the network selection shall be done in such a way that the QoS requirements of the various connections (e.g., minimum required bitrate or maximum tolerated delay) are satisfied, and, to univocally define such QoS requirements, the concept of QoS-flow was introduced [4]. Each PDU session is divided into several QoS-flows, each characterized by a standardized set of QoS requirements depending on its service characteristics [4], leading to the identification of, as of now, 22 different QoS flow types identified by a corresponding 5G QoS Identifier (5QI).

Several approaches were already studied for the problems of traffic steering and network selection, ranging from Multiple Attribute Decision Making (MADM) [5], [6] and fuzzy logic [7] to game theory [8],[9]. MDPs and RL were also already explored, for example, in [10] and [11].

An important feature for traffic steering in 5G networks is the ability to support multiple bitrates for the supported services. An example of such a service is multi-codec video streaming, whose streaming quality should be dynamically varied based on user needs, preferences and network conditions. As introduced in [12] for generic resource allocation problems, a possible approach is to associate utility functions to the services to model how the amount of assigned resources affects the user satisfaction.

From the modelling perspective, this work takes into account 5G requirements by supporting a heterogeneous network scenario while also being able to offer flexibility in service requirements. In this last respect, the network controller presented in this section aims at maximizing the perceived connection quality of the network users, thanks to an *ad hoc* utility function-based modelling of the performances of the considered classes of services, inspired by [12] and expanding the modelling of [11], which defined user satisfaction based on a throughput threshold. From the MDP and RL perspective, the proposed solution expands the algorithm presented in [10], thanks to the concepts of state aggregation and ADP, aimed at avoiding the so-called curse of dimensionality (e.g., see [13]), which affects the solutions based on Dynamic Programming (DP) and tabular RL in realistic scenarios.

The evaluation of RL approaches relies on the availability of a realistic environment, i.e., in our case, a representative 5G network simulator. At the end of this section, it is reported a preliminary evaluation of the proposed concepts by using numerical simulations considering a limited set of 5G network characteristics. Moreover, by considering some of the neglected characteristics, such as, for instance, the user mobility, the developed RL algorithm would probably not be adequate but its ideas will serve as a starting point for other improvements that could lead to the solution that will be deployed in the project PoC.

### 5.1.2 Preliminaries on Markov decision processes and reinforcement learning

A MDP is a discrete-time stochastic control process defined by the tuple  $\{S, A, T, r, \Sigma, \gamma\}$ , where  $S$  is a discrete, finite state set,  $A = \cup_{s \in S} A_s$  is the finite action set in state  $s$ ,  $T$  is the state transition probability matrix,  $r$  is the reward function, such that  $r(s, a, s')$  is the immediate reward obtained in  $s$  when action  $a \in A_s$  is taken and state  $s'$  is the next state,  $\Sigma$  is the initial state distribution over the state space  $S$ , and  $\gamma \in (0, 1)$  is the discount factor, which weights immediate rewards versus future rewards. Standard MDP definitions rely on the Markovian (or memory-less) property and on the stationary distribution of the stochastic process. Under these assumptions, the transition probabilities are stationary.

A stationary policy  $u$  is a mapping of each state to an action, i.e.,  $u(s) = a, s \in S, a \in A_s$ . The MDP problem is aimed at finding an optimal policy  $u^*: S \rightarrow A$  that maximizes, in the long run, the expected discounted reward:

$$R(u) := E_{u, \Sigma} \left\{ \sum_{t=0, 1, \dots, \infty} \gamma^t r(s(t), a(t), s(t+1)) \right\}, \quad (5.1.1)$$

where  $s(t)$  and  $a(t)$  denote the state and the action at time  $t$ , respectively, and  $E_{u, \Sigma} \{ \cdot \}$  denotes the expected value under policy  $u$  with initial state distribution  $\Sigma$ .

The value function  $V_u(s)$  is the expected discounted reward starting from  $s$  and following policy  $u$  thereafter, and the action-value function  $Q_u(s, a)$  is the expected discounted reward, starting from  $s$ , taking action  $a \in A_s$  and following policy  $u$  thereafter:

$$V_u(s) := E_u \left\{ \sum_{t=0, 1, \dots, \infty} \gamma^t \cdot r(s(t), a(t), s(t+1)) \mid s(0) = s \right\}, \quad (5.1.2)$$



$$Q_u(s, a) := E_u\{\sum_{t=0,1,\dots,\infty} \gamma^t \cdot r(s(t), a(t), s(t+1)) \mid s(0) = s, a(0) = a\}, \quad (5.1.3)$$

where  $E_u\{\cdot\}$  denotes the expected value under policy  $u$ .

As mentioned at beginning of this section, the solution presented here is interested in solving the MDP by means of RL algorithms. Let the system be in a given state  $s \in S$ ; RL algorithms take an action  $a \in A_s$  based on the current policy and then, after the transition, observe the next state  $s' \in S$  and the obtained reward  $r(s, a, s')$ . Based on the observations, the RL algorithms update an estimate of the value function of state  $s$  or of the action-value function of the pair  $(s, a)$ .

RL algorithms differ by the rule used to decide the control action and by the rule used to update the value (or action-value) function. For this algorithm, the Q-learning algorithm is considered, but more complex RL algorithms are foreseen for the algorithm improvements (see the following sections). The Q-Learning update rule is

$$Q(s(t), a(t)) \leftarrow (1 - \alpha(t))Q(s(t), a(t)) + \alpha(t) \cdot \left[ r(s(t), a(t), s(t+1)) + \gamma \max_{a' \in A_{s(t+1)}} Q(s(t+1), a') \right]. \quad (5.1.4)$$

In Equation (5.1.4),  $\alpha(t) > 0$  is the learning rate and thus the key parameter for the algorithm convergence: if  $\sum_{t=1,\dots,\infty} \alpha(t) = \infty$  and  $\sum_{t=1,\dots,\infty} (\alpha(t))^2 < \infty$ , the estimate (5.1.4) converges to the optimal action-value function as  $t \rightarrow \infty$  [13]. The action is then decided based on the current estimate of the state-action value function, and, at time  $t$ , the current best policy is:

$$u(s(t)) = \operatorname{argmax}_{a \in A_s} Q(s(t), a), s \in S. \quad (5.1.5)$$

As the estimate (5.1.4) converges to the optimal action-value function, the policy (5.1.5) converges to an optimal policy.

To guarantee a certain degree of exploration of the state space set, an  $\varepsilon$ -greedy rule is followed for the action selection: in state  $s \in S$ , the current best action (5.1.5) is taken by the controller with probability  $1 - \varepsilon$ , where  $\varepsilon \in (0,1)$  is the exploration rate; a random action  $a \in A(s)$  is chosen with probability  $\varepsilon$ , i.e.:

$$u(s) \leftarrow \begin{cases} \operatorname{argmax}_{a \in A(s)} Q(s, a), & \text{with prob. } 1 - \varepsilon \\ \operatorname{rand}\{a \in A(s)\}, & \text{with prob. } \varepsilon \end{cases}, s \in S. \quad (5.1.6)$$

A large value of  $\varepsilon$  guarantees that different policies with respect to the current best one are explored, and thus avoids that the system remains stuck in a local minimum. A small value of  $\varepsilon$ , on the other hand, lets the system choose the best action based on the current estimates of the action-value function and favours the exploitation of the current best policy.

The choice of  $\alpha(t)$  and  $\varepsilon$  depends on the specific application.

### 5.1.3 Problem modelling

Let  $I$  be the set of UEs connected within a RAN, let  $K$  be the set of different services available to each UE, let  $P$  be the set of different APs of the RAN and let  $P^i \subseteq P$  be the set of APs available to UE  $i \in I$ .

Each AP is characterized by the amount of available resources in terms of bitrate, denoted with  $W_p, p \in P$ . Similarly, the services are characterized in terms of required bitrate. Different types of services are considered: elastic services, such as web browsing, for which the user-perceived quality improves with the assigned bitrate, and non-elastic services, such as augmented reality streams, for which a fixed bitrate is needed for the transmission. Among the elastic services, it is also possible to discern services in which the service quality varies depending on the encoding that is available at the considered bitrate, as for example multi-codec video streams.

Let  $w_{pk}$  be the amount of bitrate allocated on AP  $p$  for a service of type  $k$  and  $r_{pk}(w_{pk})$  be the perceived quality of connection experienced by the users. It is possible to characterize the three types of services as follows.

*Elastic traffic*: this kind of QoS-flow benefits from having more dedicated resources, therefore its user-perceived quality function grows with the allocated bitrate  $w_{pk}$  starting from a minimum value and up to a maximum one:

$$r_{pk}(w_{pk}) = \begin{cases} 0 & \text{if } w_{pk} < w_k^1 \\ f(w_{pk}) & \text{if } w_k^1 \leq w_{pk} \leq w_k^2 \\ f(w_k^2) & \text{otherwise} \end{cases} \quad (5.1.7)$$

An example of this kind of utility function is represented in Figure 5-1, where the utility function is proportional to the allocated bitrate, i.e.,  $r_{pk}(w_{pk}) = cw_{pk}$  up to a maximum bitrate  $w_k^1$ .

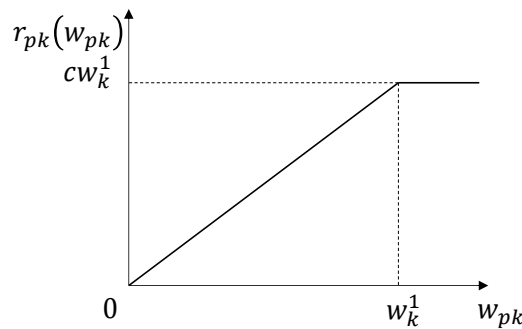


Figure 5-1: User-perceived quality of connection for service with elastic traffic

*Real-time traffic with guaranteed bitrate (fixed bitrate service)*: this kind of QoS-flows requires a fixed amount of bitrate, therefore its utility function is positive and constant if enough bitrate is allocated onto a suitable access network, 0 otherwise, as reported in Figure 5-2.

$$r_{pk}(w_{pk}) = \begin{cases} 0 & \text{if } w_{pk} < w_k^1 \\ r_{pk}^1 > 0 & \text{otherwise} \end{cases} \quad (5.1.8)$$

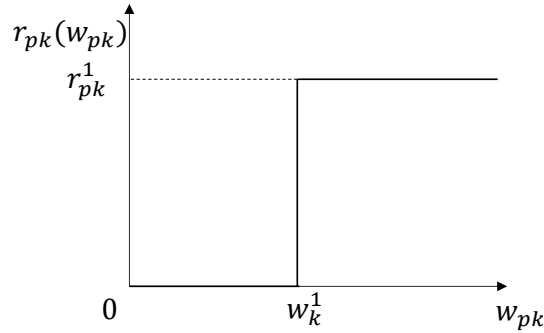


Figure 5-2: User-perceived quality of connection for service with transmission bitrate threshold

*Multi-codec traffic:* this kind of QoS-flows improves its quality depending on its encoding. The available encodings depend on the amount of resources allocated for the service, according to a distribution with multiple thresholds as reported in Figure 5-3.

$$r_{pk}(w_{pk}) = \begin{cases} 0 & \text{if } w_{pk} < w_k^1 \\ r_{pk}^1 > 0 & \text{if } w_k^1 \leq w_{pk} < w_k^2 \\ r_{pk}^2 > r_{pk}^1 & \text{if } w_k^2 \leq w_{pk} < w_k^3 \\ r_{pk}^c > r_{pk}^{c-1} & \text{if } w_{pk} \geq w_k^c \end{cases} \quad (5.1.9)$$

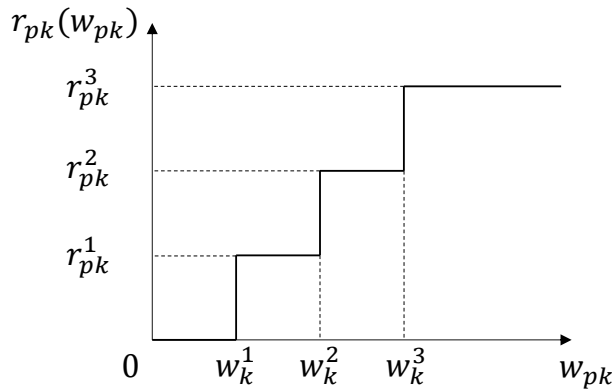


Figure 5-3: User-perceived quality of connection for service with multi-codec traffic

The reward associated with each state is represented by the cumulative QoE of the network users, obtained by summing all perceived quality of connection for all ongoing PDU sessions.

At time  $t$ , we denote with  $n_{pk}^c(t)$  the number of on-going QoS-flows of type  $k$  on AP  $p$ , considering the level  $c$  of the allocated bitrate, and with

$$n_{pk} = \sum_{c=1, \dots, C_k} n_{pk}^c,$$



where  $C_k$  is the number of bitrate threshold levels that characterize the reward function  $r_{pk}$ .

By defining  $\eta_p^1(s(t))$  as the minimum amount of bitrate required to support the on-going QoS-flows at the minimum bitrate level in state  $s$ , i.e.,

$$\eta_p^1(s(t)) = \sum_{k \in K} n_{pk} w_k^1, p \in C, \quad (5.1.10)$$

the state space  $S$  is defined as

$$S = \left\{ s = (n_{pk})_{p \in P, k \in K} \mid \eta_p^1(s) \leq W_p \right\} = \{s_1, s_2, \dots, s_{|S|}\}.$$

Since the considered resource (bitrate) is additive, and since the AP resources  $W_p$  are finite, the discrete state space  $S$  is finite as well. With little abuse of notation, the number of QoS-flows of type  $k$  on AP  $p$  in state  $s$  is denoted with  $n_{pk}(s)$ .

At each service request, the RAN controller has to decide whether to admit or reject the service request and also, in case of admission, the AP which has to transmit the service to the UE. Let  $\delta_{pk}$  be a  $|P| \cdot |K|$  vector of all zeros but the element associated with the AP  $p$  and the service type  $k$ . Then, in each state  $s$ , a request of service  $k$  can be allocated on AP  $p$  only if  $s + \delta_{pk} \in S$ ; otherwise, the request must be rejected. The action set in state  $s_j \in S$  is then defined as

$$A_{s_i} = \left\{ a = (a_{pk})_{p \in P, k \in K} \mid \begin{cases} a_{pk} \in \{0,1\} & \text{if } s_i + \delta_{pk} \in S \\ a_{pk} = 0 & \text{if } s_i + \delta_{pk} \notin S, \forall p \in P, \forall k \in K \\ \sum_{p \in P} a_{pk} \leq 1 & \forall k \in K \end{cases}, i = 1, \dots, |S|, \quad (5.1.11)$$

where  $a_{pk}$  denotes the action of admitting a request of service  $k$  on AP  $p$ . The constraint

$$\sum_{p \in P} a_{pk} \leq 1, \forall k \in K$$

in (5.1.11) states that, for each service  $k$ , the request can be either allocated to a single AP  $p$  or rejected.

For the sake of the analysis, it is assumed that, for each service type  $k \in K$ , the service requests arrive according to a Poisson distribution in time with intensity  $\nu_k$  and their duration is exponentially distributed with mean termination frequency  $\mu_k$ . Under a given policy  $u$ , when the system is in state  $s$ , the transition frequencies between states occur according to the arrival and termination frequencies and to the admission decisions:

- if the action  $u(s)$  is to allocate the service  $k$  on AP  $p$ , the transition from state  $s$  to state  $s + \delta_{pk}$  occurs with frequency  $\nu_k$ ;

- since a mapped service  $k$  on AP  $p$  terminates with termination frequency  $\mu_k$ , the transition from state  $s$  to state  $s - \delta_{ck}$  occurs with frequency  $n_{ck}(s)\mu_k$ , regardless of the policy  $u$ .

For the sake of the analysis of the MDP properties, a procedure known as *uniformization* can be applied to transform a continuous-time MDP to a discrete-time one, and it can be shown (see, e.g., [14]) that the two MDPs are equivalent. The procedure consists of defining discrete-time transitions, i.e., in defining a transition matrix  $T$ , by following a two-step procedure: in the first step, the transition frequencies of each state must be divided by constant value which is larger than the sum of the outgoing transition frequencies of any state  $s \in |S|$ ; in the second step, self-transitions are added to each state in such a way that the outgoing probability is 1. Thanks to the RL approach, the computation of the transition probabilities is not necessary.

Given a state  $s(t)$ , we are also interested in computing the bitrate which is actually used by the APs for the on-going services and in the fact that the user experience for some services can be improved if the APs assign more bitrate than the minimum required one, in order to maximize the associated reward. Therefore, an allocation procedure must be defined at every admission decision or QoS-flow termination to decide the allocation of the bitrate exceeding the minimum one  $\eta_p^1(s(t))$  to the on-going QoS-flows. In state  $s(t)$ , the bitrate allocation procedure returns  $n_{pk}^c(t)$  and  $r_{pk}^c$  (the number of QoS-flows of service  $k$  on-going on AP  $p$  with granted bitrate level  $c$  and their associated rewards), for all  $p \in P$ ,  $k \in K$  and  $c = 1, \dots, C_k$ . It is then possible to associate each state  $s(t)$  with an amount of bitrate required at the time  $t$  by all the allocated QoS-flows on AP  $p$ :

$$\eta_p(s(t)) = \sum_{k \in K} \sum_{c=1, \dots, C_k} n_{pk}^c(t) w_k^c, p \in P. \quad (5.1.12)$$

Consistent with this idea, we define a state-dependent reward  $r(s)$  in state  $s$  considering the whole capacity allocation and not only the minimum one  $\eta_p^1(s)$ :

$$r(s) = \sum_{p \in P} \sum_{k \in K} \sum_{c=1, \dots, C_k} n_{pk}^c(t) r_{pk}^c, \quad (5.1.13)$$

The described stationary MDP is ergodic and unichain (i.e., under all stationary policies, it is aperiodic and has a single recurrent class and possibly a non-empty set of transient states, see [15] for details), since the transitions are stochastic – and, therefore, aperiodic – and the transitions due to service terminations are always positive and independent of the policy. Thus, the expected state sojourn times under policy  $u$ , denoted with  $y_u^1$ 's, exist and are finite. For our controller, the interest is in maximizing the expected discounted reward (5.1.1)<sup>2</sup>, computed as

$$R(u) = (1 - \gamma) \sum_{j=1, \dots, N} y_u(s_j) r(s_j), u \in U. \quad (5.1.14)$$

#### 5.1.4 Proposed RL-based traffic steering algorithm

We note that the described MDP could be computed by means of DP algorithms or of its linear programming formulation [15], which however would require the knowledge of the transition matrix. Even if the transition matrix were known, the MDP model described in Section 5.1.3

<sup>2</sup> With some awareness, the proposed method is applicable also to the undiscounted and finite-horizon cases.

would not be tractable by standard DP methods due to scalability reasons. Therefore, this solution proposes to apply ADP techniques to reduce the problem dimension and RL to obtain a data-driven algorithm.

Concerning ADP, we reduce the state space dimension by aggregating the states with similar minimum capacity allocation. For every AP  $p \in P$ , by defining a granularity  $\Delta_p$  and using  $\lfloor \cdot \rfloor$  as the truncation operator to the next lower integer, the aggregate state set with  $\lfloor \frac{W_p}{\Delta_p} \rfloor$  bitrate levels is obtained as

$$\tilde{S} = \left\{ \tilde{s} = (l_p)_{p \in P} \mid l_p = \left\lfloor \frac{\sum_{k \in K} n_{pk} w_k^1}{\Delta_p} \right\rfloor, \text{ with } (n_{pk})_{k \in K} \text{ s. t. } \sum_{k \in K} n_{pk} w_k^1 < W_p, \forall p \in P \right\}. \quad (5.1.15)$$

Clearly, the number of states decreases as the granularities  $\Delta_p$ 's grow.

Due to the state space aggregation, also the action space (which we recall is dependent on the state) needs to be changed. Since different bitrate levels can be associated to a single aggregate state  $\tilde{s}$ , it might happen that, at two time instants  $t'$  and  $t''$ , for a given service  $c$ , the system is in state  $\tilde{s}$  with minimum load  $\eta_p(\tilde{s}(t')) w_k^1 < W_p - w_{pk}$  and  $\eta_p(\tilde{s}(t'')) w_k^1 > W_p - w_{pk}$ , respectively: only at time  $t''$ , the system could accept a new service  $k$  request. Since the action set is associated to a state, standard approaches would be either to consider the admission action for service  $c$  as not available in state  $\tilde{s}$ , regardless of the availability of the necessary capacity, or to “disaggregate” these states.

To account for these occurrences without increasing the dimension of the state space, we define a state-dependent action space which also depends on the actual measured AP transmission bitrate:

$$\tilde{A}(\tilde{s}(t)) = \left\{ (\tilde{a}_{pk})_{p \in P, k \in K} \mid \begin{array}{ll} \tilde{a}_{pk} \in \{0, 1\} & \text{if } \eta_p(\tilde{s}(t)) w_k^1 < W_p - w_{pk} \\ \tilde{a}_{pk} = 0 & \text{otherwise, } \forall k \in K \\ \sum_{p \in P} \tilde{a}_{pk} \leq 1 & \forall k \in K \end{array} \right\}. \quad (5.1.16)$$

Correspondingly, at time  $t$ , the observed reward in state  $\tilde{s}(t)$  is computed as

$$r(t) = \sum_{p \in P} \sum_{k \in K} \sum_{c=1, \dots, C_k} n_{pk}^c(t) r_{pk}^c, \quad (5.1.17)$$

The action set approximations introduced so far required modifications of the standard Q-learning algorithm, since an action might not be available at every visit of state  $\tilde{s}$ . In particular, the update rule (5.1.4) of the Q-table needs to be modified. At time  $t$ , for each state  $\tilde{s}$  and service  $k$ , let  $n_{\tilde{s}}(t)$  be the number of visits of state  $\tilde{s}$  and  $n_{\tilde{s}, \tilde{a}}(t)$  be the number of visits of state  $\tilde{s}$  when the action  $\tilde{a}_{pk}$  was available. Let  $\tilde{s}(t) = \tilde{s}'$  and  $\tilde{a}(t) = \tilde{a}'$ ; the quantity  $N_{\tilde{s}', \tilde{a}'}(t) = \frac{n_{\tilde{s}'}(t)}{n_{\tilde{s}', \tilde{a}'}(t)}$  is then used in the update rule:

$$Q(\tilde{s}', \tilde{a}') = (1 - \alpha(t))Q(\tilde{s}', \tilde{a}') + \alpha(t) \left( r(t) + \left( \sum_{n=1, \dots, \lfloor N_{\tilde{s}', \tilde{a}'} \rfloor} \gamma^n + \gamma^{(N_{\tilde{s}', \tilde{a}'} - \lfloor N_{\tilde{s}', \tilde{a}'} \rfloor)} \right) \right).$$

$$\max_{\tilde{a} \in \tilde{A}(\tilde{s}(t+1))} Q(\tilde{s}(t+1), \tilde{a}) \quad (5.1.18)$$

In this way, the actions which are less often available in a state  $\tilde{s}'$  are not penalized. An example is presented to clarify the proposed update rule. Let  $\tilde{a}'$  an action that is always available in state  $\tilde{s}'$ , i.e.,  $N_{\tilde{s}', \tilde{a}'} = 1$ . When  $\tilde{s}(t) = \tilde{s}'$  and  $\tilde{a}(t) = \tilde{a}'$ , the standard Q-learning update rule is enforced and  $Q(\tilde{s}', \tilde{a}')$  is updated using the best next-value of  $Q$  discounted by  $\gamma$ :

$$Q(\tilde{s}', \tilde{a}') \leftarrow (1 - \alpha(t))Q(\tilde{s}', \tilde{a}') + \alpha(t) \left[ r(t) + \gamma \max_{\tilde{a} \in \tilde{A}(\tilde{s}(t+1))} Q(\tilde{s}(t+1), \tilde{a}) \right].$$

Conversely, let another action  $\tilde{a}''$  be available, on the average, about half of the times the state  $\tilde{s}'$  is visited, with  $N_{\tilde{s}', \tilde{a}''} = 2.3$ . The rule (18) is such that, if  $\tilde{s}(t) = \tilde{s}'$  and  $\tilde{a}(t) = \tilde{a}''$ ,  $Q(\tilde{s}', \tilde{a}'')$  is updated using the best next-value of  $Q$  discounted by a larger value of  $\gamma$ , as if the state-action pair was visited 2.3 times:

$$Q(\tilde{s}', \tilde{a}'') \leftarrow (1 - \alpha(t))Q(\tilde{s}', \tilde{a}'') + \alpha(t) \left[ r(t) + (\gamma + \gamma^2 + \gamma^{0.3}) \max_{\tilde{a} \in \tilde{A}(\tilde{s}(t+1))} Q(\tilde{s}(t+1), \tilde{a}) \right]$$

The proposed rule avoids that action  $\tilde{a}''$  is not chosen even if available because of infrequent visits.

### 5.1.5 Simulations of the RL traffic steering controller

We assume that the three considered services are prioritized, with the constant bitrate service having the maximum priority and the elastic service having the minimum priority, and that, for each service, the service sessions have the same priority. Therefore, for a given AP  $p$ , if  $\eta_p^1(s) < W_p$ , the remaining bitrate is allocated in a fair way to the multi-coded services; if all the multi-coded service instances receive their maximum bitrate, the remaining resource is equally shared among the elastic service instances. The pseudo-code in Table 5-1 details the bitrate implemented allocation algorithm, which allocates the available resources to the services with strict priorities (e.g., all service connections  $k + 1$  are served before service connections  $k$ ,  $k = 1, \dots, K - 1$ ) and in a fair way among the connections of the same service.

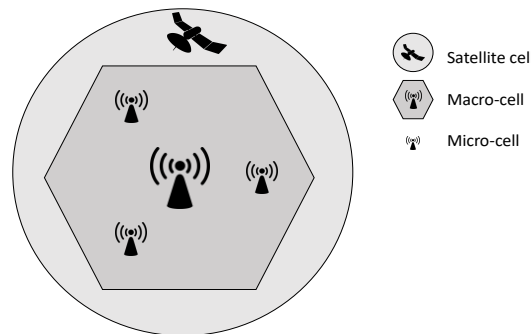


Figure 5-4: Reference scenario used in the simulations

**Table 5-1: Bitrate allocation algorithm for AP  $p \in P$  in state  $s \in S$**

<p>Assign the minimum required bitrate to all the on-going connections of AP <math>p</math>:</p> $n_{pk}^c(s) = \begin{cases} n_{pk} & \text{if } c = 1 \\ 0 & \text{otherwise} \end{cases}, k = 1, \dots, K$ <p>For <math>k = 1, \dots, K</math></p> <p>Set <math>c = 2</math></p> <p>While <math>n_{pk}^c = n_{pk}</math> and <math>c \leq C_k</math></p> <p>    Compute the bandwidth already assigned to all the connections of each service</p> $\eta_p(s) = \sum_{k \in K} \sum_{c=1, \dots, C_k} n_{pk}^c w_k^c$ <p>    Share the leftover bandwidth <math>W_p - \eta_p(s(t))</math> in a round-robin fashion among the <math>n_{pk}^{c-1}</math> multi-codec PDU sessions on-going on AP <math>p</math> (after the round, all the connections codec are upgraded only if <math>n_{pk}^c = n_{pk}</math>)</p> <p>    Evaluate <math>r_{pk}^c</math></p> <p>    Set <math>c = c + 1</math></p> <p>End</p> <p>End</p>
--

The simulations ran over a scenario of 1 hour, during which the various service types arrive according to Poisson distributions of mean values  $2s$ ,  $6s$  and  $4s$  for service 1, 2 and 3, respectively. Service dwelling time was determined, for each service type, according to exponential distributions of mean values  $30s$ ,  $120s$  and  $90s$ .

The considered RAN is reported in Figure 5-4 and is characterized by the presence of three micro-cells, one macro-cell and the availability of satellite coverage in the whole area. The UE of the connection requests was uniformly distributed in the area covered by the macro-cell; depending on the position of the UE, it can connect to either one of the micro-cell, or to a pair of micro-cells or to all three of them. It was assumed that micro-cells offer 2GBps, whereas the acro-cell and the satellite cell are limited to 1GBps.

The reward associated to the elastic services is set as

$$r_{p1}(w_{p1}) = \begin{cases} 0 & \text{if } w_{p1} < 0.01 \text{ GBps} \\ 200w_{p1} & \text{if } 0.01 \text{ GBps} \leq w_{p1} < 0.1 \text{ GBps} \\ 20 & \text{if } w_{p1} \geq 0.1 \text{ GBps otherwise} \end{cases}$$

Multi-codec services are characterized by a reward of

$$r_{p2}(w_{p2}) = \begin{cases} 0 & \text{if } w_{p2} < 0.1 \text{ GBps} \\ 6 & \text{if } 0.1 \text{ GBps} \leq w_{p2} < 0.12 \text{ GBps} \\ 12 & \text{if } 0.12 \text{ GBps} \leq w_{p2} < 0.18 \text{ GBps} \\ 25 & \text{if } w_{p2} \geq 0.18 \text{ GBps} \end{cases}$$

Finally, fixed bitrate services have the following reward:

$$r_{p3}(w_{p3}) = \begin{cases} 0 & \text{if } w_{p3} < 0.2 \text{ GBps} \\ 20 & \text{if } w_{p3} \geq 0.2 \text{ GBps} \end{cases}$$

Additionally, the rewards were scaled depending on the AP on which their corresponding service was allocated, to capture also the trade-off between user satisfaction and the operative cost incurred by the operator, which changes depending on RATs and the specific APs. The scaling factors were set to 0.75, 1, 1.25 for the micro cells, 2 for the macro-cell and 0.5 for the satellite.

The design parameters of the Q-Learning controller were set as  $\gamma = 0.9$ ,  $\varepsilon = 0.05$  and  $\alpha(t) = 1/(1 + \lfloor \tau(t)/100 \rfloor)$  where  $\tau(t)$  represents the number of connection requests processed by the system at time  $t$ .

The simulation results in terms of total cumulative reward, averaged over 100 simulation runs, showed that the proposed RL controller attains an increase in performance of around 5.3% compared to a baseline Least-Loaded (LL) balancer controller, which allocates each new connection on the currently least-loaded AP.

Figure 5-5 shows how the strategies of the two controllers (RL and LL) affect the connection blocking rates. It can be noticed that, not being forced to accept any incoming call on the least loaded APs, the RL controller manages to attain a higher reward by reserving the most profitable (in terms of reward scaling factor) network resources for the services associated with the highest rewards. From the figure, it can be observed that, using the RL controller, the rejection rate for services of type 3 halves, meaning that overall the network resources are better managed.

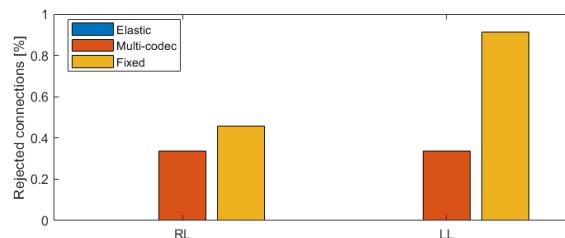


Figure 5-5: Percentage of blocked connections for the RL and the LL controllers

## 5.2 Capacity-constrained Wardrop equilibria for multi-connectivity in 5G networks

As introduced previously, the problem of *traffic steering* was given the dynamic properties that typically characterise *traffic switching*. In the scope of 5G-ALLSTAR, the traffic flow control problem consists of (i) dynamically choosing which APs shall serve each UE and (ii) dynamically deciding how much traffic relevant to each UE shall be routed through each of the serving APs. The present algorithm focuses on the downlink direction, i.e., it refers to the traffic transmitted from the core network to the UEs via the APs; nevertheless, similar considerations apply when considering the uplink direction.

In this section, the performance of the network APs is measured in terms of *latency functions* that capture the communication power required for each AP to serve the various *commodities*. In the considered 5G scenario, such commodities are mapped into the QoS-flows, which we recall are streams of data that i) share the same Quality of Service (QoS) requirements and (ii) are directed to UEs served by the same set of APs. Each latency function accounts for the transmission power required by a given AP to serve a given QoS-Flow meeting its QoS requirements. Then, the objective of the load balancing algorithm is to dynamically *steer* the downlink traffic in such a way that the values of the latency functions are equalized. This is considered a desirable situation where the QoS requirements of all the in-progress QoS-flows are satisfied in a fair way, and the degree of satisfaction of the network user (and consequently their QoE) is in an ideal condition.

The described scenario is typical in adversarial routing and load balancing problems, as the various connections are not concerned with the overall network state and aim at optimising their own, individual, performances.

The two main problems in the algorithm development are i) the fact that the latency functions are not known a priori, but can be only measured, ii) the fact that a distributed approach is needed since a centralized approach would require too much control traffic to exchange information among the potentially thousands of UEs.

A distributed, non-cooperative and dynamic load balancing algorithm is consequently developed in the context of adversarial network equilibria; specifically, the algorithm considers every single packet included in a QoS-flow as an *agent* able to make a decision regarding its routing, and specifically the AP it is assigned to. Such decisions are based on the measurements of the latency functions, obtained from the observation of the transmission power of the APs over which the commodity is routed, and are made unilaterally in an adversarial framework, with no concern for the overall system performance.

Load balancing is a particular case of routing with unitary paths. In adversarial (or selfish) routing, the control algorithms are aimed at leading the network into convenient equilibrium states. One of such states is known in mean-field game theory as Wardrop equilibrium (which can be regarded as a Nash equilibrium for infinite players [16]): in the considered scenario, if the network is in a Wardrop equilibrium, the latencies of the used RATs are equalized.

For this algorithm, we also consider the fact that the capacities of the RATs are limited, and therefore the algorithm objective is to equalize the latencies of the used RATs which are not saturated. This generalization of the Wardrop equilibrium in capacitated networks is known as Beckmann user equilibrium [17].

The main motivations behind the proposed control approach are then (i) to design a dynamic adversarial load balancing algorithm and to prove, using Lyapunov and invariance principle arguments, how the difference equation governing the global state of the system converges to an approximated Beckmann equilibrium, and (ii) to show the effectiveness of such an approach through its application to a simulated 5G network scenario.

Finally, we note that, since the algorithm is developed within the research framework of selfish routing, it can be applied to several problems and scenarios other than 5G multi-connectivity.



The rest of this section is organized as follows: Section 5.2.1 and 5.2.2 present the state-of-the-art on Wardrop load balancing and the proposed novelties for multi-connectivity; Section 5.2.3 details the algorithm and the convergence proof; Section 5.2.8 shows the simulation results; Section 5.2.1 motivates the choice of a distributed adversarial load-balancing algorithm to address the multi-connectivity problem in 5G networks, whereas Section 5.2.2 summarizes the works in the literature relevant to dynamic selfish routing and load balancing and the proposed innovations.

### 5.2.1 Multi-connectivity and traffic steering in 5G networks

The controller of this section addresses the problem of traffic steering by modelling it as a load-balancing problem.

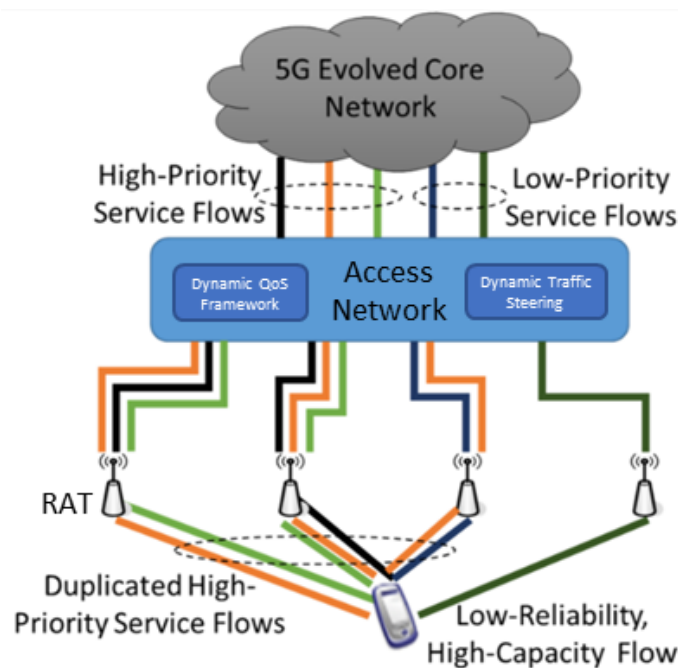


Figure 5-6: Dynamic Traffic Steering framework from [18]

The approach is hence compliant with the latest developments of the 5G architecture (see Figure 5-6), as designed by 5GPPP in [18] and already discussed in deliverable D4.1. We recall that in the scope of 5G-ALLSTAR multi-connectivity comprises the concept of *dynamic traffic steering*, which envisages the ability of dynamically steering the traffic, partitioned into QoS-flows, in real-time, among the various available APs of the RATs, based on feedbacks on the current AP performances.

In this framework, QoS-flows may be duplicated over different APs to increase their resiliency (i.e., *traffic splitting*), while other ones may be split over multiple RATs to increase their throughput or to better meet their QoS requirements.

As already introduced in D4.1, in the 5G architecture the traffic steering problem is solved in three different ways: (i) with a user-centric approach, where each UE decides its connection preferences according to local measures of some performance indicator; (ii) in a Radio Access Network (RAN)-assisted fashion, in which the decision is still made by the UEs but the RAN provides them with additional information on the network state; (iii) with a RAN-controlled approach, where all decisions are made by the RAN, which is a centralised unit by nature, or delegated to the distributed control units that govern the single APs.

Several works study the problem of multi-connectivity in the heterogeneous network framework proposed by 5G, from both architectural [19], [20] and algorithmic [3], [21], [22] points of view. Multi-connectivity enables the problem of optimally steering the network traffic over the available



APs, in such a way that the QoS requirements of the various QoS-flows are met [1], [6]. Common solutions utilise the concept of utility and latency functions, as in this work, to capture the network performances [3], [23], [24]. Several works in the literature also employ game-theoretic approaches for the AP selection, typically in adversarial frameworks, as [3], [9],[8] leading the networks to Nash equilibrium states.

Regarding game-theoretic solutions, one possible modelling choice is to have an adversarial game between the users, as in [9], [25] that envisage a setup similar to the one used in this section. In such scenarios, the users compete to attain the best connection quality while eventually also minimising their costs. An alternative approach is to set up a game between the various network operators, each controlling a set of APs as in [24],[8] and focusing on their economic performances.

The algorithm proposed here utilises *differential game theory*, a branch of game theory that studies dynamical systems, and shares some of the characteristics of the previously mentioned works, as the adversarial nature of its equilibrium. The control algorithm designed will be proven to drive the communication network state to a convenient equilibrium state (the Beckmann equilibrium), and this convergence will be attained by following an explicit discrete-time control law, with no need for round-games or price/cost bidding auctions. Contrary to optimisation-based works, the proposed control law is also suitable to steer the traffic flows in real-time, and, being a distributed decision process, it does not require any significant control traffic overhead.

The previous aspects, together with the explicit inclusion of constraints on the available transmission capacity, makes the proposed approach a suitable candidate for the deployment in 5G scenarios, in which the low latency and power consumption aspects are of crucial importance. With reference to the mentioned dynamic traffic steering framework [18], the algorithm can be implemented in RAN-assisted and in the RAN-controlled configurations: in the former case, the algorithm would run in the UEs based on the information received by the RAN; in the latter case, the algorithm would run directly in the RAN or in the distributed controllers of the various APs.

### 5.2.2 Adversarial load balancing in 5G networks and Beckmann equilibria

The problem of optimally distributing flow is one of the most fundamental, and yet challenging, aspects of any network operation. In the framework of selfish routing, the network flow is formed by a stream of infinitely-many decision-making agent [26] that competes for attaining the best connection performance, without consideration for the congestion, and consequent performance degradation, that their decisions cause to the other agents.

Wardrop equilibria [27] were then introduced, as they describe a network state in which no single agent can unilaterally improve its performances (e.g., travel time, as in the original Wardrop formulation). Being an adversarial kind of equilibria, the overall network performance is not optimised and the performance loss is referred to as *price of anarchy* in the literature [28]. The concept of Wardrop equilibrium has been extended to various families of networks, among which the capacitated ones [17], [29]–[31], and problems, as the load balancing one [32]–[34]. Even if Wardrop equilibria can be computed by centralized algorithms in polynomial time [35], for the low connection latency promised by 5G – and the consequent agile and fast traffic steering requirements – distributed approaches are more suitable, motivating the objective of developing a dynamic algorithm that converges to a Wardrop equilibria.

Based on simple difference equations derived from the flow conservations laws, the algorithm of this section proposes a load balancing solution over the nodes of a dynamical network that represents the 5G infrastructure [36], [37], consisting in the connections between several APs and their users with the core network. In doing so, the algorithm takes into account that the amount of traffic each AP can support is limited, or *capacitated*, due to transmission power constraints and, in general, resource scarcity. This limitation implies that the user equilibrium to which the network will converge may not be in principle the Wardrop equilibrium [31], which is defined for unconstrained networks. Several works [17], [29]–[31] extended the original formulation of the Wardrop user equilibrium, which corresponds to a situation in which all the latencies

of each commodity are equalised, to deal with capacitated networks. The resulting *Beckman user equilibrium* works as such that the latencies of all the unsaturated APs of each commodity are equalised. Differently from [17], [29]–[31], we propose a dynamic algorithm which will be proven to converge to a Beckmann equilibrium.

Regarding dynamic load balancing solutions for Wardrop equilibria in the literature, several works utilise the concepts of *learning* and *exploration* to cope with the limited feedback information that the decision-making agents have access to. To attain a better knowledge of the system state and dynamics, the agents *sample* different flow distribution strategies and then *exploit* the learned system characteristics to converge to optimal states. The authors of [38] present an asynchronous and distributed algorithm that employs reinforcement learning to update transmission probabilities, based on an estimation of the network edges latencies. In [39], an iterative and distributed learning solution is proven to converge to a Wardrop equilibrium state using Lyapunov arguments, as in the following.

An important contribution has been given by Fischer et al. in [40]–[42]. In [40] and [42], a round-based algorithm is developed to solve a game among the various commodities, aimed at redistributing the traffic flow and reaching an approximated Wardrop equilibrium. In [41], a similar set up is analysed assuming that the information available to the agents may be stale. In [43], a dynamic discrete-time load-balancing algorithm is presented in the context of Virtual Private Networks (VPNs), which converges to an approximate Wardrop equilibrium.

The present work extends the results from the literature mainly in two directions:

- i) the convergence properties of the algorithm are studied in the multi-commodity case, a requirement for application in the 5G framework, that was not explicitly discussed in the cited works;
- ii) the algorithm analysis and design are extended to the case of capacitated networks, enabling the application of the solution to more realistic case studies in several domains.

### 5.2.3 Proposed Wardrop load balancing algorithm

Section 5.2.4. models the 5G traffic steering problem as a load balancing one; its first subsection describes the basic definitions needed for the algorithm analysis, whereas the second subsection contains some remarks regarding Lyapunov stability; the third and final subsection presents the load balancing algorithm and the convergence proof.

### 5.2.4 5G traffic steering as a dynamic load-balancing problem

In the dynamic multi-connectivity framework of 5G networks [18], each UE selects the serving APs for its QoS-flows. When the traffic of multiple UEs is characterized by the same QoS requirements and can be served by the same set of APs, such traffic is often considered as a single aggregated QoS-flow to simplify the network control. Consistently with 5G specifications, we consider these aggregated QoS-flows to be the commodities that flow over the 5G infrastructure.

The various APs are modelled as resource (capacity) providers, among which the load balancing algorithm will distribute the QoS-flows. Each AP is characterised by a different latency function for each of the commodities it may serve. The latency functions are assumed to be representative of the transmission power required for serving the QoS-flows.

Each commodity  $i$  in the set of commodities  $\mathcal{I}$  is assumed to be characterised by a constant bitrate  $\lambda^i$ , a reasonable assumption for aggregated QoS-flows over a limited time window, and the state variable of the system at time  $k$ ,  $x_p^i[k]$ , represents the bitrate of commodity  $i$  assigned to the AP  $p$  in the set  $\mathcal{P}^i$  of the APs available to commodity  $i$ . The amount of bitrate that an AP  $p$  can sustain is limited by its maximum transmission power, causing the network to be capacity

constrained. Additionally, in the framework of *network slicing*, in which third-party tenants manage a set of network resources, such constraints could also be imposed by the network operator as representative of the amount of resources that characterise the network slice.

Regarding the latency functions, in the problem formulation we consider the fact that the commodities are generally characterised by different latency functions since each QoS-flow is characterized by specific QoS requirements and since each AP has its specific characteristics (e.g., depending on antenna gains, frequency constraints, ...). As introduced above, the selected quantity to be captured by the latency functions in the proposed scenario is the AP transmission power, as it is the main limiting factor of the downlink cell capacity, i.e. of the downlink bit rate which can be handled by the AP serving the cell. Moreover, AP transmission power also accounts for the Bit Error Rate (BER), which is a fundamental QoS requirement, since, for a given bitrate, QoS-flows requiring a lower BER need higher AP transmission power (as further specified in simulation section). At a Beckmann user equilibrium, each of the commodities will have its traffic distributed in such a way that the amount of power required to transmit each of its packets over the unconstrained APs is equalised.

### 5.2.5 Preliminaries on Wardrop and Beckmann equilibria

As anticipated in Section 5.1.1, this section further develops a well-known model for selfish routing [40], where an infinite population of agents carries an infinitesimal amount of load each and builds on the previous work [43] concerning distributed load balancing algorithms. The proposed control scheme relies on common assumptions on the latency functions.

*Assumption 1.* The latency functions  $l_p^i(\xi)$  are positive, non-decreasing and Lipschitz continuous, for  $\xi \in [0, c_p]$ , where  $c_p$  is the capacity of provider  $p$ , for all  $p \in \mathcal{P}$  and for all  $i \in \mathcal{I}$ . Furthermore, the maximum Lipschitz constant of all the  $l_p^i$ 's is denoted as  $\bar{\beta} = \max_{p \in \mathcal{P}, i \in \mathcal{I}} \beta_p^i$ .

The assumption is not restrictive in real use-cases since the provider performances decrease with their load.

In non-capacitated algorithms, the set of feasible states is defined as

$$\mathcal{X} = \left\{ \mathbf{x} = (x_p)_{p \in \mathcal{P}} \mid x_p = \sum_{i \in \mathcal{I}} x_p^i, x_p^i \geq 0, \forall p \in \mathcal{P}, \sum_{p \in \mathcal{P}^i} x_p^i = \lambda^i, \forall i \in \mathcal{I} \right\}, \quad (5.2.1)$$

A flow  $\mathbf{x} \in \mathcal{X}$  is at a Wardrop equilibrium if, for each commodity, the latencies of the loaded providers are equalized, i.e., if  $l_p^i(\mathbf{x}) \leq l_q^i(\mathbf{x})$  for all  $p \in \mathcal{P}^i$  such that  $x_p^i > 0$ , for all  $q \in \mathcal{P}^i$  and for all  $i \in \mathcal{I}$ .

By defining the Beckmann-McGuire-Winsten potential

$$\Phi(\mathbf{x}) = \sum_{i \in \mathcal{I}} \sum_{p \in \mathcal{P}^i} \int_0^{x_p^i} l_p^i(\xi) d\xi, \quad (5.2.2)$$

the Wardrop equilibria are the solutions of the optimization problem

$$\min_{\mathbf{x} \in \mathcal{X}} \Phi(\mathbf{x}). \quad (5.2.3)$$

Capacity-constrained networks are characterized by the additional capacity constraints

$$x_p \leq c_p, \forall p \in \mathcal{P}. \quad (5.2.4)$$

A flow  $x \in \mathcal{X}$  is feasible if constraints (5.2.4) hold, and the set of feasible states is defined as

$$\mathcal{X}_{CP} = \{x \in \mathcal{X} \mid x_p \leq c_p, \forall p \in \mathcal{P}\}. \quad (5.2.5)$$

Considering commodity  $i \in \mathcal{I}$  under a flow  $x \in \mathcal{X}_{CP}$ , provider  $p \in \mathcal{P}^i$  is defined as *capacity-constrained or saturated* if  $x_p = c_p$ .

A flow  $x \in \mathcal{X}_{CP}$  is at a Beckmann user equilibrium if, for each commodity, the latencies of the loaded and unconstrained providers are equalized, i.e., more precisely:

*Definition 1* ([17]). A flow  $x \in \mathcal{X}_{CP}$  is at a Beckmann user equilibrium if  $l_p^i(x) \leq l_q^i(x)$  for all  $p \in \mathcal{P}^i$  such that  $x_p^i > 0$ , for all  $q \in \mathcal{P}^i$  such that  $x_q < c_q$  and for all  $i \in \mathcal{I}$ .

The set of equilibria is then

$$\mathcal{X}_{eq} = \{x \in \mathcal{X}_{CP} \mid l_p^i(x) \leq l_q^i(x), \forall p \in \mathcal{P}^i \text{ s.t. } x_p^i > 0, \forall q \in \mathcal{P}^i \text{ s.t. } x_q < c_q, \forall i \in \mathcal{I}\}. \quad (5.2.6)$$

Let us consider the minimization problem (5.2.3) with constraints (5.2.4), hereinafter referred to capacity-constrained problem (CP). The Beckman user equilibria [30] are the optimal solutions of the CP.

Property 1 [17]. If the set of feasible solutions  $\mathcal{X}_{CP}$  of the CP is nonempty, the optimization problem consists in minimizing a convex function over a nonempty polytope and, thus, the set of optimal flows  $\mathcal{X}_{eq}$  is nonempty and convex.

### 5.2.6 Preliminaries on Lyapunov stability

The algorithm convergence proof relies on the LaSalle invariance principle for discrete-time nonlinear systems ([44], [45]).

*Definition 2.*  $\mathcal{L}: \mathcal{X} \rightarrow \mathbb{R}$  is a candidate Lyapunov function for a discrete-time nonlinear system  $x[k+1] = f(x[k])$  if

- i)  $\mathcal{L} \in \mathcal{C}^1$  and  $\mathcal{L}(x)$  is bounded below and positive definite on  $\mathcal{X}$ ;
- ii) If  $x_{eq} \in \mathcal{X}_{eq}$ , where  $\mathcal{X}_{eq}$  is the set of equilibrium points,  $\mathcal{L}(x_{eq}) = 0$  and  $\mathcal{L}(x) > 0$  if  $x \notin \mathcal{X}_{eq}$ ;
- iii) Along forward trajectories,  $\mathcal{L}$  satisfies

$$\Delta\mathcal{L}(x[k]) := \mathcal{L}(f(x[k])) - \mathcal{L}(x[k]) \leq 0, k = 0, 1, 2, \dots$$

*Theorem 1* ([44]). Let  $\mathcal{L}(x)$  be a candidate Lyapunov function for the discrete-time nonlinear system  $x[k + 1] = f(x[k])$ . Then, any bounded trajectory tends to the largest invariant subset  $\mathcal{M}$  contained in the set of points defined by  $\Delta\mathcal{L}(x) = 0$ .

### 5.2.7 Capacitated load balancing algorithm and convergence proof

For each commodity  $i \in \mathcal{I}$ , the control action consists in the decision, at time  $k$ , of *migrating* part of the flow mapped onto a given provider  $p$  to another provider  $q$ , with  $p, q \in \mathcal{P}^i$ . By denoting the rate of such migration with  $r_{pq}^i[k]$ , the system dynamics is written as

$$x[k + 1] = f(x[k]), k = 0, 1, 2, \dots \quad (5.2.7)$$

with

$$x_p[k] = \sum_{i \in \mathcal{I}} x_p^i[k], \quad (5.2.8)$$

$$x_p^i[k + 1] = x_p^i[k] + \tau \sum_{q \in \mathcal{P}^i} (r_{qp}^i[k] - r_{pq}^i[k]), \quad (5.2.9)$$

and with feasible initial conditions

$$x[0] \in \mathcal{X}_{CP}. \quad (5.2.10)$$

for all  $p, q \in \mathcal{P}^i$  and  $i \in \mathcal{I}$

The proposed controller builds on the dynamic algorithm in [43], which expresses the migration rate as

$$r_{pq}^i[k] = x_p^i[k] \sigma^i \mu_{pq}^i[k], \quad (5.2.11)$$

where  $\sigma^i$  is a positive migration gain and  $\mu_{pq}^i[k]$  is the migration policy, representing the decision of whether (if it is positive) or not (if it is equal to zero) migrate some flow from provider  $p$  to provider  $q$ .

As in [43] for the Wardrop equilibria, approximated Beckmann user equilibria are defined.

*Definition 3.* The set of  $\varepsilon$ -Beckmann user equilibria is defined as

$$\mathcal{X}_{eq}^\varepsilon = \left\{ \mathbf{x} \in \mathcal{X}_{CP} \mid l_p^i(\mathbf{x}) \leq l_q^i(\mathbf{x}) + \varepsilon, \forall p \in \mathcal{P}^i \text{ s.t. } x_p^i > 0, \forall q \in \mathcal{P}^i \text{ s.t. } x_q \leq c_q - \frac{\varepsilon}{2\bar{\beta}}, \forall i \in \mathcal{J} \right\}. \quad (5.2.12)$$

where  $\varepsilon \geq 0$  represents a maximum tolerated latency mismatch.

*Remark 1.* The defined sets are such that  $\mathcal{X}_{eq}^\varepsilon \xrightarrow{\varepsilon \rightarrow 0} \mathcal{X}_{eq}$  and  $\mathcal{X}_{eq} \subseteq \mathcal{X}_{eq}^\varepsilon \subseteq \mathcal{X}_{CP}$ : the objective of the controller is then, starting from a physically admissible state in  $\mathcal{X}_{CP}$ , to reach an approximated equilibrium state in  $\mathcal{X}_{eq}^\varepsilon$ , whose degree of approximation with respect to the equilibrium state in  $\mathcal{X}_{eq}$  reduces with  $\varepsilon$ .

The tolerance  $\varepsilon$  is introduced since the kind of migration rates of equation (5.2.11) cannot guarantee convergence in the discrete-time case, however small the sampling period [41]. A flow  $\mathbf{x} \in \mathcal{X}_{CP}$  is then at  $\varepsilon$ -Beckman equilibrium if, for each commodity, the latencies of the loaded and  $\varepsilon$ -unconstrained providers are equalized, where we define a provider  $p \in \mathcal{P}^i$  to be  $\varepsilon$ -unconstrained if  $x_p < c_p - \frac{\varepsilon}{2\bar{\beta}}$ .

In the proposed algorithm, the migration decision is defined as

$$\mu_{pq}^i[k] = \begin{cases} 0 & \text{if } l_p^i(\mathbf{x}[k]) - l_q^i(\mathbf{x}[k]) \geq \varepsilon \text{ or if } x_q[k] \geq c_q - \frac{\varepsilon}{2\bar{\beta}}. \\ 1 & \text{otherwise} \end{cases} \quad (5.2.13)$$

The controlled system dynamics, hereafter denoted as load-balancing (5.2.20) dynamics, is then expressed by equations (5.2.9), (5.2.11), (5.2.13), with control gains set as

$$\sigma^i = \frac{\varepsilon}{2\tau\bar{\beta}\lambda^i|\mathcal{P}^i-1||\mathcal{J}|}, \quad (5.2.14)$$

and with the tolerance set as

$$0 < \varepsilon \leq \min_{i \in \mathcal{J}} \bar{\beta}\lambda^i|\mathcal{J}|. \quad (5.2.15)$$

*Remark 2.* The approximated capacity-constrained user equilibria are such that, for each commodity, the latencies of the loaded and  $\varepsilon$ -unconstrained providers are equalized within the tolerance  $\varepsilon$ . Then, for a given equilibrium flow  $\mathbf{x} \in \mathcal{X}_{eq}^\varepsilon$  and for each commodity  $i \in \mathcal{J}$ , three classes of providers exist: the unloaded providers  $p \in \mathcal{P}^i$  such that  $x_p^i = 0$ ; the  $\varepsilon$ -constrained providers  $p \in \mathcal{P}^i$  such that  $x_p > c_p - \frac{\varepsilon}{2\bar{\beta}}$ ; the  $\varepsilon$ -unconstrained providers, whose latencies are equalized.

The convergence property of the algorithm relies on the following 3 lemmata.

*Lemma 1.* Under Assumption 1, considering the LB dynamics, the latency variation of a provider  $p \in \mathcal{P}^i$  in one time-step is bounded by

$$|l_p^i(f(\mathbf{x}[k])) - l_p^i(\mathbf{x}[k])| \leq \frac{\varepsilon}{2}. \quad (5.2.16)$$

*Proof.* Considering the generic commodity  $i \in \mathcal{J}$ , provider  $p \in \mathcal{P}^i$  and time  $k$ , the maximum latency decrease occurs when no commodities migrate their populations from the other providers to provider  $p$ :

$$\begin{aligned} & l_p^i(x_p[k+1]) \\ &= l_p^i(x_p[k] + \tau \sum_{q \in \mathcal{P}} (r_{qp}[k] - r_{pq}[k])) \\ &\geq l_p^i(x_p[k] - \tau \sum_{p \in \mathcal{P}} r_{pq}[k]). \end{aligned} \quad (5.2.17)$$

Since  $\beta_p^i$  is the Lipschitz constant of the function  $l_p^i(\cdot)$  between 0 and  $c_p$ , it follows that

$$l_p^i(\mathbf{x}[k+1]) \geq l_p^i(\mathbf{x}[k]) - \tau \beta_p^i \sum_{q \in \mathcal{P}} r_{pq}[k]. \quad (5.2.18)$$

Considering Equations (5.2.11) and (5.2.14), the last term of Equation (5.2.18) is written as

$$\begin{aligned} \tau \beta_p^i \sum_{q \in \mathcal{P}} r_{pq}[k] &= \tau \beta_p^i \sum_{j \in \mathcal{J}} \sum_{q \in \mathcal{P}^j} r_{pq}^j[k] \\ &= \tau \beta_p^i \sum_{j \in \mathcal{J}} \sum_{q \in \mathcal{P}^j} x_p^j[k] \sigma^j \mu_{pq}^j[k] = \\ &= \sum_{j \in \mathcal{J}} \tau \beta_p^i x_p^j[k] \sigma^j \sum_{q \in \mathcal{P}^j} \mu_{pq}^j[k] = \\ &= \sum_{j \in \mathcal{J}} \tau \beta_p^i x_p^j[k] \frac{\varepsilon}{2\tau\beta\lambda^j|\mathcal{P}^j-1||\mathcal{J}|} \sum_{q \in \mathcal{P}^j} \mu_{pq}^j[k] = \\ &\leq \sum_{j \in \mathcal{J}} \frac{\varepsilon}{2^{|\mathcal{J}|}} = \frac{\varepsilon}{2}, \end{aligned} \quad (5.2.19)$$

where the inequality holds since  $x_p^j[k] \leq \lambda^i$ ,  $\beta_p^i \leq \bar{\beta}$  and since, recalling Equation (5.2.13), there are at most  $|\mathcal{P}^j - 1|$  terms equal to 1 in  $\sum_{q \in \mathcal{P}^j} \mu_{pq}^j[k]$ . It follows that

$$l_p^i(\mathbf{x}[k+1]) \geq l_p^i(\mathbf{x}[k]) - \frac{\varepsilon}{2}. \quad (5.2.20)$$

Similarly, the maximum latency increase occurs when no commodities migrate their populations from provider  $p$  to other providers:



$$l_p^i(\mathbf{x}[k+1]) \leq l_p^i(\mathbf{x}[k]) + \tau \beta_p^i \sum_{q \in \mathcal{P}} r_{qp}[k], \quad (5.2.21)$$

which yields

$$l_p^i(\mathbf{x}[k+1]) \leq l_p^i(\mathbf{x}[k]) + \frac{\varepsilon}{2}. \quad (5.2.22)$$

■

*Lemma 2.*  $\mathcal{X}_{CP}$  is a positively invariant set for the LB dynamics.

*Proof.* We need to show that, for all  $k \geq 0$ , for all  $p \in \mathcal{P}^i$  and for all  $i \in \mathcal{J}$ , i)  $\sum_{p \in \mathcal{P}^i} x_p^i[k] = \lambda^i$ , ii)  $x_p^i[k] \geq 0$ , iii)  $x_p[k] \leq c_p$ .

Considering that  $x[0] \in \mathcal{X}_{CP}$ , equations (5.2.9), (5.2.11) and (5.2.8) yield that the population remains constant, since

$$\begin{aligned} x_p^i[k+1] - x_p^i[k] &= \sum_{p \in \mathcal{P}^i} \sum_{q \in \mathcal{P}^i} (r_{qp}^i[k] - r_{pq}^i[k]) \\ &= \sum_{p \in \mathcal{P}^i} \sum_{q \in \mathcal{P}^i} r_{qp}^i[k] - \sum_{q \in \mathcal{P}^i} \sum_{p \in \mathcal{P}^i} r_{pq}^i[k] = 0, \end{aligned} \quad (5.2.23)$$

and thus that  $\sum_{p \in \mathcal{P}^i} x_p^i[k] = \sum_{p \in \mathcal{P}^i} x_p^i[0] = \lambda^i, \forall k \geq 0$ .

Given that  $x_p^i[0] \geq 0$ , it is proven below by induction that  $x_p^i[k] \geq 0, \forall k \geq 0$ . Assuming that  $x_p^i[k] \geq 0$ , for a given  $k$ , it is sufficient to prove that

$$x_p^i[k+1] = x_p^i[k] + \tau \sum_{q \in \mathcal{P}^i} (r_{qp}^i[k] - r_{pq}^i[k]) \geq 0, \forall p \in \mathcal{P}^i. \quad (5.2.24)$$

If  $x_p^i[k] = 0$ , it follows that  $r_{pq}^i[k] = 0$  and thus Equation (5.2.24) yields  $x_p^i[k+1] \geq 0$ .

If  $x_p^i[k] > 0$ , from Equation (5.2.11) it follows that  $r_{pq}^i[k] \geq 0$ . Thus, the following inequality holds (in the worst case, no providers migrate part of their population to a provider  $p$ ):

$$x_p^i[k+1] \geq x_p^i[k] - \tau \sum_{q \in \mathcal{P}^i} r_{pq}^i[k]. \quad (5.2.25)$$

A sufficient condition for inequality (5.2.31) to hold is then

$$x_p^i[k] - \tau \sum_{q \in \mathcal{P}^i} r_{pq}^i[k] \geq 0. \quad (5.2.26)$$



Recalling Equations (5.2.11) and (5.2.13), Eq. (5.2.26) is written as

$$\begin{aligned} x_p^i[k] - \tau \sum_{q \in \mathcal{P}^i} r_{pq}^i[k] &= x_p^i[k] - \tau \sum_{q \in \mathcal{P}^i} x_p^i[k] \sigma^i \mu_{pq}^i[k] = x_p^i[k] (1 - \tau \sigma^i \sum_{q \in \mathcal{P}^i} \mu_{pq}^i[k]) \geq \\ &\geq x_p^i[k] (1 - \tau \sigma^i (|\mathcal{P}^i| - 1)), \end{aligned} \quad (5.2.27)$$

where the inequality holds since the summation has at most  $(|\mathcal{P}^i| - 1)$  terms equal to 1. In the case  $x_p^i[k] > 0$ , Equations (5.2.14) and (5.2.15) are sufficient for Equation (5.2.27) to be non-negative;

Given that  $x_p[0] \leq c_p$ , it is proven below by induction that  $x_p[k] \leq c_p, \forall k \geq 0$ . Assuming that  $x_p[k] \leq c_p$ , for a given  $k$ , it is sufficient to prove that

$$x_p[k+1] = x_p[k] + \tau \sum_{i \in \mathcal{J}} \sum_{q \in \mathcal{P}^i} (r_{qp}^i[k] - r_{pq}^i[k]) \leq c_p, \forall p \in \mathcal{P}^i. \quad (5.2.28)$$

If  $x_p[k] \geq c_p - \frac{\varepsilon}{2\beta}$  equation (5.2.13) entails that  $r_{qp}^i[k] = 0$  for all  $q \in \mathcal{P}^i$  and  $i \in \mathcal{J}$  and, thus, from equation (5.2.9), that  $x_p[k+1] \leq x_p[k]$ .

Otherwise, if  $x_p[k] < c_p - \frac{\varepsilon}{2\beta}$ , we consider that

$$\begin{aligned} x_p[k+1] &\leq x_p[k] + \tau \sum_{i \in \mathcal{J}} \sum_{q \in \mathcal{P}^i} r_{qp}^i[k] = \\ &= x_p[k] + \tau \sum_{i \in \mathcal{J}} x^i[k] \sigma^i \sum_{q \in \mathcal{P}^i} \mu_{qp}^i[k] = \\ &\leq x_p[k] + \sum_{i \in \mathcal{J}} \frac{\varepsilon}{2\beta|\mathcal{J}|} = x_p[k] + \frac{\varepsilon}{2\beta} \end{aligned} \quad (5.2.29)$$

■

*Lemma 3.* The function

$$\mathcal{L}(\mathbf{x}) := \Phi(\mathbf{x}) - \Phi_{min}. \quad (5.2.30)$$

is a candidate Lyapunov function for the LB dynamics.

*Proof.* The function  $\mathcal{L}(\mathbf{x})$  is positive definite in  $\mathcal{X}_{CP}$  by definition, since  $\Phi_{min}$  is the minimum value of  $\Phi(\mathbf{x})$  for all the minimizers of the CP.

Let  $\Delta\mathcal{L}(\mathbf{x}[k])$  denote the difference of the Lyapunov function  $\mathcal{L}(\mathbf{x})$  along the solutions of the controlled system:

$$\begin{aligned}
\Delta\mathcal{L}(\mathbf{x}[k]) &= \mathcal{L}(\mathbf{x}[k+1]) - \mathcal{L}(\mathbf{x}[k]) \\
&= \sum_{i \in \mathcal{J}} \sum_{p \in \mathcal{P}} \int_{x_p[k]}^{x_p[k+1]} l_p^i(\xi) d\xi \\
&\leq \sum_{i \in \mathcal{J}} \sum_{p \in \mathcal{P}} (x_p^i[k+1] - x_p^i[k]) l_p^i(x_p^i[k+1]) \\
&= \sum_{i \in \mathcal{J}} \sum_{p \in \mathcal{P}} \tau (\sum_{q \in \mathcal{P}} r_{qp}^i[k] - \sum_{q \in \mathcal{P}} r_{pq}^i[k]) l_p^i(x[k+1]) \\
&= \tau \sum_{i \in \mathcal{J}} \sum_{p \in \mathcal{P}} \sum_{q \in \mathcal{P}} r_{pq}^i[k] (l_p^i(x[k+1]) - l_q^i(x[k+1])).
\end{aligned} \tag{5.2.31}$$

where the inequality holds from geometric considerations: If  $x_p[k+1] > x_p[k]$ , recalling that the  $l_p^i$ 's are nondecreasing functions, the definite integral  $\int_{x_p[k]}^{x_p[k+1]} l_p^i(\xi) d\xi$  is smaller than the quantity  $(x_p[k+1] - x_p[k]) l_p^i(x_p[k+1])$ ; conversely, if  $x_p[k+1] < x_p[k]$ , the integral  $\int_{x_p[k+1]}^{x_p[k]} l_p^i(\xi) d\xi$  is larger than the quantity  $(x_p[k] - x_p[k+1]) l_p^i(x_p[k+1])$ .

Analysing each term of the inner summation, two cases hold: if  $r_{qp}^i(t) = 0$  the term is null, otherwise, if  $r_{qp}^i(t) > 0$ , the term is negative. In fact, it is shown below that, if  $r_{pq}^i[k] > 0$ , it holds that  $l_p^i(x[k+1]) - l_q^i(x[k+1]) > 0$ .

Lemma 1 states that

$$\begin{aligned}
&l_p^i(x[k+1]) - l_q^i(x[k+1]) \\
&\geq \left( l_p^i(x[k]) - \frac{\varepsilon}{2} \right) - \left( l_q^i(x[k]) \mp \frac{\varepsilon}{2} \right) = l_p^i(x[k]) - l_q^i(x[k]) - \varepsilon > 0,
\end{aligned} \tag{5.2.32}$$

where the inequality holds since a necessary condition for  $r_{pq}^i[k] > 0$  is that  $l_p^i(x[k]) - l_q^i(x[k]) > \varepsilon$  (see equation (5.2.13)). ■

Finally, the following theorem proves the convergence towards an approximated Beckmann user equilibrium.

**Theorem 1** The trajectories of the LB dynamics asymptotically tend to the set of equilibria  $\mathcal{X}_{eq}^\varepsilon$ .

*Proof.* Given that Lemma 2 states that  $\mathcal{L}(\mathbf{x})$  is a candidate Lyapunov function for the LB dynamics, the proof relies on the LaSalle invariance principle of Theorem 1, i.e., on showing that  $\mathcal{X}_{eq}^\varepsilon$  is the maximum invariant set where  $\Delta\mathcal{L} = 0$ .

Let  $\mathbf{x} \in \mathcal{X}_{eq}^\varepsilon$  and  $\mathbf{x}[0] = \mathbf{x}$ . By comparing definition (5.2.6) and equation (5.2.13), it holds that  $r_{pq}^i[k] = 0$  for all  $p, q \in \mathcal{P}^i$  and  $i \in \mathcal{J}$ , which entails i) that  $\mathbf{x}[k] = \mathbf{x}[0] = \mathbf{x}_{eq} \in \mathcal{X}_{eq}^\varepsilon$  for all  $k > 0$ , i.e., that  $\mathcal{X}_{eq}^\varepsilon$  is a positively invariant set, and ii) that  $\Delta\mathcal{L}(\mathbf{x}[k]) = 0$  in  $\mathcal{X}_{eq}^\varepsilon$  (see equation (5.2.30)).

To show that  $\mathcal{X}_{eq}^\varepsilon$  is the maximum set where  $\Delta\mathcal{L}(\mathbf{x}[k]) = 0$ , it is proven below that  $\Delta\mathcal{L}(\mathbf{x}[k]) < 0$  if  $\mathbf{x}[k] = \mathbf{x}$ , with  $\mathbf{x} \notin \mathcal{X}_{eq}^\varepsilon$ . In fact, by definition (5.2.12), in this case there exist at least one pair of providers  $p, q \in \mathcal{P}^i$  and a commodity  $i \in \mathcal{J}$  such that  $l_p^i(x[k]) - l_q^i(x[k]) > \varepsilon$ , with  $x_p^i[k] > 0$  and  $x_q[k] < c_q - \frac{\varepsilon}{2\beta}$ , which, in turn, yields  $r_{pq}^i[k] > 0$  (see equations (5.2.11), (5.2.14) and

(5.2.13)). Having established that  $r_{pq}^i[k] > 0$  with  $l_p^i(x(t)) - l_q^i(x(t)) > \varepsilon$ , it follows that the corresponding term of the inner summation of equation (5.2.30) is negative, which is a sufficient condition for  $\Delta\mathcal{L}(x[k]) < 0$  (recalling that, in the proof of Lemma 2, it is shown that the terms of equation (5.2.30) are nonpositive). ■

### 5.2.8 Numerical simulation

This section reports the simulation setup and results in sections 5.2.9 and 5.2.10, respectively.

#### 5.2.9 Simulation setup

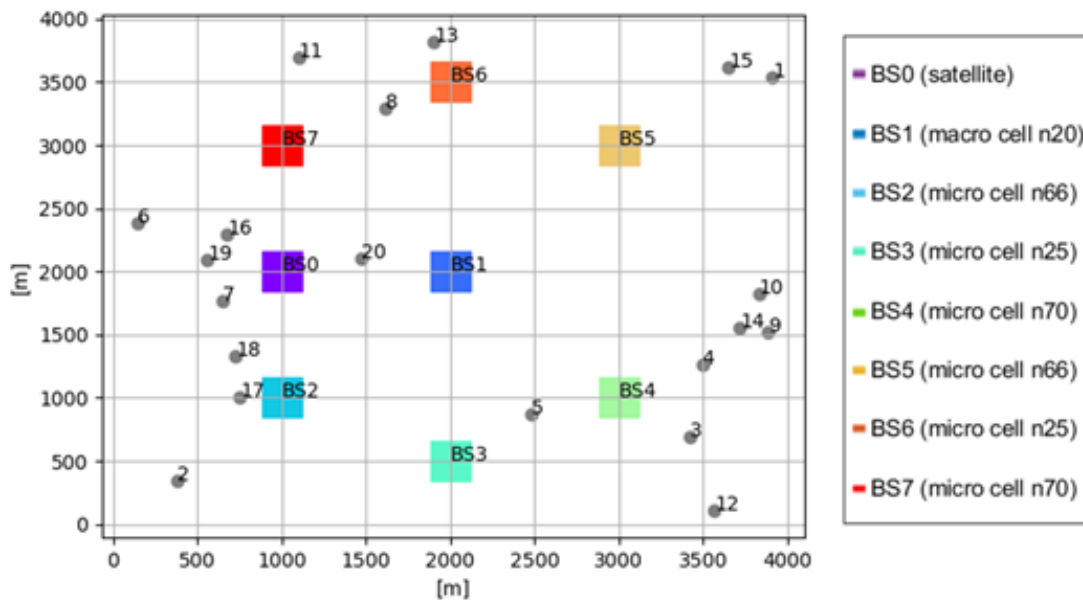


Figure 5-7 Network scenario

For the validation of the proposed algorithm, in the second issue of D4.2 we report the results of a simulation conducted on the open-source network simulator of the project (available in [46]), able to model different AP technologies, connection protocols and interference models in a multi-connectivity scenario. We consider the network depicted in Figure 5-7, consisting of a  $4 \times 4$  Km area covered by a macro cell (provider BS1), a satellite (provider BS0) and six micro cells (BS2-BS7).

Table 5-2 Characteristics of micro and macro cells

Operating band N°	Carrier frequency (GHz)	Bandwidth (MHz)
n20	0.8	20
n25	1.9	40
n66	1.7	40
n70	2	25

A total of 20 UEs/commodities (grey dots in Figure 5-7) were randomly distributed in the area, each requiring a constant load  $\lambda^i = 50 \text{ Mbps}$ . The implemented interference model is taken from [47] and the frequency characteristics of the terrestrial APs are summarised in Table 5-2 [48], [49].

Regarding the satellite AP, we considered a Time Division Multiplexing (TDM) as in the example 6.6.2 of [50]. The satellite parameters are adapted in order to have at least 1bit per symbol with typical SNR values [51], [52]. According to the TDM frame structure used, it is possible to allocate only blocks of 64 symbols ( $1 \mu\text{s}$ ). Moreover, each allocation must consider a header made of 288 symbols and a spacing between allocations of 64 symbols. Additional implementation details and updates can be found in [46].

We considered as latency functions  $l_p^i$  the number of resource blocks utilised by the commodity  $i$  on the access point  $p$ . This particular choice will drive the network towards a state in which each connection equalises the resource block usage over its available unsaturated APs  $p \in \mathcal{P}^i$ .

Note that several different choices could be made for the latency function, spacing from quantities that capture connection reliability, to transmission delay and user satisfaction, as the only requirements that such functions must satisfy are represented by Assumption 1, which open the possibility of considering a large family of functions (e.g., including polynomial or exponential ones). We mention that in non-ideal cases, the value of  $\bar{\beta}$  in (5.2.14) may need to be estimated by observing the growth of the selected latency functions during different network operative conditions.

To allow a fair comparison with the terrestrial AP resource blocks, the assumptions made for the satellite imply that its latency function is equal to the number of its allocated symbols divided by 64. Additionally, each AP was associated to a multiplicative scaling factor for their latency functions to model different operating costs. In particular, the satellite was given the highest factor (0.5), the macro cell was given a medium value (0.2) and the lowest weight was associated to micro cells (0.1). Regarding the capacitated nature of the considered network, we assume that the network operator dedicated to the controlled slice  $200 \text{ Mbps}$  on all micro cells, save for BS4 that was capacitated at  $55 \text{ Mbps}$ .

Concerning the parameters of the controller, the choice of latency functions leads to an experimentally determined value  $\bar{\beta} = 2.44$ , the latency tolerance is selected as  $\varepsilon = 0.5$  and the sampling time as  $\tau = 10^{-3} \text{ s}$ . The resulting values for  $\sigma^i$  are in the range  $[0.02, 0.05]$ .

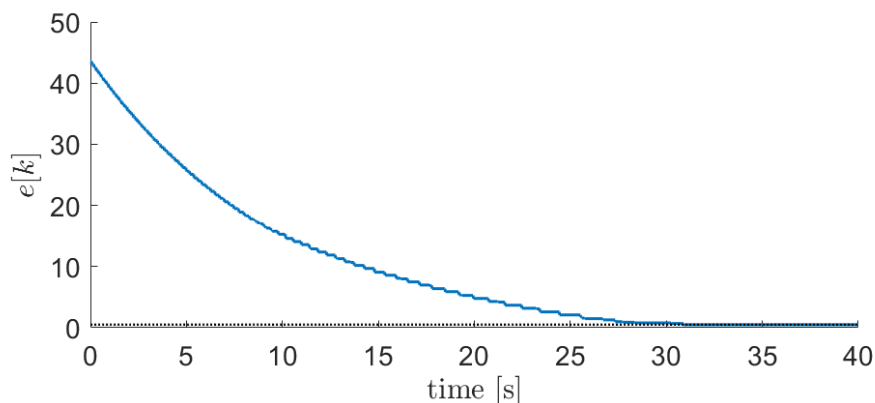


Figure 5-8 Maximum latency mismatch during the simulation (dotted line: tolerance  $\varepsilon$ ).

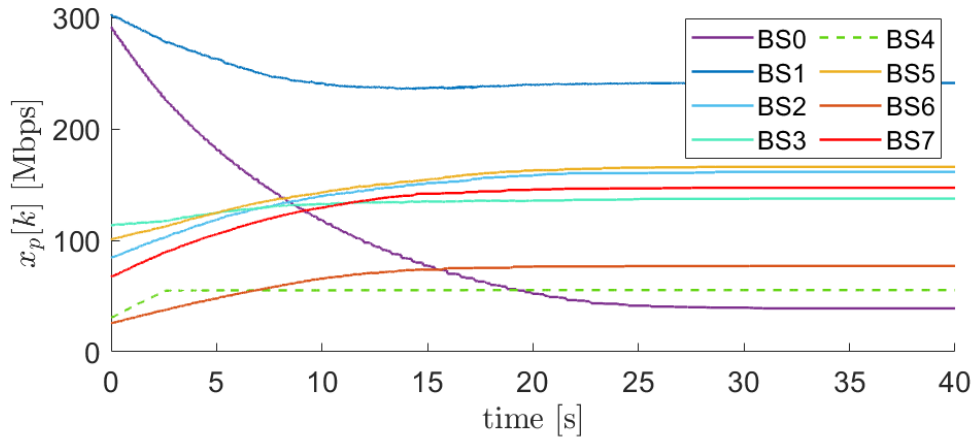


Figure 5-9 Network state in terms of total bitrate allocated on the various APs (solid lines: unconstrained providers, dashed line: constrained provider).

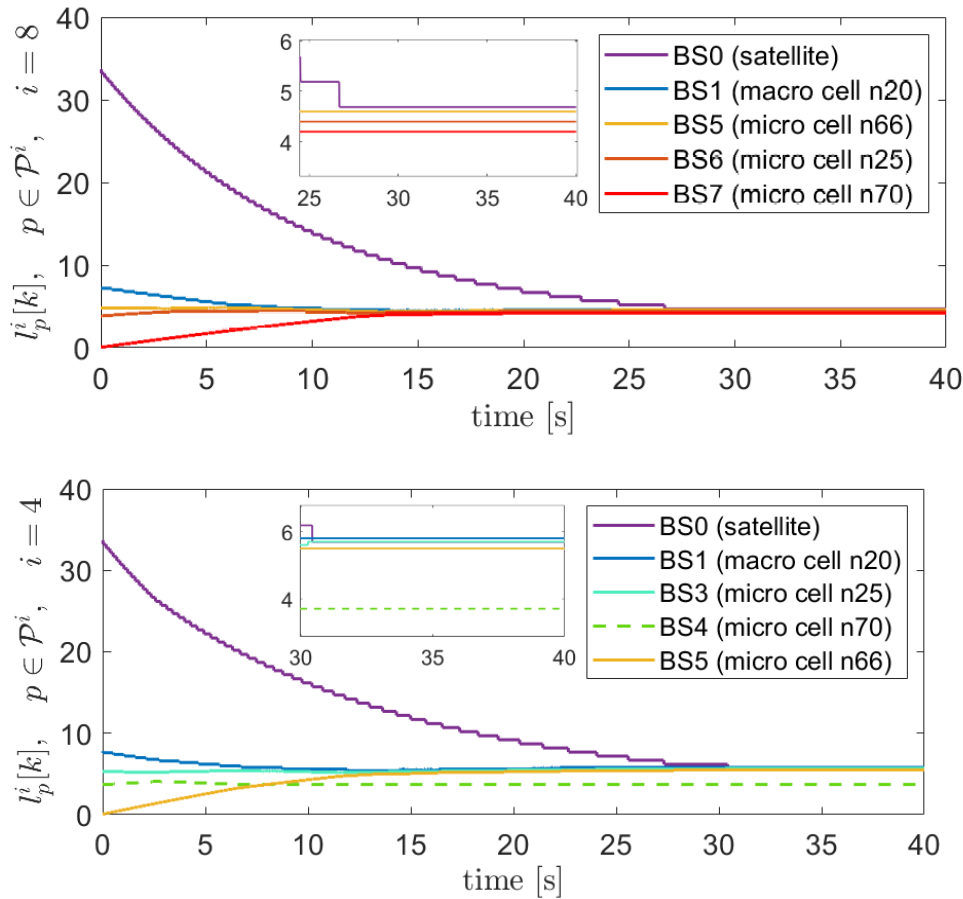


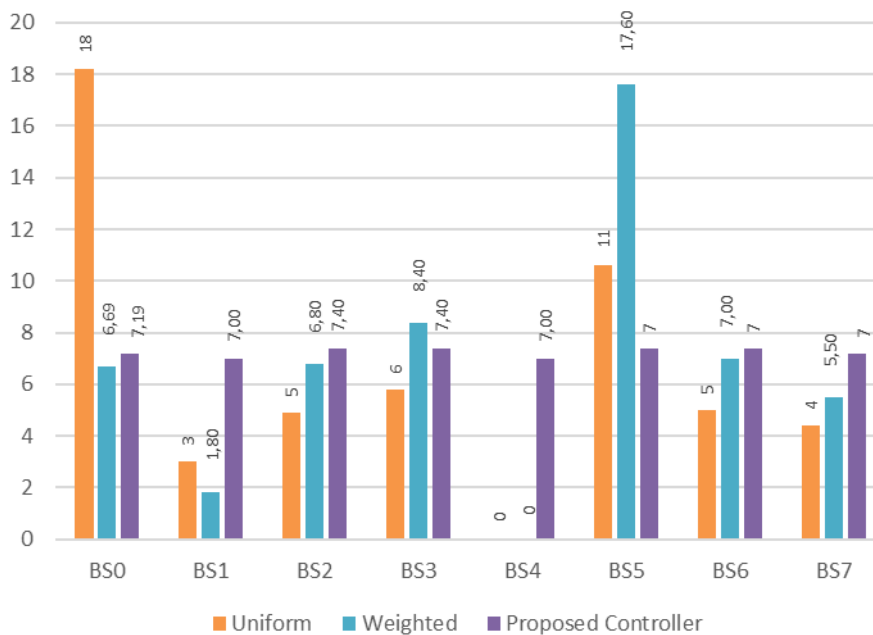
Figure 5-10: Commodity latency examples during the simulation (solid lines: unconstrained providers used by the commodity; dashed line: constrained providers; zoomed sub-plots to show the convergence within the tolerance  $\epsilon = 0.5$ ).

### 5.2.10 Simulation results

Simulation runs were initialized by distributing uniformly the load of the commodities over  $|\mathcal{P}^i| - 1$  of their available APs, selected randomly. This trivial load distribution provides a challenging

scenario to test the algorithm, as the latencies of the various AP will start with very different values, as they are neglected for the initial load distribution. Nevertheless, as we will see, the convergence times of the algorithm are acceptable and can be reduced by at least one order of magnitude with more intelligent starting allocations. In the PoC of the project, if this algorithm is selected for implementation, we will deploy a better initial load distribution as the purpose of the demonstrator will be the provision of a resilient and efficient multi-connectivity solution, whereas in this simulation we aim at stressing and demonstrating the algorithm mathematical properties in a realistic environment.

The reported simulations showed a convergence time to an  $\varepsilon$ -Beckmann equilibrium in the order of 30s, averaged over 25 runs. It is worth remarking that such convergence time is not related to the 5G QoS requirements, as it is assumed that the various access points are able to provide the proper QoS level (e.g., connection latency, average BER, reliability level,...) if their capacities are not violated.



**Figure 5-11 Comparison for the three considered algorithms of the latencies over APs for commodity  $i = 20$ .**

Figure 5-8 shows, for an example run, how the maximum latency mismatch over all the commodities, defined as

$$e[k] = \max_{i \in J} \left\{ \max_{p \in \mathcal{P}^i | x_p^i[k] > 0} l_p(x_p[k]) - \min_{q \in \mathcal{P}^i | x_q^i[k] < c_q - \frac{\varepsilon}{2\beta}} l_q(x_q[k]) \right\},$$

decreases with time and, even if the initial conditions are quite unbalanced, with  $e[k] > 40$ , after 30s  $e[k]$  is already below the threshold  $\varepsilon$ .

For the example run, Figure 5-10 reports the evolution of the latencies that characterise the commodities 4 and 8, for all of their available APs. The upper plot shows that the latencies of the APs available to QoS-Flow 8 converge to a common value, as expected, within the threshold  $\varepsilon$ ; in particular, we can notice how the commodity rapidly starts using the (initially unused) micro cell BS7 and rapidly discharges the satellite. The lower plot of Figure 5-10 shows the

latencies of the QoS-Flow 4 and highlights that the latency of micro-cell BS4 does not converge to the latencies of the other used APs: the reason is that the AP becomes  $\varepsilon$ -saturated after about 3s (see Figure 5-9) – thus, by definition, the population of QoS-Flow 4 still converges to an  $\varepsilon$ -Beckmann equilibrium. Note that the latency associated to BS4 starts higher than its final value, as the commodity migrates towards BS5, but remains the lowest latency for the commodity 4 from 10s onwards, as the other QoS flows already  $\varepsilon$ -saturated BS4 (i.e., no bitrate can be migrated to it).

Finally, Figure 5-9 shows the population dynamics over the APs, highlighting how the macro cell is the most utilised AP, while all the micro cells allocate a similar amount of bitrate. The satellite, whose latency was the most penalised as it is the most costly connection technology, is rapidly discharged.

For the sake of comparison, in Figure 5-11 we benchmark the proposed controller against two classic examples of load balancing solutions in heterogeneous networks. The figure reports the latency functions values experienced by the commodity  $i = 20$  over the eight APs in the set  $\mathcal{P}^{20}$ . The choice of the commodity 20 was related to the fact that it is one of the closest to the centre of the considered area, as depicted in Figure 5-7, and can hence be served by any AP. The first benchmarking algorithm (“uniform”) uniformly distributes the bitrate demand over the various available and unsaturated APs. Due to the capacitated nature of the network, we assume that this controller distributes the load of the commodities according to their index  $i$ , so that the traffic of commodity  $i = 1$  is the first one to be allocated while the one of commodity  $i = 20$  is the last.

The second algorithm (“weighted”) distributes the bitrate considering the scaling factors associated to the latencies of the APs (0.5 for the satellite, 0.2 for the macro cell and 0.1 for micro cells), so that for every unit of traffic allocated on the satellite 5 are allocated on the micro cells and 2.5 on the macro cell.

From the analysis of the figure, one can note that the proposed controller – in the figure, the values are the ones achieved after convergence ( $\sim 30$ s) – successfully equalises the latencies up to the threshold  $\varepsilon = 0.5$ . On the contrary, the other two controllers fail to allocate any bitrate on BS4, as it was already saturated by the other commodities that were prioritised. The uniform distribution causes the first controller to experience a very high latency on the satellite (BS0), while the distance and consequent low signal-to-noise-ratio causes the weighted controller to allocate too much bitrate on BS5 (this behaviour is further amplified by the fact that BS5 is a micro cell associated to a scaling factor of 0.1), requiring a significant amount of resource blocks.

Overall, we can conclude that the proposed controller better balances the usage of network resources because it is a feedback-based solution that steers the traffic flow based on online measurements of the latency functions. The main limitation of the proposed approach is related to the availability of the measurements needed to compute the steering decisions (i.e., the latency values in terms of assigned resource blocks), whose impact on the control traffic overhead is to be evaluated considering the control traffic already necessary for the different access technologies, and the estimation of  $\bar{\beta}$  which, however, can be performed starting from the channel models and the expected traffic that the network is designed to support. Regarding the complexity of the algorithm, the computation overhead is negligible since the control law (5.2.11) only involves basic operations (summations, multiplications and comparison between real numbers) that remain limited in number even for RANs with a high number of APs.

To conclude, we mention that the two benchmarking algorithms discussed above could be used to initialise the network resource allocation, speeding up the convergence time, and we will focus on identifying a more suitable initial allocation strategy for the project’s PoC.



### 5.3 Network selection based on Markov games and friend-or-foe RL

The solution presented in this section can be classified as a control strategy of the RAN-controlled category, characterised by the distribution of the control logic over the various RATs controllers that regulate the APs connection admittance logics, so that the network resources available for the connection are optimally exploited.

From a methodological point of view, several approaches were investigated in the literature for the network selection problem, ranging from solutions based on Multiple Attribute Decision Making (MADM) [5], [6], to fuzzy logic control systems [7], and game theory-based approaches [8], [9].

Additionally, Markov Decision Processes (MDPs) and Reinforcement Learning (RL) were tested, among the others, in [10] and [11].

The proposed approach utilizes results from both RL and game theory in a multi-agent framework. The problem will be modelled in such a way that the distributed RAT controllers will compete among each other for being selected to serve the connection requests. The overall goal of the control strategy will be the optimal usage of network resources, without relying on centralized control approaches – as, for example, with a common least loaded allocation logic which assigns the upcoming connections to the RAT with the lowest resource usage. In this regard, the present algorithm employs the so-called “friend-or-foe Q-learning” algorithm to govern the network according to an adversarial Nash strategy.

#### 5.3.1 Markov games and Nash equilibria

A Markov game among  $N$  players is defined as the tuple [53]  $\langle S, \mathcal{A}, T, \mathcal{R}, \gamma \rangle$ , where:

$S$  is the finite state space.

$\mathcal{A} = \{A_i, i = 1, \dots, N\}$  is the collection of the action sets available to the various players  $i$ .

$T(s, a_1, a_2, \dots, a_N, s') : S \times A_1 \times A_2 \times \dots \times A_N \times S \rightarrow \mathbb{R}$  is the state transition function, which describes the transition probability between the two states  $s$  and  $s'$  when the agents take the actions  $a_1, a_2, \dots, a_N$ .

$\mathcal{R} = \{R_i(s, a_1, a_2, \dots, a_N) : S \times A_1 \times A_2 \times \dots \times A_N \rightarrow \mathbb{R}, A_i = \{a_i\}\}$  is the collection of reward functions that attribute a reward to each agent when they take actions  $a_1, a_2, \dots, a_N$  and the system is in state  $s$ .

$0 \leq \gamma < 1$  is the discount factor that captures the trade-off between short-term and long-term performances sought by the agents.

In the present approach we consider the so-called *general sum* games, meaning that no assumption is made on the cumulative reward attained by the agents, contrary to zero-sum games.

A policy  $\pi_i(s) : S \rightarrow \mathbb{R}^{\#(A_i)}$  is a function that maps the state of the system into a probability distribution over the actions of player  $i$ . Each player is associated with a (state,action)-value function  $Q_i$  [54], [55], defined as

$$Q_i(s, a_1, a_2, \dots, a_N) = R_i(s, a_1, a_2, \dots, a_N) + \gamma \sum_{s'} T(s, a_1, a_2, \dots, a_N, s') Q_i(s', \pi_1, \pi_2, \dots, \pi_N), \quad (5.3.1)$$

in which  $Q_i(s', \pi_1, \pi_2, \dots, \pi_N)$  is the weighted sum of the  $Q_i$ 's according to the policies  $\pi_i$ 's. For their definition, the (state, action)-value functions represent the expected discounted reward attained over time by the players starting from state  $s$ , taking actions  $a_1, a_2, \dots, a_N$  and following the policies  $\pi_1, \pi_2, \dots, \pi_N$  from there on. The goal of the controllers that will determine the policy of each agent is the one of maximizing its own value function *unilaterally* (i.e., without cooperation).



An important concept to introduce in the framework of Markov games is the one of adversarial Nash equilibria, which are a set of policies  $\pi_1, \pi_2, \dots, \pi_i', \dots, \pi_N$  characterized by the following two properties [55]:

- no player can improve its policy unilaterally, i.e.,

$$R_i(s, \pi_1, \pi_2, \dots, \pi_i, \dots, \pi_N) \geq R_i(s, \pi_1, \pi_2, \dots, \pi_i', \dots, \pi_N);$$

- no player sees its reward lowered by a change in the policies of the other players, i.e.,

$$R_i(s, \pi_1, \pi_2, \dots, \pi_i, \dots, \pi_N) \leq R_i(s, \pi_1', \pi_2', \dots, \pi_i, \dots, \pi_N').$$

### 5.3.2 Multi-agent reinforcement learning

In scenarios in which the agents are not provided with a complete and accurate model of the system, model-free control solutions as Reinforcement Learning (RL) [54] have to be implemented to attain the desired system behaviour.

The attractiveness of RL in multi-agent scenarios is due to the fact that it allows the agent  $i$  behaviour, described by its policy  $\pi_i$ , to adapt to the strategy employed by the other agents. This capability becomes of crucial importance when the various agents compete one against the other, as each agent has no incentive to share information regarding its own configuration with the others. Nevertheless, RL also allows the agent to learn about the environment characteristics by directly interacting with it, meaning that no explicit knowledge of the functions  $T$  and  $R_j, j = 1, \dots, N$  is assumed or necessary for reaching an optimal control strategy.

In this section, the friend-or-foe Q-learning algorithm from [55] is employed in its adversarial variant. The additional degree of information that agent  $i$  requires other than the feedback observation of the tuple  $\langle s, a_1, \dots, a_N, s', r_i \rangle$  is the classification of the other players as either friends (cooperating agents that try to maximise their rewards jointly) or foes (competing agents that try to maximise their own rewards unilaterally and, consequently, to minimise player  $i$ 's reward).

In this algorithm, each agent  $i$  learns its (state, action)-value function  $Q_i$  according to the following rule:

$$Q_i(s, a_1, a_2, \dots, a_N) = (1 - \alpha(t))Q_i(s, a_1, a_2, \dots, a_N) + \alpha(t)(r_i + \gamma \text{Nash}_i(s, Q_1, Q_2, \dots, Q_N)), \quad (5.3.2)$$

where  $\text{Nash}_i(s, Q_1, Q_2, \dots, Q_N)$  is computed as ([55])

$$\text{Nash}_i(s, Q_1, Q_2, \dots, Q_N) = \max_{\pi \in \Pi(A_1 \times \dots \times A_k)} \min_{[a_k, \dots, a_N] \in (A_{k+1} \times \dots \times A_N)} \sum_{[a_k, \dots, a_N] \in (A_{k+1} \times \dots \times A_N)} \pi(a_1) \dots \pi(a_k) Q_i(s, a_1, \dots, a_N). \quad (5.3.3)$$

In (5.3.3), it is assumed, without loss of generality, that the players  $1, \dots, k$  cooperate with agent  $i$  and players  $k + 1, \dots, N$  are its foes. The sequence  $0 \leq \alpha(t) < 1$  represents the evolution, over time, of the *learning rate* of the agents. Under the hypothesis that that  $\sum_t \alpha(t) = +\infty$  and  $\sum_t \alpha(t)^2 < +\infty$  [54], Theorem 6 in [55] proves that foe Q-learning (i.e., in the case in which all agents are foes) converges to an adversarial equilibrium, provided that such an equilibrium exists.

### 5.3.3 Modelling network selection as a Markov game

As already introduced, the network selection problem will be modelled as a Markov game in which each AP is a competing player.

### 5.3.4 State space

The state of the network can be represented by the percentage of occupied resources on each of the APs. In order to have a finite number of states, a possible solution is to quantize the percentage of resources with a factor  $q$ . The set of states is then defined as:

$$S = \left\{ [s_1, s_2, \dots, s_n], s_i = nq, 0 \leq s_i \leq 1, n \in \{0, \dots, 1\} \cdot \frac{100}{q} \right\},$$

meaning that there are  $(q + 1)^n$  different states.

### 5.3.5 Action space

The actions available to each of the agents regard the decision of whether to accept or decline the allocation of the incoming connection. Assuming that  $m$  different service classes are available to network users, a total of  $2^m$  actions are required to model all the possible different choices. Note that some actions might be unavailable since the APs could decide to accept only services of certain classes.

The action set of user  $i$  is then defined as:

$$A_i = \left\{ [a^1, a^2, \dots, a^m], a^j \in \{0, 1\} \right\}.$$

### 5.3.6 Reward functions

The reward that is given to each agent  $i$  for successfully allocating a service of class  $j$  depends on the service characteristics and on the amount of resources involved in the allocation.

Assuming that the service requires  $t_j$  resources, it is possible to model the reward as

$$r_i = \alpha_{ij} t_j \frac{B_i + t_j}{C_i},$$

where  $B_i$  and  $C_i$  represent the amount of resources occupied on AP  $i$  before the new allocation and the total capacity of the AP  $i$ , respectively. The factor  $\alpha_{ij}$  serves the purpose of prioritizing certain services over other ones and/or modelling the fact that some APs are more appropriate, in terms of Quality of Service, for certain services.

The structure of the reward allows incentivizing the agent to allocate all of their resources, while also dedicating them to the most prioritized services.

When a new service request arrives at the agents, each of them selects its action and consequently takes the decision of being available for the allocation or not. One agent is sampled randomly from the list of available ones, and the allocation procedure continues. The agents that offered their availability to allocate the incoming service but were not selected for the actual allocation receive a small negative reward to disincentivize the behaviour of always offering the allocation availability. Furthermore, an agent that offered the allocation but was not able to fulfil it due to a scarcity of resources is given a highly negative reward to penalize its behaviour and the connection is discarded.

To avoid that all agents reject the less rewarding services, a negative reward is also given to all the agents if no agent offers its availability for the new allocation.

### 5.3.7 $\epsilon$ -greedy policy selection

A fundamental concept in RL is the trade-off between knowledge exploitation and environment exploration. The update of the  $Q_i$  tables (5.3.2) and the solution of the *maximin* problem (5.3.3) represent, respectively, the process of learning from experience, or knowledge acquiring, and its exploitation to derive a proper strategy for the player. To provide the players with an adequate

degree of exploration, the action selection is subject to the following rule, known as  $\varepsilon$ -greedy selection:

$$a_i = \begin{cases} \operatorname{argmax}_{a_i}(Nash_i), & \text{with probability } 1 - \varepsilon \\ \text{randomly chosen in the set } A_i, & \text{with probability } \varepsilon \end{cases} \quad (5.3.4)$$

where  $Nash_i$  is the operator described in (5.3.3), in the case in which all the APs are assumed to compete one with each other and, hence, there is no friend player that cooperates with the agent  $i$ . As suggested in [56], a possible refinement to (5.3.4) is to consider a decreasing sequence of values for  $\varepsilon$ , modelling the fact that the agent benefits more from the exploration process at the beginning, while knowledge exploitation becomes more effective as the agent experienced the system evolution and its possible states several times.

### 5.3.8 Maximin linear programming formulation

In general, a *maximin* optimization problem takes the following form [57]:

$$\begin{aligned} \max_{j=1, \dots, n} \min_{x_j} J(x_j) &= c_j x_j, \\ \text{s. t.} \quad A_{eq} x &= g, \\ A_{ub} x &\leq b, \end{aligned}$$

where  $c_j \geq 0$  are scalars,  $A_{eq}$ ,  $A_{ub}$  are matrices and  $g$  and  $b$  vectors of appropriate dimensions.

It is well known that such a formulation is equivalent to the following LP problem [57]:

$$\begin{aligned} \max_{z \in \mathbb{R}} J(x_j) &= c_j x_j, \\ \text{s. t.} \quad A_{eq} x &= g, \\ A_{ub} x &\leq b, \\ z &\leq c_j x_j, j = 1, \dots, n, \end{aligned}$$

where  $z$  is an unknown scalar that is bounded by the smallest value  $c_j x_j$  by means of the additional third constraint.

In the context of Foe-Q-Learning, the *maximin* problem that appears in (5.3.3) becomes

$$\begin{aligned} \max_{\pi \in \Pi(A_1 \times \dots \times A_k)} \min_{[a_k, \dots, a_N] \in (A_{k+1} \times \dots \times A_N)} \\ \sum_{[a_1, \dots, a_k] \in A_1 \times \dots \times A_k} \pi(a_1) \dots \pi(a_k) Q_i[s, a_1, \dots, a_N] \\ \text{s. t.} \quad \pi(a_i) \geq 0 \quad \forall a_i \in A_j, j = 1, \dots, k \\ \sum_{i=1}^k \pi(a_i) = 1, \end{aligned}$$

leading to an equivalent LP formulation of the form:

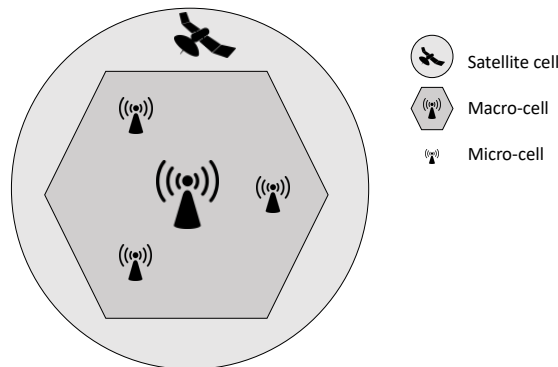
$$\begin{aligned} \max_{\pi \in \Pi(A_1 \times \dots \times A_k)} z \\ \text{s. t.} \\ h_i = \sum_{[a_1, \dots, a_k] \in A_1 \times \dots \times A_k} \pi(a_1) \dots \pi(a_k) Q_i[s, a_1, \dots, a_N], \\ \quad \forall a_k, \dots, a_N \in (A_{k+1} \times \dots \times A_N), \\ \pi(a_i) \geq 0 \quad \forall a_i \in A_1, \dots, A_k \\ \sum_{i=1}^k \pi(a_i) = 1 \\ z \leq h_i \quad \forall [a_k, \dots, a_N] \in (A_{k+1} \times \dots \times A_N). \end{aligned}$$

The following table reports the pseudo-code of the Network Selection Algorithm.

**Table 5-3: Algorithm pseudo-code**

<p>Initialize <math>Q_i(s, a), i = 1, \dots, n</math>, arbitrarily.</p> <p>For each connection request (episode) do:</p> <ul style="list-style-type: none"> <li>• Each player observes the state of the system and the class <math>j</math> of the upcoming connection</li> <li>• Each player selects its action <math>a_i</math> with an <math>\epsilon</math>-greedy policy based on their <math>Nash_i</math> function</li> <li>• The connection is allocated using the resources of one of the players that selected an action with <math>a^j = 1</math>. If no player was available for the allocation one is selected randomly.</li> <li>• All players receive a reward <math>r_i</math> as described in section 5.3.6.</li> <li>• All players update their <math>Q_i</math> table according to eq. (5.3.2).</li> </ul> <p>end</p>
--

### 5.3.9 Simulations



**Figure 5-12: Connection area covered by 3 different RATs**

The scenario considered in the simulation is shown in Figure 5-12, and consists of an area covered by three different RATs. The number of agents considered is then  $N = 3$ . The resource considered for the connection is throughput, and each RAT had a maximum capacity of 1000 Mbps. Two service classes were modelled, the first characterised by a resource request of  $t_1 = 1$  Mbps and the latter by  $t_2 = 5$  Mbps. The parameters  $\alpha_{ij}$  were set differently for each simulation.

A total of 2000 user requests were generated, where each request had a 0.8 probability of being a new connection and 0.2 of being the end of a connection, with a consequent resource deallocation. The connection requests were uniformly distributed over the two service classes.

Regarding the RL-based controller parameters,  $\alpha(t)$  was set as  $\alpha(t) = 1/(1 + \lfloor t/10 \rfloor)$ , where  $\lfloor \cdot \rfloor$  represents the lower-integer operator, and  $\epsilon(t)$  halved every 100 iterations starting from  $\epsilon(0) = 0.6$ . Finally, the discount factor  $\gamma$  was set to 0.9.

The simulative scenario considered is a simplified one, but maintains the dimensioning and the key characteristics of the test cases that are envisaged to be developed in the scope of the 5G-ALLSTAR project.

### 5.3.10 Simulation one – baseline least loaded controller

The baseline controller considered as a benchmark follows a least-loaded AP logic, as it assigns the upcoming connections to the APs with the lowest relative resource usage. Such a controller is centralised by nature, as it requires a complete knowledge of the state of the system.

### 5.3.11 Simulation two – no service prioritisation

In this simulation it was assumed that  $\alpha_{ij} = 1 \forall i, j$ , meaning that the reward of the agents depends only on the amount of allocated resources and no priority was given to any of the two service classes.

### 5.3.12 Simulation three – different rewards

The controllers trained in this simulation received a reward for the allocations characterised by  $\alpha_{i,1} = 2$  and  $\alpha_{i,2} = 0.2, \forall i$ . This choice makes the per-bps reward higher for the first class of service.

### 5.3.13 Simulation results

From the analysis of Figure 5-13 and Figure 5-14, it is possible to see how the two RL agents behave differently from the centralised least loaded controller. In particular, the controller of simulation 2 tends to uniformly accept the two services, in line with the fact that they were characterised by the same amount of reward per-Mbps, while the second RL controller (simulation three) favours the allocation of services of the first class. Overall, both the least loaded controller and the controller of simulation 2 blocked a total of 409 Mbps, meaning that the first RL solution fully exploits its available resources. The second RL controller, on the contrary, allocates a slightly lower amount throughput, blocking a total of 463 Mbps. This different behaviour is due to the fact that the agents obtain, for the same amount of resources, a different pay-off depending on the service class. In fact, due to the choice of the parameters  $\alpha_{ij}$ , the services of the first class provide ten times the amount of reward per Mbps with respect to the other. Even if the second RL controller blocked more Mbps, this translated in an improvement in performances, measured in terms of its cumulative total reward, of approximately 10%.

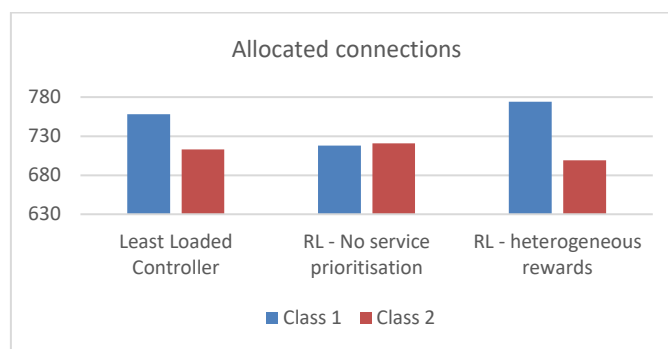
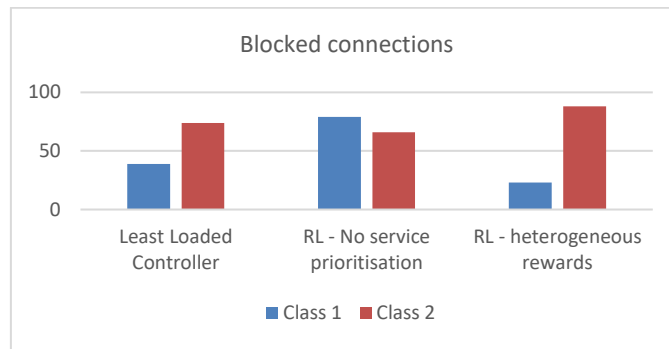
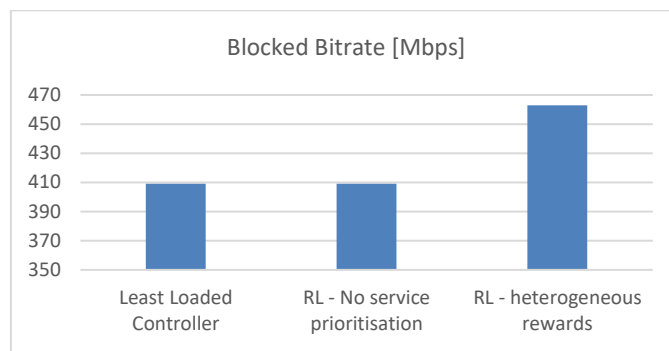


Figure 5-13: Number of allocated connections for the three controllers, divided by service class



**Figure 5-14: Number of blocked connection for the three controllers, divided by service class**



**Figure 5-15: Total amount of bitrate of the blocked connections**

### 5.4 Maximal load balancing algorithm and preliminary simulation results

The proposed algorithm is a load balancing algorithm that offloads traffic from overloaded RANs depending on the data traffic type. Load measurement of the terrestrial RAN is calculated based on Radio Resource Usage Ratio (RRUR) from the Physical Resource Blocks (PRBs). For a given time,  $T$ , the RRUR of a terrestrial RAN is calculated as:

$$\beta = \frac{1}{T \cdot w} \sum_{\tau \in (t-T, t)} (\rho \times \delta)_{\tau}$$

where  $w$  is total bandwidth of terrestrial RAN,  $\rho$  is the number of allocated PRBs and  $\delta$  is sub-carrier spacing transmitted at time  $t$ .

The proposed algorithm for load balancing runs in each gNB. Based on bandwidth usage ratio,  $\beta$ , the algorithm determines whether the gNB is overloaded or not by comparing it with a bandwidth usage threshold. Once the cell is overloaded, gNB sorted the UEs in increasing order of RSRP values based on the measurement reports. After that, the algorithm takes the first user from the list and triggers a traffic flow classification algorithm to classify delay tolerant and delay sensitive flows of the use based on the packet header. If a user has a delay tolerant flow, the algorithm offloads the delay tolerant flow to satellite links and the NTN delivers the data to that user. The algorithm continually checks the bandwidth usage ratio of the cell to determine whether it is still overloaded or not. The algorithm offloads the delay tolerant flows to satellite links until the gNB becomes unburdened. Figure 5-16 shows the flow diagram of the proposed algorithm. The proposed algorithm offload the UEs to satellite based on terrestrial RAN status such that each UE receives its required data rate.

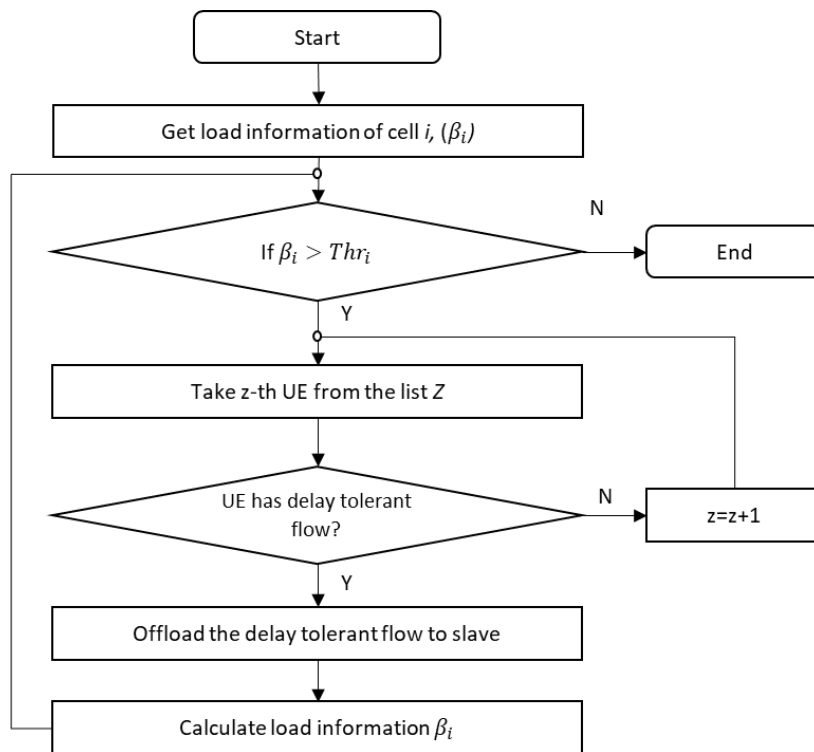
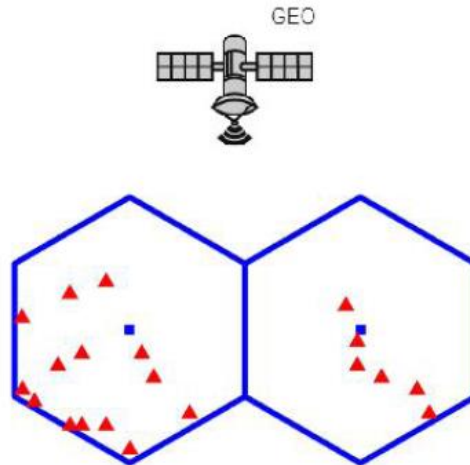


Figure 5-16: Flow diagram of proposed MLB algorithm

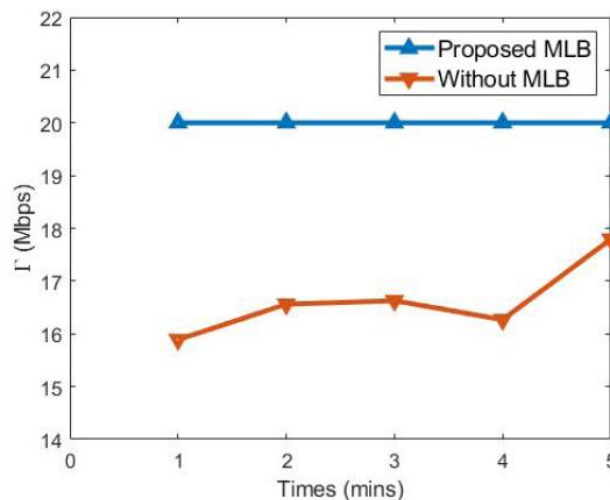
To evaluate the performance of the proposed algorithm, we consider a multi-RAT network composed of two LTE macro cells and a NTN GEO satellite. The macro cells were deployed in a hexagonal pattern within the coverage of the satellite network. We set the transmission power

to 46 dBm for the macro cells and 20 dBw for the satellite. The bandwidth for each macro cell was set to 20 MHz and the satellite operates in the Ka-band. For the macro cells, the path loss was modelled using a non-line of sight propagation loss. We randomly deployed 20 static UEs in the network and the required data rates for each of the UEs was set to 1 Mbps. We consider that 80% of the required data-rates are delay tolerant and the rest is due to low-latency delay sensitive data. The following figure shows the considered simulation environment, where the GEO satellite is plotted only for illustration purpose. The simulation was performed for a duration equivalent to five minutes of real-time network operation.



**Figure 5-17: Simulation environment for MLB algorithms**

The following figure shows the performance of the proposed algorithm. While no load balancing is considered, the overloaded macro cells cannot allocate the required resources to the users due to a lack of available resources in those cells; therefore, the user data rate is compromised in those overloaded cells. By offloading delay tolerant flows through the satellite network, the proposed algorithm enables the overloaded cells to allocate the required resources to users. Therefore, network throughput is increased when the proposed load balancing algorithm is adopted. Results show that with the proposed algorithm, the network achieved an average throughput of 20 Mbps.



**Figure 5-18: Simulation results with throughput enhancements**



## 5.5 Reinforcement learning based resource optimization for mobile relays

When millimetre wave communication is used for vehicular communication on the road, mobile relays with vehicle-to-vehicle (V2V) links are essential to overcome radio propagation shadowing or blocking problems. 3GPP standardization body determined that some part of uplink resources can be used for V2V link and therefore, interference between uplink and V2V link must be tightly controlled. Moreover, the interference power from satellite link for multi-connectivity can affect the V2V link performance. A reinforcement learning algorithm is adopted for rapid varying wireless channel environments, as interference-aware resource optimization is a critical factor for system performance and user QoE.

With performing interference control efficiently, the proposed algorithm optimizes radio resource allocation for each radio station by learning its optimal policy through adapting their actions by observing the rewards from the environment, which is a generic procedure for reinforcement learning. Reinforcement learning, especially multi-agent reinforcement learning based resource assignment technology is suitable for distributed control-based techniques for multiple access of radio link connections by multiple radio stations. The basic steps of the algorithm are given in the following:

- 1) The agent, as the controller, takes an action according to the initial random policy;
- 2) After the occurrence of an environment state transition, the reward for the action is given to the agent;
- 3) The agent performs a policy evaluation based on the state transition and reward occurred by its action;
- 4) The agent improves the policy based on the step 3). Note that step 3) and 4) can be performed at the same time;
- 5) The agent takes the next action according to the improved policy and continues with step 3).

Reinforcement learning is a learning process of the agent that finds the optimum policy maximizing the rewards with optimum state transitions in the course of test actions. Therefore, our radio resource allocation model is as follows.

- 1) Agent: multi-connectivity device that can perform cellular communication such as D2C (Device to Cellular) D2D (Device to Device) and satellite communication at the same time.
- 2) Action: device's behaviour determining its own frequency resource block and transmit power which is denoted by following notation.

$$A = \{(i, j); i \in \text{possible frequency block groups}, j \in \text{possible transmit power}\}$$

- 3) State: radio environment state set has the element of the vector  $(h, I, N, L, U)$  where  $h$  is radio channel information,  $I$  is interference power,  $N$  is frequency block group allocating to interfering devices,  $L$  is available capacity, and  $U$  is available delay.
- 4) Reward: the composite function of which arguments are the radio capacity and the delay. Its output can be the negative valued penalty for the arguments beyond the limit of the parameters set for satisfying QoS or QoE.

The proposed algorithm optimizes the radio resource allocation policy in the multiple radio link environment with reinforcement learning based Interference control algorithm. The analysed scenario is illustrated in the following figure.

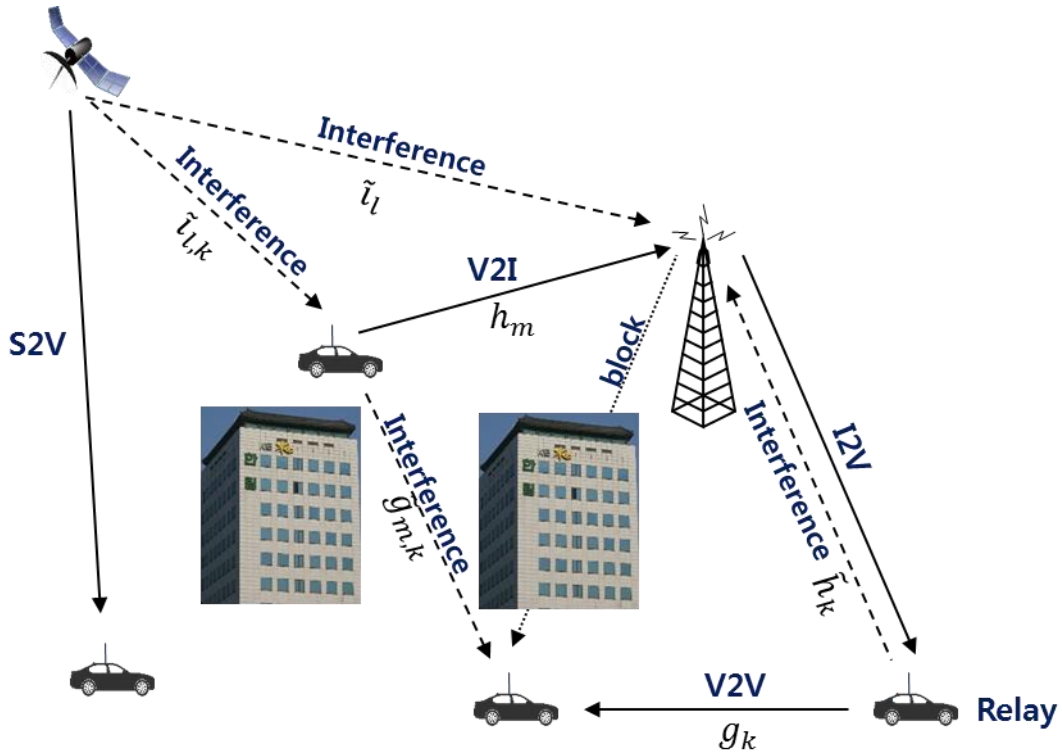


Figure 5-19: The analysis scenario for the proposed algorithm: the radio resource allocation policy with RL based interference control

In the figure, each device is the agent that controls each V2V (a.k.a., D2D) link and autonomously determines its sub-band and transmit power level for the communication. The proposed algorithm is a distributed control technique by which each agent D2D device controls its action set (resource and transmit power) values and gives V2I (a.k.a., D2C) devices the least interference while satisfying the delay limit (latency requirement). In order to control the inference that occurred in the whole radio environment, each device shares its controlling information such as its own resource allocation, transmit power, and measured radio channel conditions.

For system model, the signal to noise and interference ratio for D2C link is defined as

$$\gamma_m^c = \frac{P_m^c h_m}{\sigma^2 + \sum_{k \in K} \rho_{m,k} P_d^k \tilde{h}_k + \sum_{l \in L} \rho_{m,l} P_d^l \tilde{l}_l}$$

where  $P_m^c$ ,  $P_d^k$ ,  $P_d^l$  are the transmit powers for m-th D2C device, k-th D2D device and l-th satellite device respectively. Moreover,  $h_m$ ,  $\tilde{h}_k$ ,  $\tilde{l}_l$  are the radio channel gains arriving at cell site from m-th D2C device, k-th D2D device, and l-th satellite device, respectively.  $\sigma^2$  is the noise power and  $\rho_{m,k}$  and  $\rho_{m,l}$  are indicators for spectrum allocation for D2D device and satellite device, respectively.  $\rho_{m,k} = 1$  indicates spectrum allocation and  $\rho_{m,k} = 0$  indicates no spectrum allocation. Indicator  $\rho_{m,l}$  has the same meaning. Therefore, the capacity for m-th D2C device is given by

$$C_m^c = W \cdot \log(1 + \gamma_m)$$

where  $W$  denotes the bandwidth.

With the same analysis, the signal to noise and interference ratio for D2D link is given by

$$\gamma_k^d = \frac{P_d^k \cdot g_k}{\sigma^2 + G_c + G_d + G_f}$$

with

$$G_c = \sum_{m \in M} \rho_{m,k} P_m^c \tilde{g}_{m,k},$$

$$G_d = \sum_{m \in M} \sum_{k' \in k, k' \neq k} \rho_{m,k} \rho_{m,k'} P_{k'}^d \tilde{g}_{k',k}^d$$

and

$$G_f = \sum_{l \in L} \rho_{m,l} P_m^l \tilde{g}_{m,l}$$

where,  $\tilde{g}_{m,k}$ ,  $\tilde{g}_{k',k}^d$ ,  $\tilde{g}_{m,l}$  are interference channel gains from the interfering devices of D2C, other D2Ds and satellite devices. Therefore, the capacity for  $k$ -th D2D device is given by

$$C_k^d = W \cdot \log(1 + \gamma_k^d).$$

The detailed model of the proposed algorithm based on reinforcement learning is described in the following. At each time  $t$ , the D2D device, as the agent observes a state  $s_t$  defined by the arguments, and described in the above capacity equation. The device takes an action by selecting frequency resources and its transmission power. Then, the agent enters a new state  $s_{t+1}$  and evaluates and updates its policy in order to maximize the reward

$$r_t = \lambda_c \sum_{m \in M} C_m^c + \lambda_d \sum_{k \in K} C_k^d + \lambda_f \sum_{l \in L} C_l^f - \lambda_p (T_0 - U_t)$$

where  $T_0$  is the time constraint,  $U_t$  is the measured delay and  $\lambda_c$ ,  $\lambda_d$ ,  $\lambda_f$ ,  $\lambda_p$  are the weighting factors. According to the service type and intrinsic QoE, the four terms can be adjusted appropriately. For example, for delay critical services, the system can set  $\lambda_p \gg \lambda_c + \lambda_d + \lambda_f$ . The system follows Markov Decision Process (MDP) and the target of the reinforcement learning is to maximize the return values consisting of expected cumulative future discounted rewards [49]. Therefore, the proposed algorithm can adopt any reinforcement learning algorithm such as Q-learning, deep Q-learning, and so on.

## 5.6 AHP (Analytic Hierarchy Process) and Cooperative Differential Games for Multi-connectivity

This section reports a tentative approach studied in the scope of T4.2 which proved to be useful for the development of the other algorithms presented before. A preliminary formulation is here reported and its eventual refinements and developments will be detailed in D4.3.

The proposed methodology consists of a combination of Analytic Hierarchical Process and Cooperative Differential Game. The roles of the two techniques can be summarized as follow:

- AHP takes the responsibility of the network selection process, evaluating the affinity of each network with respect to the users' connections, considering the different network characteristics (e.g., cost, delay, latency, coverage, ...) and the actual users' QoS;
- Cooperative Differential Game is used to decides the networks' resources allocation. The networks are considered as cooperative agents, that optimally distribute their resources between users considering the output of the network selection process.

### 5.6.1 AHP (Analytic Hierarchy Process)

Network selection is a complicated process where several attributes should be considered. AHP [1] is a hierarchical model that can involve in the decision process all the criteria and all the candidate solutions considering the target.

The hierarchy reported in Figure 4-19, which details the decision process and its structure: starting from the level 1 (target object) the decision-makers are able to define the preferences for each attribute to reach the target; at level 2 (attributes) the decision-makers are able to assign a weight for each attribute with a pairwise comparison following the preferences defined at level 1; finally, in level 3 the weights of the attributes are used to evaluate the grades of the different options.

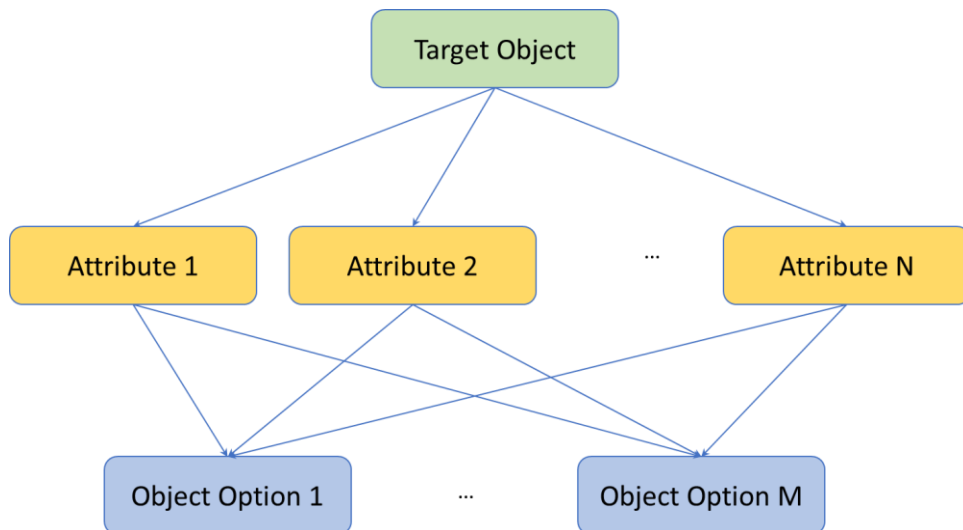


Figure 5-20: AHP Hierarchy

The AHP can be summarized in the following steps:

1. Define the decision problem: Construct a hierarchy structure in which the upper layer contains the desiderata, the intermediate layers contain the criteria and the possible choices are placed in the lowest layer.
2. Compute the vector of attribute scores: construct a pairwise comparison matrix whose elements are defined by Saaty's scale [64] (e.g., Figure 5-21). The pairwise matrix is composed by two pairwise comparisons: the first comparison is between the attributes

and a  $n \times n$  matrix  $A$  is computed, where  $n$  is the number of attributes. Each element  $a_{ij}$  of the matrix represents the importance of the  $j$ -th attribute relative to the  $i$ -th attribute. The matrix  $A$  has the following properties:

- a) each element  $a_{ij}$  with  $i > j$  is defined following the table of Figure 5-21
- b) if  $a_{ij} > 1$  then  $a_{ji} < 1$ ;
- c) if  $a_{ij} = 1$  then  $a_{ji} = 1$ ;
- d)  $a_{ij}a_{ji} = 1$ ;

the matrix constructed respecting the properties above is called a reciprocal matrix. The next step is the matrix normalization:

$$\bar{a}_{ij} = \frac{a_{ij}}{\sum_{l=1}^n a_{lj}}$$

Finally, the Attributes Priority Vector  $w_i$  (where  $i$  is the attribute index), can be computed as:

$$w_i = \frac{\sum_{l=1}^n \bar{a}_{il}}{n}$$

The Fundamental Scale for Pairwise Comparisons		
Intensity of Importance	Definition	Explanation
1	Equal importance	Two elements contribute equally to the objective
3	Moderate importance	Experience and judgment moderately favor one element over another
5	Strong importance	Experience and judgment strongly favor one element over another
7	Very strong importance	One element is favored very strongly over another; its dominance is demonstrated in practice
9	Extreme importance	The evidence favoring one element over another is of the highest possible order of affirmation
Intensities of 2, 4, 6, and 8 can be used to express intermediate values. Intensities of 1.1, 1.2, 1.3, etc. can be used for elements that are very close in importance.		

Figure 5-21: Saaty's Scale for Pairwise comparison

The comparison matrix is said to be consistent if and only if  $\lambda_{max} = n$ , where  $\lambda_{max}$  is the greater eigenvalue of the comparison matrix, and  $n$  is the matrix dimension.

A consistent comparison matrix implies that the Attributes Priority Vector  $w$  is unique, because is the eigenvector of the principal eigenvalue  $\lambda_{max}$  of the matrix.

In order to check the consistency, in [65] is defined the Consistency Index

$$CI = \frac{(\lambda_{max} - n)}{(n - 1)}$$

This index is compared with the Random Consistency Index, that is the average of CIs computed from randomly generated reciprocal matrix.

The Random Consistency Index computed for 500 matrices for each  $n$  is reported in Figure 5-22.

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

**Figure 5-22: Random Consistency Index for various  $n$**

The comparison between the two indexes is evaluated with the Consistency Ratio defined as:

$$CR = \frac{CI}{RI}$$

If the CR is smaller or equal to 0.1 the inconsistency is acceptable, otherwise, the subjective judgement needs to be revisited.

3. Compute the matrix of options scores. In the same way of the attribute score, one has to derive the so-called option score matrix. The option score matrix is an  $m \times n$  matrix, where  $m$  is the number of options and  $n$  is the number of attributes. For each attribute  $j$  a  $m \times m$  matrix  $B^j$  is constructed, where each element  $b_{ik}^j$  of the matrix is the comparison of the option  $i$ -th with respect to the option  $k$ -th considering the attribute  $j$ . The value of each element  $b_{ik}^j$  is chosen as defined above for the  $a_{ij}$  values, and each matrix  $B^j$  has the same properties of the matrix  $A$ . After normalization and consistency check, for each matrix  $B^j$  it is possible to define the score vector  $s^j$ . The entire procedure is the same as for the matrix  $A$  in the previous step.

Now, the options scores matrix  $S$  can be defined as

$$S = (s^1, s^2, \dots, s^n)$$

4. Compute the grades of the options: the last step is the association of a single grade for each option. The result is stored in a vector  $v$  defined as

$$v = Sw$$

Where the  $i$ -th entry of  $v$  is the grade of the  $i$ -th option.

### 5.6.2 Cooperative Differential Game

As introduced in section 5.2, differential games are a class of dynamic games that represent a framework in which it is possible to utilise results from optimal control theory. The main difference between the differential game and the classic optimal control is the presence of several independent control actions that drive the state evolution. Furthermore, these inputs are managed by different players and the objective function is no longer a single one, but each player can have a different objective function.

To obtain a mathematical model of the game the following elements must be defined:

- a set of differential equations, to model the state evolution of the system of interest;
- a set of state and control constraints, to model the physical constraints and the performances constraints on the variables of interest;
- the objective functions, to model the target performances.

The differential game can be:

- cooperative, in which all the players have the same objective function;

- non-cooperative or adversarial, in which each player has his own objective function.

For this algorithm a cooperative differential games is considered, while in section 5.2 we considered an adversarial one, meaning that the players coordinate their strategies in view of optimizing a collective objective function.

The general discrete time differential game, played on a time interval  $[k_0; K]$ , between  $m$  players, assumes the following form [66]:

$$\min_u J(x, u) = \sum_{k=k_0}^K x^T(k)Q(k)x(k) - R(k)^T u(k) \quad s. t.$$

$$x(k + 1) = f(x(k), u(k), d(k))$$

$$0 \leq u(k) \leq \bar{u}; \quad 0 \leq x(k) \leq \bar{x}$$

$$x(k_0) = x_0; \quad x(K) = x_f$$

$$x^T = [x_1, x_2, \dots, x_n]; \quad u^T = [u_1, u_2, \dots, u_m]; \quad d^T = [d_1, d_2, \dots, d_p];$$

### 5.6.3 Multi-Connectivity model for AHP and Cooperative Differential Game

In this section, the multi-connectivity problem is modelled as a AHP and Differential Game. To define the AHP problem model is possible to specialize the hierarchical structure in Figure 5-20 to the multi-connectivity model. Specialising the three layers one has:

1. Target Object: the target is the selection of the network that better fits the QoS Requirements, thus, the target object is a network that fully satisfy the actual QoS;
2. The Attributes: in the network selection process we need to consider several network attributes. Typically, each network has its own characteristics, such as, cost, security, packet delay, packet loss, bandwidth, etc, the values of these attributes can have large difference between networks. An example of attributes and their priority is shown in Figure 5-23

Value	Security	Cost	Delay	Bandwidth	Loss
Satellite	High	High	High	High	Low
WiFi	Low	Low	Low	Medium	Medium

Figure 5-23 Attribute example

On the other hand, the different services can have different requirements in terms of these attributes, an example is shown in table ...

Impact	Security	Cost	Delay	Bandwidth	Loss
VoIP	Medium	Medium	High	Low	Low
Video Download	Low	Medium	Low	High	Medium

Figure 5-24 Requirements examples

3. The Object Options: the possible option of the selection are all the access network that are in actual coverage of the user of interest. In this identification process, the actual performances of the access networks are not considered, but we consider only the intrinsic characteristics (attribute), to rank/select the networks without considering the resources, but only considering the QoS requirement.



Defined all the needed elements for the AHP the step defined before can be performed. In particular, the two comparison matrices defined above can be prestored for each type of service and networks, since the service requirements and the network characteristics can be considered static.

A further refinement can be the consideration of the user experience of the past connections: the users' feedback can be used to modify these comparison matrices. With this in mind a complete QoE Management framework can be considered, in particular all the users can be clustered on the basis of their personal characteristics and each of this cluster can have associated, for each service, the appropriate attribute scores and option scores as defined in the AHP, and these matrices will be updated during the time on the basis of the users' feedback.

The output of the AHP, i.e., the users' preferences about the different networks, will be the input of the differential game problem. In this control approach the resources to allocate are the bandwidths of the networks along the users. This allocation shall consider the users' (or services) requirements and the actual networks performances (at least the networks allocable bandwidth). With this objective, and the assumption that the networks never saturates but only limited resources (bandwidth) to assign, the differential game can be modelled as follows:

$$\min_u J(x, u) = \sum_{k=k_0}^K x^T(k)Q(k)x(k) - R(k)^T u(k)$$

$$x_i(k+1) = x_i(k) - \sum_{l=1}^m u_{li}(k) + d_i(k) \quad i = 1, 2, \dots, n$$

$$0 \leq \sum_{i=1}^n u_{ji}(k) \leq \bar{u}_j \quad j = 1, 2, \dots, m$$

$$x(k)^T = [x_1, x_2, \dots, x_n]$$

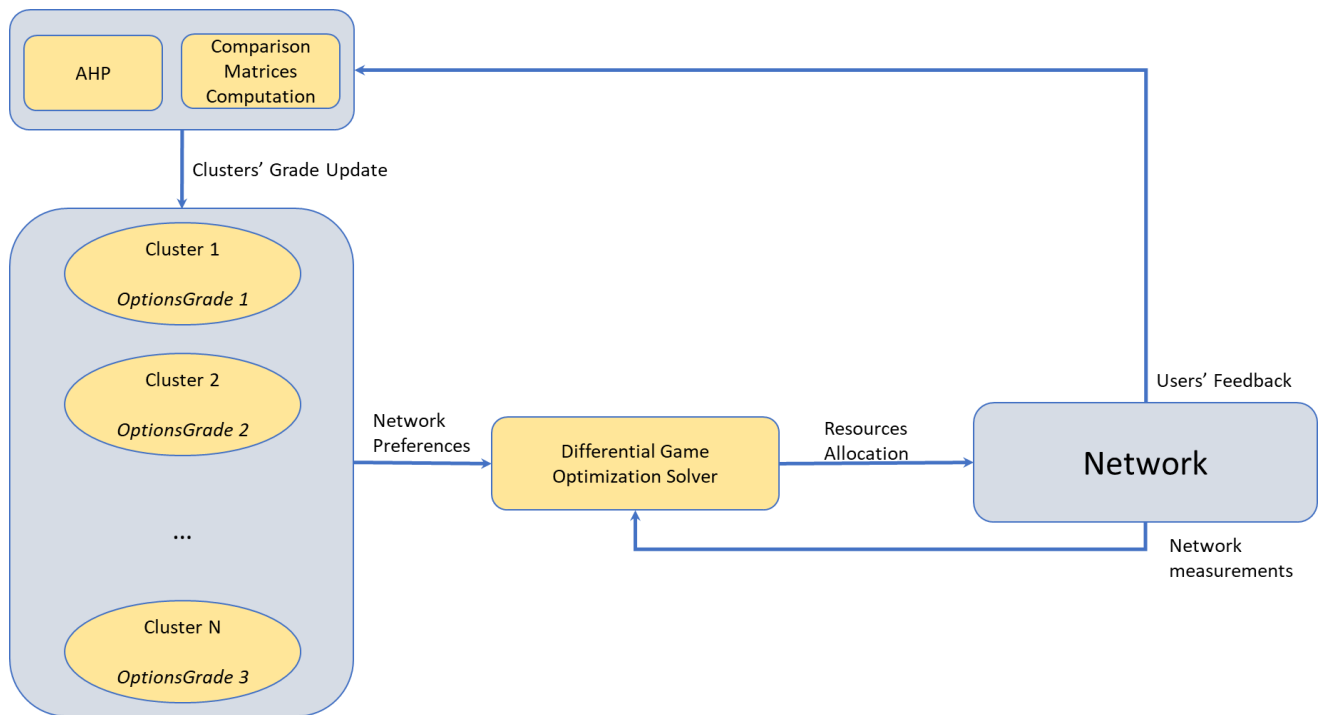
$$u(k)^T = [u_{11}, \dots, u_{1n}, \dots, u_{m1}, \dots, u_{mn}]$$

Where

- $x \in \mathbb{R}^n$  is the state vector, composed by the tail of users;
- $u \in \mathbb{R}^{mn}$  is the control vector, each  $u_{i..}$  is a vector containing the resources that the  $i$ -th networks allocates for all the users;
- $Q \in \mathbb{R}^{n \times n}$  a positive semi-definite matrix that indicates the weights of the state elements on the cost function;
- $R \in \mathbb{R}^{n \cdot m}$  the grade vector of each user  $n$  for each network  $m$ ;
- $d \in \mathbb{R}^n$  the data coming from the application layer of the users;

and the evolution of the vector  $d$  depends by the users' services. The vector  $R$  will be periodically updated by the AHP processing block and the matrix  $Q$  is defined according to the users' priority, e.g., a user with a better contract sill receive most importance in the cost function.

The block diagram of the method described above is shown in



**Figure 5-25 Block diagram of the AHP algorithm**

## 5.7 ZeroSum Games and Minimax for Multi-Connectivity

This section reports a tentative approach studied in the scope of T4.2 which proved to be useful for the development of the other algorithms presented before. A preliminary formulation is here reported and its eventual refinements and developments will be detailed in D4.3.

As introduced in sections 4.3 and 4.7, in differential games the players can be cooperative or non-cooperative. In the case of cooperative players, their respective control actions, are chosen to maximize a common utility function, while in the non-cooperative scenario, each player chose a control action in order to maximize his own utility function.

A special case of non-cooperative differential game is the so-called two-player zero-sum differential game, widely used in control applications and in robust control theory.

It is possible to model a two-player zero-sum differential game as a game in which the cost function is common to the players, but, one player selects his control action to maximize and the other player to minimize such a function.

A general zero-sum dynamic game is a triple  $\{\mathcal{H}; \mathcal{U}; \mathcal{V}\}$ , where given the state equation

$$\begin{aligned} x_{k+1} &= f_k(x_k, u_k, v_k) \\ x_1 &= \hat{x} \\ k &\in [1, K] := \{1, \dots, K\} \end{aligned}$$

and the finite-horizon cost function given by

$$J(u, w) = \sum_{k=1}^K g_k(x_{k+1}, u_k, v_k, x_k)$$

which is to be minimized by Player 1 and maximized by Player 2, choosing the control vector  $u_{[1,K]}$  and  $v_{[1,K]}$  respectively.

The state  $x_k$  takes values in  $R^n$ , the control vector of Player 1 takes value in  $\mathcal{U} \subset R^{m_1}$  and the control vector of Player 2 takes value in  $\mathcal{V} \subset R^{m_2}$ . The players choose the control actions  $u_{[1,K]}$  and  $v_{[1,K]}$  respectively minimizing and maximizing the Hamiltonian function of the problem  $\mathcal{H}$  [67].

In this non-cooperative framework, a solution of interest is the so-called Nash Equilibrium [67]. Such equilibrium is reached when no player can improve his payoff by a unilateral change of his strategy, as long as the other player sticks to the equilibrium strategy.

In particular, the pair strategies  $(u^*, v^*)$  is a Nash Equilibrium of the game if, for every  $u \in \mathcal{U}$  and  $v \in \mathcal{V}$

$$\mathcal{H}(u, v^*) \leq \mathcal{H}(u^*, v^*) \leq \mathcal{H}(u^*, v)$$

The existence of the Nash Equilibrium is guaranteed under continuity and convexity assumption from [67]:

**Theorem:** Assume that the sets  $\mathcal{U}, \mathcal{V}$  are compact and convex in  $R^m$ , the payoff function  $\mathcal{H}(u, v)$  is:

- continuous in  $(u, v) \in \mathcal{U} \times \mathcal{V}$ ;
- concave function of  $u$  for each fixed  $v \in \mathcal{V}$
- concave function of  $v$  for each fixed  $u \in \mathcal{U}$

Then, the non-cooperative game admits Nash Equilibrium.

Since this problem is a generalization of an optimal control problem, it can be solved using a generalized version of the Pontryagin Maximum Principle, called Nash-Pontryagin Minimax Principle [68], or using an extension of the Hamilton-Jacobi-Isaacs theory [68].

Our proposed solution is based on the Duality Theory [69].

In particular, the Nash Equilibrium  $(u^*, v^*)$  as defined above, is a saddle-point of the function  $\mathcal{H}(u, v)$ . Given two sets  $\mathcal{U}$  and  $\mathcal{V}$  and  $\mathcal{H}: \mathcal{U} \times \mathcal{V} \rightarrow R$ , the weak duality inequality is:

$$\sup_{v \in \mathcal{V}} \inf_{u \in \mathcal{U}} \mathcal{H}(u, v) \leq \inf_{u \in \mathcal{U}} \sup_{v \in \mathcal{V}} \mathcal{H}(u, v)$$

It is possible to define the primal and dual problem as follow. Given a point  $(u_0, v_0) \in \mathcal{U} \times \mathcal{V}$ , starting from the previous inequality, the following holds:

$$\inf_{u \in \mathcal{U}} \mathcal{H}(u, v_0) \leq \mathcal{H}(u_0, v_0) \leq \sup_{v \in \mathcal{V}} \mathcal{H}(u_0, v)$$

Let define primal and dual values, respectively for  $u \in \mathcal{U}$  and  $v \in \mathcal{V}$ ,  $p(u)$  and  $d(v)$  as:

$$p(u) := \sup_{v \in \mathcal{V}} \mathcal{H}(u, v)$$
$$d(v) := \inf_{u \in \mathcal{U}} \mathcal{H}(u, v)$$

finally, the primal and dual problem can be defined as

$$\min_{u \in \mathcal{U}} p(u)$$
$$\max_{v \in \mathcal{V}} d(v)$$

The weak duality inequality then states that

$$\max_{v \in \mathcal{V}} d(v) \leq \min_{u \in \mathcal{U}} p(u)$$

let  $(u^*, v^*) \in \mathcal{U} \times \mathcal{V}$  be a saddle-point of  $\mathcal{H}: \mathcal{U} \times \mathcal{V} \rightarrow R$ . Then, by definition

$$\mathcal{H}(u^*, v) \leq \mathcal{H}(u^*, v^*) \leq \mathcal{H}(u, v^*)$$

Which, assuming that the conditions of the theorem reported before [67] holds, is equivalent to

$$\sup_{v \in \mathcal{V}} \mathcal{H}(u^*, v) = \mathcal{H}(u^*, v^*) = \inf_{u \in \mathcal{U}} \mathcal{H}(u, v^*).$$

The last equation, combined with the weak duality inequality, shows that:

$$\begin{aligned} u^* &= \arg \min_{u \in \mathcal{U}} p(u) \\ v^* &= \arg \max_{v \in \mathcal{V}} d(v) \\ \min_{u \in \mathcal{U}} p(u) &= \max_{v \in \mathcal{V}} d(v) = \mathcal{H}(u^*, v^*) \end{aligned}$$

The multi-connectivity problem will be divided in primal and dual problems. Each of the two can be solved separately by the players, which is possible with the assumption that each player knows the admissible control actions of the other player and the actual value of the state.

### 5.7.1 Network Dynamics and Utility Function for Multi-Connectivity Modelling

The reference problem is the bandwidth allocation problem in dual-connectivity scenario, i.e. the user can be connected simultaneously to two different access networks with different characteristics. As an example, a user can be connected to a satellite network, a terrestrial network or both simultaneously.

The choice of how much user data must be sent through a network or the other needs to take into account the congestion of the two networks and how much the actual user service fits with the network characteristics. In our case, the bandwidth allocation problem is considered taking into account one user with a time-varying traffic profile and two networks with constant back-haul capacity.

The network dynamics of the problem described above is expressed by the following equations:

$$\begin{aligned} x_{k+1}^1 &= x_k^1 + d_k - u_k - v_k \\ x_{k+1}^2 &= x_k^2 + u_k - K_u \\ x_{k+1}^3 &= x_k^3 + u_k - K_v \end{aligned}$$

- $x_k^1$  is the number of packets in the tail of the user at time  $k$ ;
- $x_k^2$  is the number of packets in the tail of the terrestrial access network at time  $k$ ;
- $x_k^3$  is the number of packets in the tail of the satellite access network at time  $k$ ;
- $d_k$  is the number of packets arriving at time  $k$  from the application layer of the user equipment;
- $u_k$  is the number of packets allocated in the terrestrial network at time  $k$ ;
- $v_k$  is the number of packets allocated in the satellite network at time  $k$ ;
- $K_u$  it represents the number of packets transmitted through the terrestrial network back-haul at each time;
- $K_v$  it represents the number of packets transmitted through the satellite network back-haul at each time.

The variables  $x_k^i$  are bounded by the size of the tails of the user equipment and the access networks. The variable  $d_k$  assume values during the connection time according to the actual user equipment application or service, e.g. the packets generated by a chat application have different characteristics with respect to the packets generated by a video streaming application in terms of amount of packets rate, value range and value variations. The variables  $u_k$  and  $v_k$  are bounded by the maximum available capacity, of the two access networks, in terms of number of packets at time  $k$ . The constants  $K_u$  and  $K_v$  are constant values that represent the number of packets that the two access networks are able to unload at each time instant.

On the basis of the description above the problem constraints at time  $k$  are:

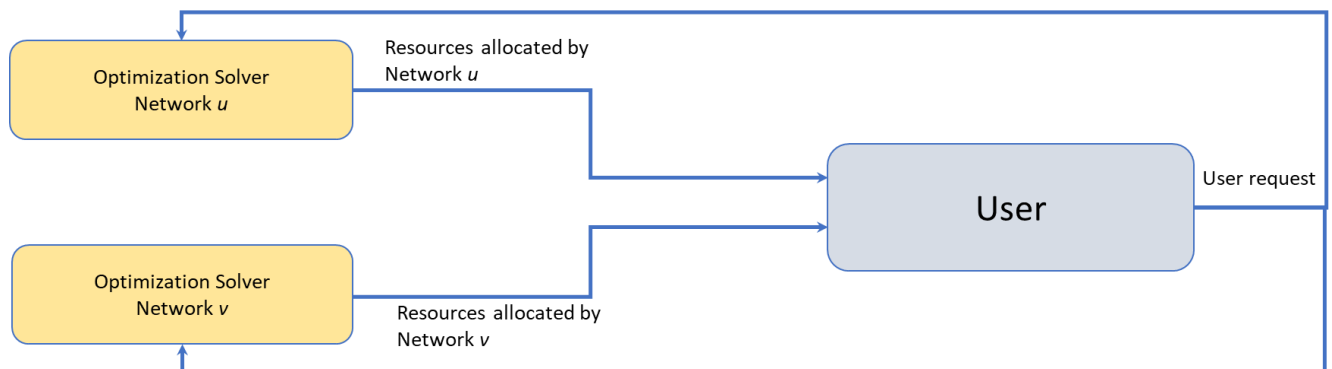
$$\begin{aligned} 0 \leq x_k^i &\leq \bar{x}^i \quad i = 1,2,3 \\ 0 \leq u_k &\leq \bar{u} \\ 0 \leq v_k &\leq \bar{v} \end{aligned}$$

The final element to define the Zero-Sum Game is the cost function. The objective of the players is to run out the packets coming from the UE, avoiding the overflow of their queues. Furthermore, they are interested to receive the maximum reward per packet by the UE, this can be considered taking into account the QoS characteristics of the UE application and the QoS characteristics of the access network. The needs described previously can be modelled by the following cost function:

$$J(x_{k+1}, u_k, v_k) = x_{k+1}^T R x_{k+1} + \mu_u u_k - \mu_v v_k$$

with R positive definite matrix, constructed such that if minimized with respect to  $u_k$  the user and network 'u' tails (i.e.,  $x_{k+1}^1$  and  $x_{k+1}^2$ ) are minimized and if it is maximized with respect to  $v_k$  the user and the network 'v' tails (i.e.,  $x_{k+1}^1$  and  $x_{k+1}^3$ ) are again minimized. Finally,  $\mu_u$  and  $\mu_v$  are values between zero and one, considered constant for a given user service traffic and are proportional respectively to how much the user service of the considered traffic fits with the network 'u' and with the network 'v'.

Figure 5-26 reports a functional scheme for the algorithm.



**Figure 5-26 Block diagram of the minimax algorithm**

## 6 Framework for QoE-aware Multi-Connectivity

This section will discuss the integration of the QoE functionalities into the traffic flow control algorithm solutions presented in Section 3 and, overall, with the entire 5G-ALLSTAR system.

### 6.1 QoE role and functional architecture in 5G-ALLSTAR

The QoE/QoS management is a fundamental part of a “Personalisation System” of the 5G-ALLSTAR system, a functional module envisaged in the grant agreement and already introduced in D4.1.

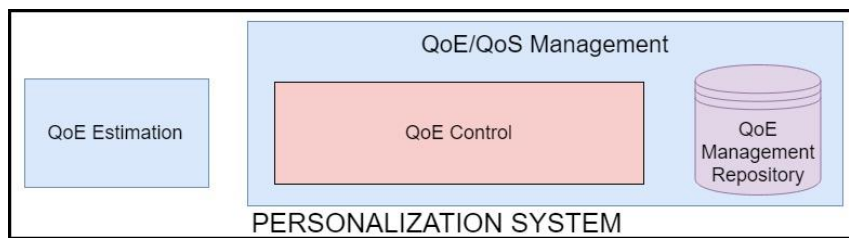


Figure 6-1: Personalization System from D4.1

The 5G-ALLSTAR personalization system aims at providing a set of non-standardized indicators regarding the connection preferences of the user, depending on its profile and on the specific service/QoS flow characteristics. Such connection preference indicators will be used as an additional input to the traffic flow controller module, so that its algorithms may be tuned and personalised to the specific user, allowing to archive a better QoE for all users thanks to the improvements associated with multi-connectivity in multi-RAT networks.

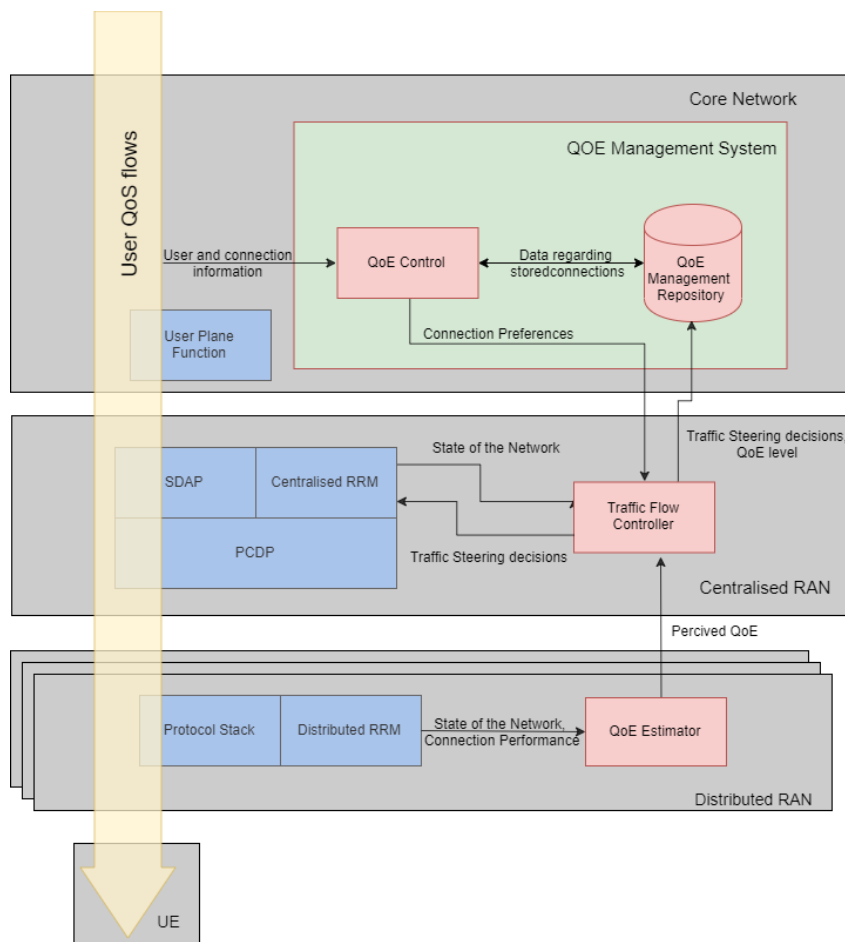
The personalization system is a part of the 5G-ALLSTAR target system, and from a functional point of view it can be seen as a single module that provides a set of APIs (Application Programming Interfaces) to offer its non-standardized services to the standardized core network functionalities (e.g., SMF, AMF), allowing the integration of advanced QoE-based services into the network without requiring significant modification to the network itself and its functioning.

The three individual sub-modules reported in Figure 6-2 were detailed in D4.1 and are briefly introduced in the following:

- The **QoE control** module is the core of the personalisation system and provides the connection preferences to the various PDU sessions that are steered by the traffic flow controller, assigning them a so-called “enriched QoS profile”.
- The **QoE management repository** represents the knowledge base of the system and gathers all the historic data regarding the QoE performance, from whose analysis the QoE control module is able to infer the user profile and its connection preferences.
- The **QoE estimation** is a distributed module that infers the QoE level, in a personalised way, for the QoS flow under analysis, starting from a local (D-RAN level) measurement of the connection performance in terms of QoS indicators.

For a more detailed description of the specific functionalities of the sub-modules, the reader is referred to D4.1 “*Mapping of the multi-connectivity functions onto the 5G network architecture*”. Figure 6-2 reports a detailed and updated version of the functional architecture of the personalisation system and its logical information exchanges.





**Figure 6-2: QoE management system functional architecture**

The purpose of the personalisation system is to provide support to, and extend the capability of, the traffic flow controller. From Figure 6-2 it can be seen that the traffic flow controller takes the perceived QoE level of the on-going PDU session as an additional feedback, so that it can update its traffic steering decision dynamically to maximize the QoE level. The traffic flow controller also gathers information regarding the connection preferences of the individual user, so that it can act as a personalized controller, and stores its decisions and collected inputs, into the dedicated knowledge base.

All of the three submodules will be implemented in an adequate form into the project PoC to demonstrate the QoE-aware capabilities of multi-connectivity. In particular, to emulate a collection of the historical data regarding the QoE level associated to past PDU sessions, the consortium designed a questionnaire aiming at inferring a set of potential user profiles. Those profiles will then be mapped onto proper connection preferences that will be fed into the implemented traffic flow control algorithms to make their behaviour adaptive and user-centric/personalised. A dedicated module will also be deployed, in order to evaluate the perceived QoE level of the tested PDU sessions, starting from a measure of their QoS and the emulated user's profile.

The following section details the procedure for the design, and the preliminary results, of the questionnaire.

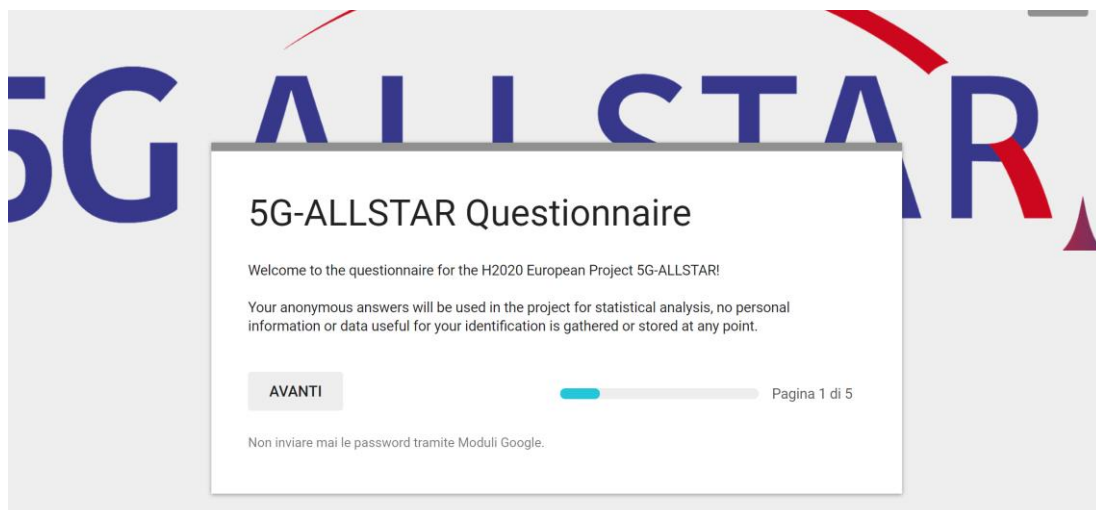
## 6.2 Questionnaire

To identify the QoS parameters that affect the most users' QoE in multi-media delivery services, the consortium designed a questionnaire for the general public. The questions were grouped into four sections as follows:

- Section 1, containing questions regarding the user preferences in watching movies via smartphone and tablet;
- Section 2, in which the user had to evaluate a video streaming that tried to deceive her/him regarding its resolution;
- Section 3, where a bit rate and resolution degradation was added into a video stream and the user was asked about its perceived quality;
- Section 4, in which a video stream was subject to severe audio delays.

Starting from the analysis of the answers collected, the algorithm(s) that will be demonstrated in the project PoC will be adapted and tuned accordingly, so that their behaviour is coherent with the QoE sensitivity analysis that can be inferred for multi-media delivery.

The rationale for introducing three video streams in the questionnaire that either try to deceive the user or that manifest service degradation problems was that, depending on the user answers, it could be possible to identify which are the key QoS parameters that shall be prioritised in scenarios with a scarcity of network resources to preserve, as much as possible, the overall QoE. It is worth recalling that the project demonstrator for the system multi-connectivity capabilities is expected to take its traffic-flow control decisions based not only on a measured feedback regarding the network state, but also depending on the specific user connection preferences, that are likely to directly correlate to the QoS parameters that affect the most her/his perceived QoE.

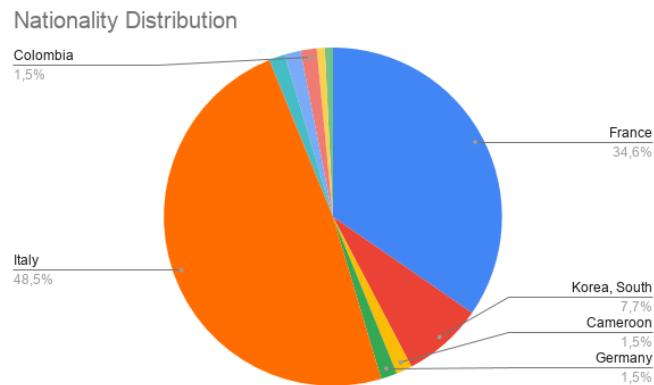


**Figure 6-3: 5G-ALLSTAR questionnaire landing page**

At the end of M18, a total of 131 answers were collected, mainly gathered after dissemination events (e.g., seminars) and social media, where the landing page of Figure 6-3 was presented. The geographic distribution of the users is reported in Figure 6-4.

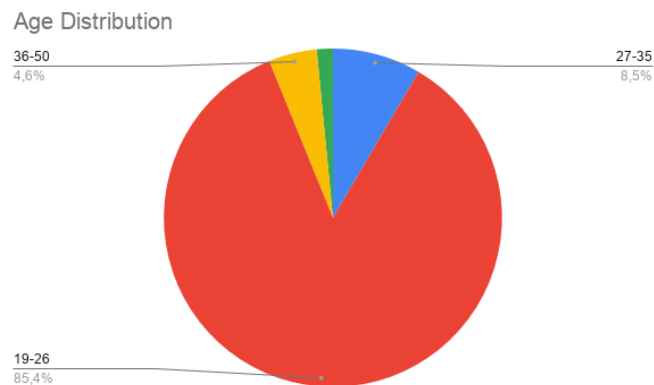
After more answers are collected, the data will be analysed to study the presence of user clusters and associating them to a set of user profiles, depending on which the QoE estimator and the traffic flow controller will be automatically configured in a personalised way.

### 6.2.1 Answers report to the first section of the questionnaire



**Figure 6-4: Nationality Distribution of the Questionnaire**

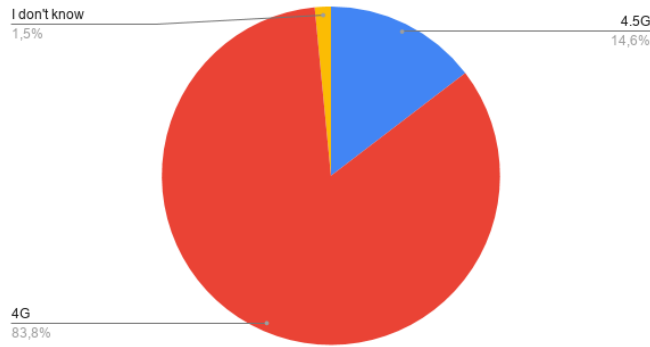
A significant number of users was collected by the academic partners, as demonstrated by both the nationality distribution (CRAT is based in Italy, while CEA and GEM are based in France) and the average age of the users, reported in Figure 6-5.



**Figure 6-5: Age Distribution of the Questionnaire**

Most of the users had a 4G smartphone as emerges from Figure 6-6, with a limited portion of the users having a 4.5 connection. This particular answer may be an indicator that several users could be not aware of the complete technical specifications of their smartphones, as most connections in Italy and France that jointly amount for around 80% of the answers, are 4.5G enabled.

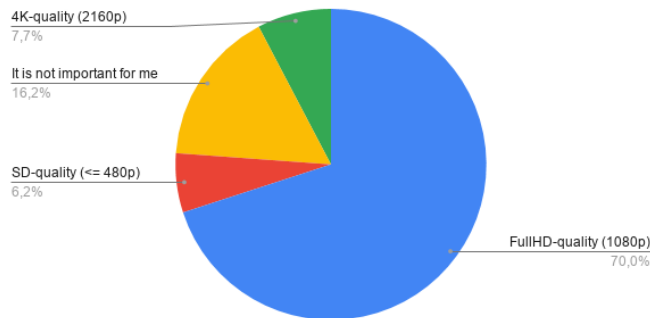
What is the connection used by your smartphone/tablet?



**Figure 6-6: Answers to question 1 pie chart**

From Figure 6-7 it can be seen that almost all users showed to be satisfied with Full HD video streams, even if a non-negligible percentage of users was not interested in resolution at all. Figure 6-8 highlights that most users utilise video streams in mobility for long routes, while around 30% of them watches media during short trips.

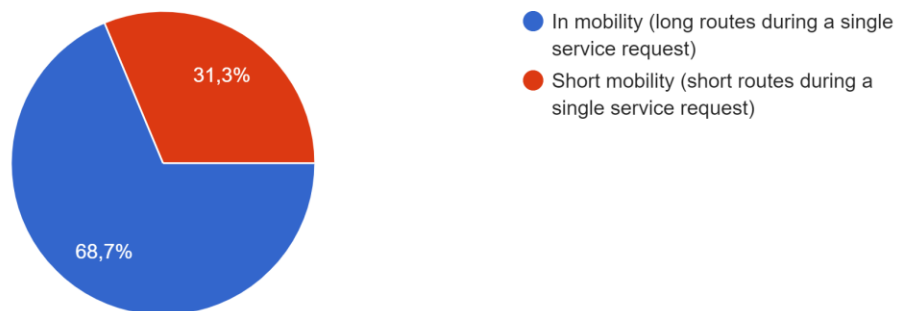
What type of service do you typically require for watching movies by using your smartphone/tablet?



**Figure 6-7: Answers to question 2 pie chart**

When you mainly use your smartphone/tablet for watching movie you are:

131 risposte



**Figure 6-8: Answers to question 3 pie chart**

Figure 6-9 reports that around one third of the users take into high consideration delays during live events, while the others would accept severe delays, with a preference of having it limited to less than half a minute.

If during a live event, what you are seeing is not in real-time but it is delayed with respect to the original event time ...ost 20 seconds), this would be for you:

131 risposte

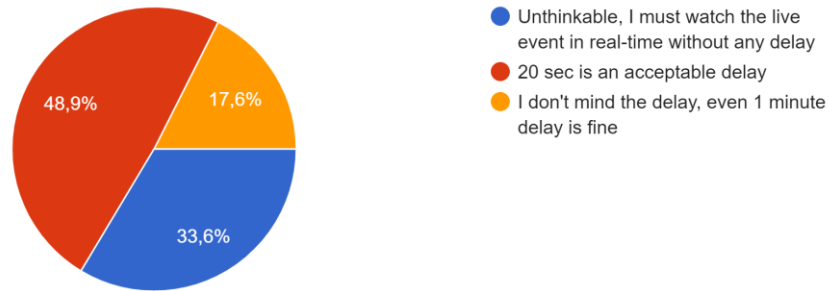


Figure 6-9: Answers to question 4 pie chart

Figure 6-10 shows that most users have into high consideration the battery life of their smartphone during work time, in line with the replies regarding the length of their trips of Figure 6-8. The majority of the users still requested a trade-off between performance and battery consumption, while only 5% of the replies neglected completely the video quality for the sake of autonomy. Figure 6-11, on the contrary, highlights that for entertainment the majority of the users would disregard power consumption in favour of video quality.

If you could choose, during your working time, you would choose a video streaming technology that requires:

131 risposte

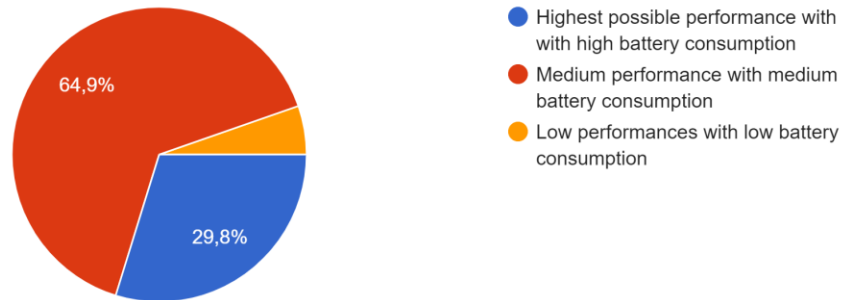


Figure 6-10: Answers to question 5 pie chart

If you could choose, during your pleasure time, you would choose a video streaming technology that requires:

132 risposte

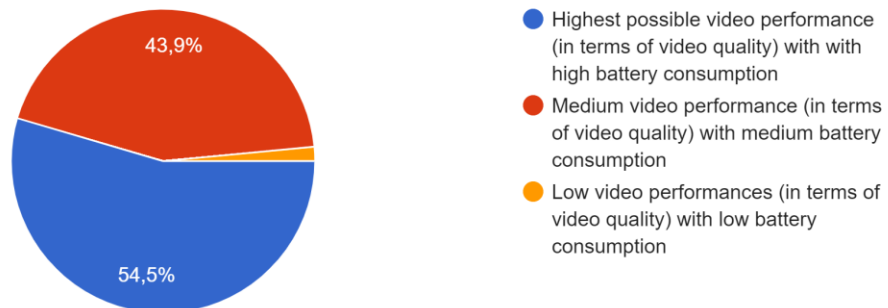
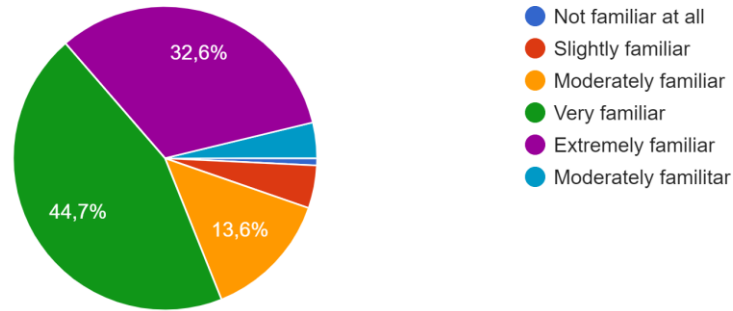


Figure 6-11: Answers to question 6 pie chart

Figure 6-12 shows that the users consider themselves very familiar with their smartphones and that they utilise it quite often and continuously during the day.

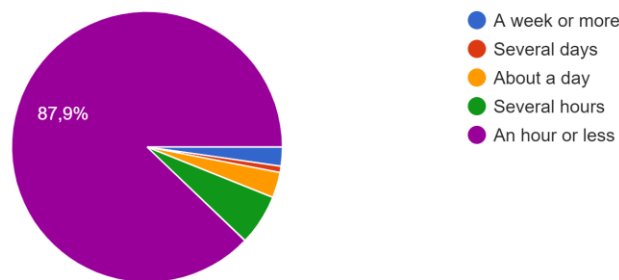
**How familiar are you with your own smartphone/tablet?**

132 risposte



**How much time has passed since the last time you have used your smartphone?**

132 risposte



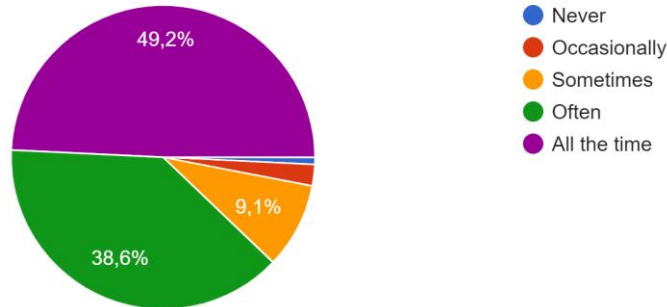
**Figure 6-12: Answers to questions 7 and 8 pie charts**

The same concept is highlighted in Figure 6-13, where the majority of the users state to utilise their smartphone for more than 2 hours a day.



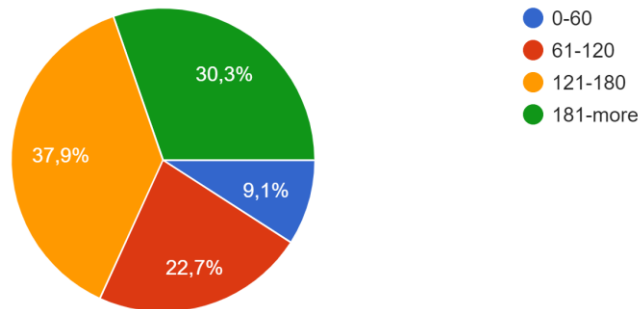
On average, how frequently do you use your smartphone?

132 risposte



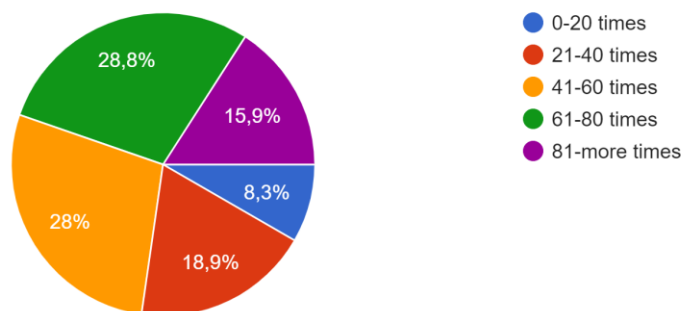
Considering your average behaviour over the previous month, how much time (in minutes) you typically spend on your smartphone in a day?

132 risposte



How many times do you check your smartphone on a typical day? (give your best guess)

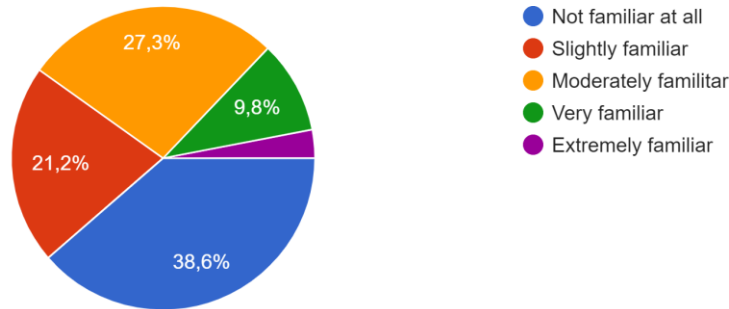
132 risposte



**Figure 6-13: Answers to questions 9, 10 and 11 pie charts**

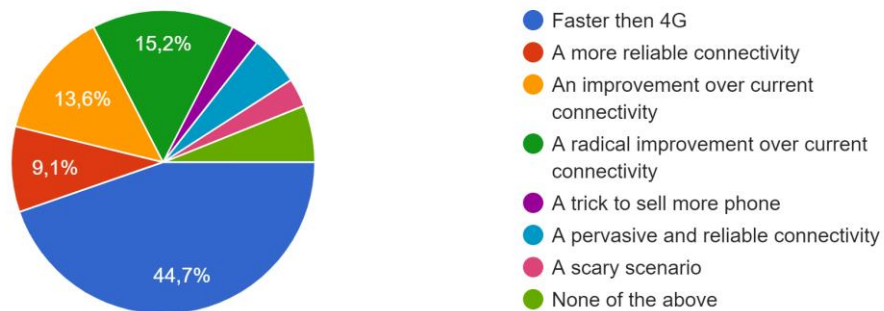
### How familiar are you with 5G?

132 risposte



### What is the main advantage of 5G for you?

132 risposte



**Figure 6-14: Answers to questions 12 and 13 pie charts**

To conclude the first section, Figure 6-14, details the level of confidence of the users regarding 5G. Over 50% of the users are not confident with the technology, showing that general public dissemination events on 5G are still needed. The majority of the users identified in the throughput increase the main innovation of 5G.

**6.2.2 Answers report to the second section of the questionnaire**

The purpose of this section of the seminar is to test the sensitivity of the users QoE to the resolution of the video stream. The user was deceived regarding the resolution of the video by the inclusion of a 4K logo and a fake resolution on the bottom right, as the native quality of the stream was 720p.

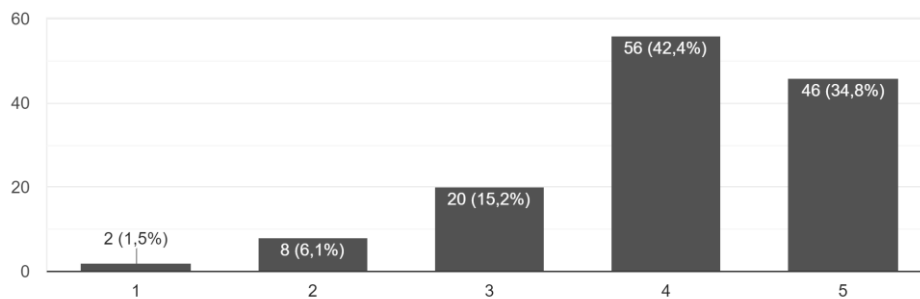


**Figure 6-15: Screenshot of the first video shown during the questionnaire**

The video chosen was characterised by vivid colours and images that seem to have a lot of details, despite their resolution as the one in the screenshot of Figure 6-15.

Please express your degree of satisfaction concerning the streaming quality (1 bad experience, 5 excellent experience)

132 risposte



**Figure 6-16: Bar chart of question 1, second section**

As reported in Figure 6-16, the users showed an overall satisfaction in the streaming quality, despite the fact that the resolution was severely lower than declared, to the point that visual artefacts and compression noise started to appear.

Please express your degree of satisfaction concerning the audio quality (1 bad experience, 5 excellent experience)

132 risposte

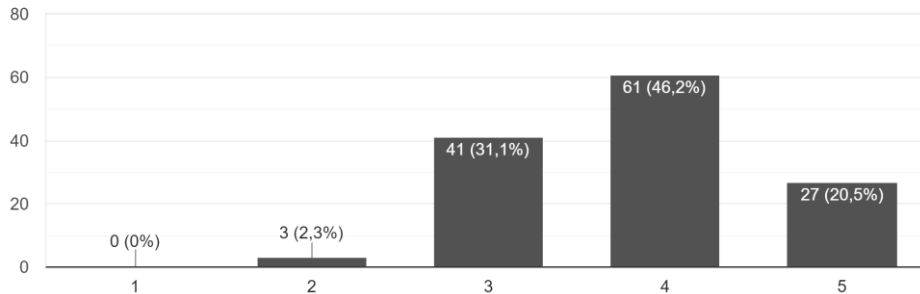


Figure 6-17: Bar chart of question 2, second section

A similar degree of satisfaction is observed for the audio in Figure 6-17, even if no service degradation was present in this case.

Most of the users were not able to discern the native quality of the stream: several users (32%) were tricked into believing that the resolution was 4K, while only 20% identified the correct resolution. It is worth mentioning that the Full HD answer may have been selected also by several users that thought to have seen a 4K stream on their Full HD phone screen.

What kind of resolution do you think the video has?

132 risposte

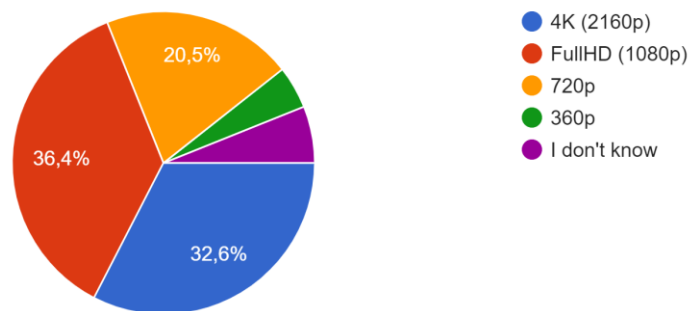
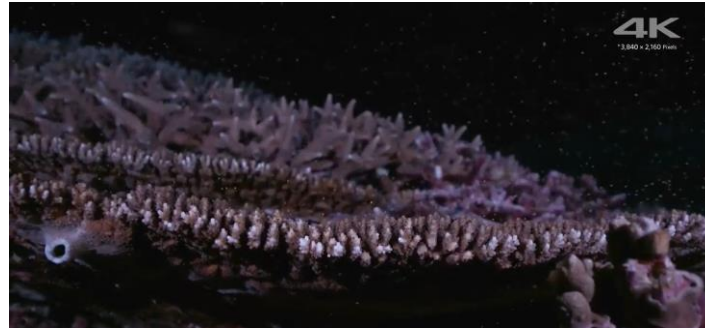


Figure 6-18: Pie chart of question 3, second section

### 6.2.3 Answers report to the third section of the questionnaire

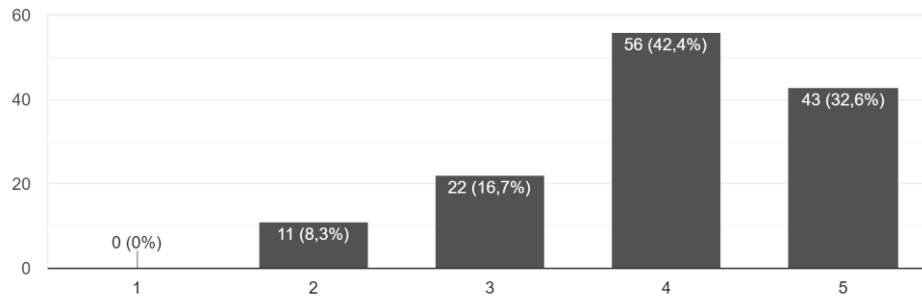
In this section of the questionnaire, the users were asked the same questions as the previous one, in order to evaluate a video stream in which the resolution and the bitrate degraded dynamically. The stream was natively Full HD, while it reached 720p in a few instances (as in the screenshot of Figure 6-19) and had a time-varying bitrate to model the behaviour of an adaptive encoding solution as the one that will be tested in the demonstrator of the project. Once again, a fake 4K logo and resolution were included in the top right.



**Figure 6-19: Screenshot of the second video shown during the questionnaire**

Please express your degree of satisfaction concerning the streaming quality (1 bad experience, 5 excellent experience)

132 risposte

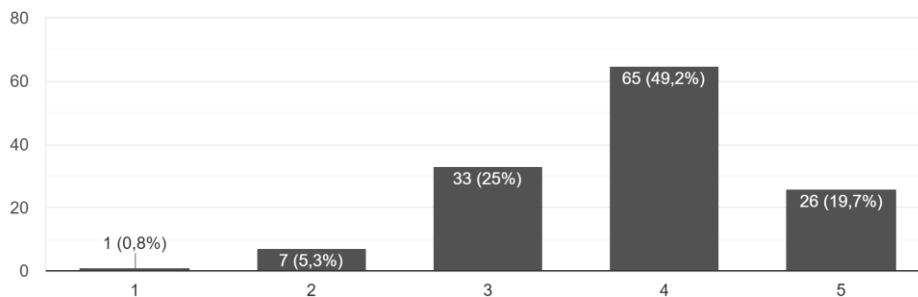


**Figure 6-20: Pie chart of question 1, third section**

As it is possible to observe in Figure 6-20, the perceived video quality did not show any significant drop even if the native resolution of the video was Full HD. The same observations can be done also for the audio evaluation, as reported in Figure 6-21.

Please express your degree of satisfaction concerning the audio quality (1 bad experience, 5 excellent experience)

132 risposte



**Figure 6-21: Bar chart of question 2, third section**

In this second test, several users perceived a better resolution of the video, and as a consequence, the number of 4K answers increased in the third question, as reported in Figure 6-22.

What kind of resolution do you think the video has?

132 risposte

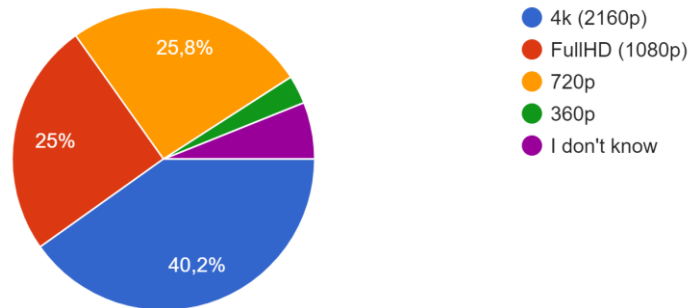


Figure 6-22: Pie chart of question 3, third section

### 6.2.4 Answers report to the fourth section of the questionnaire

The final video in the questionnaire had a twofold role: on the one hand it was aimed at testing the importance of the synonyms between the video and the audio QoS flows, while on the other it was used to discern the users that likely lost interest while completing the questionnaire. This result was obtained by adding a severe (~0.5s) second long delay between the video and the audio stream that consisted in a very short speech.

Please express your degree of satisfaction concerning the streaming quality (1 bad experience, 5 excellent experience)

132 risposte

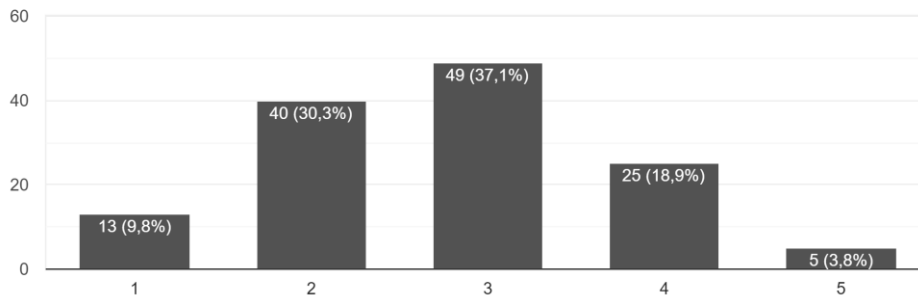


Figure 6-23: Bar chart of question 1, fourth section

Figure 6-23 highlights that the presence of delays in the audio flow impacted the perceived quality of the overall stream. In Figure 6-24 we can observe that the number of users that did not answer with care to the questionnaire is overall limited, as the vast majority of the users were aware of the delays and only 7 picked the excellent experience answer and did not perceive delays as in Figure 6-25.

Please express your degree of satisfaction concerning the audio quality (1 bad experience, 5 excellent experience)

132 risposte

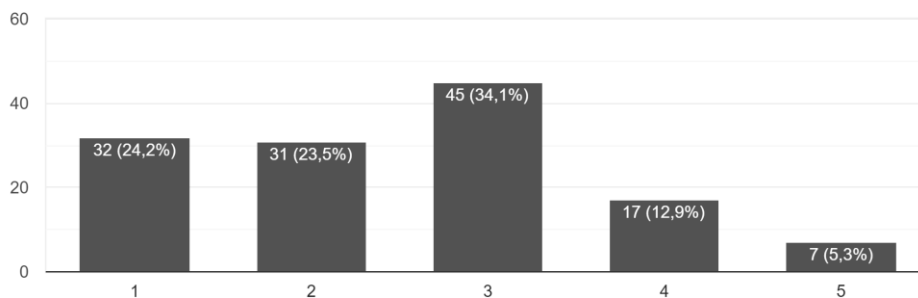
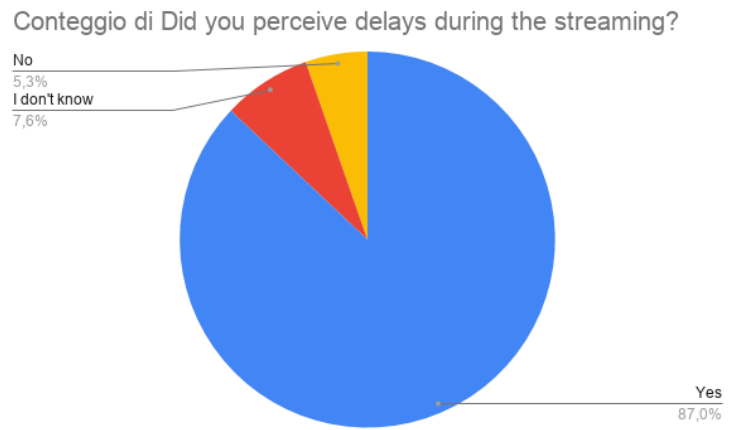


Figure 6-24: Bar chart of question 2, fourth section





**Figure 6-25: Pie chart of question 3, fourth section**

### 6.2.5 Preliminary analysis of the questionnaire answers

This section reports the result of the first analysis conducted on the data gathered by the questionnaire. Future studies will be conducted once the amount of answers has grown, so we limit the study for now to the investigation regarding the potential presence of clusters and correlations of interest. The characterisation of the user clusters and a more detailed analysis on the data gathered will be included in D4.3 at M30.

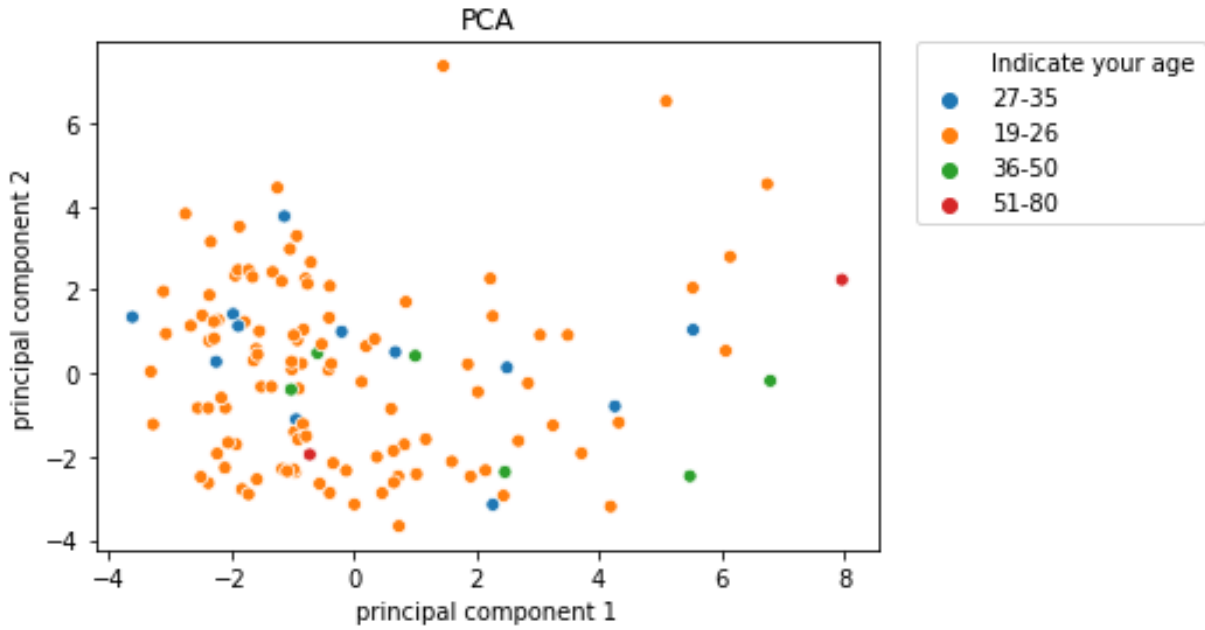


Figure 6-26: Principal component analysis, age

This is a first representation of the questionnaire data in a low-dimensional plot made using Principal Component Analysis (PCA). PCA is an orthogonal linear transformation of data coordinates into new coordinates such that the first coordinate is the coordinate with the largest variance of the data, the second coordinate is the coordinate with the second largest variance and so on.

Given the data matrix  $X_{(n \times m)}$  where its rows represent the questionnaires and its columns represent the questions of the questionnaire, the goal is to find a coordinate transformation matrix  $W_{(m \times m)}$  where each unit row vector  $w_j = (w_{j1}, \dots, w_{jm})$  transforms each row vector  $x_i$  of  $X$  into the new coordinate system  $T = XW$  (where each element is defined by  $t_{ij} = x_i w_j^T$ ).

The first component of  $T$  has to maximize the variance, so  $w_1$  has to satisfy

$$w_1 = \operatorname{argmax}_{\|w\|=1} \sum_i t_{i1}^2 = \operatorname{argmax}_{\|w\|=1} \sum_i (x_i w^T)^2$$

The  $k$ -th component of  $T$  can be computed subtracting the first  $k - 1$  components from  $X$ ,

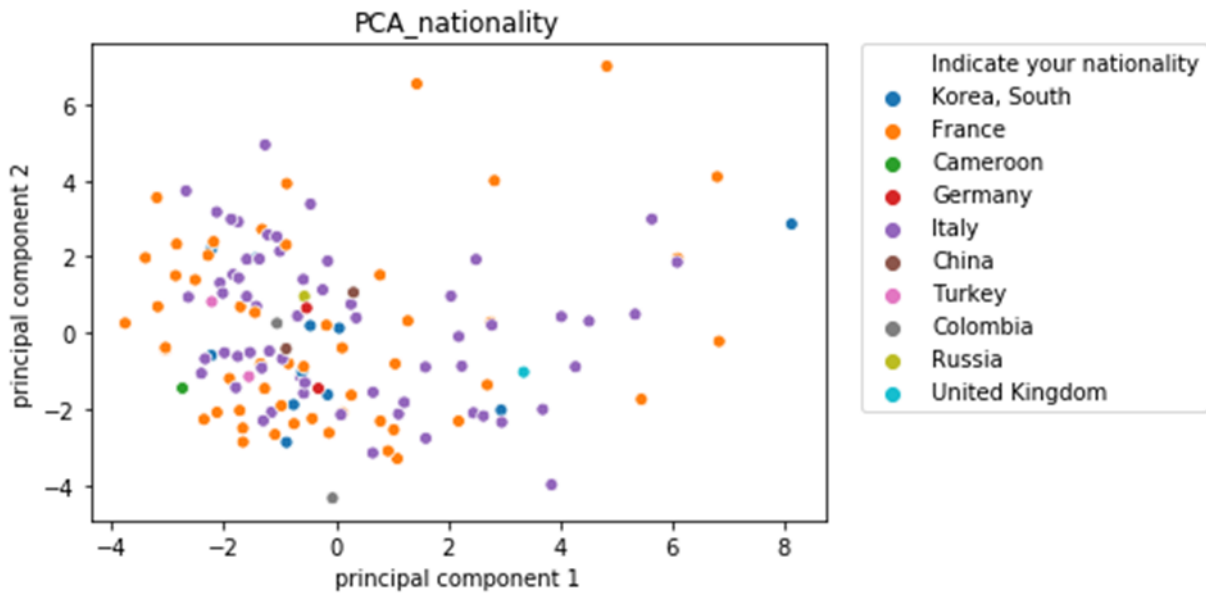
$$\widehat{X}_k = X - \sum_{s=1}^{k-1} X w_s w_s^T,$$

and then repeating the procedure above.

For this representation we took only the first two components, in order to plot the questionnaire data in a 2D-graph.

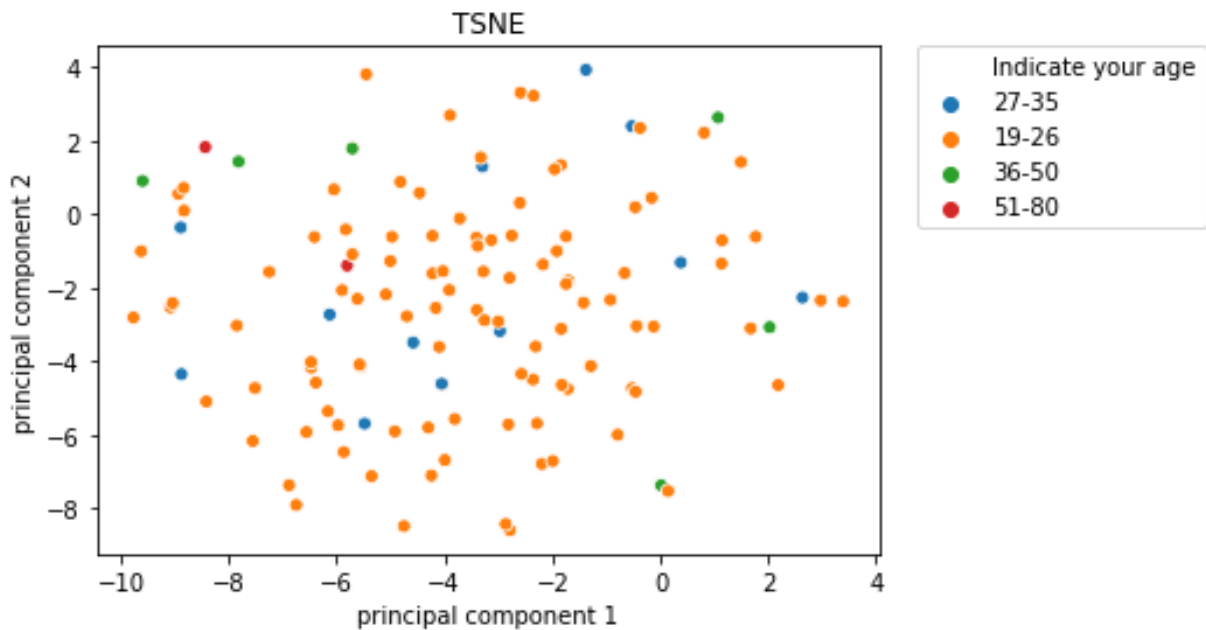
From the plot reported in Figure 6-26 it is impossible to see possible clusters related to the age of the participants. Regarding the two largest groups (19-26 and 27-35) it is possible to see that

the data are all scattered in the plot and there is no sign of clusters (at least using PCA, there still might be clusters in the original high-dimensional space).



**Figure 6-27: Principal Component analysis, nationality**

Labelling data with their nationality does not show any cluster, since also in this case there is no visual separation of data according to their nationality, as detailed in Figure 6-27. As before, this does not imply that there is no cluster in the high-dimensional space.



**Figure 6-28: TSNE algorithm results, age labelling**

This second set of plots is made using the t-distributed stochastic neighbour embedding (t-SNE) algorithm. This dimensionality reduction algorithm represents similar data by close points and dissimilar data as distant points. Given a dataset  $X = \{x_1, \dots, x_N\}$ , the t-SNE algorithm computes probabilities  $p_{ij}$ , proportional to the similarities of data  $x_i$  and  $x_j$ :

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}, \quad p_{j|i} = \frac{e^{-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}}}{\sum_{k \neq i} (e^{-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}})}$$

$$p_{ij} = 0 \quad \text{if } i = j.$$

where  $p_{j|i}$  represents the probability that  $x_i$  would pick  $x_j$  as its neighbour considering a Gaussian distribution centered at  $x_i$ .

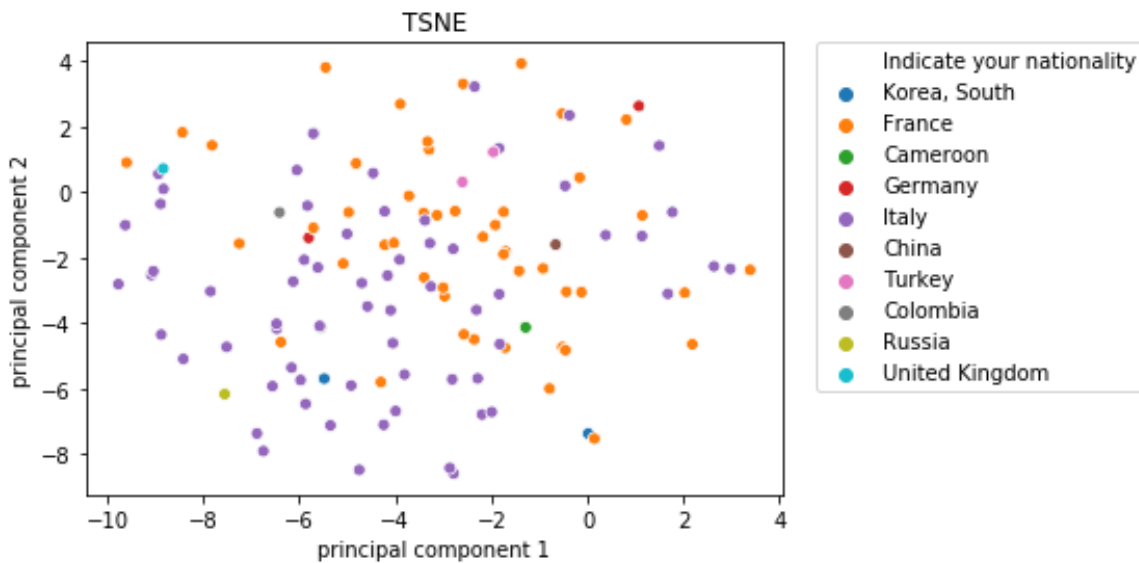
The t-SNE algorithm has to find an  $l$ -dimensional map (in our case  $l = 2$ , since we want a 2D map)  $y_1, \dots, y_N$  with similarities  $q_{ij}$  in the new low-dimensional space close as much as possible to  $p_{ij}$ .

$$q_{ij} = \frac{1}{1 + \|y_i - y_j\|^2} \frac{1}{\sum_{k \neq i} \frac{1}{1 + \|y_i - y_k\|^2}}$$

This can be done by minimizing the Kullback-Leibler divergence of distribution  $Q$  and  $P$ :

$$KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

From the t-SNE representation with ages it is not possible to see any cluster, since also in this case different ages are spread in the plot.

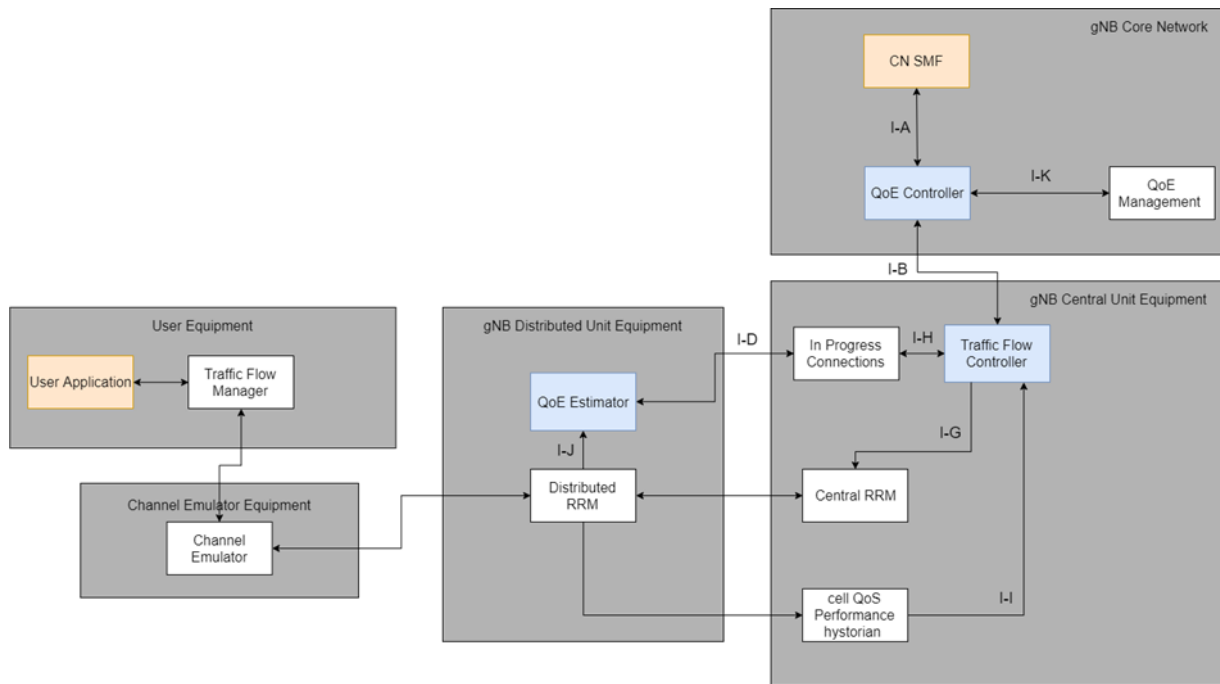


**Figure 6-29: TSNE algorithm results, nationality labelling**

In the second t-SNE plot, instead, is possible to notice a data separation by nationality. In fact almost all the Italian data are in the bottom-left part of the plot, while almost all the French data are in the top-right part of the plot. This could be interesting because the nationality can be a promising factor to cluster users.

## 7 Details of the multi-connectivity interfaces

This section represents an update of the interfaces described in D4.1, with a focus on the interfaces that will be deployed in the project PoC, whereas D4.1 preliminary defined the complete set of interfaces among all the functionalities related to multi-connectivity.

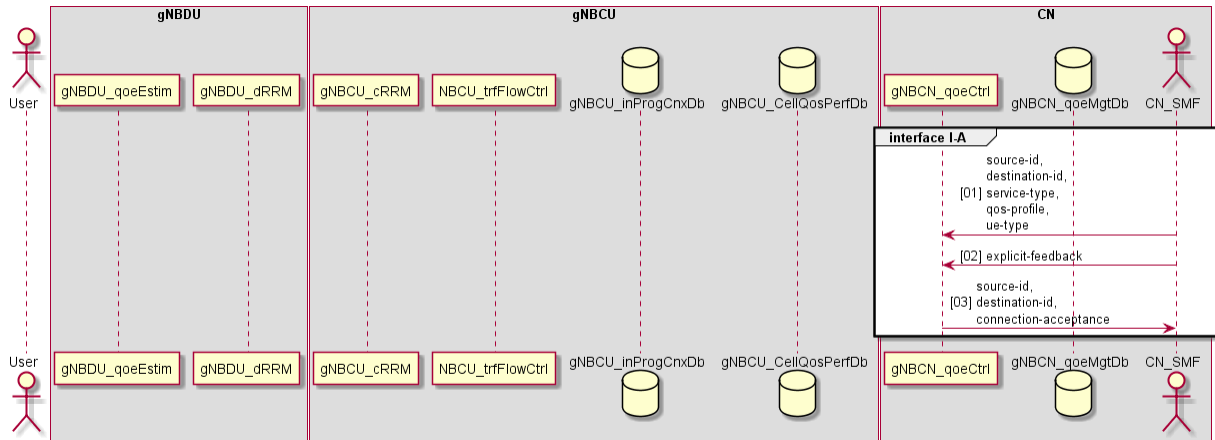


**Figure 7-1: Preliminary architecture of the project EU-PoC**

In this section, a preliminary JSON-like description of the interfaces related to the modules developed in WP4 is presented, together with their associated UML flow-charts. In this regard, Figure 7-1 represents a preliminary version of the architecture of the European PoC, in which the modules related to multi-connectivity and QoE management are highlighted in blue.

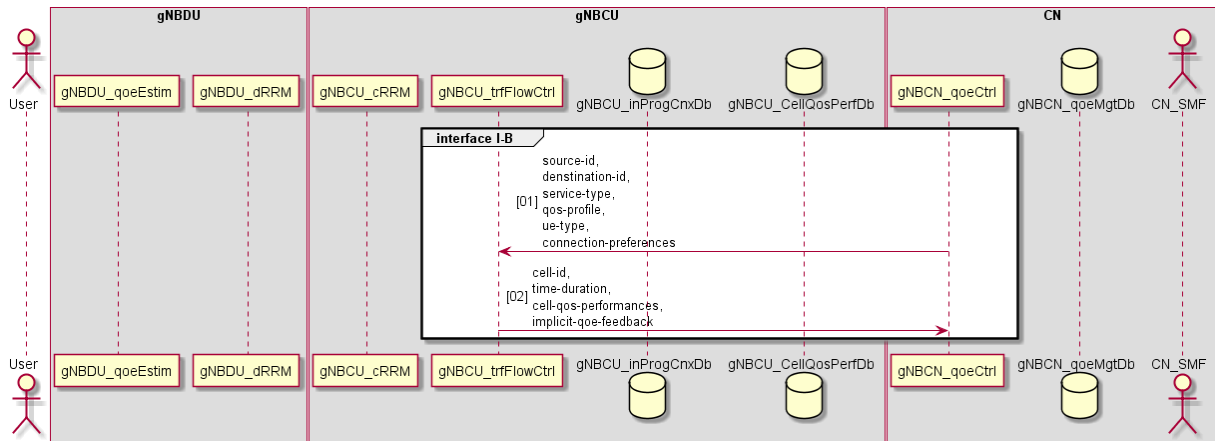
A refined and more detailed discussion on the PoC and its interfaces will be included in the upcoming deliverables of WP5.

**7.1 UML flow chart of the multi-connectivity interfaces**



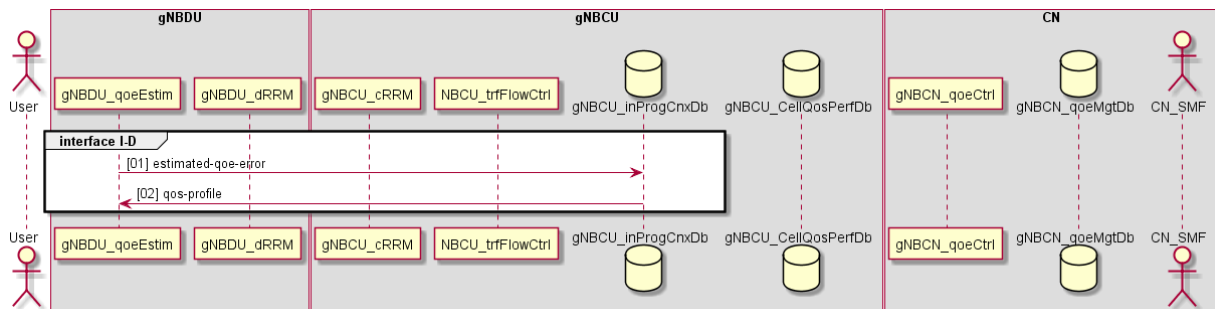
**Figure 7-2: Interface IA flow chart**

Interface IA is used by the core network to notify the QoE controller that there is a new user equipment connection to be managed.



**Figure 7-3: Interface IB flow chart**

Interface IB is used by the QoE controller to communicate to the traffic flow controller the new connection established together with the connection preferences it estimates for the UE; in such a way the traffic flow controller will be able to split the UE data in the various paths accordingly with the connection preferences.



**Figure 7-4: Interface ID flow chart**

Interface ID is user by the QoE estimator in order to store the estimated QoE error of the current connection in the InProgressConnections Database (DB).

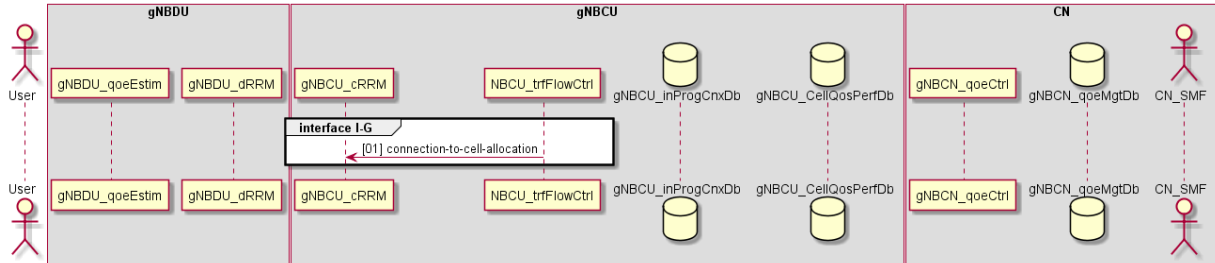


Figure 7-5: Interface IG flow chart

Interface IG is used by the traffic flow controller to instruct the central Radio Resource Management (cRRM) on the percentage of traffic relevant to the considered connection that has to be managed by each cell of the gNB-CU area.

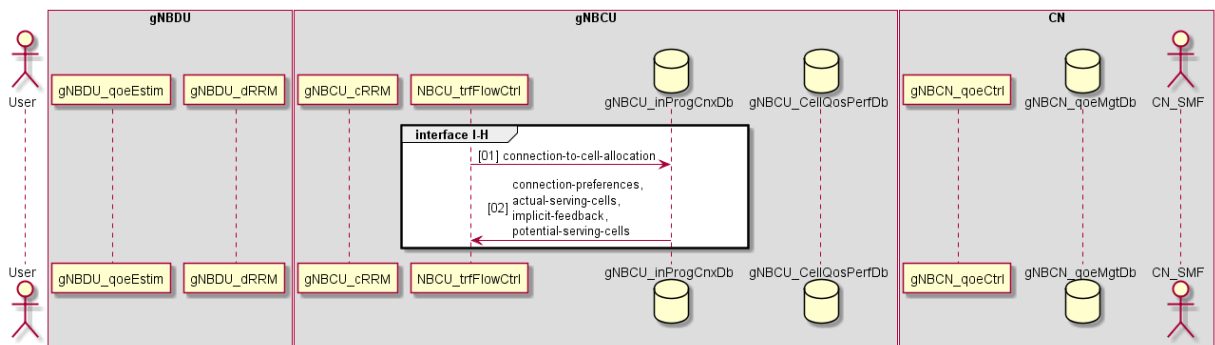


Figure 7-6: Interface IH flow chart

Interface IH is used by the traffic flow controller in order to store the data it has just sent to the cRRM (interface IG) in the InProgressConnections DB and to get the potential serving cells for the current connection.

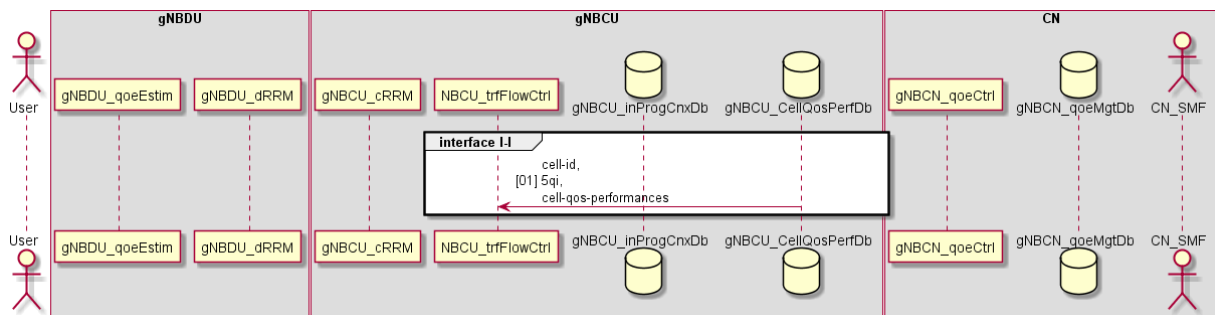


Figure 7-7: Interface II flow chart

Interface II is used by the traffic flow controller to read the actual QoS parameters measured in the current connection from the CellQoSPerformances DB.

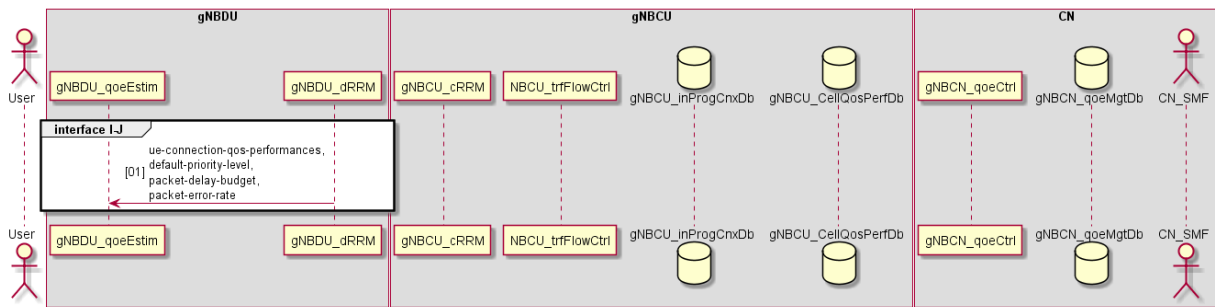


Figure 7-8: Interface IJ flow chart

Interface IJ is used by the QoE estimator in order to get QoS measurements from the distributed Radio Resource Management (dRRM) that will be used to compute the estimated QoE error.

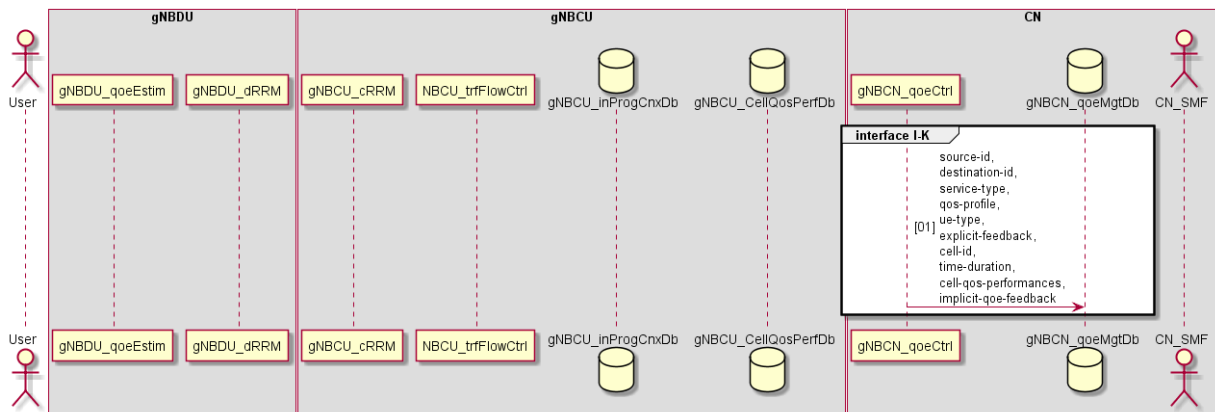


Figure 7-9: Interface IK flow chart

Interface IK is used by the QoE controller to store the information on the current connection given both by the core network (interface IA) and by the traffic flow controller (interface IB).

The remaining interfaces are not developed in WP4 and will be fully detailed in the deliverables of WP5, together with the updated version of the ones presented in this document.



## 8 Conclusions and future works

The present document reported the intermediate results of WP4 as a whole and in particular of T4.2, together with the refinement of the activities related to the multi-connectivity architecture.

The set of algorithms developed for multi-connectivity was detailed and evaluated. The document highlighted the different scopes and characteristics of the presented algorithms with the aim of driving the algorithm selection for the PoC of the project. Indeed, their testing shall now continue in the remaining activities of T4.2 and T4.3 aimed at selecting the most suitable solutions for the PoC.

Furthermore, this document presented the open source network simulator, developed from scratch with the aim of testing multi-RAT algorithms in the context of 5G communications. The development of the simulator will continue in the remaining months of WP4, with the goal of releasing a stable version that may gain traction in the research community. Additional activities are also scheduled to complete the QoE sensitivity analysis conducted on the basis of the project questionnaire.

The selection and tailoring of one or more of the algorithms to the project PoC will be the main focus of D4.3 and the remaining activities of WP4 as a whole, and for this reason the partners are working in strict cooperation with WP5 to integrate the functionalities of WP4 into the testbeds of the projects and the 8k video streaming platform.

## 9 References

- [1] M. Dryjanski and M. Szydelko, "A unified traffic steering framework for LTE radio access network coordination," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 84–92, Jul. 2016, doi: 10.1109/MCOM.2016.7509383.
- [2] A. Al Sabbagh, R. Braun, and M. Abolhasan, "A comprehensive survey on rat selection algorithms for heterogeneous networks," *World Acad. Sci. Eng. Technol.*, vol. 73, pp. 141–145, 2011.
- [3] L. Wang and G.-S. G. S. Kuo, "Mathematical Modeling for Network Selection in Heterogeneous Wireless Networks — A Tutorial," *IEEE Commun. Surv. Tutorials*, vol. 15, no. 1, pp. 271–292, 2013, doi: 10.1109/SURV.2012.010912.00044.
- [4] 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.
- [5] L. Hui, W. Ma, and S. Zhai, "A novel approach for radio resource management in multi-dimensional heterogeneous 5G networks," *J. Commun. Inf. Networks*, vol. 1, no. 2, pp. 77–83, Aug. 2016, doi: 10.1007/BF03391559.
- [6] 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.*, vol. 23, no. 6, pp. 40–46, Dec. 2016, doi: 10.1109/MWC.2016.1600059WC.
- [7] A. Wilson, A. Lenaghan, and R. Malyan, "Optimising Wireless Access Network Selection to Maintain QoS in Heterogeneous Wireless Environments," in *International Symposium on Wireless Personal Multimedia Communications 2005 (WPMC 2005)*, 2005, pp. 1236–1240.
- [8] J. Antoniou and A. Pitsillides, "4G Converged Environment: Modeling Network Selection as a Game," in *2007 16th IST Mobile and Wireless Communications Summit*, Jul. 2007, pp. 1–5, doi: 10.1109/ISTMWC.2007.4299242.
- [9] 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, p. 6, doi: 10.4108/ICST.VALUETOOLS2008.4341.
- [10] X. Gelabert, J. Perez-Romero, O. Sallent, and R. Agustí, "A Markovian Approach to Radio Access Technology Selection in Heterogeneous Multiaccess/Multiservice Wireless Networks," *IEEE Trans. Mob. Comput.*, vol. 7, no. 10, pp. 1257–1270, Oct. 2008, doi: 10.1109/TMC.2008.50.
- [11] N. Vučević, J. Pérez-Romero, O. Sallent, and R. Agustí, "Reinforcement learning for joint radio resource management in LTE-UMTS scenarios," *Comput. Networks*, vol. 55, no. 7, pp. 1487–1497, May 2011, doi: 10.1016/J.COMNET.2010.12.029.
- [12] S. Kalyanasundaram, E. K. P. Chong, and N. B. Shroff, "Optimal resource allocation in multi-class networks with user-specified utility functions," *Comput. Networks*, vol. 38, no. 5, pp. 613–630, Apr. 2002, doi: 10.1016/S1389-1286(01)00275-4.
- [13] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction*. MIT Press, Cambridge, MA, 1998.
- [14] D. P. Bertsekas, *Dynamic Programming deterministic and stochastic models*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1987.
- [15] M. L. Puterman, *Markov Decision Processes*. New York: John Wiley & Sons, Inc., 1994.
- [16] H. Kameda, J. Li, C. Kim, and Y. Zhang, *Optimal Load Balancing in Distributed Computer Systems*. London: Springer London, 1997.

- [17] 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, doi: 10.1287/moor.1040.0098.
- [18] 5GPPP Architecture Working Group, "5GPPP Architecture Working Group View on 5G Architecture," 2017.
- [19] A. Morgado, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, "A survey of 5G technologies: regulatory, standardization and industrial perspectives," *Digit. Commun. Networks*, vol. 4, no. 2, pp. 87–97, Apr. 2018, doi: 10.1016/J.DCAN.2017.09.010.
- [20] I. Da Silva *et al.*, "Tight Integration of New 5G Air Interface and LTE to Fulfill 5G Requirements," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, May 2015, pp. 1–5, doi: 10.1109/VTCSpring.2015.7146134.
- [21] J.-O. Kim, "Feedback-based traffic splitting for wireless terminals with multi-radio devices," *IEEE Trans. Consum. Electron.*, vol. 56, no. 2, pp. 476–482, May 2010, doi: 10.1109/TCE.2010.5505958.
- [22] A. A. Sabbagh, R. Braun, and M. Abolhasan, "A comprehensive survey on rat selection algorithms for heterogeneous networks," *World Acad. Sci. Eng. Technol.*, 2011.
- [23] Q.-T. Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine, "On utility models for access network selection in wireless heterogeneous networks," in *NOMS 2008 - 2008 IEEE Network Operations and Management Symposium*, 2008, pp. 144–151, doi: 10.1109/NOMS.2008.4575128.
- [24] Ho Chan, Pingyi Fan, and Zhigang Cao, "A utility-based network selection scheme for multiple services in heterogeneous networks," in *2005 International Conference on Wireless Networks, Communications and Mobile Computing*, vol. 2, pp. 1175–1180, doi: 10.1109/WIRLES.2005.1549578.
- [25] D. Niyato and E. Hossain, "Dynamics of Network Selection in Heterogeneous Wireless Networks: An Evolutionary Game Approach," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 2008–2017, May 2009, doi: 10.1109/TVT.2008.2004588.
- [26] T. Roughgarden and É. Tardos, "How bad is selfish routing?," *J. ACM*, vol. 49, no. 2, pp. 236–259, Mar. 2002, doi: 10.1145/506147.506153.
- [27] J. G. Wardrop, "Some Theoretical Aspects of Road Traffic Research," *Proc. Inst. Civ. Eng.*, vol. 1, no. 3, pp. 325–362, May 1952, doi: 10.1680/ipeds.1952.11259.
- [28] T. Roughgarden, *Selfish routing and the price of anarchy*. MIT Press, 2005.
- [29] H. Z. Aashtiani, H. Poorzahedy, and M. Nourinejad, "Extending Wardrop's First Principle for Capacitated Networks," *Transp. A Transp. Sci.*, pp. 1–30, 2018, doi: 10.1080/23249935.2018.1504255.
- [30] A. S. Schulz and N. S. Moses, "On the Performance of User Equilibria in Traffic Networks," in *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2003, pp. 86–87.
- [31] 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. 2004, doi: 10.1287/opre.1030.0091.
- [32] H. Kameda, J. Li, C. Kim, and Y. Zhang, "Optimal load balancing in distributed computer systems," 2012.
- [33] D. Grosu and A. T. Chronopoulos, "Noncooperative load balancing in distributed systems," *J. Parallel Distrib. Comput.*, vol. 65, no. 9, pp. 1022–1034, Sep. 2005, doi: 10.1016/j.jpdc.2005.05.001.
- [34] E. Altman, H. Kameda, and Y. Hosokawa, "Nash Equilibria in Load Balancing in

- Distributed Computer Systems,” *Int. Game Theory Rev.*, vol. 04, no. 02, pp. 91–100, Jun. 2002, doi: 10.1142/s0219198902000574.
- [35] M. Beckmann, C. B. McGuire, and C. B. Winsten, “Studies in the Economics of Transportation,” 1956.
- [36] G. Como, K. Savla, D. Acemoglu, M. A. Dahleh, and E. Frazzoli, “Robust Distributed Routing in Dynamical Networks—Part I: Locally Responsive Policies and Weak Resilience,” *IEEE Trans. Automat. Contr.*, vol. 58, no. 2, pp. 317–332, Feb. 2013, doi: 10.1109/TAC.2012.2209951.
- [37] G. Como, K. Savla, D. Acemoglu, M. A. Dahleh, and E. Frazzoli, “Robust Distributed Routing in Dynamical Networks—Part II: Strong Resilience, Equilibrium Selection and Cascaded Failures,” *IEEE Trans. Automat. Contr.*, vol. 58, no. 2, pp. 333–348, Feb. 2013, doi: 10.1109/TAC.2012.2209975.
- [38] V. S. Borkar and P. R. Kumar, “Dynamic Cesaro-Wardrop equilibration in networks,” *IEEE Trans. Automat. Contr.*, vol. 48, no. 3, pp. 382–396, Mar. 2003, doi: 10.1109/TAC.2003.809145.
- [39] D. Barth, O. Bournez, O. Boussaton, and J. Cohen, “Distributed Learning of Wardrop Equilibria,” in *Unconventional Computing*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 19–32.
- [40] S. Fischer, L. Olbrich, and B. Vöcking, “Approximating Wardrop equilibria with finitely many agents,” *Distrib. Comput.*, vol. 21, no. 2, pp. 129–139, Jul. 2008, doi: 10.1007/s00446-008-0057-1.
- [41] S. Fischer and B. Vöcking, “Adaptive routing with stale information,” *Theor. Comput. Sci.*, vol. 410, no. 36, pp. 3357–3371, Aug. 2009, doi: 10.1016/J.TCS.2008.01.055.
- [42] S. Fischer, H. Räcke, and B. Vöcking, “Fast Convergence to Wardrop Equilibria by Adaptive Sampling Methods,” *SIAM J. Comput.*, vol. 39, no. 8, pp. 3700–3735, 2010, doi: 10.1137/090746720.
- [43] A. Pietrabissa *et al.*, “Lyapunov-Based Design of a Distributed Wardrop Load-Balancing Algorithm With Application to Software-Defined Networking,” *IEEE Trans. Control Syst. Technol.*, pp. 1–13, 2018, doi: 10.1109/TCST.2018.2842044.
- [44] V. Sundarapandian, “An invariance principle for discrete-time nonlinear systems,” *Appl. Math. Lett.*, vol. 16, no. 1, pp. 85–91, 2003, doi: 10.1016/S0893-9659(02)00148-9.
- [45] W. Mei and F. Bullo, “LaSalle Invariance Principle for Discrete-time Dynamical Systems: A Concise and Self-contained Tutorial,” *arXiv Prepr. arXiv1710.03710*, Oct. 2017.
- [46] Emanuele De Santis *et al.*, “5G-ALLSTAR Wireless Network Simulator,” 2020. .
- [47] N. Tabia, A. Gondran, O. Baala, and A. Caminada, “Interference model and evaluation in LTE networks,” in *2011 4th Joint IFIP Wireless and Mobile Networking Conference (WMNC 2011)*, Oct. 2011, pp. 1–6, doi: 10.1109/WMNC.2011.6097237.
- [48] TSGR, “TS 138 101-1 - V15.3.0 - 5G; NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone (3GPP TS 38.101-1 version 15.3.0 Release 15),” 2018.
- [49] TSGR, “TS 138 101-2 - V15.3.0 - 5G; NR; User Equipment (UE) radio transmission and reception; Part 2: Range 2 Standalone (3GPP TS 38.101-2 version 15.3.0 Release 15),” 2018.
- [50] G. Maral and M. Bousquet, *Satellite Communications Systems - Systems, Techniques and Technologies*. Chichester, UK: John Wiley & Sons, Ltd, 2009.
- [51] H. Fenech, S. Amos, A. Tomatis, and V. Soumholphakdy, “High throughput satellite

- systems: An analytical approach," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 192–202, Jan. 2015, doi: 10.1109/TAES.2014.130450.
- [52] R. Swinford and B. Grau, "High Throughput Satellites: Delivering future capacity needs," *Arthur D. Little*, 2015. .
- [53] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Machine Learning Proceedings 1994*, 1994, pp. 157–163.
- [54] R. Sutton and A. Barto, *Reinforcement learning: An introduction*. MIT Press, 2018.
- [55] M. L. Littman, "Friend-or-Foe Q-learning in General-Sum Games," 2003.
- [56] F. Liberati *et al.*, "Stochastic and exact methods for service mapping in virtualized network infrastructures," *Int. J. Netw. Manag.*, vol. 27, no. 6, p. e1985, Nov. 2017, doi: 10.1002/nem.1985.
- [57] R. K. Ahuja, "Minimax linear programming problem," *Oper. Res. Lett.*, vol. 4, no. 3, pp. 131–134, 1985, doi: 10.1016/0167-6377(85)90017-3.
- [58] H. Khalili, P. S. Khodashenas, C. Fernandez, D. Guija, K. Liolis, C. Politis, G Atkinson, J. Cahill, R. King, M. Kavanagh, B. T. Jou, and O. Vidal, "Benefits and Challenges of Software Defined Satellite-5G Communication," in Proc. of the 15th Annual Conference on Wireless On-demand Network Systems and Services (WONS), Wengen, Switzerland, 2019, pp. 1-4.
- [59] H. Khalili, P. S. Khodashenas, D. Guija, and S. Siddiqui, "Introducing Terrestrial Satellite Resource Orchestration Layer", in the 21st International Conference on Proc. of Transparent Optical Networks (ICTON) 2019, pp. 1-4, 2019.
- [60] 3GPP TR 28.808, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Study on management and orchestration aspects with integrated satellite components in a 5G network (Release 16)" Oct. 2019
- [61] G. Giambene, S. Kota, and P. Pillai, "Satellite-5G Integration: A Network Perspective," *IEEE Network*, vol. 32, no. 5, pp. 25-31, Sept./Oct. 2018.
- [62] ITU-R S.2460-0 "Key elements for integration of satellite systems into Next Generation Access Technologies" Jul. 2019
- [63] 3rd Generation Partnership Project: Technical Specification Group Radio Access Network: Study LTE-Based V2X Services: (Release 14), Standard 3GPP TR 36.885 V2.0.0, Jun. 2016.
- [64] Saaty, Thomas L. "Decision making with the analytic hierarchy process." *International journal of services sciences* 1.1 (2008): 83-98.
- [65] Saaty, Thomas L. What is the analytic hierarchy process? In: *Mathematical models for decision support*. Springer, Berlin, Heidelberg, 1988. p. 109-121.
- [66] M Tamer Basar and Georges Zaccour. *Handbook of dynamic game theory*. English (US). Springer International Publishing, Aug. 2018. isbn: 9783319443737. doi:10.1007/978-3-319-44374-4.
- [67] Bressan, Alberto. "Noncooperative differential games. a tutorial." Department of Mathematics, Penn State University (2010).
- [68] Geering, Hans P. *Optimal control with engineering applications*. Springer, 2007.
- [69] Bonnans, Joseph Frédéric. *Convex and stochastic optimization*. Springer, 2019.
- [70] 3GPP TR 38.913, Study on scenarios and requirements for next generation access technologies
- [71] 3GPP TR 38.901 Study on channel model for frequencies from 0.5 to 100 GHz. Status:

Under change control

- [72] 3GPP TR 25.942 Radio Frequency (RF) system scenarios
  - [73] 3GPP TR 38.811 Study on New Radio (NR) to support non-terrestrial networks
  - [74] 3GPP TR 38.104, Table 5.3.3-1: Minimum guardband [kHz] (FR1) and Table: 5.3.3-2: Minimum guardband [kHz] (FR2)
  - [75] [http://www.cost.eu/domains\\_actions/ict/Actions/231](http://www.cost.eu/domains_actions/ict/Actions/231)
  - [76] [http://www.lx.it.pt/cost231/final\\_report.htm](http://www.lx.it.pt/cost231/final_report.htm)
-



## ANNEX Preliminary JSON-like multi-connectivity interface description

```

----- message laTbdMsg1:
{
  "msg-type":{
    "type": "string",
    "value": "IA-connect-request"
  },
  "msg-src":{
    "type": "string",
    "value": "CN_SMF"
  },
  "msg-dst":{
    "type": "string",
    "value": "gNBCN_qoeCtrl"
  },
  "parameters":{
    {
      "source-id":{
        "type": "ip-address",
        "address": "192.168.1.1" #CRAT:this indicates the address format to be used
      }
    },
    {
      "destination-id":{
        "type": "ip-address",
        "address": "192.168.1.1" #CRAT:this indicates the address format to be used
      }
    },
    {
      "service-type":{
        "type": "unsigned integer16",
        "description": "Service type ID" # cde is there a description by previous value?
      }
    },
    {
      "qos-profile":{
        "5qi":{
          "type": "unsigned integer16",
          "enum": [
            1, 2, 3, 4,
            65, 66, 67, 75,
            5, 6, 7, 8, 9,
            69, 70, 79, 80,
            81, 82, 83, 84,
            85
          ],
          "description": "5G QoS Indicator"
        },
        "arp":{
          "type": "unsigned integer16",
          "max_value": 15,
          "min_value": 1,
          "description": "Allocation and Retention Priority - 1 highest priority, 15 lowest priority"
        },
        "rqa":{
          "type": "boolean",
          "optional": "true",
          "description": "Reflective QoS Attribute"
        },
        "gfbr_ul":{
          "type": "unsigned integer64",
          "description": "Guaranteed Flow Bit Rate Uplink"
        },
        "gfbr_dl":{
          "type": "unsigned integer64",
          "description": "Guaranteed Flow Bit Rate Downlink"
        },
        "mfbr_ul":{
          "type": "unsigned integer64",
          "description": "Maximum Flow Bit Rate Uplink"
        },
        "mfbr_dl":{
          "type": "unsigned integer64",
          "description": "Maximum Flow Bit Rate Downlink"
        },
        "mplr_ul":{
          "type": "float64",
          "min_value": 0,
          "max_value": 1,
          "description": "Maximum Packet Loss Rate Uplink"
        },
        "mplr_dl":{
          "type": "float64",
          "min_value": 0,
          "max_value": 1,
          "description": "Maximum Packet Loss Rate Downlink"
        }
      }
    },
    {
      "ue-type":{
        "type": "unsigned integer16",
        "description": "UE type ID" # cde is there a description by previous value?
      }
    }
  }
}

```

```

}
----- message laTbdMsg2:
{
  "msg-type":{
    "type": "string",
    "value": "IA-connect-acceptance"
  },
  "msg-src":{
    "type": "string",
    "value": "CN_SMF"
  },
  "msg-dst":{
    "type": "string",
    "value": "gNBCN_qoeCtrl"
  },
  "parameters":{
    {
      "explicit-feedback":{
        "type": "float64",
        "description": "feedback related to the Perceived QoE directly provided by the users involved in the connection"
      }
    }
  }
}
----- message laTbdMsg3:
{
  "msg-type":{
    "type": "string",
    "value": "IA-connect-preference"
  },
  "msg-src":{
    "type": "string",
    "value": "gNBCN_qoeCtrl"
  },
  "msg-dst":{
    "type": "string",
    "value": "CN_SMF"
  },
  "parameters":{
    {
      "source-id":{
        "type": "ip-address",
        "address": "192.168.1.1" #CRAT:this indicates the address format to be used
      }
    },
    {
      "destination-id":{
        "type": "ip-address",
        "address": "192.168.1.1" #CRAT:this indicates the address format to be used
      }
    },
    {
      "connection-acceptance":{
        "type": "boolean"
        "description": "OK (1) if connection is accepted, KO (0) otherwise"
      }
    }
  }
}
----- message lbTbdMsg1:
{
  "msg-type":{
    "type": "string",
    "value": "IB-connect-preference"
  },
  "msg-src":{
    "type": "string",
    "value": "gNBCN_qoeCtrl"
  },
  "msg-dst":{
    "type": "string",
    "value": "gNBCU_trfFlowCtrl"
  },
  "parameters":{
    {
      "source-id":{
        "type": "ip-address",
        "address": "192.168.1.1" #CRAT:this indicates the address format to be used
      }
    },
    {
      "destination-id":{
        "type": "ip-address",
        "address": "192.168.1.1" #CRAT:this indicates the address format to be used
      }
    },
    {
      "service-type":{
        "type": "unsigned integer16",
        "description": "Service type ID" # cde is there a description by previous value?
      }
    }
  },
  {
    "qos-profile":{
      "5qi":{
        "type": "unsigned integer16",
        "enum": [
          1, 2, 3, 4,
          65, 66, 67, 75,
          5, 6, 7, 8, 9,
          69, 70, 79, 80,

```



```

81, 82, 83, 84,
85
},
"description": "5G QoS Indicator"
},
"arp":{
"type": "unsigned integer16",
"max_value": 15,
"min_value": 1,
"description": "Allocation and Retention Priority - 1 highest priority, 15 lowest priority"
},
"rqa":{
"type": "boolean",
"optional": "true",
"description": "Reflective QoS Attribute"
},
"gfbr_ul":{
"type": "unsigned integer64",
"description": "Guaranteed Flow Bit Rate Uplink"
},
"gfbr_dl":{
"type": "unsigned integer64",
"description": "Guaranteed Flow Bit Rate Downlink"
},
"mfbr_ul":{
"type": "unsigned integer64",
"description": "Maximum Flow Bit Rate Uplink"
},
"mfbr_dl":{
"type": "unsigned integer64",
"description": "Maximum Flow Bit Rate Downlink"
},
"mplr_ul":{
"type": "float64",
"min_value": 0,
"max_value": 1,
"description": "Maximum Packet Loss Rate Uplink"
},
"mplr_dl":{
"type": "float64",
"min_value": 0,
"max_value": 1,
"description": "Maximum Packet Loss Rate Downlink"
}
}
},
{
"ue-type":{
"type": "unsigned integer16",
"description": "UE type ID" # cde is there a description by previous value?
}
},
{
"connection-preferences":{
"type": "array",
"values":{
"type": "float64",
"min_value": 0,
"max_value": 1
},
"description": "personalized connection requirements aimed at satisfying even the subjective
QoE of the user handling the connection in question (e.g. connection quality, cost, mobility, battery consumption, ...)"
}
}
}
}
}
----- message IbTbdMsg2:
{
"msg-type":{
"type": "string",
"value": "IB-connect-qoe-history"
},
"msg-src":{
"type": "string",
"value": "gNBCU_trfFlowCtrl"
},
"msg-dst":{
"type": "string",
"value": "gNBCN_qoeCtrl"
},
"parameters":{
{
"cell-id":{
"type": "unsigned integer64",
"description": "Cell identification number" # cde is there a description by previous value?
}
},
{
"time-duration":{
"type": "unsigned integer64",
"description": "duration of the time interval in which the cell has served the connection in question"
}
},
{
"cell-qos-performance":{
"qos-parameters": {
"arp":{
"type": "unsigned integer16",
"max_value": 15,
"min_value": 1,
"description": "Allocation and Retention Priority - 1 highest priority, 15 lowest priority"
}
}
}
}
}
}
}
}

```











```

"type": "boolean",
"optional": "true",
"description": "Reflective QoS Attribute"
},
"gfbr_ul":{
"type": "unsigned integer64",
"description": "Guaranteed Flow Bit Rate Uplink"
},
"gfbr_dl":{
"type": "unsigned integer64",
"description": "Guaranteed Flow Bit Rate Downlink"
},
"mfbr_ul":{
"type": "unsigned integer64",
"description": "Maximum Flow Bit Rate Uplink"
},
"mfbr_dl":{
"type": "unsigned integer64",
"description": "Maximum Flow Bit Rate Downlink"
},
"mplr_ul":{
"type": "float64",
"min_value": 0,
"max_value": 1,
"description": "Maximum Packet Loss Rate Uplink"
},
"mplr_dl":{
"type": "float64",
"min_value": 0,
"max_value": 1,
"description": "Maximum Packet Loss Rate Downlink"
}
},
{
"ue-type":{
"type": "unsigned integer16",
"description": "UE type ID" # cde is there a description by previous value?
}
}
{
"explicit-feedback":{
"type": "float64",
"description": "feedback related to the Perceived QoE directly provided by the users involved in the connection"
}
}
{
"cell-id":{
"type": "unsigned integer64",
"description": "Cell identification number" # cde is there a description by previous value?
}
}
{
"time-duration":{
"type": "unsigned integer64",
"description": "duration of the time interval in which the cell has served the connection in question"
}
},
{
"cell-qos-performance":{
"qos-parameters": {
"arp":{
"type": "unsigned integer16",
"max_value": 15,
"min_value": 1,
"description": "Allocation and Retention Priority - 1 highest priority, 15 lowest priority"
},
"rqa":{
"type": "boolean",
"optional": "true",
"description": "Reflective QoS Attribute"
},
"gfbr_ul":{
"type": "unsigned integer64",
"description": "Guaranteed Flow Bit Rate Uplink"
},
"gfbr_dl":{
"type": "unsigned integer64",
"description": "Guaranteed Flow Bit Rate Downlink"
},
"mfbr_ul":{
"type": "unsigned integer64",
"description": "Maximum Flow Bit Rate Uplink"
},
"mfbr_dl":{
"type": "unsigned integer64",
"description": "Maximum Flow Bit Rate Downlink"
},
"mplr_ul":{
"type": "float64",
"min_value": 0,
"max_value": 1,
"description": "Maximum Packet Loss Rate Uplink"
},
"mplr_dl":{
"type": "float64",
"min_value": 0,
"max_value": 1,
"description": "Maximum Packet Loss Rate Downlink"
}
}
}
}

```

```
"description": "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"  
}  
},  
{  
  "implicit-qoe-feedback":{  
    "type": "float64",  
    "description": "feedback related to the Perceived QoE computed by the QoE Estimation"  
  }  
}  
}  
}
```