Ref. Ares(2019)7930590 - 30/12/2019

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:	3131/12/2019	Version:	1.0
Start date of Project:	01/07/2018	Duration:	36 months

Deliverable D4.2
Design and simulation of the multi-RAT load balancing algorithms

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

Document Number:	H2020-EUK-815323/5G-ALLSTAR/D4.2
Document Title:	Design and simulation of the multi-RAT load balancing algorithms
Editor(s):	Alessandro Giuseppi (CRAT)
Authors:	Alessandro Giuseppi (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 subcontract agreement with ETRI: Sungoh Kwon, Syed Maaz Shahid, Seyoum Yemane Teklay
Dissemination Level:	PU
Contractual Date of Delivery:	31/12/2019
Security:	Public
Status:	Draft
Version:	1.0
File Name:	5G-ALLSTAR_D4.2_proof_revAP.docx3

Abstract

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)

Executive Summary

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, 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

List of Figures

List of Tables

List of Abbreviations

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

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

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

SN	Secondary Node
SNR	Signal to Noise Ratio
UE	User Equipment
UL	Uplink
UP	User Plane
UPF	User Plane Function
WLAN	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.

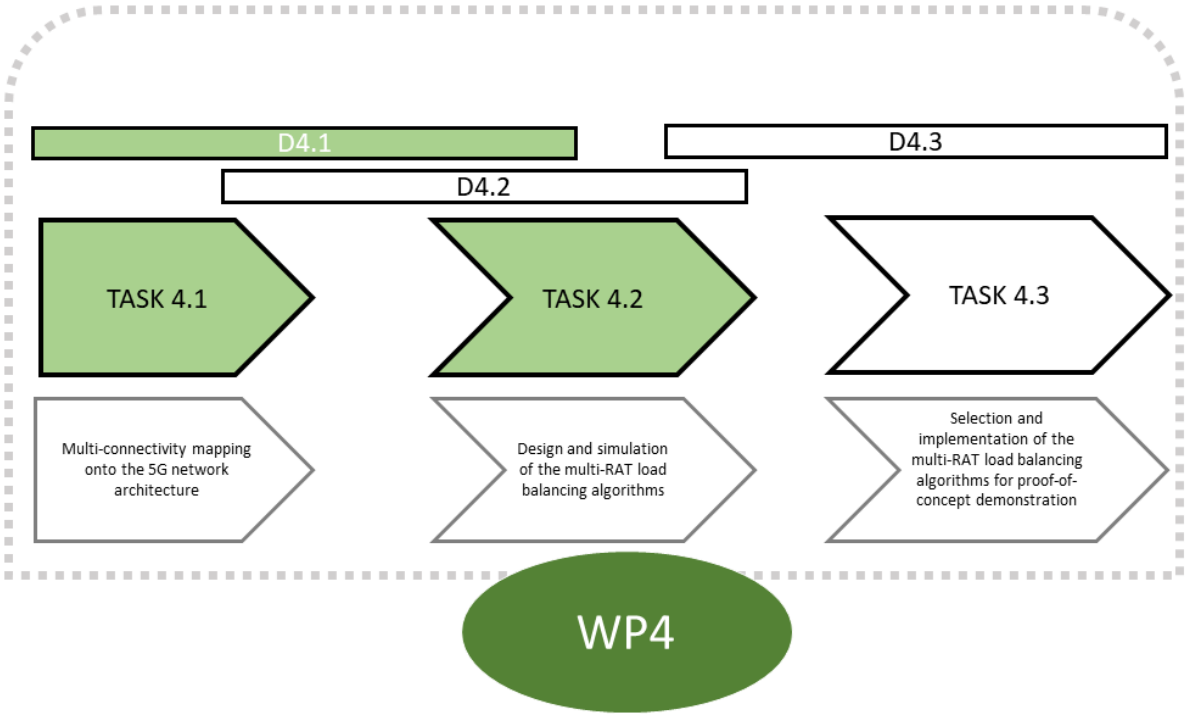


Figure 1-1: Work package 4 plan

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, proving 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.

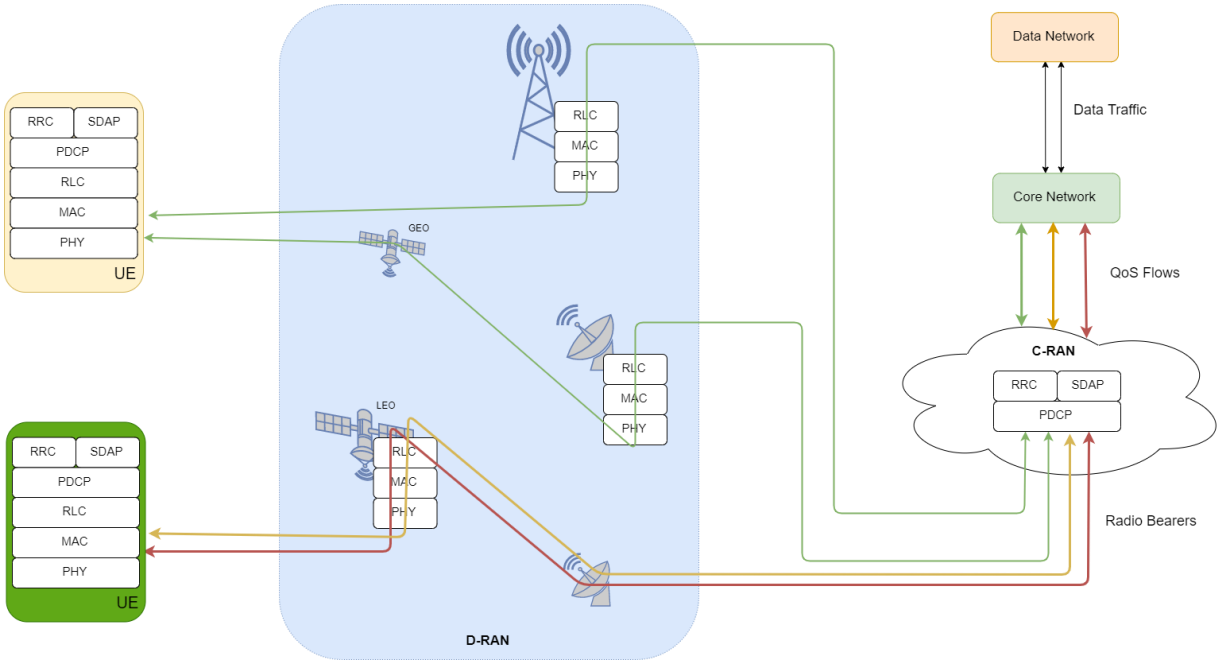


Figure 2-1: Multi-Connectivity target physical architecture

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 space 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.

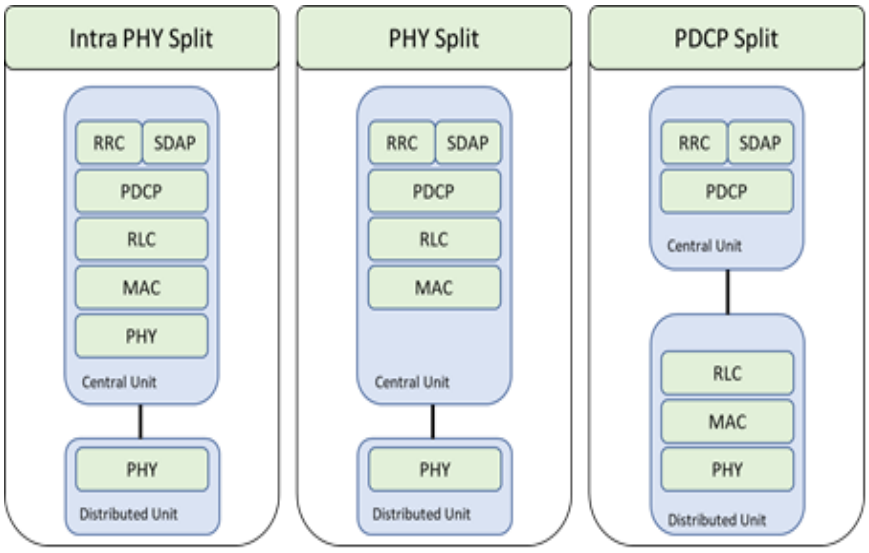


Figure 2-2: Multi-connectivity protocol split options

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.

In general, satellites cover extremely wide areas, usually beyond the reasonable scope of a single ground C-RAN. Having a satellite dedicated to a controller deployed in a C-RAN that does not have an adequate coverage would lead to the loss of one of the key advantages of satellite connections, i.e., their wide area availability. On the other hand, the same connection resources cannot be controlled by multiple C-RANs, as they may make decisions that may be in conflict (e.g., cumulatively allocating more resources than available). 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 it oversees only a portion of resources so that no conflict on their usage may arise;
- The deployment of a wide-area C-RAN in the control plane, that oversees the functioning of multiple C-RANs of the user plane taking their traffic flow control decision without requiring to directly interact with the QoS flows.

The algorithms developed are transparent to this decision, as in the first case they would only be provided with fewer network resources for the satellite RAT, while in the second case they would have at their disposal the resources coming from several C-RANs.

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 [57]:

- 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 [56] that exemplify roles of satellites in future networks.

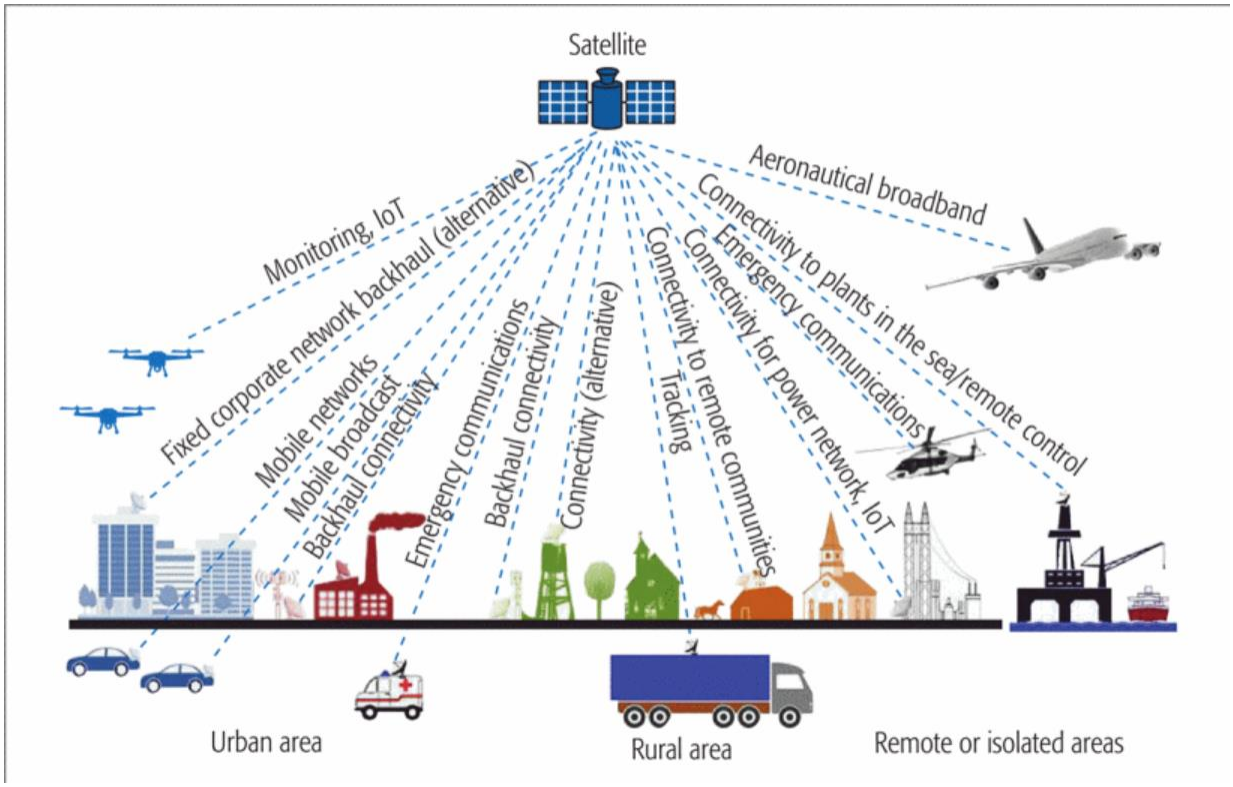


Figure 2-3: Satellite 5G use cases [56]

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.

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 [65], 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 [65] 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 [66], 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 [67] and [68] 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.

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 log10(d) (d in m)	24 + 45 log (d+20) (d in m)	32.45+20 log10(fc)+20log10(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 [53] and recommended MANO[1]-like framework for rapid provisioning of network services and management and orchestration of heterogeneous terrestrial and satellite resources in [54]. Reference management and orchestration architectures for integrating satellite components into terrestrial 5G are specified by 3GPP in [55]. These reference architectures are shown in Figure 3-1 and Figure 3-2 below.

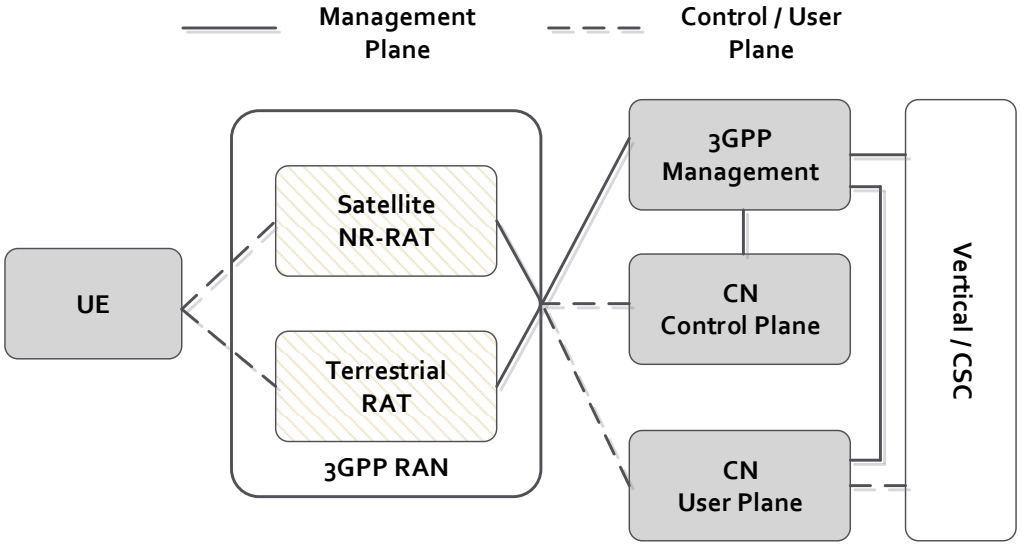


Figure 3-1: Reference architecture for the management of a satellite NR-RAT

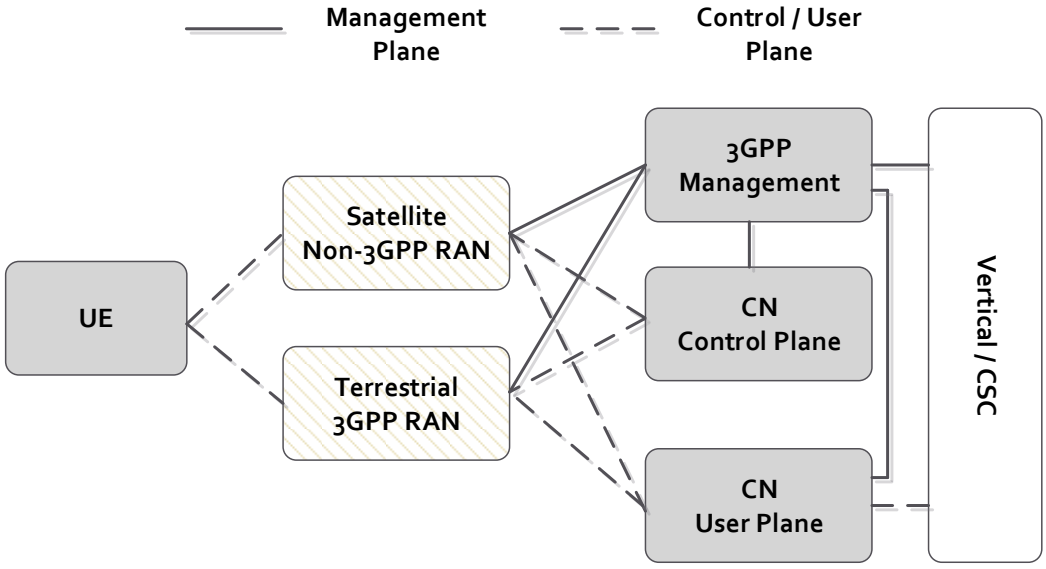


Figure 3-2: Reference architecture for the management of a non-3GPP satellite RAN

[1] https://www.etsi.org/technologies/nfv/open-source-mano

4 Proposed algorithms for multi-connectivity

This section presents the traffic flow control algorithms designed in T4.2 and their preliminary evaluation based on numerical simulations.

In particular, the first algorithm presented in Section 4.2 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 4.3 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 4.4, 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 4.5 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 4.6 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 4.7 and 4.8, 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.

4.1 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 redefined 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

4.2 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 4.2.1 presents the state of the art of traffic steering in 5G networks, as well as the proposed novelties; while Section 4.2.2 provides the required background regarding MDPs; Sections 4.2.3 and 4.2.4 present respectively the problem modelling and the network selection algorithm; while Section 4.2.5 shows some simulation results to validate the proposed algorithm.

4.2.1 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.

4.2.2 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, $A = \bigcup_{s \in S} A_s$ 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, Σ 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 memoryless) 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 $u^*: S \to A$ that maximizes, in the long run, the expected discounted reward:

$$R(u) \coloneqq E_{u,\Sigma}\{\sum_{t=0,1,\dots,\infty} \gamma^t r(s(t), a(t), s(t+1))\}, \tag{4.2.1}$$

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 Σ.

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:

$$V_u(s) \coloneqq E_u\{\sum_{t=0,1,\dots,\infty} \gamma^t \cdot r(s(t), a(t), s(t+1)) \mid s(0) = s\}, \tag{4.2.2}$$

$$Q_u(s, a) \coloneqq E_u\{\sum_{t=0,1,\dots,\infty} \gamma^t \cdot r(s(t), a(t), s(t+1)) \mid s(0) = s, a(0) = a\}, \tag{4.2.3}$$

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

$$Q\big(s(t), a(t)\big) \leftarrow \big(1 - \alpha(t)\big)Q\big(s(t), a(t)\big) + \alpha(t) \cdot \left[r\big(s(t), a(t), s(t+1)\big) + \gamma \max_{a' \in A_{s(t+1)}} Q(s(t+1), a')\right]. \quad (4.2.4)$$

In Equation (4.2.4), $\alpha(t) > 0$ is the learning rate and thus the key parameter for the algorithm convergence: if $\sum_{t=1,\dots,\infty} \alpha(t) = \infty$ and $\sum_{t=1,\dots,\infty} \big(\alpha(t)\big)^2 < \infty$, the estimate (4.2.4) converges to the optimal action-value function as $t \to \infty$ [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:

$$u(s(t)) = \underset{a \in A_s}{\operatorname{argmax}}\, Q(s(t), a)\,, s \in S. \quad (4.2.5)$$

As the estimate (4.2.4) converges to the optimal action-value function, the policy (4.2.5) converges to an optimal policy.

To guarantee a certain degree of exploration of the state space set, an ε-greedy rule is followed for the action selection: in state $s \in S$, the current best action (4.2.5) 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 ε, i.e.:

$$u(s) \leftarrow \begin{cases} \underset{a \in A(s)}{\operatorname{argmax}}\, Q(\mathrm{s}, \mathrm{a})\,, \text{with prob. } 1 - \varepsilon \\ \mathrm{rand}\{a \in A(s)\}, \text{with prob. } \varepsilon \end{cases}, s \in S. \quad (4.2.6)$$

A large value of ε 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 ε, 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 ε depends on the specific application.

4.2.3 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:

$$r_{pk}(w_{pk}) = \begin{cases} 0 & \text{if } w_{pk} < w_k^1 \\ f(w_{pk}) & \text{if } w_k^1 \le w_{pk} \le w_k^2 \\ f(w_k^2) & \text{otherwise} \end{cases}. \tag{4.2.7}$$

An example of this kind of utility function is represented in Figure 4-1, 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.

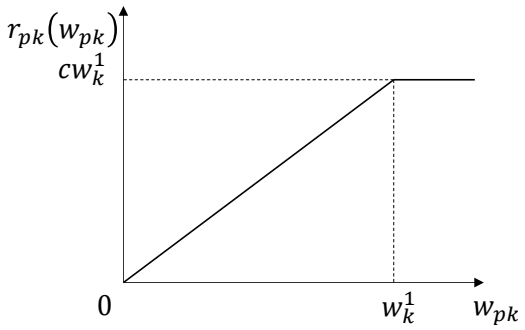


Figure 4-1: User-perceived quality of connection for service with elastic traffic

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 4-2.

$$r_{pk}(w_{pk}) = \begin{cases} 0 \text{ if } w_{pk} < w_k^1 \\ r_{pk}^1 > 0 \ \ otherwise \end{cases}, \tag{4.2.8}$$

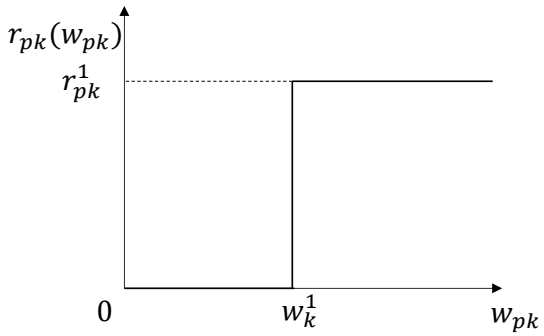


Figure 4-2: User-perceived quality of connection for service with transmission bitrate threshold

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 4-3.

$$r_{pk}(w_{pk}) = \begin{cases} 0 & \text{if } w_{pk} < w_k^1 \\ r_{pk}^1 > 0 & \text{if } w_k^1 \le w_{pk} < w_k^2 \\ r_{pk}^2 > r_{pk}^1 & \text{if } w_k^2 \le w_{pk} < w_k^3, \\ r_{pk}^c > r_{pk}^{c-1} & \text{if } w_{pk} \ge w_k^C \end{cases} \tag{4.2.9}$$

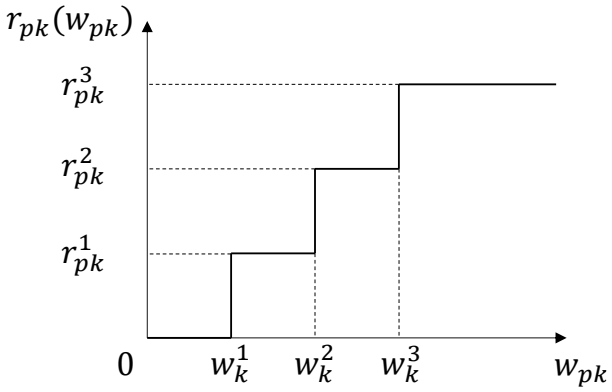


Figure 4-3: User-perceived quality of connection for service with multi-codec traffic

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,

$$\eta_p^1(s(t)) = \sum_{k\in K} n_{pk} w_k^1, p \in C, \tag{4.2.10}$$

the state space *S* is defined as

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

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 δ_{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

$$A_{s_i} = \left\{ a = (a_{pk})_{p\in P, k\in K} \middle| \begin{matrix} a_{pk} \in \{0,1\} & \text{if } s_i + \delta_{pk} \in S \\ a_{pk} = 0 & \text{if } s_i + \delta_{pk} \notin S, \forall p \in P, \forall k \in K \\ \sum_{p\in P} a_{pk} \le 1 & \forall k \in K \end{matrix} \right\}, i = 1, \dots, |S|, \tag{4.2.11}$$

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

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

in (4.2.11) 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 ν_k and their duration is exponentially distributed with mean termination frequency μ_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 ν_k;

- since a mapped service k on AP p terminates with termination frequency μ_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 $\eta_p^1\big(s(t)\big)$ 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:

$$\eta_p\big(s(t)\big) = \sum_{k\in K}\sum_{c=1,\dots,C_k} n_{pk}^c(t) w_k^c, p \in P. \tag{4.2.12}$$

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 $\eta_p^1(s)$:

$$r(s) = \sum_{p\in P}\sum_{k\in K}\sum_{c=1,\dots,C_k} n_{pk}^c(t) r_{pk}^c, \tag{4.2.13}$$

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's, exist and are finite. For our controller, the interest is in maximizing the expected discounted reward (4.2.1)[2], computed as

$$R(u) = (1-\gamma)\sum_{j=1,\dots,N} y_u\big(s_j\big) r\big(s_j\big), u \in U. \tag{4.2.14}$$

4.2.4 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 4.2.3

[2] 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 Δ_p and using $\lfloor\ \rfloor$ as the truncation operator to the next lower integer, the aggregate state set with $\left\lfloor \frac{W_p}{\Delta_p} \right\rfloor$ bitrate levels is obtained as

$$\tilde{S} = \left\{ \tilde{s} = \left(l_p\right)_{p\in P} \middle| l_p = \left\lfloor \frac{\sum_{k\in K} n_{pk} w_k^1}{\Delta_p} \right\rfloor, \text{with } \left(n_{pk}\right)_{k\in K} \text{ s.t.} \sum_{k\in K} n_{pk} w_k^1 < W_p, \forall p \in P \right\}. \quad (4.2.15)$$

Clearly, the number of states decreases as the granularities Δ_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 $\eta_p\left(\tilde{s}(t')\right) w_k^1 < W_p - w_{pk}$ and $\eta_p\left(s(t'')\right) w_k^1 > W_p - w_{pk}$, 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:

$$\tilde{A}\left(\tilde{s}(t)\right) = \left\{ \left(\tilde{a}_{pk}\right)_{p\in P, k\in K} \middle| \begin{matrix} \tilde{a}_{pk} \in \{0,1\} & \text{if } \eta_p\left(\tilde{s}(t)\right) w_k^1 < W_p - w_{pk} \\ \tilde{a}_{pk} = 0 & \text{otherwise}, \forall k \in K \\ \sum_{p\in P} \tilde{a}_{pk} \leq 1 & \forall k \in K \end{matrix} \right\}. \quad (4.2.16)$$

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

$$r(t) = \sum_{p\in P} \sum_{k\in K} \sum_{c=1,\dots,C_k} n_{pk}^c(t) r_{pk}^c, \quad (4.2.17)$$

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 (4.2.4) 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:

$$Q(\tilde{s}',\tilde{a}') = \big(1-\alpha(t)\big)Q(\tilde{s}',\tilde{a}') + \alpha(t)\left(r(t) + \left(\sum\nolimits_{n=1,\dots,\left\lfloor N_{\tilde{s}',\tilde{a}'}\right\rfloor} \gamma^n + \gamma^{\left(N_{\tilde{s}',\tilde{a}'} - \left\lfloor N_{\tilde{s}',\tilde{a}'}\right\rfloor\right)}\right)\right) \cdot \max_{\tilde{a}\in\tilde{A}(\tilde{s}(t+1))} Q(\tilde{s}(t+1),\tilde{a}) \qquad (4.2.18)$$

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 γ:

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

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 γ, as if the state-action pair was visited 2.3 times:

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

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

4.2.5 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 4-2 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.

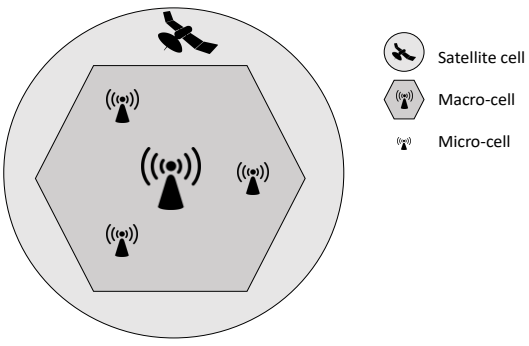


Figure 4-4: Reference scenario used in the simulations

Table 4-2: Bitrate allocation algorithm for AP $p \in P$ in state $s \in S$

Assign the minimum required bitrate to all the on-going connections of AP p:

$$n_{pk}^{c}(s) = \begin{cases} n_{pk} \text{ if } c = 1 \\ 0 \text{ otherwise} \end{cases}, k = 1, \dots, K$$

For $k = 1, \dots, K$

Set $c = 2$

While $n_{pk}^{c} = n_{pk}$ and $c \leq C_k$

Compute the bandwidth already assigned to all the connections of each service

$$\eta_p(s) = \sum_{k \in K} \sum_{c=1,\dots,C_k} n_{pk}^{c} w_k^{c}$$

Share the leftover bandwidth $W_p - \eta_p(s(t))$ in a round-robin fashion among the n_{pk}^{c-1} multi-codec PDU sessions on-going on AP p (after the round, all the con nections codec are upgraded only if $n_{pk}^{c} = n_{pk}$)

Evaluate r_{pk}^{c}

Set $c = c + 1$

End

End

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 4-4 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

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

Multi-codec services are characterized by a reward of

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

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 4-5 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.

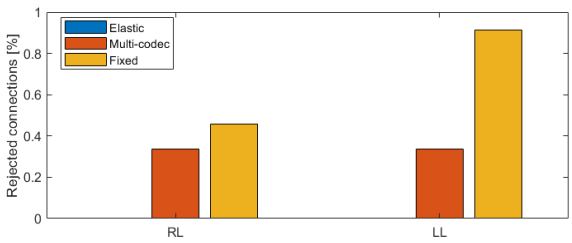


Figure 4-5: Percentage of blocked connections for the RL and the LL controllers

4.3 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 4.3.1 and 4.3.2 present the state-of-the-art on Wardrop load balancing and the proposed novelties for multi-connectivity; Section 4.3.3 details the algorithm and the convergence proof; Section 4.3.8 shows the simulation results; Section 4.3.1 motivates the choice of a distributed adversarial load-balancing algorithm to address the multi-connectivity problem in 5G networks, whereas Section 4.3.2 summarizes the works in the literature relevant to dynamic selfish routing and load balancing and the proposed innovations.

4.3.1 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.

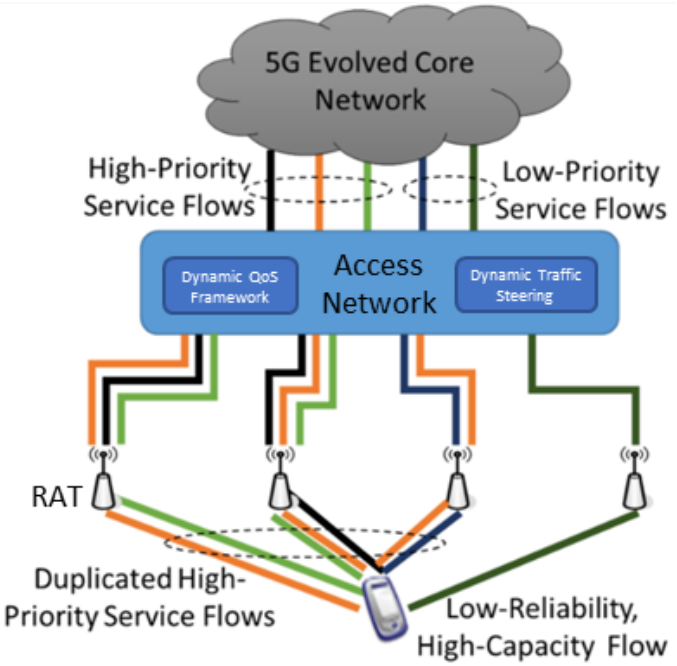


Figure 4-6: Dynamic Traffic Steering framework from [18]

The approach is hence compliant with the latest developments of the 5G architecture (see Figure 4-6), 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], [23]. Common solutions utilise the concept of utility and latency functions, as in this work, to capture the network performances [3], [24], [25]. 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], [26] 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 [25],[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.

4.3.2 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 [27] 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 [28] 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 [29]. The concept of Wardrop equilibrium has been extended to various families of networks, among which the capacitated ones [17], [30]–[32], and problems, as the load balancing one [33]–[35]. Even if Wardrop equilibria can be computed by centralized algorithms in polynomial time [36], 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 [37], [38], 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 [32], which is defined for unconstrained networks. Several works [17], [30]–[32] 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], [30]–[32], 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 [39] present an asynchronous and distributed algorithm that employs reinforcement learning to update transmission probabilities, based on an estimation of the network edges latencies. In [40], 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 [41]–[43]. In [41] and [43], 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 [42], a similar set up is analysed assuming that the information available to the agents may be stale. In [44], 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.

4.3.3 Proposed Wardrop load balancing algorithm

Section 4.3.4. 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.

4.3.4 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 λ^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.

4.3.5 Preliminaries on Wardrop and Beckmann equilibria

As anticipated in Section 4.2.1, this section further develops a well-known model for selfish routing [41], where an infinite population of agents carries an infinitesimal amount of load each and builds on the previous work [44] 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 Lipchitz constant of all the l_p^i's is denoted as $\bar{\beta} = \max_{p \in \mathcal{P}^i, i \in \mathcal{I}} \beta_p^i$.

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

$$\mathcal{X} = \left\{ \boldsymbol{x} = (x_p)_{p \in \mathcal{P}} | x_p = \sum_{i \in \mathcal{I}} x_p^i,\ x_p^i \geq 0, \forall p \in \mathcal{P}^i,\ \sum_{p \in \mathcal{P}^i} x_p^i = \lambda^i, \forall i \in \mathcal{I} \right\}, \tag{4.3.1}$$

A flow $\boldsymbol{x} \in \mathcal{X}$ is at a Wardrop equilibrium if, for each commodity, the latencies of the loaded providers are equalized, i.e., if $l_p^i(\boldsymbol{x}) \leq l_q^i(\boldsymbol{x})$ 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

$$\Phi(\boldsymbol{x}) = \sum_{i \in \mathcal{I}} \sum_{p \in \mathcal{P}^i} \int_0^{x_p^i} l_p^i(\xi) d\xi, \tag{4.3.2}$$

the Wardrop equilibria are the solutions of the optimization problem

$$\min_{\boldsymbol{x} \in \mathcal{X}} \Phi(\boldsymbol{x}). \tag{4.3.3}$$

Capacity-constrained networks are characterized by the additional capacity constraints

$$x_p \leq c_p, \forall p \in \mathcal{P}. \tag{4.3.4}$$

A flow $\boldsymbol{x} \in \mathcal{X}$ is feasible if constraints (4.3.4) hold, and the set of feasible states is defined as

$$\mathcal{X}_{CP} = \{\boldsymbol{x} \in \mathcal{X} \,|x_p \leq c_p, \forall p \in \mathcal{P}\}. \tag{4.3.5}$$

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

A flow $\boldsymbol{x} \in \mathcal{X}_{CP}$ 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 $\boldsymbol{x} \in \mathcal{X}_{CP}$ is at a Beckmann user equilibrium if $l_p^i(\boldsymbol{x}) \leq l_q^i(\boldsymbol{x})$ 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

$$\mathcal{X}_{eq} = \{\boldsymbol{x} \in \mathcal{X}_{CP} | l_p^i(\boldsymbol{x}) \leq l_q^i(\boldsymbol{x}), \forall p \in \mathcal{P}^i \text{ s.t. } x_p^i > 0, \forall q \in \mathcal{P}^i \text{ s.t. } x_q < c_q, \forall i \in \mathcal{I}\}. \tag{4.3.6}$$

Let us consider the minimization problem (4.3.3) with constraints (4.3.4), hereinafter referred to capacity-constrained problem (CP). The Beckman user equilibria [31] 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.

4.3.6 Preliminaries on Lyapunov stability

The algorithm convergence proof relies on the LaSalle invariance principle for discrete-time nonlinear systems ([45], [46]).

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

i) $\mathcal{L} \in \mathcal{C}^1$ and $\mathcal{L}(\boldsymbol{x})$ is bounded below and positive definite on $\mathcal{X}$;

iii) If $\boldsymbol{x}_{eq} \in \mathcal{X}_{eq}$, where $\mathcal{X}_{eq}$ is the set of equilibrium points, $\mathcal{L}(\boldsymbol{x}_{eq}) = 0$ and $\mathcal{L}(\boldsymbol{x}) > 0$ if $\boldsymbol{x} \notin \mathcal{X}_{eq}$;

iii) Along forward trajectories, $\mathcal{L}$ satisfies

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

Theorem 1 ([45]). Let $\mathcal{L}(\boldsymbol{x})$ be a candidate Lyapunov function for the discrete-time nonlinear system $\boldsymbol{x}[k+1] = f(\boldsymbol{x}[k])$. Then, any bounded trajectory tends to the largest invariant subset $\mathcal{M}$ contained in the set of points defined by $\Delta\mathcal{L}(\boldsymbol{x}) = 0$.

4.3.7 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

$$\boldsymbol{x}[k+1] = f(\boldsymbol{x}[k]), k = 0,1,2,\dots \tag{4.3.7}$$

with

$$x_p[k] = \sum_{i\in\mathcal{I}} x_p^i[k], \tag{4.3.8}$$

$$x_p^i[k+1] = x_p^i[k] + \tau \sum_{q\in\mathcal{P}^i}\big(r_{qp}^i[k] - r_{pq}^i[k]\big), \tag{4.3.9}$$

and with feasible initial conditions

$$\boldsymbol{x}[0] \in \mathcal{X}_{CP}. \tag{4.3.10}$$

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

The proposed controller builds on the dynamic algorithm in [44], which expresses the migration rate as

$$r_{pq}^i[k] = x_p^i[k]\sigma^i\mu_{pq}^i[k], \tag{4.3.11}$$

where σ^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 [44] for the Wardrop equilibria, approximated Beckmann user equilibria are defined.

Definition 3. The set of ε-Beckmann user equilibria is defined as

$$\mathcal{X}_{eq}^{\varepsilon} = \left\{ \boldsymbol{x} \in \mathcal{X}_{CP} | l_p^i(\boldsymbol{x}) \leq l_q^i(\boldsymbol{x}) + \varepsilon, \forall p \in \mathcal{P}^i \text{ s.t. } x_p^i > 0, \forall q \in \mathcal{P}^i \text{ s.t. } x_q \leq c_q - \frac{\varepsilon}{2\bar{\beta}}, \forall i \in \mathcal{I} \right\}. \quad (4.3.12)$$

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

Remark 1. The defined sets are such that $\mathcal{X}_{eq}^{\varepsilon} \underset{\varepsilon \to 0}{\longrightarrow} \mathcal{X}_{eq}$ 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 ε.

The tolerance ε is introduced since the kind of migration rates of equation (4.3.11) cannot guarantee convergence in the discrete-time case, however small the sampling period [42]. A flow $\boldsymbol{x} \in \mathcal{X}_{CP}$ is then at ε-Beckman equilibrium if, for each commodity, the latencies of the loaded and ε-unconstrained providers are equalized, where we define a provider $p \in \mathcal{P}^i$ to be ε-unconstrained if $x_p < c_p - \frac{\varepsilon}{2\bar{\beta}}$.

In the proposed algorithm, the migration decision is defined as

$$\mu_{pq}^i[k] = \begin{cases} 0 & \text{if } l_p^i(\boldsymbol{x}[k]) - l_q^i(\boldsymbol{x}[k]) \geq \varepsilon \text{ or if } x_q[k] \geq c_q - \frac{\varepsilon}{2\bar{\beta}}. \\ 1 & \text{otherwise} \end{cases} \quad (4.3.13)$$

The controlled system dynamics, hereafter denoted as load-balancing (4.3.20) dynamics, is then expressed by equations (4.3.9), (4.3.11), (4.3.13), with control gains set as

$$\sigma^i = \frac{\varepsilon}{2\tau\bar{\beta}\lambda^i|\mathcal{P}^i - 1||\mathcal{I}|}, \quad (4.3.14)$$

and with the tolerance set as

$$0 < \varepsilon \leq \min_{i \in \mathcal{I}} \bar{\beta}\lambda^i|\mathcal{I}|. \quad (4.3.15)$$

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

$$\left|l_p^i\left(f(\boldsymbol{x}[k])\right) - l_p^i(\boldsymbol{x}[k])\right| \leq \frac{\varepsilon}{2}. \tag{4.3.16}$$

Proof. Considering the generic commodity $i \in \mathcal{I}$, 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:

$$\begin{aligned} & l_p^i\left(x_p[k+1]\right) \\ &= l_p^i\left(x_p[k] + \tau \textstyle\sum_{q\in\mathcal{P}}\left(r_{qp}[k] - r_{pq}[k]\right)\right) \\ &\geq l_p^i\left(x_p[k] - \tau \textstyle\sum_{p\in\mathcal{P}} r_{pq}[k]\right). \end{aligned} \tag{4.3.17}$$

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

$$l_p^i(\boldsymbol{x}[k+1]) \geq l_p^i(\boldsymbol{x}[k]) - \tau\beta_p^i \textstyle\sum_{q\in\mathcal{P}} r_{pq}[k]. \tag{4.3.18}$$

Considering Equations (4.3.11) and (4.3.14), the last term of Equation (4.3.18) is written as

$$\begin{aligned} \tau\beta_p^i \textstyle\sum_{q\in\mathcal{P}} r_{pq}[k] &= \tau\beta_p^i \textstyle\sum_{j\in\mathcal{J}} \sum_{q\in\mathcal{P}^j} r_{pq}^j[k] \\ &= \tau\beta_p^i \textstyle\sum_{j\in\mathcal{J}} \sum_{q\in\mathcal{P}^j} x_p^j[k]\sigma^j\mu_{pq}^j[k] = \\ &= \textstyle\sum_{j\in\mathcal{J}} \tau\beta_p^i x_p^j[k]\sigma^j \sum_{q\in\mathcal{P}^j} \mu_{pq}^j[k] = \\ &= \textstyle\sum_{j\in\mathcal{J}} \tau\beta_p^i x_p^j[k] \frac{\varepsilon}{2\tau\bar{\beta}\lambda^j|\mathcal{P}^j-1||\mathcal{J}|} \sum_{q\in\mathcal{P}^j} \mu_{pq}^j[k] = \\ &\leq \textstyle\sum_{j\in\mathcal{J}} \frac{\varepsilon}{2|\mathcal{J}|} = \frac{\varepsilon}{2}, \end{aligned} \tag{4.3.19}$$

where the inequality holds since $x_p^j[k] \leq \lambda^i$, $\beta_p^i \leq \bar{\beta}$ and since, recalling Equation (4.3.13), there are at most $\left|\mathcal{P}^j - 1\right|$ terms equal to 1 in $\sum_{q\in\mathcal{P}} \mu_{pq}^j[k]$. It follows that

$$l_p^i(\boldsymbol{x}[k+1]) \geq l_p^i(\boldsymbol{x}[k]) - \frac{\varepsilon}{2}. \tag{4.3.20}$$

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

$$l_p^i(\boldsymbol{x}[k+1]) \leq l_p^i(\boldsymbol{x}[k]) + \tau\beta_p^i \sum_{q\in\mathcal{P}} r_{qp}[k], \tag{4.3.21}$$

which yields

$$l_p^i(\boldsymbol{x}[k+1]) \leq l_p^i(\boldsymbol{x}[k]) + \frac{\varepsilon}{2}. \tag{4.3.22}$$

■

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 $i \in \mathcal{I}$, 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 (4.3.9), (4.3.11) and (4.3.8) yield that the population remains constant, since

$$\begin{aligned} x_p^i[k+1] - x_p^i[k] &= \sum_{p\in\mathcal{P}^i}\sum_{q\in\mathcal{P}^i}\left(r_{qp}^i[k] - r_{pq}^i[k]\right) \\ &= \sum_{p\in\mathcal{P}^i}\sum_{q\in\mathcal{P}^i} r_{qp}^i[k] - \sum_{q\in\mathcal{P}^i}\sum_{p\in\mathcal{P}^i} r_{qp}^i[k] = 0, \end{aligned} \tag{4.3.23}$$

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

$$x_p^i[k+1] = x_p^i[k] + \tau\sum_{q\in\mathcal{P}^i}\left(r_{qp}^i[k] - r_{pq}^i[k]\right) \geq 0, \forall p \in \mathcal{P}^i. \tag{4.3.24}$$

If $x_p^i[k] = 0$, it follows that $r_{pq}^i[k] = 0$ and thus Equation (4.3.24) yields $x_p^i[k+1] \geq 0$.

If $x_p^i[k] > 0$, from Equation (4.3.11) 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):

$$x_p^i[k+1] \geq x_p^i[k] - \tau\sum_{q\in\mathcal{P}^i} r_{pq}^i[k]. \tag{4.3.25}$$

A sufficient condition for inequality (4.3.31) to hold is then

$$x_p^i[k] - \tau\sum_{q\in\mathcal{P}^i} r_{pq}^i[k] \geq 0. \tag{4.3.26}$$

Recalling Equations (4.3.11) and (4.3.13), Eq. (4.3.26) is written as

$$x_p^i[k] - \tau \sum_{q\in\mathcal{P}^i} r_{pq}^i[k] = x_p^i[k] - \tau \sum_{q\in\mathcal{P}^i} x_p^i[k]\sigma^i \mu_{pq}^i[k] = x_p^i[k]\left(1 - \tau\sigma^i \sum_{q\in\mathcal{P}^i} \mu_{pq}^i[k]\right) \geq$$

$$\geq x_p^i[k]\left(1 - \tau\sigma^i\left(\left|\mathcal{P}^i\right| - 1\right)\right), \tag{4.3.27}$$

where the inequality holds since the summation has at most $\left(\left|\mathcal{P}^i\right| - 1\right)$ terms equal to 1. In the case $x_p^i[k] > 0$, Equations (4.3.14) and (4.3.15) are sufficient for Equation (4.3.27) 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

$$x_p[k+1] = x_p[k] + \tau \sum_{i\in\mathcal{I}} \sum_{q\in\mathcal{P}^i}\left(r_{qp}^i[k] - r_{pq}^i[k]\right) \leq c_p, \forall p \in \mathcal{P}^i. \tag{4.3.28}$$

If $x_p[k] \geq c_p - \frac{\varepsilon}{2\bar{\beta}}$ equation (4.3.13) entails that $r_{qp}^i[k] = 0$ for all $q \in \mathcal{P}^i$ and $i \in \mathcal{I}$ and, thus, from equation (4.3.9), that $x_p[k+1] \leq x_p[k]$.

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

$$x_p[k+1] \leq x_p[k] + \tau \sum_{i\in\mathcal{I}} \sum_{q\in\mathcal{P}^i} r_{qp}^i[k] =$$
$$= x_p[k] + \tau \sum_{i\in\mathcal{I}} x^i[k]\sigma^i \sum_{q\in\mathcal{P}^i} \mu_{qp}^i[k] =$$

$$\leq x_p[k] + \sum_{i\in\mathcal{I}} \frac{\varepsilon}{2\bar{\beta}|\mathcal{I}|} = x_p[k] + \frac{\varepsilon}{2\bar{\beta}} \tag{4.3.29}$$

■

Lemma 3. The function

$$\mathcal{L}(\boldsymbol{x}) \coloneqq \Phi(\boldsymbol{x}) - \Phi_{min}. \tag{4.3.30}$$

is a candidate Lyapunov function for the LB dynamics.

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

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

$$\begin{aligned}\Delta\mathcal{L}(\boldsymbol{x}[k]) &= \mathcal{L}(\boldsymbol{x}[k+1]) - \mathcal{L}(\boldsymbol{x}[k]) \\ &= \sum_{i\in\mathcal{I}}\sum_{p\in\mathcal{P}}\int_{x_p[k]}^{x_p[k+1]} l_p^i(\xi)d\xi \\ &\leq \sum_{i\in\mathcal{I}}\sum_{p\in\mathcal{P}}\left(x_p^i[k+1]-x_p^i[k]\right)l_p^i\left(x_p^i[k+1]\right) \\ &= \sum_{i\in\mathcal{I}}\sum_{p\in\mathcal{P}}\tau\left(\sum_{q\in\mathcal{P}} r_{qp}^i[k] - \sum_{q\in\mathcal{P}} r_{pq}^i[k]\right)l_p^i(\boldsymbol{x}[k+1]) \\ &= \tau\sum_{i\in\mathcal{I}}\sum_{p\in\mathcal{P}}\sum_{q\in\mathcal{P}} r_{pq}^i[k]\left(l_q^i(\boldsymbol{x}[k+1]) - l_p^i(\boldsymbol{x}[k+1])\right). \end{aligned} \tag{4.3.31}$$

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 $\left(x_p[k+1]-x_p[k]\right)l_p^i\left(x_p[k+1]\right)$; 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 $\left(x_p[k]-x_p[k+1]\right)l_p^i\left(x_p[k+1]\right)$.

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 $l_p^i(\boldsymbol{x}[k+1]) - l_q^i(\boldsymbol{x}[k+1]) > 0$.

Lemma 1 states that

$$\begin{aligned} & l_p^i(\boldsymbol{x}[k+1]) - l_q^i(\boldsymbol{x}[k+1]) \\ & \geq \left(l_p^i(\boldsymbol{x}[k]) - \tfrac{\varepsilon}{2}\right) - \left(l_q^i(\boldsymbol{x}[k]) \mp \tfrac{\varepsilon}{2}\right) = l_p^i(\boldsymbol{x}[k]) - l_q^i(\boldsymbol{x}[k]) - \varepsilon > 0, \end{aligned} \tag{4.3.32}$$

where the inequality holds since a necessary condition for $r_{pq}^i[k] > 0$ is that $l_p^i(\boldsymbol{x}[k]) - l_q^i(\boldsymbol{x}[k]) > \varepsilon$ (see equation (4.3.13)). ■

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 $\mathcal{L}(\boldsymbol{x})$ 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 $\boldsymbol{x} \in \mathcal{X}_{eq}^{\varepsilon}$ and $\boldsymbol{x}[0] = \boldsymbol{x}$. By comparing definition (4.3.6) and equation (4.3.13), it holds that $r_{pq}^i[k] = 0$ for all $p, q \in \mathcal{P}^i$ and $i \in \mathcal{I}$, which entails i) that $\boldsymbol{x}[k] = \boldsymbol{x}[0] = \boldsymbol{x}_{eq} \in \mathcal{X}_{eq}^{\varepsilon}$ for all $k > 0$, i.e., that $\mathcal{X}_{eq}^{\varepsilon}$ is a positively invariant set, and ii) that $\Delta\mathcal{L}(\boldsymbol{x}[k]) = 0$ in $\mathcal{X}_{eq}^{\varepsilon}$ (see equation (4.3.30)).

To show that $\mathcal{X}_{eq}^{\varepsilon}$ is the maximum set where $\Delta\mathcal{L}(\boldsymbol{x}[k]) = 0$, it is proven below that $\Delta\mathcal{L}(\boldsymbol{x}[k]) < 0$ if $\boldsymbol{x}[k] = \boldsymbol{x}$, with $\boldsymbol{x} \notin \mathcal{X}_{eq}$. In fact, by definition (4.3.12), in this case there exist at least one pair of providers $p, q \in \mathcal{P}^i$ and a commodity $i \in \mathcal{I}$ such that $l_p^i(\boldsymbol{x}[k]) - l_q^i(\boldsymbol{x}[k]) > \varepsilon$, with $x_p^i[k] > 0$ and $x_q[k] < c_q - \frac{\varepsilon}{2\bar{\beta}}$, which, in turn, yields $r_{pq}^i[k] > 0$ (see equations (4.3.11), (4.3.14) and

(4.3.13)). Having established that $r_{pq}^{i}[k] > 0$ with $l_p^i(\boldsymbol{x}(t)) - l_q^i(\boldsymbol{x}(t)) > \varepsilon$, it follows that the corresponding term of the inner summation of equation (4.3.30) is negative, which is a sufficient condition for $\Delta\mathcal{L}(\boldsymbol{x}[k]) < 0$ (recalling that, in the proof of Lemma 2, it is shown that the terms of equation (4.3.30) are nonpositive). ■

4.3.8 Numerical simulation

This section reports the simulation setup and results in sections 4.3.9 and 4.3.10, respectively.

4.3.9 Simulation setup

For the validation of the proposed algorithm, we consider the network depicted in Figure 4-7, consisting in an area covered by two macro-cells (providers 7 and 15). In each of the coverage zones of the two macro-cells there are six micro-cells (providers $1-6$ and $9-14$), while a high-capacity satellite (provider 8) offers connectivity in the whole area.

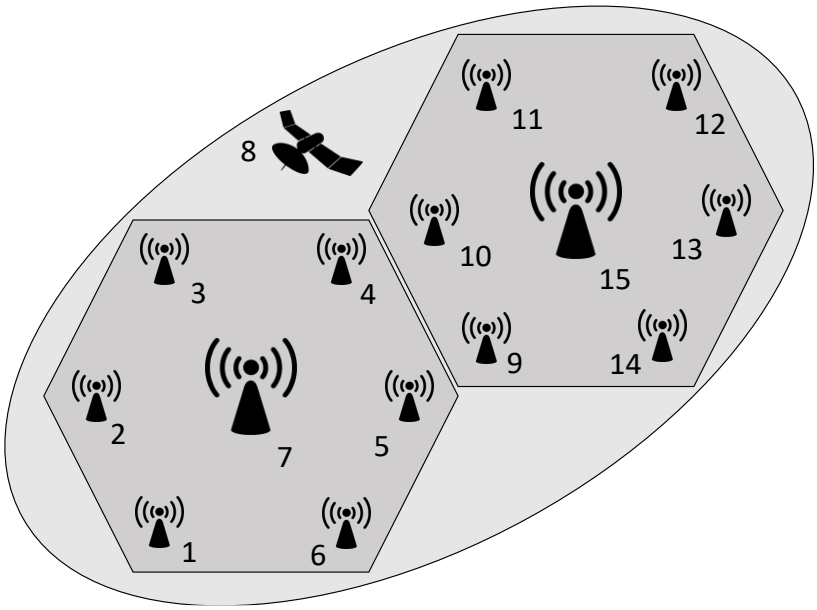


Figure 4-7: Network representation

A total of 20 aggregated QoS-flows (commodities) are transmitting in the area, each one characterised by a different set of APs available for routing its traffic. As introduced in Section 4.3.4, since the QoS-flows represent the aggregated traffic received by all the UEs in a given area that share the same available APs, it is reasonable to assume their cumulative bitrate to be equal to a constant $\lambda^i \in [4{,}13]$ Gbps, for each commodity i. In the example scenario, all the commodities can be served by the satellite connection, twelve commodities by a pair of adjacent micro-cells (e.g., (1,2), (2,3), …, in Figure 4-7) and by the macro-cell which covers the micro-cells, eight commodities by a group of three adjacent micro cells (e.g., (1,2,3), (2,3,4), …, in Figure 4-7) and by their corresponding macro cell.

Following the Shannon–Hartley theorem, the amount of power required for an AP to satisfy its required total bitrate at time k over a provider p, i.e., $x_p[k] = \sum_{i\in\mathcal{I}} x_p^i[k]$, is

$$P_p^i(x_p[k]) = \left(2^{\frac{x_p[k]}{W_p}} - 1\right)\frac{N_0}{G_p^i}, \tag{4.3.33}$$

in which W_p represents the AP bandwidth, N_0 captures the thermal noise and G_p^i is the inverse of the path loss L_p^i that characterises the connections of QoS-Flow i in AP p. It is fundamental remarking that the path loss L_p^i (and hence G_p^i) also accounts for the BER guaranteed to the i-th QoS-flow: the guaranteed BER determines the minimum energy-per-bit, E_b^i, which should be received by the UE, which, in turn, is reflected on the path loss parameter L_p^i. For instance, higher requirements of the i-th QoS-flow in terms of BER, i.e., of E_b^i, other conditions being equal, can be taken into account by increasing L_p^i and hence the required AP transmission power.

In (4.3.31), the effects of interference were neglected for simplicity, a reasonable assumption in the considered simplified scenario also in view of the overall interference reduction envisaged in 5G thanks to beamforming [47]. Furthermore, all the QoS-flows within a given AP were assumed to be characterised by the same path loss, due to their geographic proximity and to the presence of layer 1 power-control compensating mechanisms. The modelled path losses were calculated according to the standard free-space path loss expression

$$L_p^i = 92.45 + \log_{10}(f_p) + 20\log_{10}(d_p^i),$$

in which f_p represents the carrier frequency measured in GHz and d_p^i is the *virtual* connection distance, in km, of the UEs of QoS-Flow i from AP p, which, as explained above, can differ from the actual distance to reflect the BER requirements of QoS-flow i.

Considering the transmission power (4.3.31) as the latency function, i.e., $l_p^i[k] = P_p^i(x_p[k])$, the algorithm objective becomes the one of routing the traffic of each QoS-flow over the different RATs in such a way to minimise, unilaterally, their power consumption. This scenario is representative of a situation in which the various QoS-flows, in principle owned by different tenants, are charged by the network operators depending on their energy usage. Several different choices could be made for the latency function, ranging 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.

Regarding the capacitated nature of the considered network, one has to consider that the amount of power available to the antennas of a network slice is limited by a value $\bar{P}_p$ due to either the network operator resource allocation policy, in terms of dedicated bitrate in the various APs, or the equipment limitation. Considering an average value $\bar{G}_p$ for the G_p^i's, his power constraint implies that the bitrate offered by the AP p is limited by

$$c_p = W_p \log_2\left(1 + \bar{P}_p \frac{\bar{G}_p}{N_0}\right).$$

Additionally, the satellite connection was assumed to be characterised by an isotropic gain of $60\ \mathrm{dBi}$.

Table 4-3 reports the numerical values that characterise the network under study. We further imposed that the micro-cell 5 maximum bitrate was capacitated at $14\ \mathrm{Gbps}$ as not the whole transmission power was dedicated to the considered network slice by the operator.

Table 4-3: Characteristics of micro, macro and satellite cells

Param.	*Micro-cell*	*Macro-cell*	*Satellite*
$\overline{P_p}$	1 W	41 W	2 W
d_p	[0.1,0.2] km	[0.2,0.3] km	35000 km
f_p	21 GHz	2.1 GHz	20 GHz
W_p	1 GHz	0.2 GHz	0.8 GHz
c_p	17.84 Gbps	4.4 Gbps	2 Gbps

Concerning the algorithm parameters, the choice of latency functions leads to the value $\bar{\beta} = 3.4\ \mathrm{W/GHz}$, the latency tolerance is selected as $\varepsilon = 0.01\ \mathrm{W}$ and the sampling time as $\tau = 10^{-3}\mathrm{s}$.

4.3.10 Simulation results

Simulation runs were initialized without load on the macro-cells and with a random load distribution over the micro-cells and the satellite, and showed a convergence time to an ε-Beckmann equilibrium in the order of $150s$, averaged over 25 runs. Figure 4-8 shows, for an example run, how the maximum latency mismatch over all the commodities, defined as

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

decreases with time and, even if the initial conditions are quite unbalanced with $e[k] > 2\mathrm{W}$, after $100s$ $e[k]$ is already below $0.03\ \mathrm{W}$.

As examples of simulation results, Figure 4-9 reports the evolution of the latencies that characterise the commodities 9, 11 and 5, for all of their available APs. The plot at the top shows that the latencies of the APs available to QoS-flow 9 converge to a common value, as expected, within the threshold ε; in particular, we can notice how the commodity rapidly starts using the (initially unused) available macro-cell. The middle plot shows that QoS-flow 11 avoids using the satellite, even if it is available in $\mathcal{P}^i$, in order to achieve a lower convergence latency value. Finally, the plot at the bottom shows the latencies of the QoS-flow 5 and highlights that the latency of micro-cell 5 does not converge to the latencies of the other used APs (micro-cell 6, macro-cell 7): the reason is that, in the simulation scenario, the available capacity of micro-cell 5 is lower than the one of the other micro-cells and becomes ε-saturated after about $60s$ – thus, by definition, the population of QoS-flow 5 still converges to an ε-Beckmann equilibrium.

Finally, Figure 4-10 shows the population dynamics over the APs, highlighting how most of the traffic is routed over the micro cells, while still utilising the macro cells and satellite to balance the overall power consumption.

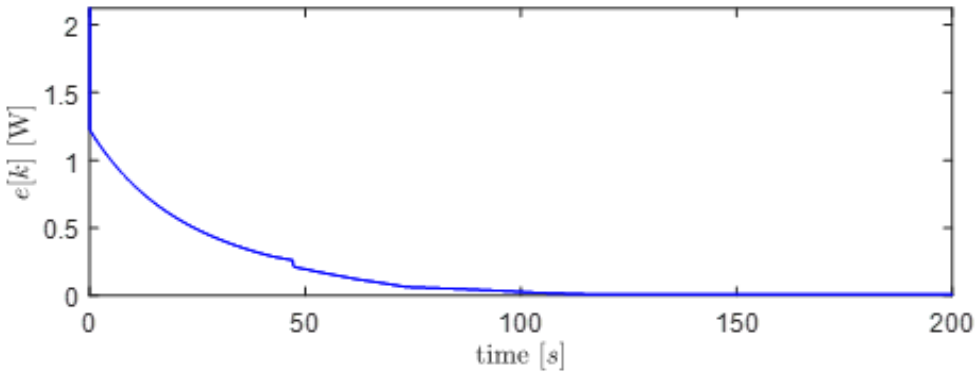


Figure 4-8: Maximum latency mismatch during the simulation

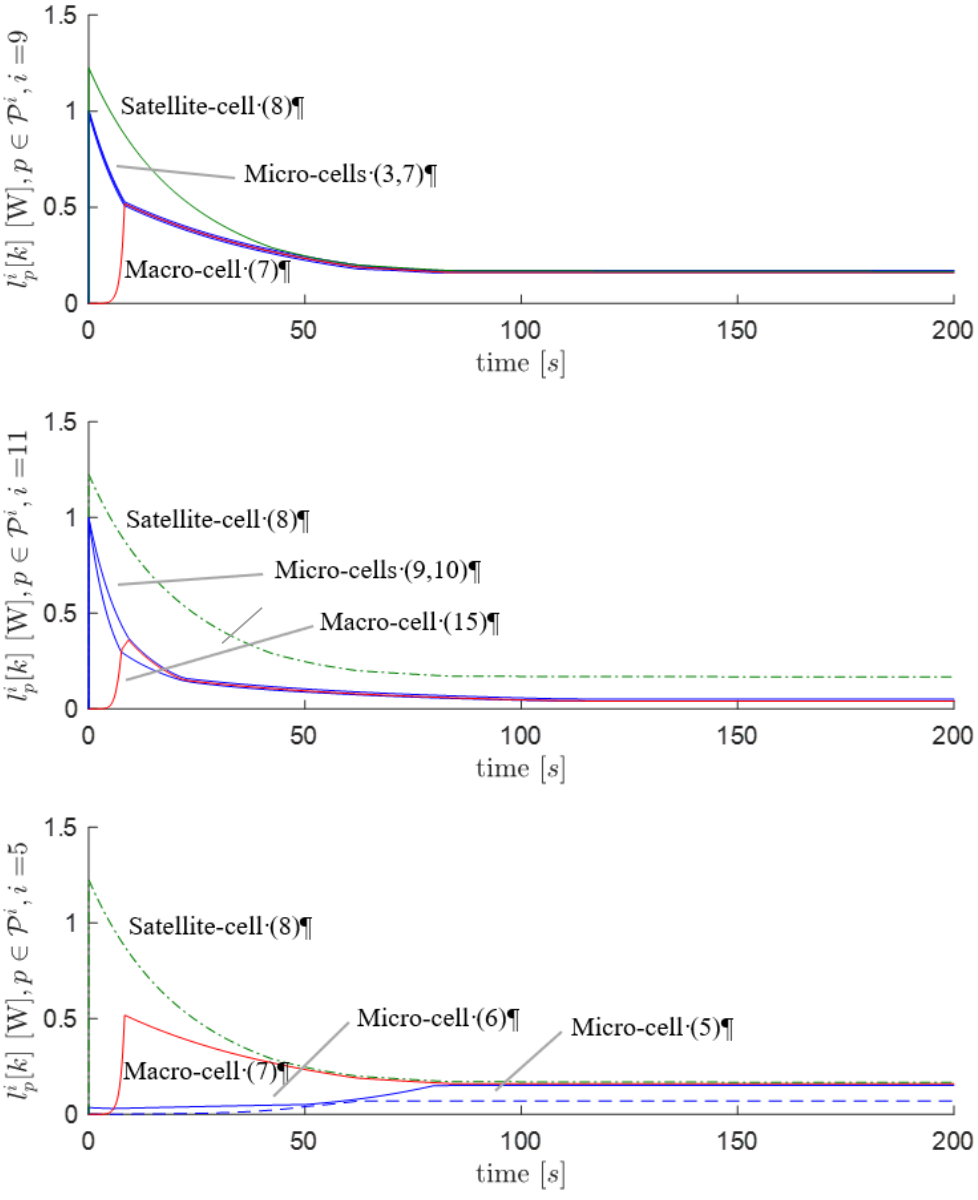


Figure 4-9: Commodity latency examples during the simulation (solid lines: unconstrained providers used by the commodity; dash-dotted lines: unconstrained providers not used by the commodity; dashed lines: constrained providers)

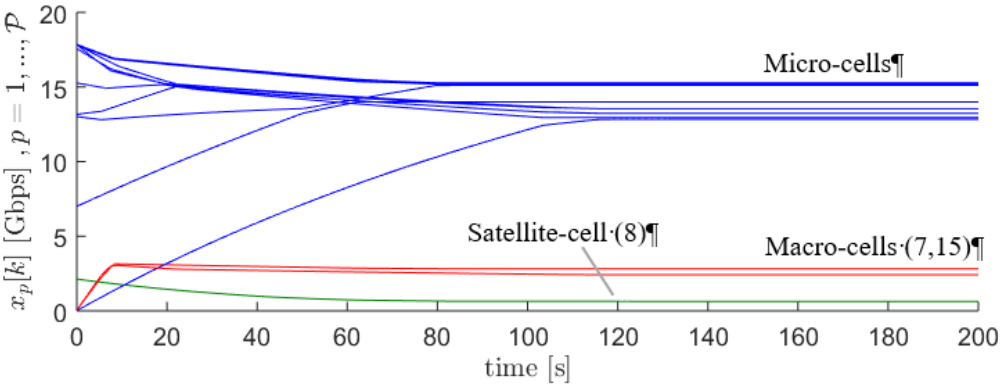


Figure 4-10: Access points loads

4.4 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.

4.4.1 Markov games and Nash equilibria

A Markov game among N players is defined as the tuple [48] $\langle S, \mathcal{A}, T, \mathcal{R}, \gamma \rangle$, 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.

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

$\mathcal{R} = \{R_i(s, a_1, a_2, \dots, a_N) : S \times A_1 \times A_2 \times \dots \times A_N \to \mathbb{R},\ A_i = \{a_i\}\}$ 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 $\pi_i(s) : S \to \mathbb{R}^{\#(A_i)}$ 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 [49], [50], defined as

$$Q_i(s, a_1, a_2, \dots, a_N) = R_i(s, a_1, a_2, \dots, a_N) + \gamma \sum_{s'} T(s, a_1, a_2, \dots, a_N, s') Q_i(s', \pi_1, \pi_2, \dots, \pi_N), \qquad (4.4.1)$$

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 π_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 [50]:

- 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.,

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

4.4.2 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) [49] 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 π_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 [50] 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:

$$Q_i(s, a_1, a_2, \dots, a_N) = \big(1 - \alpha(t)\big) Q_i(s, a_1, a_2, \dots, a_N)$$
$$+ \alpha(t)\big(r_i + \gamma\, Nash_i(s, Q_1, Q_2, \dots, Q_N)\big), \qquad (4.4.2)$$

where $Nash_i(s, Q_1, Q_2, \dots, Q_N)$ is computed as ([50])

$$Nash_i(s, Q_1, Q_2, \dots, Q_N) = \max_{\pi \in \Pi(A_1 \times \dots \times A_k)} \min_{[a_k, \dots, a_N] \in (A_{k+1} \times \dots \times A_N)} \sum_{[a_k, \dots, a_N] \in (A_{k+1} \times \dots \times A_N)} \pi(a_1) \cdots \pi(a_k) Q_i(s, a_1, \dots, a_N). \qquad (4.4.3)$$

In (4.4.3), 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 $\sum_t \alpha(t)^2 < +\infty$ [49], Theorem 6 in [50] 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.

4.4.3 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.

4.4.4 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:

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

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

4.4.5 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:

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

4.4.6 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 α_{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.

4.4.7 ε-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 (4.4.2) and the solution of the *maximin* problem (4.4.3) 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 ε-gready selection:

$$a_i = \begin{cases} \text{argmax}_{a_i}(Nash_i), \text{with probability } 1-\varepsilon \\ \text{randomly chosen in the set } A_i, \text{with probability } \varepsilon \end{cases}, \quad (4.4.4)$$

where $Nash_i$ is the operator described in (4.4.3), 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 [51], a possible refinement to (4.4.4) is to consider a decreasing sequence of values for ε, 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.

4.4.8 Maximin linear programming formulation

In general, a *maximin* optimization problem takes the following form [52]:

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

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 [52]:

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

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 (4.4.3) becomes

$$\max_{\pi \in \Pi(A_1 \times \dots \times A_k)} \min_{[a_k,\dots,a_N] \in (A_{k+1} \times \dots \times A_N)} \sum_{[a_1,\dots,a_k] \in A_1 \times \dots \times A_k} \pi(a_1) \cdots \pi(a_k) \, Q_i[s, a_1, \dots, a_N]$$

$$\begin{aligned} s.t. \quad & \pi(a_i) \geq 0 \; \forall \; a_i \in A_j, j = 1, .., k \\ & \textstyle\sum_{i=1}^{k} \pi(a_i) = 1, \end{aligned}$$

leading to an equivalent LP formulation of the form:

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

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

Table 4-4: Algorithm pseudo-code

Initialize $Q_i(s,a), i = 1, \dots, n$, arbitrarily.
For each connection request (episode) do:

- Each player observes the state of the system and the class j of the upcoming connection
- Each player selects its action a_i with an ε-greedy policy based on their $Nash_i$ function
- The connection is allocated using the resources of one of the players that selected an action with $a^j = 1$. If no player was available for the allocation one is selected randomly.
- All players receive a reward r_i as described in section 4.4.6.
- All players update their Q_i table according to eq. (4.4.2).

end

4.4.9 Simulations

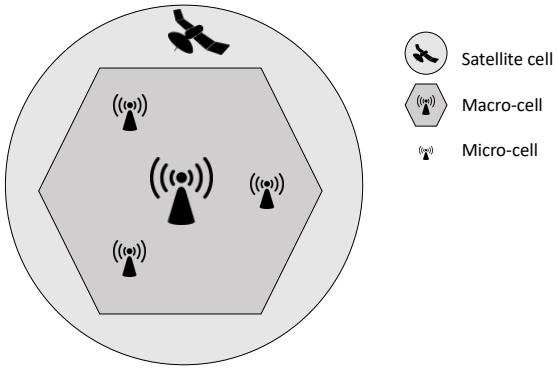


Figure 4-11: Connection area covered by 3 different RATs

The scenario considered in the simulation is shown in Figure 4-11, 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 $t_1 = 1\ \mathrm{Mbps}$ and the latter by $t_2 = 5\ \mathrm{Mbps}$. The parameters α_{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 $\lfloor\ \rfloor$ represents the lower-integer operator, and $\varepsilon(t)$ halved every 100 iterations starting from $\varepsilon(0) = 0.6$. Finally, the discount factor γ 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.

4.4.10 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.

4.4.11 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.

4.4.12 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.

4.4.13 Simulation results

From the analysis of Figure 4-12 and Figure 4-13, 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 α_{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%.

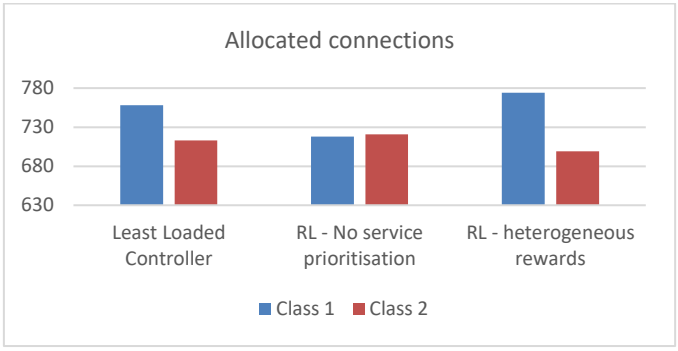


Figure 4-12: Number of allocated connections for the three controllers, divided by service class

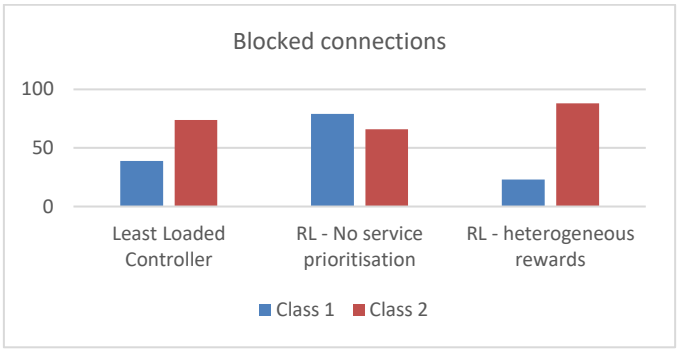


Figure 4-13: Number of blocked connection for the three controllers, divided by service class

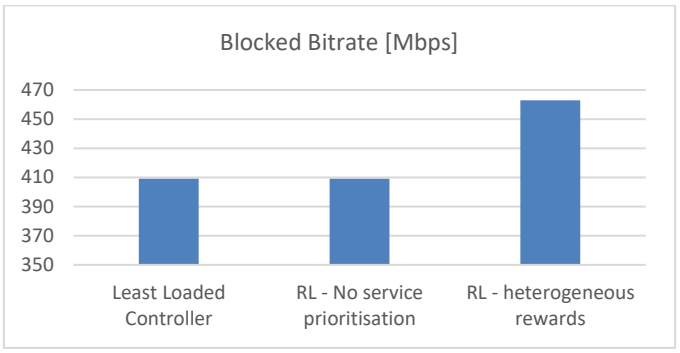


Figure 4-14: Total amount of bitrate of the blocked connections

4.5 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:

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

where w is total bandwidth of terrestrial RAN, ρ is the number of allocated PRBs and δ is sub-carrier spacing transmitted at time t.

The proposed algorithm for load balancing runs in each gNB. Based on bandwidth usage ratio, β, 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 form 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 4-15 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.

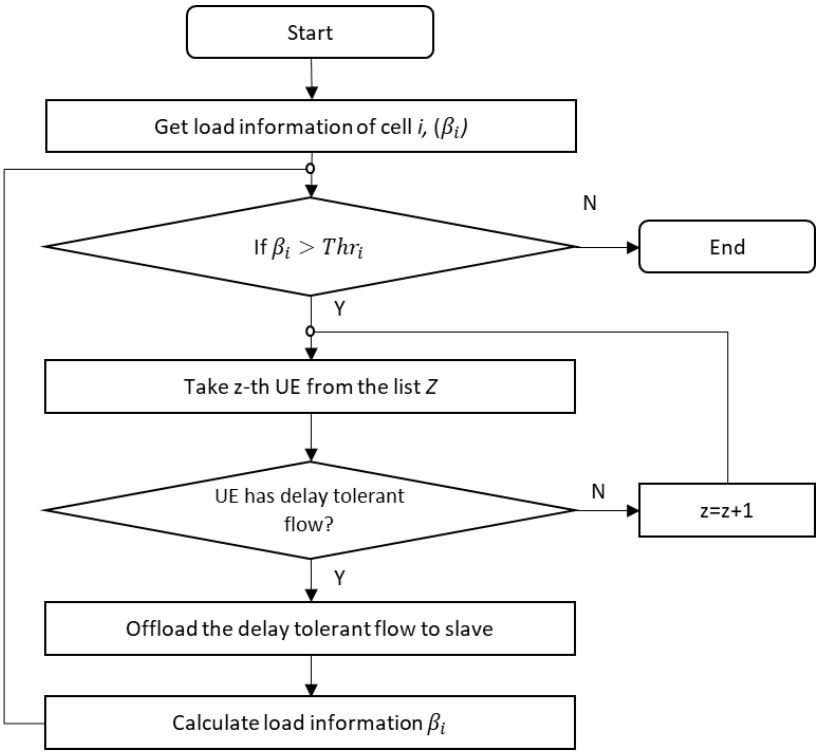


Figure 4-15: Flow diagram of proposed MLB algorithm

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.

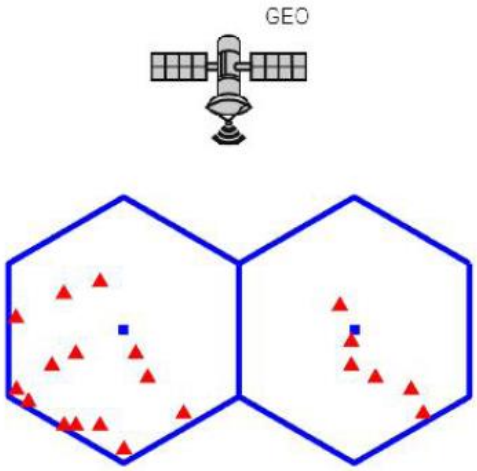


Figure 4-16: Simulation environment for MLB algorithms

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.

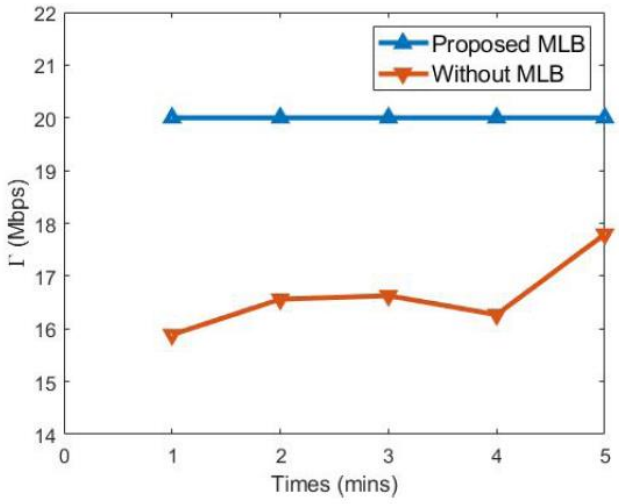


Figure 4-17: Simulation results with throughput enhancements

4.6 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.

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

3) State: radio environment state set has the element of the vector $(\mathrm{h}, \mathrm{I}, \mathrm{N}, \mathrm{L}, \mathrm{U})$ 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.

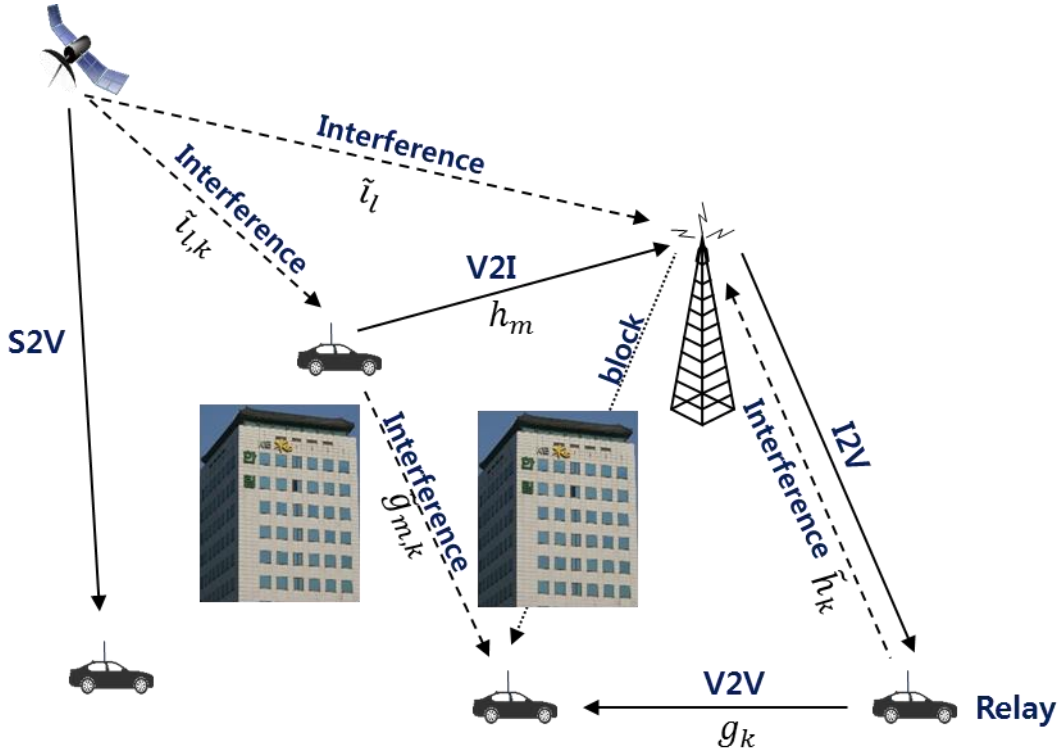


Figure 4-18: The analysis scenario for the proposed algorithm: the radio resource allocation policy with RL based interference control

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

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

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 1-th satellite device, respectively. σ^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

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

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 λ_c, λ_d, λ_f, λ_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 Decission 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.

4.7 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 Co-operative 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.

4.7.1 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.

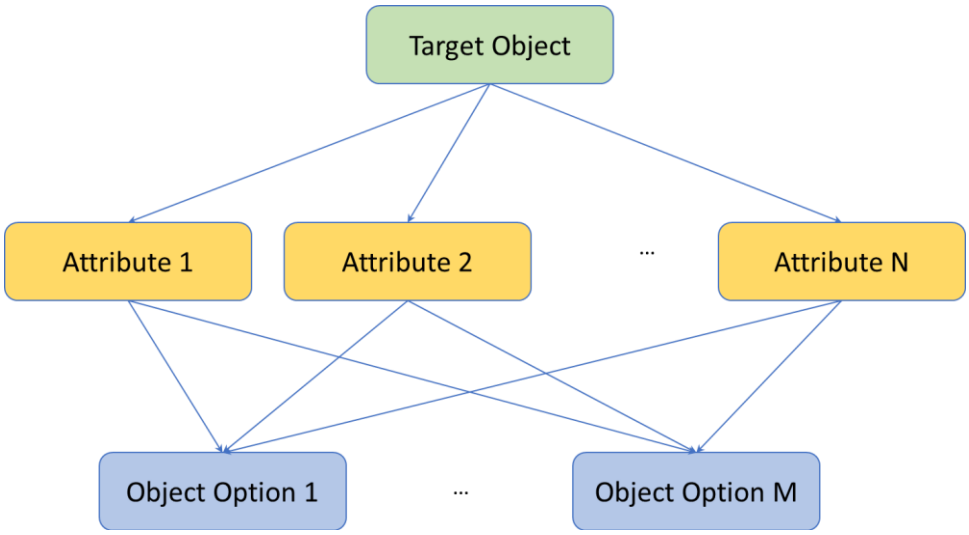


Figure 4-19: AHP Hierarchy

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 choises 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 [59] (e.g., Figure 4-20). 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 k-th attribute. The matrix A has the following properties:

a) each element a_{ij} with $i > j$ is defined following the table of Figure 4-20
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 4-20: Saaty's Scale for Pairwise comparison

The comparison matrix is said to be consistent if and only if $\lambda_{max} = n$, where λ_{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 $\boldsymbol{w}$ is unique, because is the eigenvector of the principal eigenvalue λ_{max} of the matrix.

In order to check the consistency, in [60] 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 4-21.

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 4-21: 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 $\boldsymbol{v}$ defined as

$$v = Sw$$

   Where the i-th entry of $\boldsymbol{v}$ is the grade of the i-th option.

4.7.2 Cooperative Differential Game

As introduced in section 4.3, 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 4.3 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 [61]:

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

$$x(k+1) = f\big(x(k), u(k), d(k)\big)$$
$$0 \le u(k) \le \bar{u}; \quad 0 \le x(k) \le \bar{x}$$
$$x(k_0) = x_0; \quad x(K) = x_f$$
$$x^T = [x_1, x_2, \dots, x_n]; \quad u^T = [u_1, u_2, \dots, u_m]; \quad d^T = \big[d_1, d_2, \dots, d_p\big];$$

4.7.3 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 4-19 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 4-22

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

Figure 4-22 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 4-23 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:

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

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

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

$$x(k)^T = [x_1, x_2, \dots, x_n]$$
$$u(k)^T = [u_{11}, \dots, u_{1n}, \dots, u_{m1}, \dots, u_{mn}]$$

Where

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

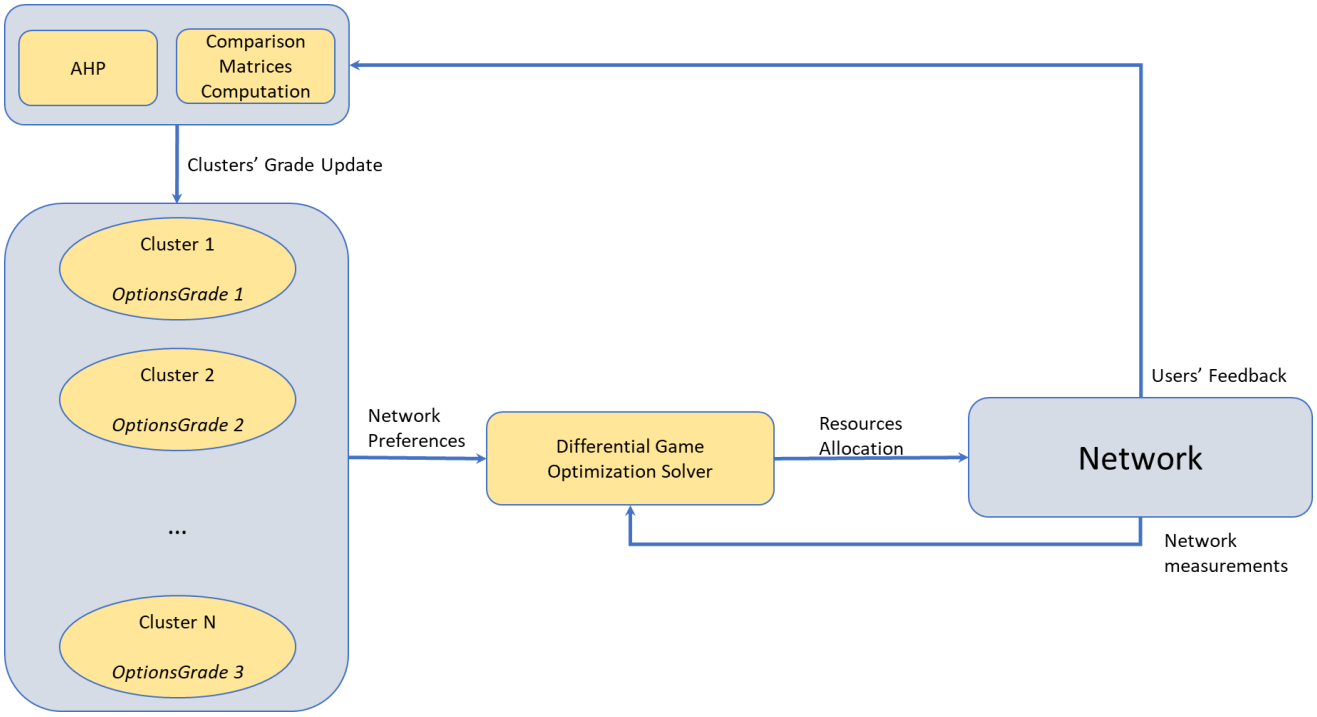


Figure 4-24 Block diagram of the AHP algorithm

4.8 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

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

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 $\mathcal{H}$ [62].

In this non-cooperative framework, a solution of interest is the so-called Nash Equilibrium [62]. 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 [62]:

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 [63], or using an extension of the Hamilton-Jacobi-Isaacs theory [63].

Our proposed solution is based on the Duality Theory [64].

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 $\mathcal{H}: \mathcal{U} \times \mathcal{V} \to R$, 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 inequalitiy, 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:

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

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 $\mathcal{H}: \mathcal{U} \times \mathcal{V} \to R$. 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 [62] hold, 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:

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

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.

4.8.1 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:

$$x_{k+1}^1 = x_k^1 + d_k - u_k - v_k$$
$$x_{k+1}^2 = x_k^2 + u_k - K_u$$
$$x_{k+1}^3 = x_k^3 + u_k - K_v$$

- 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:

$$0 \leq x_k^i \leq \overline{x}^i \quad i = 1,2,3$$
$$0 \leq u_k \leq \overline{u}$$
$$0 \leq v_k \leq \overline{v}$$

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 tails. 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, μ_u and μ_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 4-25 reports a functional scheme for the algorithm.

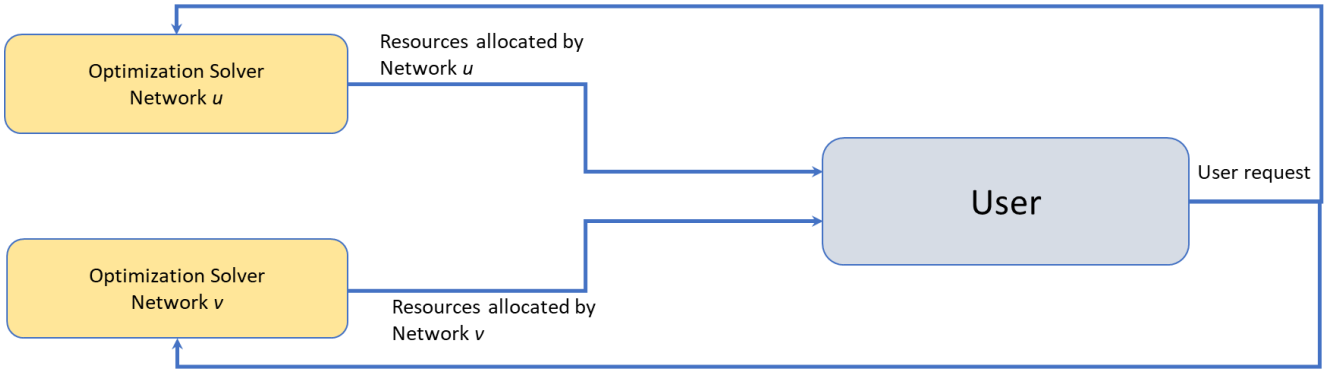


Figure 4-25 Block diagram of the minimax algorithm

5 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.

5.1 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.

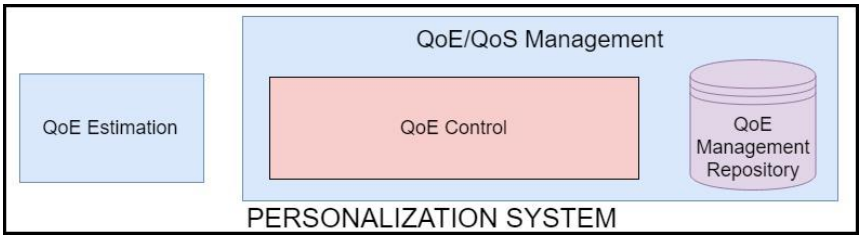


Figure 5-1: Personalization System from D4.1

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 5-2 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 net-work architecture*”. Figure 5-2 reports a detailed and updated version of the functional architecture of the personalisation system and its logical information exchanges.

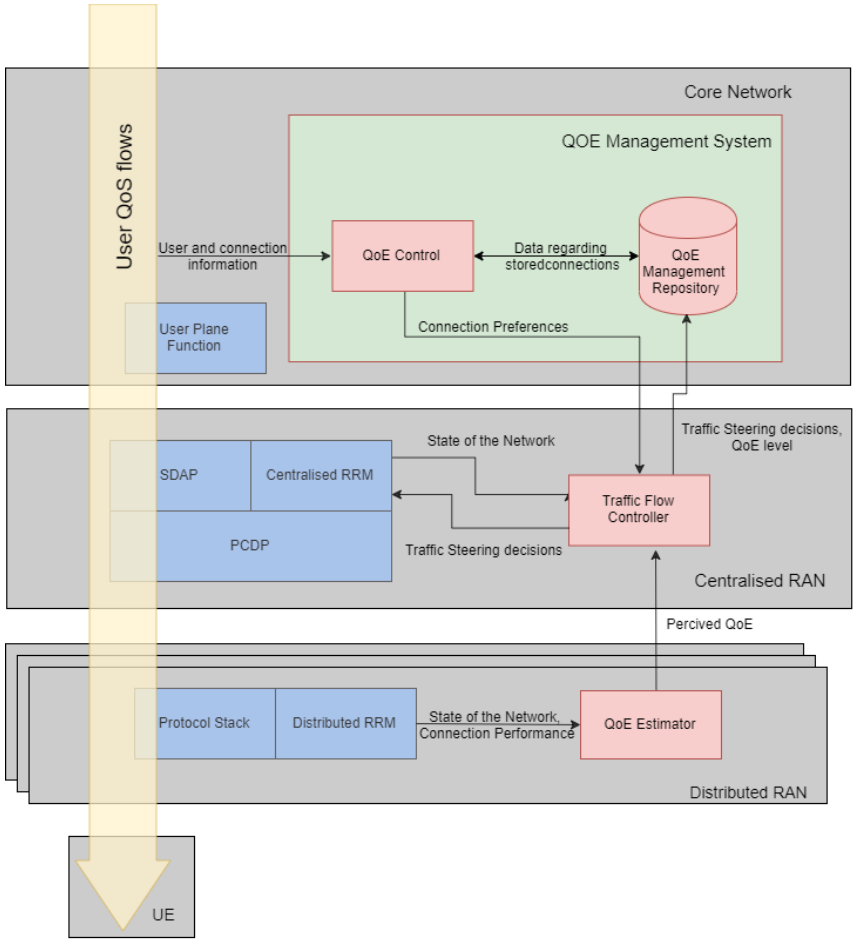


Figure 5-2: QoE management system functional architecture

The purpose of the personalisation system is to provide support to, and extend the capability of, the traffic flow controller. From Figure 5-2 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.

5.2 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.

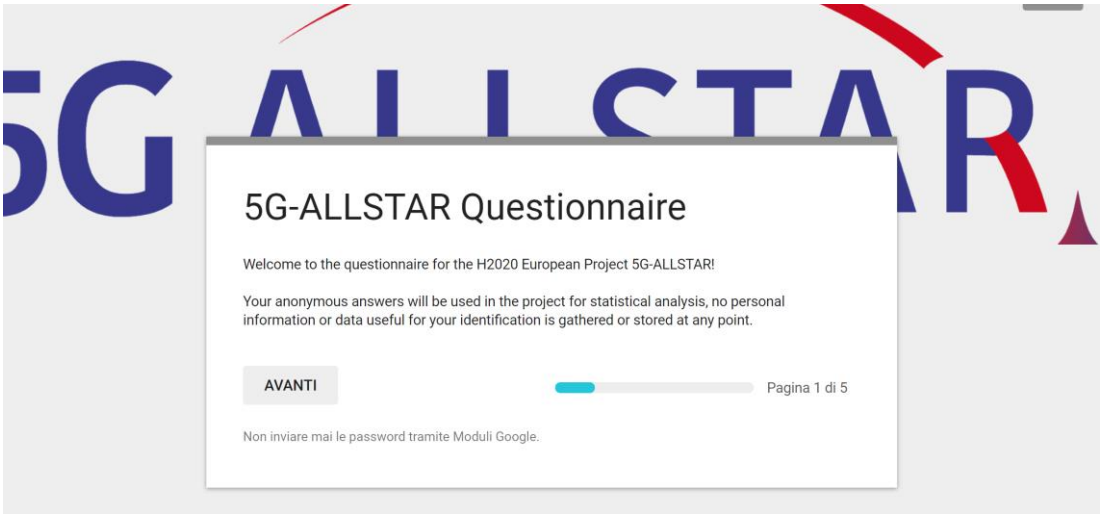


Figure 5-3: 5G-ALLSTAR questionnaire landing page

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 5-3 was presented. The geographic distribution of the users is reported in Figure 5-4.

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.

5.2.1 Answers report to the first section of the questionnaire

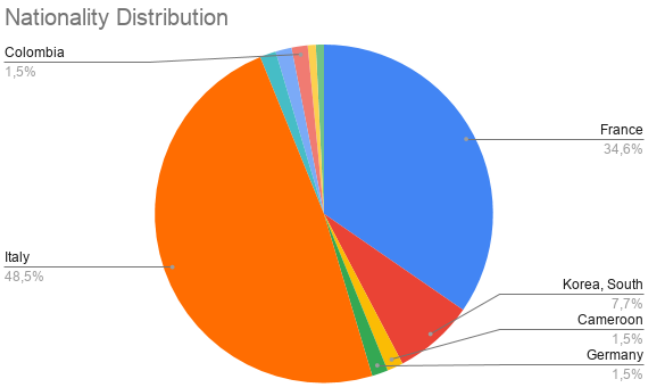


Figure 5-4: Nationality Distribution of the Questionnaire

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 5-5.

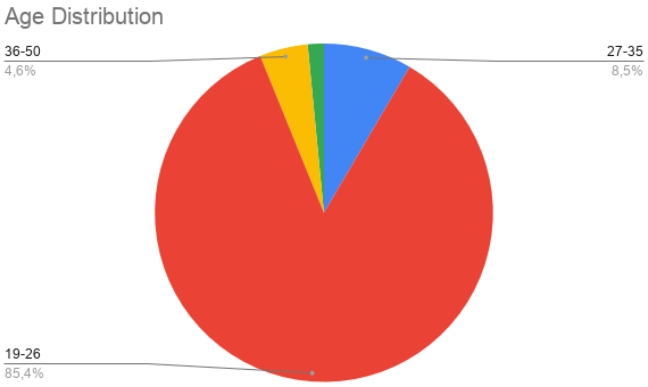


Figure 5-5: Age Distribution of the Questionnaire

Most of the users had a 4G smartphone as emerges from Figure 5-6, 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.

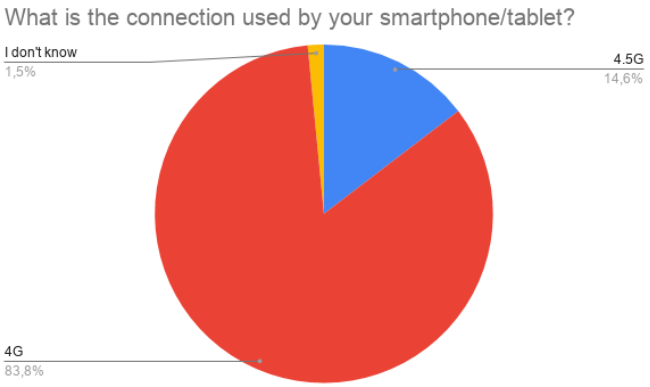


Figure 5-6: Answers to question 1 pie chart

From Figure 5-7 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 5-8 highlights that most users utilise video streams in mobility for long routes, while around 30% of them watches media during short trips.

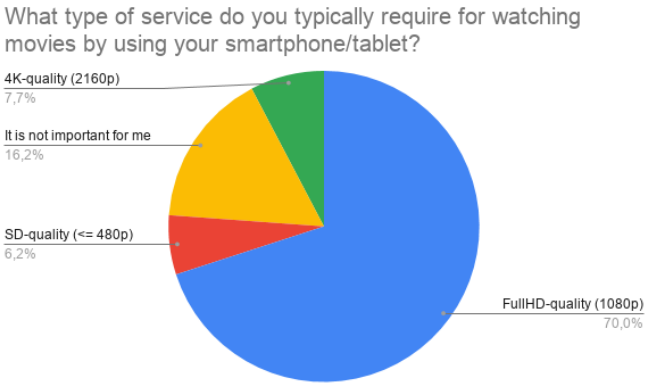


Figure 5-7: Answers to question 2 pie chart

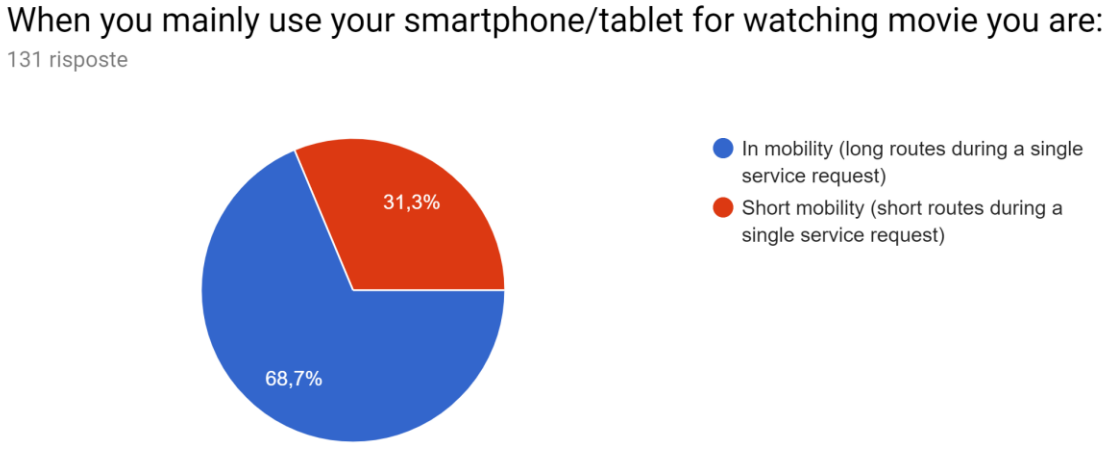


Figure 5-8: Answers to question 3 pie chart

Figure 5-9 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.

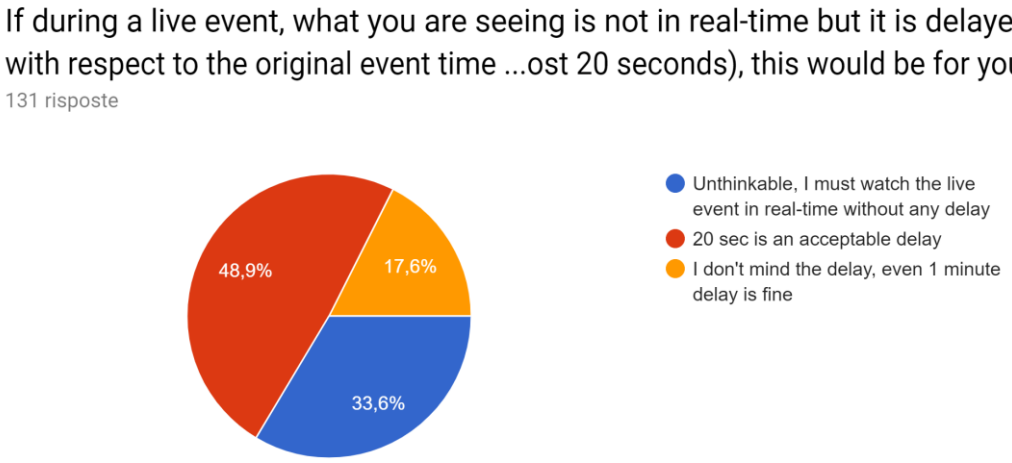


Figure 5-9: Answers to question 4 pie chart

Figure 5-10 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 5-8. 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 5-11, on the contrary, highlights that for entertainment the majority of the users would disregard power consumption in favour of video quality.

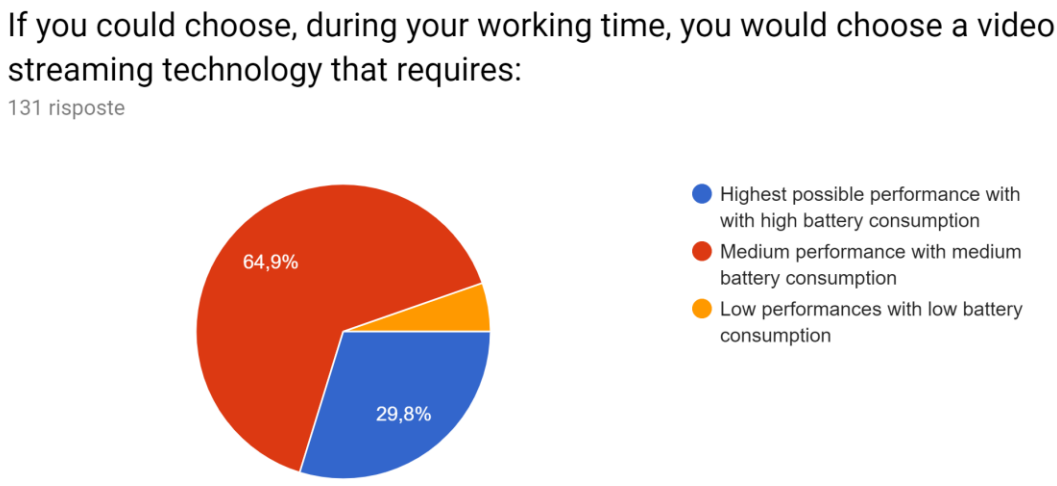


Figure 5-10: Answers to question 5 pie chart

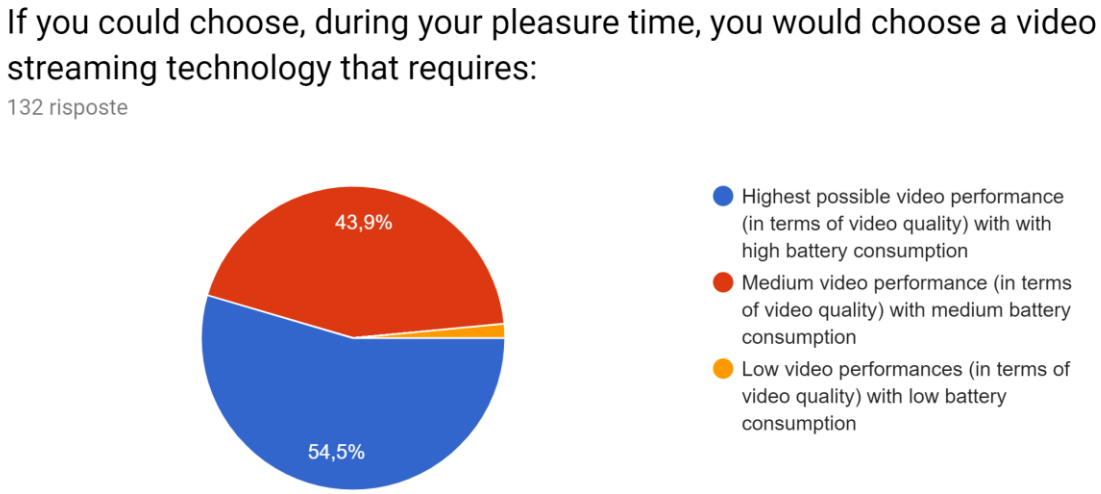


Figure 5-11: Answers to question 6 pie chart

Figure 5-12 shows that the users consider themselves very familiar with their smartphones and that they utilise it quite often and continuously during the day.

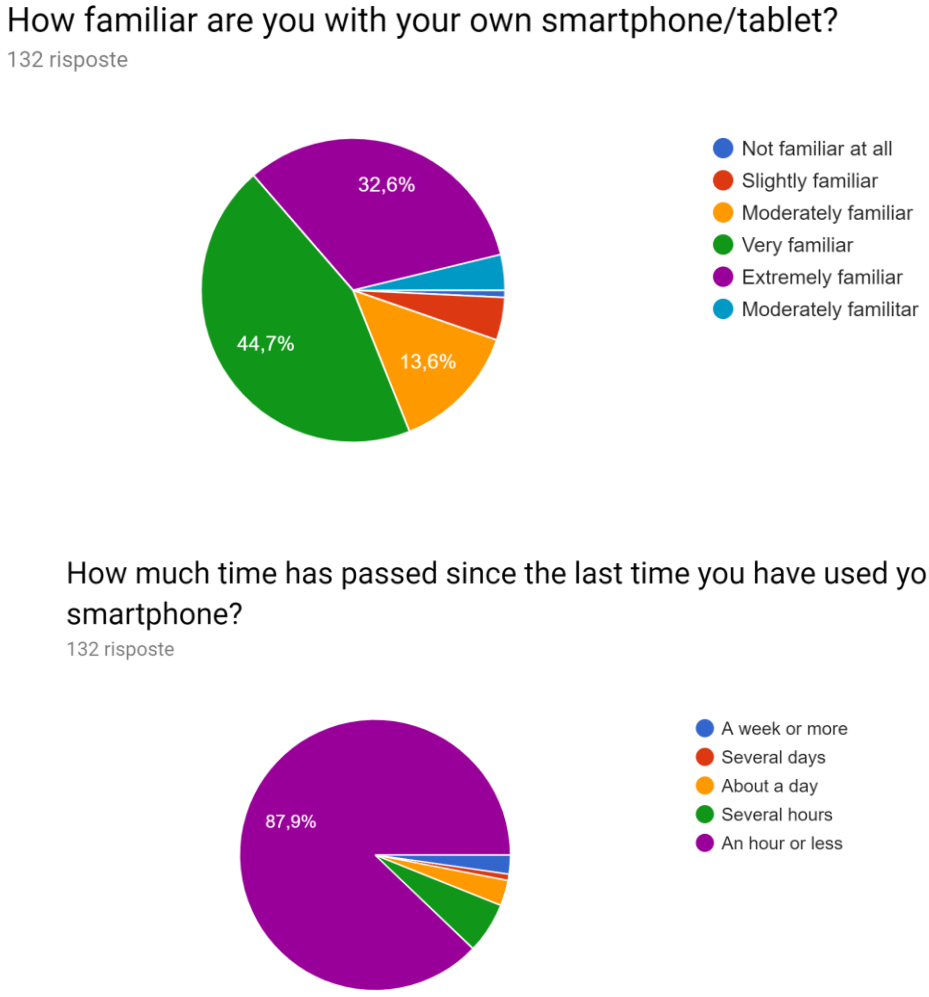
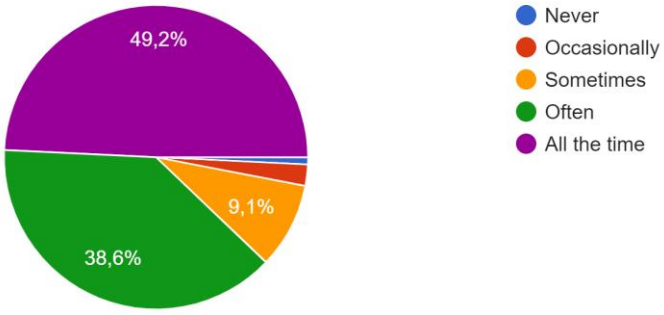


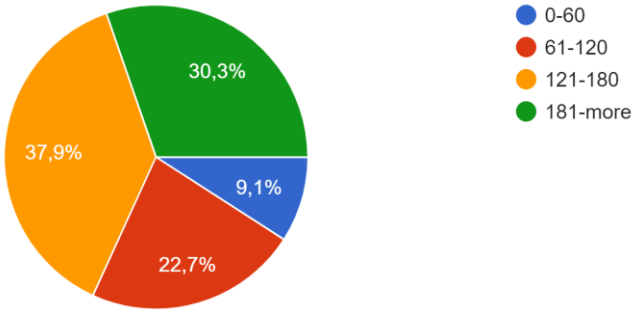
Figure 5-12: Answers to questions 7 and 8 pie charts

The same concept is highlighted in Figure 5-13, where the majority of the users state to utilise their smartphone for more than 2 hours a day.

On avarage, 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

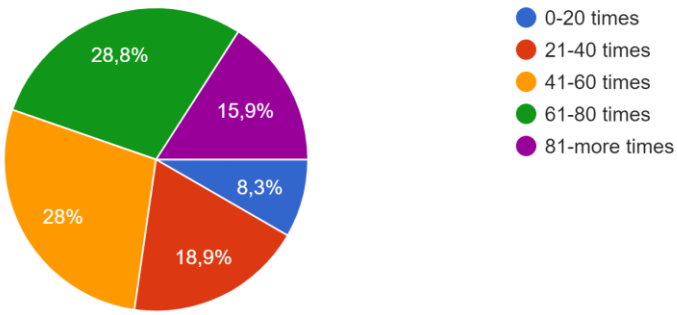


Figure 5-13: Answers to questions 9, 10 and 11 pie charts

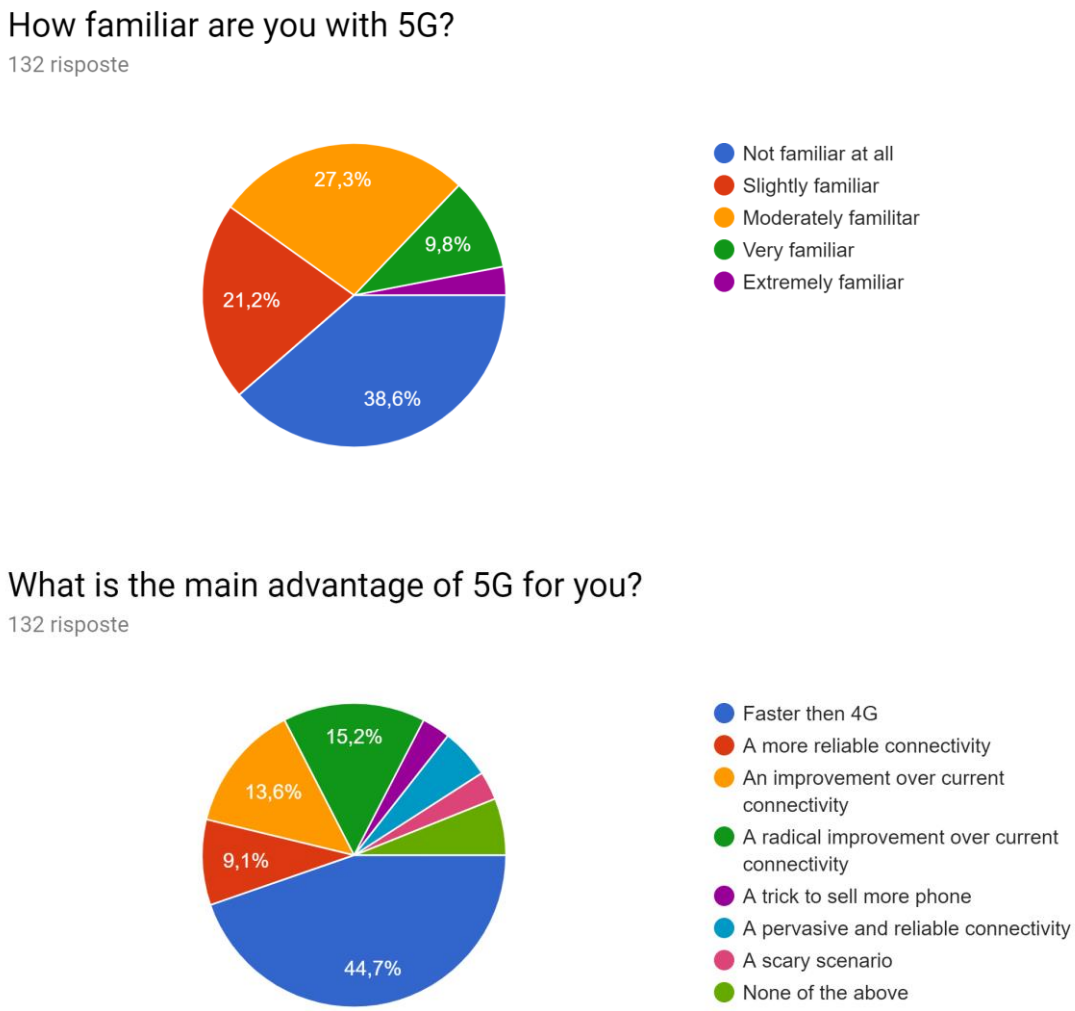


Figure 5-14: Answers to questions 12 and 13 pie charts

To conclude the first section, Figure 5-14, 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 through-put increase the main innovation of 5G.

5.2.2 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.

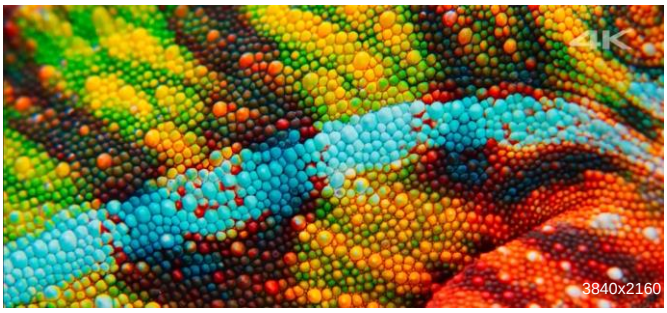


Figure 5-15: Screenshot of the first video shown during the questionnaire

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 5-15.

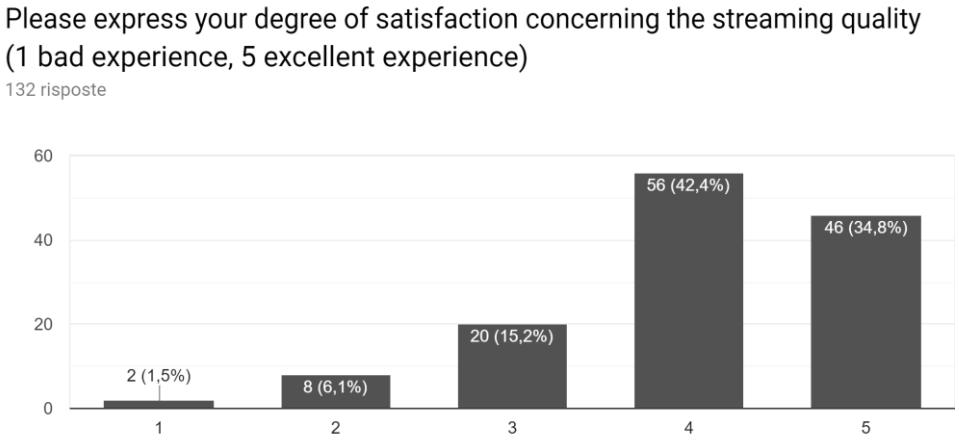


Figure 5-16: Bar chart of question 1, second section

As reported in Figure 5-16, 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.

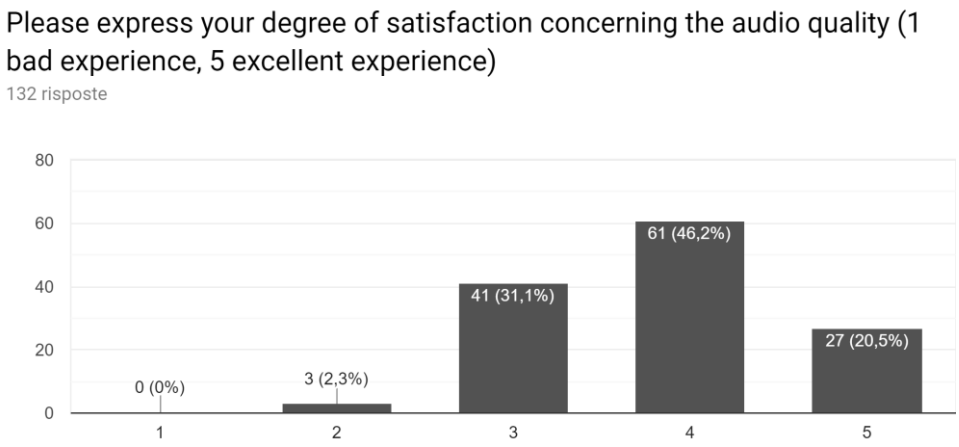


Figure 5-17: Bar chart of question 2, second section

A similar degree of satisfaction is observed for the audio in Figure 5-17, 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.

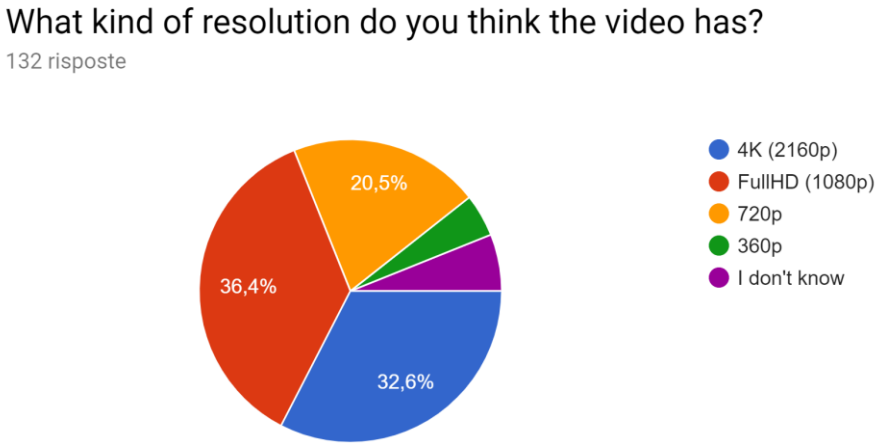


Figure 5-18: Pie chart of question 3, second section

5.2.3 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 5-19) 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 5-19: Screenshot of the second video shown during the questionnaire

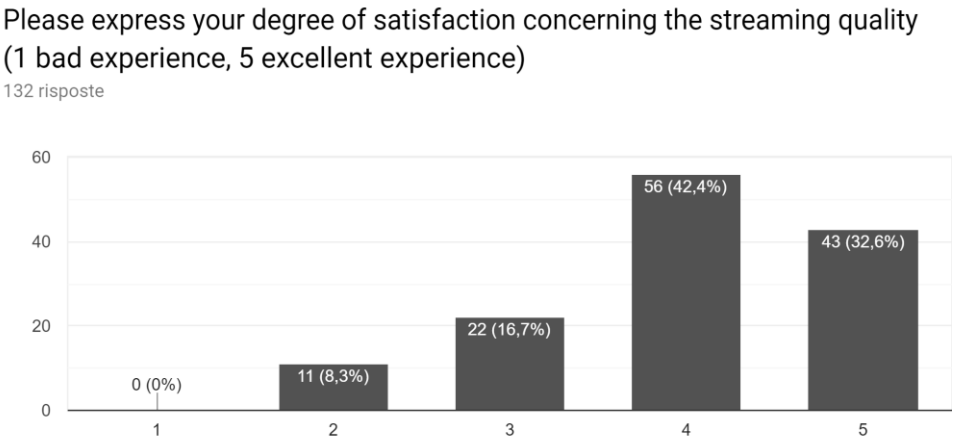


Figure 5-20: Pie chart of question 1, third section

As it is possible to observe in Figure 5-20, 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 5-21.

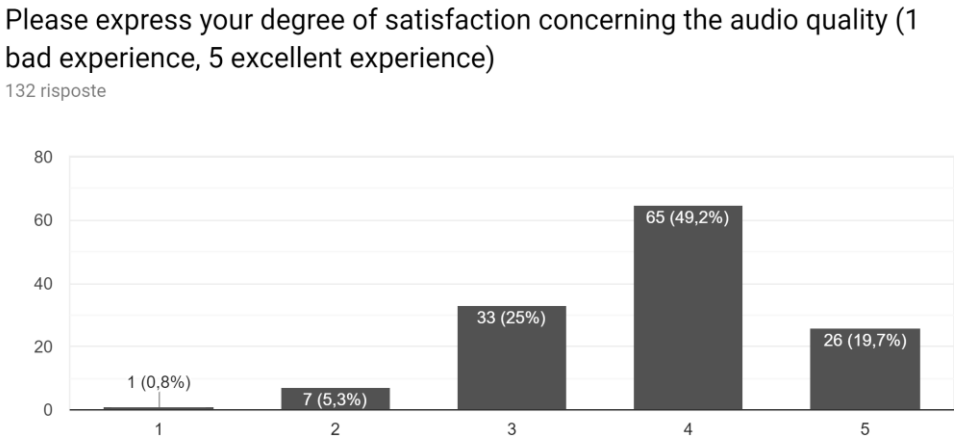


Figure 5-21: Bar chart of question 2, third section

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 5-22.

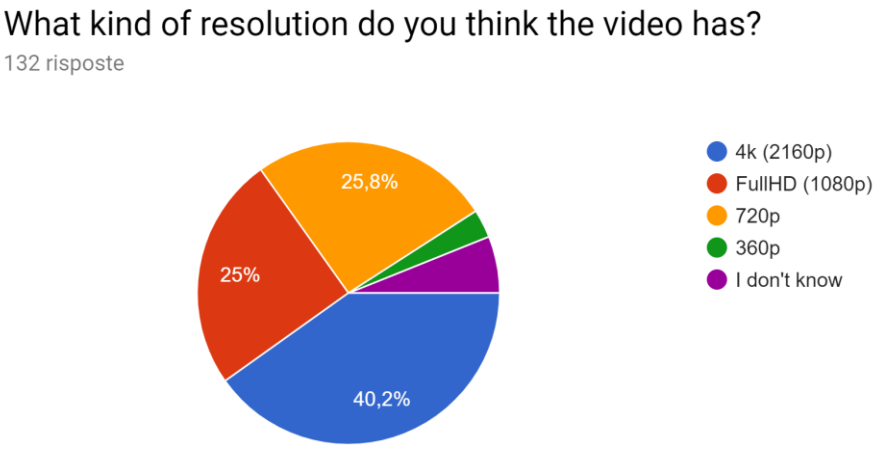


Figure 5-22: Pie chart of question 3, third section

5.2.4 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.

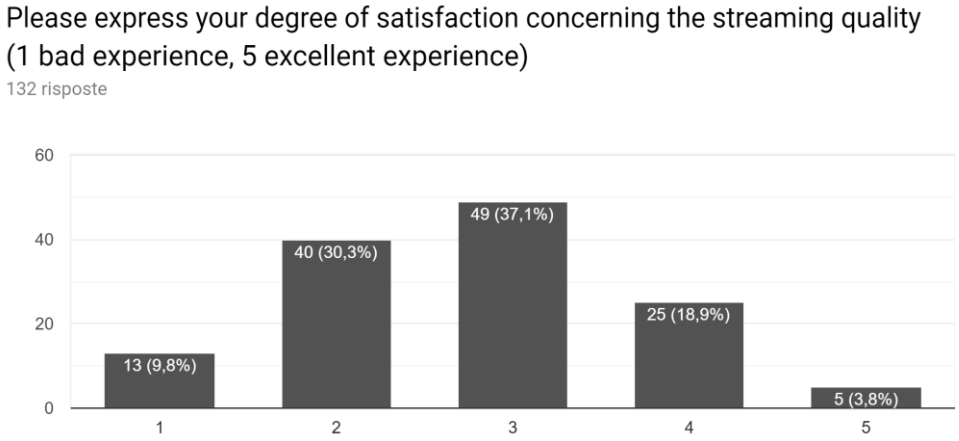


Figure 5-23: Bar chart of question 1, fourth section

Figure 5-23 highlights that the presence of delays in the audio flow impacted the perceived quality of the overall stream. In Figure 5-24 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 5-25.

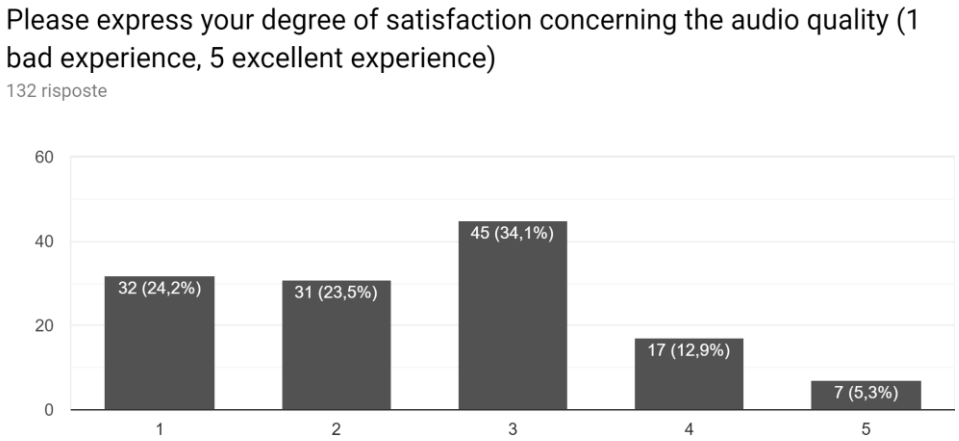


Figure 5-24: Bar chart of question 2, fourth section

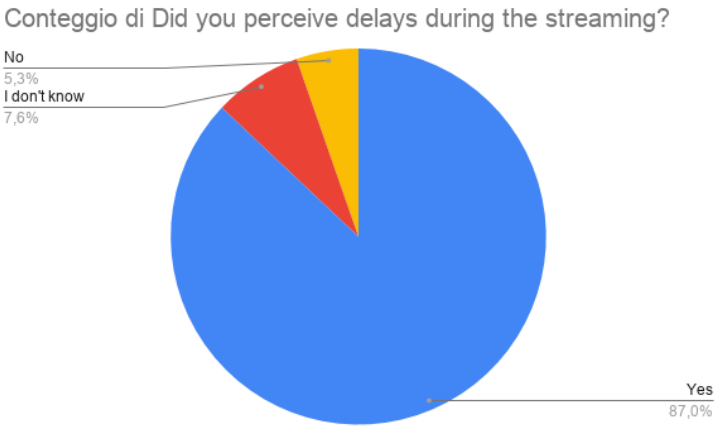


Figure 5-25: Pie chart of question 3, fourth section

5.2.5 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.

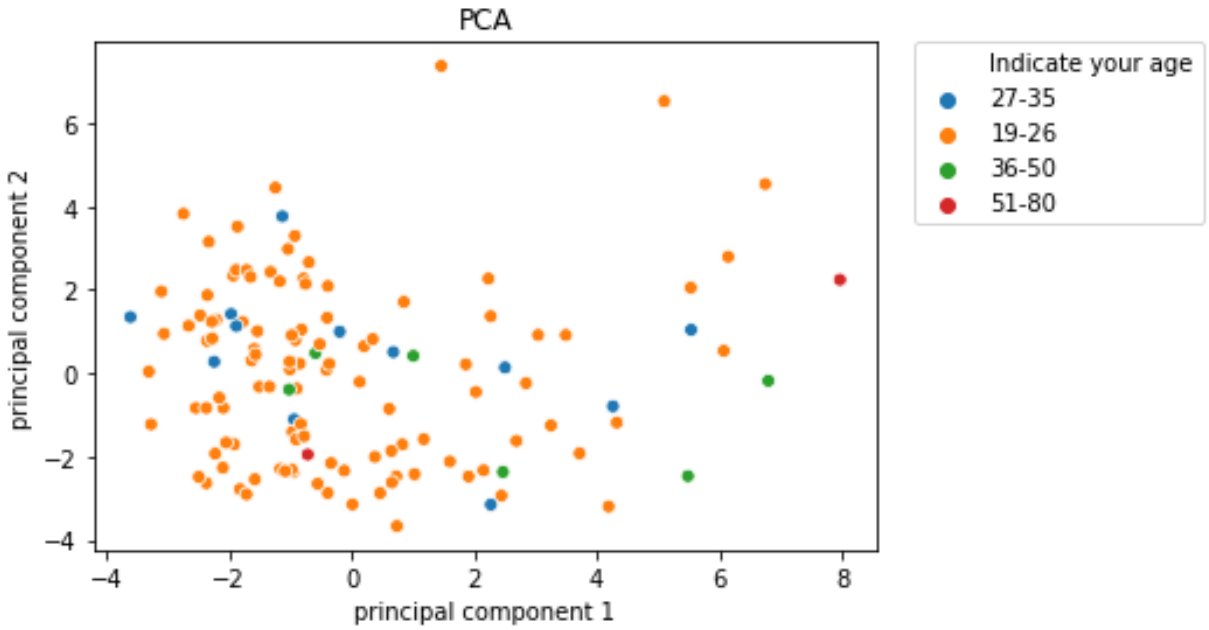


Figure 5-26: Principal component analysis, age

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

The first component of $\boldsymbol{T}$ has to maximize the variance, so $\boldsymbol{w_1}$ has to satisfy

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

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

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

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 5-26 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).

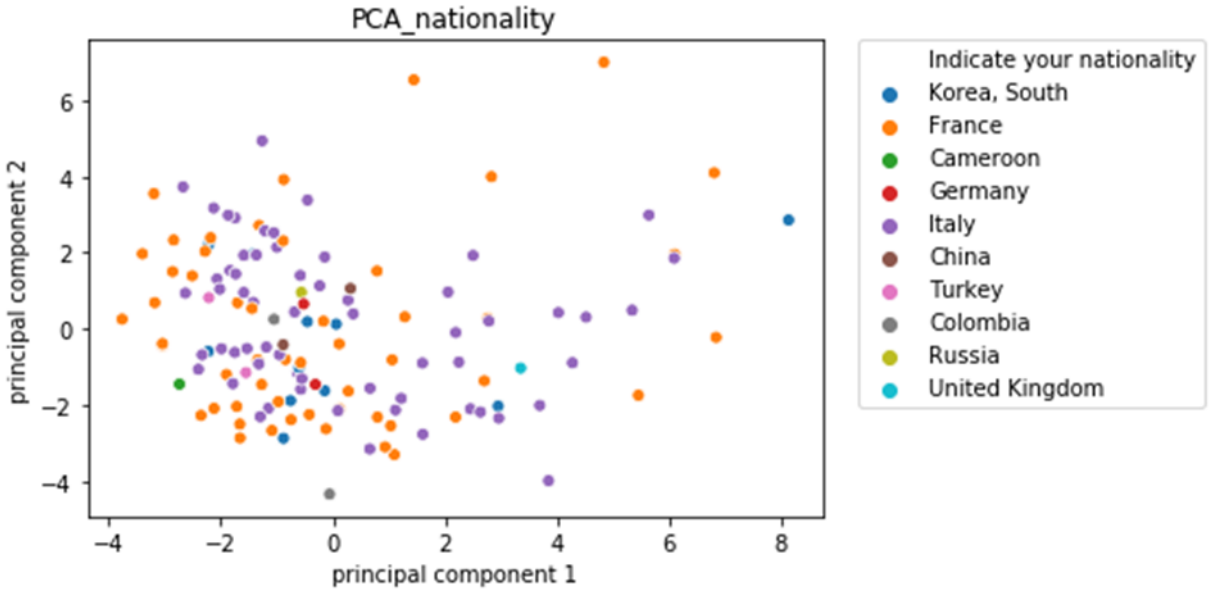


Figure 5-27: Principal Component analysis, nationality

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 5-27. As before, this does not imply that there is no cluster in the high-dimensional space.

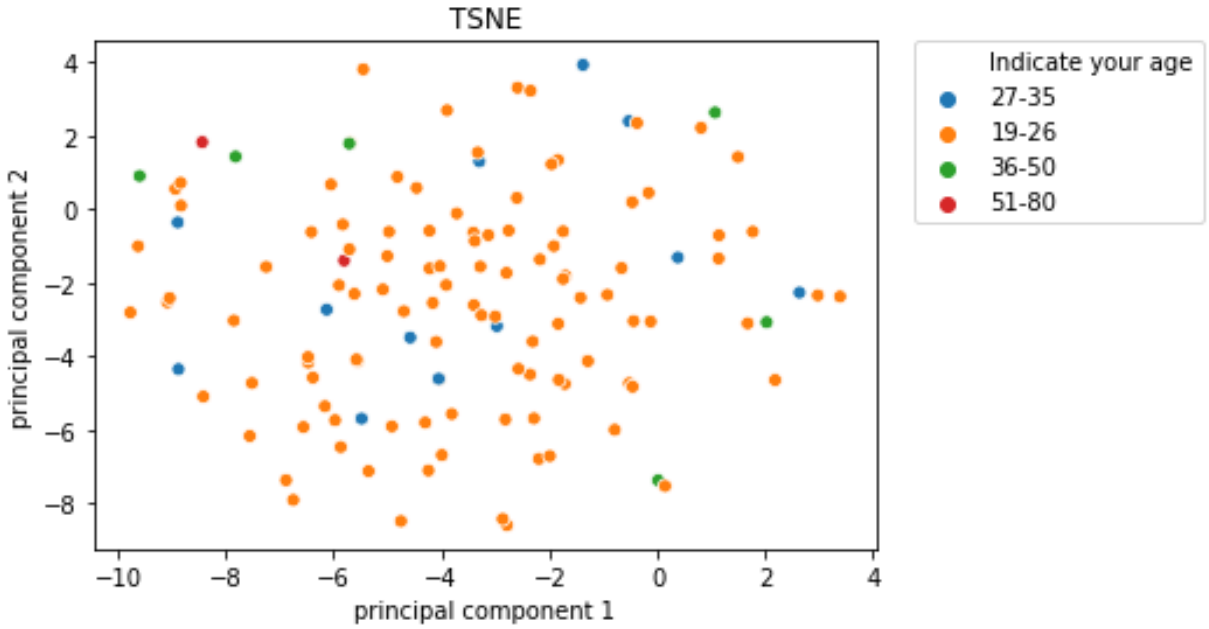


Figure 5-28: TSNE algorithm results, age labelling

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 $\boldsymbol{X} = \{x_1, \dots, x_N\}$, the t-SNE algorithm computes probabilities p_{ij}, proportional to the similarities of data x_i and x_j:

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

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

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) $y_1, \ldots, y_N$ with similarities q_{ij} in the new low-dimensional space close as much as possible to p_{ij}.

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

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.

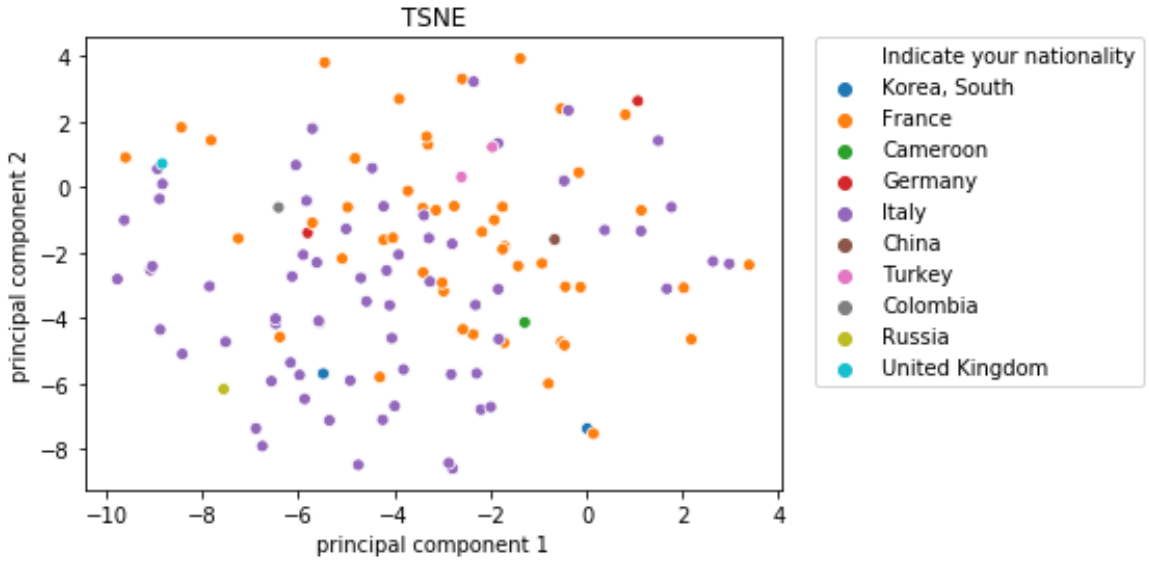


Figure 5-29: TSNE algorithm results, nationality labelling

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.

6 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.

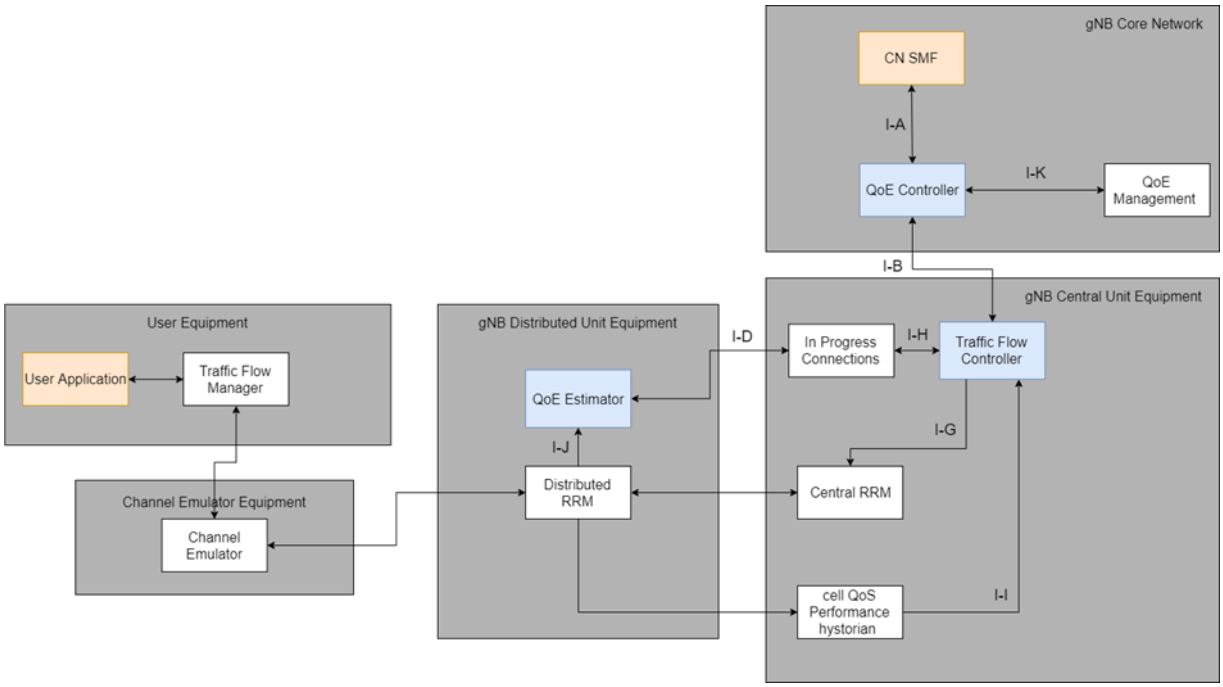


Figure 6-1: Preliminary architecture of the project EU-PoC

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 6-1 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.

6.1 UML flow chart of the multi-connectivity interfaces

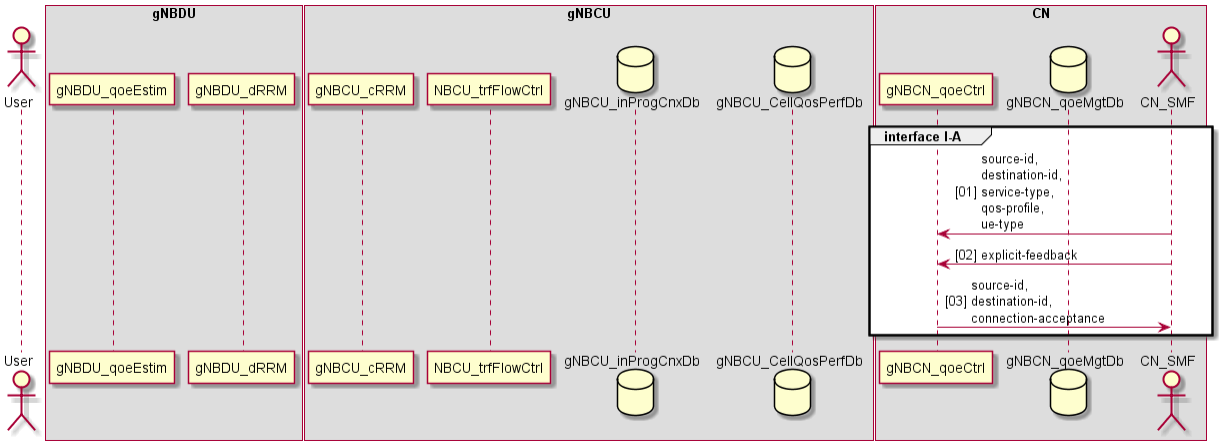


Figure 6-2: Interface IA flow chart

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

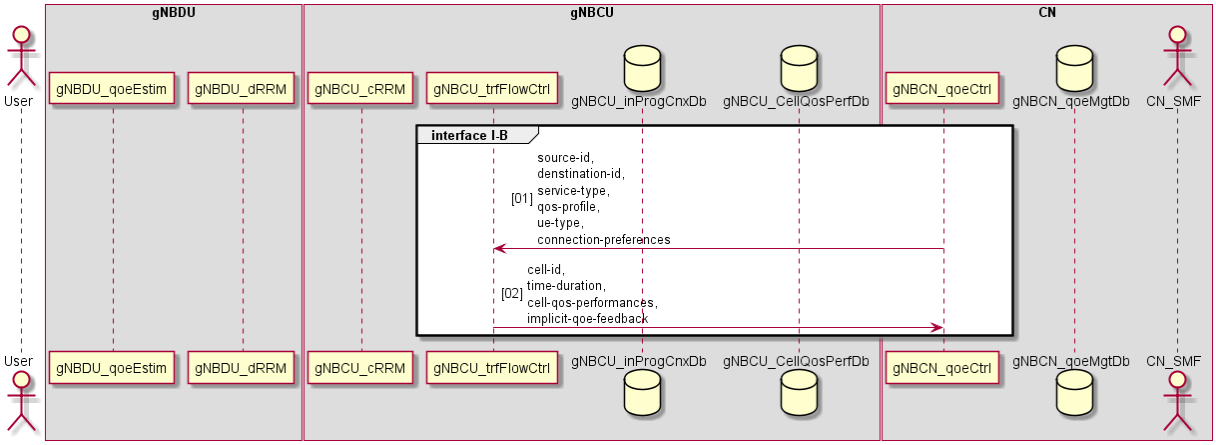


Figure 6-3: Interface IB flow chart

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.

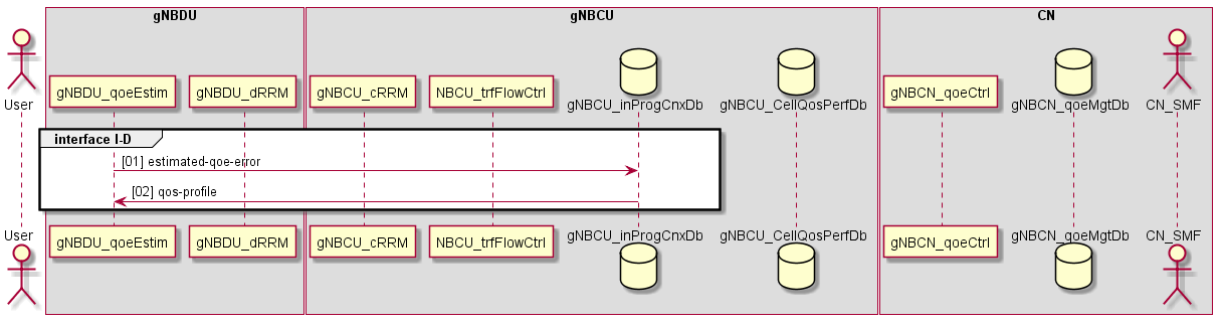


Figure 6-4: Interface ID flow chart

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).

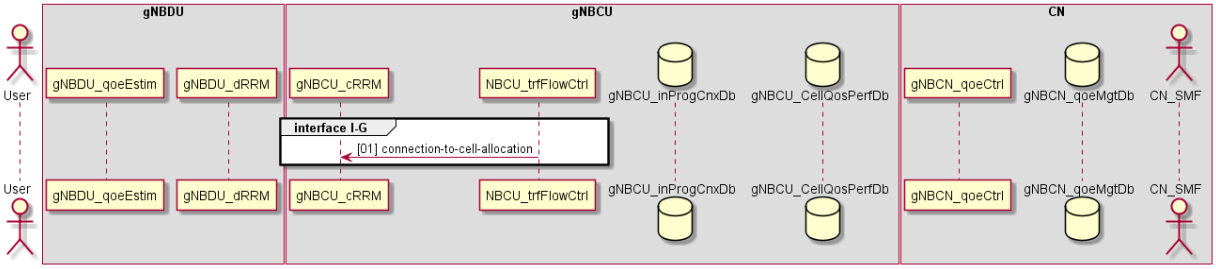


Figure 6-5: Interface IG flow chart

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.

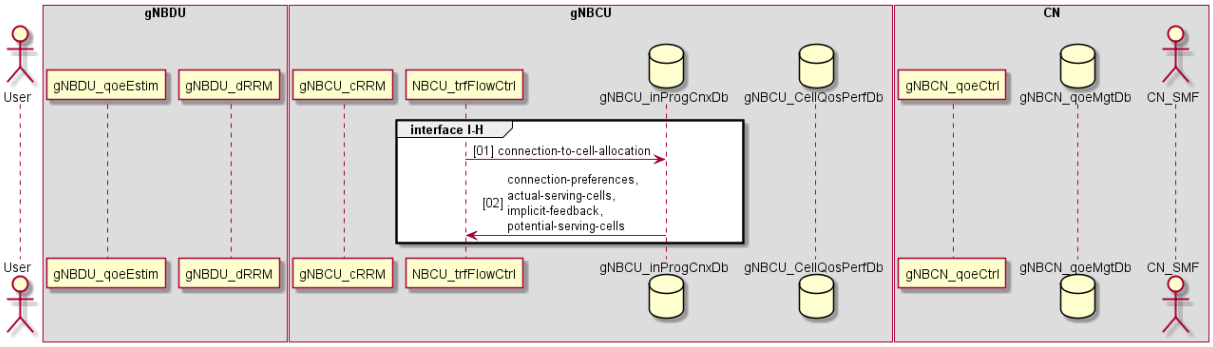


Figure 6-6: Interface IH flow chart

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.

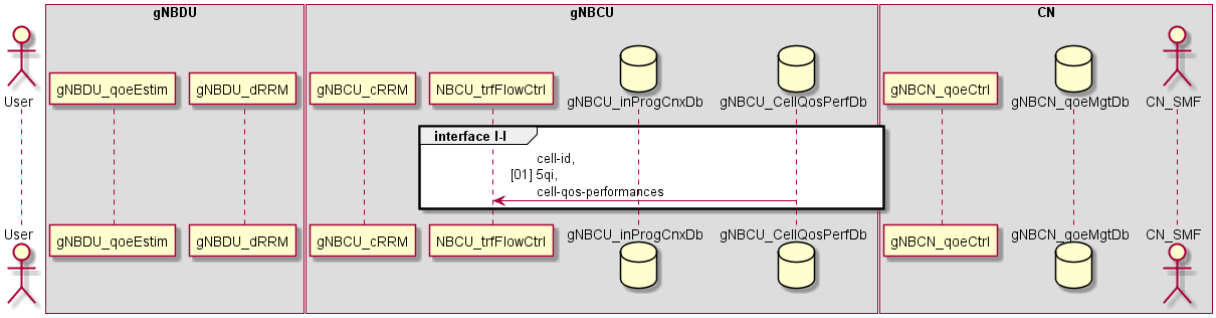


Figure 6-7: Interface II flow chart

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

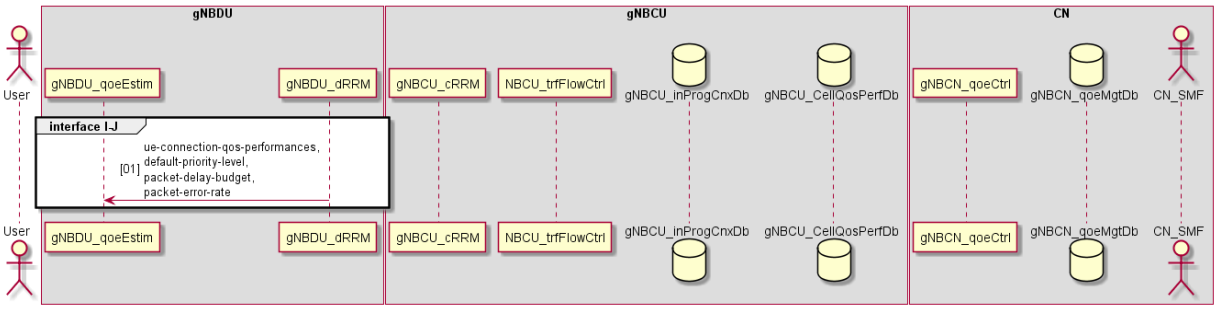


Figure 6-8: Interface IJ flow chart

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.

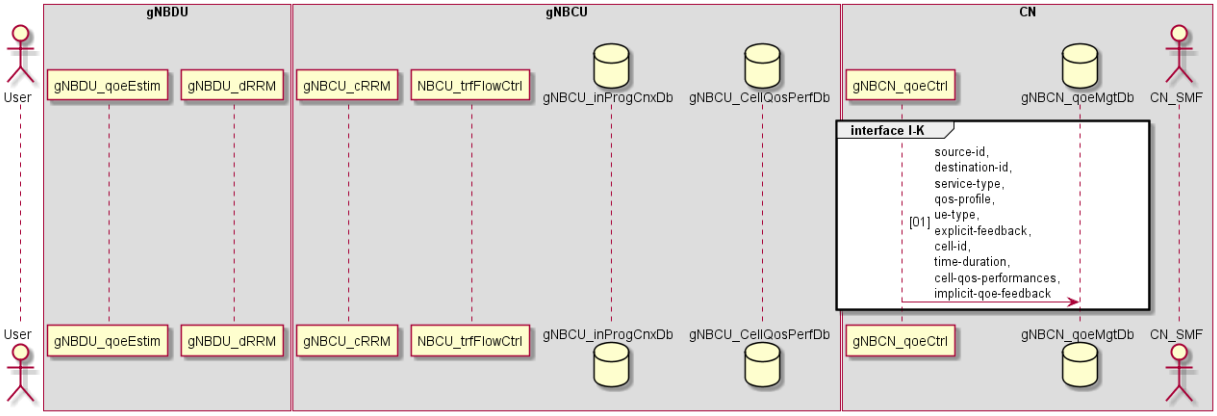


Figure 6-9**: Interface IK flow chart**

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.

7 Conclusions and future works

The present document reports 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 concepts of traffic steering, splitting and switching were presented and several algorithms were detailed and preliminarily evaluated on the basis of numerical simulations, as envisaged in the grant agreement. In particular, most of the algorithms presented are built on the basis of results from classic game theory and control theory, with a few also utilising results from reinforcement learning.

A total of seven network control approaches were presented, and three of which were already submitted for publication in scientific conferences and journals, while other two are in preparation.

The activities related to the QoE management of the project were also discussed, together with a preliminary analysis of the answers gathered by a questionnaire designed by the consortium for investigating the relation between QoS and QoE for the general public.

The future works that will be completed in the remaining months of the projects will focus on the refinement and consequent selection, starting from a more detailed set of testing and the GAP analysis reported in this deliverable, of one (or more) adequate algorithms to be deployed in the PoC of the project to demonstrate the functionalities developed in the entirety of WP4.

Eventual improvements to the algorithms, and their finalization, will be reported together with the implementation details in D4.3.

8 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.

[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.

[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.

[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.

[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, 2007, pp. 1–5.

[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.

[10] X. Gelabert, J. Perez-Romero, O. Sallent, and R. Agusti, “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.

[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.

[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.

[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.

[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.

[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), 2015, pp. 1–5.

[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.

[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] 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.

[24] 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.

[25] 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.

[26] 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.

[27] T. Roughgarden and É. Tardos, “How bad is selfish routing?,” J. ACM, vol. 49, no. 2, pp. 236–259, Mar. 2002.

[28] J. G. Wardrop, “Some Theoretical Aspects of Road Traffic Research,” Proc. Inst. Civ. Eng., vol. 1, no. 3, pp. 325–362, May 1952.

[29] T. Roughgarden, Selfish routing and the price of anarchy. MIT Press, 2005.

[30] H. Z. Aashtiani, H. Poorzahedy, and M. Nourinejad, “Extending Wardrop’s First Principle for Capacitated Networks,” Transp. A Transp. Sci., pp. 1–30, 2018.

[31] 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.

[32] 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.

[33] H. Kameda, J. Li, C. Kim, and Y. Zhang, “Optimal load balancing in distributed computer systems,” 2012.

[34] 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.

[35] 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.

[36] M. Beckmann, C. B. McGuire, and C. B. Winsten, “Studies in the Economics of Transportation,” 1956.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] S. Fischer and B. Vöcking, "Adaptive routing with stale information," Theor. Comput. Sci., vol. 410, no. 36, pp. 3357–3371, Aug. 2009.

[43] 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.

[44] 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.

[45] V. Sundarapandian, "An invariance principle for discrete-time nonlinear systems," Appl. Math. Lett., vol. 16, no. 1, pp. 85–91, 2003.

[46] 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.

[47] S. Nordholm, J. Nordberg, I. Claesson, and S. Nordebo, "Beamforming and Interference Cancellation for Capacity Gain in Mobile Networks," Ann. Oper. Res., vol. 98, no. 1/4, pp. 235–253, 2000.

[48] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in Machine Learning Proceedings 1994, 1994, pp. 157–163.

[49] R. Sutton and A. Barto, Reinforcement learning: An introduction. MIT Press, 2018.

[50] M. L. Littman, "Friend-or-Foe Q-learning in General-Sum Games," 2003.

[51] 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.

[52] R. K. Ahuja, "Minimax linear programming problem," Oper. Res. Lett., vol. 4, no. 3, pp. 131–134, 1985.

[53] 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.

[54] 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.

[55] 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

[56] 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.

[57] ITU-R S.2460-0 "Key elements for integration of satellite systems into Next Generation Access Technologies" Jul. 2019

[58] 3rd Generation Partnership Project: Techical Specification Group Radio Access Network: Study LTE-Based V2X Services: (Release 14), Standard 3GPP TR 36.885 V2.0.0, Jun. 2016.

[59] Saaty, Thomas L. "Decision making with the analytic hierarchy process." International journal of services sciences 1.1 (2008): 83-98.

[60] Saaty, Thomas L. What is the analytic hierarchy process? In: Mathematical models for decision support. Springer, Berlin, Heidelberg, 1988. p. 109-121.

[61] 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.

[62] Bressan, Alberto. "Noncooperative differential games. a tutorial." Department of Mathematics, Penn State University (2010).

[63] Geering, Hans P. Optimal control with engineering applications. Springer, 2007.

[64] Bonnans, Joseph Frédéric. Convex and stochastic optimization. Springer, 2019.

[65] 3GPP TR 38.913, Study on scenarios and requirements for next generation access technologies

[66] 3GPP TR 38.901 Study on channel model for frequencies from 0.5 to 100 GHz. Status: Under change control

[67] 3GPP TR 25.942 Radio Frequency (RF) system scenarios

[68] 3GPP TR 38.811 Study on New Radio (NR) to support non-terrestrial networks

ANNEX Preliminary JSON-like multi-connectivity interface description

```
---------- message IaTbdMsg1:
{
 "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 IaTbdMsg2:
{
 "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 IaTbdMsg3:
{
 "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 IbTbdMsg1:
{
 "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"
```

```
        },
        "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"
    }
   }
  }
}
------------- message IcTbdMsg1:
# TO FILLED BY CRAT, CEA and FhG IIS              # cde no CRAT input
 {
  "msg-type":{
    "IC-tbdmsg1"
  },
  "msg-src":{
    "gNBDU"
  },
  "msg-dst":{
    "UEs",
    "RAN"
  },
  "parameters":{
   {
     "bearer-configuration":{
       "tbd",
       "tbd"
     }
   },
   {
     "measurement-configuration":{
       "tbd",
       "tbd"
     }
   }
  }
}
------------- message IcTbdMsg2:
# TO FILLED BY CRAT, CEA and FhG IIS              # cde no CRAT input
 {
  "msg-type":{
    "IC-tbdmsg2"
  },
  "msg-src":{
    "UEs",
    "RAN"
  },
  "msg-dst":{
    "gNBDU"
  },
  "parameters":{
   {
     "measurement-report":{
       "tbd",
       "tbd"
     }
   }
  }
}

------------- message IdTbdMsg1:
# TO FILLED BY CRAT
 {
```

```
   "msg-type":{
     "type": "string",
     "value": "ID-tbdmsg1"
   },
   "msg-src":{
     "type": "string",
     "value": "gNBDU_qoeEstim"
   },
   "msg-dst":{
     "type": "string",
     "value": "gNBCU_inProgCnxDb"
   },
   "parameters":{
     {
       "estimated-qoe-error":{
         "type": "float64",
         "description": "Estimated QoE error computed by the QoE Estimator"
       }
     }
   }
}

---------- message IdTbdMsg2:
# TO FILLED BY CRAT
{
   "msg-type":{
     "type": "string",
     "value": "ID-tbdmsg2"
   },
   "msg-src":{
     "type": "string",
     "value": "gNBCU_inProgCnxDb"
   },
   "msg-dst":{
     "type": "string",
     "value": "gNBDU_qoeEstim"
   },
   "parameters":{
     {
       "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"
         }
       }
     }
   }
}

---------- message IeTbdMsg1:
# TO FILLED BY CRAT (?)       # cde no CRAT input
 {
  "msg-type":{
    "IE-tbdmsg1"
```

```
  },
  "msg-src":{
    "gNBCU_cRMM"
  },
  "msg-dst":{
    "gNBDU_dRMM"
  },
  "parameters":{
    {
      "bearer-configuration":{
        "tbd",
        "tbd"
      }
    }
  }
}
---------- message IeTbdMsg2:
# TO FILLED BY CRAT (?)       # cde no CRAT input
 {
  "msg-type":{
    "IE-tbdmsg2"
  },
  "msg-src":{
    "gNBDU_dRMM"
  },
  "msg-dst":{
    "gNBCU_cRMM"
  },
  "parameters":{
    {
      "measurement-reports":{
        "tbd",
        "tbd"
      }
    }
  }
}
---------- message IgTbdMsg1:
 {
  "msg-type":{
    "type": "string",
    "value": "IG-connections-to-cells-allocation"
  },
  "msg-src":{
    "type": "string",
    "value": "gNBCU_trfFlowCtrl"
  },
  "msg-dst":{
    "type": "string",
    "value": "gNBCU_cRMM"
  },
  "parameters":{
    {
      "connections-to-cells-allocation":{
        "type": "array",
        "items": {
          "type": "float64",
          "max_value": 1,
          "min_value": 0
        },
        "description": "percentage of the traffic relevant to the considered connection which has to be managed by each of
        the Cells of the gNB-CU area"
      }
    }
  }
}
---------- message IhTbdMsg1:
 {
  "msg-type":{
    "type": "string",
    "value": "IH-connections-to-cells-allocation"
  },
  "msg-src":{
    "type": "string",
    "value": "gNBCU_trfFlowCtrl"
  },
  "msg-dst":{
    "type": "string",
    "value": "gNBCU_inProgCnxDb"
  },
  "parameters":{
    {
      "connection-to-cells-allocation":{
        "type": "array",
        "items": {
          "type": "float64",
          "max_value": 1,
          "min_value": 0
        },
        "description": "percentage of the traffic relevant to the considered connection which has to be managed by each of
        the Cells of the gNB-CU area"
      }
    }
  }
}
---------- message IhTbdMsg2:
 {
  "msg-type":{
    "type": "string",
    "value": "IH-connections-ongoinginformation"
  },
```

```
  "msg-src":{
    "type": "string",
    "value": "gNBCU_inProgCnxDb"
  },
  "msg-dst":{
    "type": "string",
    "value": "gNBCU_trfFlowCtrl"
  },
  "parameters":{
    {
      "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, ...)"
      }
    },
    {
      "actual-serving-cells":{
        "type": "array",
        "values":{
          "cell-id":{
            "type": "unsigned integer64",
            "min_value": 0,
            "description": "Cell identification number" # cde is there a description by previous value?
          }
        }
      }
    },
    {
      "implicit-feedback":{
        "type": "float64",
        "description": "feedbacks related to the Perceived QoE computed by the QoE Estimation"
      }
    },
    {
      "potential-serving-cells":{
        "type": "array",
        "values":{
          "cell-id":{
            "type": "unsigned integer64",
            "min_value": 0,
            "description": "Cell identification number" # cde is there a description by previous value?
          }
        }
      }
    }
  }
}
---------- message IiTbdMsg1:
{
  "msg-type":{
    "type": "string",
    "value": "Ii-cell-status"
  },
  "msg-src":{
    "type": "string",
    "value": "gNBCU_CellQosPerfDb"
  },
  "msg-dst":{
    "type": "string",
    "value": "gNBCU_trfFlowCtrl"
  },
  "parameters":{
    {
      "cell-id":{
        "type": "unsigned integer64",
        "min_value": 0,
        "description": "Cell identification number" # cde is there a description by previous value?
      }
    },
    {
      "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"
      }
    },
    {
      "cell-qos-performances":{
        "resource_type":{
          "type":"boolean",
          "description": "GBR or non-GBR"
        },
        "default-priority-level":{
          "type": "unsigned integer64"
        },
        "packet-delay-budget":{
          "type": "unsigned integer64"
        },
```

```
        "packet-error-rate":{
          "type": "float64"
        },

        "description": "Measured QoS parameters"
      }
    }
  }
}
-------------- message IjTbdMsg1:
{
 "msg-type":{
   "type": "string",
   "value": "IJ-tbdmsg1"
 },
 "msg-src":{
   "type": "string",
   "value":  "gNBDU_dRRM"
 },
 "msg-dst":{
   "type": "string",
   "value": "gNBDU_qoeEstim"
 },
 "parameters":{
   {
     "ue-connection-qos-performances":{
       "resource_type":{
         "type":"boolean",
         "description": "GBR or non-GBR"
       },
       "default-priority-level":{
         "type": "unsigned integer64"
       },
       "packet-delay-budget":{
         "type": "unsigned integer64"
       },
       "packet-error-rate":{
         "type": "float64"
       },

       "description": "Requested QoS parameters"
     }
   }
  }
}

-------------- message IkTbdMsg1:
{
 "msg-type":{
   "type": "string",
   "value": "IK-tbdmsg1"
 },
 "msg-src":{
   "type": "string",
   "value": "gNBCN_qoeCtrl"
 },
 "msg-dst":{
   "type": "string",
   "value": "gNBCN_qoeMgtDb"
 },
 "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?
    }
   }
   {
    "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"
      }
    }
  }
 }
}
```