Sapienza University of Rome

Department of Computer, Control, and Management Engineering Antonio Ruberti
PhD in Automatic Control, Bioengineering and Operations Research

Thesis For The Degree Of Doctor Of Philosophy

Network Control Systems for 5G Multi-Connectivity

Advisor
Prof. Francesco Delli Priscoli

Co-Advisor
Dr. Alessandro Giuseppi

Candidate
Antonio Ornatelli
1465696

Academic Year MMXVIII-MMXX (XXXIII cycle)

Thesis defended on 13 July 2021
in front of a Board of Examiners composed by:

Prof. Fabio Tardella, Sapienza University of Rome (chairman)
Prof. Giuseppe Baselli, Polytechnic University of Milan
Prof. Stefano Panzieri, Roma Tre University

Network Control Systems for 5G Multi-Connectivity
Ph.D. thesis. Sapienza - University of Rome

Author's email: antonio.ornatelli@uniroma1.it

Dedicated to
my family

Acknowledgements

A dutiful and heartfelt thanks go to my Tutor, Professor Francesco Delli Priscoli, always available and attentive to my academic, professional, and human growth. A special thanks go to Alessandro Giuseppi, for the continuous technical support during these years, the friendship and the continuous confrontations since the first year of this long university career. In the same way, I would like to thank Federico Lisi, colleague and friend, who with his advice helped me to grow both as a professional and as a person. I would like to thank all the colleagues of the Network Control Laboratory, Antonio Pietrabissa, Alessandro Di Giorgio, Martina Panfili, Vincenzo Suraci, Francesco Liberati, Andrea Tortorelli, and Roberto Germanà.

A special thanks go to all the people I have met over the years in Thales Alenia Space, where I spent most of this experience. First of all, I would like to thank Giuseppe Tomasicchio who allowed me to start this journey and for being a guide for my first steps in the company, also I would like to thank Vincenzo Marziale for being an always available and precious point of reference in my professional growth. Thanks also go to Guglielmo Lulli and Felice Rosato, who were for me two lighthouses in the fantastic night sea of digital electronics.

A deep thanks go to the friends of the B.O.S.E., Marco La Ferla, Felice Rosato, Luca Zuccaro, and Dino Quaranta for their friendship, continuous support, and the ability to lighten the long working days with a smile.

I would also like to thank the people who made possible the start of my new adventure in an important company as Enel is, in particular, I would like to thank Mike Fife, and Fernando Gargiulo.

Finally, I would like to thank my family, my parents who have always supported me in my life, my wife Jessica, who supports and bears me every day, and my son Giacomo, who makes every day brighter and happier.

Antonio Ornatelli, Rome, March 2021

Ringraziamenti

Un doveroso e sentito ringraziamento va al mio Tutor, il Professore Francesco Delli Priscoli, sempre disponibile ed attento alla mia crescita didattica, professionale ed umana. Un grazie speciale va ad Alessandro Giuseppi, per il continuo supporto tecnico durante questi anni, per l'amicizia e i continui confronti sin dal primo anno di questo lungo percorso universitario. Allo stesso modo vorrei ringraziare Federico Lisi collega e amico, che con i suoi consigli mi ha aiutato a crescere sia come professionista che come persona. Vorrei inoltre ringraziare tutti i colleghi del Network Control Laboratory, Antonio Pietrabissa, Alessandro Di Giorgio, Martina Panfili, Vincenzo Suraci, Francesco Liberati, Andrea Tortorelli e Roberto Germanà.
Un ringraziamento speciale va a tutte le persone che ho conosciuto in questi anni in Thales Alenia Space, dove ho trascorso gran parte di questa esperienza. Vorrei ringraziare prima di tutti Giuseppe Tomasicchio che mi ha dato l'opportunità di iniziare questo percorso e per essere stato una guida per i miei primi passi in azienda, inoltre vorrei ringraziare Vincenzo Marziale per esere stato un punto di riferimento sempre disponibile e prezioso nella mia crescita professionale. Un grazie va inoltre a Guglielmo Lulli e a Felice Rosato, che sono stati per me due fari nel fantastico mare notturno dell'elettronica digitale.
Un ringraziamento profondo va agli amici del B.O.S.E., Marco La Ferla, Felice Rosato, Luca Zuccaro e Dino Quaranta per la loro amicizia, il loro continuo supporto e la capacità di alleggerire le lunghe giornate di lavoro con un sorriso.
Vorrei inoltre ringraziare le persone che hanno reso possibile l'inizio di una nuova avventura in una azienda importante come Enel, in particolare vorrei ringraziare Mike Fife e Fernando Gargiulo.
Infine, vorrei ringraziare la mia famiglia, i miei genitori che mi hanno sempre supportato nel mio percorso di vita, mia moglie Jessica, che ogni giorno mi supporta e sopporta, e mio figlio Giacomo, che rende ogni giornata più luminosa e allegra.

Antonio Ornatelli, Roma, Marzo 2021

Abstract

This thesis aims to develop a network control framework when considering heterogeneous access network technologies. The scenario of interest is when several services are using a common shared and large-scale communication network and the users who are benefiting from these services have different connection requirements. The reason behind the need for a common network satisfying several service requirements at the same time, is the continuous transformation of physical infrastructures such as energy distribution, entertainment, communication, transportation, and many others, in Cyber-Physical Systems interacting with each other. This transformation increases the need of the different industries in the deployment of large-scale communication networks, to connect the high number of sensors and actuators integrated into the physical infrastructure, introducing the so-called Networked Control System. In this framework, when sensors and actuators are used to perform and act decision-making on large-scale systems, several problems introduced by the networking arises. In particular, when unpredictable and uncontrollable packet drop, latency, or data corruption occur, either between sensors or actuators and the autonomous decision-maker, the resulting control actions may perform undesired effects on the controlled physical system introducing instability. These problems can be considered during the autonomous decision-maker design, implying higher design and implementation complexity, that can degrade the performances concerning the controlled physical systems. Another solution is the development of an underlying network control able to provide the needed networking performances to the decision-maker as a service. This solution allows the decision-maker to see the network as a black box that meets its requirements. This network control problem can be faced considering different aspects because of the high number of problems that can be addressed to have an efficient and reliable network, one of the main problems to be addressed is resource allocation. This work deal with the resource allocation considering multi-connectivity, defined as the management of the network resources when the users can connect simultaneously to different access points, even belonging to different radio technologies, to satisfy users' requirements, and considering network constraints and efficiency. Furthermore, in this work the reference network is the fifth-generation technology standard for broadband cellular networks, known as 5G Network, standardized by the 3GPP association. The 5G Network is designed to provide a general-purpose connectivity platform, considered as the key enabling technology to support markets such as automotive, energy, food and agriculture, city management, government, healthcare, manufacturing, and public transportation, through the digital transformation of their infrastructure and business processes, providing a large scale, efficient and reliable communication platform, able to satisfy several users' requirements such as high reliability, ultra-low latency, high bandwidth, and mobility. The thesis presents the definition of network architecture, compliant with the 3GPP standards and able to accommodate multi-connectivity algorithms, and control algorithms able to manage the multi-connectivity considering both users' requirements and network conditions, using either model-based and data-driven techniques, using centralized, distributed and hierarchical solutions, allowing the deployment of the control algorithms either in the cloud or at the edge of the network, on the basis of the resources and performances needed in each particular scenario.

Keywords: Network Control, Multi-Connectivity, 5G Heterogeneous Networks, QoS Requirements, Resource Allocation, Model-Based Decision Making, Data-Driven Decision Making.

Contents

List of Figures

Acronyms

3GPP 3rd Generation Partnership Project

5G fifth-generation of mobile networks

5G-ALLSTAR 5G AgiLe and fLexible integration of SaTellite And cellular

5GPPP 5G Infrastructure Public Private Partnership

AHP Analytic Hierarchy Process

AIV Air Interface Variant

AMF Access and Mobility Management Function

AP Access Point

ASPC Adaptive Step Power Control

BER Bit Error Rate

BS Base Station

C-RAN Cloud Radio Access Network

CDMA Code Division Multiple Access

CI Consistency Index

CN Core Network

CogNet Cognitive Networks

CP Control Plane

CPS Cyber-Physical Systems

CR Consistency Ratio

CU Central Unit

DAE Differential Algebraic Equations

DL Down Link

DRB Data Radio Bearer

DU Distributed Unit

E-UTRA Evolved Universal Mobile Telecommunications System

E2E End-to-End

FLC Fuzzy Logic Controller

FS Fast Switch

GEO Geostationary Earth Orbit

gNB Next generation NodeB

GRA Grey Relational Analysis

HARQ Hybrid Automatic Repeat Request

HH Hard Handover

HMM Hidden Markov Model

I-MPC Iterative Model Predictive Control

IoT internet of Things

IP internet Protocol

KPI Key Performance Indicator

LEO Low Earth Orbit

LQ Linear Quadratic

LTE Long Term Evolution

MAC Madium Access Control

MADM Multi Attribute Decision Making

MARL Multi-Agent RL

MC Multi-Connectivity

MCG Master Cells Group

MdO Multi-domain Orchestrator

MDP Markov Decision Process

ML Machine Learning

MMDP multi-agent MDP

MN Main Node

MPC Model Predictive Control

N-NF Nearest Not-Full

NAS Non-Access Stratum

NCS Networked Control Systems

NE Nash Equilibrium

NFV Network Function Virtualization

NG New Generation

NR New Radio

NTN Non-Terrestrial Network

PDCP Packet Data Convergence Protocol

PDR Packet Detection Rules

PDU Protocol Data Unit

PHY Physical Layer

POMDP Partially Observable Markov Decision Process

QoE Quality of Experience

QoS Quality-of-Service

RAN Radio Access Network

RAT Radio Access Technology

RF Radio Frequency

RI Random Consistency Index

RL Reinforcement Learning

RLC Radio Link Control

RRC Radio Resource Control

RRM Radio Resource Management

SCG Secondary Cells Group

SDAP Service Data Adaptation Protocol

SDN Software Defined Networking

SINR Signal-to-Interference-plus-Noise Ratio

SMF Session Management Function

SN Secondary Node

TD Temporal-Difference

UE User Equipment

UL Up Link

UP User Plane

UPF User Plane Function

Uu Air Interface Physical Layer

WLAN Wireless Local Area Network

Chapter 1

Introduction

1.1 Scenario and Motivations

The growing population's needs and demands in terms of energy, transportation, communications, food, and entertainments services, made the respective infrastructures (e.g., Power Grids, Intelligent Transportation Systems, Satellite Constellations, etc) a complex interconnection between computing devices and physical systems, facilitating real-time plant monitoring, closed-loop process management and human decision-makers interaction with large scale and complex infrastructures. This evolution imposes the deployment of the so-called *Cyber-Physical Systems (CPS)*, depicted in the figure 1.1. The term CPS [1] was coined by Helen Gill at the National Science Foundation in the U.S. to refer to the integration of computation with physical processes.

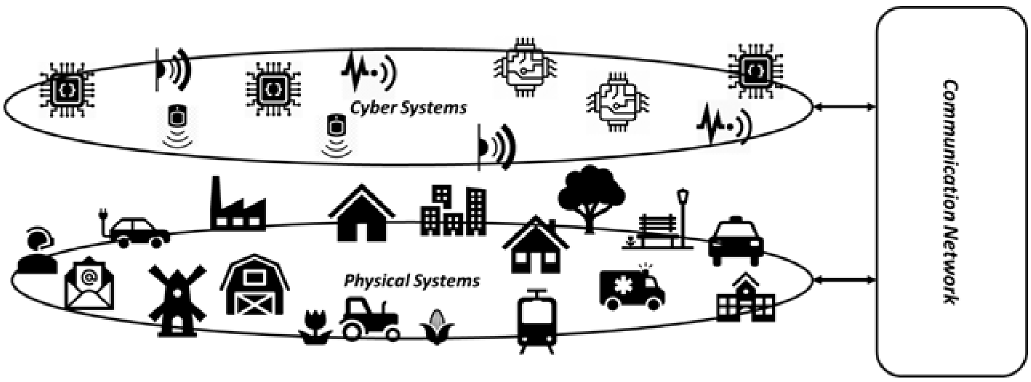
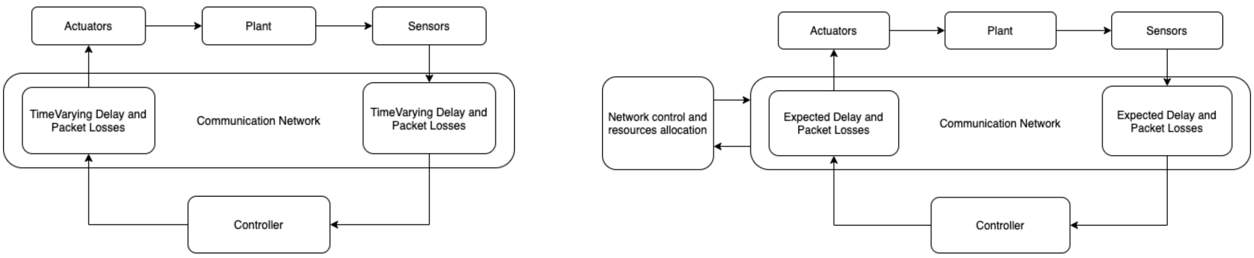


Figure 1.1: Cyber-Physical Systems

CPS can be considered as the *intersection between physical and cyber systems, involving communication and networking aspects.* The integration of such systems, from mathematical point of view, makes the meeting of different disciplines happen [2]: continuous time modelling introduced by the physical systems, discrete time modelling introduced by the cyber systems and probabilistic and game theoretic modelling introduced by the networking; making the modelling, the design and the analysis of the CPS a challenging task. Said task is made worse by the resilience needs [3], that implies the use of automatic control strategies and diagnostic to react to security attacks and random faults in both networking and computational devices.

The introduction of CPS quickly increased the interest to, and the development of a special kind of control engineering discipline: Networked Control Systems (NCS). NCS [4, 5], can be defined as a control system in which the feedback loop is closed via a communication link over a network, possibly shared with other computing devices. In NCS, the control signalling, i.e. input reference signal, control signal and plant output, are exchanged via real-time networks, connecting sensors, actuators, plant and computing devices, introducing large physical space deployment, interconnection complexity and real-time needs. Without considering network topology or human interaction in the loop, NCS can be categorized in [4]: i) time-critical applications such as military, space, remote guidance, etc; ii) non-real-time control such as data storage, sensor data collection, e-mail, etc. However, network reliability is fundamental in both cases because network behaviour effects stability, safety and performances of the NCS. In light of the above observations Quality-of-Service (QoS) management in the underlying networks becames crucial to avoid network unreliability and time-dependent QoS degradation, in terms of delays, jitter, losses, and errors due to communication interference, queues congestion, and poor routing or scheduling strategies. The consideration of communication delay, bandwidth limitation, and resource scheduling add a burdening amount of complexity in the controller design process, in [5] the use of a middleware platform is proposed. This middleware platform is an abstraction layer over the underlying computing devices and the communication network, that allows the controller designer to take care only about the plant to be controlled, transferring all the communication and computing issues to a set of services provided by the middleware, the QoS management is one of these services. However, to guarantee a QoS level it is needed to control the communication network itself, to counteract interference, congestion, and link performances degradation. To summarize a standard NCS is depicted in figure 1.2a, where the controller shall be designed considering the plant to be controlled and the time-varying network behavior. On the other hand, in figure 1.2b, a network controller is considered, and the plant controller can be designed only considering the process to be controlled because the network controller drives the communication network in the required conditions.



(a) Networked Control System

(b) Networked Control System Network Controller

Figure 1.2: Networked Control System solutions

Communication networks are widely used as a tool in modern control systems as described above, just as the reverse is true. Control techniques are applied either in wireless or wired communication networks [6–16], for load balancing, power control, flow control, energy management, admission control, etc. Since the new generation communication networks involve several entities, large scale deployment and heterogeneous technologies to satisfy heterogeneous users' requirements in a dynamic environment, the management has higher complexity. Indeed, the needs of self-managing and self-adapting networks arise, with the introduction of the so-called Cognitive Networks (CogNet) [17]. CogNet are defined in [17] as *"... a network that can perceive current network conditions, and then plan, decide and act on those conditions. The network can learn from these adaptations and use them to make future decisions, all while considering end-to-end goals"*. To make the networks "cognitive" several advanced control methods are being used, involving optimization, game theory, and machine learning techniques [18, 19].

A perfect example of middleware, designed and developed as a CogNet [20], is the 5G [21] designed to satisfy the wide range of needs of vertical industries (remote control, automotive, energy, healthcare, agriculture, etc). In this prospective, the objective of the 5G mobile network design is to provide a multi-service network, aimed at satisfying, on the same common physical infrastructure, simultaneously different and conflicting requirements of several services. Network Slicing is the reference paradigm to provide the needed network flexibility, [22] supporting several defined QoS levels simultaneously and keeping them logically isolated by the same physical network. This is possible by using software virtualization and employing cognitive resources management across multiple heterogeneous networks all of which have their own peculiarities and requirements. An example can be considered from the 3rd Generation Partnership Project (3GPP) standard that regulates the 5G design and development [23], considering two NCS for different purposes and with different QoS defined in the standard: i) Process Automation, with packet delay budget 50 milliseconds and packet error rate 10-3; ii) Intelligent Transport System, with packet delay budget 10 milliseconds and packet error rate 10-5. These two services will be able to use the same physical network, experiencing two different QoS satisfying their requirements, this tanks to a cognitive network controller able to manage the needed resources of the underlying shared network, preventing and mitigating the undesired network's behavior.

1.2 5G Network

The 5G system [24], is divided in Core Network (CN) and Radio Access Network (RAN), the CN include the set of functions for the operation of 5G, the RAN include the radio technologies and functions to connect users to the 5G-CN. CN and RAN are divided in Control Plane (CP) and User Plane (UP), the CP is the set of functions and protocols that provide the control and management functions and convey the control signalling through the network, the UP is the set of functions/elements and protocols that covey the user data through the network and is configured by the CP functions. The 5G network adopts also the Software Defined Networking (SDN) technology: an approach that facilitates network management and enables programmatically efficient network configuration to improve network performance and monitoring; and the Network Function Virtualization (NFV): a network architecture concept that uses the technologies of IT virtualization to virtualize entire classes of network node functions into building blocks that may connect, or chain together, to create communication services. Indeed, in the 3GPP standard the 5G system architecture is defined to support technologies such as NFV and SDN, providing a service-based interaction between the network functionalities. The advantages to have a separated CP/UP and NFV architecture, include the capability to develop and upgrade the two plane's functions independently. Moreover, this separation allows the capability to select/use the CP functions independently by the selected/used UP functions and vice versa. This allows, using SDN technology, different configurations on the same network, also at the same time, allowing Network slicing, i.e. multiple virtual networks, customized to meet the specific needs of applications, services, devices, customers or operators, created on top of a common shared physical infrastructure.

Figure 1.3 shows the 5G System Architecture with the main Network Functionalities, in particular for the CN the main functionalities are:

- Access and Mobility Management Function (AMF): Acts as the termination point for the Non-Access Stratum (NAS) signaling, and acts the mobility management;
- Session Management Function (SMF): Supports the establishment, modification, and release of a data session, configuration of traffic steering policies at the User Plane Function (UPF), UE internet Protocol (IP) address allocation, and policy enforcement.
- UPF: Serves as the anchor point for Intra/inter-Radio Access Technology (RAT) mobility, packet routing, traffic reporting, handles user plane QoS;

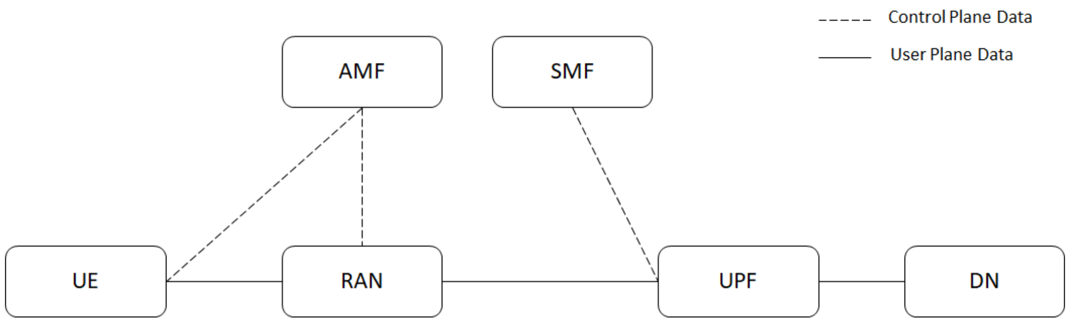


Figure 1.3: 5G System Architecture

In figure 1.4 the RAN is represented [25]. The RANs are composed of gNBs, the gNBs can be interconnected, and each gNB contains a CP and UP, each gNB can be divided into central unit gNB-Central Unit (CU) and Distributed Unit (DU), these units contain a set of functionalities of the control plane and user plane, based on the needs different split of the gNB functionalities among central and distributed units can be done. Each of these components can communicate with the other using standardized service-based interfaces.

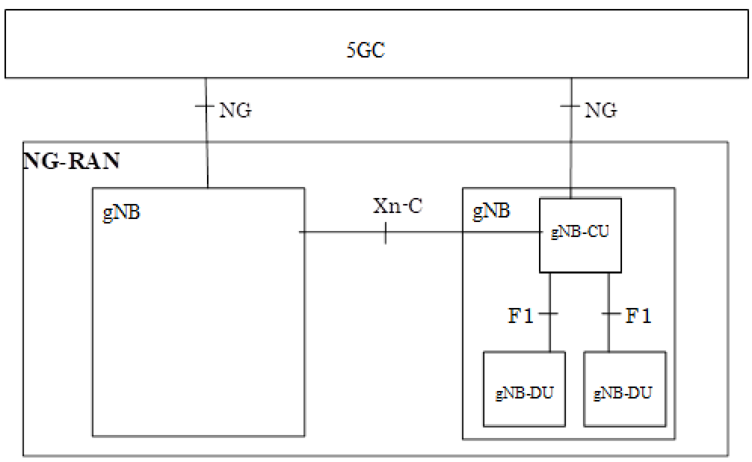


Figure 1.4: RAN overall architecture

The RANs, using the gNBs, contain the New Radio (NR) protocol stack, and its first role is to connect the UEs to the Data Network (i.e., Internet and Data/Application Servers). The RAN's main functionalities are [26]:

- Functions for Radio Resource Management: Radio Bearer Control, Radio Admission Control, Connection Mobility Control, Dynamic allocation of resources to UEs in both uplink and downlink (scheduling);
- IP header compression, encryption and integrity protection of data;
- Routing of User Plane data towards UPF(s);
- Routing of Control Plane information towards AMF;
- Connection setup and release;
- Measurement and measurement reporting configuration for mobility and scheduling;
- Transport level packet marking in the uplink;
- QoS Flow management and mapping to data radio bearers;

The 5G networks are founded on service-based traffic management using different levels of QoS differentiation [24]. The QoS provides a way to differentiate the characteristics of the services data flows, in the 5G network the QoS forwarding behavior is defined by the 5QI, this number defines different attributes such as packet delay, packet error rate, and priority level. The QoS Flows are defined to differentiate the type of traffic in the connection, for example in a video conference between two PCs is created one connection between the PCs, this connection is composed of three QoS flows: one QoS flow for the audio, one QoS flow for the video and one QoS flow for the data; each QoS flow has different QoS attributes defined by its 5QI. The data flows are managed at different levels in the network as reported in figure 1.5 from [24]. The first step is in the CogNet, where the UPF assign to each data flow (typically an IP flow) a 5QI value, considering the Packet Detection Rules (PDR)s, where a PDR is a set of the attribute of the IP packet to identify a specific packet data flow (e.g. PDR={source_address=10.0.0.1, destination_address=20.0.0.5, traffic_class=2}; 5QI=3). After this each marked QoS flow arrives at the RAN, there the flows are distributed among the different radio access resources through the Data Radio Bearer (DRB), where a DRB is a radio path (the set of resources and their configuration) between the access network and the user equipment.

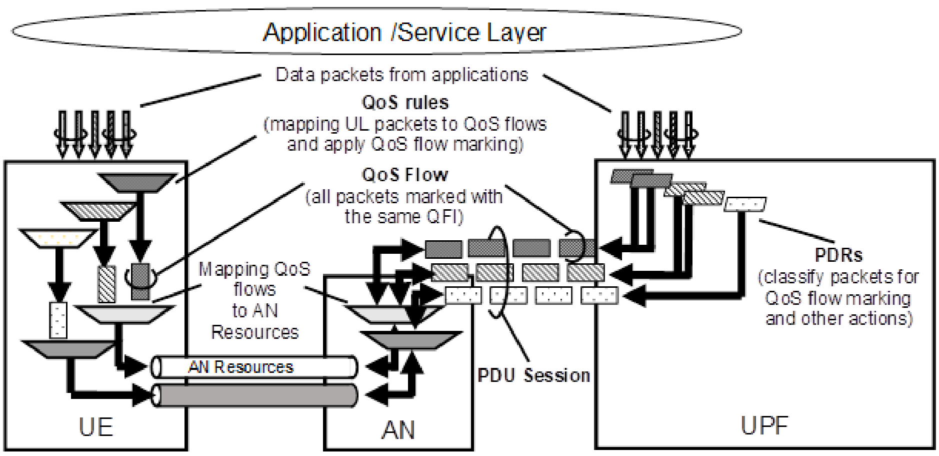


Figure 1.5: The principle for classification and User Plane marking for QoS Flows and mapping to AN Resources

5G network aim is the provision of a common shared communication network able to satisfy at the same time different service requirements. In particular, the 5G network will integrate networking, computing, and storage resources into one unified infrastructure, providing through resources virtualization the needed service to satisfy the users' requirements [22]. This approach allows the unification of all the resources in a common platform that, with a dynamic and optimized management, can make the most of these resources and promote the convergence of the fixed, mobile, and broadcast service in a common infrastructure. The network is then proposed as a middleware platform for all possible services, that are classified into five main categories [27]:

1. *Enhanced Mobile Broadband*: data-driven services requiring high data rates across a wide coverage area, e.g. Mobile Broadband, UHD/Hologram, High-mobility, Virtual Presence;

2. *Critical Communications*: low latency and high-reliability services, e.g. Interactive Game/Sports, Industrial Control, Remote Control, Drone/Robot/Vehicle, Emergency;

3. *Massive Machine Type Communications*: services with a high amount of UEs in a relatively small area, with sporadic and low power transmissions, e.g. Subway/Stadium Service, eHealth, Wearables, Inventory Control;

4. *Enhancement Vehicle-to-Everything*: services for moving vehicles communicating with several systems, e.g. Autonomous Driving, safety and non-safety aspects associated with the vehicle;

5. *Network Operation*: services required by some network components or layers, e.g. Network Slicing, Routing, Migration and Inter-working, Energy Saving;

to allow these heterogeneous and complex service requirements the networks shall be provided with advanced functionalities such as [27]:

- *Flexible application traffic routing*: the network shall be able to transfer these data traffic flexibly and efficiently;

- *Flexibility and scalability*: the system shall flexibly scale with various levels of demand to avoid localized under-utilization of resources;

- *Best Connection per Traffic Type*: the network shall be able to select the best connection based on the traffic application;

- *multiple RAT connectivity and RAT selection*: Multiple RAT connectivity is beneficial for increasing throughput and reliability. The network capability to select which data flow goes over which RAT benefits both network and users, providing load balancing between networks and higher QoS for the users;

- *Higher User Mobility*: the network mobility management shall be seamless and effective, to provide entertainments in high mobility environments such as trains, cars and airplanes, and connectivity in the automotive sector;

- *Connectivity Everywhere*: the network shall dramatically reduce the area where the connectivity is not provided, the use of new RATs such as satellite is considered;

- *Context Awareness to support network elasticity*: context-awareness could enable rapid network configuration and provide the expected experience for the multiple services/applications. The information gathered by sensors (e.g., accelerometer, gyroscope, magnetometer, barometer, proximity sensor, GPS) and connectivity technologies (e.g., Bluetooth, WIFI, NFC) installed on the UEs can be not only useful to the Apps installed in the smartphone but also to the networking technologies;

- *Priority, QoS and Policy Control*: To cope with diverse service requirements, it is imperative that intelligent decisions are made at the network such as allocation of resources, scheduling of resources, and adapt the network to meet these service requirements;

Indeed, a key enabling technology to satisfy the different services required in the 5G network (e. g., multimedia, augmented reality, healthcare, internet of Things (IoT), automotive, manufacturing, ...) is the use of Heterogeneous Multi-Access Radio Technologies, that exploit the characteristics of the different technologies, for example terrestrial radio technologies can provide low latency (10 milliseconds) and high throughput (Gbps data rates) in short-range (hundreds of meters), but the performances are very poor in longer area (e. g. hundreds of kbps in tens of kilometres), in the other hand, the satellite provides hundreds of Mbps in very large area (at least thousands of kilometres), but introducing long latency (about 50 milliseconds in low earth orbit and 600 milliseconds in geostationary earth orbit). Furthermore, satellite can provide coverage in geographical areas and not only in population centres, service in case of disaster that causes outage of terrestrial network, and a combination of more access points can allow higher throughput and reliability. Among the various problems that need to be addressed for the design of a fully operable Multi-Connectivity system, a fundamental role is played by resource management algorithms, as controlling the number of resources dedicated to the various connections over the available APs directly impacts the network efficiency and performances. In fact, optimally balancing the number of resources used over the available APs allows the network operator to reduce power consumption, enhance the system resiliency to faults or malicious attacks, and may also provide higher QoS levels to the network users.

1.3 Scope of the thesis

This thesis deals with the problem of communication networks control, to provide an underlying common network able to satisfy several QoS requirements and verticals' needs (e.g., automotive, industrial automation, eHealth, energy, agriculture) with shared network's resources provided by heterogeneous radio technologies (e.g., cellular, Wireless Local Area Network (WLAN) and satellite), adopting control techniques able to make the network's resources management "cognitive", considering the definition provided previously. In particular, the problem of resource control in 5G heterogeneous networks in the multi-connectivity scenario is treated, to design architectures and algorithms to exploit two main aspects:

- the capability of the network to manage the resources in *intelligent, adaptive and optimized* way, considering both real-time network performances, historical users' information, and current users' requirements;

- the exploitation of the *Heterogeneous Multi-Access Radio Technology* to improve the network efficiency and performances, and the users' experience.

Moreover, the thesis is focused on the RAN resources, since the wireless access is the bottleneck of the network, and the Core Network is typically developed with optical wired technology, which implies over-sizing of the core network resources. However, the CN data, such as historical users' information and applications' information, are involved in the decision process. The work is developed considering the 3GPP standard as a reference, to provide architectural and algorithmic solutions that could be directly implemented in the 5G networks either in the cloud or at the edge of the network.

The thesis has been carried out in the framework of the 5G AgiLe and fLexible integration of SaTellite And cellular (5G-ALLSTAR) project [28], which has received funding from the European Union Horzon2020 EU-Korea program and is supported by the Institute for Information & communications Technology Promotion (IITP) grant funded by the Korean government.

The main contributions of the thesis are published in the 5G-ALLSTAR project deliverables and in the following conference proceedings:

- F. Lisi, G. Losquadro, A. Tortorelli, A. Ornatelli, and M. Donsante, "Multi-Connectivity in 5G terrestrial-Satellite Networks: the 5G-ALLSTAR Solution," *2019 25th Ka and Broadband Communications Conference (KA)*, Sorrento, Italy, 2019, ISSN-2573-6124.

- A. Ornatelli, A. Giuseppi, V. Suraci, and A. Tortorelli, "User-aware centralized resource allocation in heterogeneous networks," *2020 28th Mediterranean Conference on Control and Automation (MED)*, Saint-Raphaël, France, 2020, pp. 292-298, DOI: 10.1109/MED48518.2020.9183080.

- A. Ornatelli, A. Tortorelli and A. Giuseppi, "Iterative MPC for Energy Management and Load Balancing in 5G Heterogeneous Networks," *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, New York City, NY, 2020, pp. 0467-0471, doi: 10.1109/UEMCON51285.2020.9298113.

- A. Ornatelli, A. Tortorelli and F. Liberati, "A Distributed Reinforcement Learning approach for Power Control in Wireless Networks," 2021 IEEE World AI IoT Congress (AIIoT), 2021, pp. 0275-0281, doi: 10.1109/AIIoT52608.2021.9454208.

- A. Ornatelli, A. Tortorelli, A. Giuseppi, and F. Delli Priscoli, "Hierarchical RL for Load Balancing and QoS Management in Multi-Access Networks," *2021 29th Mediterranean Conference on Control and Automation (MED)*, Bari, Italy, 2021.

Furthermore the author's research activities led to the publication of the following articles which have not been reported in the present thesis:

- A. Di Giorgio, A. Giuseppi, F. Liberali, A. Ornatelli, A. Rabezzano and L. R. Celsi, "On the optimization of energy storage system placement for protecting power transmission grids against dynamic load altering attacks," *2017 25th Mediterranean Conference on Control and Automation (MED)*, Valletta, 2017, pp. 986-992, doi: 10.1109/MED.2017.7984247.

- G. Tomasicchio, P. Conforto, G. Losquadro, M. La Ferla, and A. Ornatelli, "Small sat mission for maritime surveillance, based on SAR, VDES/AIS and COMINT/ELINT integrated solutions," *2017 23rd Ka and Broadband Communications Conference (KA)*, Trieste, Italy, 2017, ISSN-2573-6124.

- Looking at NB-IoT over LEO satellite systems: design and evaluation of a service-oriented solution, *Submitted at IEEE Internet of Things Journal.*

1.4 Thesis outline

The rest of the thesis is organized as follow:

- Chapter 2 presents a state of the art focused on 5G Multi-Connectivity, exploring network standardized architectures and management approaches, as well as resource allocation;

- Chapter 3 presents the proposed network architecture, designed to be compliant with the standardized 5G network and to easily accommodate Multi-connectivity control algorithms;

- Chapter 4 presents two model-based Multi-Connectivity control strategies. The two problems are formulated as optimization problems, and are solved with a centralized solutions, considering the model of the controlled process known;

- Chapter 5 presents two data-driven Multi-Connectivity control strategies. The two problems are modelled and solved using the Reinforcement Learning method, implementing hierarchical and distributed solutions, placing most of the computing process at edge of the network, were the data are generated and collected.

- Chapter 6 presents the conclusions and the future work.

Chapter 2

State of the Art

This chapter provides the state of the art concerning multi-RAT 5G network from a twofold perspective: from one side it is dedicated to present the state of the art related to multi-RAT management methods and network architectures starting from the 3GPP standard, from the other side it is dedicated to present the state of the art related to algorithms for addressing the network selection and resource allocation problem.

2.1 Multi-RAT Integration and Management

Multi-RAT access network, or heterogeneous access network, is considered to be the key enabling technology to satisfy the 5G requirements, such as high data rate, ultra-low latency, and reliability. To make efficient use of all the available network resources, multi-connectivity has been proposed to simultaneously connect, and orchestrate multiple different radio access technologies. Multi-connectivity opens the possibility of routing, or steering, the network traffic (now divided in QoS flows) over different RATs at the same time. This scenario has the possibility to include in future mobile communication networks new technologies, such as satellite systems, and obtain the most suitable QoS connection characteristics (e.g., in terms of reliability, throughput, latency, etc.) depending on the nature of the data stream considered, where a fundamental role is played by resource management algorithms, as controlling the number of resources dedicated to the various connections over the available APs directly impacts on the network efficiency and performances. In fact, optimally balancing the number of resources used over the available APs allows the network operator to reduce power consumption, enhance the system resiliency and provide higher QoS levels to the network users. In general, multi-connectivity can be defined as *the capability to configure a UE to utilize resources provided by different nodes, which are characterized by different access technologies*. In [29] three multi-RAT integration methods are presented:

- Application Layer Integration: it consists of a higher-layer interface, providing information exchange between UEs and content provider, over multiple RATs. This solution can be easily implemented, but it is application-dependent and may not fully take into account the network state, which leads to sub-optimal exploitation of resources, especially if the network state is observed to vary dynamically;

- Core-Network-Based Integration: this solution is proposed by 3GPP for cellular/WLAN integration based on inter-working between core networks. In this case, the RAT selection is made considering operators' policy, but the overall network selection decision remains in control of the UE. The UE is then able to take its decisions considering operator policies, radio links performances, and user preferences. It is worth remarking that typically the UEs only has local knowledge about the network conditions, resulting in sub-optimal decisions.

- RAN-Based Integration: this solution is proposed by 3GPP in 5G/Long Term Evolution (LTE) dual-connectivity and allows coordination between the RATs using dedicated interfaces. The cooperation level between the different RATs is constrained by the back-haul links. The benefits of this solution are the adaptation of the decisions to dynamic variations in the radio link conditions, consequently minimizing session interruptions or packet drops.

After the multi-RAT integration methods, another decision is about the management of these connectivity, in [29] the main approches are presented:

- User-Centric Approach: with this solution the UE is continuously monitoring the radio links conditions, and, considering thresholds-based performance parameters (e.g. Signal-to-Interference-plus-Noise Ratio (SINR) and Bit Error Rate (BER)), it performs the selection. The UEs can consider other RATs characteristics (e.g. coverage, reliability and battery consumption) to better satisfy the user experience. This approach is limited to the local UE knowledge;

- RAN-Assisted Approach: The RAN-assisted approach employs network assistance from the RAN to the UEs for RAT selection decisions. In this way the UEs have a better knowledge of the network status in order to have better decisions. An example of assistance parameters can be network load, RAT utilization, expected resources allocation;

- RAN-Controlled Approach: the above-mentioned schemes are user-centric by nature, resulting in sub-optimal decisions from the overall system performances point of view. The RAN-controlled approach places the multi-RAT control in the radio networks. In this approach the RAN can assign the UEs to certain RATs. Such a solution can be distributed across RATs or may utilize a central unit that manages radio resources across several cells/RATs. The UEs, in this solution, are configured to report radio measurements on their local radio environment. This solution is adopted by 3GPP for addressing NR-LTE dual-connectivity.

The RAN-Based Integration and RAN-Controlled approaches have been adopted in the Metis-II project [30] and also promoted in the 5G Infrastructure Public Private Partnership (5GPPP) document [21]. The Metis-II project was aimed at designing an access network architecture and proving technical enablers for efficient integration between RATs. The project proposes an architecture as depicted in figure 2.1 (b) composed by several Air Interface Variant (AIV) and a central unit with AIV-agnostic functions, that based on the real-time feedback provided by the AIVs, is capable to steer the QoS Flows dynamically on the different AIVs.

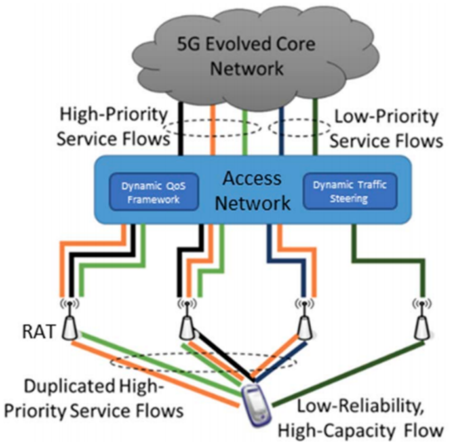
Further considerations can be done about the RAN-Based Integration choice, since the latter can be managed in several ways, in [31] three types of RAN-Based Integration are presented:

- *Hard Handover (HH)* enables users with poor coverage to switch to another RAT which can provide better coverage. The HH requires quite extensive Radio Resource Control (Radio Resource Control (RRC)) and CogNet signalling, as well as cell search and synchronization, which results in relatively long interruption delays. Another drawback is the low reliability since the users can be connected to only one RAT at a time.

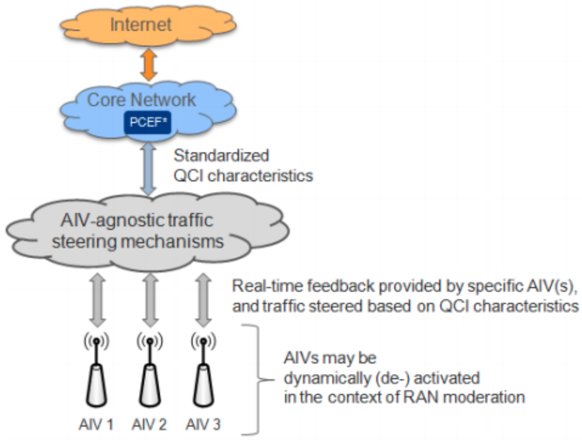
- *Fast Switch (FS)* assumes to have common Control Plane between different technologies. In FS no signaling is required for User Plane switch, and it can respond almost instantaneously when the channel quality variations occur. Another benefit brought by the use of FS consists in the increased reliability since the users can be connected to multiple RATs at each time, due to the use of one control plane connected to all the RATs. Obviously, even FS has some drawbacks, these are due to the presence of greater overhead for the increased control signaling.

- *User Plane Aggregation* assumes to have both UP and CP connected to all the RATs, and that the UP data is aggregated at a central anchor point. The benefits, in this case, are increased throughput, resources pooling, and the support for reliable seamless mobility. Notice that these benefits may be limited, due to the different latency and throughput of the involved RATs.

Typically, the UP Aggregation can be considered as the best choice to increase throughput and reliability, while decreasing signaling and switching time. However, the adoption of UP Aggregation imposes some architectural constraints such as a common entity to accommodate the shared functionalities for both user and control plane. Moreover, the traffic steering among several RATs implies even more difficult decisions process with respect the other two methods.



(a) Dynamic Traffic Steering framework in 5GPPP [21]



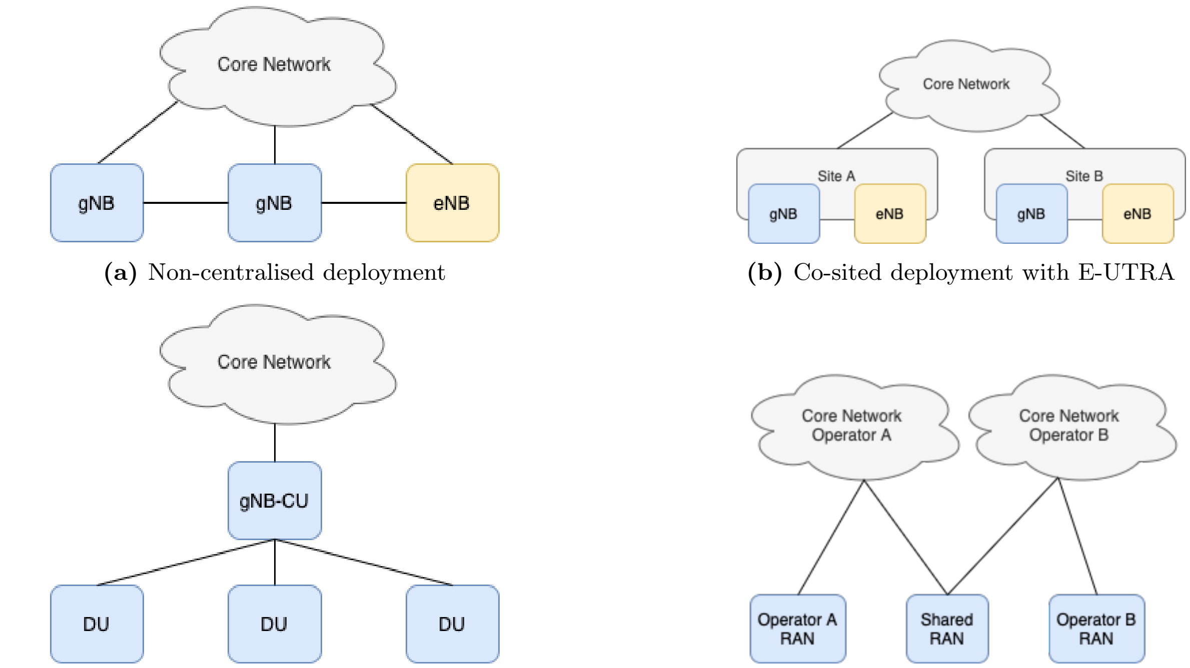
(b) Architecture for traffic steering and RAN moderation in Metis-II [30]

Figure 2.1: Dynamic traffic steering in Multi-Connectivity

2.2 3GPP Multi-Connectivity Architectures

The starting point to define a multi-connectivity solution is the definition of the RAN architecture. The 3GPP standard, in [32], define the RAN architecture in case of heterogeneous deployment, assuming gNBs characterized by any radio technology in the same geographical area, able to communicate each other via Base Station (BS) interface. The heterogeneous deployments are divided in:

- *Non-centralised deployment*, figure 2.2a, the full protocol stack is supported at the gNB, the gNBs can be connected to any transport network (RAN-CN interface) in order to reach the CogNet and assuming the gNB able to connect with other gNB characterized by any radio technology via dedicated interfaces (inter-BS interface).
- *Co-sited deployment with Evolved Universal Mobile Telecommunications System (E-UTRA)*, figure 2.2b, the 5G functionalities (i.e., NR) are co-site with LTE functionalities (i.e., E-UTRA) either as part of the same base station or as multiple base stations at the same site. In this deployment, it is desirable to fully utilise all radio resources assigned to both RATs by means of load balancing or connectivity via multiple RATs.
- *Centralized deployment*, figure 2.2c, NR should support centralization of the upper layers of the NR radio stacks. Different protocol split options between the Central Unit and lower layers of gNB nodes may be possible. The functional split between the Central Unit and lower layers of gNB nodes may depend on the connectivity between CU and DUs. High performance between the CU and lower layers of gNB nodes, e.g. optical networks, can enable advanced CoMP schemes and scheduling optimization, which could be useful in high capacity scenarios, or scenarios where cross cell coordination is beneficial. Low-performance connectivity between the CU and lower layers of gNB nodes can enable the higher protocol layers of the NR radio stacks to be supported in the CU, since the higher protocol layers have lower performance requirements in terms of communication bandwidth, delay, synchronization, and jitter.
- *Shared RAN deployment*, figure 2.2d, NR should support shared RAN deployments, supporting multiple hosted Core Operators. The Shared RAN could cover large geographical areas, as in the case of national or regional network sharing. The Shared RAN coverage could also be heterogeneous, i.e. limited to few or many smaller areas, for example in the case of Shared in-building RANs. A shared RAN should be able to efficiently interoperate with a non-shared RAN. Each Core Operator may have their own non-shared RAN serving areas adjacent to the Shared RAN. Mobility between the non-shared RAN and the Shared RAN shall be supported in a way at least as good as for LTE.



(a) Non-centralised deployment

(b) Co-sited deployment with E-UTRA

(c) Centralized deployment

(d) Shared RAN deployment

Figure 2.2: 3GPP RAN deployments

After the deployment, the other key point in the multi-connectivity architecture is the definition of the functional split between the entities. In particular the split between the CU and DUs is considered in the 3GPP with several options, reported in figure 2.3

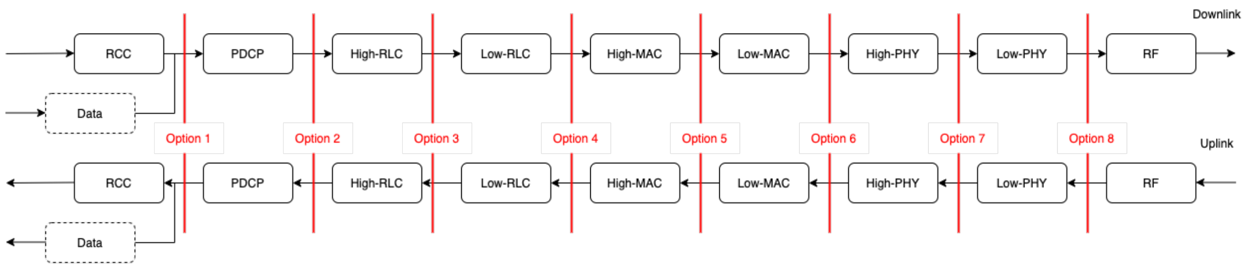


Figure 2.3: Function Split between central and distributed unit

The choice of how to split NR functions in the architecture depends on several factors related to the RAN deployment, the involved radio technologies, the transport network between the CU and DUs and the service requirements. Some examples of such factors are:

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

Some of the possible choices are reported with an emphasis on advantages and disadvantages of each choice, considering the required performances of transport network (i.e. the fronthaul) between CU and DUs (e.g., synchronization, latency, throughput) and the resources management capability (e.g., load balancing, joint transmission, traffic aggregation):

- Option 1: In this split option, RRC is in the central unit. Packet Data Convergence Protocol (PDCP), Radio Link Control (RLC), Madium Access Control (MAC), physical layer and Radio Frequency (RF) are in the distributed unit, thus the entire user plane is in the distributed unit. This option allows a separate U-plane while having a centralised RRC/Radio Resource Management (RRM). It may in some circumstances provide benefits in handling some edge computing or low latency use cases where the user data needs to be located close to the transmission point. Because of the separation of RRC and PDCP, securing the interface in practical deployments may or may not affect the performance of this option. User plane aggregation can be tricky.

- Option 2: : In this split option, RRC, PDCP are in the central unit. RLC, MAC, physical layer and RF are in the distributed unit. This option will allow traffic aggregation from DUs transmission points to be centralized. Additionally, it can facilitate the management of traffic load between DUs transmission points. Fundamentals for achieving a PDCP-RLC split have already been standardized for LTE Dual Connectivity. Therefore this split option should be the most straightforward option to standardize and the incremental effort required to standardize it should be relatively small.

- Option 6: The MAC and upper layers are in the central unit (CU). Physical Layer (PHY) layer and RF are in the DU. The interface between the CU and DUs carries data, configuration, and scheduling-related information (e.g. MCS, Layer Mapping, Beamforming, Antenna Configuration, resource block allocation, etc.) and measurements. This option will allow traffic aggregation from DUs transmission points to be centralized. Additionally, it can facilitate the management of traffic load between DUs transmission points. This option is expected to reduce the fronthaul requirements in terms of throughput to the base-band bit-rates as the payload for Option 6 is transport block bits. Joint Transmission, centralized scheduling and resource pooling for layers including and above MAC are possible since the MAC is in CU. This split may require subframe-level timing interactions between MAC layer in CU and PHY layers in DUs. Round trip fronthaul delay may affect Hybrid Automatic Repeat Request (HARQ) timing and scheduling.

- Option 7: Multiple realizations of this option are possible, including asymmetrical options which allow to obtain benefits of different sub-options for Up Link (UL) and Down Link (DL) independently (e.g. Option 7-1 is used in the UL and Option 7-2 is used in the DL). A compression technique may be able to reduce the required transport bandwidth between the DU and CU. This option will allow traffic aggregation from DUs transmission points to be centralized. Additionally, it can facilitate the management of traffic load between DUs transmission points. These options are expected to reduce the fronthaul requirements in terms of throughput. Centralized scheduling, Joint processing (both transmit and receive) are possible with these options as MAC is in CU. This split may require subframe-level timing interactions between part of PHY layer in CU and part of PHY layer in DUs.

However, the best solution is not a fixed split, but the use of software-defined technology, to provide the flexibility to move RAN functions between the central unit and distributed unit adaptively for different use cases, depending on the factors above.

2.2.1 3GPP Dual-Connectivity

In the 3GPP standard, a particular case of Multi-Connectivity is already standardized considering the coexistence between 5G and LTE. This standardized solution is called Multi-RAT Dual Connectivity [33], in which the multiple Tx/Rx UEs may be configured to use resources provided by two nodes: the first node provides E-UTRA access (i.e., LTE), while the second node provides NR access (i.e., 5G). In 3GPP Dual-Connectivity, as reported in Figure 2.4, the two RATs involved in the connection are identified as Master and Secondary Nodes (Main Node (MN) and Secondary Node (SN)), the functional split is the option-2 discussed above and the architecture can be referred to the non-centralised deployment. In the scheme, three bearer types across the Air Interface Physical Layer (Uu) interface are defined: i) the MCG bearers, only the MCG radio resources are involved; ii) the SCG bearers, only SCG radio resources are involved; iii) the Split bearers, both MCG and SCG radio resources are involved. Each Service Data Adaptation Protocol (SDAP) entity is placed in one RAT. The MN decides which QoS flow should be assigned to each SDAP entity. The MN or SN node, that hosts the SDAP entity, for a given QoS flow decides how to map it to DRBs.

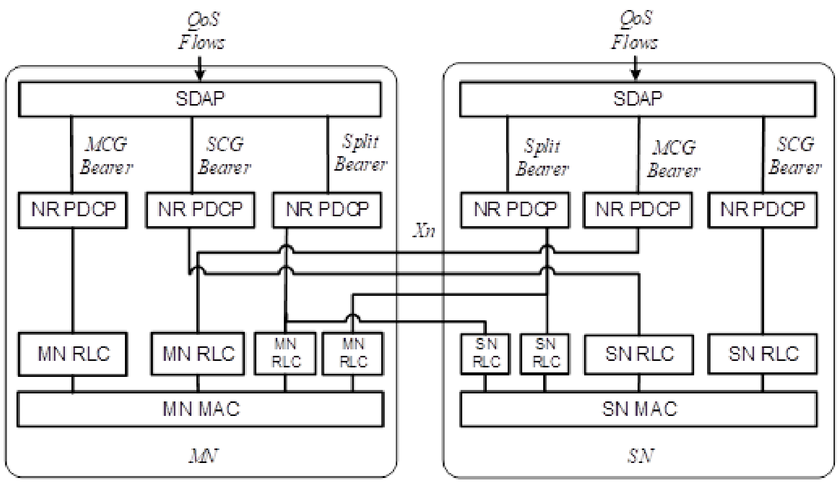


Figure 2.4: Network side protocol termination options for MCG, SCG and split bearers in MR-DC with 5GC

2.2.2 3GPP Non-Terrestrial Networks for Multi-Connectivity

The key aspect of the Multi-Connectivity is the diversity of the involved radio technologies. To exploit this advantage, in the 3GPP, there was the introduction of the *Non-Terrestrial Network (NTN)* as a new concurrent access network. In [34] the advantages of the NTNs integration in terms of services are divided in:

- *Service Continuity* providing coverage in geographical areas and not only in population centers;
- *Service Ubiquity* in case of a disaster that causes outage of terrestrial network or economic rationales such as not enough revenues, some areas can be unserved or underserved by terrestrial network and the satellite access can provide coverage for such areas;

- *Service Scalability* in comparison with a terrestrial network, a satellite network has a large coverage, satellites are therefore efficient in multicasting or broadcasting a similar content over a large area.

To provide these advantages satellites can be used as a bridge to connect terrestrial-RAN to the 5G-CN, or to extend RAN with remote terrestrial access, but, the most interesting case is the use of the satellite technologies in the Access Network, providing hybrid Satellite-Terrestrial RAN. The real challenge, in this case, is the integration between the two different technologies (terrestrial and satellite networks). The NTNs differ from typical cellular networks in terms of network functional architecture, as well as deployment scenarios and actual coverage which may span across several countries. An appropriated mapping of NTNs physical entities onto logical NG-RAN architecture can minimize integration issues. In [35] the logical RAN architecture of three main architectures have been proposed considering bent pipe or regenerative satellite payload, to connect UE to CN via NTN-RAN:

- Figure 2.5 shows the bent pipe payload, where the network is composed of a gNB on the ground that contains User and Control Plane functionalities and the satellite is used as a relay, with only amplification, frequency conversion, and filtering functionalities (PHY-low layer). This architecture can be used when the satellite has a stringent constraint in terms of power consumption, weight, etc. On the other hand, this architecture has disadvantages in terms of Control Plane traffic (i.e., a big amount of control signals is sent along the paths) and real-time functionalities (i.e., the control information used by real-time functions has Round Trip Time (RTT) delay);

- Figure 2.6 shows the regenerative payload. In this case the functionalities are split between a gNB-CU on ground and gNB-DU on the satellite. The real-time function and the lower-layer protocol stack (e.g., PHY, MAC, RLC) are placed on-board, to avoid the problem on control signalling.

- Figure 2.7 shows the regenerative payload, all the network functionalities are placed on-board. In this case, the integration with terrestrial networks can be actually challenging due to the difficulties of communication and synchronization between terrestrial and non-terrestrial network functionalities.

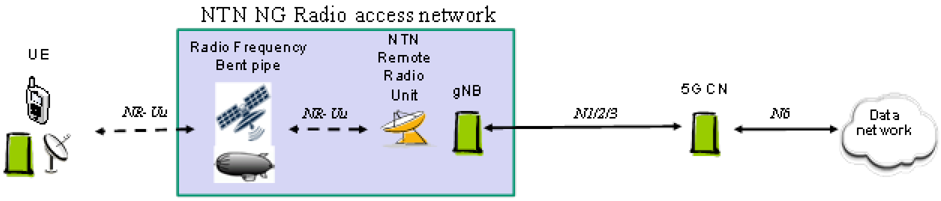


Figure 2.5: NG RAN architecture in Non-Terrestrial network with bent pipe payload

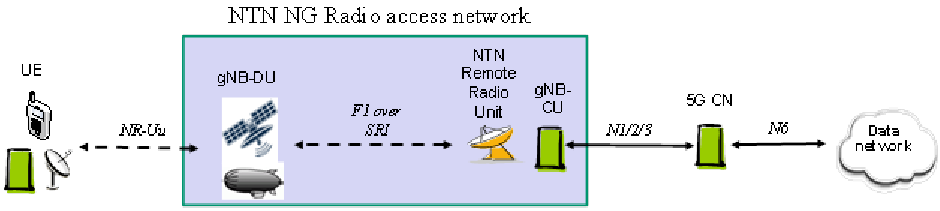


Figure 2.6: NG RAN architecture in Non-Terrestrial network with gNB-DU processed payload (SRI refers to Satellite Radio Interface)

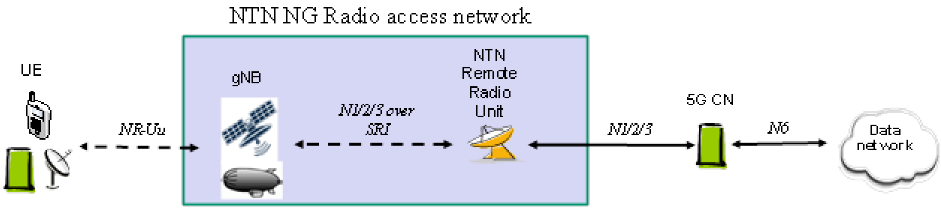


Figure 2.7: NG RAN architecture in Non-Terrestrial network with gNB processed payload (SRI refers to Satellite Radio Interface)

2.3 Resources Management Algorithms for Multi-Connectivity

The most recent discussions about Multi-Connectivity concern the selection of the appropriate technologies considering services and user requirements and the management of the selected resources, optimally steering the traffic between the access networks. In [36, 37], the multi-connectivity control problem was investigated, where the main multi-connectivity problem is identified in the *traffic steering* problem. Traffic steering is defined as the function of distributing the traffic load optimally across different network entities and spectrum bands. However, this traffic distribution shall be performed considering operator and user preferences, the characteristics of the different RATs, such as coverage, latency, and capacity, and the different traffic characteristics (i.e., the QoS requirements), introducing the need of optimally associate users and access networks. Therefore, the goal of the traffic steering is to efficiently deliver capacity to the UEs satisfying the QoS/QoE requirements, considering the network management operations, such as energy-saving, load balancing, interference minimization, and congestion control. These functions can be performed in a centralized or distributed fashion, and in both cases, information about the entire network status is needed. However, centralized coordination can achieve optimal performances, though it introduces problems in the real-time control of the RATs. On the other hand, a distributed implementation can only perform traffic steering based on local information, typically achieving sub-optimal solutions.

In this framework, the Multi-Connectivity Control can be divided into the *Network Selection* and *Resources Management (or Network Control)* processes. This separation is applicable even if the multi-connectivity control is faced by a single entity executing a single algorithm. In this section, the solutions proposed in the literature for the network selection and the resource management problems are reported.

2.3.1 Network Selection

The main decision for traffic steering in multi-connectivity scenarios is represented by the RAT selection issue [38–40], also known as network selection problem in heterogeneous networks. The problem consists in the selection of the most appropriate access network with characteristics able to satisfy the users' requirements. These selections can be performed by considering different network features as for instance: the mobility of the network nodes, the QoS attributes, the network reliability, and the energy constraints. The algorithms capable to perform the RAT selection are evaluated by considering their characteristics such as computational complexity, implementation complexity, distributed or centralized deployment with either open or closed-loop solutions, and dynamic or static behavior, considering model-based or data-driven techniques. Many methods are proposed in the literature with several advantages or disadvantages in terms of performance as well as problem tractability. A comprehensive comparison between available solutions in the literature is provided by [38], where the authors show that tractable methods, in terms of implementation and execution, are weak from the performance point of view. Indeed, methods such as utility theory, Multi Attribute Decision Making (MADM) and fuzzy logic are typically static. On the other hand, methods able to provide a dynamic solution, such as optimization, game theory, and Markov chain, have some disadvantages considering implementation and execution effort.

The easier network selection strategy, considering the implementation complexity, is the Utility theory. It is based on the concept of "utility" and its associated "utility functions". In [38, 41, 42], the utility functions were characterized by different attributes, chosen to capture the networks' and users' characteristics, depending on the use case requirements. Each of the chosen attributes of interest is then associated with a function, whose type is selected so that it models correctly the behavior of its associated characteristic, common choices for these functions are linear, sigmoidal, exponential, logarithmic, and polynomial. The overall utility function is then the summation of all the attribute functions, and the goal of the network control is to maximize (or minimize) these sum of functions.

In general, this approach can be hard in the design phase, since it requires the selection of the attributes of interest and the associated utility functions that better model the considered scenario. This process is analyzed in [42], where three utility functions are designed to analyze different user attitudes to risk for economic benefits and delay preferences. The choice of attributes and their associated functions are taken to represent three different users risk profiles:

1. U1, risk-neutral users, that is the standard, in-between users;
2. U2, risk-seeking users, that pay more but experience less delay;
3. U3, risk-averse users, that pay less but experience more delay.

The choice of the different utility functions modifies the connection experience for the users. In particular, for smaller file sizes, the price difference between U3 and U2 seems negligible, making U2 the best risk profile to account for the smaller percentage of transfers exceeding the time deadline. For large files, the price difference may deter low budget users from this choice.

Another method, typically easy to be implemented, is the MADM method [38, 39, 43, 44]. It is an approach based on the selection of different alternatives, each characterized by different attributes. The problem can be expressed in matrix form, where columns indicate attributes and rows indicate alternatives. Furthermore, the attributes shall be weighted, to represent the importance of each attribute in the choice, and normalization is needed to make the different attributes' measurement units homogeneous. In [39], a selection network scheme is presented, based on two different MADM approaches, to select the best network in an integrated cellular/wireless LAN system, to provide the user with the best available QoS during the connection. The proposed scheme comprises two parts: 1. an Analytic Hierarchy Process (AHP) to decide the relative weights of evaluative criteria set according to user preferences and service applications; 2. a Grey Relational Analysis (GRA) to rank the network alternatives because it is faster and simpler to implement than AHP. The simulation results reveal that the proposed network selection scheme can efficiently decide the trade-off among user preference, service application, and network condition. The main disadvantage of this method is that it is static and open-loop.

Fuzzy logic [38, 40, 45] is a suggestive solution to the network selection problem. It is based on the idea that an object or state cannot be assigned univocally to a class, but instead, it is associated with a number between 0 and 1 for each of the available classes, representing the confidence of the classification. This method can be used to model the complex heterogeneous network system, and potentially can be coupled with other methods, e.g. MADM. The workflow of this method begins with the fuzzifier, which maps the selected attributes into a fuzzy set. The second step is made by the fuzzy inference engine, that with the help of the fuzzy rule base, produces the maps that connect the input fuzzy set to the output fuzzy set. The output fuzzy set is then processed by the defuzzifier that takes the control decisions. In [45] the authors adopted an algorithm based on a Fuzzy Logic Controller (FLC) to evaluate fitness ranking of candidate networks. The selection is made in three phases: 1) preselection, to eliminate unsuitable networks; 2) discovery; 3) decision making. The discovery phase is based on FLC, where the fuzzifier maps the selected variables (network data rate, SINR and application requirement data rate) into the fuzzy set (i.e. the fuzzifier). Then, the inference engine uses these mapping functions as input to the predefined logic rule base (i.e. the inference engine). Finally, the defuzzifier obtains the overall ranking through defuzzification with a weighted average method. This method can be implemented considering a closed-loop solution, considering the dynamical behavior of the system. However, the design and development of such a method can be challenging.

The Game Theory proposes one of the most challenging but complete solutions for the network selection problem. In Game Theory approaches, [38, 40, 46, 47], the problem is modeled with a set of players/agents (i.e. the decision-makers), and a set of possible actions, defined as the strategy set of actions. The basis of this approach is the concept that each action, performed by a player, affects the actions of other players. In general, each player tries to maximize its own utility or payoff, in an adversarial, non-cooperative, framework. However, cooperative games are also considered.
Several solutions, or equilibrium point, can be defined for the different games, however, the typically considered optimal solution is the so-called Nash Equilibrium (NE), which is defined as the combination of strategies containing the "best" strategy for every player, in the sense that no player can improve its performances unilaterally changing its action. In the literature several network selection games can be found:

- games between users, where the users can be cooperative or non-cooperative, and each user try to exploit as much as possible the network resources, considering its utility in case of the non-cooperative game, or the utility of all the users in case of the cooperative game;

- games between access networks, in this case, the different access networks shall decide the action to be taken. This can be either cooperative or non-cooperative. Non-cooperative when each network is managed by a different operator and each operator tries to optimize its own network. Cooperative in case the actions of each network is performed considering the performances of all the networks;

- game between users and networks, where the set of users and the set of networks are modeled by two different players. In this case, the users' strategies are the selection of their favorable network, while the networks' strategies are the selection of their favorable users. In this last case, the NE can be reached only if the networks and the users correspondingly select each other.

In [47] the network selection problem is modeled as a non-cooperative game between access networks. These access networks compete to decide the sub-set of service to admit from a set of services. This selection is made maximizing a payoff that considers the users' satisfaction. Indeed, each access network admits the services that indicate the higher preferences towards the specific access network.

Combinatorial optimization [38, 48] approaches search for an optimum object in a finite collection of objects. These approaches are characterized by a cost function and a set of constraints and can be a good choice in the case of convex formulation, which can be solved easily and, potentially, in a distributed way. It [48], optimal traffic splitting and aggregation solution are proposed. The problem is modelled as a maximization problem, whose cost function depends on the users' throughput and whose constraints imply that the user throughput is bounded by the sum of individual rate across all RATs. The peak rate of each RAT is considered known through the channel state information feedback. The algorithm was designed to be implemented in a distributed way, and the simulations showed an increase in throughput. However, this solution is typically computationally hard.

Markov chains [38, 49, 50] are one of the most common tools for decision making and are used for addressing several different problems. This approach is based on several models (e.g., Markov Decision Process (MDP), Partially Observable Markov Decision Process (POMDP), Hidden Markov Model (HMM)), in order to better represent the environment in which the decisions are taken. Markov Decision Process (MDP) is the model that better fits most of the environment since it is used to describe and model a stochastic dynamical system. MDP is defined by a state space, a set of available actions for each state, a state transition probability function, and a reward function. When the system is in a given state, the decision-maker selects an action, which causes the system to evolve to the next state according to the transition probability function. The decision-maker then collects the reward associated with the new state, which is an index of its performance in the new state. In [49], a Markov chain approach for network selection is used. In the formulation, each state is defined by the users' satisfaction probability for each radio technology. In each state, the possible actions are defined by the choice of radio technology to use. The proposed learning algorithm follows the Reinforcement Learning paradigm, with the reward function defined as the ratio between the sum of the weighted users' satisfaction for each technology over the total number of users. The algorithm is compared with baseline load balancing solutions, in a scenario with two different user profiles corresponding to low and high-bit-rate services. The results demonstrate the ability of the proposed algorithm to adapt to current system conditions while achieving high user satisfaction, indeed, the algorithm showed better behavior and lower signaling requirements compared with two load balancing algorithms. However, this solution has several disadvantages, in particular regarding the dimensionality of the state and action spaces, with resulting difficulty in the learning process and the decisions exploration.

2.3.2 Resources Management

A suitable modeling framework for network resource management is represented by the so-called dynamical communication networks [6, 7], which are networks whose evolution is described by differential (or difference) equations derived from the general mass conservation law. The fundamental reasoning behind such a modeling framework is that any flow, be it a physical quantity or a stream of information, is conserved over the network links and nodes, and consequently its distribution follows simple physical-inspired laws. However, the modeled system shall be controlled following some Key Performance Indicator (KPI)s, a typical mathematical tool used for the KPIs modeling and evaluation is the utility function [41, 42], formally defined as a monotonically non-decreasing function that captures a certain performance, associated with the network state. Common use cases are, for example, the modeling of users' queues that shall be minimized to avoid congestion, or of user satisfaction as a function of the network resources assigned concerning some specific RATs. The combination of dynamical networks with utility functions allows the investigation of optimization algorithms for the dynamical control of traffic flows, as the utility functions may be utilized to capture several different objectives, while the differential representation of the network enables a realistic analysis of its time evolution, in terms of its congestion state and overall connection quality. The resource management in multi-connectivity can then be seen as a multi-objective network control problem, characterized by several utility functions capturing its performances and costs, as well as physical and logical constraints, such as link capacities and user preferences. Furthermore, to have a representative model of all the system behaviors and limitations, some constraints shall be considered in the model, since when dealing with real networks that run on physical equipment, it is not reasonable to assume the number of network resources to be unlimited. Capacitated Networks have been extensively studied in the literature [8, 51, 52], and typical solution for controllers that utilise utility functions is to associate extremely negative values (i.e., high costs) to states that violate the physical/logical limitations of the controlled network. On the other hand, several control methodologies directly take into account constraints in their optimization, such as MPC [53, 54] and Optimal control for Differential Algebraic Equations (DAE) [55, 56].

Other than the methodologies introduced so-far, which are related to traditional topics of Control Theory and Networked Systems, an interesting frontier that will be studied is one of the control solutions derived from the integration of Machine Learning (ML) and Game Theory solutions. In this regard, several works [57, 58] studied the joint effects of ML and Control Theory to enhance the system performance, as ML solutions can capture also aspects of the system that are hard to model, thanks to their ability to adapt or learn, from the interaction with the controlled system. Several other works, on the other hand, propose controllers fully based on methodologies as Reinforcement Learning [49, 59–63]. Furthermore, Game-Theory is proposed in several work as a network control tool for traffic flow control [12, 14, 64–66], network formation [67, 68], network selection [69] and power control [18]. Furthermore, under adequate hypotheses, game-theoretic problems can be tRANslated into an equivalent differential (or difference) equation-based formulation, in the field of deterministic evolutionary dynamics [70, 71] or standard differential game [72, 73], allowing once again the control solution to be based on strong, mathematically based, properties. However, the application of Game Theory in engineering applications, such as network control, can be very challenging because existing game-theoretic models are not tailored to cope with engineering-specific issues such as modeling time-varying wireless channels, developing performance functions (i.e., utility functions) that depend on restrictive communication metrics (e.g., Shannon' limit, queues level, channel delay, SINR), and respecting to certain standards (e.g., 3GPP), since it was developed in economical and social sciences.

Chapter 3

Proposed Architecture

In this chapter, the network architecture for multi-connectivity control purposes is detailed, as designed in this thesis in the framework of the 5G-ALLSTAR project, previously presented with its different level of details in [74–76]. In the following the aspects already discussed in Section 2.2 are used and recalled.

Several architectures, defined by different choices on logical entities and interfaces between them, can be defined for the multi-connectivity control purpose, considering heterogeneous technologies with the involvement of satellites. These architectures shall provide the capability to take decisions and act these decisions, regarding traffic steering, splitting, and switching between the different technologies, in an efficient and fast way. To satisfy this requirement the candidate architectures shall have particular characteristics, that can be summarized in coordination and cooperation between the control planes and a direct and fast link between the user planes of the different technologies. Three main architectural solutions, based on the 3GPP deployments, with these characteristics, can be proposed:

- Fully centralized solution: with common control and user plane for the different technologies, e.g. a central gNB with control and user plane functionalities that connects with the user equipment via different technologies such as transparent payload satellites and terrestrial radio technologies;

- Fully distributed solution: each technology has its own control and user plane, considering several gNB each for technology, e.g. regenerative payload satellites and terrestrial base station, with control and user plane data exchange between the gNBs;

- Partially distributed solution: there is a central entity with common functionalities such as common control plane functionalities and higher user plane protocol layers, considering a gNB-CU with one common control plane and several user planes, connected to several gNB-DU, e.g. regenerative payload satellites and terrestrial radio unit with limited processing capabilities both connected to a common central unit.

These different solutions can be adopted based on the performances that each architecture provides. Preliminary considerations can be done considering the real-time and the data exchange needs of the functionalities to provide multi-connectivity, in particular, the fully centralized solution can be the best from the point of view of implementation and optimality in the control decision processes, but it can introduce several problems in the network for the high amount of information exchange between the central unit and the radio unit, furthermore, the real-time functionalities can suffer the distance with the radio units and the needs of high computing power can be a problem. The fully distributed solution is the best from the point of view of the real-time needs of the functionalities, but the distribution of these functionalities can introduce problems in the computation of the control decisions because of the loss of knowledge about all the other distributed units. Other problem can be on the control and user data exchange between the distributed units, that can congestion the access network and the coordination and synchronization between the different processing units can be hard. The partially distributed solution mix the advantages of both the previous, the data exchange between the entities of the access network decrease in this solution, furthermore, the real-time functionalities can be implemented in the distributed units, but the common functionalities that need the information about all the network will be developed on the central unit to collect all the information about the radio units in the access network, in the other hand, this solution can be more complex, e.g. because of synchronization problems, integration of common central functionalities and distributed technology-dependent functionalities, etc.

3.1 Network Architecture for Multi-Connectivity Control

Several works, out of the 3GPP, have faced with the problem of Multi-Connectivity from an architectural perspective. The main topic of discussion is the functional split among RAN components [77–79], considering integration of user and control plane aspects [31, 80–83]. Starting from the 3GPP architectures, the gNB can be composed of a Central Unit (CU), also called either Central-RAN or Cloud-RAN [84, 85], and of several Distributed Units (DUs). This functional split between the Central and Distributed unit has the purpose of placing the cooperative, and technology independent, RAN functionalities in a central node, so that they can benefit from the advantages of centralization (i.e., centralized decisions, high computation power available), allowing flexible RATs selection/switch, such as "fast switch". The selection of the functional placement, i.e. the decision about which function should be placed in either the central or the distributed units, is a crucial point that defines the whole traffic flow control system. From the control plane perspective, the ideal scenario would be to put the whole set of functionalities that are technology independent as well as Non-Real-Time and low bit rate, (e.g. traffic steering, spectrum sharing, etc...) in the central unit to have a complete view of the system, allowing optimal decision making. In this case, the distributed units have technology-dependent, Real-Time, and high bit rate functionalities to meet their requirements.

The figure 3.1 reports a possible reference physical architecture for Multi-Connectivity (MC) introduced in [76], the depicted scenario is characterized by two Protocol Data Unit (PDU) sessions, namely PDU1 and PDU2, used to show the data flow defined in the architecture:

- PDU1 (the green line in Figure 3.1) is duplicated over two different Access Points belonging to different RATs so that the connection resiliency is increased;
- PDU2 is composed of two different QoS flows (red and yellow lines in Figure 3.1) that, based on the analysis of the network state, can be routed independently from one another, to satisfy their QoS requirements accordingly.

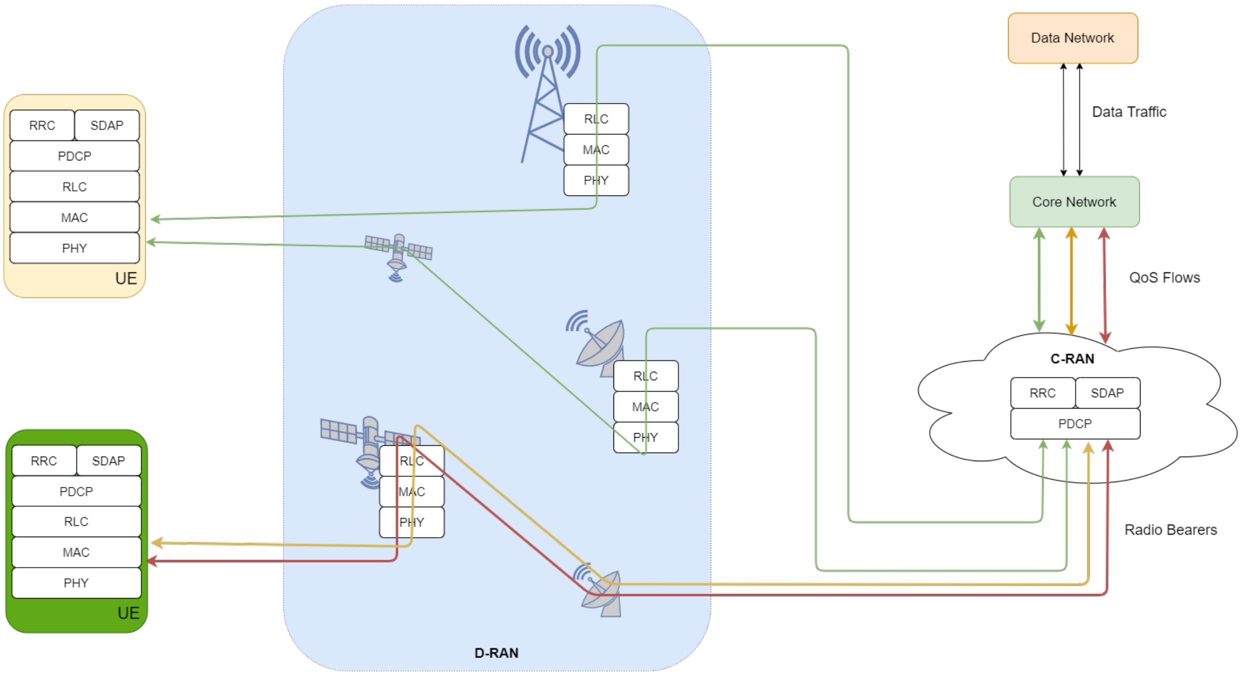


Figure 3.1: Multi-Connectivity reference scenario physical architecture

The idea is to provide an End-to-End (E2E) framework, that can be easily integrated with the standard 5G functionalities (e.g., SMF, AMF, RRM, etc.), exploiting this integration to manage the connections from the higher protocol levels to the lower protocol level, using CN and RAN information, such as, user and PDU session data from the CN and real-time measurement from the RAN. The objective is the management of all this information to allow fast and dynamic control of the connection flows, ensuring the best mapping between the high-level connection flows and the low-level radio resources.

The integration between the standard 3GPP network functionalities and the 5G-ALLSTAR multi-connectivity system is one of the fundamental aspects to be considered, in this respect, the typical QoS flows handling will be guaranteed and strengthened, considering as a reference the QoS management standardized in 3GPP for 5G networks, presented in Section 1.2, where the fundamental steps are:

1. mapping of application data (PDU Session data) in QoS Flows, performed in the CN by the UPF, and respecting the SMF configuration;

2. mapping of QoS Flows in the Access Resources, performed in the RAN.

However, the standard 3GPP QoS management approach does not allow the exploitation of users and applications information in the RAN during the resource assignment, making fail the possibility to assign lower-level resources in a Quality of Experience (QoE) fashion. The 5G-ALLSTAR architecture improves this aspect by making use of user and application information in the RAN. In this perspective, the providers' connections and users' information typically stored and managed in the CN and used for the PDU Session to QoS Flow mapping, accordingly to the 3GPP standard, will be processed and enriched with a specific QoE Management System, to be used to strongly influence the resource assignment in the RAN. On the other hand, this resources assignment shall be constrained by the QoS of the ongoing connections, the actual status of the network as well as by users' characteristics. This approach implies the needs of the information that is managed at RRM level in 3GPP standard, such as, measurement provision and configuration functionalities, at the same level of users and connections information. Further improvements in the 5G-ALLSTAR architecture is the strong cooperation and coordination between the several RATs involved in the connections, where a key feature is the load balancing between their resources and the user aware RATs selection, to guaranteed an optimal resource management, improving power consumption, network reliability, and QoS requirements satisfaction. The considerations above require that, based on the network status, the controllers try to perform the best associations between QoS Flows and RATs from the QoE point of view, but it will be constrained by the RATs congestion and the link performances.

Summarizing, as detailed in [74, 76], the chosen architecture uses a RAN-Based Integration, to allow the RATs cooperation and coordination, fast reaction to the radio link conditions variation and user plane flows anchoring between the RATs to allow Fast Switch between them. Furthermore, a RAN-Controlled approach is used, to avoid sub-optimal solutions, typical in other approaches such as the User-Centric approach, and encouraging the dynamic adaptation of the resources allocation and configuration based on the actual networks' condition measurements.

The considerations above, drive the architecture design considering a common CU that contains the Control Plane cooperation and coordination functionalities together with the User Plane integration and switching functionalities since the adoption of UP Aggregation and/or Fast Switch imposes the architectural constraints to have a common entity to accommodate the shared functionalities. Solutions with a common central entity are already considered in 3GPP standard, in [86] as Centralized Deployment, with the need of a decision on the functional split between CU and DUs, previously discussed in Section 2.2. In the proposed architecture, the decision regarding the functional split is driven mainly by power consumption and latency constraints: for the Control Plane the technology-independent and asynchronous (Non-Real-Time) functionalities (e.g., Measurements Configuration, Radio Bearers Control, QoS Flows to Radio Bearers mapping, Connection Mobility Control) are placed in the CU and the technology-dependent and synchronous functionalities (Radio Admission Control, Beamforming Configuration) in the DU; for the UP is adopted the PDCP split, that allows the connection switching and aggregation at PDCP level, placed in the CU, facilitating the management of traffic load between RATs, without low latency and high throughput requirements that are present in case of Intr-PHY or PHY-MAC split. This split, identified as Option 2 in 3GPP, is the more promising for standardization in 5G, since it is already implemented in LTE Dual Connectivity and the incremental effort for the needed functionalities can be minimal.

The considerations above apply for the resource allocation, and for the QoS flows allocation among the RATs (i.e., Flow Control) that shall be performed in Dynamic Real-Time and Measurement-Based fashion. On the other hand, the personalization system introduced does not require Real-Time capabilities and needs very high storage and processing power capabilities. These characteristics move the personalization-related functionalities placement in the CN. The overall personalization system is composed of a repository of the historical connections data for each user, and a processing block that, based on the stored information, can synthesize for each user, or a homogeneous group of users (i.e., a cluster), a set of user characteristics in terms of user's Connection Preferences. These Connection Preferences specify the user's needs not in terms of additional QoS constraints, but in terms of personal user's preferences, expressed over a set of non-standardized parameters such as battery consumption, connection cost, and mobility. The stored information is updated based on the user's feedback at the end of each connection and is used for Traffic Flow Control as described in [76]. In 5G-ALLSTAR, the Traffic Flow Controller performs its choices based on the QoS constraints and network status, trying to satisfy in the best possible way the Connection Preferences of each user, in a multi-objective fashion. The final building block needed to allow a multi-connectivity architecture with a personalization system is a measurement of the actual users' experience (QoE). This measurement links QoS and user experience, considering additional attributes that are specific to the user (or a cluster). The QoE level can be used by the Flow Control as an additional input for its APs selection and traffic steering decisions. The resulting functional architecture is depicted in figure 3.2, which represents a functional overview of the modules developed in this thesis and in the 5G-ALLSTAR project and their interaction, as detailed in [74–76]:

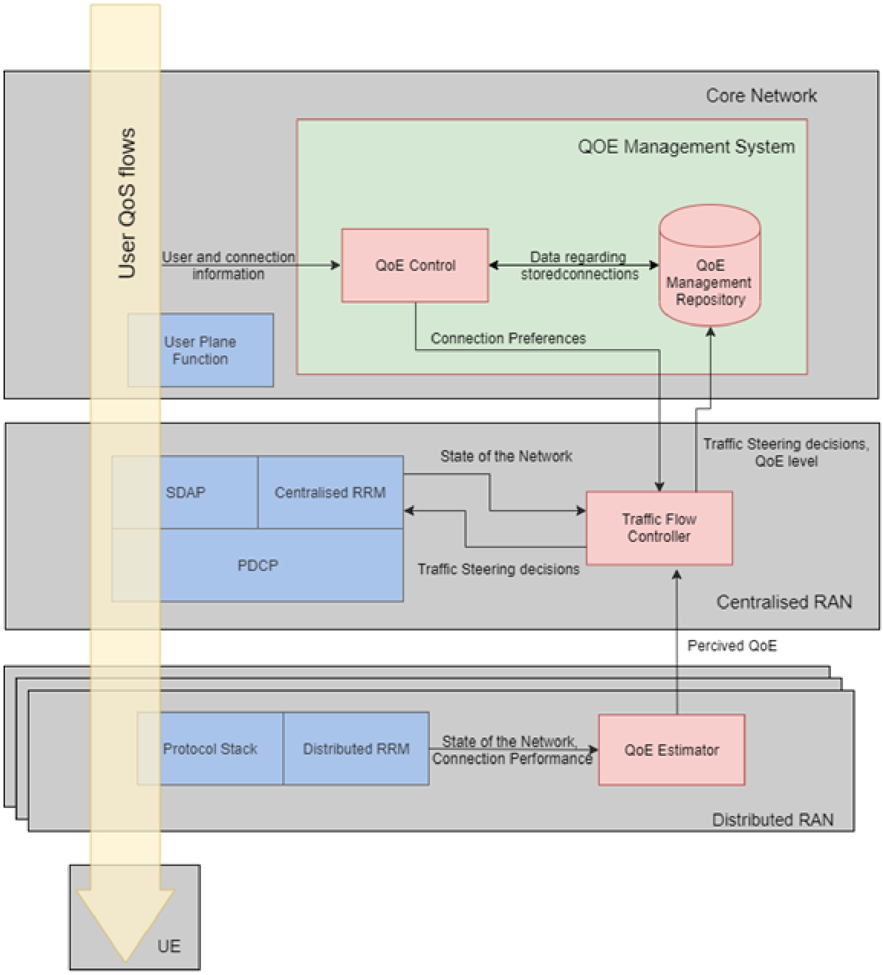


Figure 3.2: Proposed 5G Functional Architecture for Multi-Connectivity

Although the Option 2 (or PDCP split) is the reference split option for the 5G-ALLSTAR architecture, it is worth mentioning that the architecture in figure 3.2 does not explicitly detail how the protocol stack was split to enable multi-connectivity. The reason is the non-applicability of this solution for some scenarios, (e.g., multi-connectivity between some non-terrestrial and terrestrial networks) that for implementation reasons enable only different splits, for example when a different type of satellites are involved in the RAN as discussed in Section 2.2 and depicted in figure 3.1. However, three split options are considered to cover all the possible needs depending on the RATs characteristics, these options are reported in figure 3.3, the ideal architecture shall be flexible to dynamic change and integration between these three different splits at the same time, using the software-defined solution as already discussed in [20, 22, 86, 87].

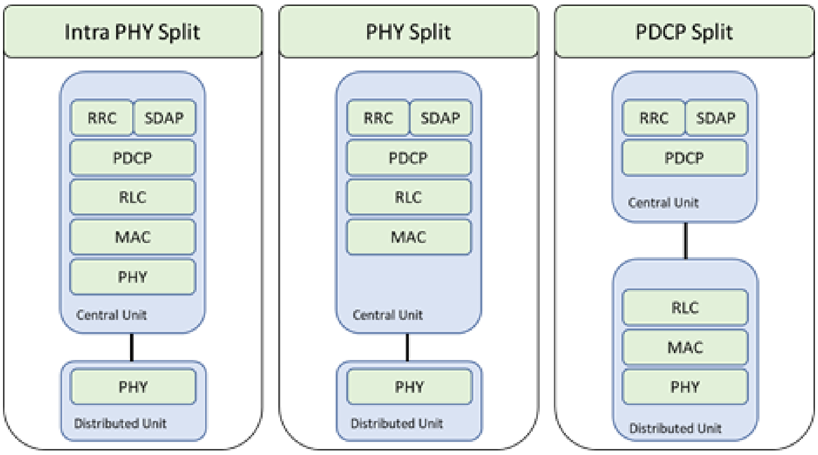


Figure 3.3: Multi-connectivity protocol split - considered options

Indeed, a particular wariness shall be considered where a satellite is involved as a RAT integrated with terrestrial RATs, as discussed in [35] and already reported in Section 2.2, several considerations can be done considering the satellite involvement: the PDCP split 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 Low Earth Orbit (LEO) satellites. Regarding the inclusion of Geostationary Earth Orbit (GEO) satellites in the multi-connectivity framework, it is highlighted in figure 3.1 how they act as a transparent means of communication, as their DU protocol stack is deployed on dedicated ground equipment considering the high latency between gNB or UE and satellite. In general, satellites cover extremely wide areas, usually beyond the reasonable scope of a single ground Cloud Radio Access Network (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 conflict (e.g., cumulatively allocating more resources than available), as also clarified in 3GPP [25]. While this is not the main topic of this thesis, only to justify the consideration of non-terrestrial networks in this work, two different solutions to this issue have been identified and suggested, 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.

Both solutions are transparent to the algorithms for traffic control deployed in this thesis, and they do not impose particular modifications to the proposed architecture. The more promising solution is the one involving network virtualization, and thus network slicing. In [87] Virtualization benefits in the terrestrial-satellite network, integration is presented. Virtualization is considered the key enabler that allows the reproduction of network functions as logical entities that can be assembled in any topology, allowing a dynamical assignment of the resources regardless of technology. Furthermore, virtualization, in the proposed architecture, allows higher protocol layers and related network functions, that are very similar, to be integrated into a common logical central entity. This integration reduce infrastructure and device complexity, and allows the placement of a controller in this common central entity, to supports the single control plane of the physical nodes, including the multi-connectivity control algorithm. On top of this logical level, there is a need for a Multi-domain Orchestrator (MdO) that keeps updated information about the underlying satellite and terrestrial domains and hosts the logic to orchestrate resources and services across the domains. In this thesis, the actions of the orchestrator are considered executed, but the orchestrator himself is not a topic of the thesis.

Finally, the procedure during a connection in the proposed architecture is shown in Figure 3.4, where the particular case of a DL communication is presented. Such procedure involves the following entities, already identified in figure 3.2:

- the UE;
- two gNB-DUs, where the gNB-DUs are characterized by different radio technologies;
- the C-RAN, composed by Flow Control and User Plane upper layer;
- the CN, composed by the QoE Management functionalities.

The procedure represents a scenario in which a DL stream comes from the Data Network to the UE. The procedure steps, as reported in Figure 3.4, are:

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

The actors involved in the procedure, and flows between them, are reported exchange diagram in figure 3.4:

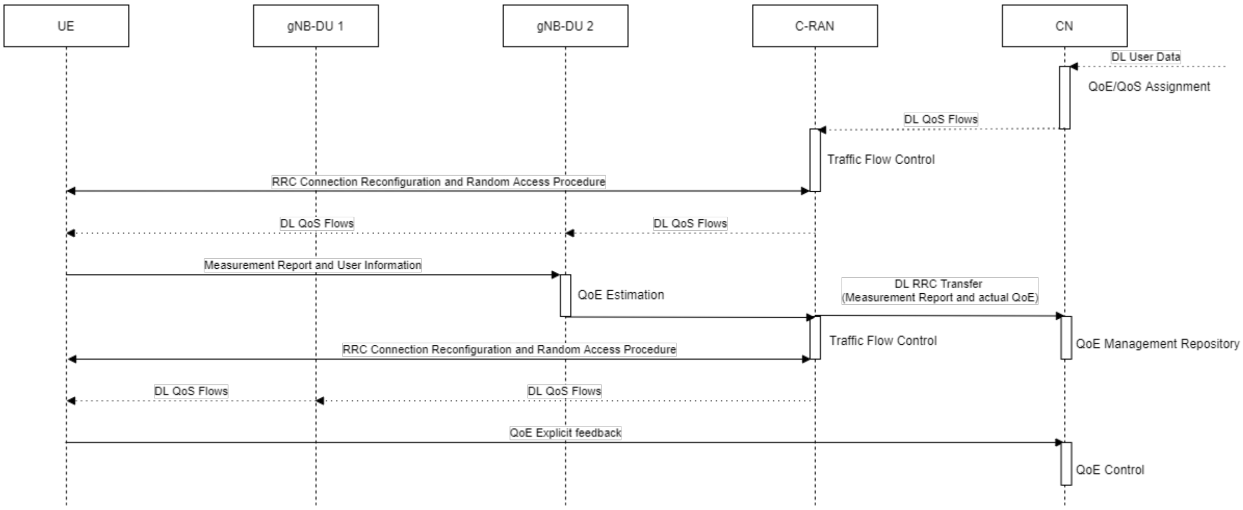


Figure 3.4: Exchange Diagram - Downlink procedure in the proposed architecture

In the following of this thesis, the proposed solution in figure 3.2, is used as a reference for the placement of the developed algorithms in the 5G network.

Chapter 4

Model-Based Multi-Connectivity Control

In this chapter two model-based multi-connectivity control strategies are presented.

The first approach [88] is based on a two-step decision process, to allocate the bandwidth of the network maximizing the users' satisfaction dividing the problem in *Network Ranking* and *Bandwidth Allocation*. The two-steps proposed method allows the separation of the two decision processes that have two different execution rates and solution methodologies, the Network Ranking is executed with a low rate (i.e. minutes) in the Core Network considering high-level data about the past users' connections, the problem is solved using the AHP, that solves the network selection process by evaluating the affinity of each network concerning the users' connections and considering the different network characteristics and the actual users' QoS. The Bandwidth Allocation is executed in real-time, considering the high-level and low-rate results of the Network Ranking process and the real-time measurements from the RAN, this problem is solved dynamically, in a centralized and cooperative fashion trying to satisfy the users' resources request using the actual networks available resources and optimizing the network performances using a Linear Quadratic Difference Game formulation.

The second approach [89] is based on an iterative solution of the resources assignment problem formulated as a MPC, considering load balancing and energy management network aspects. The original resource management problem is modeled as a Quadratic MPC, considering traffic storage and transmission costs, queues' levels, and network dynamic evolution. The problem formulation is then further expanded to avoid assumptions about the disturbances and a fixed horizon, and an iterative solution able to find the optimal resource allocation in the optimal horizon length is provided.

4.1 User-Aware Centralized Resources allocation in Heterogeneous Networks

In the following section, RAN-Controlled approach and RAN-Based Integration are considered as presented in Section 2.1. The RAN-Based Integration allows the cooperation and coordination among RATs while the RAN-Controlled Approach allows to avoid sub-optimal solutions typical of User-Centric approaches. To support these approaches, the considered architecture consists of a CU, containing the CP cooperation and coordination functionalities together with the UP integration and switching functionalities. Such architecture is referred to as Centralized Deployment, presented in Section 2.2. Concerning the network selection problem, the proposed approach is user-aware: it envisages a personalization system whose functionalities are placed in the CogNet. Such a system consists of a repository of the historical connections data for each user and a processing block that, based on the stored information, can synthesize, for each user (or clusters, i.e. homogeneous groups of users), a set of user characteristics in terms of user's Connection Preferences. These Connection Preferences specify the user's needs not in terms of additional QoS constraints, but in terms of personal user's preferences expressed over a set of non-standardized parameters as battery consumption, connection costs, mobility, etc. The stored information is updated based on the user's feedback at the end of each connection and is used for resource allocation. The main contributions of this work are the following:

- The modeling of the heterogeneous network selection problem as a AHP;
- The design of a two-step method that allows the separation of the two decision processes that have two different execution rates;
- the modeling of the bandwidth allocation problem as a Cooperative Linear Quadratic Difference Game between networks, able to consider affinity between networks and users.

The multi-connectivity problem, when considering the user requirements and network status, requires solving two distinct (but related) problems:

- Network selection process, aimed at assigning users, or a portion of their connection traffic, with one or more access networks. The selection process is typically performed statically, and the users-networks association is maintained during the connection life cycle;
- Bandwidth allocation process, aimed at allocating the available resources while matching users' requirements. This problem is solved dynamically, in centralized/distributed and cooperative/non-cooperative fashion trying to satisfy the users' resources request using the currently available network's resources and optimizing the network performances [18, 90].

To solve these two problems, the proposed methodology combines the AHP and a Cooperative Linear Quadratic Difference Game with the following roles:

- AHP solves the network selection process by evaluating the affinity of each network concerning the users' connections and considering the different network characteristics and the actual users' Connection Preferences;
- Cooperative Difference Game solves the networks' resources allocation problem by considering the networks as cooperative agents that optimally distribute their resources between users considering the output of the network selection process.

The resulting approach is thus a model-driven, dynamic algorithm with prioritization of objectives and performances based on the users' preferences and networks' characteristics. The two stages are not solved simultaneously since the AHP stage involves non-real-time data (users and networks information) typically stored in the CogNet by the network operators and updated with a time greater than a connection duration, in the other hand the differential game stage involves real-time data in his computations, typically stored in RAN and updated in real-time accordingly to the connection requirements and the network's conditions.

4.1.1 Analytic Hierarchy Process

The AHP is a methodology used to perform autonomous multi-criteria decisions in complex environments [91, 92]. AHP can involve in the decision process all the criteria and all the candidate solutions to reach a given target. The hierarchy structure in figure 4.1, define the AHP process, consisting of three layers: (i) the target objective, (ii) several evaluation criteria, and (iii) several possible options. This structure drives the decision process starting from level 1 (target object), where the decision-makers define the preferences for each attribute to reach the target. At level 2 (attributes) the decision-makers assign a weight for each attribute with a pairwise comparison, following the preferences defined at level 1. Finally, the weights of the attributes are used to evaluate the grades of the different options.

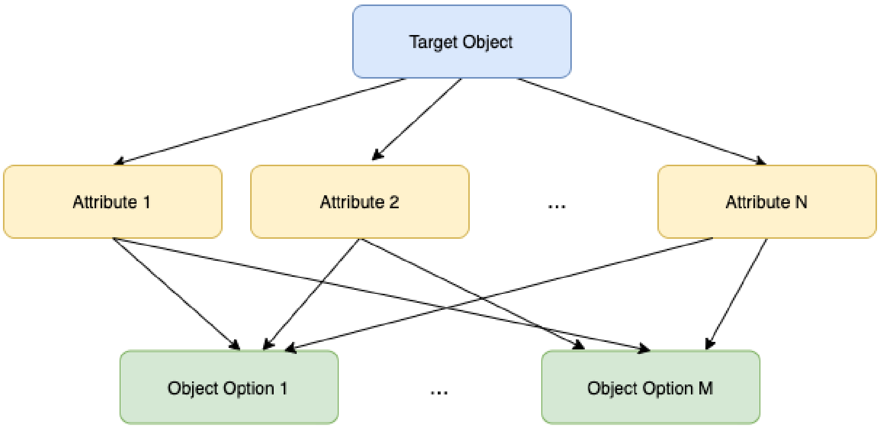


Figure 4.1: AHP Hierarchical Structure

More in detail, the AHP can be summarized in the following steps:

1. Definition of the decision problem. Defining the hierarchical structure, i.e. the target objective, the n evaluation criteria (or attributes) and the m available options identified in the problem;

2. Computation of the vector of attribute scores w. At first, the evaluation criteria are pairwise compared by domain experts, or by autonomous predefined system, using the Saaty's scale (table 4.1) to select the values a_{ij} of the nxn matrix A, where n is the number of attributes, and the element a_{ij} represents the importance of the $i-th$ attribute compared to the $j-th$ attribute. Moreover, the matrix A has the following properties:

   - each element a_{ij} is defined following the Saaty's table;
   - if $a_{ij} > 1$ then $a_{ji} < 1$;
   - if $a_{ij} = 1$ then $a_{ji} = 1$;
   - $a_{ij}a_{ji} = 1$;

   the matrix constructed respecting the properties above is called a reciprocal matrix.

   With the matrix A defined it is possible to compute the elements of the Attribute scores vector w_i, where i is the $i-th$ criteria, as:

$$w_i = \frac{\sum_{l=1}^{n} \bar{a}_{il}}{n} \tag{4.1}$$

where $\bar{a}_{il}$ is the normalized value of the $il-th$ entry of matrix A. The computed column vector w contains the information about the importance of each attribute in the decision process, i.e. w determines the decision criteria weights.

3. Consistency check. In [92] it has been proposed an automatic procedure to avoid inconsistencies (or judgment errors) in the comparison matrix A. In particular, the matrix A is said to be consistent if and only if its greater eigenvalue λ_{max} is equal to n. To evaluate the degree of inconsistency of the comparison matrix, it is possible to define the Consistency Index (CI) as:

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{4.2}$$

By considering this index together with the Random Consistency Index (RI), computed as the average of CIs associated to randomly generated comparison matrices, it is possible to define the Consistency Ratio (CR):

$$CR = \frac{CI}{RI} \tag{4.3}$$

If the CR is smaller or equal to 0.1 the inconsistency is acceptable, otherwise, the subjective judgment needs to be reconsidered, in particular, if the subjective judgment is performed by an automatic process this index can evaluate its definition.

4. Computation of options' scores vector v. For each criterion j, the option score matrix B^j is a mxm matrix whose entries b^j_{ik} are computed by comparing (e.g., according to table 4.1) option i and option k with respect to the $j-th$ criterion. The consistency of matrices B^j can be checked with the procedure described in the previous step. The score of the $i-th$ options with respect to the $j-th$ criteria is computed as:

$$s^j_i = \frac{\sum_{l=1}^{n} \bar{b}^j_k}{n} \tag{4.4}$$

where $\bar{b}^j_k$ is the normalized value of the $jk-th$ entry of matrix B^j. The column vector $s^j = [s^j_1, \dots, s^j_m]^T$ stores the scores of the m options with respect to the $j-th$ criterion. The vector v containing the final scores assigned to the m options is computed as $v = Sw$ where $S = [s^1, \dots, s^n]$ is a mxn matrix that has as columns the scores of the options concerning each criterion. Indeed, the score vector v contains the score of the options concerning each criterion (i.e., the matrix S) weighted by the importance of criteria for the decision process (i.e., the vector w computed at step 2).

Table 4.1: Saaty's Scale

The fundamental scale of Pairwise Comparison		
Intensity	Definition	Explanation
1	Equal	Two attributes contribute equally
3	Moderate	Experience or judgment moderately favor one of the two attributes
5	Strong	Experience or judgment strongly favor one of the two attributes
7	Very Strong	The dominance of the favored attribute is demonstrated in practice
9	Extreme	The evidence favors one element with the highest order of affirmation

4.1.2 Linear Quadratic Difference Game

The differential or difference game [56] is a class of dynamic games that represent a generalization of the optimal control theory. The main difference between differential games and optimal control is the presence of several independent control actions that drive the state evolution. These inputs are managed by different players and there is not a single objective function as each player has its own. Differential games are characterized by:

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

Differential games can be cooperative, where all the players have the same objective function, or non-cooperative, where each player has his own objective function. In this work, the former class of games is considered. That is, players, coordinate their strategies given optimizing a collective objective function: selfish behaviors are rewarded less than collaborative ones. The problem of interest can be modelled as a Discrete-Time Linear Quadratic (LQ) differential game [73] whose general formulation for two players is:

$$u^* = \arg\min_u J(x,u,k) = x^T(K)Fx(K) + \sum_{k=k_0}^{K-1} x^T(k)Qx(k) + u_1^T(k)R_1u_1(k) + u_2^T(k)R_2u_2(k) \quad (4.5)$$

s.t.

$$x(k+1) = Ax(k) + B_1u_1(k) + B_2u_2(k); \quad x(k_0) = x_0; \ \ x(K) = x_f \quad (4.6)$$

$$0 \leq x(k) \leq \bar{x}; \quad 0 \leq u_1(k) \leq \bar{u}_1; \quad 0 \leq u_2(k) \leq \bar{u}_2; \quad \forall k \in [k_0, K] \quad (4.7)$$

$$F > 0; \quad Q > 0; \quad R_1 > 0; \quad R_2 > 0 \quad (4.8)$$

$$x^T(k) = [x_1(k), \ldots, x_n(k)] \quad (4.9)$$

$$u_1^T(k) = [u_1^1(k), \ldots, u_1^m(k)]; \quad u_2^T(k) = [u_2^1(k), \ldots, u_2^m(k)] \quad (4.10)$$

$$d^T(k) = [d_1(k), \ldots, d_p(k)] \quad (4.11)$$

where K is the final time, $x(k)$ is the variable state of the dynamical system to be controlled at time k (with x_0 and x_f respectively initial and final state values), $u_1(k)$ and $u_2(k)$ are the vectors of variables which can be manipulated by two players, $d(k)$ models disturbances at time k and F, Q, R_1 and R_2 are weight matrices. The matrices A, B_1 and B_2 are the system dynamic matrix and the input matrix of the two players, respectively. The values $\bar{u}_1$, $baru_2$ and $\bar{x}$ are the upper bounds of the variables $x(k)$, $u_1(k)$ and $u_2(k)$ respectively.

4.1.3 Problem Formulation

The mathematical formulation of the proposed methodology is developed dividing the problem into two sub-problems able to decouple the two process involved in the decision, each one with its own data production rate. The two sub problems are identified in network selection problem formalized as an AHP model whereas, and user-aware bandwidth allocation problem formalized as a cooperative LQ difference game.

In figure 4.2 the two separated control loops are shown, underlying the different rate between the network selection problem, solved by the *AHP Processor* starting from the users' feedback (i.e., a low rate measurement), and the bandwidth allocation problem, solved by the *DIfferential Game Solver* that is an optimization solver that allocate the bandwidth using feedback about the network status (i.e., an high rate measurement).

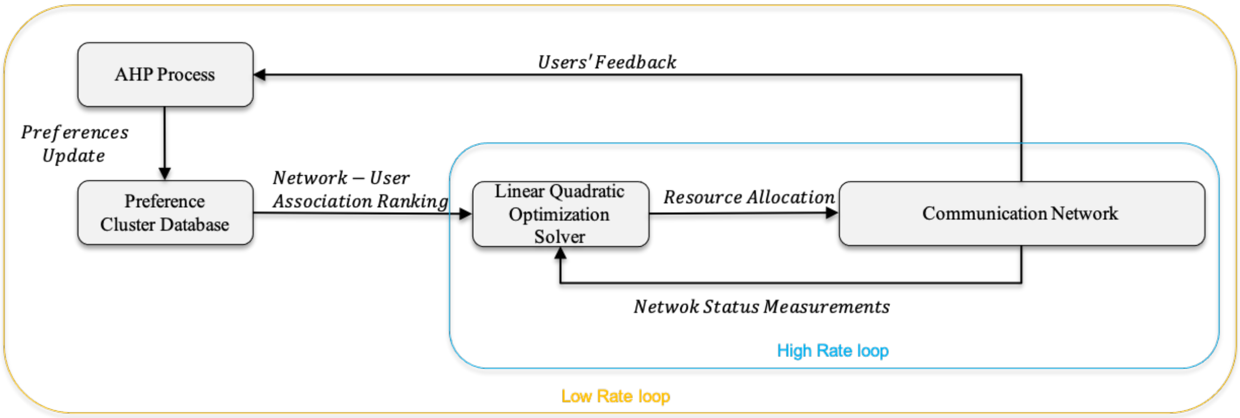


Figure 4.2: Control Schema

AHP-based access network ranking

As described in Section 4.1.1, to use the AHP it is necessary to define the target objective, the evaluation criteria (attributes), and the available options. The target objective, in the considered scenario, is the selection of the most suitable network considering user preferences. The criteria against which access networks can be evaluated can be defined in terms of (sub) sets of intrinsic network characteristics (e.g., cost, security, packet delay, packet loss, bandwidth, coverage). In heterogeneous networks, the variance between the values of these characteristics can be significant. For example, satellite access networks are typically characterized by high-security levels, delays, costs, coverage, and bandwidth while terrestrial access networks offer low costs and delays with limited security, coverage, and bandwidth. The options, in a heterogeneous network, are represented by all the available access networks associated with given users. Such networks are characterized in terms of the above mentioned intrinsic characteristics. Based on the defined hierarchical structure, it is possible to apply the steps defined in Section 4.1.2. In particular, the two comparison matrices can be computed and stored by the network operators or the service providers, to be used during the connections. The outcome of this first step is thus a matrix R whose generic entry r_{ij} represents the score assigned to the choice of associating user/cluster i to network j. Summarizing, to model the network ranking problem as an AHP problem the three layers can be defined as follow:

- *target objective*: define the affinity between access networks' characteristics and users' requirements;

- *evaluation criteria*: defined by QoS/QoE characteristics, such as Bandwidth (BW), Latency (L), Packet Loss (PL), Cost (C), Mobility (M) and Battery (B);
- *possible options*: defined as the available access networks, such as Satellite, 5G and WLAN.

A further refinement of such static assignment considers users' past interactions: implicit and explicit users' feedbacks, made based on the experienced performances, can be used to update the comparison matrices, namely the judgments between the possible options. This leads to the definition of a dynamic QoE Management framework.

Bandwidth allocation as a Cooperative Difference Game

The second step of the proposed methodology consists of deciding how to allocate the available resources within the different access networks. Based on the static association between access networks and users, a cooperative game is set up to dynamically solve the resource allocation problem. Such allocation is performed considering users' (or services) requirements and the networks' status. In the assumption that networks do not have full queues (but only limited resources in terms of bandwidth), the backhaul capacity of each access network is assumed to be greater than its allocated bandwidths. The cooperative difference game can be formulated as a particular case of (4.5)-(4.8), the main differences are about the final state value constraint since we are interested in minimizing that value and not to have a predefined final value. Furthermore, is considered the only linear contribution of the control variables in the cost function, because the aim is to maximize the usage of the network with higher grade, using the result of the AHP as weight matrices R_i, that is used as a negative contribution in the cost function to be minimized, to maximize their contributions, on the contrary, the use of the quadratic expression of the control variables u_i leads to minimizing the assigned resources, that can result in the network under use, implying users' queue growth. The resulting LQ formulation is:

$$[u^*, v^*, w^*] = \arg\min_{u,v,w} J(x(k), u(k), v(k), w(k), k) = \sum_{k=k_0}^{K} x^T(k)Qx(k) - u(k)^T R_u - v(k)^T R_v - w(k)^T R_w \tag{4.12}$$

s.t.

$$x(k+1) = x(k) - u(k) - v(k) - w(k) + d(k), \quad x(k_0) = x_0 \tag{4.13}$$

$$0 \leq x(k) \leq \bar{x} \tag{4.14}$$

$$0 \leq \sum_{i=1}^{p} u_i(k) \leq \bar{u}; \quad 0 \leq \sum_{i=1}^{p} v_i(k) \leq \bar{v}; \quad 0 \leq \sum_{i=1}^{p} w_i(k) \leq \bar{w} \tag{4.15}$$

$$u_i(k) \geq 0; \quad v_i(k) \geq 0; \quad w_i(k) \geq 0; \quad Q > 0; \quad R_u > 0; \quad R_u > 0; \quad R_w > 0 \tag{4.16}$$

$$x^T(k) = [x_1(k), \ldots, x_p(k)] \tag{4.17}$$

$$u^T(k) = [u_1(k), \dots, u_p(k)]; \quad v^T(k) = [v_1(k), \dots, v_p(k)]; \quad w^T(k) = [w_1(k), \dots, w_p(k)] \tag{4.18}$$

$$d^T(k) = [d_1(k), \dots, d_p(k)] \tag{4.19}$$

where K is the final time in the optimization window, namely the time in which the information about the network's conditions can be considered static, $x(k)$ is the state of the evolving state of the queues (with x_0 initial values) and $u(k)$, $v(k)$ and $w(k)$ vectors are the variables which can be manipulated by the three players (i.e., the networks), in particular the single $i-th$ element of each of these vectors, is the bandwidth allocated by the network to the $i-th$ user (or cluster of users). The traffic arriving at the users' queues is modeled as a disturbance $d(k)$ and its evolution depends on the service relative to the cluster itself. The values $\bar{x}, \bar{u}, \bar{v}$ and $\bar{w}$ are the upper bounds of the respective variables. The matrices $R_{.}$ are extracted by the matrix $R = [R_u, R_v, R_w]$, namely the output of the first step of the proposed methodology. The weight matrix Q allows associating to given user higher relevance in the cost function allowing to capture different contractual positions.

4.1.4 Simulation and Results

The selection criteria considered in the simulations are: Bandwidth (BW), Latency (L), Packet Loss (PL), Cost (C), Mobility (M) and Battery (B). The options are Satellite, 5GTerrestrial, and WLAN. The Satellite network is characterized by high BW, M, C, B, and L but low PL; the 5G-Terrestrial network by low L and medium levels of all the attributes; the WLAN network by low C, L, B, M and medium levels of the other attributes, applying AHP this assignment results consistent and produce the attribute priority between options (i.e., access networks) in figure 4.3. The target objective consists of steering the traffic of three clusters of homogeneous users whose preferences, in order of relevance are: Cluster 1: BW, PL, M; Cluster 2: L, C, M; Cluster 3: L, C, B, applying AHP this assignment results in the following clusters' score of the attributes (i.e., QoS indicators) in figure 4.4.

The time step is assumed to be 1 unit of time while the clusters' queues are assumed to start with 2 Mb stored. The bandwidth limits for the considered networks are 20Mpbs for Satellite, 10 Mbps for 5GTerrestrial, and 6 Mbps for WLAN. The resulting AHP option score (i.e., networks ranking) according to the attribute priority and the attribute score is presented in figure 4.5. The clusters are characterized by different services, each service requires a different amount of data traffic; in figure 4.6 the traffic required by the three clusters is depicted. The network ranking capturing the clusters' preferences and the clusters' demand in terms of bandwidth, depicted in Fig. 5, drives the cooperative LQ differential game described in 4.1.3. In particular, the networks cooperate to match clusters' demands while considering their preferences. This cooperation is possible in virtue of the chosen multi-connectivity integration and management strategies and the network architecture (i.e. RAN-based integration and RAN-Controlled management, see section 3.1).

Simulations show that all the bandwidth of the satellite network is allocated to C1 (see figure 4.7) and, when its demand exceeds the satellite bandwidth limit, traffic is steered on the 5GTerrestrial (see figure 4.8). This result is in line with the AHP ranking: C1 prefers the satellite and, in the second instance, the WLAN (see figure 4.5) which, however, is congested. The bandwidth assigned

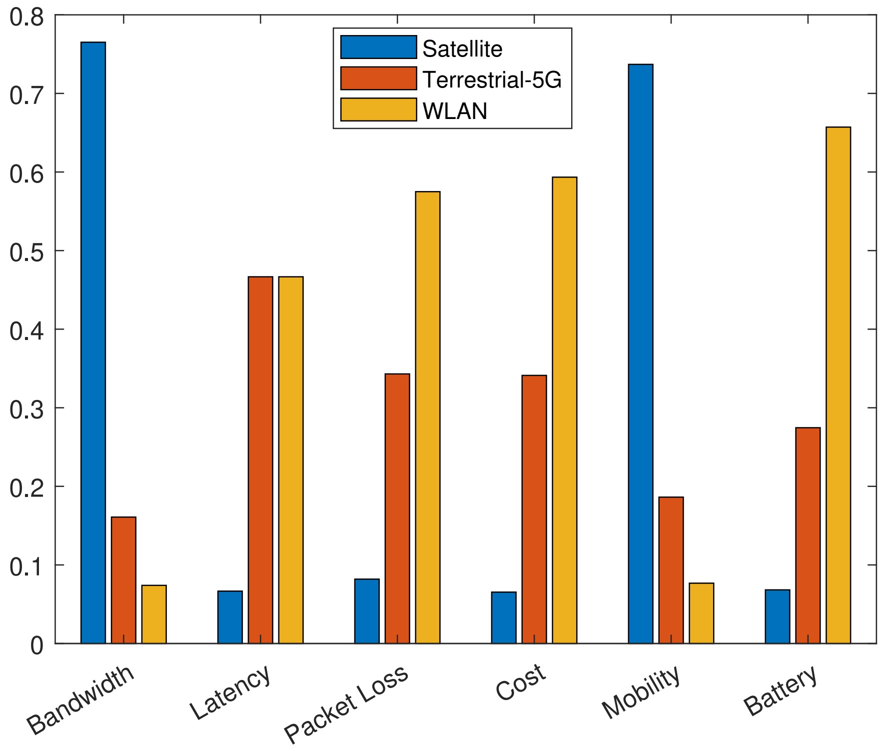


Figure 4.3: AHP Attribute Priority Vector Values

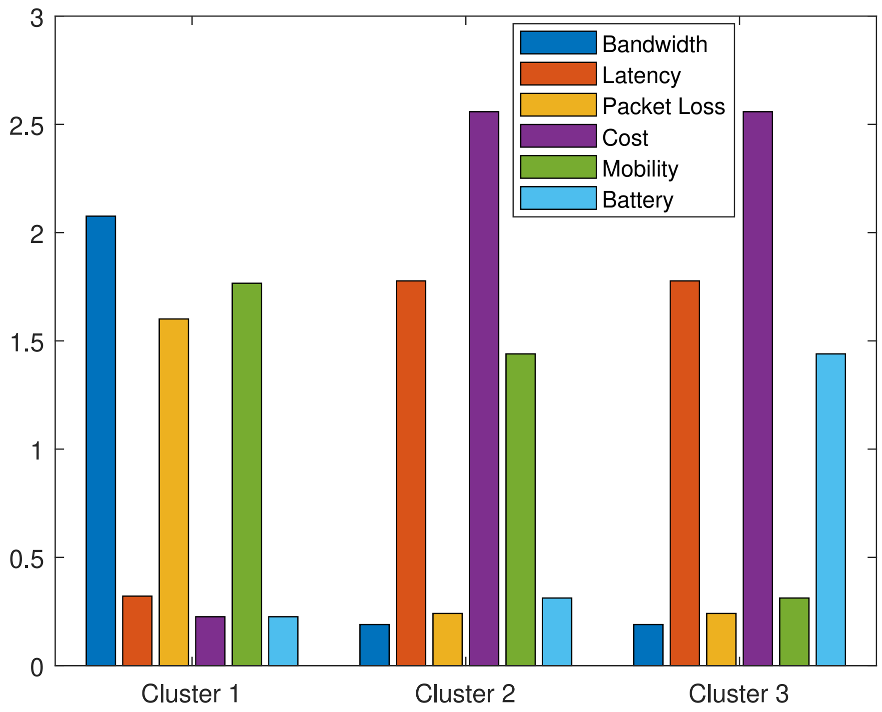


Figure 4.4: AHP Attribute Score Vector Values

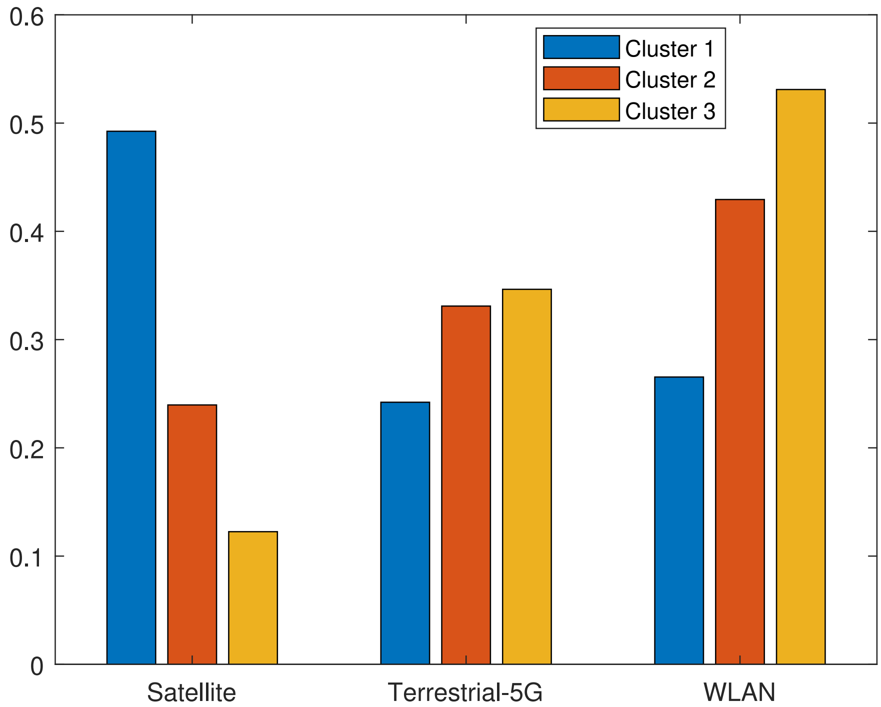


Figure 4.5: AHP Networks Ranking

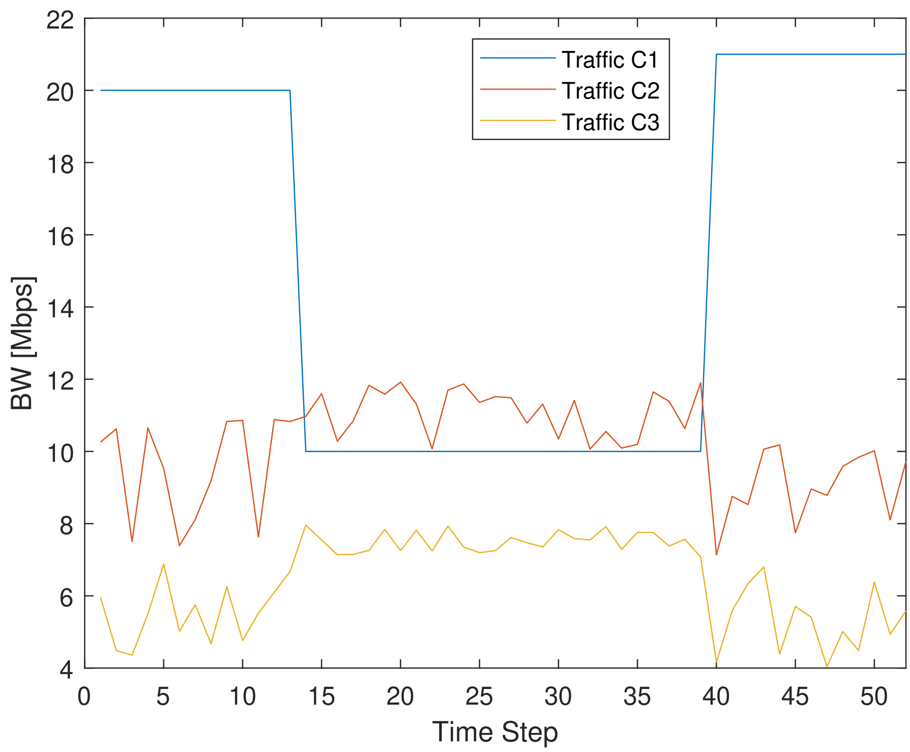


Figure 4.6: Traffic required by the different clusters

to cluster 2 is shown in figure 4.9, showing that the 5G-Terrestrial network is the most used by cluster 2; when the latter is congested, the traffic is steered on the satellite network since the WLAN is saturated by C3. This result is in line with the AHP ranking: C2 prefers the WLAN which, however, is completely used by C3 (for which it represents the first choice) and thus traffic is steered on the 5G-Terrestrial network which represents C2's second choice. The bandwidth assigned to cluster 3 is shown in figure 4.10, due to the strong preference of C3 for the WLAN network, its traffic is mostly sent on such network; when the WLAN has no bandwidth to assign, the 5G-Terrestrial is used, which represents the C3's second choice. It is possible to note that, when the Satellite has free bandwidth and the WLAN is congested, the proposed algorithm decision is to reallocate the 5G-Terrestrial bandwidth previously allocated to C2 in favor of C3; C2's demand is then satisfied with the satellite bandwidth which, instead, represent the least favorite option for C3. The aggregate allocated bandwidth of each network is depicted in figure 4.7, as shown, the networks' bandwidth limits are respected, and only the 5G-Terrestrial isn't fully allocated. This is since the 5G-Terrestrial is the only network that isn't the first choice of some cluster. The observed behavior shows that the clusters with a great preference for a specific network are favored against the "undecided" clusters.

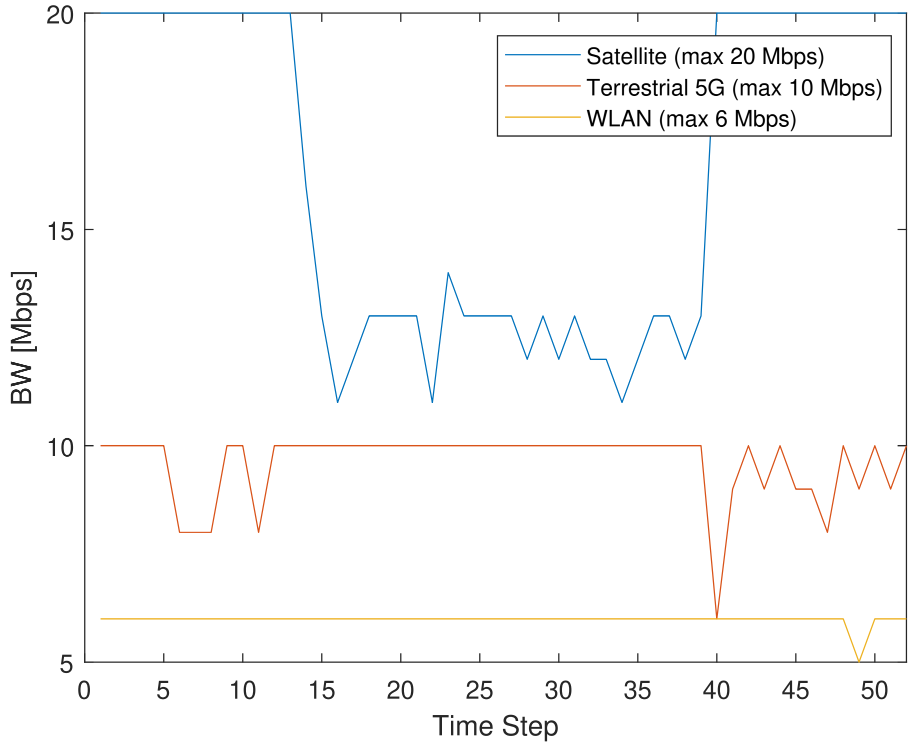


Figure 4.7: Allocated bandwidth per network

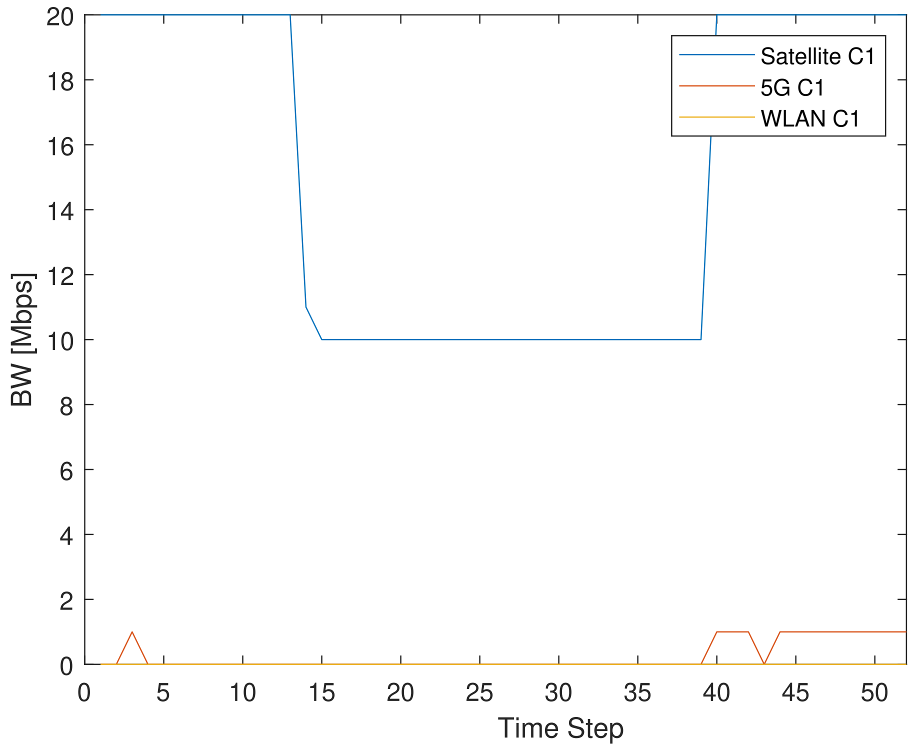


Figure 4.8: Bandwidth allocation for cluster 1

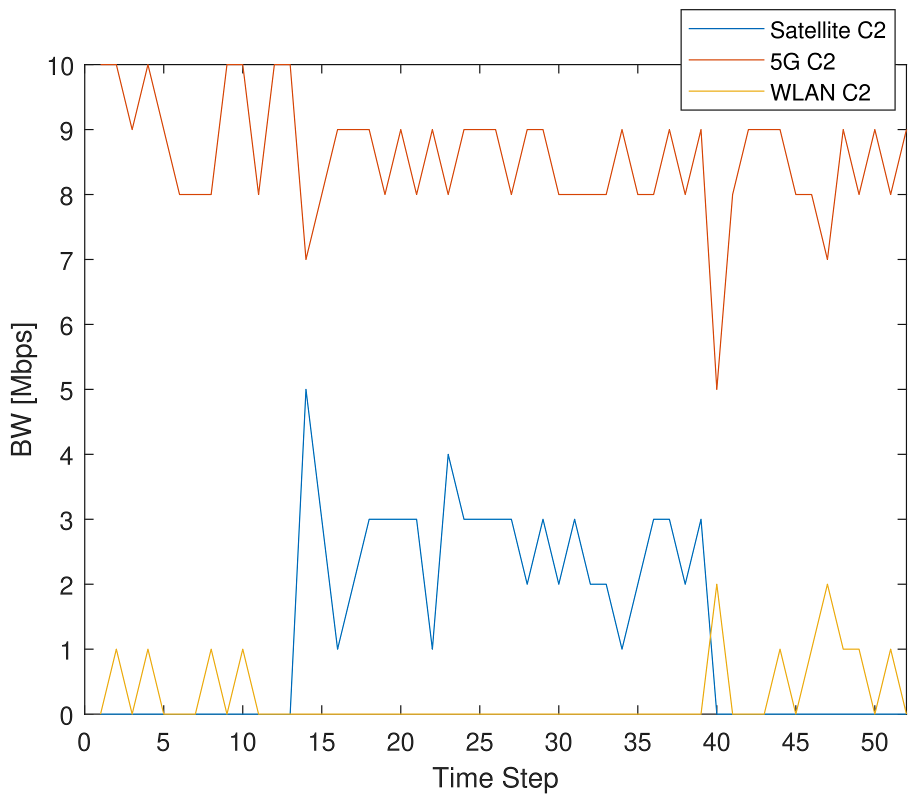


Figure 4.9: Bandwidth allocation for cluster 2

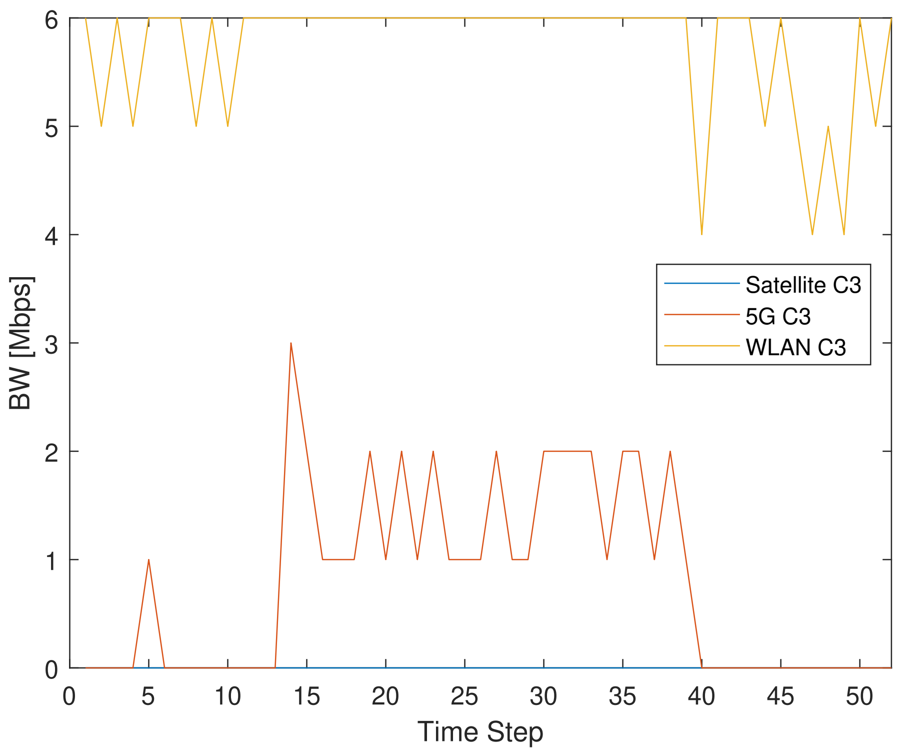


Figure 4.10: Bandwidth allocation for cluster 3

4.2 Iterative MPC for Energy Management and Load Balancing in 5G Heterogeneous Networks

The problem addressed consists of fairly distributing loads in multi-access networks to optimize the network's energy consumption. This section is presented a control strategy to optimize the transmission power in a 5G heterogeneous network, by defining a dynamic algorithm for queue management. The methodology selected for this purpose is based on MPC and the problem is modeled as a linear dynamical system with a quadratic cost function to be optimized. The main contributions of this work are the following:

- The modeling of a heterogeneous network with multiple RATs as a dynamical system in which the controller of the UEs regulate their transmission powers to sustain the connections;
- The design of a centralized Iterative MPC control system for the optimal network resource management in a Multi-Connectivity setting;
- The validation of the proposed approach with standard solutions to evaluate its performance.

In the considered scenario, users can transmit and/or store the traffic produced by their applications (e.g., chats, browsing, video streaming) using different technologies (e.g., WLAN, LTE, etc.). These two actions are characterized by different costs defined in terms of the consumed energy which depends on the user's storage technology, transmission technology (i.e., the selected Access Network Technology), and distance from the AP. Users are also characterized in terms of a maximum storage capacity and a maximum transmission bandwidth due to constraints imposed by the physical transmission channel used and the APs' limited processing capabilities. The objective is thus to allocate to the users the APs' available bandwidth respecting users' storage capacity while minimizing the energy consumption of the overall network. To solve this problem, an Iterative Model Predictive Control (I-MPC) problem is set up to optimize the network's energy consumption in a centralized way. The main advantage of using I-MPC compared to standard MPC solutions is the ability of the controller to determine, at each time step, the optimal prediction window length.

4.2.1 Model Predictive Control

The model-based approach discussed above can be solved using several methods (e.g., see [54, 93, 94]). According to the MPC paradigm, the controller shall compute the solution to the constrained optimization problem and then apply the first element of the optimal control vector to the system and discard its remaining entries. At the following sampling time, the MPC controller shall re-evaluate the optimal control actions, in what is called a receding horizon approach. This iterative procedure is what gives MPC both the characteristics of optimal open-loop control and feedback-based closed-loop control systems.
In figure 4.11 the MPC block schema is represented. It is composed of a model-based optimizer, i.e., a block containing the optimization algorithm, the model of the process used to predict its likely evolution, and the optimization solver. The optimization algorithm can produce a control action (i.e., the input of the process $u(t)$) based on the desired output (i.e., $r(t)$) and the actual process states (i.e., the process measurements $x(t)$), these measurements are used as the initial state of the prediction model. Therefore, an optimization solver is used when the Explicit MPC approach is adopted, i.e. when the optimization algorithm is solved at each iteration using a specific commercial solver. On the other hand, the optimization problem can be solved offline, in this case, the approach is called Implicit MPC, and it has the advantage of being faster in the control action computation, but it is harder to design because it needs a closed-form solution of the optimization problem.

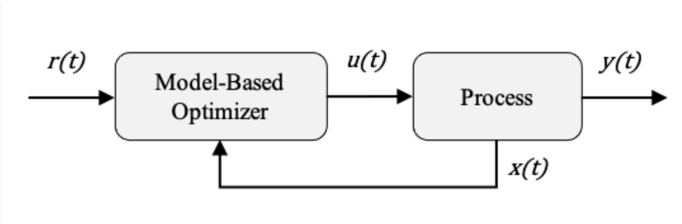


Figure 4.11: MPC control schema

The peculiarity of this approach is the capability to compute a control action at time t_0 that optimizes the process behavior in a prediction window $[t_0, T]$, where T is called prediction horizon. Note that only the control action at the first time instant (i.e., $u^*(t_0)$) is used as input of the process, all the other computed control actions (i.e., $u^*(t) \;\; \forall t \in (t_0, T]$) are discarded. In the next time step, the prediction window is moved towards the future of one step. This behavior gives to the MPC the name of receding horizon technique.
Summarizing the main steps of the MPC are at each time instant t:

1. get new measurements to update the estimate of the current state of the process $x(t)$;

2. solve the optimization problem with respect to the prediction horizon to compute $u^*(t)\ \ \forall t \in [t_0, T]$;

3. apply only the first optimal control action to the process, i.e. $u(t) = u^*(t_0)$, discard the remaining samples.

4.2.2 Problem Formulation

The problem can be modeled as a quadratic optimization problem where the cost function J represents the network's consumed energy. As already mentioned, the consumed energy can be defined in terms of the total storage (E_s) and transmission (E_{Tx}) costs. The former models the storage energy consumption and is a constant depending on each user's storage technology. The transmission cost, instead, depends on (i) the transmission technology, (ii) the distance $d_{i,j}$ between the $i-th$ user and the $j-th$ AP and (iii) the medium used for the transmission. Hence, the transmission cost associated to the $i-th$ user and the $j-th$ AP, referred to as $E^{i,j}_{Tx}$, can be defined as

$$E^{i,j}_{Tx} = C_f + C_s(d_{i,j}) \tag{4.20}$$

where C_f is a constant depending on the $i-th$ user's transmission technology and C_s, the function of the distance between the user and the AP, depends on the medium used for the transmission. The resulting cost function can be defined as:

$$J(q,u) = \sum_{i=1}^{N}(q_i^2(k)E_s^i + \sum_{j=1}^{M} u_{i,j}^2(k)E^{i,j}_{Tx}) \tag{4.21}$$

where N is the number of users, M is the number of APs, $q_i(k)$ is the traffic stored by the $i-th$ user at time k, and $u_{i,j}(k)$ is the traffic transmitted by the $i-th$ user towards the $j-th$ AP. The optimization problem described can be formulated as a MPC problem as follows:

$$u^* = \arg\min_u J(q,u,K) = \sum_{k=0}^{K-1}\sum_{i=1}^{N} J(q,u) \tag{4.22}$$

$$0 \leq q_i(k) \leq Q_i \quad \forall i,k \tag{4.23}$$

$$0 \leq \sum_{i=1}^{N} u_{i,j}(k) \leq U_j \quad \forall j,k \tag{4.24}$$

$$q_i(k+1) = q_i(k) + d_i(k) - \sum_{j=1}^{M} u_{i,j}(k) \quad \forall i,k \tag{4.25}$$

where K is the time horizon over which the optimization is performed and (4.22) specifies that the transmission energy of all the users is the objective function to be minimized, (4.23) guarantees that, for each user i and each time instant k, the stored traffic $q_i(k)$ does not exceed the maximum storage capability Q_i, (4.24) guarantees that for each AP j and each time instant k, the traffic received by AP j by all users does not exceed the maximum AP's capacity U_j, and (4.25) models the dynamic evolution of the storage level of user i which depends on the traffic stored and transmitted at time k (i.e., $q_i(k)$ and $u_{i,j}(k)$, respectively) and on $d_i(k)$ which represents the traffic generated by the user at time k. The mathematical model (4.22)-(4.25) consists in a LQ-MPC with fixed time horizon K as described in 4.2.1. Concerning equation (4.25), it should be noted that, for the considered problem, the traffic $d_i(k)$ generated by the user represents a disturbance which should be estimated for each $k > 0$. Errors in the estimation of such variables translate into a poor description of users' storage level evolution and, in turn, of the network's dynamics. Furthermore, a fixed time horizon may lead to slower network performances. Indeed, by tailoring the time horizon at each time step, it is possible to find the optimal load distribution while minimizing the time at which users' queues are empty. To address the two above-mentioned issues, an iterative implementation of the LQ-MPC described by (4.22)-(4.25) has been proposed. The proposed implementation is aimed at exploiting the available knowledge about the users' generated traffic $d_i(k)$ to improve the quality of solutions generated by the model (4.22)-(4.25). To achieve this purpose, at each instant of time the disturbances $d_i(k)$ are measured and the optimization is performed by taking in consideration only the measured values. In other words, at each given time $\bar{k}$, the status of each user is gathered and their generated traffic $d_i^m = d_i(\bar{k})$ is measured. For exploiting such knowledge, the model (4.22)-(4.25) can be extended as follows. Let d_i^m be the traffic generated by the $i-th$ user at time $\bar{k}$ and, without loss of generality, assume $\bar{k} = 0$. Then, the above-mentioned issues can be addressed by considering the following additional constraints:

$$d_i(0) = d_i^m, \quad \forall i \tag{4.26}$$

$$d_i(k) = 0, \quad \forall k > 0 \tag{4.27}$$

$$q_i(K) = 0 \tag{4.28}$$

where (4.26) allows to take into consideration the measured values of the disturbances, (4.27) specifies that the optimization is performed without taking into consideration future unknown values of the disturbances, and (4.28) specifies that at the end of the time horizon users' queues should be empty. It should be noted that the formulation (4.22)-(4.28) is heavily impacted by the choice of the time horizon. Indeed, if the time horizon is too small, then the problem may result unfeasible since there would not be enough time to empty users' queues (constraint (4.28)). On the other hand, large time horizons do not allow to select optimal control actions to set the users' queues empty. To address this problem, it is possible to adopt an iterative solution for finding the most suitable time horizon at each time step. The pseudo-code of the proposed solution is reported in Algorithm 1, where $(\bar{q}, \bar{u}, \bar{K})$ is the optimal load distribution and sent traffic for each user i over the time window $\bar{K}$, $\bar{K}_0$ is the initial time horizon and $d_i^m = (d_1^m, \ldots, d_N^m)^T$ is the measured disturbance at time $k = 0$ for each user i.

4.2.3 Simulation and Results

To validate the proposed I-MPC solution, several simulations have been performed focusing on (i) the fairness of the load distribution among the users and (ii) the energy consumption. The

Algorithm 1: Iterative MPC

Result: $(\bar{q}, \bar{u}, \bar{K})$
$\widetilde{K} \leftarrow \bar{K}_0$;
$d(0) \leftarrow d_i^m$;
$d(k) \leftarrow 0_N, \quad k > k_o$;
$(\bar{q}, \bar{u}, \bar{K}) \leftarrow \arg\min_{q,u,\widetilde{K}} J(q, u, \widetilde{K}) \;\; s.t. \;\; (4.22) - (4.28)$;
if $(\bar{q}, \bar{u}, \bar{K})$ *is feasible* **then**
  while $\widetilde{K} > 0$ **do**
    $\widetilde{K} \leftarrow \widetilde{K} - 1$;
    $(\widetilde{q}, \widetilde{u}, \bar{K}) \leftarrow \arg\min_{q,u,\widetilde{K}} J(q, u, \widetilde{K}) \;\; s.t. \;\; (4.22) - (4.28)$;
    if $(\widetilde{q}, \widetilde{u}, \bar{K})$ *is feasible* **then**
      $(\bar{q}, \bar{u}, \bar{K}) \leftarrow (\widetilde{q}, \widetilde{u}, \widetilde{K})$;
    else
      return $(\bar{q}, \bar{u}, \bar{K})$;
    end
  end
else
  return empty solution // Problem is unfeasible or $\bar{K}_0$ is too small;
end

performances of the proposed approach, i.e. the model described by (4.22)-(4.28) implemented according to the algorithm reported in 1, are compared with those of a standard MPC, i.e. the model described by (4.22)-(4.25), and a Load Balancing Controller based on the *Most Loaded* principle (at each time step k, users with higher queue levels are assigned with the needed bandwidth).
Simulations have been performed assuming 6 users and 3 APs. The storage costs E_s^i (expressed in energy unit) have been assumed equal for all users:

$$E_s = [1.5,\ 1.5,\ 1.5,\ 1.5,\ 1.5,\ 1.5]^T \tag{4.29}$$

The transmission costs (expressed in energy unit), depending on the network's topology, transmission technology, and used medium are

$$E_{Tx} = \begin{bmatrix} 10 & 8 & 10 & 1 & 10 & 60 \\ 3 & 20 & 3 & 8 & 30 & 8 \\ 40 & 4 & 2 & 3 & 10 & 40 \end{bmatrix}^T \tag{4.30}$$

The users' initial storage levels $q\,(0) = q_0$ (expressed in packets) are:

$$q_0 = [5.8,\ 1.9,\ 1.6,\ 5.1,\ 1.7,\ 1.1]^T \tag{4.31}$$

The maximum storage levels Q and APs' capacities U (both expressed in packets) are

$$Q = [15,\ 15,\ 15,\ 15,\ 15,\ 15]^T, \quad U = [19,\ 7\ ,10]^T \tag{4.32}$$

Finally, the traffic d_i (expressed in packets per time step) to be sent by each user has been considered as a random, piece-wise continuous signal uniformly distributed between 1 and 7.

As already mentioned, simulations are aimed at studying the performances of three control strategies concerning the fairness of the (i) load distribution and (ii) energy consumption. Concerning the first aspect, in figure 4.12 it is shown the mean storage level perusers. As expected, the Most Loaded approach guarantees the fairest load distribution. However, the amount of traffic stored is higher concerning the I-MPC and MPC approaches. In presence of traffic spikes, this aspect could translate into network congestion or higher packets' waiting times. On the other hand, the I-MPC performs better than the MPC concerning the fairness of load balancing. More in detail, the MPC tends to allocate the APs' to users with lower transmission costs, without considering the storage level. The I-MPC, instead, can take into consideration also users' loads due to constraint (4.28) and the adopted algorithm (see 1), i.e. the imposition of emptying users' queues in the minimum time steps. Concerning the performances of the three approaches concerning the energy consumption problem, in figure 4.13 it is shown the total consumed energy per user. It can be observed that, as expected, the Most Loaded approach performs very poorly since it does not take into consideration energy-related aspects. On the other hand, the performances of the I-MPC and MPC approaches are quite similar though the latter behaves slightly better at the expenses of a significantly higher storage level. Furthermore, as highlighted in figure 4.2.1, the I-MPC guarantees lower peaks of energy consumption for all users. This aspect is particularly relevant in IoT scenarios which envisage a multitude of connected devices with limited processing and battery capabilities, that may have limitations on their maximum transmission power output.

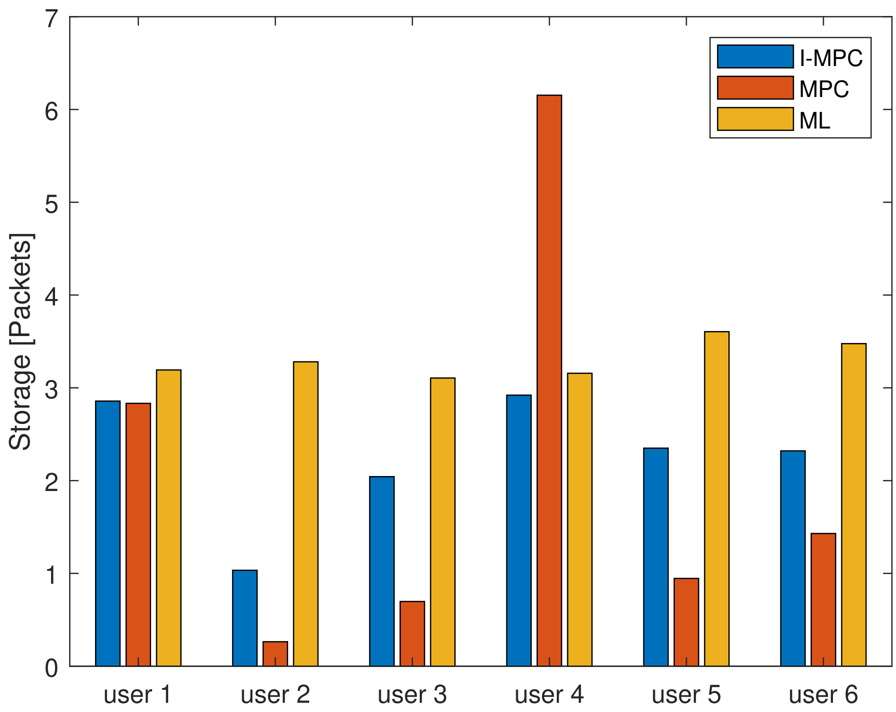


Figure 4.12: Mean Storage level per User

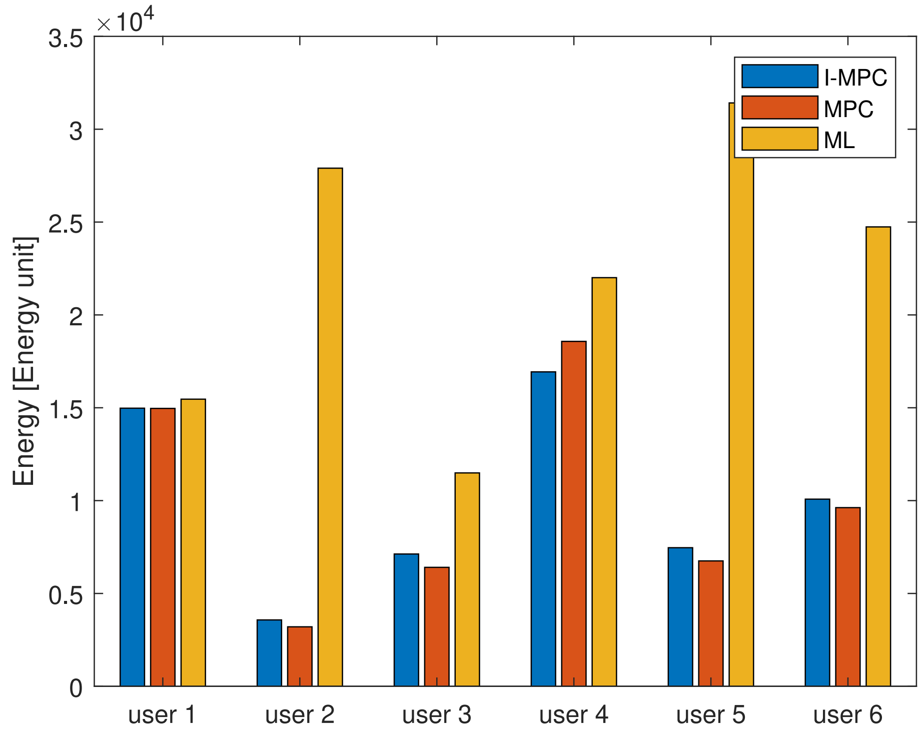


Figure 4.13: Total Energy Consumption per User

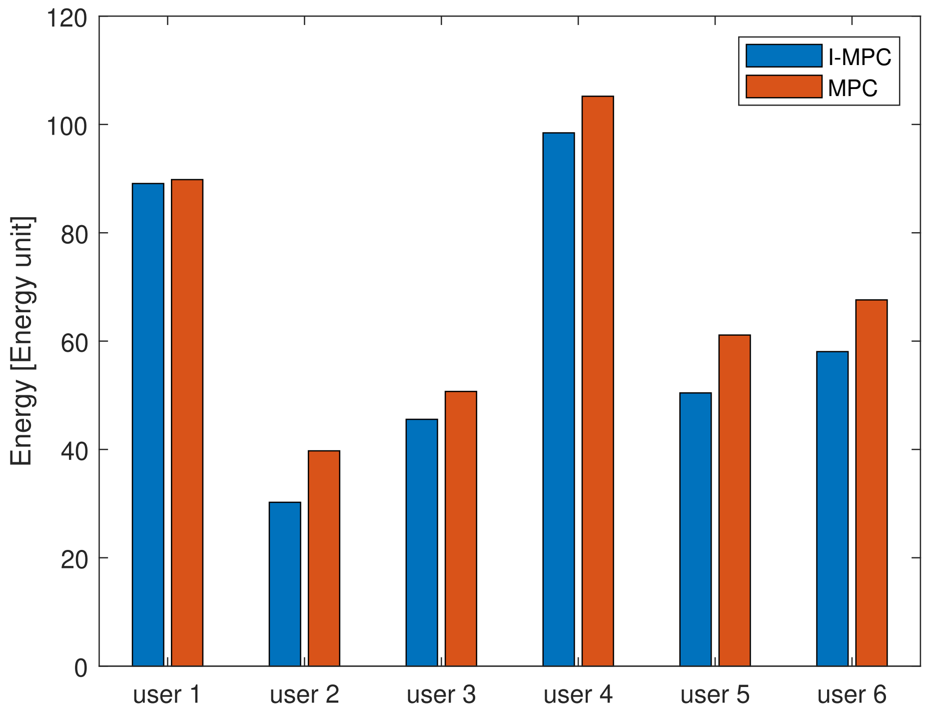


Figure 4.14: Max Energy Consumption for time instant per user

Chapter 5

Data-Driven Multi-Connectivity Control

In this chapter two data-driven multi-connectivity control strategies are presented.

The first approach is a Reinforcement Learning based hierarchical control strategy with the capability of simultaneously tacking the load balancing and QoS management problems in a scalable, dynamic, and closed-loop way. The underlying idea is to associate a Local Controller to each AP. That Local Controlled are coordinated by a single Global Controller located in the C-RAN. The use of a Central Controller allows to coordinate the Local Controllers without information exchange between them, and on the other hand, allows the reduction of the Local Controllers' actions and state space.

The second approach [95] is a fully distributed Reinforcement Learning algorithm for controlling the power consumption in wireless networks. More in detail, the goal consists in defining the best strategy to be adopted for managing, in an efficient way, transmission power. The proposed approach uses a consensus protocol to share information in the network, namely the min-consensus.

5.1 Hierarchical RL for Load Balancing and QoS Management in Multi-Access Networks

In this section, a hierarchical reinforcement learning algorithm to provide a scalable, implementable, and powerful solution to the load balancing and mobility management problem is presented.
In 5G networks, according to the architecture proposed in Section 3.1, the flow of packets relevant to a given connection can be dynamically split into several sub-flows that follow different paths (i.e., network access) and are recombined at the destination. The proposed algorithm deals with Load Balancing problem, maximizing the exploitation of the air interface of a 5G multi-access network. The algorithm aims to provide to the users the best connectivity, considering several radio access networks, possibly involving heterogeneous technologies, and considering the air interface band which is used to link UE and AP as a valuable and limited resource. In this respect, the Load Balancing problem which is dealt with, consists of dynamically selecting the most appropriate routing over the air interface (i.e. in the run from/to UE to AP) of the packets relevant to connections in progress, taking into account the QoS profile of such connections, i.e. trying to respect its constraints in terms of throughput, latency, BER, jitter, and mobility. However, the management of multi-access networks can be challenging, due to the larger number of factors to consider associating each user with the best access, namely users' requirements and network condition.

The methods proposed in the literature to solve the above-mentioned problem, recalled in Section 2.3, can result in some cases not adequate because of their computational complexity or poor performances. In the following, a hierarchical implementation of a reinforcement learning algorithm with dynamic and adaptive characteristics, but with low implementation and computational complexity, because the hierarchical structure, is presented.

5.1.1 Reinforcement Learning

The reinforcement learning problem [19, 49, 59, 60, 96] is the problem of learning from interaction to achieve a goal. The learner and decision-maker is called *agent* (i.e., the controller), it interacts continuously with the *environment* (i.e., plant): the agent select *actions* (i.e., control signal) and the environment change its state reacting to the performed actions. The environment, changing its state produces a *reward* (i.e., feedback), namely a numerical value that the agent tries to maximize over the time selecting the appropriate actions. The agent-environment interaction is depicted in figure 5.1

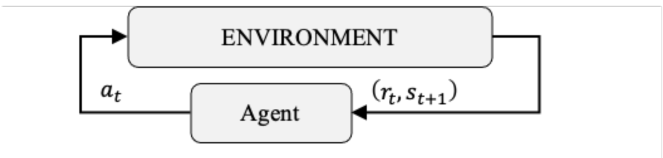


Figure 5.1: Agent-Environment interactions in RL

RL, as defined in [96] is simultaneously a problem, a class of solution methods for that class of problems, and the field that studies these problems and their solution methods. In particular, RL is a machine learning approach, focused on a goal-direct learning approach based on interactions with the environment. The main difference between RL and other machine learning methods is that the learner is not told which actions to take but instead must discover which actions yield the most reward by trying them out directly in the target environment. The objective of the RL is how to learn the best mapping between situations and actions, to maximize a reward signal. RL can be considered a close-loop problem because the actions of the learning agent influence its later inputs. To be able to perform the RL method as presented above, the learning agent must be able to sense the state of the environment and must be able to take actions that affect the state. The agent also must have a goal or goals relating to the state of the environment. The formulation is intended to include just these three aspects *sensation, action, and goal* in their simplest possible forms.

The RL problem is typically formulated as a MDP, namely a RL task that satisfies the Markov property [96]. The Markov property is respected if the environment's response at $t + 1$ depends only on the state and action representations at t:

$$s_{t+1} = \delta(s_t, a_t); \quad r_t = r(s_t, a_t) \tag{5.1}$$

An MDP can be modeled as a tuple:

$$MDP = (S, A, \delta, r) \tag{5.2}$$

where

- S is a finite set of states;
- A is a finite set of actions;
- $\delta = P(S \times A \times S)$ is a probability distribution over transitions, can be expressed as $P(s'|s, a)$ that is the conditional probability of the successor state, given the current state and the current action;
- $r : S \times A \times S \rightarrow \mathcal{R}$ is a reward function.

Given an MDP the goal is to find the *optimal policy.* Where the policy is a function $\pi : S \times A$, and the optimality is defined concerning maximizing the expected value of cumulative discounted reward:

$$V^{\pi}(s_1) = E[\bar{r}_1 + \gamma\bar{r}_2 + \gamma^2\bar{r}_3 + \dots] \tag{5.3}$$

where $\bar{r}_t = r(s_t, a_t, s_{t+1})$, $a_t = \pi(s_t)$, and $\gamma = [0, 1]$ is the discount factor for future rewards. The optimal policy can be defined as:

$$\pi^* = \arg\max_{\pi} V^{\pi}(s), \qquad \forall s \in S \tag{5.4}$$

The RL modeled using MDP can be optimally solved by using several different techniques, the most used method is dynamic programming. Dynamic programming methods, such as the Bellman Equation, can be used only in the case of perfect knowledge of the environment. This means that the agent knows the δ (transition probability functions) and the r (the rewards) for each state and then it achieves the best possible cumulative rewards. When these models are known the RL can be used to solve planning and reasoning problems or more in general for decision-making in a known environment.

However, in several applications, such as the control of communication networks treated in this thesis, the environment in which the agent plays is completely unknown by the agent itself. In this case, the RL can be used to solve learning problems, with the aim to use the learned model for decision-making in unknown environment.

This is the case the agent shall perform actions by *exploring* the environment. To solve a RL, if the agent has no knowledge of the environment, the dynamic programming techniques are not suitable, tush Temporal-Difference (TD) methods, such as Q-Learniing or SARSA [96], can be used. The Q-Learning [97] algorithm allows agents to act optimally with Markovian model, learning by the consequences of action without having knowledge about the transition and reward functions, representing a fundamental algorithm for *model-free* RL solution.

Q-Learning aims to estimate the optimal value for each policy, using the agent experience during the learning process to improve the estimate. Once the agent has acquired a perfect knowledge of the transition and immediate reward function it can provide the optimal policy, and it can calculate the optimal action for each state.

with $\epsilon \in (0, 1)$, used to adopt an $\epsilon - greedystrategy$ for the exploration of all the states and the exploitation of the learned model, typically ϵ can decrease over time (i.e., first exploration, then exploitation). Furthermore, γ is the discount factor and α is the learning rate.

The fundamental requirement for Q-Learning convergence to an optimal solution is that the agent must visit each state often enough [96, 97]. This result implies that (near to) optimal solutions

Algorithm 2: The Q-Learning Algorithm

Result: $\widetilde{Q}(s,a)$
$\forall s \in S$, $\forall a \in A$, initialize table entry $\widetilde{Q}(s,a) \leftarrow 0$;
observe current state s_t;
while *true* **do**
  select a_t random action with probability ϵ (exploration);
  select the best action (*i.e.*, $a_t = \arg\max_{a'} \widetilde{Q}(s_t, a')$) with probability $1-\epsilon$ (exploitation);
  execute a_t;
  receive immediate reward r_t;
  observe the new state s_{t+1};
  update the table entry of $\widetilde{Q}(s_t, a_t)$ as;
  $\widetilde{Q}(s_t, a_t) \leftarrow \widetilde{Q}(s_t, a_t) + \alpha_t[r_t + \gamma max_{a'} \widetilde{Q}(s_{t+1}, a') - \widetilde{Q}(s_t, a_t)]$
end

can be found in the case of a MDP with a limited number of states and actions, but the problem can be intractable when states and actions number increase (or are continuous set). This because the sequence of visits of each state, permits to the agent to achieve the suitable experience only if each visit is performed in a distinct episode. This requirement makes important the modeling of agent, environment, and their interactions, since the state and actions set can be reduced if the problem is modeled as needed.

5.1.2 Problem Formulation

The load balancing problem considered in this section concerns the dynamic allocation of connections to the available APs considering QoS constraints. With connection, it is meant a mono-directional flow of packets characterized by the same origin and the same destination. Each connection is assumed to have a specific QoS profile (inherited by the packets of said connection) consisting of a set of constraints regarding, for example, the throughput, latency, BER, jitter, mobility and so on.

To address this problem, a hierarchical control architecture has been envisaged (see Figure 5.2). The underlying idea is to associate a Local Controller to each AP and, consequently, to the cell covered by said AP. A set of local controllers willing to share their resources, referred to as Local Controller Sharing Set, are then coordinated by a single Global Controller located in the C-RAN. The cells covered by the APs associated to the local controllers in the Local Controller Sharing Set define the Global Controller Area. From now on, the symbol i will be used to refer to a generic AP in a given Local Controller Sharing Set and I will denote the number of Local Controllers coordinated by a given Global Controller. Concerning QoS profiles, it is assumed that in a pre-operational phase the most suitable number of profiles P has been identified. Such operation can be performed, for instance, using k-means algorithms exploiting available historical data sets. The number of QoS profiles should be identified trading off QoS personalization (increasing the number of profiles) and complexity (reducing the number of profiles). Profiling is expected to provide a set of classification rules allowing, whenever a new connection c is triggered, to classify this connection as belonging to a given profile p. A fundamental QoS constraint characterizing a given profile p relates to the minimum bit rate Y_p which, in any traffic condition, must be assigned to any connection belonging to the profile p.

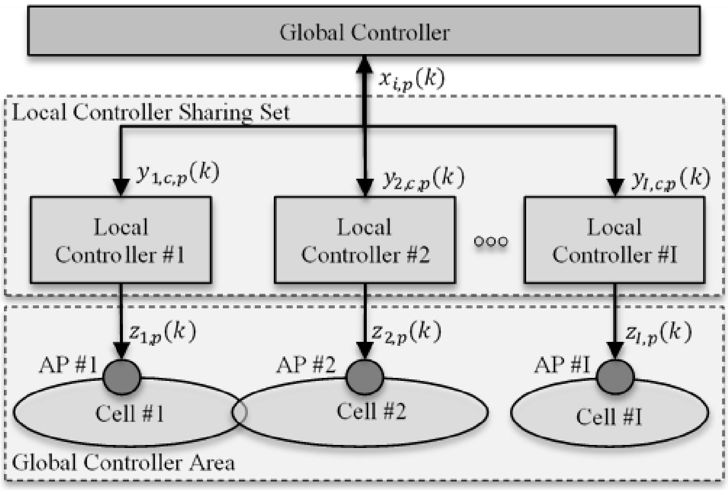


Figure 5.2: Control Architecture

At each discrete time k, the following variables are of interest for the considered problem:

- $A_{p,c}(k)$ denoting the set of APs whose cells cover a given UE involved in a connection c belonging to profile p; note that it is assumed that the Global Controller is aware of (i) the profile of each in-progress connection in the Global Controller Area and (ii) the sets of APs covering each connection c of any profile p (i.e., the Global Controller has a complete view of the sets $A_{p,c}(k)$ for all c and p);

- $x_{i,p}(k)$ representing the target bit rate assignment computed by the $i-th$ local controller for connections belonging to profile p (i.e., represents the sum of bit rates that should be assigned to connections of profile p served at time k by the $i-th$ AP); such target value, transmitted to the Global Controller, is computed independently by each local controller based on the performances experienced by all the in-progress connections served by the $i-th$ AP;

- $y_{i,p,c}(k)$ representing the actual bit rate assigned from the Global Controller to the $i-th$ local controller for connections belonging to profile p; note that, at any time, the actual bit rate $y_{i,p,c}(k)$ assigned by the Global Controller to the $i-th$ Local Controller for connection c of profile p must be higher of Y_p;

- $z_{i,p}(k)$ denoting the overall bit rate assigned from the Global Controller to the $i-th$ local controller for all connections belonging to profile p.

Note that the goal of the Global Controller is to compute the actual bit rates to be assigned to the local controllers in such a way that they approach, as far as possible, the target bit rates computed by the local controllers (i.e., ideally it should be $z_{i,p}(k) \cong x_{i,p}(k)$ since the latter tries to optimize the i-th AP's capacity utilization from a local perspective).

Local Controllers

The proposed control strategy demands part of the computation to the Local Controllers. As anticipated, each local controller solves, independently, a RL problem for addressing the load balancing problem and transmit to the Global Controller the target bit rate assignment $x_{i,p}(k)$. In the following, the RL problem that each local controller has to solve will be formalized. For sake of clarity, only the uplink (i.e., the flow of packets from the UE to the AP) will be considered; the extension to downlinks is straightforward.

State space modelling The state of a generic AP is defined as the measured traffic level for each profile p. Said traffic level is assumed to be described by L discrete levels. By doing so, the state of each local controller $s_i(k) \in \mathbb{R}^P$ can be described as a row vector:

$$s_i(k) = \begin{bmatrix} l_{i,1} & \dots & l_{i,P} \end{bmatrix} \in \mathbb{R}^P \tag{5.5}$$

where the generic scalar entry $l_{i,p}$ of the state represents the traffic level experienced by the generic AP for connections of profile p. Note that the state space, as it has been defined, guarantees high flexibility since it is possible to trade-off state's description capabilities and computational costs by increasing or decreasing the number of discrete levels L, respectively. Indeed, with this modelling choices, the number of possible states is equal to L^P which is a relatively small number. As a final remark, note that it is possible to generalize the proposed formulation by considering different discrete traffic levels for each profile (in this case, of course, one should define the number of discrete levels L_p for each profile p).

Action space modelling From the problem modeling, it follows that the control variables are the target bit rate assignments $x_{i,p}(k)$ computed by the local controllers. In order to induce a smooth convergence between the target and actual bit rate assignments, i.e., $x_{i,p}(k)$ and $z_{i,p}(k)$, respectively, and reduce the dimension of the local controllers' action spaces, instead of directly considering the target bit rate assignments $x_{i,p}(k)$ it is possible to consider as control actions $a_i(k) \in \mathbb{R}^P$ the following row vectors:

$$a_i(k) = \begin{bmatrix} \lambda_{i,1} & \dots & \lambda_{i,P} \end{bmatrix} \in \mathbb{R}^P \tag{5.6}$$

where the generic scalar entry $\lambda_{i,p}$ represents the target variation, with respect to the previous discrete time instant, of the bit rates assigned from the $i-th$ local controller to connections of profile p; furthermore, such variations are limited to a small number of discrete levels Λ. With this modelling choices, the total number of possible states is Λ^P which is a relatively small number. Note that it is possible to generalize the proposed formulation by considering different discrete traffic levels for each profile (in this case, of course, one should define the number of discrete levels Λ_p for each profile p).

Rewards shaping The objective of the control strategy proposed consists in (i) keeping each cell, as far as possible, far from congestion for any profile p and (ii) assuring that the cell's capacity is exploited. Following on these considerations, it is possible to consider rewards depending on the state and on the profile as follows:

$$r_{i,p}(k) = b_{i,p}(s_p(i)) \tag{5.7}$$

where $s_p(i)$ is the generic entry of the $i-th$ local controller state (see 5.5) and $b_{i,p}(\bullet)$ are functions of the state shaped in such a way to provide hard penalizations to congested states and mild penalization to idle states.

Global Controller

The control problem, from the Global Controller point of view, consists in providing to the local controllers the actual bit rate assignments $y_{i,p,c}(k)$ (i.e., for each local controller i and each (c,p) couple connection-profile) based on the received target variations of bit rate assignments $\lambda_{i,p}$ (see equation 5.6). Hence, the Global Controller must minimize the following performance index $J(k)$

$$J(k) = \sum_{i=1}^{I}\sum_{p=1}^{P}\sum_{c=1}^{C_p(k)} \left(y_{i,p,c}(k) - x_{i,p}(k)\right)^2 = \sum_{i=1}^{I}\sum_{p=1}^{P} \left(z_{i,p}(k) - x_{i,p}(k)\right)^2 \tag{5.8}$$

while guaranteeing, for each local controller i and each connection-profile couple, that

$$\sum_{i \in A_{p,c}(k)} y_{i,p,c}(k) \geq Y_p \tag{5.9}$$

$$y_{i,p,c}(k) = 0 \text{ if } i \notin A_{p,c}(k) \tag{5.10}$$

$$y_{i,p,c}(k) \geq 0 \tag{5.11}$$

where

- 5.8 is the performance index to be minimized; note that said index is lower when the target and actual bit rate assignments are closer;
- 5.9 allows to satisfy QoS constraints in terms of the minimum bit rate required by each profile p (the structure of equation 5.9 can be replicated to take into account additional QoS constraints);
- 5.10 specifies that, if the $i-th$ AP does not cover the connection c of profile p at time k, since the Global Controller cannot assign a bit rate to said connection-profile couple;
- 5.11 specifies that the actual bit rate assignments cannot be negative.

The actual bit rate assignments $y_{i,p,c}(k)$, computed by the Global Controller as output of the optimization problem 5.8-5.11, are transmitted at each discrete time k to the UEs through the serving APs. Said UEs can transmit toward the APs at a bit rate $T_{i,p,c}(k)$ which is not higher than $y_{i,p,c}(k)$.

5.1.3 Simulation and Results

The considered case study concerns load balancing and mobility management. More in detail, the problem consists in associating APs and moving UEs taking into account spatial considerations. Hence, the goal is to allocate enough bandwidth to theUEs and to exploit as much as possible APs capacities (avoiding congestion) while considering the distance between given APs and UEs. In other words, a given AP should assign less bandwidth to UEs whose distance is higher than a fixed threshold. This condition allows to reduce the power consumption required for transmission and, at the same time, to minimize the probability that an AP assigns bandwidth to a UE which is likely to exit from its coverage area. To simultaneously tackle the load balancing and mobility management problems, the local controllers' state defined in 5.1.2 is augmented and has two components:

- the first component, $s_{i,1}$, represents the distance between the $i-th$ AP and the UEs;
- the second component, $s_{i,2}$, represent the traffic level as defined in equation 5.5.

The considered scenario envisages the presence of two UEs exploiting the resources provided by four cells (i.e., four APs and, consequently, four local controllers). The two moving UEs are assumed to belong to two different QoS profiles (hence, the number of UE profiles considered by the Global Controller is $P = 2$). In other words, it is assumed that the two moving UEs have different QoS requirements due to their different motion characteristics. In this case, said requirements can be referred to as QoS mobility profiles since the UEs' motion degrades several QoS indicators such as error rate, energy consumption and service continuity. Note that this scenario can be generalized by considering, instead of single UEs, clusters of moving UEs involving several connections. Indeed, nowadays, vehicles are equipped with a wide set of sensors for guidance support, multimedia systems and, also, passengers' terminals. The local controllers can consider all the connections in the same vehicle (cluster) as belonging to the same QoS mobility profile p.

As depicted in Figure 5.3, it is assumed that the two moving UEs (represented as a white and a black car) change their coverage area during the simulations. More in detailed, for $k = 1, \ldots, 200$ the black car is covered by the cells associated to APs #1 and #3 while the white car is covered by the cells associated to APs #2 and #3 and, for $k > 200$, vice versa.

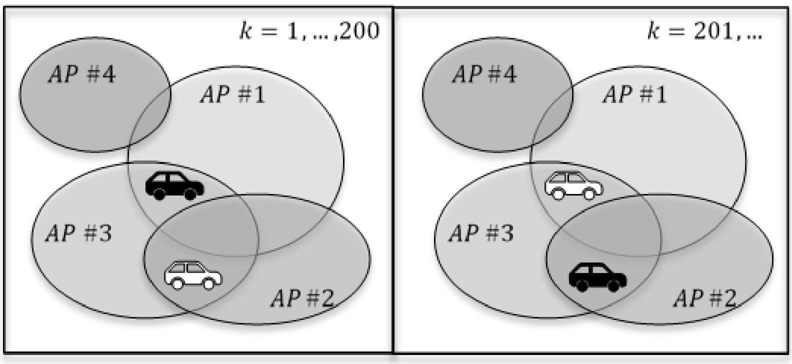


Figure 5.3: APs' coverage and UEs' position (left: $k \leq 200$; right: $k > 200$)

Concerning the local controllers' state definition, let

- ${\rho^{\text{distance}}_{\text{high}}}_{i,p}$ and ${\rho^{\text{distance}}_{\text{low}}}_{i,p}$ be two fixed thresholds specifying the maximum distance, between the $i-th$ AP and given UE;
- ${\rho^{\text{traffic}}_{\text{high}}}_{i,p}$ and ${\rho^{\text{traffic}}_{\text{low}}}_{i,p}$ be two fixed thresholds specifying the traffic level above which the $i-th$ local controller is considered overloaded with respect to connections of profile p and vice versa, respectively.

Said thresholds allow to define a limited set of discrete levels for characterizing the local controllers' congestion level (i.e., used to solve the load balancing problem) and the convenience for a given local controller to serve a generic connection-profile couple (i.e., used to perform mobility management). On the ground of these considerations, it is possible to define the generic entries of the two components of the augmented local controllers state ($d^{(j)}_{i,p}(k)$ and $l_{i,p}(k)$ respectively) as

$$d^{(j)}_{i,p}(k) = \begin{cases} 0 & \text{if } \delta^{(j)}_{i,p}(k) < \rho^{\text{distance}}_{\text{low}\ \ i,p} \\ 1 & \text{if } \rho^{\text{distance}}_{\text{low}\ \ i,p} \le \delta^{(j)}_{i,p}(k) \le \rho^{\text{distance}}_{\text{high}\ \ i,p} \\ 2 & \text{if } \delta^{(j)}_{i,p} > \rho^{\text{distance}}_{\text{high}\ \ i,p} \end{cases} \tag{5.12}$$

$$l_{i,p}(k) = \begin{cases} 0 & \text{if } t_{i,p}(k) < \rho^{\text{traffic}}_{\text{low}\ \ i,p} \\ 1 & \text{if } \rho^{\text{traffic}}_{\text{low}\ \ i,p} \le t_{i,p}(k) \le \rho^{\text{traffic}}_{\text{high}\ \ i,p} \\ 2 & \text{if } t_{i,p} > \rho^{\text{traffic}}_{\text{high}\ \ i,p} \end{cases} \tag{5.13}$$

where $d^{(j)}_{i,p}(k)$ and $l_{i,p}(k)$ represent the $i-th$ local controller state with respect to the mobility management and load balancing problems, respectively, and $\delta^{(j)}_{i,p}(k)$ and $t_{i,p}(k)$ are the relative distance between the $i-th$ local controller's AP and the $j-th$ UE and the traffic level experienced by the $i-th$ local controller with respect to connections of profile p, respectively. The rewards defined in equation 5.7 can be particularized for the considered scenario for taking into account the augmented state defined in equations 5.12-5.13 as follows:

$$r_{i,p}(k) = \sum_{j=1}^{J} \frac{K_1}{1 + e^{\alpha_j * d^{(j)}_{i,p}(k)}} \ * \ t_{i,p} + \frac{K_2}{1 + e^{-\beta * l_{i,p}(k)}} \tag{5.14}$$

where K_1 and K_2 are two constants used to weight the two state's components and α_j and β are positive constants. Rewards have been shaped using sigmoid functions since this function, by properly setting α and β, is able to represent the effect of the distance and traffic level on the system performances. This is possible thanks to the sigmoid structure which allows the definition of three input variable intervals (characterized by α and β) returning low, medium, or high values of the output (i.e., the reward value). Concerning the action space, it is assumed that the $i-th$ local controller can either (i) increment the bandwidth allocated to connections of profile π of a positive constant Δ, (ii) do not vary the allocated bandwidth or (iii) decrements the allocated bandwidth of $-\Delta$. It follows that the dimension of the action space is $\Lambda^P = 9$ since there are $P = 2$ QoS (mobility) profiles and $\Lambda = 3$ discrete levels of allocated bandwidth variations (i.e., $+\Delta$,0,$-\Delta$).

In the simulations, the proposed RL-based hierarchical control strategy is compared with a Nearest Not-Full (N-NF) controller allocating bandwidth to the UEs through the nearest, not congested, AP.

Simulations show that the proposed hierarchical control strategy is able to effectively tackle both the load balancing and mobility management problems. More in detail, concerning load balancing, from 5.4 it is clear that the proposed RL-based approach guarantees a fair load distribution between APs while the N-NF approach under/overloads them. Furthermore, the proposed RL approach is able to keep APs' loads in the optimal range defined by the upper and lower thresholds (i.e., $\rho^{\text{traffic}}_{\text{low}\ \ i,p}$ and $\rho^{\text{traffic}}_{\text{high}}$) which in the figure are represented by the green and red lines, respectively. Concerning mobility management, when the two UEs change their coverage area (i.e., at $k = 200$), the N-NF approach experiences drastic changes in the allocated bandwidth (which translates in performing heavy handover procedures that degrades the system performances) while the proposed RL approach guarantees a smoother transition (see 5.4). Said transition stops when the loads reach a new equilibrium in terms of loads and of the mutual distances between APs and UEs. 5.5 shows the bandwidth allocation for the two UEs computed by the proposed RL and N-NF approaches. As it can be seen, both approaches vary the bandwidth allocation considering the relative position between the APs and UEs. However, the proposed RL approach is able to exploit all the available resources and to avoid congestion while respecting the QoS (mobility) constraints.

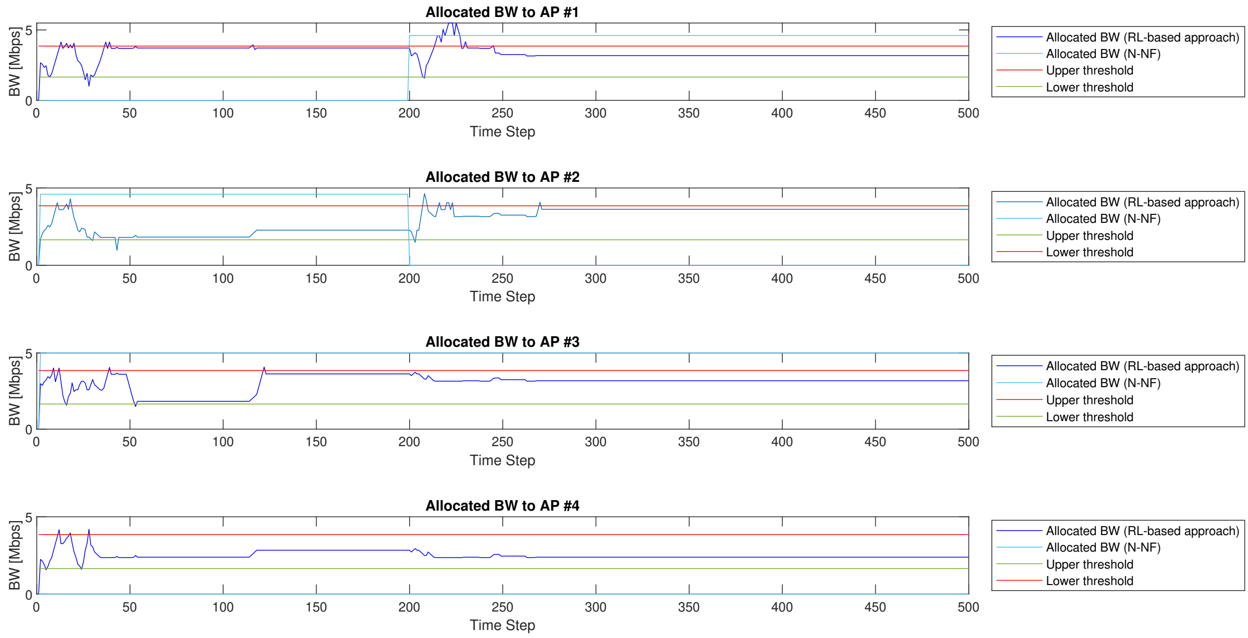


Figure 5.4: APs allocated bandwidth

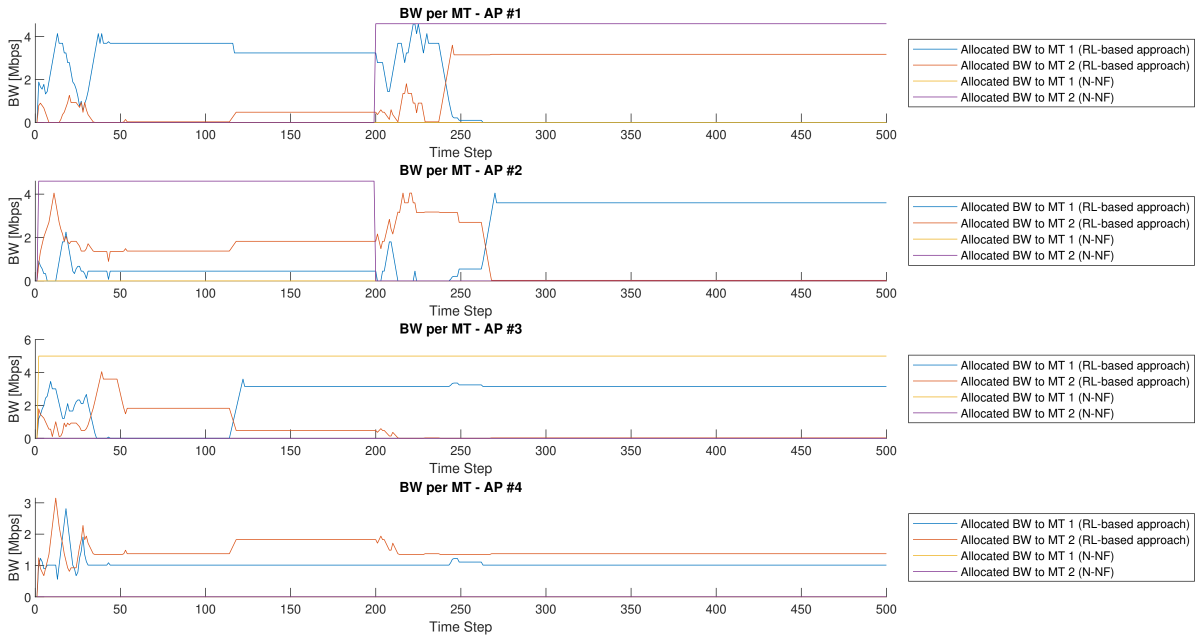


Figure 5.5: UEs allocated bandwidth

5.2 A Distributed Reinforcement Learning approach for Power Control in Wireless Networks

In this section, a fully distributed solution for the power control problem in wireless networks is presented. Power control [98–102] is typically referred to as the problem of finding the best strategy required to adjust, correct and manage in an efficient way the transmission power between APs and UEs (i.e. downlink for transmission from AP to UE, uplink for transmission from UE to AP). In the following, the model of the system and the formulation of its solution as an optimization model are provided. Typically, the power control solutions, such as Distance-Based Power Allocation, Distributed Balancing Power Control, Multiple-step SIR based power control with fixed step size, Adaptive Step Power Control, are based on knowledge about the network topology, the UEs positions and in some case about the transmission power of each network entity. Sometime they are based on a target SINR, that can be over or underestimated in a dynamic environment. In the following, starting from a monolithic optimization model, a model-free and robust distributed solution is formulated using RL [96, 103, 104] and min-consensus techniques [105, 106]. The proposed approach provides a solution to the above-mentioned problem that is scalable and low complexity. Moreover, the proposed solution can be deployed in all the possible network architectures, even if the processing capacity at the edge (i.e., at the APs) is very low and if there is a low communication capacity between the APs and between APs and the CN.

5.2.1 Multi-Agent Reinforcement Learning

In section 5.1.1 single-agent RL is presented. Most of the practical application of RL, such as CPS [107], finance [108], sensor and communication networks [109], and social science [110], can't be modelled as a single-agent RL introducing the needs of Multi-Agent RL (MARL). Specifically, [103, 104] MARL addresses the sequential decision-making problem of multiple autonomous agents that operate in a common environment, each of which aims to optimize its own long-term return by interacting with the environment and other agents using classical RL techniques (e.g., Q-Learning described in Algorithm 2). The reference schema is depicted in figure 5.6.

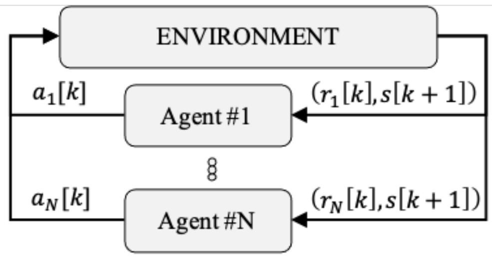


Figure 5.6: Agent-Environment interactions in RL

In MARL the involved agents can operate as a single agent following the description in 5.1.1, however they can perform their decisions in three different settings:

- *cooperative setting*, all agents usually share a common reward function, i.e., $r_1 = r_2 = \ldots = r_N = r$. This model is referred to as multi-agent MDP (MMDP) in the machine learning and artificial intelligence community [111], and Markov teams or team Markov games in the control or game theory community [112, 113]. In this setting, the value function and Q-function are identical to all agents, which thus enables the single-agent RL algorithms, e.g., Q-learning (see Algorithm 2), to be applied. The global optimum for cooperation is a Nash equilibrium of the game. Besides the common-reward model, another slightly more general and surging model for cooperative MARL considers team-average reward [114].
- *competitive setting*, in MARL is typically modeled as zero-sum Markov games, i.e. all the sum of the reward of all the agents sum to zero ($\sum_{i \in N} r_i(s, a, s') = 0 \ \ \forall \ (s, a, s') \in S \times A \times S$). For ease of algorithm design and analysis, most literature focused on two agents that compete against each other [115], where clearly the reward of one agent is exactly the loss of the other.
- *mixed setting*, it is also known as the general-sum game setting, where no restriction is imposed on the goal and relationship among agents [116]. Each agent is self-interested, whose reward may be conflicting with others'.

The other interesting feature of MARL the way the agents interact each other. Three different reference structures are:

- *Centralized structure*, there exists a central controller that can aggregate information from the agents and even design policies for all agents. The information exchanged between the central controller and the agents can thus include both some private observations from the agents, and the local policies designed for each agent from the controller. This solution implies scalability limitation and low reliability since the central controller can be a single point of failure, hoever this solution allow global optimality and doesn't need a communication network between agents;

- *Decentralized with networked agents structure*, agents are connected via a possibly time-varying communication network, so that the local information can spread across the network, by information exchange with only each agent's neighbors. This solution is the most reliable and scalable, but the convergence time can be higher depending by the network topology;
- *Fully decentralized structure*, the agents are full decentralized, with no explicit information exchange with each other. Instead, each agent makes decisions based on its local observations, without any coordination and/or aggregation of data.

5.2.2 Min-Consensus

Recently consensus algorithms have attracted renewed attention because they can be exploited for distributed cooperative control in a wide range of applications such as energy network, communication network, autonomous vehicles and social networks. The main advantage of these algorithms is the capability of information exchange via local interactions [105], this because in many applications, since the large-scale deployment and the communication channels constraints, the capability of exchange information between each couple of agents in the network is very challenging. The networked system is than characterized by the network topology that allows the local information exchange, but also by the agents' dynamic input-output characteristics and by the interaction protocols. Local interaction protocols are [105]:

- Consensus, having agents come to a global agreement on a state value;
- Formations: making the agents move to a desired geometric shape;
- Assignments: deciding a fair assignment of tasks among multiple agents;
- Coverage: producing maximally spread networks without making them disconnected or exhibit "holes" in their coverage;
- Flocking/Swarming: making the agents exhibit behaviors observed in nature, such as flocking birds, schooling fish, or swarming social insects;
- Social Networks and Games: analyzing how the outcomes of games and social interactions are influenced by the underlying interaction topology; and
- Distributed Estimation: organizing a group of sensors to collectively estimate a random phenomena of interest.

In this work the interaction protocol is a particular type of consensus protocol, namely the *min-consensus* [106].
To analyze the min-consensus algorithm the following graph theoretic concepts are introduced. The topology of the network is represented by a *communication graph* denoted as $\mathcal{D} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of agents (vertices in the graph) and $\mathcal{E}$ is the set of communication paths between the agents (set of directed edges in the graph): if $(i, j) \in \mathcal{E}$, then the agent i is able to transmit information to the agent j. Furthermore, is possible to define the state graph, namely the graph that comprises only the state variables as vertices (i.e., $\mathcal{V} = \{x_1, ..., x_n\}$ as state vertices) and a set of directed edges between the state vertices.

The agent's local knowledge abut the network is defined by $\mathcal{N}_i^-$, that is the set of in-neighbors of agent i, defined as the set of agents $j \neq i$ such that $(j,i) \in \mathcal{E}$ for all i.
A digraph $\mathcal{D}$ is said to be strongly connected if there exists a directed path between any two vertices.

Given a communication graph $\mathcal{D}(\mathcal{V}, \mathcal{E})$, the min-consensus algorithm consists in the following update rule for each agent $i \in \mathcal{V}$:

$$x_i(k+1) = \min_{j \in \bar{\mathcal{N}}_i^-} x_j(k) \quad \forall i \quad \in \quad 1, \dots, n \tag{5.15}$$

where $\bar{\mathcal{N}}_i^- = \mathcal{N}_i^- \cup \{i\}$. Min-consensus is achieved [106], if there exist an instant of time k for which $x_i(\bar{k}) = x_j(\bar{k}) \;\; \forall i,j \in \{1,\dots,n\}$ if $\bar{k} \leq k$ and $\forall$ $\mathbf{x}_0 = [x_1(0), \dots, x_n(0)]^T$. In particular, the value of k suffices to be equal to the length of the longest shortest path between any pair of agents in the digraph:
Lemma: [106] If $\mathcal{D}$ is strongly connected, then min-consensus is achieved.

5.2.3 Problem Formulation

In the following the downlink power control problem is considered, the same considerations hold for the uplink scenario. In wide-band systems, when Code Division Multiple Access (CDMA) techniques are considered, the DL-interference management, performed by an AP, becomes the problem of allocating different powers to different UE as a function of the amount of interference coming from the transmissions of the near APs [117].
The interference effecting an UE can be measured by means of the SINR [118]. The SINR can be defined as the ratio between the power of the useful signal over the power of the unwanted signals plus the noise. Equation (5.16) represents the SINR value at UE j, when the transmitting AP i is located at x_i and is transmitting with power P_i. Then, the SINR at a the receiving UE, located at x_j is given by:

$$SINR_{i,j} = \frac{P_i L(x_i, x_j)}{\sum_{k \in \mathbb{S}_{AP}, k \neq i} P_k L(x_k, x_j) + W N_0} \tag{5.16}$$

where $L(x,y)$ is the path loss function when the transmitter is at x and the receiver is at y, N_0 is the thermal noise spectral density at the receiver, W is the channel bandwidth, and $\mathbb{S}_{AP}$ is the set of neighbor APs.

SINR is a good interference indicator since it can be used to define the effect of the interference on the communication, that is able to reduce the amount of information that can be transmitted. By applying the Shannon capacity theorem the total channel capacity is:

$$C_{i,j} = W \log(1 + SINR_{i,j}) \tag{5.17}$$

where $C_{i,j}$ is the channel capacity (or maximum rate of data) in bits per second, W is the channel bandwidth in Herz, and $SINR_{i,j}$ is the signal-to-interference-plus-noise ratio as defined in equation (5.16). What equation (5.17) says is that higher the SINR and more the channel bandwidth, the higher the possible data rate. Indeed, the power control problem, aims at maximizing the SINR at each UE, this is a problem that needs coordination between APs and knowledge of the path loss (i.e., the positions) between each AP-UE couple to reach the optimality. However, this is not easy in real applications, due to UEs mobility, APs synchronization and communication constraints and many other practical reasons. In real applications, typically a lower bound of the SINR is provided to the APs for each UE, this is based on the QoS level to be satisfied for the connection. Indeed, this value can be under or overestimated in case of particular scenario, causing underutilization of the resources or unfesibility of the problem. The proposed approach tries to avoid that problem by maximizing the lower SINR in the network, with the aim to exploit the network's resources with the maximum efficiencies. The mathematical model of the proposed problem can be formalized as a *minimax* optimization problem:

$$\mathbf{P}(k) = \arg\max_{\mathbf{P}\in\mathbb{R}^{N\times M}} \big(\min_{i\in\mathbb{S}_{AP}, j\in\mathbb{S}_{UE}} SINR_{i,j} \big) \tag{5.18}$$

s.t.

$$SINR_{i,j} = \frac{P_{i,j}(k)L(x_i,x_j)}{\sum_{k\in\mathbb{S}_{AP},k\neq i} P_k(k)L(x_k,x_j) + WN_0} \quad \forall i\in\mathbb{S}_{AP}, \quad \forall j\in\mathbb{S}_{UE} \tag{5.19}$$

$$P^i_{min} \leq P_{i,j}(k) \leq P^i_{max} \quad \forall i\in\mathbb{S}_{AP}, \quad \forall j\in\mathbb{S}_{UE} \tag{5.20}$$

Where $\mathbf{P} \in \mathbb{R}^{N\times M}$, with N and M respectively number of APs and UEs in the network. It is the control vector, containing the transmission power of each AP towards each UE. $\mathbb{S}_{AP}$ and $\mathbb{S}_{UE}$ are respectively the set of APs and the set of the UEs in the network. P^i_{min} and P^i_{max} is the minimum and maximum power of AP i for a single connection, and $SINR_{i,j}$ is the SINR at UE j connected to AP i as defined above.
However, problem (5.18)-(5.20) needs information about all the networks entities and the channel characteristics between the entities to be solved, moreover if the solver is implemented in a centralized fashion it is not scalable (i.e., an AP or a UE can't be added to the system without changes in the algorithm) and require a lot of computation power in large scale networks (i.e., the case of modern urban networks). These limitations make the use of this solution unusable in the target applications, such as IoT scenario where a hughe amount of sensors are continuously connected or disconnected following the battery saving policies, or in dense urban scenario where the UEs mobility and the channel characteristics has high time varying behavior. Tush, an approximate solution of this minimax problem can be found in distributed fashion, introducing scalability, only local information knowledge need and applicability to large networks by using a particular MARL (see section 5.2.1) setting and using min-consensus multi-agent technique (see section 5.2.2). The problem formulation will be subject of the following hypothesis:

- each AP transmits to one UE (for notation simplification), reasonable since it can be considered the UE with lower SINR in the AP coverage;
- x_i is the location of the UE that is connected to AP i;
- the SINR is measured by the UE, that send the measured to the connected AP, as specified in the 3GPP standard;
- each AP can know the minimum SINR in in the network (i.e., for all the APs) with only local information, this is always true imposing that the APs are strongly connected, since with this condition min-consensus converges in finite number of steps [105].

The solution can be implemented using MARL techniques, considering a fully cooperative setting with distributed networked agents (see section 5.2.1), with agents' state space, action space and reward of AP i are defined as follow: the main information about the state is relative to the UEs position in the coverage area.

$$\mathbf{s}_i(k) = [x_{i,1} \dots x_{i,M}] \in \mathbb{R}^M \tag{5.21}$$

where a generic $x_{i,m}$ is the distance between the AP i and the UE m, furthermore, such distances are limited to a small number of discrete levels X. With this modelling choices, the total number of possible states is X^M which is a relatively small number. The information about the UEs' position is easy to retrieve by all the APs, because several methods are implemented in the real network to do it, e.g. using beamforming techniques or GPS.
From the problem formulation (5.18)-(5.20), it follows that the control variables are the transmission power of the APs towards the connected UEs. In order to reduce the dimension of the action spaces, instead of directly considering the transmission power $P_{i,j}$ it is possible to consider as control action $a_i(k) \in \mathbb{R}^M$ the following row vectors:

$$\mathbf{a}_i(k) = [\delta_{i,1} \dots \delta_{i,M}] \in \mathbb{R}^M \tag{5.22}$$

where the generic scalar entry $\delta_{i,m}$ represents the variation, either negative or positive, with respect to the previous discrete time instant, of the transmission power from the AP i to UE m. In this way the transmission power at time $t+1$ is $P_{i,m}(t+1) = max\{min\{P_{i,m}(t) + \delta_{i,m}, P^i_{min}\}, P^i_{max}\}$, to respect constraint 5.18 of the problem formulation; furthermore, such variations are limited to a small number of discrete levels Δ. With this modelling choices, the total number of possible states is Δ^M which is a relatively small number.
Finally, the reward function is common to all the agents, is equal to a function of the minimum SINR level between all the UEs, i.e. $SINR^* = \min_i SINR_i$.

$$r^i = f(SINR^*) \quad \forall i \in \mathbb{S}_{AP} \tag{5.23}$$

this value can be found by the networked agents (i.e. the APs) composing the network with only local knowledge, this make the solution cooperative and distributed, thus scalable since an AP can be added to the set of agent when active and removed if deactivated. In particular, each AP knows the SINR of each connected UE, and sharing this knowledge only with the neighbours APs, the minimum in all the network will be known by all the APs in a finite number of interaction accordingly to the min-consensus algorithm presented in section 5.2.2.

In particular, the reward function is defined as follow:

$$r^i = \frac{k}{1 + e^{\alpha SINR^*}} \tag{5.24}$$

where K is a constant used to weight the reward effect, and α is positive constant used to define the shape of the function. Rewards have been shaped using sigmoid functions since this function, by properly setting α, is able to represent the effect of the SINR level on the system performances. This is possible thanks to the sigmoid structure which allows the definition of three input variable intervals (characterized by α) returning low, medium, or high values of the output (i.e., the reward value).

5.2.4 Simulation and Results

In the following a comparison between proposed solution, i.e. minimax-RL, and Adaptive Step Power Control (ASPC) is presented. ASPC [100], uses adaptive step sizes to achieve faster convergence to the target SINR if compared with standard methods. In this method step size depends on the previous state, tush it has memory of one or more previous steps. The basic mechanism is: if one power-up/down command is required, then the step is δ; if two or more consecutive power-up commands are required, the step size is $\mu * \delta$; if two or more consecutive power-down commands are required, the step size is $\nu * \delta$. Furthermore, the algorithm proposed in [100] add a change in the value of μ and ν if alternated command sequences appear, such as up-down-up (or down-up-down). The detailed algorithm steps are reported in Algorithm 3.

Algorithm 3: Adaptive Step Power Control

Result: $P_{i,m}(k)$
UE_m measure the current SINR level $SINR_m(k)$;
$SINR_m(k)$ is transmitted to AP_i;
if $SINR_m(k) > TH_m$ **then**
  if $SINR_m(k-1) > TH_m$ **then**
    $P_{i,m}(k) = P_{i,m}(k-1) - \delta\mu$
  else
    $P_{i,m}(k) = P_{i,m}(k-1) - \delta$
  end
else
  if $SINR_m(k-1) < TH_m$ **then**
    $P_{i,m}(k) = P_{i,m}(k-1) + \delta\nu$
  else
    $P_{i,m}(k) = P_{i,m}(k-1) + \delta$
  end
end
if $(SINR_m(k) > TH_m \ \& \ SINR_m(k) < TH_m \ \& \ SINR_m(k) > TH_m)$ **then**
  $\mu = \frac{\mu}{\lambda_\mu}$
end
if $(SINR_m(k) < TH_m \ \& \ SINR_m(k) > TH_m \ \& \ SINR_m(k) < TH_m)$ **then**
  $\nu = \frac{\nu}{\lambda_\nu}$
end
return $P_{i,m}(k)$;

The comparison is done with three different SINR target levels, namely three different ASPC's thresholds, that considering the optimal target SINR are defined in order to be one underestimated, one overestimated and one optimal for each UE. The scenario is composed by 4 APs and 4 UEs. Each UE is connected to an AP, a smooth variation in the UEs' position is performed at step 200, i.e. each UE remains connected to the same AP but changing its position around it.

Figure 5.7 depicts the total network bandwidth used under the proposed approach and the three different implementations of the ASPC algorithm, referred in the figure as the threshold TH1 that is the underestimated threshold, the TH2 that is the overestimated threshold and TH3 that is the optimal threshold. As it can be seen, the proposed approach (purple dotted line) is able to achieve higher values of used bandwidth meaning that the networks' resources are efficiently exploited. This is due to the APs' cooperative behaviours allowing to reach a given SINR level without allocating higher power transmission where not necessary. Figure 5.8 shows the network's efficiency, defined as the ratio between the transmission power total network throughput and the total network transmission power. As it can be observed, the performances of the proposed approach are worse in the initial time steps because the time needed to converge to the optimal policy, that in the long run, outperforms the ASPC algorithms. Furthermore, the higher performances of the proposed approach can be observed when UEs' positions, at time step $k = 200$, change. Indeed, thanks to the AP's coordination envisaged by the proposed approach, the network is able react to such changes guaranteeing an efficient usage of networks' resources, without model-based bias.

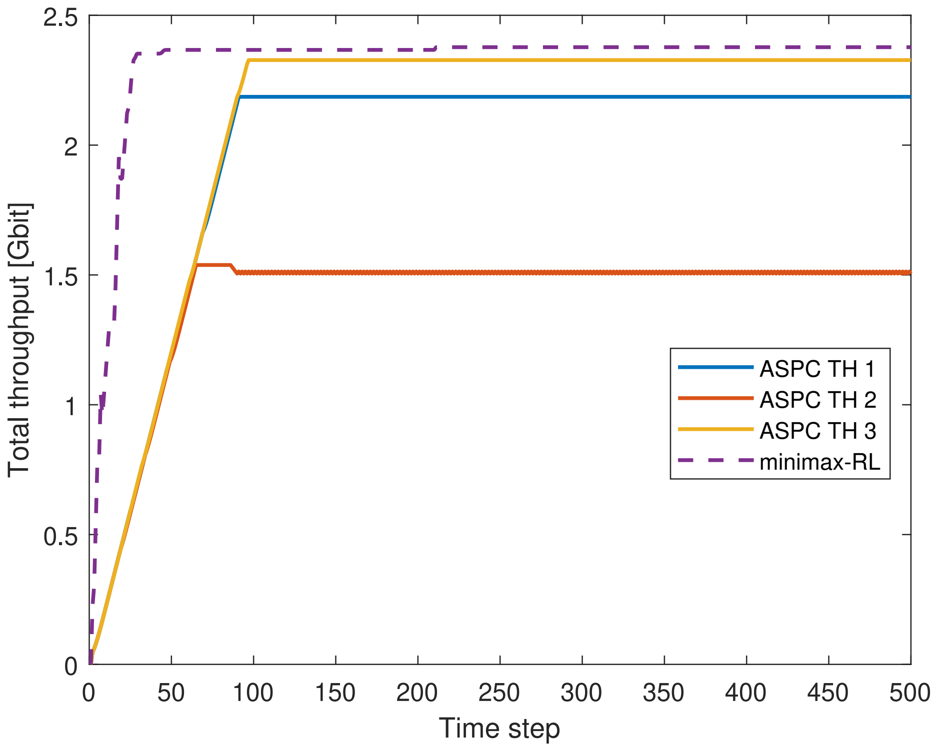


Figure 5.7: Total network's exploited bandwidth

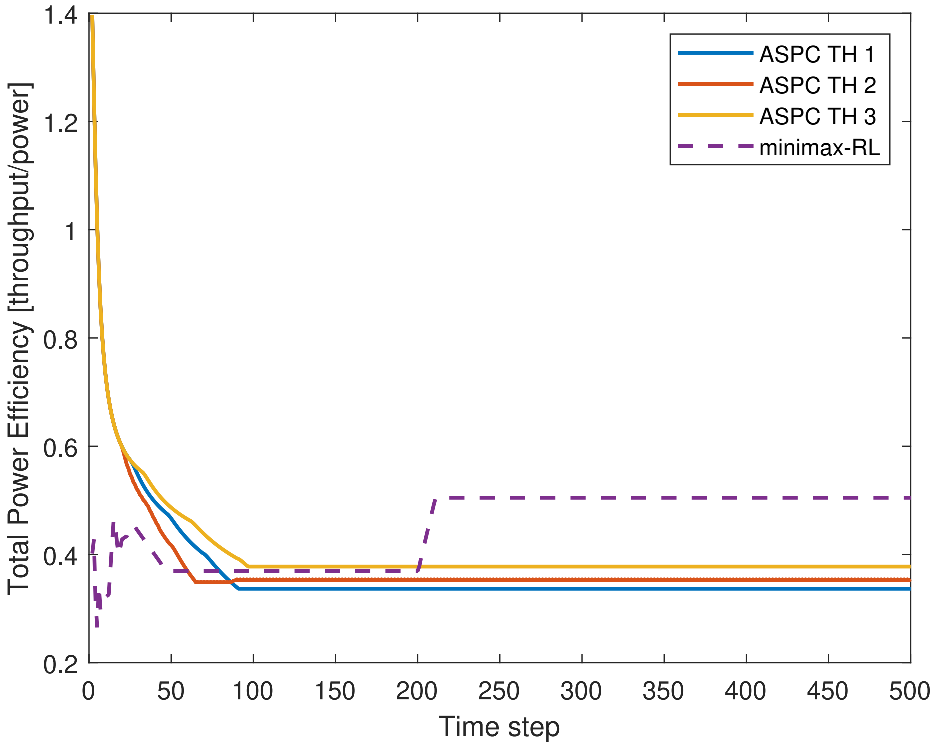


Figure 5.8: Total network's efficiency

Chapter 6

Conclusion

The research reported in this thesis is motivated by the needs of the CPSs to have an underlying communication network able to satisfy the requirements of the infrastructure operators, while respecting the characteristics of physical processes, sensors, actuators, and controllers composing such CPSs. The thesis is focused on a particular network designed with the scope to support the infrastructure operators' networking needs, this network is the fifth-generation of mobile networks, known as 5G network. In the thesis different control methods to manage the Multi-Connectivity resources in 5G network are provided, with the definition of a general architecture, and the design and validation of model-based and data-driven algorithms, as well as their centralized or distributed implementation. The developed algorithms provide the capability to control a network offering Multi-Connectivity placing the controller either in the cloud or at the edge of the network, based on the needs of each particular scenario. Furthermore, the proposed solutions are compliant with the 3GPP standard that regulates the 5G deployment, this implies that these solutions can be considered ready to be studied for implementation on the network.

Deeper simulations of the proposed algorithms have been left for the future. Future works concern the development of more accurate simulators, including real scenarios such as dense networks deployment, to study the behavior of the algorithms in large networks with high amounts of traffic and users. Furthermore, the effect of the algorithms' control actions on the service satisfaction can be studied, this implies the development of a simulator able to characterize the network traffic with needs of the services, such as delay-sensitive or high bandwidth services, and the development of a metric to evaluate these effect.

Bibliography

[1] Edward Ashford Lee and Sanjit A Seshia. *Introduction to embedded systems: A cyber-physical systems approach*. Mit Press, 2016.

[2] Walid M Taha, Abd-Elhamid M Taha, and Johan Thunberg. *Cyber-Physical Systems: A Model-Based Approach*. Springer Nature, 2020.

[3] S Amin, G Schwartz, S Shankar, T Samad, and A Annaswamy. Resilient Cyber-Physical Systems. *The Impact of Control Technology–2nd Ed.(IEEE CSS)*, 2014.

[4] Fei-Yue Wang and Derong Liu. Networked control systems. *Theory and Applications*, 2008.

[5] Alberto Bemporad, Maurice Heemels, Mikael Johansson, et al. *Networked control systems*, volume 406. Springer, 2010.

[6] Giacomo Como, Ketan Savla, Daron Acemoglu, Munther A Dahleh, and Emilio Frazzoli. Robust distributed routing in dynamical networks—Part i: Locally responsive policies and weak resilience. *IEEE Transactions on Automatic Control*, 58(2):317–332, 2012.

[7] Giacomo Como, Ketan Savla, Daron Acemoglu, Munther A Dahleh, and Emilio Frazzoli. Robust distributed routing in dynamical networks–part ii: Strong resilience, equilibrium selection and cascaded failures. *IEEE Transactions on Automatic Control*, 58(2):333–348, 2012.

[8] José R Correa, Andreas S Schulz, and Nicolás E Stier-Moses. Selfish routing in capacitated networks. *Mathematics of Operations Research*, 29(4):961–976, 2004.

[9] Andrea Baiocchi, Francesco Delli Priscoli, Francesco Grilli, and Fabrizio Sestini. The geometric dynamic channel allocation as a practical strategy in mobile networks with bursty user mobility. *IEEE Transactions on Vehicular Technology*, 44(1):14–23, 1995.

[10] Francesco Delli Priscoli and Antonio Pietrabissa. Design of a bandwidth-on-demand (BoD) protocol for satellite networks modelled as time-delay systems. *Automatica*, 40(5):729–741, 2004.

[11] Guido Oddi, Antonio Pietrabissa, Francesco Delli Priscoli, and Vincenzo Suraci. A decentralized load balancing algorithm for heterogeneous wireless access networks. In *WTC 2014; World Telecommunications Congress 2014*, pages 1–6. VDE, 2014.

[12] Alessandro Giuseppi and Antonio Pietrabissa. Wardrop equilibrium in discrete-time selfish routing with time-varying bounded delays. *IEEE Transactions on Automatic Control*, 2020.

[13] G Goodwin, M Cea, K Lau, T Wigren, T Samad, and A Annaswamy. Control Challenges in Mobile Communications. *The Impact of Control Technology–2nd Ed.(IEEE CSS)*, 2014.

[14] Eitan Altman and Tamer Basar. Multiuser rate-based flow control. *IEEE Transactions on Communications*, 46(7):940–949, 1998.

[15] Man-Tung T Hsiao and Aurel A Lazar. ptimal decentralized flow control of Markovian queueing networks with multiple controllers. *Performance evaluation*, 13(3):181–204, 1991.

[16] Christos G Cassandras, Tao Wang, and Sepideh Pourazarm. Optimal routing and energy allocation for lifetime maximization of wireless sensor networks with nonideal batteries. *IEEE Transactions on Control of Network Systems*, 1(1):86–98, 2014.

[17] Ryan W Thomas, Daniel H Friend, Luiz A DaSilva, and Allen B MacKenzie. Cognitive networks. In *Cognitive radio, software defined radio, and adaptive wireless systems*, pages 17–41. Springer, 2007.

[18] Zhu Han, Dusit Niyato, Walid Saad, Tamer Başar, and Are Hjørungnes. *Game theory in wireless and communication networks: theory, models, and applications*. Cambridge university press, 2012.

[19] Osvaldo Simeone. A very brief introduction to machine learning with applications to communication systems. *IEEE Transactions on Cognitive Communications and Networking*, 4(4): 648–664, 2018.

[20] R. Mullins and M. T. Barros. 5GPPP Working Group on Network Management and QoS, Cognitive Network Management for 5G. *5GPPP*, 2017.

[21] Simone Redana, Ömer Bulakci, Anastasios Zafeiropoulos, Anastasius Gavras, Anna Tzanakaki, Antonino Albanese, Apostolos Kousaridas, Avi Weit, Bessem Sayadi, Boris Tiomela Jou, et al. 5G PPP Architecture Working Group: View on 5G Architecture. *5GPPP*, 2019.

[22] SN WG. Software Network Working Group: Vision on Software Networks and 5G. *5GPPP*, 2017.

[23] 3GPP TS 23.502 Technical Specification Group Services and System Aspects. Procedures for the 5G System; Stage 2. *3rd Generation Partnership Project*, 2019.

[24] 3GPP TS 23.501 Technical Specification Group Services and System Aspects. System Architecture for the 5G System; Stage 2. *3rd Generation Partnership Project*, 2018.

[25] 3GPP TS 38.401 Technical Specification Group Radio Access Network. NG-RAN; Architecture description. *3rd Generation Partnership Project*, 2018.

[26] 3GPP TS 38.300 Technical Specification Group Radio Access Network. NR; NR and NG-RAN Overall Description; Stage 2. *3rd Generation Partnership Project*, 2018.

[27] 3GPP TS 22.891 Technical Specification Group Services and System Aspects. Feasibility Study on New Services and Markets Technology Enablers; Stage 1. *3rd Generation Partnership Project*, 2019.

[28] 5G-ALLSTAR team. 5G ALL-STAR vision document: Vision, Scope and Goals. *5G ALL-STAR D2.1*, 2018.

[29] Sergey Andreev, Mikhail Gerasimenko, Olga Galinina, Yevgeni Koucheryavy, Nageen Himayat, Shu-Ping Yeh, and Shilpa Talwar. Intelligent access network selection in converged multi-radio heterogeneous networks. *IEEE wireless communications*, 21(6):86–96, 2014.

[30] SE El Ayoubi, M Boldi, Ö Bulakci, P Spapis, M Schellmann, P Marsch, M Säily, JF Monserrat, T Rosowski, G Zimmermann, et al. Preliminary views and initial considerations on 5g ran architecture and functional design. *White Pap. METIS II*, pages 1–27, 2016.

[31] Athul Prasad, Fernando Sanchez Moya, Marten Ericson, Roberto Fantini, and Omer Bulakci. Enabling RAN moderation and dynamic traffic steering in 5G. In *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, pages 1–6. IEEE, 2016.

[32] 3GPP TR 38.801 Technical Specification Group Radio Access Network. Study on new radio access technology: Radio access architecture and interfaces. *3rd Generation Partnership Project*, 2017.

[33] 3GPP TS 37.340 Technical Specification Group Radio Access Network. Evolved Universal Terrestrial Radio Access (E-UTRA) and NR; Multi-connectivity; Stage 2. *3rd Generation Partnership Project*, 2018.

[34] 3GPP TR 22.822 Technical Specification Group Services and System Aspects. Study on using Satellite Access in 5G; Stage 1. *3rd Generation Partnership Project*, 2018.

[35] 3GPP TR 38.811 Technical Specification Group Radio Access Network. NR; Study on New Radio (NR) to support non-terrestrial networks; Stage 2. *3rd Generation Partnership Project*, 2018.

[36] Ning Zhang, Shan Zhang, Shaohua Wu, Ju Ren, Jon W Mark, and Xuemin Shen. Beyond coexistence: Traffic steering in LTE networks with unlicensed bands. *IEEE wireless communications*, 23(6):40–46, 2016.

[37] Marcin Dryjanski and Michal Szydelko. A unified traffic steering framework for LTE radio access network coordination. *IEEE Communications Magazine*, 54(7):84–92, 2016.

[38] Lusheng Wang and Geng-Sheng GS Kuo. Mathematical modeling for network selection in heterogeneous wireless networks—A tutorial. *IEEE Communications Surveys & Tutorials*, 15 (1):271–292, 2012.

[39] Qingyang Song and Abbas Jamalipour. Network selection in an integrated wireless LAN and UMTS environment using mathematical modeling and computing techniques. *IEEE wireless communications*, 12(3):42–48, 2005.

[40] AA Sabbagh, Robin Braun, and Mehran Abolhasan. A comprehensive survey on rat selection algorithms for heterogeneous networks. *World Academy of Science, Engineering and Technology*, 2011.

[41] Quoc-Thinh Nguyen-Vuong, Yacine Ghamri-Doudane, and Nazim Agoulmine. On utility models for access network selection in wireless heterogeneous networks. In *NOMS 2008-2008 IEEE Network Operations and Management Symposium*, pages 144–151. IEEE, 2008.

[42] Olga Ormond, John Murphy, and Gabriel-Miro Muntean. Utility-based intelligent network selection in beyond 3G systems. In *2006 IEEE international conference on communications*, volume 4, pages 1831–1836. IEEE, 2006.

[43] Leifang Hui, Wei Ma, and Shenghua Zhai. A novel approach for radio resource management in multi-dimensional heterogeneous 5G networks. *Journal of Communications and Information Networks*, 1(2):77–83, 2016.

[44] Farooq Bari and Victor Leung. Multi-attribute network selection by iterative TOPSIS for heterogeneous wireless access. In *2007 4th IEEE consumer communications and networking conference*, pages 808–812. IEEE, 2007.

[45] A Wilson, Andrew Lenaghan, and Ron Malyan. Optimising wireless access network selection to maintain qos in heterogeneous wireless environments. In *wireless personal multimedia communications*, pages 18–22, 2005.

[46] Matteo Cesana, Nicola Gatti, and Ilaria 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*, pages 1–10, 2008.

[47] Josephina Antoniou and Andreas Pitsillides. 4G converged environment: Modeling network selection as a game. In *2007 16th IST Mobile and Wireless Communications Summit*, pages 1–5. IEEE, 2007.

[48] Sarabjot Singh, Shu-ping Yeh, Nageen Himayat, and Shilpa Talwar. Optimal traffic aggregation in multi-RAT heterogeneous wireless networks. In *2016 IEEE International Conference on Communications Workshops (ICC)*, pages 626–631. IEEE, 2016.

[49] Nemanja Vučević, Jordi Pérez-Romero, Oriol Sallent, and Ramon Agustí. Reinforcement learning for joint radio resource management in LTE-UMTS scenarios. *Computer Networks*, 55(7):1487–1497, 2011.

[50] Xavier Gelabert, Jordi Pérez-Romero, Oriol Sallent, and Ramon Agustí. A Markovian approach to radio access technology selection in heterogeneous multiaccess/multiservice wireless networks. *IEEE transactions on mobile computing*, 7(10):1257–1270, 2008.

[51] Bernard Gendron, Teodor Gabriel Crainic, and Antonio Frangioni. Multicommodity capacitated network design. In *Telecommunications network planning*, pages 1–19. Springer, 1999.

[52] Patrice Marcotte, Sang Nguyen, and Alexandre Schoeb. A strategic flow model of traffic assignment in static capacitated networks. *Operations Research*, 52(2):191–212, 2004.

[53] James Blake Rawlings and David Q Mayne. *Model predictive control: Theory and design*. Nob Hill Pub., 2009.

[54] Alberto Bemporad. Model predictive control design: New trends and tools. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 6678–6683. IEEE, 2006.

[55] ML Bell and RWH Sargent. Optimal control of inequality constrained DAE systems. *Computers & Chemical Engineering*, 24(11):2385–2404, 2000.

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

[57] Chin-Teng Lin, C. S. George Lee, et al. Neural-network-based fuzzy logic control and decision system. *IEEE Transactions on computers*, 40(12):1320–1336, 1991.

[58] Tao Zhang, Shuzhi Sam Ge, and Chang Chieh Hang. Adaptive neural network control for strict-feedback nonlinear systems using backstepping design. *Automatica*, 36(12):1835–1846, 2000.

[59] Richard S Sutton, Andrew G Barto, and Ronald J Williams. Reinforcement learning is direct adaptive optimal control. *IEEE Control Systems Magazine*, 12(2):19–22, 1992.

[60] Frank L Lewis and Derong Liu. *Reinforcement learning and approximate dynamic programming for feedback control*, volume 17. John Wiley & Sons, 2013.

[61] Antonio Pietrabissa, Francesco Delli Priscoli, Alessandro Di Giorgio, Alessandro Giuseppi, Martina Panfili, and Vincenzo Suraci. An approximate dynamic programming approach to resource management in multi-cloud scenarios. *International Journal of Control*, 90(3):492–503, 2017.

[62] Francesco Delli Priscoli, Alessandro Di Giorgio, Federico Lisi, Salvatore Monaco, Antonio Pietrabissa, Lorenzo Ricciardi Celsi, and Vincenzo Suraci. Multi-agent quality of experience control. *International Journal of Control, Automation and Systems*, 15(2):892–904, 2017.

[63] Pietrabissa Antonio, Battilotti Stefano, Facchinei Francisco, Giuseppi Alessandro, Oddi Guido, Panfili Martina, and Suraci Vincenzo. Resource management in multi-cloud scenarios via reinforcement learning. In *2015 34th Chinese Control Conference (CCC)*, pages 9084–9089. IEEE, 2015.

[64] Byung-Wook Wie. A differential game approach to the dynamic mixed behavior traffic network equilibrium problem. *European Journal of Operational Research*, 83(1):117–136, 1995.

[65] Yannis A Korilis and Aurel A Lazar. On the existence of equilibria in noncooperative optimal flow control. *Journal of the ACM (JACM)*, 42(3):584–613, 1995.

[66] Zhi-Mi Cheng, Xian-Wei Zhou, Yan Ding, and Fu-Hong Lin. A Cooperative Differential Game Model for Multiuser Rate-Based Flow Control. *Wireless personal communications*, 72 (2):1173–1186, 2013.

[67] Walid Saad, Zhu Han, Mérouane Debbah, and Are Hjorungnes. Network formation games for distributed uplink tree construction in IEEE 802.16 j networks. In *IEEE GLOBECOM 2008-2008 IEEE Global Telecommunications Conference*, pages 1–5. IEEE, 2008.

[68] Esteban Arcaute, Ramesh Johari, and Shie Mannor. Network formation: Bilateral contracting and myopic dynamics. *IEEE Transactions on Automatic Control*, 54(8):1765–1778, 2009.

[69] Dusit Niyato and Ekram Hossain. Dynamics of network selection in heterogeneous wireless networks: An evolutionary game approach. *IEEE transactions on vehicular technology*, 58(4), 2008.

[70] William H Sandholm. *Population games and evolutionary dynamics*. MIT press, 2010.

[71] Bin Wu, Da Zhou, and Long Wang. Evolutionary dynamics on stochastic evolving networks for multiple-strategy games. *Physical Review E*, 84(4):046111, 2011.

[72] Tamer Başar and Pierre Bernhard. *H-infinity optimal control and related minimax design problems: a dynamic game approach*. Springer Science & Business Media, 2008.

[73] Jacob Engwerda. Linear quadratic differential games: an overview. In *Advances in dynamic games and their applications*, pages 1–34. Springer, 2009.

[74] 5G-ALLSTAR team. Mapping of the multi-connectivity functions onto the 5G network architecture. *5G-ALLSTAR D4.1*, 2019.

[75] F. Lisi, G Losquadro, A. Tortorelli, A Ornatelli, and M Donsante. Multi-Connectivity in 5G terrestrial-Satellite Networks: the 5G-ALLSTAR Solution. *25th Ka and Broadband Communications Conference*, Sorrento, Italy, 2019.

[76] 5G-ALLSTAR team. Design and simulation of the multi-rat load balancing algorithms. *5G-ALLSTAR D4.2*, 2019.

[77] Caner Kilinc, Marten Ericson, Patrik Rugeland, Icaro Da Silva, Ali Zaidi, Osman Aydin, Venkatkumar Venkatasubramanian, Miltiades C Filippou, Marco Mezzavilla, Nandish Kuruvatti, et al. 5G Multi-RAT integration evaluations using a common PDCP layer. In *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, 2017.

[78] Sen Xu, Meng Hou, Yu Fu, Honglian Bian, and Cheng Gao. Improved Fast Centralized Retransmission Scheme for High-Layer Functional Split in 5G Network. In *Journal of Physics: Conference Series*, 2018.

[79] Azad Ravanshid, Peter Rost, Diomidis S Michalopoulos, Vinh V Phan, Hajo Bakker, Danish Aziz, Shreya Tayade, Hans D Schotten, Stan Wong, and Oliver Holland. Multi-connectivity functional architectures in 5G. In *2016 IEEE international conference on communications workshops (ICC)*, pages 187–192. IEEE, 2016.

[80] Subramanya Chandrashekar, Andreas Maeder, Cinzia Sartori, Thomas Höhne, Benny Vejlgaard, and Devaki Chandramouli. 5g multi-RAT multi-connectivity architecture. In *2016 IEEE International Conference on Communications Workshops (ICC)*, pages 180–186. IEEE, 2016.

[81] Diomidis S Michalopoulos, Ingo Viering, and Lei Du. User-plane multi-connectivity aspects in 5G. In *2016 23rd International Conference on Telecommunications (ICT)*, pages 1–5. IEEE, 2016.

[82] Icaro Da Silva, Gunnar Mildh, Johan Rune, Pontus Wallentin, Jari Vikberg, Paul Schliwa-Bertling, and Rui Fan. Tight integration of new 5G air interface and LTE to fulfill 5G requirements. In *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, 2015.

[83] António Morgado, Kazi Mohammed Saidul Huq, Shahid Mumtaz, and Jonathan Rodriguez. A survey of 5G technologies: regulatory, standardization and industrial perspectives. *Digital Communications and Networks*, 4(2):87–97, 2018.

[84] Aleksandra Checko, Henrik L Christiansen, Ying Yan, Lara Scolari, Georgios Kardaras, Michael S Berger, and Lars Dittmann. Cloud RAN for mobile networks—A technology overview. *IEEE Communications surveys & tutorials*, 17(1):405–426, 2014.

[85] Ericsson. Cloud RAN Architecture for 5G Cloud. *White Paper*, 2016.

[86] 3GPP TS 38.801 Technical Specification Group Radio Access Network. NR; Study on new radio access technology: Radio access architecture and interfaces; Stage 2. *3rd Generation Partnership Project*, 2018.

[87] Giovanni Giambene, Sastri Kota, and Prashant Pillai. Satellite-5G integration: A network perspective. *IEEE Network*, 32(5):25–31, 2018.

[88] Antonio Ornatelli, Alessandro Giuseppi, Vincenzo Suraci, and Andrea Tortorelli. User-aware centralized resource allocation in heterogeneous networks. In *2020 28th Mediterranean Conference on Control and Automation (MED)*, pages 292–298. IEEE, 2020.

[89] Antonio Ornatelli, Andrea Tortorelli, and Alessandro Giuseppi. Iterative MPC for Energy Management and Load Balancing in 5G Heterogeneous Networks. In *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pages 0467–0471. IEEE, 2020.

[90] Vanita Rana, Indu Bala, and Neelu Jain. Resource allocation models for cognitive radio networks: a study. *International Journal of Computer Applications*, 91(12), 2014.

[91] Thomas L Saaty. What is the analytic hierarchy process? In *Mathematical models for decision support*, pages 109–121. Springer, 1988.

[92] Thomas L Saaty. How to make a decision: the analytic hierarchy process. *European journal of operational research*, 48(1):9–26, 1990.

[93] Francesco Borrelli, Alberto Bemporad, and Manfred Morari. *Predictive control for linear and hybrid systems*. Cambridge University Press, 2017.

[94] David Q Mayne, James B Rawlings, Christopher V Rao, and Pierre OM Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36(6):789–814, 2000.

[95] Antonio Ornatelli, Andrea Tortorelli, and Francesco Liberati. A distributed reinforcement learning approach for power control in wireless networks. In *2021 IEEE World AI IoT Congress (AIIoT)*, pages 0275–0281, 2021. doi: 10.1109/AIIoT52608.2021.9454208.

[96] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction.* MIT press, 2018.

[97] Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.

[98] Loutfi Nuaymi, Philippe Godlewski, and Xavier Lagrange. Power allocation and control for the downlink in cellular CDMA networks. In *12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications. PIMRC 2001. Proceedings (Cat. No. 01TH8598).* IEEE, 2001.

[99] Chung-Ju Chang and Fang-Ching Ben. Down-link power control in DS/CDMA cellular mobile radio network. In *Proceedings of 1994 3rd IEEE International Conference on Universal Personal Communications*, pages 89–93. IEEE, 1994.

[100] Loutfi Nuaymi, Xavier Lagrange, and Philippe Godlewski. A power control algorithm for 3G WCDMA system. In *European Wireless*, pages 25–28. Citeseer, 2002.

[101] Mauricio G Cea and Graham C Goodwin. An MPC-based nonlinear quantizer for bit rate constrained networked control problems with application to inner loop power control in WCDMA. In *2011 9th IEEE International Conference on Control and Automation (ICCA)*, pages 153–158. IEEE, 2011.

[102] Fredrik Gunnarsson and Fredrik Gustafsson. Time delay compensation in power controlled cellular radio systems. *IEEE Communications Letters*, 5(7):295–297, 2001.

[103] Lucian Busoniu, Robert Babuska, and Bart De Schutter. A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(2):156–172, 2008.

[104] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *arXiv preprint arXiv:1911.10635*, 2019.

[105] Mehran Mesbahi and Magnus Egerstedt. *Graph theoretic methods in multiagent networks*, volume 33. Princeton University Press, 2010.

[106] Behrang Monajemi Nejad, Sid Ahmed Attia, and Jörg Raisch. Max-consensus in a max-plus algebraic setting: The case of fixed communication topologies. In *2009 XXII International Symposium on Information, Communication and Automation Technologies*, pages 1–7. IEEE, 2009.

[107] Jeffrey L Adler and Victor J Blue. A cooperative multi-agent transportation management and route guidance system. *Transportation Research Part C: Emerging Technologies*, 10(5-6): 433–454, 2002.

[108] Jae Won Lee, Jonghun Park, O Jangmin, Jongwoo Lee, and Euyseok Hong. A multiagent approach to q-learning for daily stock trading. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 37(6):864–877, 2007.

[109] Jorge Cortes, Sonia Martinez, Timur Karatas, and Francesco Bullo. Coverage control for mobile sensing networks. *IEEE Transactions on robotics and Automation*, 20(2):243–255, 2004.

[110] Cristiano Castelfranchi. The theory of social functions: challenges for computational social science and multi-agent learning. *Cognitive Systems Research*, 2(1):5–38, 2001.

[111] Craig Boutilier. Planning, learning and coordination in multiagent decision processes. In *TARK*, volume 96, pages 195–210. Citeseer, 1996.

[112] T Yoshikawa. Decomposition of dynamic team decision problems. *IEEE Transactions on Automatic Control*, 23(4):627–632, 1978.

[113] Xiaofeng Wang and Tuomas Sandholm. Reinforcement learning to play an optimal nash equilibrium in team markov games. *Advances in neural information processing systems*, 15: 1603–1610, 2002.

[114] Thinh Doan, Siva Maguluri, and Justin Romberg. Finite-time analysis of distributed TD(0) with linear function approximation on multi-agent reinforcement learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 1626–1635. PMLR, 09–15 Jun 2019.

[115] Michael L Littman. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*, pages 157–163. Elsevier, 1994.

[116] Junling Hu and Michael P Wellman. Nash q-learning for general-sum stochastic games. *Journal of machine learning research*, 4(Nov):1039–1069, 2003.

[117] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[118] Anurag Kumar, D Manjunath, and Joy Kuri. *Wireless networking*. Elsevier, 2008.