UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

PhD program in Signal Theory and Communications

Proposal and Evaluation of Connectivity Solutions for Beyond 5G Radio Access Networks

Doctoral thesis by:
Juan Jesús Hernández Carlón

Thesis advisor:
Dr. Jordi Pérez Romero, Dr. Oriol Sallent Roig

Signal Theory and Communications Department
Barcelona, June 2024

To my beloved mother

To my brother and sisters

To my beloved father, José

Abstract

In the context of the fifth generation (5G) of mobile communications, the number of connected devices connected is expected to increase substantially compared to previous systems (e.g. LTE). Similarly, more stringent user requirements in terms of quality of service (QoS) are also envisaged beyond the simple delivery of high data rates (e.g., in terms of latency, reliability, etc.). Delivering these services adequately represents a challenge for mobile network operators in terms of network deployment and operation. In mobile networks, coverage is one of the most important aspects as the network performance and the users' experience is entirely dependent on it. However, various factors, such as suboptimal planning, penetration losses, physical obstructions, etc., make mobile networks susceptible to coverage problems. Therefore, adopting strategies to ensure adequate coverage is critical for mobile network operators to successfully provide 5G and future (6G) services. In general, optimal connectivity conditions can be pursued through various strategies, such as coverage optimization, network capacity management, network layout upgrade, efficient radio resource management, etc. In this way, operators can guarantee the QoS requirements demanded by network users.

In recent years, artificial intelligence (AI) has revolutionized mobile network design, deployment and management processes, thus driving the development of innovative solutions, sustained also on other technological advances such as software-defined networking (SDN), network functions virtualization (NFV). In fact, the impact of these technological tools is not limited to 5G systems, but will also extend to future generations such as 6G. In this context, this thesis addresses the challenge of proposing, developing and evaluating solutions at the Radio Access Network (RAN) level with the objective of ensuring optimal connectivity conditions and thus satisfying the QoS requirements of network users. Various techniques are used to achieve this goal, with special emphasis on the use of AI techniques.

Firstly, the thesis presents a model to optimize Multi-connectivity (MC) in heterogeneous networks. MC is a key technology for managing high traffic densities and meeting stringent QoS requirements. Multi-connectivity allows users to simultaneously access the network through multiple RAN nodes, including LTE eNodeBs (eNBs) and 5G NR gNodeBs (gNBs). However, performing effective multi-connectivity management is challenging due to various factors such as propagation conditions, interference, loads of various cells, QoS metrics, etc. To address these challenges, the thesis presents a novel algorithm designed to dynamically split User Equipment (UE) traffic between different RATs and cells. The algorithm aims to satisfy QoS requirements while minimizing radio resource consumption in order to minimize the possibility of congestion in the involved cells. The proposed solution is based on AI, specifically on Deep Reinforcement Leaning (DRL) by means of the Deep Q-Network (DQN) algorithm. Through a training

phase, the model learns an optimal traffic splitting policy to be applied to each UE. The policy adapts to the current conditions of both the UE and the network. This adaptive approach improves network performance by increasing user throughput while mitigating the risk of having cell congestion.

Deepening into the problem of ensuring adequate network coverage and as a second contribution, the thesis presents a methodology for coverage optimization in 5G systems. This methodology is based on two main tasks: detection and resolution of coverage holes. For this purpose, the thesis introduces a Machine Learning based model capable of detecting and characterizing coverage holes by analyzing real network traffic data. A coverage hole becomes significant for network performance when there is a persistent presence of users in its region, as this is reflected in a degradation of both user experience and overall network performance. The proposed model considers the number of users to identify the holes that need to be resolved in order to improve the overall network performance.

Finally, as a solution to the coverage holes, the thesis considers the integration of relays in order to extend network coverage; the solution comprises both fixed and mobile relays (i.e., UEs acting as relays). The solution based on fixed relays focuses on a functionality to mitigate coverage holes by strategically placing relays in order to improve network availability. On the other hand, given the current and future availability of UEs with powerful communication and computing capabilities, device-to-device communications (D2D) emerge as a viable alternative to address coverage problems. In this respect, a solution to extend coverage based on Relay UEs (RUEs) is proposed. To achieve this effectively, a DQN-based algorithm is proposed for the intelligent activation or deactivation of the RUEs. Through a training process, the algorithm learns to activate and deactivate the RUEs strategically, providing coverage only when necessary, thus avoiding unnecessary activations that may represent excessive energy consumption. Overall, the proposed coverage optimization methodology has demonstrated its feasibility, leading to improved network and user performance.

Resumen

En el contexto de la quinta generación de comunicaciones móviles (5G), el número de dispositivos conectados a las redes ha aumentado sustancialmente en comparación con los sistemas anteriores (por ejemplo, LTE). Del mismo modo, los requisitos de los usuarios en cuanto a calidad del servicio (QoS) también han aumentado significativamente. Más allá de la mera entrega de altas velocidades de datos, la tecnología 5G considera escenarios y aplicaciones con diferentes requisitos en términos de latencia, fiabilidad, velocidad de datos, etc. Prestar estos servicios adecuadamente representa un reto para los operadores de redes móviles en términos de operación y despliegue de sus redes. En una red móvil, la cobertura es uno de los aspectos más importantes, ya que de esta depende en gran manera su rendimiento, así como el de los usuarios. Sin embargo, diversos factores, como puede ser una planificación subóptima, pérdidas de penetración, obstrucciones físicas, etc., hacen que las redes móviles sean susceptibles de sufrir problemas de cobertura. Garantizar una cobertura adecuada es fundamental a fin cumplir los estrictos requisitos de los servicios relacionados con 5G y servicios futuros (6G), de manera que los operadores de red han de adoptar estrategias que garanticen una cobertura óptima. En general, los operadores deben garantizar condiciones óptimas de conectividad para los usuarios de la red, para esto, han de implementar diversas estrategias, como, por ejemplo, la optimización de la cobertura, la gestión de la capacidad de la red, la actualización del layout de la red, hacer una gestión eficaz de los recursos radio, etc. De este modo, los operadores pueden garantizar los requisitos de calidad de servicio demandados por los usuarios de la red.

En los últimos años, la inteligencia artificial (IA) ha revolucionado los procesos de diseño, despliegue y gestión de redes móviles, impulsando así el desarrollo de soluciones innovadoras, sustentadas también en otros avances tecnológicos como las redes definidas por software (SDN), la virtualización de funciones de red (NFV). De hecho, el impacto de estas herramientas tecnológicas no se limita a los sistemas 5G, sino que se extenderá también a futuras generaciones como la 6G. En este contexto, esta tesis aborda el reto de proponer, desarrollar y evaluar soluciones a nivel RAN con el objetivo de garantizar unas condiciones óptimas de conectividad y satisfacer así los requisitos de QoS de los usuarios de la red. Para lograr este objetivo se utilizan diversas técnicas, con especial énfasis en el uso de técnicas de IA.

En primer lugar, la tesis presenta un modelo para optimizar la Multiconectividad en redes heterogéneas. La multiconectividad es una tecnología clave para gestionar elevadas densidades de tráfico y cumplir estrictos requisitos de QoS. La multiconectividad permite a los usuarios acceder simultáneamente a la red a través de múltiples nodos RAN, incluidos los eNodeB de LTE (eNB) y los gNodeB de 5G NR (gNB). Sin embargo, realizar una gestión eficaz de la multiconectividad supone un reto debido a diversos factores como las condiciones de propagación, las interferencias en los enlaces de los UE con diferentes celdas/RAT, las

cargas de las diversas celdas existentes, las métricas de QoS, etc. Para hacer frente a estos retos, la tesis presenta un novedoso algoritmo diseñado para dividir dinámicamente el tráfico de UE entre diferentes RATs y celdas. El algoritmo tiene como objetivo satisfacer los requisitos de QoS al mismo tiempo que se minimiza el consumo de recursos de radio a fin de minimizar la probabilidad de congestión en las celdas involucradas. La solución propuesta se basa en IA, concretamente en Deep Reinforcement Learning por medio del algoritmo Deep Q-Network (DQN). Mediante una fase de entrenamiento, el modelo aprende una política óptima de división del tráfico para aplicarla a cada UE. La política se adapta a las condiciones actuales tanto del UE como de la red. Este enfoque adaptativo mejora el rendimiento de la red al aumentar el throughput de los usuarios al mismo tiempo que minimiza la probabilidad de congestión de las celdas.

Profundizando en el problema de asegurar una adecuada cobertura de red y como segunda contribución, la tesis presenta una metodología para la optimización de la cobertura en sistemas 5G. Esta metodología se basa en dos tareas principales: detección y resolución de huecos de cobertura. Para ello, la tesis introduce un modelo basado en Machine Learning capaz de detectar y caracterizar huecos de cobertura mediante el análisis de datos reales de tráfico de red. Un hueco de cobertura se vuelve significativo para el rendimiento de la red cuando hay una presencia persistente de usuarios en su región, ya que esto se refleja en una degradación tanto de la experiencia del usuario como del rendimiento global de la red. El modelo propuesto tiene en cuenta el número de usuarios para identificar los huecos que hay que resolver para mejorar el rendimiento global de la red.

Por último, como solución a los huecos de cobertura, la tesis considera la integración de relays para ampliar la cobertura de la red; la solución comprende tanto relays fijos como móviles (es decir, equipos de usuario que actúan como relays). La solución basada en relays fijos se centra en una funcionalidad para mitigar los huecos de cobertura mediante la colocación estratégica de relays con el fin de mejorar la disponibilidad de la red. Por otro lado, dada la disponibilidad actual y futura de UEs con potentes capacidades de comunicación y computación, las comunicaciones dispositivo a dispositivo (D2D) emergen como una alternativa viable para abordar los problemas de cobertura. En este sentido, se propone una solución para ampliar la cobertura basada en Relay UEs (RUEs). Para conseguirlo de forma efectiva, se propone un algoritmo basado en DQN para la activación o desactivación inteligente de los RUEs. Mediante un proceso de entrenamiento, el algoritmo aprende a activar y desactivar los RUEs estratégicamente, proporcionando cobertura sólo cuando es necesario, evitando así activaciones innecesarias que pueden representar un consumo excesivo de energía. En general, la metodología de optimización de la cobertura propuesta ha demostrado su viabilidad, lo que ha permitido mejorar el rendimiento tanto de usuarios como de la red.

Acknowledgement

Firstly, I would like to express my sincere and deepest gratitude to my thesis advisors, Dr. Jordi Pérez-Romero and Dr. Oriol Sallent Roig. Thank you for your support, guidance, patience and all the time you have devoted to my Ph.D studies and the production of this dissertation. Your support has been invaluable.

I would also like to express my sincere thanks to Dr. Ferran Casadevall, whose role has been fundamental to my doctoral studies and also for the valuable conversations that have meant so much to me.

Furthermore, I am deeply grateful to all the members of the GRCM team who were always available to help. Thank you for the good times, the enriching conversations and the countless cups of coffee.

I must also thank the Spanish Ministry of Science and Innovation, as my Ph.D studies were mainly funded by the Ministry under the grant number PRE2018-084691, as well as by GRCM funding during the last months. My deepest gratitude goes to them.

Finally, I cannot overlook the invaluable support and love of my family and friends who, although far away, are always there for me. Thank you for your unwavering encouragement and love.

Contents

List of Figures

List of Tables

List of Abbreviations

3GPP	3rd Generation Partnership Project
5G	Fifth Generation
5GC	5G core
6G	Sixth Generation
AI	Artificial Intelligence
BS	Base Station
AMF	Access and Mobility Management Function
CAPEX	Capital Expenditure
CCO	Coverage and Capacity Optimization
CDF	Cumulative Density Function
CH	Coverage Hole
CN	Core Network
D2D	Device to Device
DBSCAN	Density-based Spatial Clustering of Applications with Noise
DDPG	Deep Deterministic Policy Gradients
DDQN	Double DQN
DL-SCH	Downlink Shared Channel
DNN	Deep Neural Networks
DPG	Deterministic Policy Gradients
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
eMBB	Enhanced Mobile Broadband
EMS	Element Management System
ETSI	European Telecommunications Standards Institute
FR	Frequency Range
Gbps	Gigabits Per Second
HetNets	Heterogeneous Networks
IFFT	Inverse Fast Fourier Transformation
IoT	Internet of Things
ITU	International Telecommunication Union

KPI	Key Performance Indicator
KPM	Key Performance Measurements
LTE	Long Term Evolution
MAC	Medium Access Control
MC	Multi-Connectivity
MN	Master Node
MCG	Master Cell Group
MDT	Minimization of Drive Tests
MDP	Markov Decision Process
MIMO	Multiple-Input Multiple-Output
MNO	Mobile Network Operator
ML	Machine Learning
mMTC	Massive Machine Type Communications
mmWave	Millimetre Wave
MSE	Mean Squared Error
NFV-MANO	Network Functions Virtualization Management and Orchestration
NMS	Network Management System
near-RT RIC	Near-Real time RAN Intelligent Controller
NF	Network Functions
NFV	Network Function Virtualisation
NFVO	NFV Orchestrator
NG-RAN	Next-Generation RAN
NR	New Radio
O-RAN	Open RAN
O-CU-CP	O-RAN Central Unit – Control Plane
O-CU-UP	O-RAN Central Unit – User Plane
O-DU	O-RAN Distributed Unit
O-RU	O-RAN Radio Unit
OFDMA	Orthogonal Frequency-Division Multiple Access
PDCP	Packet Data Convergence Protocol
PHY	Physical Layer
PM	Performance Management
PRB	Physical Resource Block
QoS	Quality of Service

RAN	Radio Access Network
RAT	Radio Access Technology
RB	Resource Block
RF	Radio Frequency
RLC	Radio Link Control
RL	Reinforcement Learning
RRC	Radio Resource Control
RRM	Radio Resource Management
RSRQ	Reference Signal Received Quality
RT-RIC	Real-Time RAN Intelligent Controller
RUE	Relay UE
SRB	Signalling Radio Bearer
SCG	Secondary Cell Group
SDAP	Service Data Adaptation Protocol
SDN	Software-Defined Networking
SINR	Signal to Interference Ratio
SLA	Service Level Agreement
SMDP	Semi Markov Decision Process
SMF	Session Management Function
SMO	Service Management and Orchestration
SN	Secondary Node
SON	Self-Organizing Network
TD	Time Difference
UE	User Equipment
UMa	Urban Macro-cell
UPF	User Plane Function
URLLC	Ultra-Reliable and Low-Latency Communications
VIM	Virtualized Infrastructure Manager
VNFM	VNF Manager
V2V	Vehicle-to-Vehicle Communication
V2X	Vehicle-to-Everything Communication

Chapter 1. Introduction

1.1. Fifth Generation (5G) of Mobile Networks

The emergence of 5G technology over the past few years has changed the landscape of mobile communications. As 5G networks continue to proliferate around the world, their impact will extend beyond traditional telecommunications, to sectors such as agriculture, smart cities and emergency response systems. Indeed, a future of ubiquitous connectivity is predicted with the full adoption of 5G technology, since it promises to revolutionise industries ranging from healthcare and transportation to entertainment and manufacturing [1]. Its performance capabilities in terms of reliability, low latency and high bandwidth pave the way for the proliferation of Internet of Things (IoT) devices, autonomous vehicles, augmented reality applications and much more.

5G is designed to deliver multi-gigabits per second (Gbps) peak data rates, coupled with ultra-low latency, high reliability, massive network capacity, enhanced availability and a consistent user experience [1]. These attributes are designed to support a range of usage scenarios, including wearables, smartwatches, autonomous vehicles, telemedicine, drones, and more. To bring clarity and specificity to these aspects, the International Telecommunication Union's Radiocommunication Sector (ITU-R) outlined the vision and requirements of the 5G system under the IMT-2020 framework [2]. According to the IMT-2020 document, eight parameters were identified as key capabilities, as shown in Fig. 1. These parameters served as guiding principles for the development and implementation of 5G technology.

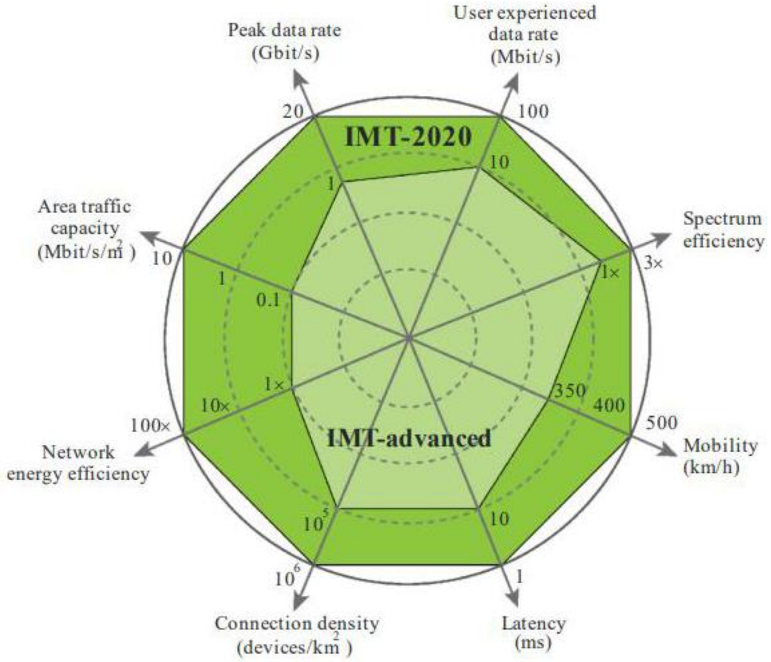

Fig. 1. Vision and requirements of 5G system (IMT-2020) [2]

To support the requirements, the 3rd Generation Partnership Project (3GPP) has standardised the 5G system architecture. This architecture encompasses the Next-Generation Radio Access Network (NG-RAN) [5], characterised by a novel radio access technology (RAT) called 5G New Radio (NR), and, in parallel, the evolution of the 5G core network, referred to as 5GC [6]. 5G NR integrates novel features and technologies to achieve the demanding target requirements of 5G. These include, among others, the use of millimetre wave (mmWave) frequencies, the deployment of massive multiple-input multiple-output (MIMO) antennas and the implementation of beamforming techniques [7].

In terms of services, the IMT-2020 vision outlines three primary usage scenarios within the 5G environment:

- **Enhanced mobile broadband (eMBB)**: eMBB is designed to deliver higher data rates, greater bandwidth, higher throughput, improved reliability and lower latency, as well as enhanced multimedia functionality to the end user. eMBB is primarily aimed at improving the user experience for applications such as video streaming, real-time gaming and virtual reality functionality by providing data rates of more than 10 Gbps.
- **Ultra-reliable and low-latency communications (URLLC):** This use case has stringent requirements for capabilities such as throughput, latency and availability. Wireless control of industrial manufacturing, self-driving cars and mission-critical applications, among many others, are examples of uRLLC use cases.
- **Massive machine type communications (mMTC):** This means a large number of connected devices transmitting a relatively small amount of non-delay-sensitive data. Devices must be low cost and have very long battery life. mMTC can be used to support Internet of Things applications.

Fig. 2. illustrates some examples of 5G usage scenarios defined in IMT-2020.

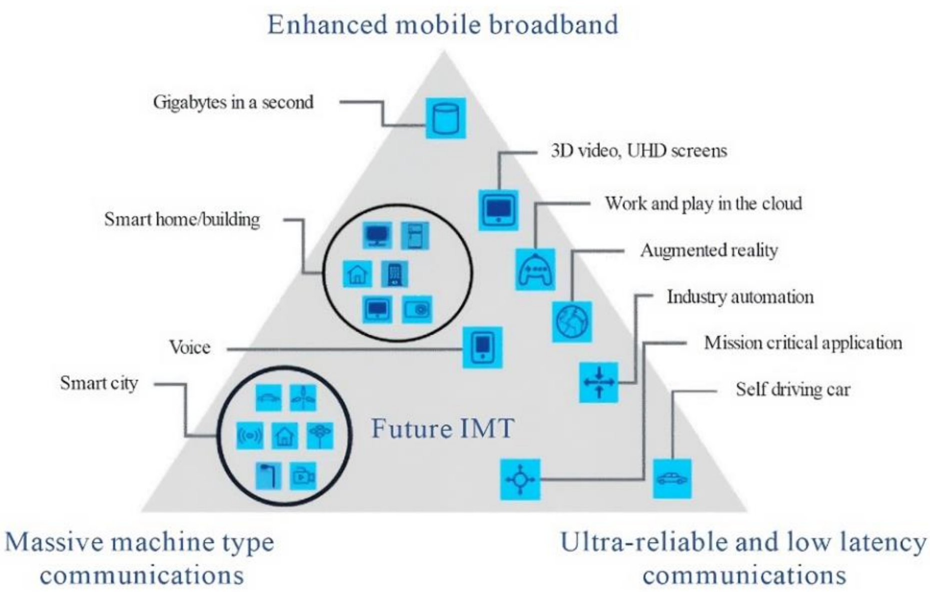

Fig. 2. 5G usage scenarios (IMT-2020) [2]

It is worth noting that, the ITU-R has already started discussions on the framework and objectives for IMT-2030 [3], which is aimed at the development of 6G. IMT-2030 is expected to support expanded and new usage scenarios compared to IMT-2020, while offering enhanced and novel capabilities. The envisaged usage scenarios include immersive communications, hyper-reliable and low-latency communications, massive communications, ubiquitous connectivity and the integration of artificial intelligence with communications. Fig. 3. illustrates the envisaged usage scenarios and the new targets for expected capabilities under IMT-2030.

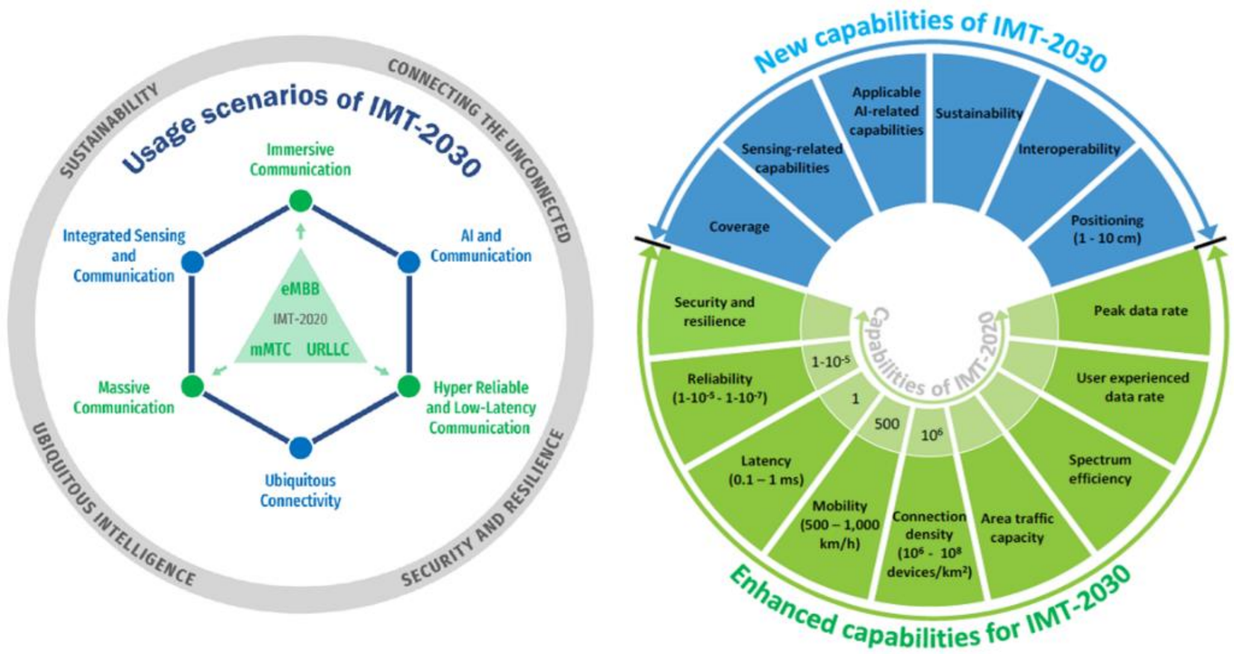

Fig. 3. Usage scenarios and capabilities (IMT-2030) [3]

1.2. Scope and Motivation

In recent years, there has been continued and substantial growth in the traffic generated by users of mobile networks. The 2023-Ericsson Mobility Report [10] projects that the average data consumption per smartphone will surpass 21 GB per month by the end of 2023 and is expected to reach 56 GB by 2028. Moreover, 5G subscriptions are forecast to surpass 5.3 billion worldwide by the end of 2029, representing more than half of all mobile subscriptions. In fact, 5G is anticipated to emerge as the dominant mobile access technology by 2028. In light of this expected growing demand in 5G services, Mobile Network Operators (MNOs) face a crucial challenge in enhancing their Radio Access Network (RAN) infrastructure to handle the expected increase in traffic.

Every service enabled by 5G networks has specific requirements that must be met for proper functioning. For example, the manufacturing industry demands low-latency, high-reliability and high-availability connectivity. Meanwhile, the financial sector requires higher data rates to collect vast amounts of high-frequency stock market data, and ultra-reliable, low-latency communications for rapid response and control in high-frequency trading [1]. To effectively meet the demands of all 5G-related services, service providers face the continuing challenge of ensuring the efficient operation of their networks. Achieving this efficiency requires optimal network planning and deployment, as well as management and optimisation processes. Typically, mobile networks struggle with coverage issues due to a variety of factors such as inadequate planning, interference, physical obstacles. etc. Ensuring adequate coverage is critical in 5G and future deployments due to the stringent QoS requirements. However, this is particularly challenging in deployments operating at mmWave frequencies. Due to their characteristics, mmWave signals are susceptible to blockage often caused by physical obstacles such as new buildings, grown-up trees, etc. As a result, there are regions where users experience inadequate signal strength, resulting in call drops and radio link failures (RLF), reducing user's performance to suboptimal levels. Without adequate coverage, it is impossible to meet the stringent demands of 5G-related services. In general, MNOs need to ensure optimal connectivity conditions for network users. To achieve this, they must approach various aspects such as optimising network coverage, managing network capacity, upgrading network layout, implementing appropriate radio resource management (RRM), etc. The advances in mobile communications field have been accompanied by significant advances in the AI field. Indeed, service providers are now integrating AI into network management, planning and optimisation processes [11], facilitating the development of AI-based models to improve efficiency. Indeed, the impact of AI is anticipated to be even greater in the 6G era than it is in 5G [3]. As a result, there is a wide scope for contributions in this area. In this context, this thesis finds space to define, develop and evaluate solutions at the RAN level to ensure optimal connectivity conditions for 5G network users, thereby meeting their QoS requirements. In addition, the thesis explores the incorporation of AI techniques in the presented solutions.

This work is divided into three main parts. The first part focuses on a solution to optimise multi-connectivity in heterogeneous networks. MC allows users to access the network through multiple RAN nodes, such as LTE and 5G NR, simultaneously. This enables network users to experience improved connectivity conditions, resulting in higher data rates, improved reliability and reduced latency to meet their QoS requirements. The MC solution is specifically designed to optimise the UE traffic splitting process, with a focus on minimising radio resource consumption (e.g. Physical Resource Blocks - PRBs), meeting user QoS requirements and avoiding cell congestion. The solution uses the DQN algorithm to dynamically learn an optimal traffic splitting policy through training stage. This policy, applied on a per-UE basis and automatically adapts to current network and user conditions, improving network performance by increasing satisfying users QoS while avoiding cell congestion.

Coverage is a key element in achieving ubiquitous connectivity, which is, in turn a key element of today's 5G networks and is expected to be critical in future 6G networks (IMT-2030). In that respect, the second part of the thesis proposes a methodology for coverage optimisation in mobile networks. The methodology divides the optimisation process into two main tasks: the detection of coverage-constrained areas and the resolution of coverage problems. For the detection of areas with coverage limitations, the thesis introduces a machine learning-based coverage hole detector (CHD). This CHD is designed to identify, delimit, and characterise coverage holes by analysing real-network user performance reports. The solution is designed to identify the coverage holes that must be resolved to improve the overall performance of the network. The third part of the thesis focuses on effectively mitigating coverage issues. In particular, as a solution to the detected coverage holes, the proposed methodology considers the use of both fixed and mobile relays to extend coverage. The fixed relay-based solution is designed to strategically place relays in order to mitigate coverage holes, thereby enhancing the network service availability and, consequently, the user experience. In contrast, the mobile relay-based solution leverages D2D communication as an option to address coverage problems. In this regard, the solution explores the use of relay UEs as an alternative to efficiently improve coverage, with the aim of enhancing the QoS of network users. Specifically, the solution focuses on functionality for an efficient activation process of the Relay UEs. The activation functionality is based on a Deep Q-Network algorithm, which, through a training stage, learns the optimal activation strategy according to the current network dynamics.

In the following, the objectives, methodology and contributions of this thesis are explained in detail.

1.3. Objectives, methodology and Thesis contributions

Based on the scope and motivation presented in previous section, the overall objective of this thesis is **to propose, develop and evaluate different solutions for the 5G RAN that guarantee optimal connectivity conditions for network users, thus satisfying their QoS requirements**.

In order to achieve the global objective of the thesis, it has been divided into two main sub-objectives:

1. First, it aims to define and develop an AI-based solution for optimising multi-connectivity in heterogeneous networks, focusing on ensuring user QoS satisfaction while effectively managing radio resources.

2. Given the critical role of network coverage in ensuring QoS satisfaction, the second sub-objective involves defining an AI-based methodology for coverage optimization in 5G networks. This sub-objective encompasses two central tasks:

    a. Detection of coverage-constrained areas through extensive analysis of real user-network performance data.

    b. Provide effective solutions to enhance coverage in the detected regions.

The general methodology for approaching the different solutions proposed in this thesis includes the following stages:

- **State-of-the-art review**: to conduct the state-of-the-art review on each topic, a variety of resources have been utilized. These include scientific articles, technical documents such as those from 3GPP, books, video lectures, and tutorials.

- **Formulation of the problem and its solution**: this stage encompasses the conception of the idea, the formulation of the problem to be addressed, and the proposed solution from both a mathematical and algorithmic point of view. The proposed solution is also formulated from the perspective of a specific AI technique.

- **Codification and validation of the solution elements:** this stage involves coding both the proposed solution and the environments for simulation and evaluation having as reference the mathematical and algorithmic approach for the solution. The main tools used in this phase include the Python programming language, TensorFlow (an open source library from Google for developing deep learning models), the MATLAB numerical computing system, and Scikit-learn (a Python machine learning library), among others. Once both the evaluation environment and the proposed model have been coded, they are subject to rigorous operational validation. The aim of this process is to identify potential errors early on that could have a negative impact on the results.

- **Assessment of proposed solution:** Once the validation process is complete, this stage implies to evaluate the proposed solution under the conditions for which the solution was designed. The data obtained from this stage is analysed by using various statistical tools and techniques, as well as graphical visualisations of the key metrics.

Based on the above context, the main thesis contributions (TC) are listed below:

[TC.1]Development and evaluation of a DQN-based algorithm for multi-connectivity optimization in heterogeneous networks. The model enables the implementation of an efficient multi-connectivity process in heterogeneous networks scenarios. It facilitates the distribution of UE traffic across different RATs and cells while ensuring QoS requirements are met and minimizing resource consumption, which reduces the probability of cells overload. The contributions associated to this model are the following:

[TC.1.1]Definition of a mathematical formulation that measures together the effectiveness of a traffic split configuration in minimizing the fraction of total bandwidth allocated to the UE while at the same time ensuring that the obtained throughput is above the required bit rate, and that the total fraction of occupied PRBs in the involved RATs/cells is lower than a given threshold.

[TC.1.2]Definition of a DQN agent capable of interacting with an environment, such as a simulated Radio Access Network, to learn a policy for deciding the most optimal traffic split configuration for a UE given both UE and network conditions at the moment of decision-making, the goal is to enable the agent to learn an effective traffic-splitting policy maximizing the result of the metric defined in TC1.1.

[TC.1.3] Mapping of the proposed solution within the context of the Open RAN (O-RAN) architecture that provide context on where the different elements of the solution can be hosted and on the interaction among them.

[TC.2]Definition, development and assessment of a Machine Learning based methodology for the optimal detection of coverage holes in 5G networks. The developed methodology enables the possibility to detect coverage holes within the region of a given base station by analysing user performance data reports. Similarly, the model provides alternatives for ensuring coverage of the detected holes. Some contributions associated to the proposed methodology are the following:

[TC.2.1]Definition of an algorithm for detecting coverage holes based on a machine learning clustering technique called DBSCAN. The algorithm consists of four main stages. The first stage involves pre-processing a large volume of real-user performance reports. The second stage optimizes the DBSCAN parameters to ensure optimal performance. In the third stage, coverage holes are detected and characterized. The final stage of the algorithm validates the sustained presence of traffic in the coverage holes to confirm which ones must be resolved.

[TC.2.2]Definition of functionality for determining the optimal placement of relay nodes to provide coverage to the detected and validated coverage holes. This functionality is based on analyzing the pathloss conditions in the surrounding areas of the coverage holes.

[TC.3]Development and assessment of a DQN-based methodology for coverage enhancement of 5G networks through optimal relay UEs activation. The proposed model allows for the possibility of intelligently activating some UEs in the network as relays to provide coverage to those regions with coverage problems, given the spatio-temporal traffic conditions. The fact that the activations are based on DQN makes the solution energy efficient, as activation is only performed when necessary. Some others contributions associated to this model are the following:

[TC.3.1]Definition of a metric to measure the efficiency of the activation process for all relay UEs. This formulation measures the combined efficiency of the mode (activated or deactivated) of all the relay UEs as a function of how users they provide coverage if active, or how users let without coverage if deactivated.

[TC.3.2]Definition of a DQN agent and the formulation of state, action and reward for DQN functioning. The defined agent was trained within a simulated 5G base station environment to handle the decision-making process of activating/deactivating all the relay UEs in an efficient manner, aligning with the criterion of TC.3.1.

[TC.3.3]Definition of the Key Performance Indicators (KPIs) for the assessment of the RUE activation functionality in terms of network performance improvements of the relay UEs activation. The energy efficiency of the solution is also measured.

1.4. Organization of the Thesis

This thesis is organized in 6 chapters. The structure of the document is illustrated in Fig. 4.

Chapter 2 provides general background concepts to contextualise the reader within the different topics covered throughout the document.

Chapter 3 presents a Deep Reinforcement Learning (DRL)-based solution aimed at optimising traffic sharing in heterogeneous network environments (e.g. LTE+5G NR) where multi-connectivity is deployed. The solution is designed to intelligently split traffic and allocate proportions of devices to different cells based on current traffic loads and radio conditions, with the ultimate goal of ensuring QoS satisfaction for network users. In addition, the solution has been designed to function as a third-party application, known as xApps, within the Open RAN (O-RAN) architecture framework. The proposed solution has been extensively evaluated under various conditions to contextualise its performance.

Chapter 4 addresses the problem of enhancing network coverage in 5G networks to achieve ubiquitous connectivity, thereby improving the user experience. To this end, a methodology is proposed that firstly focuses on detecting coverage-constrained areas. In this respect, a DBSCAN-based detector is introduced to analyze extensive user network performance data and identify coverage holes (CHs) with persistent traffic loads. As an alternative to mitigating the identified CHs, the model considers the use of fixed relays to extend the coverage of the RAN. In particular, an optimal relay positioning functionality is proposed. To evaluate the results of the proposed methodology in terms of network performance, a metric has been formulated to jointly measure the average spectral efficiency and the level of service availability in the network. The methodology has been evaluated through system-level simulations that replicate the real spatio-temporal traffic patterns collected from the Campus Nord of the Universitat Politècnica de Catalunya

Chapter 5 deals with the problem of optimally utilizing the relaying capabilities of UEs to augment the RAN in 5G deployments and beyond, particularly in those utilizing millimetre waves where radio signals may struggle to penetrate certain structures, leading to the presence of coverage holes. To address these limitations, the use of relay UEs (RUEs) is proposed as a solution to effectively extend the coverage of a cellular network. In this regard, a deep learning-based algorithm is proposed in this chapter to optimize the decision-making process regarding when RUEs should be activated and deactivated, taking into account the benefits they can provide in terms of increasing spectral efficiency and decreasing outage probability for network users. To evaluate the efficiency of the RUE activation process, a metric has been formulated to measure the effectiveness of the current mode of the RUEs (activated or deactivated) as a result of decisions made by the DQN agent. Additionally, different key performance metrics of interest have been defined to evaluate the solution. The proposed solution has been evaluated and compared with six different activation strategies, providing context on the real impact and benefits obtained.

Chapter 6 provides a summary of the work presented in this thesis and highlights the main conclusions reached. It also discusses future research directions of the work.

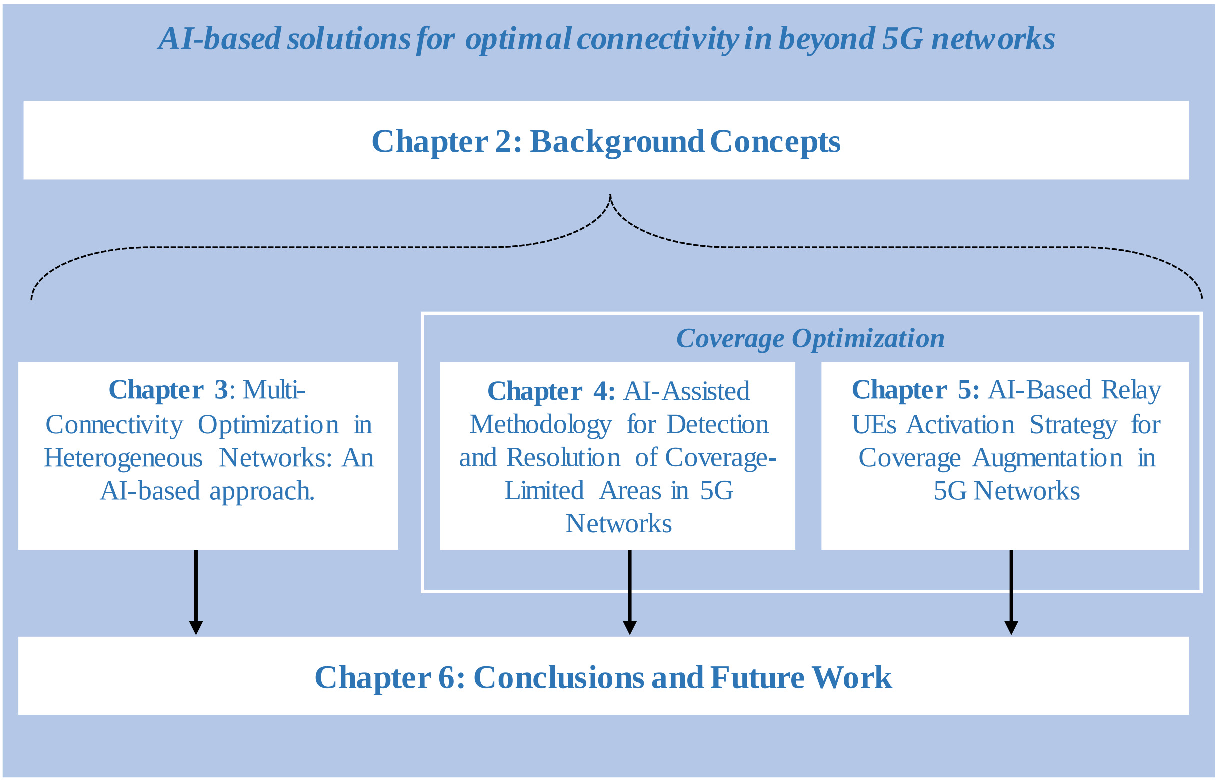

Fig. 4. Organisation of this Thesis.

1.5. List of Publications

As dissemination activities resulting from the thesis, the following papers have been published in journals (J) and conferences (C).

1.5.1. Journals

[J1] **J.J. Hernandez-Carlon,** J. Pérez-Romero, O. Sallent, I. Vilà, F. Casadevall, “A Deep Q-Network-Based Algorithm for Multi-Connectivity Optimization in Heterogeneous Cellular-Networks,” *Sensors 2022*, 22, 6179, August 2022.

[J2] **J. J. Hernández-Carlón**, J. Pérez-Romero, O. Sallent, I. Vilà, F. Casadevall, “Deep Learning-based Algorithm for Optimizing Relay User Equipment Activation in 5G Cellular Networks,” *IEEE Transactions on Vehicular Technology*, October 2023.

1.5.2. Conferences

[C1] **J. J. Hernández-Carlón,** J. Pérez-Romero, O. Sallent, I. Vilà, F.Casadevall, “A Deep Q Network-based Multi-Connectivity Algorithm for Heterogeneous 4G/5G Cellular Systems,” *18th International Conference on Artificial Intelligence Applications and Innovations (AIAI 2022)*, Crete (Greece), Virtual Conference, June 2022.

[C2] **J. J. Hernández-Carlón**, J. Pérez-Romero, O. Sallent, I. Vilá and F. Casadevall, “Deep Learning-based Multi-Connectivity Optimization in Cellular Networks,” *2022 IEEE 95th Vehicular Technology Conference*: (VTC2022-Spring), Helsinki, Finland, June 2022.

[C3]**J. J. Hernández-Carlón**, J. Pérez-Romero, O. Sallent, I. Vilà and F. Casadevall, "On the Detection and Solution of Coverage Holes in 5G Networks through Relay User Equipment: a combined DBSCAN and Deep-Q Network Approach," *2023 IEEE 97th Vehicular Technology Conference*: (VTC2023-Spring), Florence, Italy, June 2023.

1.5.3. Contribution to Research Projects

The work done during the thesis was related to the following research projects:

- **ARTIST**: "smART radIo acceSs with inTegration of user devices" funded by the Agencia Estatal de Investigación. Ref. PID2020-115104RB-I00.
- **BeGREEN**: "Beyond 5G Artificial Intelligence Assisted Energy Efficient Open Radio Access Networks". Supported in part by the Smart Networks and Services Joint Undertaking through the European Union's Horizon Europe Research and Innovation Programme under Grant: 101097083.
- **VERGE:** "AI-powered eVolution towards opEn and secuRe edGe architEctures". Supported in part by the Smart Networks and Services Joint Undertaking through the European Union's Horizon Europe Research and Innovation Programme underGrant: 101096034.
    - D3.1. Oluwatayo Y. Kolawole, Joan S. Pujol Roig (Editor). *First report on VERGE edge intelligence and network management*, December 2023.

Finally, it is important to mention that this work is also funded by the Spanish Ministry of Science and Innovation under grant ref. PRE2018-084691.

Chapter 2. Background Concepts.

2.1. Introduction

This chapter provides a brief overview of the key concepts that are behind the solutions presented throughout the thesis. Keeping these concepts in mind will enhance the reader’s understanding and the readability of the document.

2.2. 5G System Architecture

Overall, the 5G system is composed essentially by three main elements: User Equipment (UE), which itself consists of a mobile station and a USIM, the next generation RAN (NG-RAN) and the 5G core network (5GC), as shown in Fig. 5.

The NG-RAN in charge of different radio-related functionalities within the 5G system, such as scheduling, radio resource handling, retransmission protocols, coding, and various multi-antenna schemes. On the other hand, the 5GC is responsible of ensuring the proper operation of the network , encompassing functionalities such as ,session management, mobility management, access control authentication, among others important aspects that are essential for network operation.

The NG-RAN is responsible for various radio-related functionalities within the 5G system, including scheduling, radio resource handling, retransmission protocols, coding, and various multi-antenna schemes. On the other hand, the 5GC is responsible of ensuring the proper operation of the network, encompassing functionalities such as session management, mobility management, access control, authentication, charging, and other important aspects.

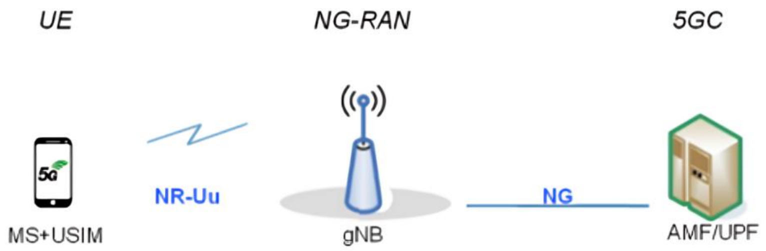

Fig. 5. Overview of the 5G system [13].

2.2.1. 5G Core Network (5GC)

The 5G core network is the heart of a 5G network since it is essential for the proper operation of the network. The 5GC establishes reliable, secure connectivity to the network for end users and provides access to its services. The 5G core network could be conceived as an evolution of the EPC used in LTE but essentially with three main enhancement compared to EPC: service-based architecture, support for network slicing, and control-plane/user-plane split [7]. The new 5GC architecture is based on what is called a Service-Based

Architecture (SBA). In this new architecture, each network function (NF) offers one or more services to other NFs via Application Programming Interfaces (API). Each NF is formed by a combination of small pieces of software code called as microservices. On high level, the 5G core network can be illustrated as shown in Fig. 6. The figure, uses a service-based representation. The different functionalities and elements in the architecture are listed below:

- Authentication Server Function (AUSF)
- Access and Mobility Management Function (AMF)
- Network Exposure Function (NEF)
- Network Repository Function (NRF)
- Network Slice Selection Function (NSSF)
- Policy Control Function (PCF)
- Session Management Function (SMF)
- Unified Data Management (UDM)
- User Plane Function (UPF)
- Application Function (AF)
- User Equipment (UE)
- (Radio) Access Network ((R)AN)

For more details and the specific description of the different Network Function of the 5GC, see section 6 of [6].

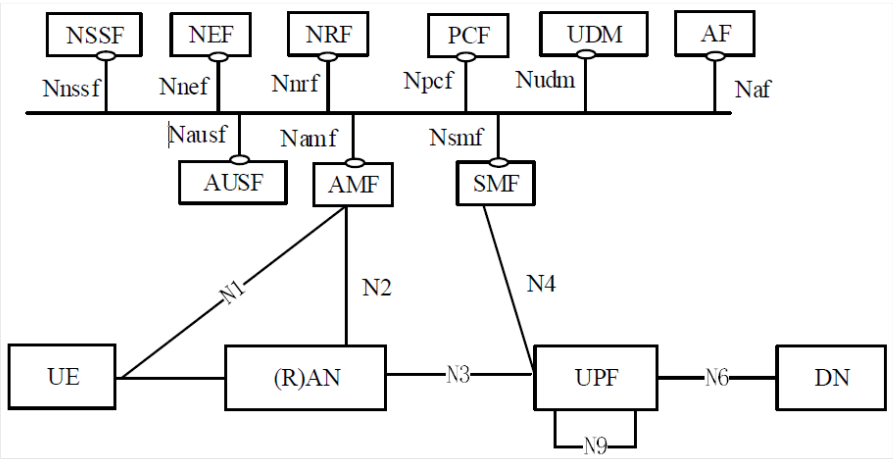

Fig. 6. High-level 5G network architecture (service-based description) [6].

2.2.2. New Generation -Radio Access Network (NG-RAN)

The Radio Access Network is a key component of a mobile network responsible for connecting devices to the core network via wireless. In the context of 5G it is designated as Next Generation Radio Access Network (NG-RAN). NG-RAN Node can be either a gNB or an ng-eNB according to the 3GPP [5]. A gNB, providing NR user plane and control plane protocol terminations towards the UE; or an ng-eNB, providing E-UTRA user plane and control plane protocol terminations towards the UE.

The gNBs and ng-eNBs are interconnected to each other by means of the Xn interface. The gNBs and ng-eNBs are also connected by means of the NG interfaces to the 5GC, more specifically, to the AMF (Access and Mobility Management Function) by means of the NG-C interface and to the UPF (User Plane Function) by means of the NG-U interface. Fig. 7 provides a general overview on the NG-RAN architecture.

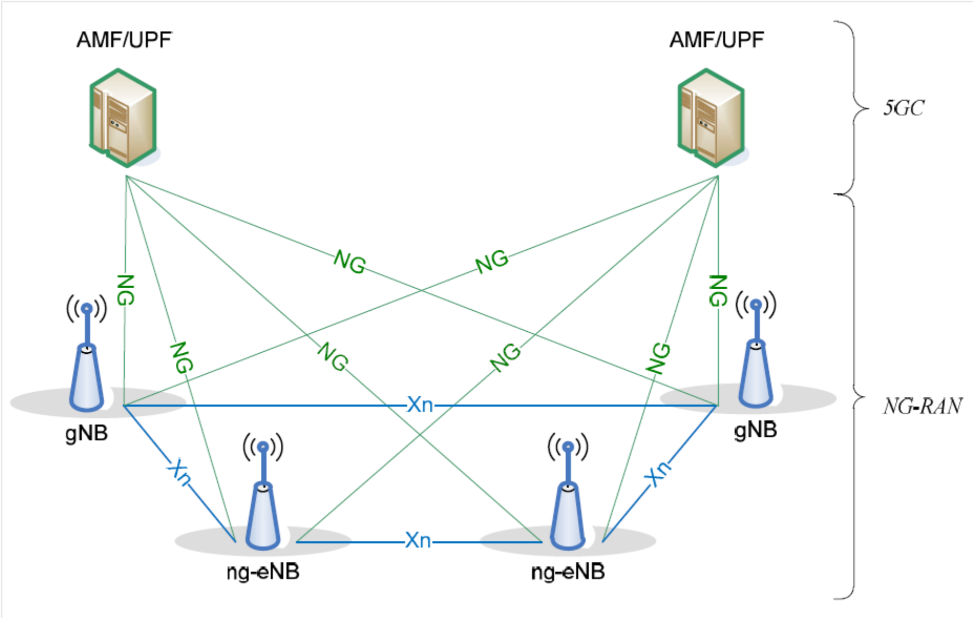

Fig. 7. NG-RAN overall architecture [5].

It is important to note that the transition process between legacy technologies (still in use) such as Long-Term Evolution - LTE and 5G takes into account different functional modes of network configuration. In this respect, 3GPP has identified a number of possible configurations that allow 5G NR elements to be connected to LTE elements, as shown in Fig. 8.

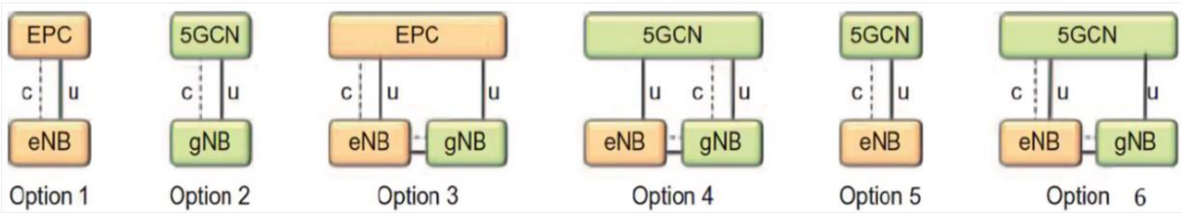

Fig. 8. Different combinations of core networks and radio-access technologies [7].

Option 3, known as "non-standalone operation", involves the use of LTE technology for control plane functions such as initial access and mobility, while NR is used exclusively for user plane data. Both the gNB and the ng-eNB have the ability to establish connections with UE over the air interface. In this setup, the gNB and ng-eNB are responsible for radio-related functions in one or more cells. In contrast, in the standalone mode shown in Option 2, the gNB is directly connected to the 5G core. Here, both user plane and control plane functions are managed by the gNB. Options 4, 5 and 7 show different ways of connecting an LTE eNB to the 5GCN.

2.2.3. Management Plane

One of the key aspects of 5G and beyond networks is virtualisation, which involves abstracting network functions and services from the underlying physical hardware and implementing them as software. This approach significantly improves the flexibility, scalability and efficiency of the network. However, proper management of these functions is critical. In this context, the European Telecommunications Standards

Institute (ETSI) has introduced a framework known as Network Functions Virtualisation Management and Orchestration (NFV-MANO) for the management and orchestration of virtualised network functions [8]. NFV-MANO consists of three main components: the NFV Orchestrator (NFVO), the VNF Manager (VNFM) and the Virtualised Infrastructure Manager (VIM). These components work together to provide a comprehensive framework for dynamic and automated management of virtualised network functions. The NFVO provides service orchestration and overall resource management, the VNFM manages the lifecycle and configuration of VNFs, and the VIM monitors the underlying physical and virtual infrastructure. This coordinated approach enables the flexible, scalable and efficient operation of virtualised networks, which is essential to meet the requirements of 5G networks (see [8]).

As shown in Fig. 9 the service management and orchestration of the network can include various functionalities such as the NFV MANO mentioned above, as well as others such as the Network Management System (NMS) and the Element Management System (EMS). These components are integral parts of network management, each serving different purposes and interacting in specific ways. The EMS is an application that manages network elements (NEs) individually, allowing granular control and monitoring of specific elements. It is responsible for performing the FCAPS functions of the NEs, where FCAPS stands for Fault, Configuration, Accounting, Performance and Security [9]. In contrast, the NMS is an application that can perform FCAPS functions for NEs in a correlated manner. This means that it manages NEs by considering both the data of individual NEs and the aggregation of all data, understanding the connections and relationships between them. As a result, the NMS can perform FCAPS functions from a comprehensive network-wide perspective [9].

Overall, management and orchestration functionalities are key to the optimal operation, maintenance and optimisation of mobile networks. In the context of 5G and beyond, these functionalities are essential to provide adequate network connectivity conditions, resulting in optimal network performance and improved end-user experience.

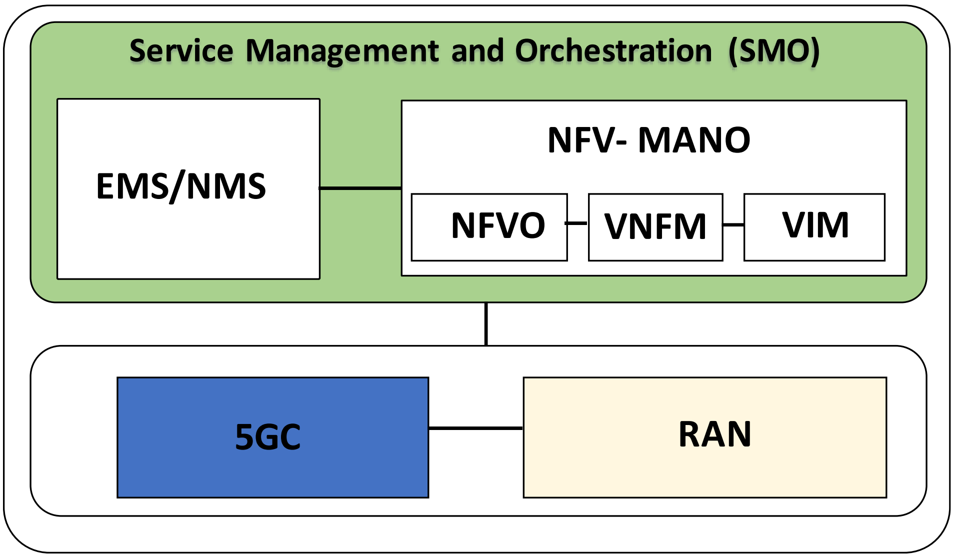

Fig. 9. Network Functionalities for Management and Orchestration

2.3. Heterogeneous Networks (HetNets)

Either 5G systems or beyond (e.g. 6G) will support applications with high data rate requirements. In addition, the number and density of devices requiring connectivity is expected to increase significantly [3]. One solution to meet the service requirements of mobile networks is to deploy heterogeneous networks (HetNets). HetNets consist of macro, micro, pico and femto cells. In turn, as a part of the heterogeneity, the network can incorporate multiple types of RATs. The concept of multi-RAT is conceived to allow network users to get access via different radio access technologies such as (2G, 3G, 4G, 5G) as well other wireless standards like Wi-Fi. For instance, in a HetNet deployment, it is possible to have macro cells operating on 4G LTE for wide-area coverage, small cells or femtocells operating on 4G or 5G to enhance coverage and capacity in dense urban areas and even Wi-Fi access points integrated into the cellular network to offload data traffic in indoor environments or high-traffic areas. The primary objective of HetNets is to improve network performance, capacity and coverage by deploying a mix of cell sizes and types tailored to the specific needs of different areas and user densities. Some of the benefits of using HetNets include

- ***Improved Coverage and Capacity*****:** By deploying a combination of macrocells and small cells, HetNets can provide better coverage and capacity in both urban and rural areas. Small cells are particularly useful in dense urban environments where macrocells alone may struggle to meet the demand for high-speed data.

- ***Enhanced User Experience*****:** HetNets help enhance the overall user experience by reducing congestion and improving network reliability and data speeds, especially in crowded areas such as stadiums, airports, and urban centres.

- ***Optimized Resource Utilization*****:** HetNets allow for more efficient use of available spectrum and network resources by dynamically adapting to changing traffic patterns and user demands. This flexibility improves spectral efficiency and helps operators maximize their investment in network infrastructure.

- ***Cost Savings*****:** Deploying small cells as part of a HetNet architecture can be more cost-effective than relying solely on traditional macrocells deployments, especially in areas with high population densities. Small cells are typically easier and less expensive to install and maintain than macrocells.

- ***Support for Diverse Applications****:* HetNets are well-suited for supporting a wide range of applications and services, including enhanced mobile broadband, massive machine-type communications (IoT), ultra-reliable low-latency communications (URLLC), and mission-critical services.

A graphic example of a heterogeneous network is presented in the figure below:

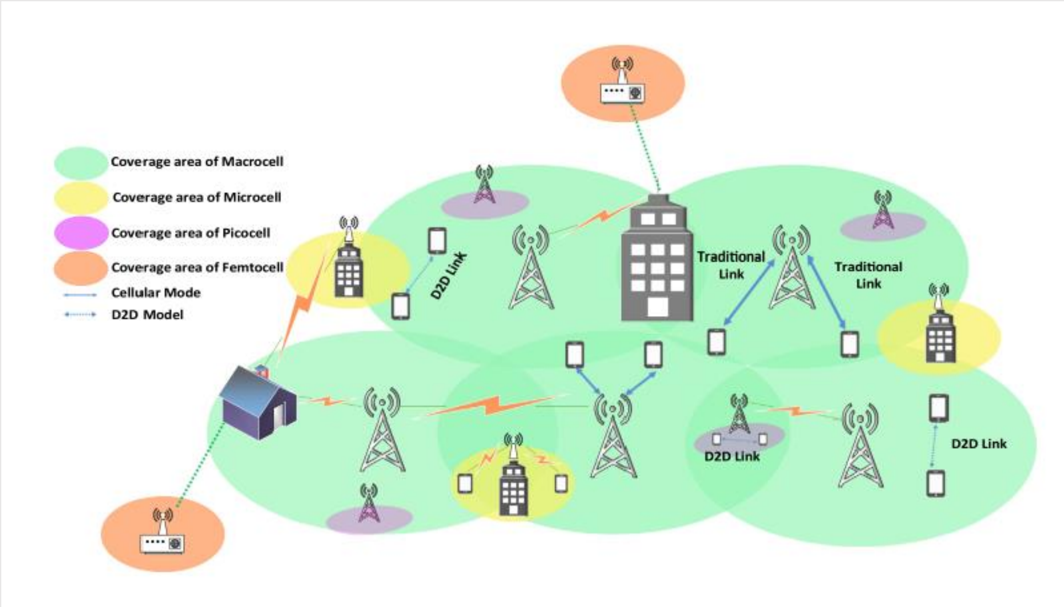

Fig. 10 Heterogeneous Network Model of [14].

2.3.1. Multi-connectivity (MC)

Multi-connectivity (MC) leverages the heterogeneous networks since it enables a UE to get connected to multiple access nodes of the network at the same time. This could offer many benefits in terms of quality of service, energy efficiency, mobility, and spectrum and interference management [15]. The central idea behind the MC concept is that a UE has connectivity with different nodes of the Radio Access Network (RAN) at the same time, e.g. eNodeBs (eNB) operating with LTE and/or gNodeBs (gNB) operating with 5G NR [7]. There is one master node (MN) responsible for the radio-access control plane and one, or in the general case multiple, secondary node(s) (SN) that provide additional user-plane links. In such manner, the UE can be provided with radio resources from distinct eNBs/gNBs, which enables a way of fulfilling the strict service requirements in terms of high data rate and ultra-reliability.

In the Third Generation Partnership Project (3GPP), MC is specified through the Multi-Radio Dual Connectivity (MR-DC) feature defined in [16] that considers different options depending on the technology used by the MN and by the SN and on the core network technology (i.e. 5G core or Evolved Packet Core (EPC)). Moreover, some 3GPP study items have addressed the MR-DC with multiple cells operating in different bands for example, [17], which considers up to 4 bands in LTE and 2 bands in 5G NR, one for the sub-6GHz Frequency Range 1 (FR1) and the other for the mmWave Frequency Range 2 (FR2). The operation of MR-DC is built upon the use of three different types of radio bearers [16], namely the Master Cell Group (MCG) bearers in which data is transmitted through the MN, the Secondary Cell Group (SCG) bearers in which data is transmitted through the SN, and the Split bearers, in which data is split between the SN and the MN at the Packet Data Convergence Protocol (PDCP) layer of the radio interface protocol stack.

2.3.2. Device-to-Device (D2D) Communication

Device-to-Device (D2D) communication is a wireless technology that allows devices to communicate directly with each other, without the need to transmit data through a network infrastructure. This technology has applications in various scenarios, including proximity-based services, where devices detect the presence of near neighbours and initiate various services, such as social applications activated by the proximity of the user. D2D also serves public safety needs by enabling devices to maintain at least local connectivity even if the radio infrastructure is damaged.

This technology was first introduced by the 3GPP focusing on Public Safety applications and proximity-based services (device discovery). Two scenarios were considered when developing the device-to-device enhancements, in coverage as well as out-of-coverage communication for public safety, and in-coverage discovery of neighbouring devices for commercial use cases, as illustrated in Fig. 11. In release 13, device-to-device communication was further enhanced with relaying solutions for extended coverage. The device-to-device design also served as the basis for the Vehicle-to-Vehicle communication (V2V) and Vehicle-to-Everything communication (V2X) work in release 14.

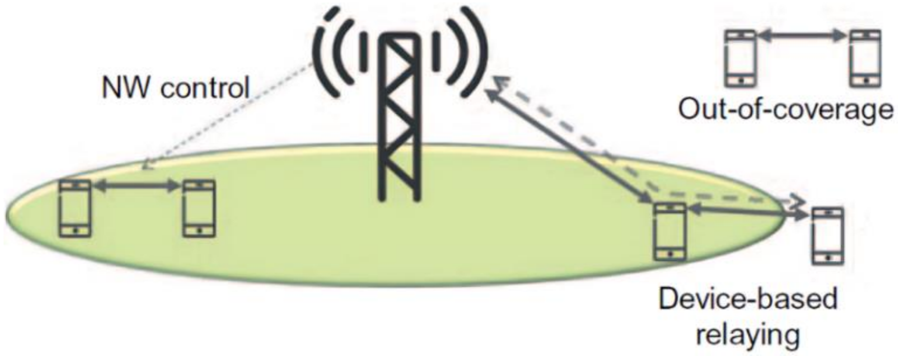

Fig. 11 Device-to-device communication scenario from [7].

2.3.3. Relay Nodes

In 5G networks, relay nodes serve as intermediate points between the UE and base stations, assisting in extending coverage and enhancing network performance. These nodes receive signals from UEs and forward them to base stations, effectively expanding the coverage area and overcoming obstacles such as terrain or building structures. Relay nodes can be strategically placed to fill coverage holes, improve signal strength in remote areas, or enhance capacity in high-traffic zones. Their deployment is particularly advantageous in scenarios where traditional base station installation is impractical or cost-prohibitive. By leveraging relay nodes, 5G networks can achieve broader coverage, better penetration, and improved quality of service for users. Additionally, relay nodes contribute to reducing energy consumption and operational costs by optimizing resource utilization and minimizing the need for additional infrastructure [18]. The growing interest in the relaying concept has led the 3GPP to introduce a new relaying technology called Integrated Access and Backhaul (IAB). This technology provides an alternative to fiber backhaul by extending 5G New Radio to support wireless backhaul.

Traditionally, relays have been utilized as a means of providing wireless backhaul links in situations where the use of wired backhaul is inaccessible or cost-ineffective. However, with the emergence of new services demanding elevated user experience requirements in mobile environments, such as trains and buses, the deployment of mobile relays is necessary and has been considered and become a topic of research [20]. Indeed, with the current technological advances in UEs, UEs with powerful communication and computing capabilities are available. Consequently, the use of UEs as relays has been considered. UE-to-network relaying, where a UE relays another UE's traffic to/from the network in a two-hop link, has been the subject of a 3GPP study item [21]. More recently, it has been included as one of the connectivity models in [22], which discusses various deployment scenarios of relay UEs and defines requirements and key performance indicators. In that respect the implementation of relay UEs in different scenarios is a currant topic of research nowadays. Fig. 12. shows an example of a network deployment scenario considering the use of relay UEs.

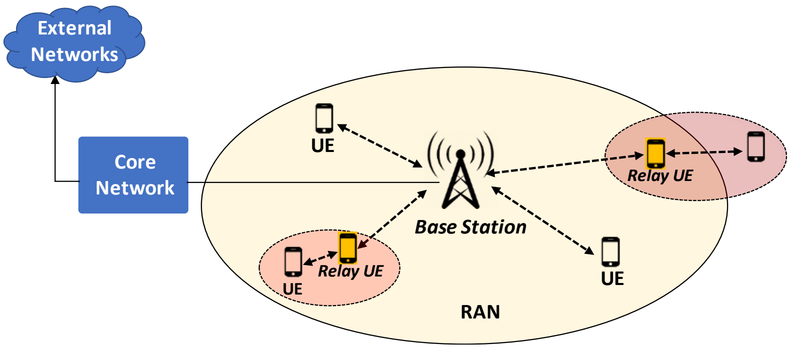

Fig. 12 Example of deployment with Relay UEs

2.4. Artificial Intelligence (AI)

According to [23], AI can be defined as "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings". Similarly, IBM defines AI as "a technology that enables computers and machines to simulate human intelligence and problem-solving skills" [24]. In general, AI is often categorised into weak AI (or narrow AI) and strong AI (or artificial general intelligence - AGI), depending on the "level of intelligence" of the system compared to humans. The current application of AI such as facial recognition, language translation, or playing a specific game like chess, employ week AI. Strong AI, on the other hand, remains a theoretical concept, in this form of AI, machines would have the same intelligence as humans, with self-awareness and consciousness. They would be able to solve problems, learn and plan for the future autonomously. However, achieving strong AI remains a major challenge and has not yet been realised [24].

AI encompasses various fields of study, including machine learning, natural language processing (NLP), text and speech synthesis, computer vision, robotics, planning, and expert systems. Fig. 13. shows a graphical classification of the primary AI domains according to [25].

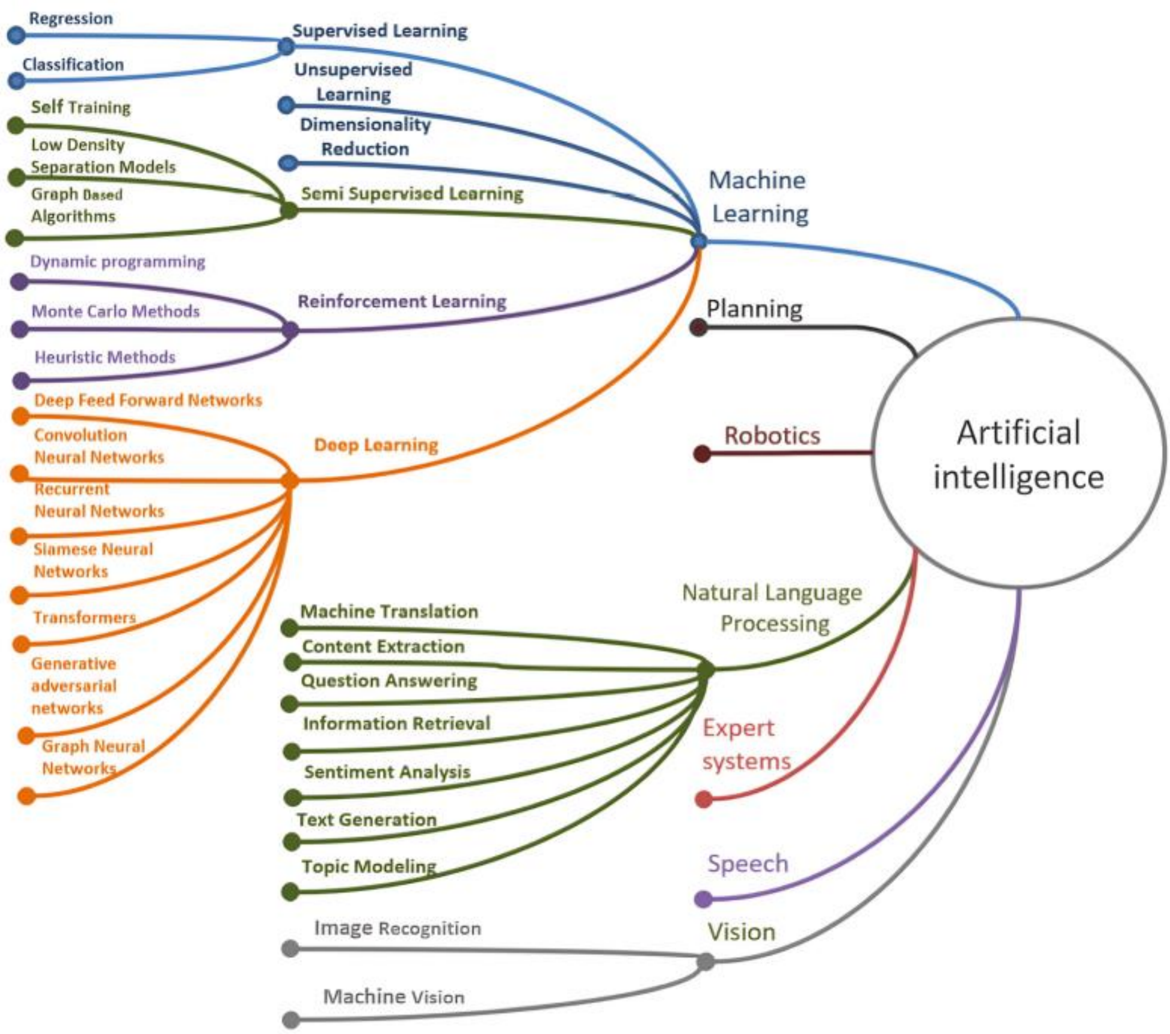

Fig. 13 AI subsections according to [25]

2.4.1. Machine Learning (ML)

According to [26], "Machine learning (ML) is a sub-discipline of AI and computer science that focuses on using data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy". Machine learning methods are categorised into different types based on the learning approach and the purpose of the algorithm. Supervised learning uses labelled data for the training where inputs correspond to desired outputs. By analysing these labelled pairs, the algorithm can make predictions about new, unseen data. In contrast, unsupervised learning works with unlabelled data, where the algorithm is tasked with identifying hidden patterns or structures in the data itself. Semi-supervised learning uses datasets containing both labelled and unlabelled data, guiding the algorithm towards independent conclusions. Finally, reinforcement learning employs a system of rewards and punishments within the dataset, providing feedback to the algorithm to learn from its experience through trial and error.

Due to their importance in this thesis, some AI concepts that are key to a better understanding of the proposed solutions are described below.

2.4.1.1. *Reinforcement Learning (RL)*

According to [27], reinforcement learning (RL) can be defined as "a type of machine learning process that focuses on decision making by autonomous agents". An autonomous agent is any system that can make decisions and act in response to its environment. In reinforcement learning, an autonomous agent learns to perform a task by trial and error and obtains reward according to the quality of the performed action. RL is in essence the relation between an agent, environment and a goal. It is based on the definitions of a Markov decision process (MDP) [28]. A reinforcement learning agent learns about a problem by interacting with its environment. The environment provides data about its current state, which the agent uses to decide on an action. If the action triggers a reward signal from the environment, the agent is motivated to repeat that action in similar future situations. This cycle continues for each new state. Gradually, through rewards and penalties, the agent learns to perform the actions in the environment that meet a defined goal. Fig. 14. Intends to graphically represent the mentioned process.

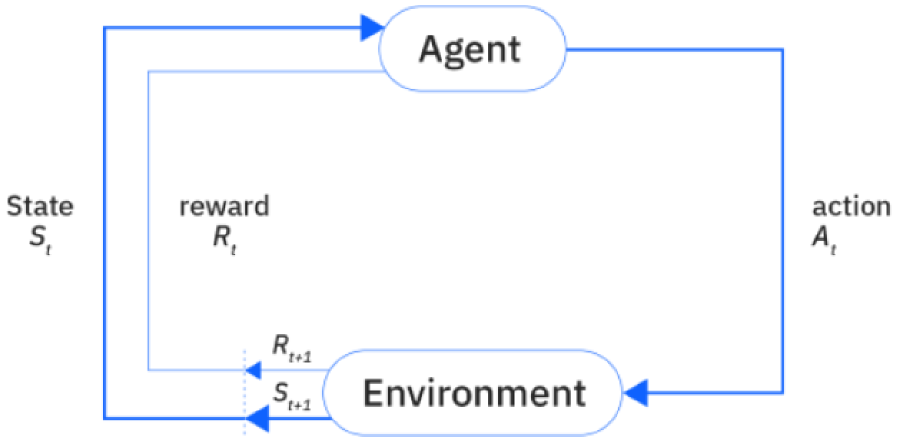

Fig. 14 Graphical representation of an MDP [27].

2.4.1.2. *Deep Learning (DL)*

According to [29] Deep learning "is a sub-domain of machine learning that uses multi-layered neural networks, called deep neural networks, to simulate the complex decision-making power of the human brain". Deep learning uses neural networks arranged in successive layers to learn from data through an iterative process. This approach is particularly valuable when trying to identify patterns in unstructured data, enabling computers to deal with poorly defined abstractions and complex problems.

Deep neural networks are composed of several layers of interconnected nodes, each building on the previous layer to refine and optimise the prediction or categorisation. This process of computation through the network is called forward propagation. The input and output layers of a deep neural network are called the visible layers. The input layer is where the deep learning model receives data for processing, and the output layer is where the final prediction or classification is made. Another key process in deep learning algorithms is the backpropagation used to compute the level of error in the predictions then adjusts the weights and biases of the function by moving backwards through the layers in an effort to train the model. The combination of forward propagation and back propagation allows a neural network to make predictions and correct any errors [29]. Therefore, over time, the algorithm becomes increasingly accurate.

Deep learning is one of the fastest growing sub-domains of AI [30]. Nowadays its application encompasses a wide variety of fields such as law enforcement, financial services, customer services, healthcare. Deep learning is also widely used in the field of mobile networks such as 5G and it is expected to become an essential technology for future 6G networks [31].

2.4.1.3. Clustering Techniques

In machine learning, a cluster is a group or collection of data points that have similar characteristics or patterns of similarity. This similarity can be based on various attributes, characteristics or measurements of the data points. Clustering, as an unsupervised learning technique, is used to group data sets that share similar characteristics. Its goal is to create groups of homogeneous data points from a heterogeneous data set. Clustering evaluates similarity using metrics such as Euclidean distance, cosine similarity, Manhattan distance and others. It then groups the points with the highest similarity scores [32]. There a wide variety of clustering algorithms most of them can be classified inside the following types: *Based on density*: this type of clustering groups data based on the concentration of points. The method relies on density in such a way that it finds the area where the data points are dense and considers it a cluster; *Distribution based*: in this approach, all data points are considered based on their membership in a cluster. That is, there is one point as the centre, and as the data distance from the centre increases, the probability that it is part of this cluster decreases; *Centroid based*: this method separates data points based on multiple centres in the data. Each point is assigned to a cluster based on the square of its distance from the centre.

The clustering techniques can be applied in a wide variety of tasks, according to [33] the principal uses are market segmentation, social network analysis, search result grouping, medical imaging, image segmentation, anomaly detection. However different clustering techniques are also used in the context of 5G and beyond mobile networks (e.g. [34] and [35]).

2.5. O-RAN overview

Open RAN (O-RAN) refers to a new approach to the design and implementation of mobile networks, particularly the RAN. Traditionally, RAN components have been monolithic entities, all-in-one solutions that implement each layer of the cellular protocol stack. They are supplied by a small number of vendors and are perceived by operators as black boxes. Open RAN aims to break this vendor lock-in by standardising the interfaces between network components so that hardware and software from different vendors can interoperate.

In 2018, the O-RAN Alliance was launched as a result of the combination of two initiatives known as the xRAN Forum and the C-RAN Alliance [36]. Its central objective is to specify and standardise the architecture and interfaces required for an Open RAN , based on four fundamental principles:

disaggregation, intelligent RAN controllers, virtualisation and open interfaces [37]. Currently, the O-RAN Alliance has a community of over 300 mobile operators, vendors, and research and academic institutions involved in the radio access network (RAN) industry [38]. In fact, it is projected that the inflection point between traditional mobile networks and Open RAN will occur around 2028 [39].

One of the principal characteristics of O-RAN architecture is the RAN disaggregation, which splits a 5G NR base station (gNB) into different functional units, namely a Central Unit (CU), a Distributed Unit (DU), and a Radio Unit (RU) (called O-CU, O-DU, and O-RU in O-RAN specifications). The O-CU can be further split into two logical components, one for the Control Plane (CP), and one for the User Plane (UP). The O-RAN architecture also considers the use of LTE technology with so-called O-eNB nodes. The O-CU hosts the upper layers of the radio interface protocol stack. These include the PDCP layer that splits the traffic in case of multi-connectivity, in addition to the Radio Resource Control (RRC) and the Service Data Adaptation Protocol (SDAP) layers for control plane and user plane, respectively, that are on top of PDCP. In turn the O-DU hosts the lower layers of the protocol stack, namely the Radio Link Control (RLC), the Medium Access Control (MAC), which hosts the scheduler in charge of allocating the PRBs to the different UEs, and the upper parts of the Physical (PHY) layer (e.g., channel coding, modulation). Finally, the O-RU hosts the lower parts of the PHY layer (e.g., Inverse Fast Fourier Transformation (IFFT) for OFDMA transmission) and the Radio Frequency (RF) functions. To illustrate this, Fig. 15. from [40] shows the evolution of the traditional base station architecture toward a virtualized gNB.

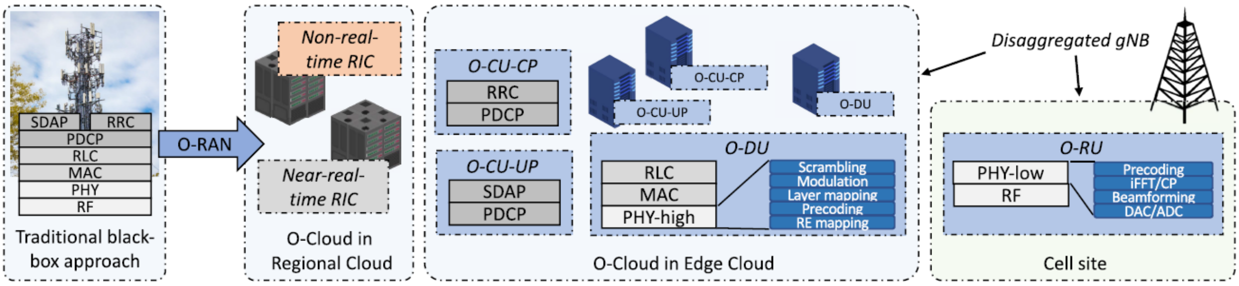

Fig. 15 Evolution of the traditional black-box base station architecture (left) toward a virtualized gNB with a functional split (right, including the CU and DU at the edge, and the RU at the cell site) [40].

A key innovation in O-RAN is the inclusion of programmable elements capable of performing closed-loop optimisation routines to manage the RAN. These are designated as RICs. Specifically, the O-RAN vision includes two logical controllers that have an isolated and centralised view of the network, thanks to data pipelines that stream and aggregate hundreds of Key Performance Measurements (KPMs) on the status of the network infrastructure (e.g. number of users, load, throughput, resource utilisation), as well as additional contextual information from sources outside the RAN. The two RICs process this data and use AI and ML algorithms to determine and apply control policies and actions on the RAN. This effectively introduces data-driven functionalities that can automatically optimise, for example, network and RAN slicing, load balancing, multi-connectivity, handover and scheduling policies, among other functions [41]. The O-RAN Alliance has drafted specifications for a non-real-time RIC that integrates with the network orchestrator and operates on a time scale greater than 1 second, and a near-real-time RIC that drives control loops with RAN nodes on a time scale between 10 milliseconds and 1 second. A central element of the O-RAN framework is the Service Management and Orchestration (SMO), the SMO handles all orchestration, management and automation processes to monitor and control RAN components. The SMO acts as an intermediary between network components. It offers a set of interfaces that enable interaction and data collection, facilitating AI/ML-driven network monitoring and control [40]. Fig. 16. shows the general O-RAN architecture. For more details on each element on the architecture, as well as the different interfaces, see [40] and [42].

In summary, O-RAN is a key technology for 5G networks and future deployments. Currently, many operators around the world have started trials and deployments of O-RAN based technology and solutions, with some in the field trial stage and others already in commercial launch [43]. Similarly, in the research and development of future 6G networks, O-RAN is a technology of major relevance being considered for multiple applications (e.g. [44] and [45]).

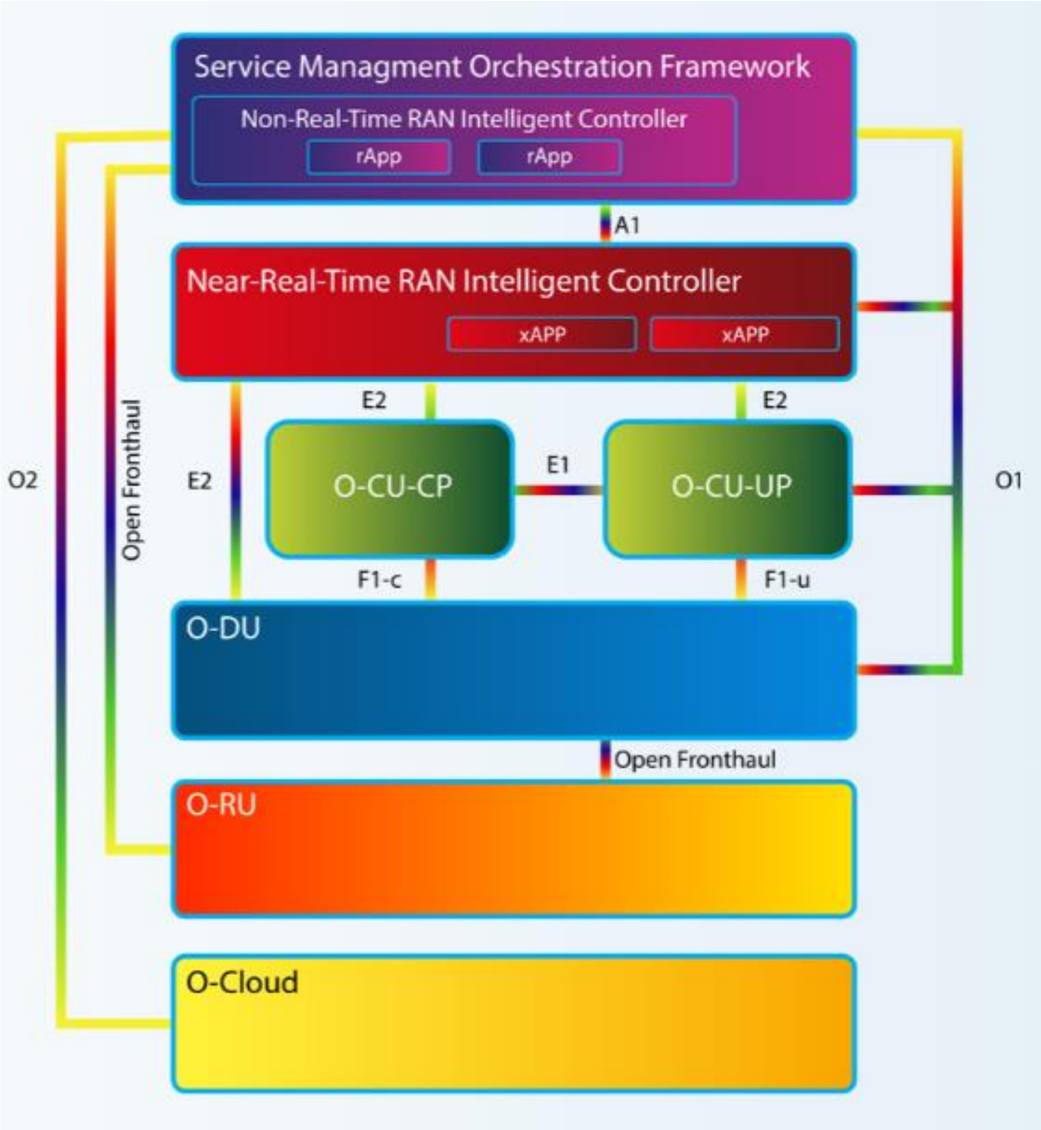

Fig. 16 O-RAN architecture [46].

2.6. Proposed AI-based solutions at RAN level

The concepts presented in the previous sections of this chapter are intended to introduce the reader to the different concepts used in the proposed solutions throughout the thesis, with the aim of ensuring smooth readability and better understanding of the solutions. Chapter 3 presents an AI-based solution for multi-connectivity (MC) optimization. The benefits of an efficient MC process are reflected in improved user experience and network performance. In Chapter 4, the proposed solution focuses on the detection and resolution of coverage-constrained areas in 5G deployments by analyzing large amounts of network performance data. The impact of this solution can be translated into better coverage for network users, thus improving the user experience. Finally, in Chapter 5, we present an AI-based solution for dynamic coverage augmentation using the relaying capabilities of some user equipment (UEs). This solution aims to provide coverage in specific regions at specific times when its deployment is needed and worthwhile. The impact of this solution is to provide an additional option for improving coverage, which has a significant impact on both network and user performance.

Overall, the impact of these different solutions is in line with the main objective of the thesis, which is to provide adequate connectivity conditions in the Radio Access Networks. Fig. 17. provides a graphical representation of the different solutions and their application in the RAN. The full details of each solution are described in the following three chapters.

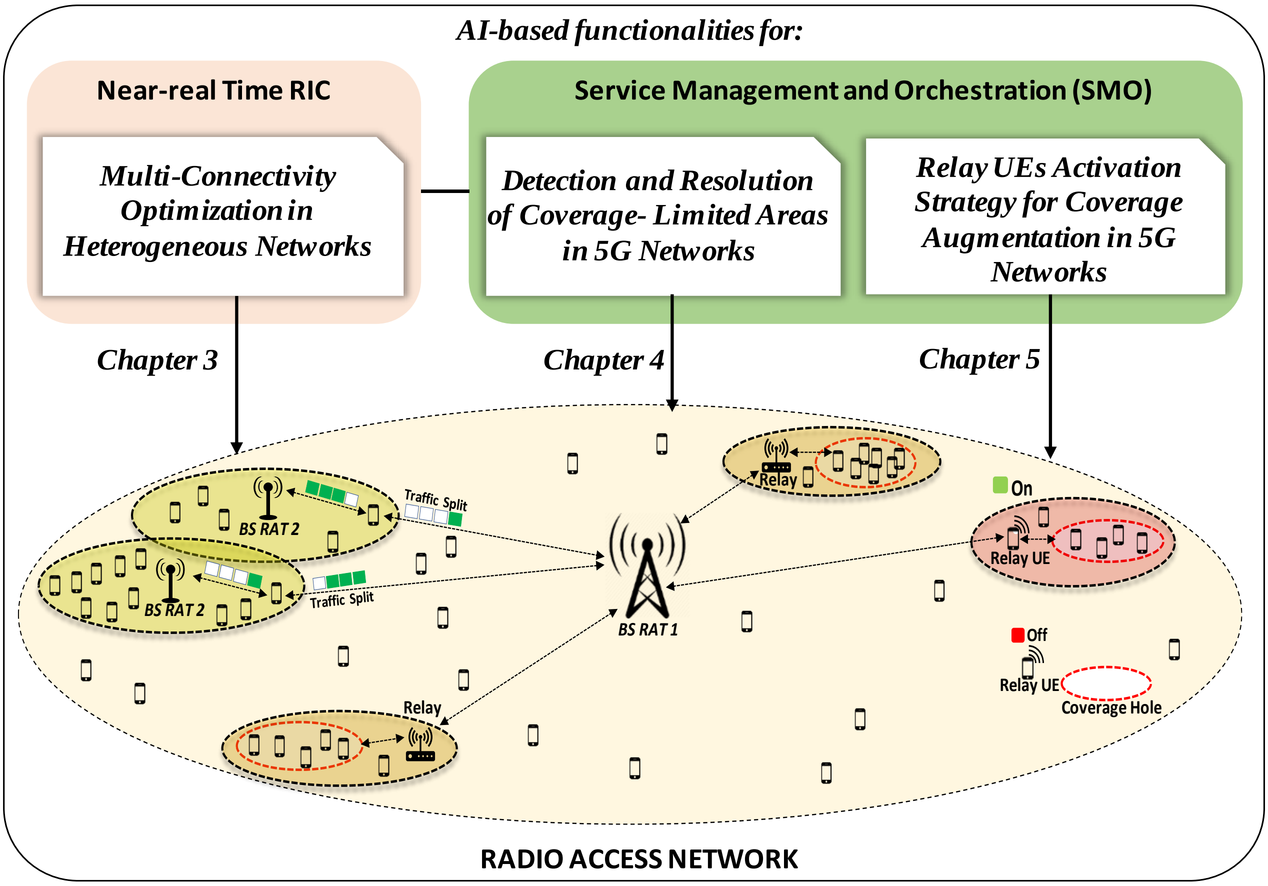

Fig. 17 Diagram of AI-based solutions at the RAN.

Chapter 3. Multi-Connectivity Optimization in Heterogeneous Networks: An AI-Based Approach.

3.1. Introduction

This chapter presents a DRL-based solution aimed at optimising traffic split in heterogeneous network environments, such as LTE+5G NR, where multi-connectivity is employed. The solution intelligently splits traffic from devices to different cells based on current traffic loads and radio conditions experienced, with the primary objective of ensuring QoS satisfaction for network users. The solution involves training a DQN agent to learnt to determine the multi-connectivity (MC) configuration that minimizes the fraction of total bandwidth allocated to the UE, while ensuring its throughput meets or exceeds the required bit rate, and simultaneously avoids congestion in the cells involved in the MC. Furthermore, the solution has been designed to function as a third-party application, known as xApps, within the Open-RAN (O-RAN) architecture framework. Extensive evaluation of the proposed solution has been carried out through system-level simulations under various conditions to provide a comprehensive understanding of its performance.

The rest of the section is organised as follows. Section 3.2 presents a Literature review on the topic as well as the contributions of the solution. Section 3.3 presents the system model and formulates the considered multi-connectivity problem, and Section 3.4 presents the proposed DQN-based solution. Different performance results are provided in Section 3.5. Finally, Section 3.6 presents the concluding remarks.

3.2. Literature Review and Contributions

With the ongoing increase of data traffic as well as the emergence of new services with high speed and reliability requirements, the Mobile Network Operators (MNO) are actively trialling and deploying the fifth generation (5G) networks, in an effort to support new vertical-driven use cases and enhanced user experiences [47]. 5G New Radio (5G NR) deployments are progressively introduced on top of existing legacy technologies, such as Long-Term Evolution (LTE), leading to heterogeneous deployments with multiple radio access technologies (RATs), different cell types, such as macrocells, indoor and outdoor small cells, and operating in a wide range of spectrum bands (e.g. sub 6 GHz bands used by all RATs and millimetre wave (mmW) bands used by 5G New Radio). Network densification has been a mainstay of LTE networks and the need for small cells will be even more critical in 5G and beyond systems due to operation in higher spectrum bands and the need to support traffic densities that are two to three orders of magnitude higher than in LTE [22]. The general industry consensus is that 5G will drive hyperdense deployments with site densities in excess of 150 sites/km^2 in urban and selected indoor scenarios [48].

In this context, multi-connectivity (MC) (see section 2.3.1) emerges as a key technology for supporting simultaneous access via LTE and 5G networks [49], and is expected also to play a key role in further network evolutions towards 6G. In the Third Generation Partnership Project (3GPP), MC is specified through the Multi-Radio Dual Connectivity (MR-DC) feature defined in [16] that considers different options depending on the technology used by the MN and by the SN and on the core network technology.

From an algorithmic perspective, the literature has considered different problems in relation to MC, such as resource allocation [50] and [51], cell and RAT selection [52] and [53], and traffic split [54]-[57]. Concerning resource allocation, a Smart Aggregated RAT Access (SARA) strategy is proposed in [50] for joint RAT selection and resource allocation in a scenario with cellular base stations and Wi-Fi access points. The solution makes use of a Semi Markov Decision Process (SMDP)-based hierarchical decision framework (HDF). In [51] the optimization problem of resource allocation in a MC scenario with 5G and LTE is formulated. Then, a solution based on two heuristic algorithms is proposed, namely a base station selection algorithm performed by the UE and a resource block algorithm executed by the base station.

The problem of secondary base station selection in MC with 5G/LTE is addressed in [52], which presents different algorithms aimed at improving robustness and performance while minimizing the energy consumption. In turn, [53] considers an LTE/WiFi scenario and proposes a network switching strategy based on a Markov Decision Process (MDP) and a value iteration algorithm to determine the set of RATs that a user is connected to in each handover window.

Finally, the problem of traffic split between different RATs is considered in [54] focusing on an LTE/Wi-Fi scenario. The paper assumes a heterogeneous network controlled by a single coordinating node that collects relevant information, decides on the best choice of RAT for all users, and advices on the actual amount of radio resources that every user may utilize on each technology. The problem is formulated analytically and a solution based on the weighted max-min algorithm is proposed. In [55] the problem of traffic split between the master and the secondary eNB in LTE with dual connectivity is modelled as a Constrained Markov Decision Process and a solution based on the Lagrangian approach is proposed. Similarly, [56] considers a scenario with 5G-LTE multi-connectivity and makes use of Lagrangian Dual Decomposition to determine the fraction of traffic transmitted through each cell that maximizes the goodput, and [57] formulates a PDCP split bearer decision problem that decides whether and how to split the traffic across multiple cells in order to meet the bandwidth requirements of user services and proposes a heuristic solution to solve the problem.

The model proposed in this chapter addresses the traffic split multi-connectivity problem in multi RAT scenarios. The target is to determine a policy to optimally distribute the traffic of a UE across the different RATs and cells by fulfilling the QoS requirements while minimizing the resource consumption of the UE and ensuring that no overload situations arise in the involved cells. The main novelty of the proposed solution with respect to prior works [54]-[57] is the use of Deep Reinforcement Learning (DRL), and in particular the Deep Q-Network (DQN) technique [58], in order to learn the traffic split policy to be applied on a per UE basis so that the amount of bandwidth used by a UE is minimized while at the same time providing the required bit rate and avoiding overload situations in the involved cells. At the time of developing the presented solution and writing the corresponding report, to the authors' knowledge, the use of DQN had not yet been considered by other researchers in the context of MC. Instead, previous works have considered different algorithmic solutions and different optimization targets, such as linear programming for ensuring fairness in [54], Lagrangian multipliers for minimizing delay in [55], Lagrangian Dual Decomposition for optimizing goodput in [56] or a heuristic algorithm for maximizing the number of served UEs in [57]. Following the trend of applying machine learning for different problems in the RAN [59] and for different applications [60], the motivation to consider DRL in MC is that DRL techniques are useful for optimizing dynamic decision-making problems in the absence of an accurate mathematical model of the operational environment. Moreover, thanks to their capability of generalizing from past experience, DRL techniques are efficient in problems that depend on a large number of input variables and in which both the inputs and the decision-making outputs can take a large range of possible values, as it is the case of the MC problem considered here.

The model presented in this chapter is based on two earlier works, available at [61] and [62]. The model presented in the chapter significantly extends these works by presenting the detailed formulation and algorithmic solution of the DQN model, the architectural framework based on the open RAN (O-RAN) architecture [40] for supporting the proposed solution, and by providing a much more exhaustive performance assessment of the algorithm considering different evaluation conditions.

3.3. Multi-connectivity model formulation

3.3.1. Problem definition

Let us consider a heterogeneous RAN where different UEs with multi-connectivity capabilities are camping. A given UE u considers M different RATs and N different cells per RAT as candidates for the multi-connectivity. Then, let us denote as $A_u=\{C_{m,n}\}$ the set of candidate cells detected by the UE u, where $C_{m,n}$ denotes the n-th cell of the m-th RAT with n=1, ..., N and m=1,..., M. It is worth mentioning that, due to the mobility of the UE, the specific detected cells in a given RAT may change with time. In this respect, it is assumed that the N candidate cells of a RAT correspond to the best N cells detected by the UE at a certain time based on measurements averaged during a time window of duration ΔT s.

Through the use of multi-connectivity, the traffic of the u-th UE is split across multiple RATs/cells of the set A_u. It is assumed that, at a certain time, the UE can be simultaneously connected to a maximum of N_{max} cells among the $M \cdot N$ candidates. The multi-connectivity configuration of the u-th UE can be expressed as the $M \times N$ matrix $\mathbf{B_u}=\{\beta_{m,n}\}$, where $\beta_{m,n}\in[0,1]$ defines the fraction of total traffic of UE u that is delivered through the n-th cell of the m-th RAT. Then, the objective is to find the optimal configuration $\mathbf{B_u}=\{\beta_{m,n}\}$ to be applied in a time window of ΔT that allows ensuring the Quality of Service (QoS) requirements with minimum resource consumption and avoiding overload situations in the different RATs/cells. In this respect, it is assumed that the QoS requirement of the user u is expressed in terms of a required bit rate R_u (b/s) to be provided.

To formalize the problem, let us denote as $T_u(\mathbf{B_u})$ the total throughput or bit rate obtained by user u during the last time window period ΔT as a result of the multi-connectivity configuration $\mathbf{B_u}$. Let us also denote $a_{m,n}(\beta_{m,n})$ as the number of physical resources in the m-th cell and n-th RAT assigned to the u-th UE to transmit the traffic corresponding to fraction $\beta_{m,n}$. Assuming LTE and 5G NR-based RATs, the physical resources correspond to the Physical Resource Blocks (PRBs), each one defined as a set of 12 contiguous subcarriers in an Orthogonal Frequency Division Multiple Access (OFDMA) access [48]. Then, considering that $b_{m,n}$ is the bandwidth of one PRB in the m-th cell and n-th RAT, the bandwidth allocated to the user u in this RAT, denoted as $\gamma(\beta_{m,n})$, is given by:

$$\gamma(\beta_{m,n}) = a_{m.n}(\beta_{m,n}).b_{m.n} \tag{1}$$

In addition, the total fraction of occupied PRBs in a RAT/cell accounting for all the UEs connected to that cell is denoted as $\rho_{m,n}(\beta_{m,n})$.

With all the above, the considered problem to be solved for the u-th UE is formally defined as:

$$\begin{gathered}\mathbf{B_u} = \underset{\mathbf{B_u}}{arg\ min}\left[\frac{1}{w_{max}}\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma(\beta_{m,n})\right] \\ \text{s.t.}\quad T_u(\mathbf{B_u}) \geq R_u\ ,\quad \rho_{m,n}(\beta_{m,n}) \leq \rho_{max}\quad \forall m,n \\ \sum_{m=1}^{M}\sum_{n=1}^{N}\beta_{m,n} = 1\end{gathered} \tag{2}$$

where w_{max} is the maximum possible bandwidth to be assigned to the user u and $\rho_{max} \in [0,1]$ is the maximum threshold established to avoid overload situations in a cell. Then, the problem in (2) intends to find the multi-connectivity configuration $\mathbf{B_u}$ that minimizes the fraction of total bandwidth allocated to the UE u while at the same time ensuring that the provided throughput is above the required bit rate, i.e. $T_u(\mathbf{B_u}) \geq R_u$ and that the total fraction of occupied PRBs in the involved RATs/cells is lower than a threshold ρ_{max}.

3.3.2. O-RAN based System Architecture for MC Configuration

In order to enforce in the network, the multi-connectivity configuration B obtained as a result of the above problem, the presented model proposes the system architecture depicted in Fig. 18. It is based on the O-RAN Alliance reference architecture [40], which complements 3GPP 5G standards with a foundation of virtualized network elements, white-box hardware and standardized interfaces that fully embrace the core principles of openness and intelligence. See section 2.5 for more details on O-RAN technology.

The multi-connectivity situation illustrated in Fig. 18. considers the downlink traffic transmitted to a UE served by two cells of 5G NR RAT m=1. The cell n=1 is at the MN while the cell n=2 is at the SN. The multi-connectivity configuration for the different users is determined by an MC controller, which can be hosted at the near-real time RAN Intelligent Controller (near-RT RIC) of the O-RAN architecture. The near-RT RIC is deployed at the edge of the network, is able to operate control loops with a periodicity between 10 ms and 1 second, and it can interact with the DUs and CUs in the RAN. A relevant characteristic of the near-RT RIC is that it supports the execution of third-party applications, referred to as *xApps* [63]. Then, the MC controller can be implemented as one of these *xApps*.

The inputs of the MC controller are different measurements (further explained in next section) from the RAT/cells as shown in Fig. 18. These measurements can be sent to the near-RT RIC through the E2 interface. At the near-RT RIC, they are sent to the *xApp* MC controller where they are collected and processed in order to be adapted to the format required by the DQN agent that constitutes the core of the MC controller and will be explained in section 3.4. The output of the MC controller is the configuration $\mathbf{B_u}$ $=\{\beta_{m,n}\}$ with the weights $\beta_{m,n}$ to be configured at the PDCP layer of the MN, which resides in the O-CU. This configuration can be established via the E2 interface that supports enabling, disabling or modifying the dual connectivity process [64]. Based on this configuration, the PDCP splits the traffic between the MN and the SN and delivers the part of the traffic of the SN via the *Xn* interface that interconnects the MN and the SN. Then, the MAC scheduler at each O-DU will allocate the necessary amount of bandwidth resources $\gamma(\beta_{m,n})$ to the UE to transmit the fraction of traffic $\beta_{m,n}$ corresponding to the cell. The specific design of the MAC scheduler is out of the scope of this work, but in general it will consider aspects such as the instantaneous propagation and interference conditions observed by the UE, the QoS requirements, the amount of UEs in the cell, etc.

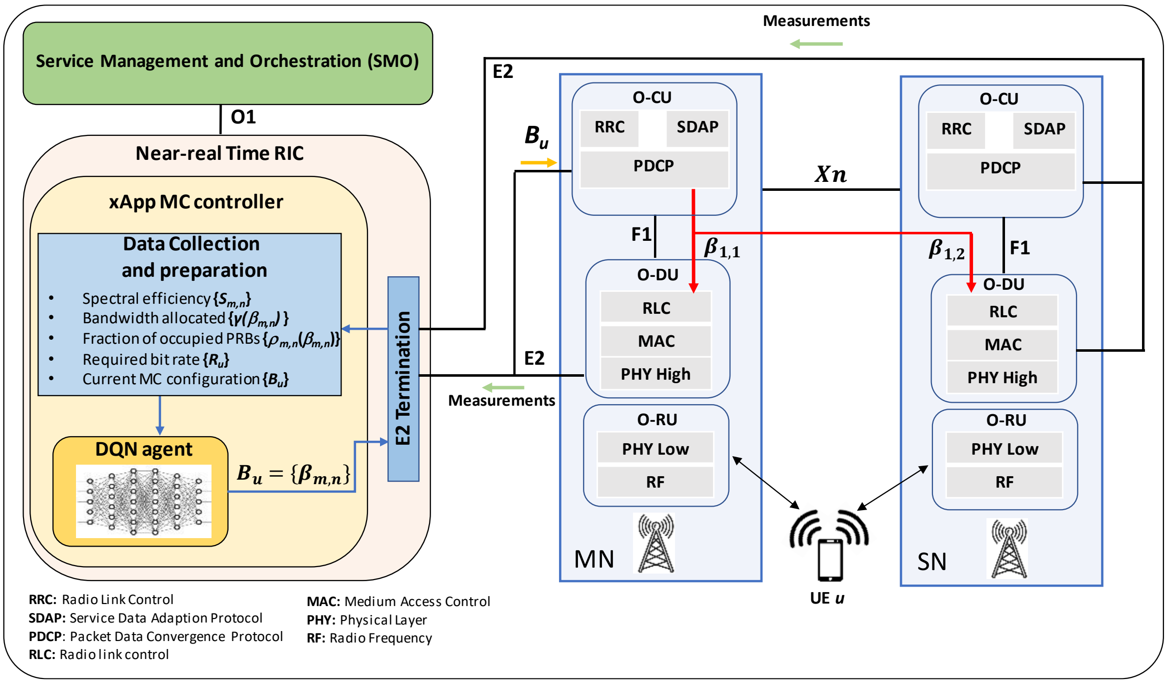

Fig. 18. O-RAN-based architectural components for supporting the MC configuration

3.4. Deep-Q Network Approach for Multi-connectivity Optimization

To address the problem defined by (2) a large number of variables must be considered, such as the propagation and interference conditions experienced by the UE in the links with the different cells/RATs, the existing load in each cell, the QoS requirements, etc. Additionally, there is also a dependence of the behaviour of the MAC layer in each cell/RAT that determines the amount of resources allocated to the UE. However, since the MC controller operates on top of the different RATs, in general it will not have a precise model of how these resource allocation techniques work to determine the value of $\gamma\left(\beta_{m,n}\right)$ and its impact on the QoS metrics. Then, given the complexity and the multiplicity of inputs to the multi-connectivity problem, DRL techniques are considered as a solid candidate for approaching it. Particularly, this proposed solution relies on the DQN algorithm, which is a model-free and value-based DRL algorithm that considers discrete action spaces. This algorithm has been already successfully applied to other problems in the RAN such as capacity sharing in [65], resource allocation in heterogeneous networks in [66] or spectrum sharing in 4G/5G networks in [67].

Other more sophisticated techniques, such as Double DQN (DDQN) or Deep Deterministic Policy Gradients (DDPG), could be considered to overcome the overestimation of the Q values in DQN. A previous work of the authors in the area of capacity sharing [68], where DQN was compared against DDQN and DDPG, suggested retaining DQN for this work. While no significant differences in terms of performance would be expected, practicality considerations such as the speed of the training process, and the number of hyperparameters to configure, may favour DQN.

In the proposed approach, the learning is a dynamic process carried out by the DQN agent located at the MC controller, which makes decisions on the multi-connectivity configurations for the different UEs. The agent operates in discrete times with granularity equal to the time window duration ΔT. These discrete times are denoted as t, t+1, ..., t+k,... At time t the DQN selects an action $\boldsymbol{a}(\boldsymbol{t})$ that contains the MC configuration to be applied for a given UE in the next time window. The action selection is based on the current state at time t, denoted as $\boldsymbol{s}(\boldsymbol{t})$**,** and on the decision-making policy available at this time. Then, as a result of applying the selected MC configuration a reward signal $r(t+1)$ is provided to the DQN agent at the end of the time window. This reward signal measures how good or bad was the last performed action $\boldsymbol{a}(\boldsymbol{t})$ according with the considered optimization criteria and, in consequence, this obtained signal is used to improve the decision-making policy. The different components of this process are detailed in the following

3.4.1. State, Action and Reward Definition

The state $\boldsymbol{s}(t)$ is a vector that includes the following components for a given UE u:

- Requirement of UE u: $R_{u.}$
- Spectral efficiency per RAT/cell $\{S_{m,n}\}$ of UE u.
- Current configuration $\mathbf{B_u}=\{\beta_{m,n}\}$, which corresponds to the configuration applied at time t-1 to the UE u.

- Fraction of occupied bandwidth resources by the UE u in each RAT/cell $\{\gamma(\beta_{m,n})\}$.
- Fraction of total occupied bandwidth resources in each RAT/cell $\{\rho_{m,n}(\beta_{m,n})\}$.

All the values $S_{m,n}$ **,** $\gamma(\beta_{m,n})$ and $\rho_{m,n}(\beta_{m,n})$ are average values measured during the last time window of duration ΔT, i.e. between discrete times t-1 and t. Notice that the state has a total of $1+4\cdot N\cdot M$ components.

Each action $\boldsymbol{a}(\boldsymbol{t})\in\boldsymbol{\mathcal{A}}$ represents a matrix $\mathbf{B_u}=\{\beta_{m,n}\}$ that corresponds to the MC configuration to be applied during the next time window ΔT. The action space $\boldsymbol{\mathcal{A}}$ includes all the possible MC configurations and is defined considering that the possible $\beta_{m,n}$ values are discretized with granularity $\Delta\beta$ and that the aggregate of all $\beta_{m,n}$ values in matrix $\mathbf{B_u}$ equals 1. Moreover, the action space also considers that a UE can be connected to a maximum of N_{max} cells. Therefore, at most N_{max} values of $\beta_{m,n}$ can be different from 0 in a certain action.

The reward $r(t+1)$ intends to measure how good or bad was the performance obtained by the last action $\boldsymbol{a(t)}$ for the state $\boldsymbol{s}(t)$ in relation to the target of the optimization. Then, considering the optimization problem (2), and that the last action $\boldsymbol{a(t)}$ is given by MC configuration $\mathbf{B}=\{\beta_{m,n}\}$, the reward is defined as:

$$r(t+1)=\left(1-\frac{1}{w_{max}}\sum_{m=1}^{M}\sum_{n=1}^{N}\gamma\left(\beta_{m,n}\right)\right)\cdot\min\left(1,\frac{T_u(\mathbf{B})}{R_u}\right)\cdot\prod_{\substack{m,n\\ \beta_{m,n}>0}}\min\left(1,\frac{\rho_{max}}{\rho_{m,n}\left(\beta_{m,n}\right)}\right) \qquad (3)$$

The first term in $r(t+1)$ captures the total bandwidth assigned to the UE u in all the cells/RATs, so the lower the amount of bandwidth assigned the higher will be the reward, and this reflects a better fulfilment of the optimization target in (2). The second term multiplicative represents a penalty introduced when the achieved throughput $T_u(\mathbf{B_u})$ is lower than the minimum requirement R_u, corresponding to the first constraint in (2). The last multiplicative term introduces a penalty for each cell/RAT in which the UE has transmitted traffic (i.e. $\beta_{m,n}>0$) and the cell is overloaded, thus capturing the second constraint in (2). Note that the values of $\gamma(\beta_{m,n})$, $\rho_{m,n}(\beta_{m,n})$ and $T_u(\mathbf{B_u})$ correspond to the averages obtained during the time window ΔT between discrete times t and $t+1$.

3.4.2. Deep-Q Network Agent training stage

The training process is used to dynamically learn the decision-making policy π that the DQN agent will use when selecting the different actions as a function of the current state. For this purpose, the DQN agent executes the DQN algorithm of [58] but particularized to the state, action and reward signals presented above.

In general, RL algorithms aim at finding the optimal policy π^* that maximizes the discounted cumulative future reward (i.e. $\sum_{j=0}^{\infty}\tau^k r(t+j+1)$, where τ is the discount rate ranging $0\leq\tau\leq1$). In value-based RL algorithms, such as DQN, this is done by obtaining the optimum action-value function $Q^*(\boldsymbol{s},\boldsymbol{a})$, which is the maximum expected discounted cumulative reward starting at time t from state $\boldsymbol{s}$, taking the action $\boldsymbol{a}$ and following the policy π, that can be expressed in a recursive form by the Bellman equation as:

$$Q^{*}(s,a) = E\left[r(t+1) + \tau \cdot \max_{a'} Q^{*}\left(s(t+1),a'\right) \middle| s(t) = s, a(t) = a,\ \pi \right] \tag{4}$$

Given $Q^{*}(\boldsymbol{s}, \boldsymbol{a})$, the optimum policy is defined by greedily selecting the action $\boldsymbol{a}$ with the highest value for each state $\boldsymbol{s}$, that is:

$$\pi^{*} = \arg\max_{a} Q^{*}(s,a) \tag{5}$$

In order to determine the optimum action-value function, DQN approximates $Q^{*}(\boldsymbol{s},\boldsymbol{a})$ by a deep neural network (DNN) with weights θ, denoted as $Q(\boldsymbol{s},\boldsymbol{a},\theta)$, which is progressively updated during the training process. This DNN takes as input the state $\boldsymbol{s}$ and provides as output the value for each possible action $\boldsymbol{a}$ in accordance with the weights θ, which define the interconnections between the different neurons. For updating $Q(\boldsymbol{s},\boldsymbol{a},\theta)$, the DQN agent is composed of the following elements:

- Evaluation DNN $Q(\boldsymbol{s},\boldsymbol{a},\theta)$: This is the main approximation of the optimum action-value function $Q^{*}(\boldsymbol{s},\boldsymbol{a})$. It is used to determine the decision-making policy π for selecting the actions, as:

$$\pi = \arg\max_{a} Q\left(s,a,\theta\right) \tag{6}$$

- Target DNN $Q(\boldsymbol{s}, \boldsymbol{a},\theta^{-})$: This is another DNN with the same structure as the evaluation DNN but with weights θ^{-}. It is used to obtain the Time Difference (TD) target $r(t+1) + \tau \max_{a'} Q\left(s(t+1),\ a',\ \theta^{-}\right)$ that is used for making successive updates of the evaluation DNN during the training. Moreover, this DNN is updated every P time steps (i.e. time windows) with the weights of the evaluation DNN, i.e. $\theta^{-}=\theta$.

- Experience dataset D: This is a dataset of length l that stores the experiences of the DQN agent. The stored experience at time t is represented by the tuple $<\boldsymbol{s}(t), \boldsymbol{a}(t), r(t+1), \boldsymbol{s}(t+1)>$, which captures the state at t, the action taken, and the resulting reward and new state at time $t+1$. The stored experiences are randomly selected during the training process to update the weights θ.

At initialization, the weights of both the evaluation and target DNNs are randomly selected. Then, they are updated as a result of the training process of the DQN agent, which is divided in two parts: the data collection and the update of weights θ.

The data collection consists in gathering experiences and storing them in the experience dataset D. For each time t, the DQN agent observes the state of the environment $\boldsymbol{s}(t)$ for a given UE and, accordingly, it triggers an action $\boldsymbol{a}(t)$ based on an ε-Greedy policy that chooses actions according to the policy π in (6) with probability 1- ε and a random action with probability ε. Then, the reward $r(t+1)$ is collected and the experience tuple $<\boldsymbol{s}(t), \boldsymbol{a}(t), r(t+1), \boldsymbol{s}(t+1)>$ is stored in the dataset. When the dataset is full (i.e. l experiences are stored), old experiences are removed from the dataset to save new ones. It is worth mentioning that during a number of InitialCollectSteps of the data collection, the actions are selected completely randomly by forcing ε=1 in order to explore several states and start filling the dataset with experiences.

The process of updating the weights θ of the evaluation DNN is executed in every time window in parallel to the data collection and it makes use of the experiences stored in the experience dataset. Specifically, for each update a mini-batch $U(D)$ of J past experiences is firstly selected randomly from the dataset. The selected experiences are denoted as e_j, j=1,...,J, and the components of tuple e_j are denoted as $<\boldsymbol{s_j},\boldsymbol{a_j},r_j,\boldsymbol{s_j}^*>$. Then, the update is performed based on the mini-batch gradient descent process. First, it computes the average mean squared error (MSE) loss $L(\theta)$ over all the J experiences in the mini-batch as:

$$L(\theta) = \underset{e_j \in U(D)}{E}\left[\left(r_j + \tau \cdot \max_{a'} Q\left(s_j^{*}, a', \theta^{-}\right) - Q\left(s_j, a_j, \theta\right)^2\right)\right] \tag{7}$$

Then, the mini-batch gradient descent of $L(\theta)$, denoted as $\nabla L(\theta)$, is obtained by differentiating $L(\theta)$ with respect to θ, which yields:

$$\nabla L(\theta) = \underset{e_j \in U(D)}{E}\left[\left(r_j + \tau \cdot \max_{a'} Q\left(s_j^{*}, a', \theta^{-}\right) - Q\left(s_j, a_j, \theta\right)\right) \cdot \nabla_{\theta} Q\left(s_j, a_j, \theta\right)\right] \tag{8}$$

Then, the weights of the evaluation DNN $Q(\boldsymbol{s},\boldsymbol{a},\theta)$ are updated as:

$$\theta \leftarrow \theta + \alpha \cdot \nabla L(\theta) \tag{9}$$

where α denotes the learning rate.

After each update of θ, the resulting $Q(\boldsymbol{s}, \boldsymbol{a},\theta)$ will be used for triggering new actions. Moreover, after P updates of θ, the weights of the target DNN are updated as $\theta^{-}=\theta$.

The training operation of the DQN-agent associated to UE u is summarized in Algorithm I, which includes the data collection (lines 3-12) and the update of the weights θ of the evaluation DNN (lines 13-21 of Algorithm I). The training duration in time steps is given by parameter MaxNumberOfTrainingSteps.

Algorithm I. DQN training for UE $\boldsymbol{u}$

```
1 Initialize DNN counter p=0.
2 For t=0… MaxNumberOfTrainingSteps
3    Collect state s(t) (see section 3.4)
4    Generate random ε’ (ε’=1 for the initial steps).
5    If ε’<ε
6      Choose randomly action a(t)
7    Else
8      Obtain action according to π.
9    End if
```

Obtain reward $r(t+1)$ and $s(t+1)$ as a result of action $\boldsymbol{a(t)}$
If D is full (l samples are stored)
    Remove the oldest one.
Store experience < $\boldsymbol{s(t)}, \boldsymbol{a(t)}, r(t+1), \boldsymbol{s(t+1)}$ > in D
Randomly sample a minibatch of experiences $U(D)$ from D of length J
Compute the loss function $L(\theta)$
Compute the mini-batch gradient descent $\nabla L(\theta)$
Update weights θ of evaluation DNN.
If p==P
  Update the weights of target DNN $\theta^{-}=\theta$ and set p=0.
Else
  p=p+1
End if
End for

3.5. Performance Evaluation

This section evaluates the performance of the proposed solution by means of system level simulations. After describing the considered scenario for the training and evaluation as well as the relevant parameters for both stages in section 3.5.1, the evolution of training process in order to obtain the policy is presented in section 3.5.2. Section 3.5.3 describes the considered benchmarking strategies for assessing the performance of the proposed MC strategy. Then, sections 3.5.4 provides the obtained performance results.

3.5.1. Scenario Description

The considered scenario is a square area of 500 m x 500 m, composed by four 5G NR cells and two LTE cells. As explained in section 3.3, $C_{m,n}$ denotes the n-th cell of the m-th RAT, being m=1 for LTE and m=2 for 5G NR. Therefore, the LTE cells are identified as $C_{1,1}$ and $C_{1,2}$ and the 5G NR cells are $C_{2,1}$, $C_{2,2}$, $C_{2,3}$, $C_{2,4}$. The relevant parameters of the cells are presented in Table 1. Fig. 19. illustrates the position of the different cells. The positions of the LTE and 5G NR cells were selected to illustrate a scenario in which 4 5G NR microcells are deployed in a denser area associated with a traffic hotspot, complementing the deployment of 2 LTE macrocells for wider coverage footprint.

The scenario assumes a non-homogeneous traffic distribution with UEs that support MC and other UEs that generate additional background traffic. The UEs that support MC follow specific trajectories moving at 1m/s along the scenario and have an active session during the whole simulation duration with a required bit rate R_u=50 Mb/s. These UEs can connect to up to N_{max}=2 cells from any of the two RATs. The background traffic generation assumes Poisson session arrivals with aggregate generation rate 0.6 sessions/s and exponentially distributed session duration with average 120s. As a result, the number of backgrounds UEs vary randomly during a simulation and the average number is 0.6·120 = 72 background UEs. A background UE remains static during a session. 50% of the background UEs are randomly located inside a square hotspot of 250 m x 250 m centered at the middle of the scenario (see Fig. 19). The rest of background UEs are randomly distributed in the whole scenario. Background UEs connect to the RAT/cell with the highest Signal to Interference and Noise Ratio (SINR). To capture the different bit rates achievable in the two technologies, when a background UE is connected to LTE, its serving cell allocates the needed resource blocks (RBs) to achieve a bit rate of 2.5 Mb/s, and when it is connected to 5G NR, the allocation is to achieve a bit rate of 40 Mb/s.

The DQN model was developed in Python using the TF-agents library [69]. The DQN model parameters are detailed in Table 2. They were selected after conducting different tests of the algorithm with different configurations, then choosing a suitable configuration with satisfactory behaviour in terms of, for example, reward performance, convergence, and stability.

TABLE I. CELLS CONFIGURATION PARAMETERS FOR MC SOLUTION

Cell configuration parameters		
Parameter	Value	
Type of RAT	LTE	5G NR
Cells position [*x*, *y*] m	[62, 250] [437,250]	[187, 125] [187,375] [312,125] [312,375]
Frequency	2100 MHz	26 GHz
Subcarrier separation	15 kHz	60 kHz
Nominal channel bandwidth	20 MHz	50 MHz
Number of available PRBs	100	66
Base station transmitted power	49 dBm	21 dBm
Base station antenna gain	5 dB	26 dB
Base station height	25 m	10 m
UE antenna gain	5 dB	10 dB
Overload threshold ρ_{max}	0.95	0.95
UE noise figure	9 dB	
UE height	1.5 m	
Path loss model	Model of Sec 7.4 of [70]	
W_{max}	95.04 MHz (corresponds to the case when MC is done with 2 cells of 5G NR)	

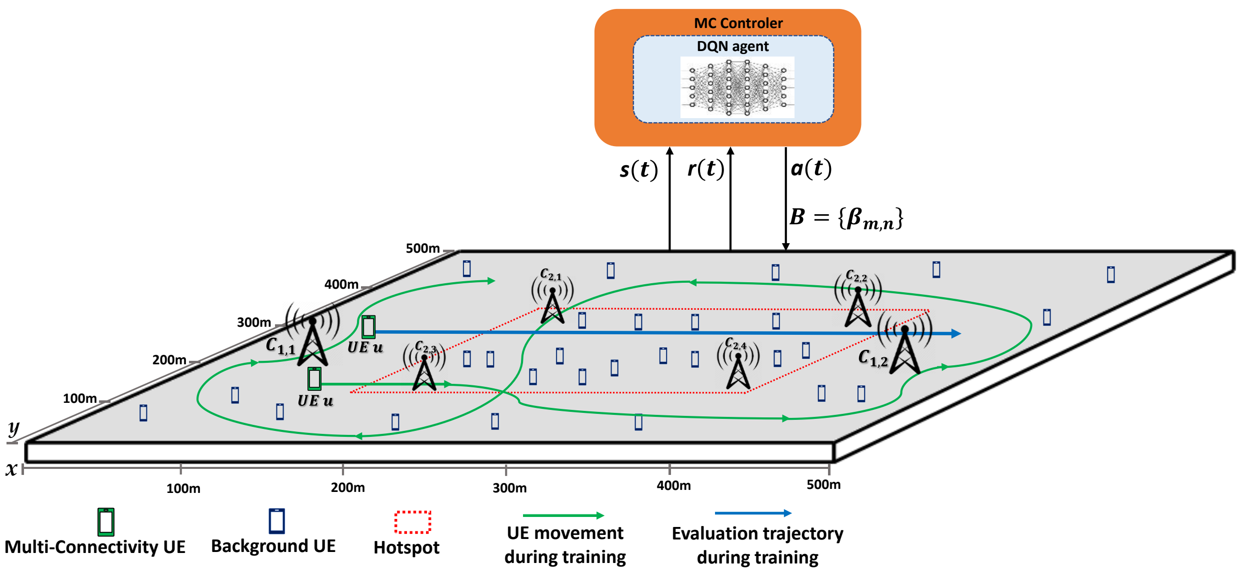

Fig. 19. Graphic representation of the scenario used for training/evaluation

TABLE II. DQN ALGORITHM CONFIGURATION PARAMETERS FOR MC SOLUTION

DQN algorithm parameters	
Parameter	Value
Initial collect steps	5000
MaxNumberOfTrainingSteps	1e6
Experience Replay buffer maximum length (l)	1e5
Mini-batch size (J)	256
DNN updating period (P)	2500 sec
Discount factor ($\boldsymbol{\tau}$)	0.9
Learning rate (α)	0.0003
ε value (ε-Greedy)	0.1
DNN architecture	Input layer: 17 nodes Two hidden layers: 100 and 50 nodes Output layer: 58 nodes
Time window (ΔT)	1 sec
Granularity $\Delta\beta$	0.1

3.5.2. Training Evolution

The training process of the DQN algorithm is performed by considering a MC-capable UE moving along the scenario following trajectories according to random walk and with required bit rate R_u=50 Mb/s, while at the same time background UEs also generate traffic as explained previously. The DQN agent decides the MC connectivity configuration of the UE and, based on the obtained rewards, the decision-making policy is progressively updated as explained in section 3.4.2. Fig. 19. intends to graphically represent the scenario used for conducting the training and evaluation processes. At the beginning of the training the UE is located in the coordinates [X_1=50, Y_1=450]. Then, it moves with speed of 1 m/s following a random walk model, in which it changes the direction (between +/- $\frac{\pi}{4}$) with probability $p_{dir} = 1/20$ at each time step. The green arrows in Fig. 19 intend to exemplify this process.

The training is executed until reaching the maximum number of training steps MaxNumberOfTrainingSteps=10^6. In order to illustrate the evolution of the learning process, the policy of the DQN agent is obtained every 2500 training steps and is applied to an evaluation scenario in which an illustrative MC-capable UE follows the blue trajectory in Fig. 19., starting from point [X_1=50, Y_1=300] and following a straight trajectory at 1 m/s up to the point [X_2=450, Y_2=300]. Fig. 20. presents the evolution of the average reward obtained with the application of these policies as a function of the number of training steps. The results show that at the beginning the average reward increases significantly, meaning that the training process is able to progressively learn better policies. Then, the reward tends to stabilize at around $40x10^4$ training steps, which gives an indication of the number of training steps needed to learn the policy. After this time, the reward fluctuations are only around 2%, which reflects the stability of the algorithm.

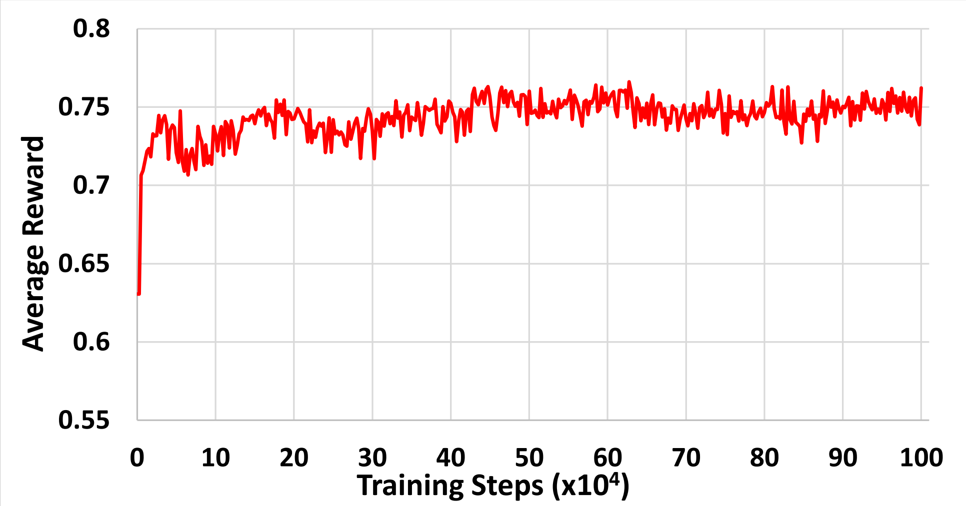

Fig. 20. Evolution of the average reward as a function of the training steps

3.5.3. Benchmarking Strategies

Aiming to evaluate the performance of the policy obtained through our proposed DQN-based method against different strategies, two reference approaches have been considered:

Optimum strategy: for a given UE, based on the set of candidate cells A_u=$\{C_{m,n}\}$ and N_{max} values, this strategy performs an exhaustive search process among all possible MC configurations at each time step and obtains the MC configuration that provides the highest reward. It is worth mentioning that it would not be a practical strategy for its implementation in real scenarios, mainly due to the large execution time for assessing all the possible MC configurations, so it is just considered as an upper bound of the DQN algorithm performance.

SINR-based strategy: in this strategy at every time step, all the traffic of the UE is served by the cell with the highest SINR value. This strategy reflects the classical approach in which a UE is served only by once cell, which is the one that provides the best quality (i.e., the highest SINR value). This criterion is considered for example in the specification [71] for cell selection procedure, as well as in the measurement report triggering criteria of [72] to determine when one cell becomes better than another one, so that this can be used when deciding a handover. In addition, different works have also considered an SINR-based criterion for determining the cell a UE is connected to in a cellular system, such as [73] and [74]. It is worth mentioning that the SINR-based strategy can be implemented making use of handover procedures for changing the cell that the UE is connected to.

3.5.4. Performance Evaluation of the DQN-based strategy

3.5.4.1. Performance for Different Trajectories

In order to assess the benefits brought by the proposed DQN-based MC strategy, this sub-section compares the performance obtained by the proposed approach against the two reference strategies mentioned before, the *optimum strategy*, and the *SINR-based strategy*. To conduct this evaluation, the DQN-based MC strategy makes use of the policy learnt after completing the training process explained in Section 3.5.2.

This policy is applied to the simulation of a UE of interest following 100 different straight trajectories of duration 400 s with different starting points and directions covering the entire evaluation scenario area. In each time window, the MC configuration according to the policy is applied and the resulting reward and throughput is measured. The same evaluation is executed when applying the two other reference strategies with the same trajectories.

Fig. 21. shows the average reward obtained for each one of the trajectories with all the considered strategies. It is observed that the DQN-based strategy achieves a performance very close to the optimum one in all the studied cases, with differences less than 1% in some of the trajectories, which confirms the good behaviour of the proposed approach. In turn, if we analyze the reward of the 100 trajectories and obtain the average for all of them, the DQN-based approach outperforms the SINR-based approach in around 13.1% but the improvement can be as big as 50% in certain trajectories (e.g., trajectory 95). This main advantage of the DQN-based strategy with respect to the classical SINR-based strategy is due to smartly splitting the traffic of the UE among the cells to avoid overload and enhance the obtained bit rate. From an implementation perspective, while the SINR-based approach can be implemented relying on handover procedures, the DQN approach requires the support of the MC feature and that the base stations handling the involved cells are interconnected through the Xn interface in case of 5G NR or the X2 interface in case of LTE. Both requirements are now commonplace in live 4G/5G networks, which rely on the MC feature for supporting the widely used 5G NSA operation.

Regarding comparison with the optimum strategy, the error obtained with the DQN-based approach was computed as the percentage of difference between the reward of both strategies. The average of this error for all the trajectories was 4.32%. Similarly, the 5th percentile of the error was 0.35% and the 95th percentile was 7.94%. This reflects that the DQN-based approach achieves a performance quite close to the optimum but without having to exhaustively search for all the possible MC configurations.

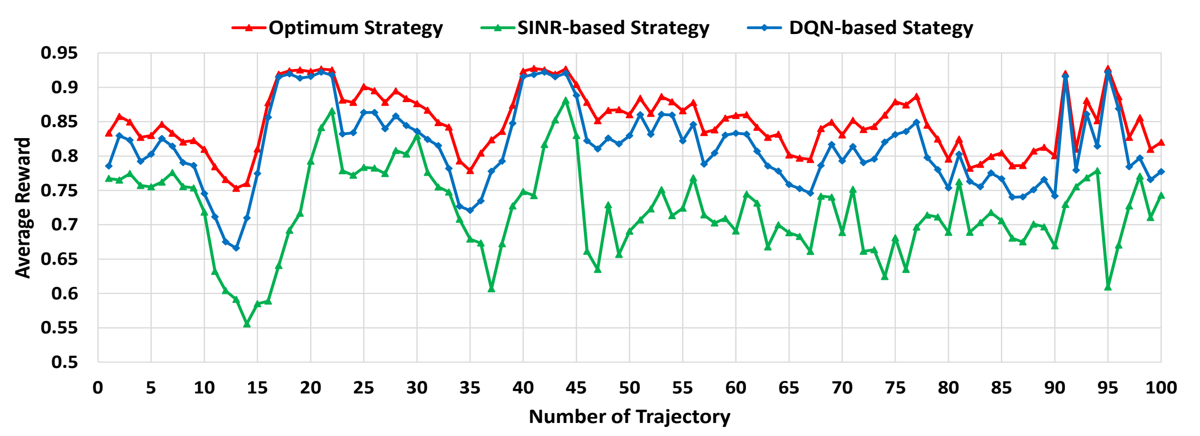

Fig. 21. Average reward for different trajectories.

In order to assess the throughput performance, Fig. 22. shows plots for each strategy of the cumulative distribution function (CDF) of the instantaneous throughput (Tu) values obtained by the UE of interest with all the considered trajectories. Again, it is observed that the DQN-based strategy achieves a close performance to the optimum strategy and clearly outperforms the SINR-based strategy. In particular, with the DQN-based approach the UE throughput achieved 90% of the time is higher or equal to 37 Mb/s, while with the SINR-based approach this value is only 18 Mb/s. Similarly, the DQN-based approach is able to provide the throughput of 50 Mb/s during around 50% of the time, while the SINR-based approach only provides it 35% of the time.

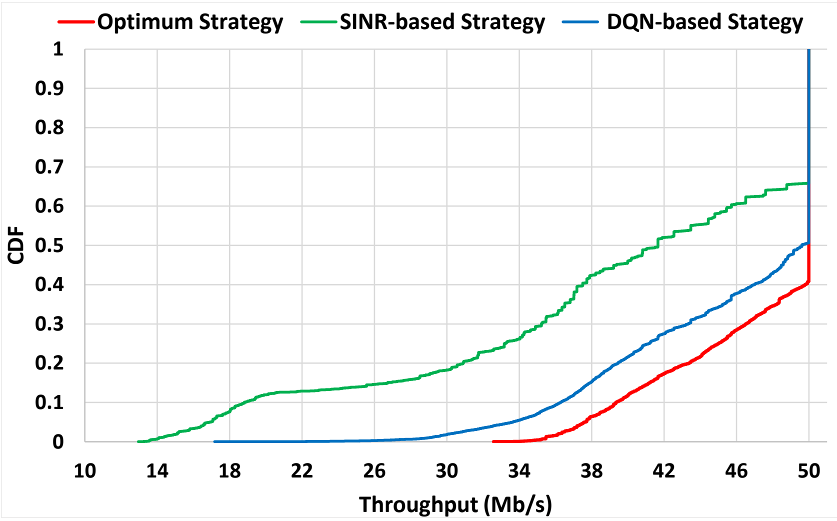

Fig. 22. CDF of the throughput achieved by the UE of interest.

Aiming to further assess the behaviour of the DQN-based strategy, we carried out a more detailed analysis focusing on a specific trajectory of a given UE starting in the point [X_1 = 485, Y_1 = 181] and following a straight trajectory at 1 m/s up to the point [X_2 = 85, Y_2 = 181]. For this specific trajectory, the SINR-based average reward was 0.634 while for the DQN approach this average reward rose to 0.835 which represents a gain of around 31%. Fig. 23. presents a more specific and detailed analysis of this trajectory during a period of 25 s between the points (X_1 = 285, Y_1 = 181) and (X_2 = 260, Y_2 = 181). In particular, Fig. 23a. represents the evolution of the SINR experienced by the UE in the LTE and 5G NR cells during this period. Since one of the NR cells has a higher SINR value, this will be the selected cell with the SINR-based strategy during the whole analyzed period. In contrast, the DQN-based strategy is able to split the traffic through the LTE and 5G NR cells in accordance with the experienced conditions in terms of signal and load. In this case, Fig. 23b shows the evolution of the values of $\beta_{m,n}$ selected by the algorithm. During the first 15 s the DQN-based strategy splits the traffic just using both LTE cells, i.e., $\beta_{1,2}$ and $\beta_{1,1}$ values are different from 0, even when those cells do not have the highest SINR values. For the remaining seconds, the traffic of the UE is delivered by using different MC configurations; for example, LTE-NR by adjusting the values of $\beta_{1,2}$ and $\beta_{2,4}$. As a result, Fig. 23c. shows the evolution of the reward with both strategies during the analyzed period, and it can be observed that the DQN-based strategy clearly outperforms the SINR-based one. In fact, if we consider average reward values for both strategies, the obtained with the SINR-based strategy is around 0.27, while the obtained with the DQN approach is 0.62, which is more than twice. It is worth mentioning that during the analyzed situation, the optimal strategy also provides a relatively low reward of 0.71. This is due to the traffic dynamics of the background UEs that lead to a high load in the cells during the analyzed time period. This high load does not allow ensuring the required bit rate with any of the combinations, which reduces the reward.

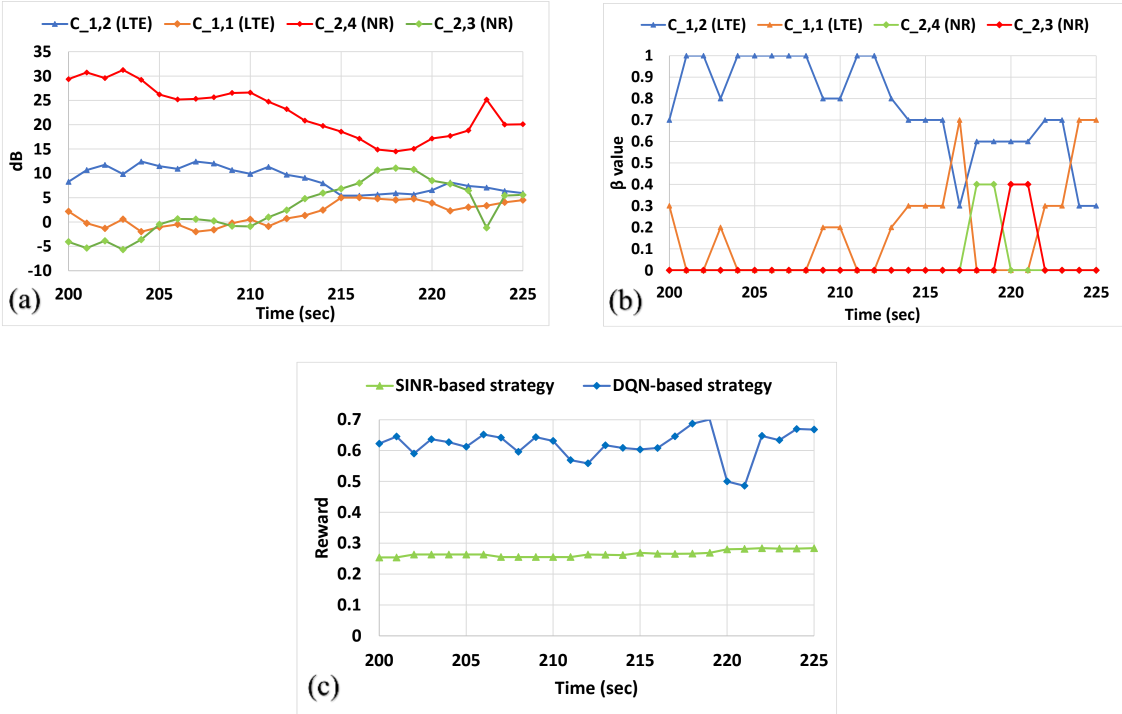

Fig. 23. (a) SINR evolution of the LTE and NR cells in the analyzed period; (b) evolution of $\beta_{m,n}$ in the detected cells by the UE during the analyzed period and (c) reward in the analyzed period.

3.5.4.2. *Statistics on the Multi-Connectivity Configuration*

As explained previously, the goal of the proposed algorithm was to find the MC configuration that maximizes the reward under any experienced conditions by the UE. If we look at the evolution of $\beta m,n$ in the detected cells by the UE during the analyzed period in Fig. 23b. we see that at some particular moments the DQN-based algorithm decided to send all the user traffic through a single cell. This means that during these moments, the algorithm decided that, based on the current state, not doing MC would result in a better performance, and this is confirmed through the reward observed in Fig. 23c. Based on this, in the following we assess the decisions made by the algorithm in relation to when to apply MC and which are the technologies involved.

If we consider that the evaluation scenario has two LTE cells ($C_{1,1}$ and $C_{1,2}$) and four 5G NR cells ($C_{2,1}$, $C_{2,2}$, $C_{2,3}$, $C_{2,4}$), with N_{max}=2, this means that the MC configuration can be done with five possible connection modes, namely no MC with LTE; no MC with NR; MC with LTE-LTE; MC with LTE-NR and MC with NR-NR. Fig. 24. presents the percentage of time that the algorithm selected each one of these modes for the same 100 trajectories studied in Section 3.5.4. It is observed that during around 56% of the time the algorithm decided not to do MC and to connect instead to only one cell of LTE or one cell of 5G NR. In contrast, during the remaining 44% of the time it decided to do MC, particularly being the combination LTE-NR, the most used approach selected in around 38% of the time. This is logic given that LTE cells have a bigger coverage area than the NR cells, but these have more bandwidth. Therefore, this connection mode could be helpful at the time of fulfilling; for example, the UE data rate requirements while the UE is moving. Given the statistics presented in Fig. 24. and after observing previously that the DQN-based solution performs much better that the SINR-based one, the importance of deciding the correct connection mode and the multi-connectivity configuration when required is clear.

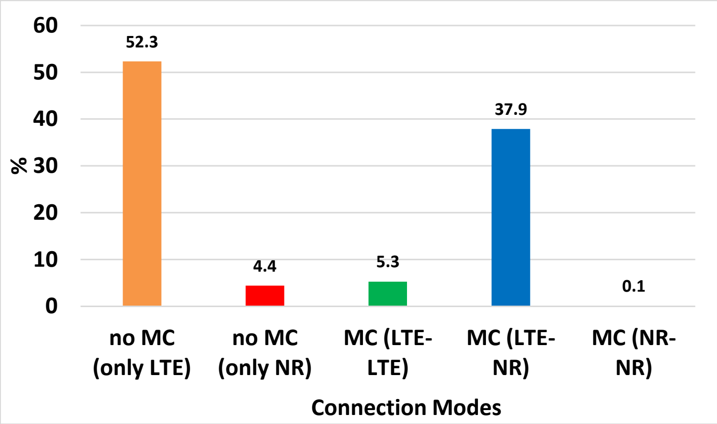

Fig. 24. Connection mode statistics for all evaluation times on the 100 trajectories studied

3.5.4.3. *Performance for Fixed Positions*

In the context of 5G, it is also pertinent to analyze situations in which the users/devices remain stationary at fixed positions, since some services have this characteristic (e.g., Fixed Wireless Access, smart cities with sensors or cameras at lamp posts). We carried out an evaluation performance of the DQN-based strategy considering the same scenario explained in Section 3.5.1 but this time the evaluation UE remained in a fixed position. In order to explore a variety of situations, we studied the performance of the UE for 400 s in 19 × 19 = 361 different positions all around the 500 × 500 m scenario, selected from a grid of locations in steps of 25m in both horizontal and vertical directions. We compared the results against the optimal and SINR-based strategies. Fig. 25. shows the CDF of the instantaneous values of throughput (T_u) and reward for all the considered positions. It is shown that, like in the evaluation with trajectories, the performance for fixed positions is close to the optimum one and clearly outperforms the SINR-based strategy. In fact, when analyzing some particular fixed positions, the DQN-based strategy is able to outperform the reward of the SINR-based one in up to 65%.

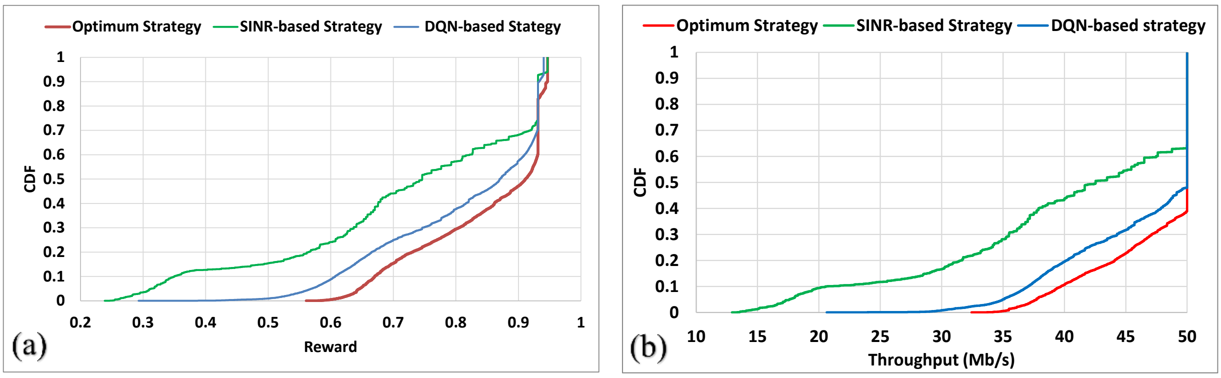

Fig. 25. (a) CDF of average reward for the UE of evaluation; (b) CDF of throughput of the evaluated UE.

Aiming to explain more in detail the situations where the DQN-based strategy can come up with more different decisions than the SINR-based approach, we focused on a specific period of time of one of the evaluated fixed positions where the UE was located in the coordinates (X = 300, Y = 350). For this position, the UE experienced SINR values in the 5G NR ($C_{2,2}$) and LTE ($C_{1,2}$) cells equal to 45.7 dB and 33.8 dB respectively. Due to these perceived values, the SINR-based strategy selected the 5G NR cell during the entire period. In contrast, as shown in Fig. 26a., the DQN-based strategy considered the current cell loads, and decided to split the traffic among the two cells. In the case of fixed positions, splitting the traffic becomes important because, if the higher-SINR-detected cell is serving other users, at some point it can get overloaded. However, by considering other aspects such as load as the DQN-based strategy does, it is possible to avoid this type of issue while improving the throughput obtained by the UE/device of interest. This effect can be seen clearly in Fig. 26b. For the 20-s analyzed period the obtained average throughput with the SINR-based strategy is 15.3 Mb/s while by using our proposed approach for the same period, the average throughput reaches 40.72 Mb/s, which represents a gain of around 166%. This result reinforces the idea about the importance of optimizing the MC configuration of the UEs.

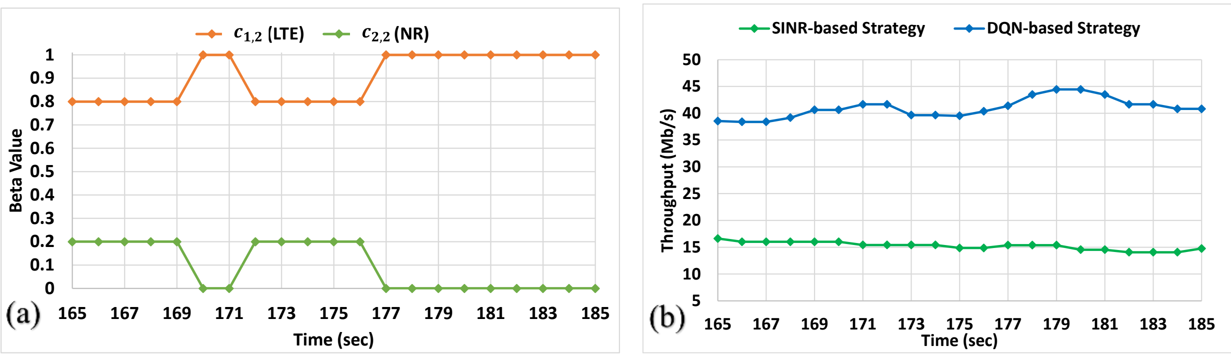

Fig. 26. (a) Evolution of $\beta_{m,n}$ in the analyzed period; (b) Obtained throughput in the analyzed period

Considering the results for all the different fixed positions, it was seen that the error of the DQN-based approach with respect to the optimum was, on average, 3.76%, the 5th percentile was 0.08% and the 95th percentile was 13.3%. Similarly, considering all the results obtained with the 100 trajectories of Section 3.5.4 and the fixed positions of Section 3.5.5, the average error was 3.9% and the distribution provides a 5th percentile of 0.09% and a 95th percentile of 12.5%. This reflects the close to optimum behaviour obtained with the proposed approach.

3.5.4.4. Performance Evaluation in scenarios with multiple MC-capable UEs

This section considers the performance of the DQN-based solution when there are multiple MC-capable UEs coexisting in the same evaluation environment and applying the learnt DQN-based policy. The study consisted of five different simulations, each one with a distinct number of MC-capable UEs, ranging from 5 to 25, all of them with a Ru = 25 Mb/s. During the simulation time each UE moved following a different trajectory and generated traffic during the whole evaluation time, equal to 400 s. The background traffic had a generation rate of 0.2 sessions/s and an exponentially distributed session duration with an average 120s, resulting in an average of 24 background UEs during a simulation.

The results presented in Fig. 22. were obtained by averaging the performance for the different UEs in each evaluation with the SINR-based strategy and with the proposed DQN-based strategy. Fig. 27a depicts the performance of our solution in terms of reward. It can be seen that the gain of DQN-based strategy with respect to SINR-based strategy increases from 11.2% for 5 UEs up to 31.7% for 25 UEs. Regarding the throughput performance, a similar assessment was done and the results are presented in Fig. 27b. They show that the proposed method outperforms the SINR-based approach for all the considered numbers of users., with a gain of around 33% for the case with 25 UEs. Notice how by applying our solution it tends to keep the obtained throughput very close to the requirement of Ru = 25 Mb/s.

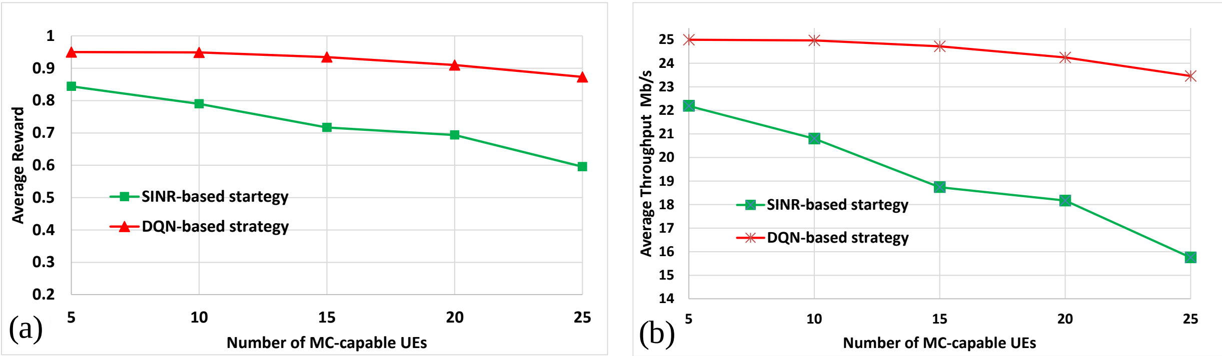

Fig. 27. (a) average reward values as a function of the number of MC-capable UEs; (b) average throughput values as a function of the number of MC-capable UEs.

3.5.4.5. *Analysis of the robustness of the learnt DQN-policy*

This section aims at evaluating the robustness of the DQN-policy when it is applied under conditions that differ from the ones that were considered during the training. For this purpose, considering that the training process has been done with a required bit rate R_u=50 Mb/s, the following results assess the generalization capability of the learnt policy when it is applied to different R_u values ranging between 15 and 65 Mb/s.

As a relevant metric for this assessment, a DQN-policy efficiency metric is considered, defined as the ratio between the reward of the DQN policy and that of the optimum strategy. Fig. 28. depicts the DQN-policy efficiency as a function of the required bit rate R_u value. The results of policy efficiency correspond to the average for the one hundred trajectories considered in the study in Section 3.5.4. For the R_u value of 50 Mb/s that was considered during the training it is observed in Fig. 28. that the efficiency of the original policy is around 95.75%. Then, it tends to decrease for higher and lower values of required bit rate. In fact, it is worth mentioning the effects of decreasing R_u, because even when the amount of required radio resources is less, the efficiency losses are higher. For example, the policy efficiency with R_u=15 Mb/s is 13.42% less than with 50 Mb/s. From the red line in Fig. 23, can be seen that efficiency variations lower than 1% are observed when the R_u value changes from 40 Mb/s up to 65 Mb/s, i.e. -10/+15 Mb/s with respect to the value used for training. This reflects that the learnt DQN policy is robust in front of variations of around 20-30% of the R_u value used in the training.

Based on these results, another training process has been conducted with the same parameters of Table II but now changing the R_u value during training. Specifically, R_u at the beginning of the training is 50 Mb/s and then it is changed between 5 Mb/s and 60 Mb/s with steps of +/- 10%. Moreover, the number training steps has been increased up to 3E6 in order to account for a higher number of possible situations to learn. The green line in Fig. 28., shows the obtained efficiency with the new learnt policy (denoted as retrained policy) in comparison to the original policy. It is clearly observed that the new policy achieves a good efficiency for all the considered RBR values.

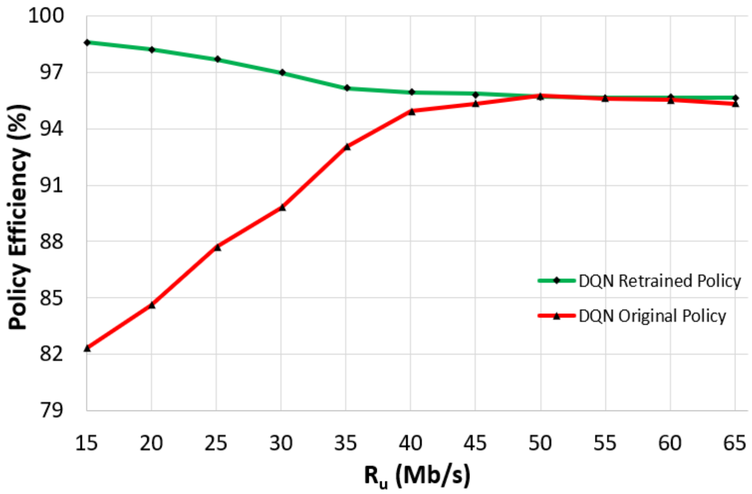

Fig. 28 DQN-Policy efficiency for different required bit rate (original vs retrained).

3.6. Concluding Remarks

In line with the overall objective of this thesis presented in chapter 1 of proposing, developing and evaluating different solutions for the 5G RAN in order to guarantee optimal connectivity conditions for network users aiming at satisfying their QoS requirements, in this chapter has been presented a novel approach based on Deep Q-Network algorithm for splitting the traffic of a UE among cells when using multi-connectivity depending on the current traffic and radio conditions experienced by the UE in the involved cells.

The proposed strategy has been evaluated and compared against the optimum case and against a classical SINR-based approach for different evaluation scenarios, involving UEs following trajectories and stationary UEs. Results have shown the capability of the DQN agent to learn a quasi-optimal policy that in certain conditions outperforms the SINR-based approach in up to 50% in terms of reward for certain trajectories. In turn, for the case of fixed positions, results have shown that the DQN-based strategy can achieve throughput gains of up to 166% with respect to the SINR-based strategy at certain times. The statistics regarding the different connection modes used by the DQN strategy have confirmed the capability of the algorithm to optimize the process of deciding when to use or not MC, to maximize the UE performance. Moreover, a performance analysis of the DQN-based strategy has been conducted when applied to different numbers of MC-capable users coexisting in the same evaluation environment. The results from this study have revealed a significant better performance for all the MC-capable users when applying the solution, which reinforces and confirms the relevance of optimizing the MC configuration. Additionally, it has been carried out a study analysing the robustness of the learnt policy when being applied with a required bit rate value different than the one that was considered during the training stage. It has been observed that the learnt policy is able to work properly with variations of the required bit rate of around 20%-30% of the value considered in the training. In turn, by conducting a training that considers a wider range of values of required bit rate, it is possible to increase the performance of the obtained policy.

Overall, the proposed solution has demonstrated adequate performance in all scenarios evaluated. Its use would imply a higher performance for network users, thus improving their connectivity conditions, which is in line with the general objective of this thesis.

Chapter 4. AI-Assisted Methodology for Detection and Resolution of Coverage-Limited Areas in 5G Networks.

4.1. Introduction

In line with the central objective of this work of providing optimal connectivity conditions for network users, this chapter addresses the problem of enhancing network coverage in 5G networks in order to achieve ubiquitous connectivity, thereby improving the user experience. The chapter proposes a methodology that first focuses on the detecting coverage constrained areas. To this end, a DBSCAN-based detector is introduced to analyse extensive user network performance data and identify, characterize, and validate the presence of coverage holes with high and sustained traffic using performance reports from network users. To mitigate the problems in the identified coverage holes, the proposed model considers the use of fixed relays to extend the coverage of RAN. In particular, an optimal relay positioning functionality is proposed. Overall, the proposed methodology in this chapter has two main goals: detecting coverage-constrained areas and effectively mitigating the coverage issues in these areas.

To evaluate the result of the proposed methodology in terms of network performance, a metric has been formulated to jointly measure the average spectral efficiency and the level of service availability in the network. The methodology has been evaluated through system-level simulations that replicate the real spatio-temporal traffic patterns collected from the Campus Nord of the Universitat Politècnica de Catalunya.

The chapter is structured as follows. Section 4.2 presents a literature review on the topic, along with a description of the contributions of the proposed methodology. Section 4.3 motivates the considered problem by presenting the scenario to be analysed and some results to highlight the coverage problems that may be found. The system model and the problem definition are described in Section 4.4, while Section 4.5 presents the proposed DBSCAN-based solution along with performance results in terms of coverage hole detection. Section 4.6 presents the relay-based solution for the detected coverage holes, along with the performance results obtained using this solution. Finally, Section 4.7 summarises the conclusions.

4.2. Literature Review and Contributions

One of the central objectives of 5G and beyond systems (e.g., 6G) is the provision of ubiquitous, seamless coverage. This is critical to meet the requirements of services associated with 5G [2], as well as those envisaged for 6G [3]. The provision of ubiquitous connectivity requires improved service availability, which, in turn, depends entirely on adequate coverage. Therefore, it is imperative to guarantee an adequate user experience. In that respect, MNOs must provide the appropriate conditions for consistent coverage across different locations, particularly indoor areas.

In their efforts to ensure a consistent coverage, MNOs face different issues such as coverage overshoots, pilot pollution and coverage holes (CHs) [75]. A CH refers to an area where the signal level of the serving cell drops below the threshold required to maintain essential services, such as the Signalling Radio Bearer (SRB) and Downlink Shared Channel (DL-SCH). Such CH can emerge due to several factors, including physical obstructions, suboptimal antenna configurations or even inadequate RF planning [76].

The presence of coverage holes can result in adverse effects for users, including call drops and radio link failures, ultimately leading to a sub-optimal user experience. If this situation occurs persistently for different users, it can lead to an increase in metrics such as call drop rate, which in turn affects the overall performance of the network. There are several methods in the literature for identifying coverage constrained areas in 5G networks, including techniques such as coverage analysis, performance measurements, drive testing and the use of crowdsourced data [75]. For example, by using the Minimization of Drive Tests (MDT) functionality [76], MNOs can collect radio network measurements to detect coverage issues, handover problems, interference problems, etc. This data can then be used to optimize the network. Indeed, the process for detecting coverage issues and implementing potential enhancement measures can be contextualized within the framework of Self-Organizing Networks (SON) as outlined by 3GPP, specifically focusing on the Coverage and Capacity Optimization (CCO) use case [77]. CCO leverages the collection of measurements, such as Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), and others, to systematically monitor the coverage within New Radio (NR) cells. Once areas of limited coverage have been identified, the MNOs can take measures to mitigate their impact. This may include deploying additional base stations, adjusting the transmitted power, or exploring techniques like relaying capabilities and others as highlighted in the survey [75].

One of the possible solutions to the coverage problems explored in the literature is the use of relays. Different authors have studied the use of relays for coverage extension. For example, in [78], a method for extending cell coverage radius was explored by obtaining the optimal relay placement position through an iterative algorithm. In [79], the authors investigated the use of relays for coverage extension and capacity enhancement. The authors presented different deployment scenarios for relay-assisted networks and analyzed their performance and benefits. In turn, 3GPP describes the characteristics of the relay-to-base station transmission in [80]. Similarly, some other authors have explored the use of UEs as relay. For example, in [81], the authors proposed an underlying D2D multi-hop relay-aided scheme to enhance the coverage and capacity of the network. In [18], the authors showed that incorporating UEs as relays can provide significant benefits to MNOs, since it reduces the need for a large number of base stations. Similarly, in [82] a functional framework for Relay UE (RUE) activation is proposed. It characterizes potential RUEs using a utility metric that assesses the improvement in network coverage when the RUE is activated. The 3GPP has also investigated the connectivity model of RUEs in [22].

In the context outlined above, the first objective of this chapter is to propose a solution that leverages the analysis of spatio-temporal traffic reports to identify coverage holes in base stations of a 5G network, while the second objective focuses on effectively mitigating these coverage holes through strategic relay placement. The proposed CH detector targets the identification and characterization of regions with coverage issues and sustained presence of traffic, including coordinates, area, density, etc. In this respect the solution starts from the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm [83] as a key tool for detection as used in a previous work that is available at [84] and introduces novel elements to achieve these goals. Specifically, the DBSCAN-based coverage hole detector is enhanced by incorporating a clustering quality measurement stage into the detection process, ensuring the application of the highest quality clustering for the coverage hole detection. In addition, the detection process is complemented by a reinforcement process that considers the repeatability of the detected clusters over time, providing additional confirmation of the CH detection. This reinforcement relies on a second execution of the DBSCAN algorithm that clusters the centroids of the coverage holes detected in different time periods. In this way, it is possible to confirm the persistent concentration of users in an area of poor coverage, thus improving the reliability of CHs detection. Moreover, this work introduces a new functionality that optimally determines the best relay placement location in accordance with the validated coverage holes considering pathloss conditions in the surrounding area to ensure a strong link between the BS and the relay. Last but not least, it is worth highlighting that the analysis is conducted considering real user distribution patterns collected in a university campus. In this way, the persistent concentration of users in an area of poor coverage is backed by realistic measurements reflecting typical daily motifs (e.g., students moving to a classroom with poor coverage for a lecture, then leaving to another classroom for another lecture, etc.).

4.3. Contextual Overview: Analyzed Scenario

In order to illustrate the problematic to be addressed, we have conducted an analysis of the propagation conditions and spatio-temporal traffic user distribution at the Campus Nord of the Universitat Politècnica de Catalunya in Barcelona, Spain, as shown in Fig. 29. The campus includes a total of 24 buildings of three floors each, denoted as A1, B1, C1, D1, A2, B2, etc. in Fig. 29. The 5G NR coverage in the scenario is provided by three base stations located in the surrounding area of the campus and identified by red icons in Fig. 29. The BSs operate at 3.7 GHz, each with a transmit power of 22 dBm, a bandwidth of 100 MHz and an antenna gain of 26 dB. The Urban Macro (UMa) model described in Section 7.4 of [70] has been used to generate a coverage map of the entire campus. In turn, the time and spatial user distribution has been extracted from available measurements of the number of users that are connected to each one of the WiFi access points distributed across the campus. This information is recorded in 15-minute intervals. By aggregating these data, we have produced a report for three different days which allowed to appreciate the real distributions of the users across the campus.

Fig. 30. illustrates the outcomes in terms of spectral efficiency and the average number of users/m^2 over the course of three days measured at 15-minute intervals for the first floor of the different buildings. The results indicate that the user distribution is non-uniform across the entire campus. Certain buildings exhibit higher user density than others, and even within the same building, varying user concentrations are observed on different days.

Google. Imagery ©2023 Airbus, Institut Cartogràfic de Catalunya, Maxar Technologies, Map data ©2023, Spain.

Fig. 29. Considered scenario of Campus Nord and its surrounding area.

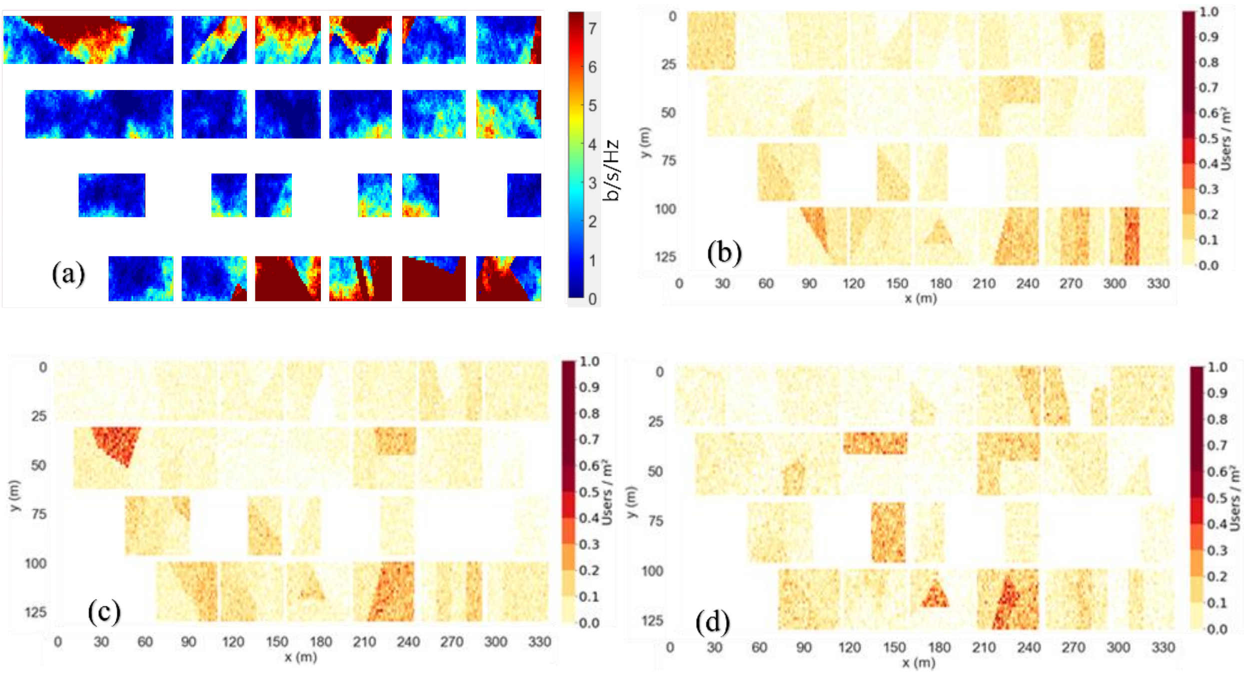

Fig. 30. Spectral efficiency map of the campus (a) and traffic spatial density in users/m2 for three different days 1 (b), 2 (c) and 3 (d).

For example, Building A6 displays fluctuating user concentrations in different days, while Buildings A1, A4, and C4 consistently exhibit elevated and sustained user density over the three days. When considering the user density in these buildings jointly with the spectral efficiency map of the campus (see Fig. 30. (a)), it is observed that the propagation conditions in Buildings A4 and A6 allow users to experience good levels of spectral efficiency. In contrast, users in buildings C4 and A1 are likely to experience significantly lower spectral efficiency values, which could result in the presence of coverage holes in areas with sustained high traffic levels. The identification of these regions of poor coverage and significant traffic constitutes indeed the problematic that is addressed by this proposed solution. In this way, this solution studies the way to detect these regions and to provide a feasible solution to mitigate the poor coverage conditions. The following section outlines the proposed solution that addresses these key aspects.

4.4. Considered System Model and Problem Definition

The considered system comprises a RAN consisting of multiple Base Stations (BS), the Core Network (CN), and the Service Management and Orchestration (SMO) functionalities, which includes aspects such as network configuration or performance monitoring. To characterize the coverage conditions experienced by the UEs, the spectral efficiency is considered. To that end, let us consider a UE within the coverage area of a given BS b. If the UE has a direct link to the BS, its spectral efficiency SD can be calculated using the Shannon formula.

$$S_D = min\left(S_{max}, log_2\left(1+SINR_{UE\text{-}BS}\right)\right) \tag{10}$$

where S_{max} value represents the highest spectral efficiency achievable by using the Maximum Modulation and Coding Scheme (MSC) defined in 5G NR [86]. Meanwhile, $SINR_{UE\text{-}BS}$ refers to the SINR in the connection between the UE and the BS.

A UE is considered to be in outage if its spectral efficiency is below a specified minimum threshold (S_{min}). In the event that S_D falls below S_{min}, it is assumed that the network service is not available. To mitigate this situation, this proposed solution considers the use of relays. Therefore, a UE has the option to establish a connection with a relay, which is assumed to have a spectral efficiency greater than S_{min}. When the UE connects with the BS through a relay, its spectral efficiency is constrained by the link with the worst conditions among the BS-Relay and Relay-UE connections. Therefore, the spectral efficiency when the UE connects through the relay, denoted as S_R, can be computed as follows:

$$S_R = min\left(S_{max}, log_2\left(1+min\left(SINR_{BS\text{-}Relay}, SINR_{Relay\text{-}UE}\right)\right)\right) \tag{11}$$

where $SINR_{BS\text{-}Relay}$ and $SINR_{Relay\text{-}UE}$ correspond to the SINR in the connection between the BS and the relay and between the relay and the UE, respectively.

Within the coverage area of the RAN, UEs may experience different levels of spectral efficiency depending on their location and propagation conditions. Then, a common problem is the existence of areas of very low spectral efficiency (so called coverage holes) that prevent that the UEs can get access to the services.

In order to assess how good, the coverage in a given scenario is, we consider two key metrics. The first one is the *service availability* of the network (η), which is defined as the probability that a UE experiences a spectral efficiency greater than the minimum level (S_{min}) that is needed to properly receive the service. Thus, at a certain period of time it can be measured as the proportion of total UEs whose spectral efficiency exceeds S_{min}. The second one is the *average spectral efficiency* (S), which can be obtained by averaging the spectral efficiencies of each UE given by (10) or (11), depending on whether the UE is connected directly to a BS or through a relay.

To optimize the coverage conditions in the scenario, the two metrics, S and η, are combined in the following global metric:

$$\delta = \alpha \cdot \eta + (1\text{-}\alpha) \cdot \frac{S}{S_{max}} \tag{12}$$

where α is used to weight the impact of the two metrics. The value of δ ranges from 0 to 1, where 1 corresponds to the best possible performance. In this context, the problem addressed is the maximisation of δ. This maximisation relies on the accurate identification of coverage holes as well as the strategic placement of relays to optimally mitigate their impact, thereby increasing the value of δ and consequently improving user and network performance.

4.4.1. Architectural components of the considered approach

The architecture of the proposed approach can be divided into three main blocks that are associated with the SMO, namely Network Monitoring, Coverage Hole Detector (CHD) and Relay Positioning (RP), as shown in Fig. 31. Once CHs have been identified and properly validated with the CHD, the proposed solution involves the deployment of relays (RP function) to enhance the coverage in these areas. Thanks to the deployed relays, the UEs will have two different connectivity options for establishing a link with the core network: either through a direct UE-BS link or by using a relay via a UE-relay link, as shown in the lower part of Fig. 31.

The CHD and RP functions are fed from network performance measurements collected by the Network Monitoring block. The CHD function is further divided into two distinct stages: data pre-processing and clustering. In the data pre-processing stage, network reports are collected and processed. Subsequently, in the clustering stage, the DBSCAN algorithm is applied to detect and validate the CHs. These detected CHs are characterized by essential information such as centroid coordinates and the distance between the centroid and the farthest point within the cluster. Following the detection of these CHs, the RP block analyzes the CHD results and evaluates the propagation conditions in the proximity of the CH regions in order to determine optimal coordinates for relays' positioning aimed at mitigating the poor coverage conditions at the CHs.

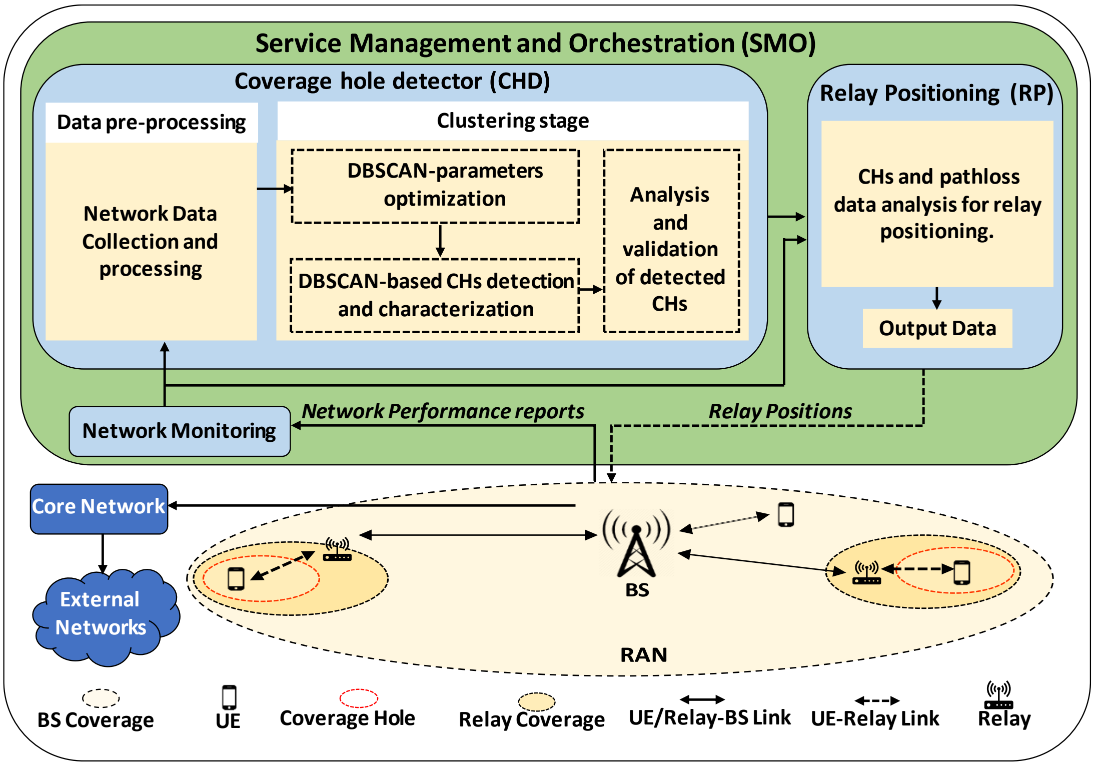

Fig. 31. Architecture of the consider approach.

4.5. Proposed Methodology for Coverage Hole Detection

The different blocks of the proposed solution are described in detail in the following sections.

4.5.1. DBSCAN-based Approach for Detection of Coverage Holes in 5G networks

4.5.1.1. Network Monitoring and Data collection

The network monitoring block is responsible for collecting data from network users such as location, session duration, start and end time of user sessions, spectral efficiency and signal to noise ratio (SNR). This can be done using, for example, RRC measurement reports [72] or MDT procedures [76]. It is assumed that the collected data is organized in daily reports denoted as K_d, where d={1,2,3,...,D} is the number of the day. These reports are delivered to the Data pre-processing block of the CHD.

4.5.1.2. Coverage Hole Detector

This block is in charge of the detection and validation of the CHs. The key technique used by the CHD is the DBSCAN clustering algorithm [83]. Its main objective is to identify clusters of points in a dataset based on their density distribution. The algorithm works based on two essential parameters: ε and σ. The ε parameter defines the radius around each point that forms a neighbourhood, while σ specifies the minimum number of points required to form a cluster. The algorithm starts by randomly selecting a point and expanding a cluster around it by adding neighbouring points within distance ε. If the cluster contains at least σ points, a new cluster is formed. Otherwise the point is marked as noise. This process is repeated until all points are either assigned to clusters or marked as noise. Further details on this process can be found in [83]. The pseudo-code of the CHD is presented in Algorithm II and it is divided in the following two stages:

1) Data pre-processing

In this sub-block the performance reports are processed and filtered to retain only data from users with signal quality levels below S_{min} (i.e. UEs in outage). This data includes their spectral efficiency S_D, location and time of session start. Each daily report is then divided into different time frames T_f (e.g. hours), numbered as f={1,2,3,...,F}. The output of this sub-block is an $F×D$ matrix **M**={$L_{f,d}$}, where $L_{f,d}$ ={l_1,l_2,l_2, ...,l_U} is a vector containing the location of all users with $S_D< S_{min}$. In this regard, U denotes total number of users in outage in the time frame f of a day d. This sub-block is also responsible for adapting the DBSCAN input data to a suitable format. In this case, this is performed by normalizing the coordinates of the users in outage using z-score normalization [87]. The process described above is carried out between lines 1-14 of Algorithm II.

2) Clustering Stage

To identify and validate the regions with a high concentration of users in outage, the DBSCAN algorithm is applied to each element $L_{f,d}$ of matrix **M**. This is done through the following three processes.

DBSCAN-parameters optimization: The choice of values for ε and σ depends on the nature of the data being analyzed and is very important as it affects the quality of the clustering and, consequently, the decisions made based on the clustering results. A small ε value could result in multiple clusters containing noise points, while a big value in the formation of one big cluster containing all points, so both cases may lead to inaccurate insights. In the case of σ, it has a direct impact in the density of points in the clusters as well as in the number of detected clusters, so its value should be also selected carefully. This process takes place between lines 15-21 of Algorithm II. The optimal ε, σ values are obtained through a fine-tuning process based on a clustering evaluation metric known as the Davies-Bouldin Index (DBI) [88], which assesses both the compactness and separation of clusters, with lower values indicating better cluster quality. The DBSCAN algorithm is applied several times to each element of M, each time with a different value of ε and σ, in the range [ε_{min}, ε_{max}] with granularity G_ε and [σ_{min}, σ_{max}] with granularity G_σ, respectively. The DBI values corresponding to each configuration (ε, σ) and element of M are retained. Then, the average of the DBIs for all elements of M with a given configuration (ε, σ) is obtained. The optimum configuration (ε_{opt}, σ_{opt}) is the one with the lowest average DBI.

DBSCAN-based CHs detection and characterization: In lines 22-24 of Algorithm II the DBSCAN is applied to each element of M using the optimal configuration (ε_{opt}, σ_{opt}) obtained in the previous process. The detected clusters are characterised by computing their centroid coordinates (C_h) and by the distance between the centroid and the farthest point within the cluster (R_h). This data forms a new F×D matrix of clustering results denoted as R, where each element is a pair (C_h, R_h).

Analysis and validation of detected CHs: In this process, the detected CHs are analysed to identify only those that present sustained presence of a large number of users in the area of the CH over different time frames or days, as these would be the ones in which a relay can be necessary. This is performed by executing the DBSCAN algorithm again but this time, instead of analysing user positions, the algorithm examines the centroid coordinates of the previously detected clusters in matrix R. The parameters of this DBSCAN execution are ε_{val}, which is the minimum distance between cluster centroids and σ_{val}, which is the minimum number of samples required to form a cluster. In practical terms, σ_{val} can be interpreted as the minimum number of days required for a cluster to be detected and considered valid. For example, if the centroid coordinates of different detected clusters are within a distance smaller or equal than ε_{val} and this occurs in at least σ_{val} different days, they are assumed to represent the same concentration of users in outage. This confirms the sustained presence of traffic and validates the coverage hole. Then, the output of this process is a dataset, denoted as H, containing only the cluster centroids C'_h of the validated CHs, as well as an average value of R'_h. This process corresponds to lines 25-28 in Algorithm II.

Algorithm II. Coverage Hole Detector

 # Data pre-processing
1 Obtain the K_d reports from Network Monitoring block.
2 Set values for ε_{min}, ε_{max}, σ_{min}, σ_{max}, ε_{val}, σ_{val} G_σ, G_ε and F
3 Classify data of K_d reports into F time frames
4 **For** d=1…D
5   **For** f=1…F
6   Obtain the spectral efficiency value S_D for all users u (1)
7   **For** u=1…U
8   **If** $S_D < S_{min}$
9   Get a vector $L_{f,d}$ containing coordinates of users u l_u
10   **End if**
11   **End for**
12   **End for**
13   Generate a matrix M={$L_{f,d}$}
14 **End for**
 # DBSCAN-parameters optimization
15 **For** each ε from ε_{min} to ε_{max} with granularity G_σ
16   **For** each σ from σ_{min} to σ_{max} with granularity G_ε
17   Apply DBSCAN to each element of M with (ε, σ)
18   Compute DBI for detected clusters with each configuration
19   **End for**
20 **End for**
21 Average DBI values to get (ε_{opt}, σ_{opt}) that ensure *min*(DBI)
 # DBSCAN-based CHs detection and characterization
22 Apply DBSCAN to each element of M with (ε_{opt}, σ_{opt})
23 Characterize detected CHs by computing their C_h, and R_h
24 Generate a new matrix R of detected and characterized CHs
 # Analysis and validation of detected CHs
25 Extract CHs centroids coordinates from R into new vector D
26 Apply DBSCAN with (ε_{val}, σ_{val}) to the data in D
27 Characterize detected/validated CHs by C'_h, R'_h
28 Store data of detected/validate CHs in a dataset H

4.6. Performance Evaluation

This section evaluates the performance of the proposed solution through system-level simulations in the scenario described in Section 4.3. For evaluation purposes, the network performance reports have been generated following a traffic generation that assumes a Poisson session process where the generation rate is adjusted based on the actual number of users in each 15-minute interval and the actual spatial user distribution measured in the Campus Nord. The considered configuration parameters of the BSs and relays are shown in Table III. The proposed solution for CHD and RP has been developed in Python by using the machine-learning library *scikit-learn* [89] for the DBSCAN implementation.

TABLE III. BS AND RELAYS CONFIGURATION PARAMETERS

Parameter	**Value**	
Type of node	BS	Relay
Frequency	3.7 GHz	3.5 GHz
Channel bandwidth	100 MHz	100 MHz
Tx power	22 dBm	-10 dBm
Tx Antenna gain	26 dB	0 dB
Height	BS1=27 m; BS2=23 m; BS3=8 m	1.5 m
UE antenna gain	10 dB	0 dB
UE noise figure	9 dB	
UE height	1.5m	
Path loss model	Model UMa of Sec 7.4 of [70]	Model InH - Office of Sec 7.4 of [70]

4.6.1. Performance Evaluation of Coverage Hole Detection Process

The detection of CHs involves the collection of D=28 days of performance reports, focusing specifically on the first floor of the 24 buildings on the Campus Nord, as shown by the yellow dotted line in Fig. 29. The considered time span for each day is between 8:00 a.m. and 8:00 p.m., divided into F=12 time frames of one hour. These data serve as input for the CHD described in Section 4.5. The parameters employed by CHD are S_{min}=0.5 b/s/Hz, ε_{min}=0.1, ε_{max}=0.4 (note that the coordinates are normalised in Section 4, so the distance ε is also normalised between 0 and 1), σ_{min}=10, σ_{max}=80, G_{σ} =0.1, G_{ε}=10, ε_{val}=0.1, σ_{val}=10 and x=3 m. Fig. 32. shows the average values of the DBI for all combinations. The graph shows that the combination that gives the minimum DBI is ε=0.3 with σ=70. Consequently, these values are identified as the optimal parameters (ε_{opt}, σ_{opt}) to be used at the DBSCAN-based CHD and validation stage. As a result, the CHD successfully detected and validated five coverage holes (H=5) with the following data: C'_1=[87.9,15.75] m with R'_1=17,1 m; C'_2=[80.2,43.3] m with R'_2=20.35 m; C'_3=[219.5,46.3] m with R'_3=30.1 m; C'_4=[71.5,75.8] m with R'_4=25 m and C'_5=[89.5, 108.7] m with R'_5=27 m. Using this data as a source, the RP block identified the optimal positions for placing the relays that are illustrated in Fig. 33.

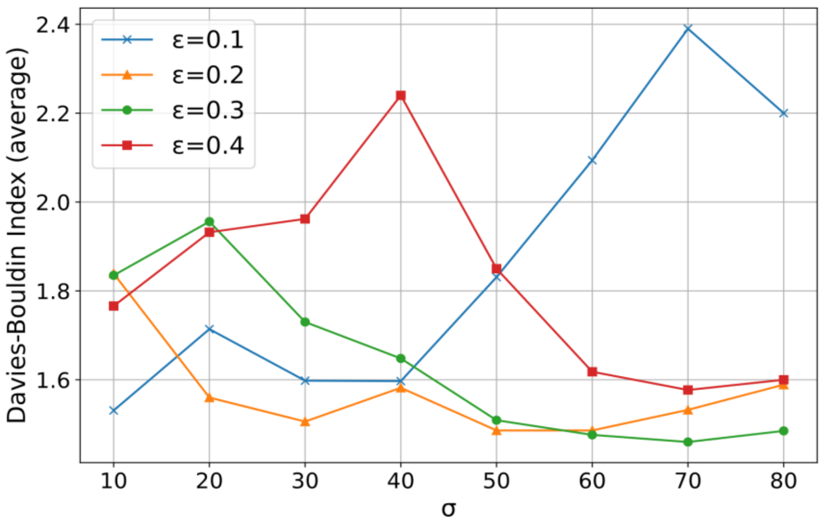

Fig. 32. Average Davies-Bouldin Index for different (ε, σ) combinations.

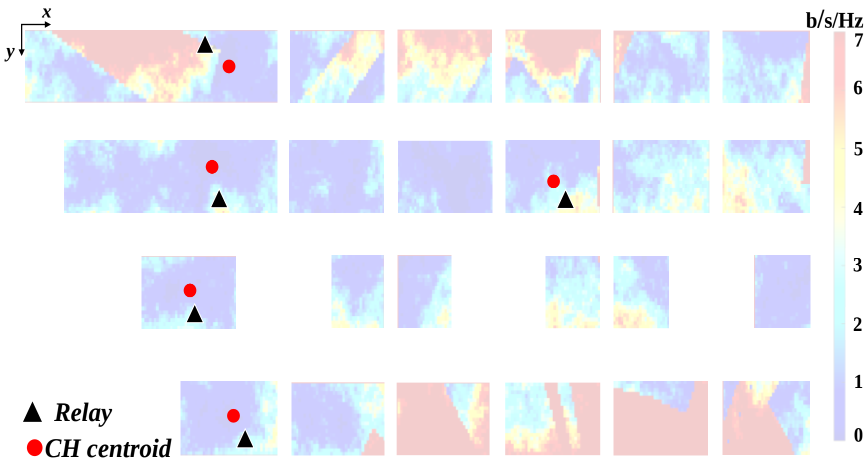

Fig. 33. Updated Network as result of CH detection and relay positioning.

4.7. Proposed Solutions for Coverage Hole Mitigation

In order to solve the coverage issues, according to the situation case, the MNOs could apply different solutions, for instance, the deployment of additional base stations, small cells, relays, etc. Given the studied case in this thesis, it is considered the use or relays for providing coverage to the constrained areas. As illustrated in Fig. 28.

4.7.1. Relay-based Solution

4.7.1.1. Relay positioning functionality

This block, whose pseudo-code is presented in Algorithm III, is responsible for deciding the optimum coordinates for a relay placement. This is done on the basis of the centroid of the validated CHs C'_h and their R'_h in dataset *H*. The RP block analyses the path loss conditions in the region defined by a circle centred at C'_h with radius R'_h+x, where *x* is a distance in meters (line 4 of Algorithm III). Then, the spectral efficiency is calculated in each 1×1m pixel of this region and, the pixel with the highest spectral efficiency is considered as the optimum position for placing a relay. This process is repeated for all detected CHs. The optimal coordinates for relay placement are the output of this block and of the overall proposed solution.

Algorithm III. Relay Positioning

Get dataset H containing data of detected/validated CHs
Set the value for x
For each validated CH in H
    compute pathloss in each 1×1m pixel inside in the circle centered at C'_h with radius R'_h+x
    compute the spectral efficiency in each 1×1m pixel inside circle centered at C'_h with radius
    obtain the coordinates of the pixel with highest spectral efficiency
End for
Gather optimal coordinates in an output file

4.7.1.2. Performance evaluation of the updated network

In this section, we assess the performance of the updated network deployment considering the relays resulting from the CHD and RP algorithms to provide coverage to the detected and validated CHs. Our simulation replicates the exact conditions of the 28 days during which data was collected for CH detection. This ensures that the results with and without relays are directly comparable. Fig. 34. compares the obtained outage probability, average spectral efficiency and δ metric in (12) for the original and updated network deployments. A notable result is the significant reduction in the probability of outage, as shown in Fig. 34a. A day-to-day comparison shows a consistent reduction across all days, ranging from a minimum of 31.6% to a maximum of 61.5% and with an average reduction of 45.6%. In terms of the spectral efficiency, a similar effect is observed in Fig. 34b., which shows that the use of relays in the identified positions consistently improves spectral efficiency over the 28-day evaluation period with an average improvement of 11.4%. The spectral efficiency and outage probability are captured together by δ in Fig. 34c., where it is observed that, thanks to the 5 new relays, the value of δ increases in all the evaluated days with an average improvement of 8.8%. Note that these results are global, considering data from the entire campus. In other words, significant improvements are achieved by strategically placing relays in just 5 out of the 24 buildings. This highlights the critical role of accurate coverage hole detection and optimal relay placement in improving network performance.

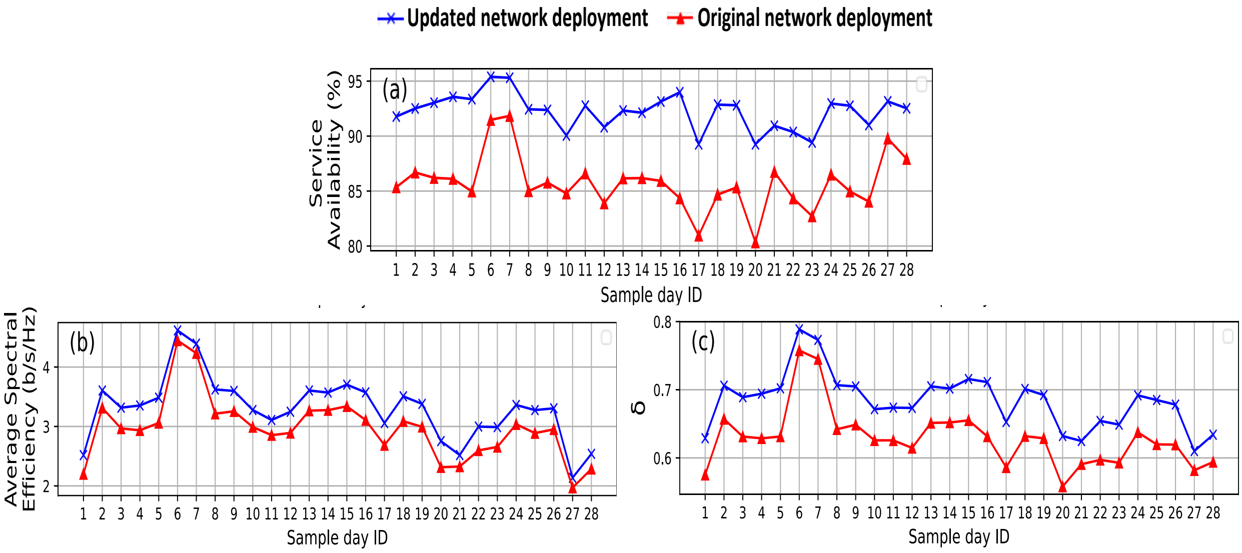

Fig. 34. Obtained performance in the studied scenario. (a) Outage probability; (b) Average spectral efficiency; (c) δ value

To deeply analyze the performance obtained in the regions where the CH were detected and validated, Fig. 30. includes the comparison of the spectral efficiency (Fig. 35a) and the outage probability (Fig. 35b) in the 5 detected and validated CH for the original and updated network. The results are obtained by averaging the sessions that occurred in each region over the 28 days studied. Regarding the spectral efficiency results in (Fig. 35a), significant improvements are obtained for the updated network deployments. For example, for CH1, the average spectral efficiency went from 0.14 b/s/Hz to 3.43 b/s/Hz, which is a substantial increase in a factor of 24.5. Similarly, this improvement factor is 9 for CH2, 11 for CH3, 6.9 for CH4 and 34.5 for CH5. Accounting for all CHs, on average the spectral efficiency was improved by a factor of 17.2.

Moreover, a notable improvement is also observed in terms of outage probability. By definition, the outage probability reached 100% in all regions of detected CHs, as in this region all the UEs experience S_D<S_{min}. However, when relays were deployed in the network, there was a significant decrease in the outage probability, as illustrated in Fig. 35b. The values decreased from 100% down to 4.5% in CH1, 1.1% in CH2, 0% in CH3, 4.9% in CH4 and 10.1% in CH5. On average, the outage probability for the detected regions was reduced in 95.8%. These significant improvements highlight the capability of the proposed solution to improve the coverage conditions in the detected CHs.

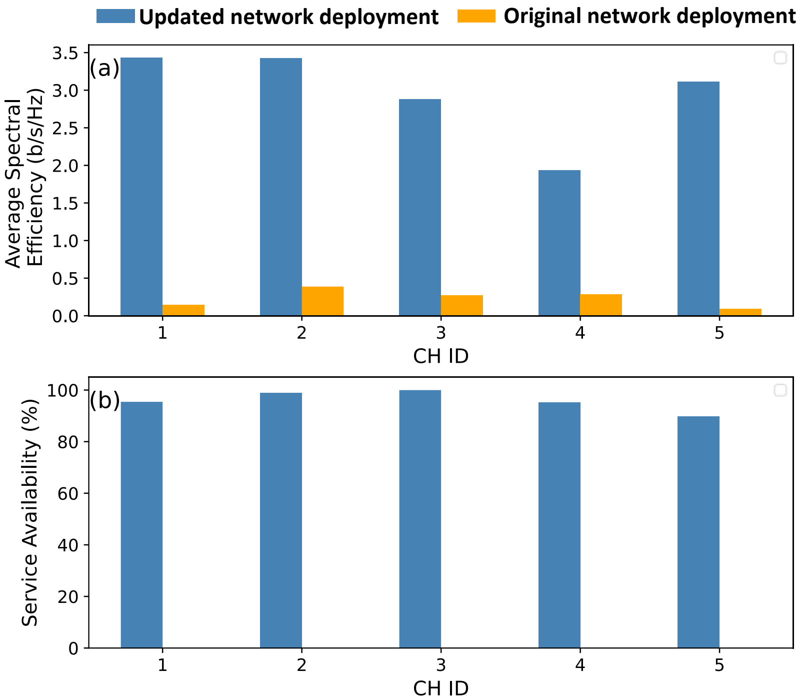

Fig. 35. Obtained performance in the detected and validated coverage holes. (a) Average spectral efficiency; (b) Outage probability

4.8. Concluding Remarks

In line with the objectives of this thesis stated in Section 1.3, and given the critical role of network coverage in ensuring QoS satisfaction of network users. In this chapter has been addressed the problem of identifying coverage-constrained areas within the area covered by a mobile network operator by means of extensive analysis of real user-network performance data. The chapter introduced a novel methodology for the detection, validation, and resolution of coverage holes in 5G and beyond networks.

It was proposed a coverage hole detector (CHD) that makes use of the DBSCAN clustering algorithm to detect, validate and characterize coverage holes with high and sustained presence of traffic based on network performance reports of the network users. For mitigating the coverage holes, the considered methodology also proposed the use of relays as a way of enhancing the coverage. To do so, the methodology incorporates a functionality that obtains the coordinates for the optimal placement of relays based on the detected coverage holes and the pathloss conditions.

To evaluate the proposed approach, different system-level simulations that replicate the real spatio-temporal traffic patterns collected from the Campus Nord of the Universitat Politècnica de Catalunya have been performed. The proposed solution successfully identified and validated five coverage holes, an updated network deployment with the strategically placed relays has been analysed. The updated network demonstrated remarkable improvements with a substantial increase in service availability as well as in the spectral efficiency. Considering the whole campus, the spectral efficiency was increased on average by 11.4% while the service availability was improved by 7.6%. Furthermore, when analysing performance at a CH level, in terms of spectral efficiency, the improvement ranged from a factor 6.9 to a factor 34.5 depending on the CH and the service availability was increased up to 95.8% on average. The obtained results highlight the important role of an accurate CH detection and optimal relay placement in improving network performance. The proposed methodology has shown significant improvements in all considered metrics, leading to improved network performance and, consequently, improved user experience through adequate connectivity conditions, the latter being the central objective of this thesis.

Chapter 5. AI-Based Relay UEs Activation Strategy for Coverage Augmentation in 5G Networks.

5.1. Introduction

In this chapter is addressed the problem of optimally using the relay capabilities of the UEs to augment the radio access network (RAN) in 5G deployments and beyond. This can be particularly useful in coverage constrained scenarios, such as those using millimetre waves, due to the difficulty radio signals penetrate some structures. This can lead to signal blockages and high penetration losses when providing outdoor-to-indoor coverage. To overcome these limitations, the use of relay UEs (RUEs) is seen as a possible solution to effectively extend the coverage of a cellular network. In that respect and, in connection with the overall objective of the thesis of providing enhanced connectivity capabilities for network users, as well as an alternative of solution for the fixed-relay based solution presented in the previous chapter. In this chapter is proposed a deep learning-based algorithm to optimize the decision regarding when RUEs should be activated and deactivated in accordance with the benefits they can provide for increasing the spectral efficiency and decreasing outage probability of network users. In order to evaluate the efficiency of the RUE activation process, a metric has been formulated to measure the effectiveness of the current mode of the RUEs (activated or deactivated) as a result of decisions made by the DQN agent. In order to evaluate the solution, key performance metrics of interest have been defined. Based on these metrics, the proposed solution has been evaluated and compared with six different activation strategies, providing context on the real impact and benefits obtained by the proposed solution.

The rest of this chapter is organized as follows: Section 5.2 provide a literature review on the studied topic as well as the contributions of the proposed model. In Section 5.3 is described the problem definition as well as the considered relay-activation problem. Section 5.4 presents the proposed DQN-based solution. Then, Section 5.5 provides the performance assessment of the proposed solution in terms of different indicators and the comparison against different benchmark solutions, including the exhaustive search strategy, genetic algorithm and a reference from the state-of-the-art. Finally, Section 5.6 summarizes the conclusions and discusses future work.

5.2. Literature Review and Contributions

In line with the increasing demands of traffic and new services in mobile networks over the years, there has been a tremendous technological evolution not only at the network infrastructure side but also at the UE side, leading to the availability of UEs with powerful communication and computational capabilities. Following these trends, in [18], is described a beyond 5G (B5G) scenario where the UE actively cooperates in the provision of network services, e.g., by relaying traffic from other UEs toward the network. The obtained results revealed that the use of relay UEs (RUEs) can be beneficial for MNOs thanks to a substantial reduction in the number of base stations to be deployed and the consequent reduction in capital expenditures.

Although the idea of using relay stations for coverage and capacity extension has already been studied in the literature (see, e.g., [90]), its practical implementation in previous technologies such as 4G has been quite limited and focused on very specific situations, such as using fixed relays for extending coverage in a tunnel. However, the idea of using relays has recently gained momentum due to the challenges in providing coverage and capacity in certain scenarios, such as in-home residential environments with mmWave frequencies, smart factories or even public safety applications. Accordingly, the Third Generation Partnership Project (3GPP) has recently introduced a new relaying technology, referred to as Integrated Access and Backhaul (IAB), which makes use of 5G New Radio (NR) technology for supporting the backhaul of a base station, thus offering an alternative to fiber backhaul [91][92]. Similarly, the use of vehicle-mounted relays that consider moving vehicles with onboard relay base stations is also a subject of a recent work item in 3GPP Release 18, whose outcomes are reported in [93]. The UE-to-network relaying feature, in which a UE relays the traffic of another UE to/from the network in a two-hop link, has also been the subject of a 3GPP study item [21] and was recently incorporated as one of the connectivity models of [22], which discusses different applicability scenarios of relay UE and defines requirements and key performance indicators. The possible applications of relay UE in [22] cover a wide variety of scenarios, such as in homes, smart farming, smart factories and even public safety applications. In addition to the work in standardization, the interest in UE-to-network relaying is also reflected by different works, such as the recent survey [94] and references therein.

To successfully realize the UE-to-network relaying concept, it is necessary to design and develop new functions in B5G systems. On the one hand, these cover top-level service layer capabilities so that mobile network operators and UE holders can interact to establish the conditions when the UE can be integrated as part of the RAN, including service level agreements (SLA) at both the business and technical levels. On the other hand, adequate management and control layer functionalities need to be developed to leverage the connectivity supplied by the RUEs. In this context, a critical functionality is "RUE activation", which is the focus of the proposed model presented in this chapter. This functionality is in charge of deciding under which conditions a UE is eligible to be activated to act as a relay and be incorporated as an additional element of the RAN.

The RUE activation functionality was studied in [95] in relation to the type of context information to be considered by this problem. The authors in [95] analyzed seven RUE activation strategies that differ in the criteria and context information used and found that the most effective strategies for reducing outage probability were those that considered the number of UEs that an RUE could serve based on the knowledge of the spectral efficiency of these UEs. Leveraging the outcomes of a previous work, a functional framework for supporting RUE activation was presented in [82] based on the characterization of each potential RUE through a utility metric that measures the coverage enhancements brought to the network when the RUE is activated.

The main contribution of the solution presented in this chapter is a new RUE activation strategy that makes use of deep reinforcement learning (DRL), and more specifically of the deep Q-network (DQN) technique, to optimally activate UEs as relays to enhance the coverage conditions of the deployed base station. The use of DQN is particularly convenient because it is able to learn decision-making policies for problems with high-dimensional state and action spaces and it is able to progressively update the learned policy based on the accumulated experience during a training process. Then, with the proposed approach, each base station is associated with a DQN agent that learns the RUE activation policy depending on the potential benefit brought by the RUEs to be activated in accordance with the network dynamics. Overall, the proposed solution aims at improving the spectral efficiency and reducing the outage in the scenario, but in a way that RUEs are only activated when they are beneficial for the network, thereby reducing the time that RUEs remain active.

To the best of our knowledge, at the time of doing developing this model, no previous works in the literature had addressed the RUE activation problem by means of DRL. Only a few works have considered machine learning (ML) methods in the context of relay-assisted networks, but they address different problems. For example, in [96], a deep-learning model is used for the user-to-relay association problem, in charge of predicting the best serving relay for a UE. The algorithm made decisions in accordance with the distances between the UE, the relay and the base station. Similarly, the authors of [97] considered the problem of selecting the set of relays that enable a data packet to travel from a start node to an end node. The problem was modelled as a Markov decision process with actions associated with specific nodes, and a Q-Learning algorithm was trained to determine the best communication path. In that context, the presented model in the chapter contributes with a novel alternative for optimally enhance the network-users connectivity conditions.

5.3. Problem Definition and Proposed Architecture

Let us consider a scenario with a 5G network infrastructure deployed by a mobile network operator, as depicted in the lower part of Fig. 31. It is composed of the RAN that includes the base stations (BS), the core network (CN) and the service management and orchestration (SMO) system that enables the configuration and performance management (PM) of the infrastructure. To address coverage-limited situations due to, e.g., high penetration losses or obstructions when using mmWave bands, the system has the possibility of activating some UEs to act as relays for other UEs. In this way, a UE can reach the core network through a direct link with the BS or by means of an RUE, as shown in Fig. 31.

The SMO includes an RUE activation management (RAM) function that determines when, where and under what conditions a UE served by a BS is suitable to be activated as an RUE to solve coverage problems. To this end, acquiring knowledge about the UEs in the area of a BS and their behavioural patterns will be relevant. For example, let us suppose a situation in which a UE is located inside an office building during working hours. Then it is highly likely that it remains stationary and is connected to its serving BS for long periods of time. If the signal quality of this UE is sufficiently good and its battery level is above a certain threshold, this UE can be considered to be a potential candidate relay to be activated. The decision on whether to activate a candidate RUE needs to trade-off different aspects, such as the global benefit of activating the RUE in accordance with the performance experienced by the UEs connected to it and the cost of activating the RUE, e.g., in terms of energy consumption. Based on this, the problem considered in this chapter is the optimization of the RUE activation decisions to ensure that RUEs are only activated when the network dynamics require them.

The RUE activation decision-making for a given BS is executed at the RUE activation controller (RAC) of the RAM function at the SMO, as depicted in the functional architectural framework of Fig. 31. It is supported by the candidate RUE database that contains the list of UEs capable of acting as RUEs in the BS. These are the UEs that stay within the BS for a sufficiently long period of time and whose owners have reached the appropriate agreements with the MNO so that they are authorized to act as RUEs. How the list is built is out of the scope of this proposed solution, based on the assumption that the identification of "home UE" associated with BSs would not be difficult to achieve (e.g., by observing that a UE is always served by the same BS at certain times, hours, etc.). Similarly, the business model-related mechanisms for engaging UE owners to let their terminals act as relays are also out of the scope of this presented solution but could be based, e.g., on giving incentives to UE owners [94].

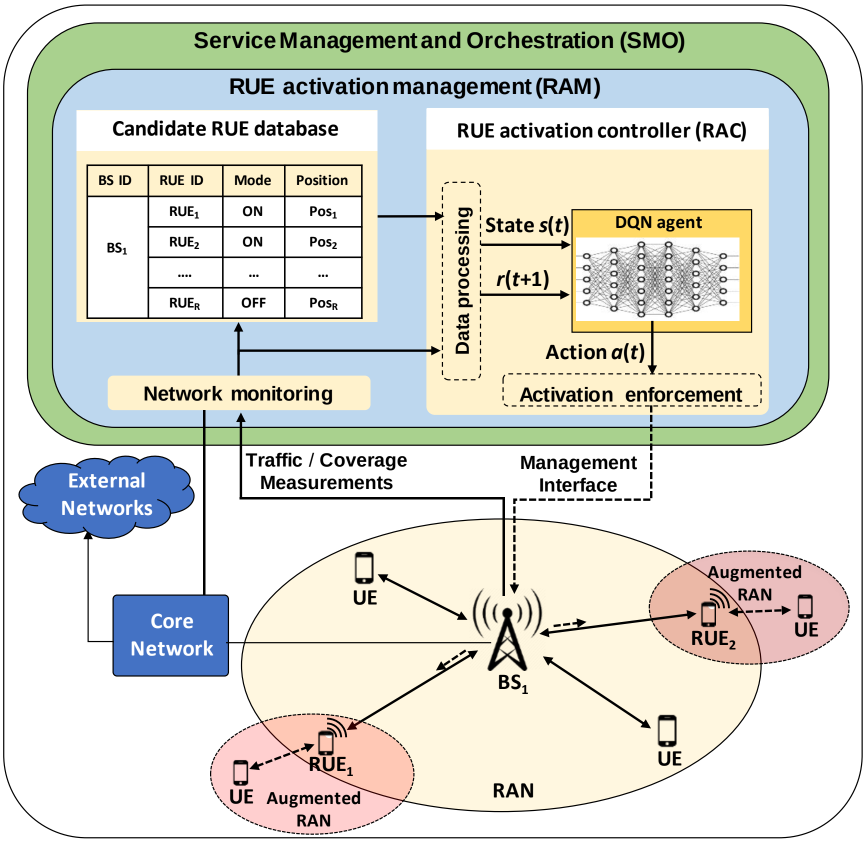

Fig. 36. Architectural components of the considered approach.

The candidate RUE database includes different fields such as the identifier of the relay, the position and the current status mode that specifies if the RUE is active or not at a certain time. These data, together with specific performance measurements collected from the network, constitute the inputs to the RAC so that it can decide to activate the RUEs at a certain time. To formalize the problem, let us first model the performance experienced by UEs. For this purpose, let us suppose a UE is located in the coverage area of a given BS. If the UE is directly connected with the BS, the spectral efficiency S_D can be obtained by using the Shannon formula as given by expression (10) in section 4.4. On the other hand, if the UE is connected to the BS via an activated RUE, the spectral efficiency is limited by the link with the worst conditions between the BS-RUE and RUE-UE links as it is expressed in equation (11) of section 4.4.
It is assumed that a UE is in outage if it is experiencing a spectral efficiency lower than a given threshold Smin that establishes the minimum requirement for proper service provisioning. It is also assumed that a UE will only try to connect to an activated RUE if it is in outage with its serving BS (i.e., $S_D<S_{min}$). In this case, the UE will try to connect to the activated RUE, providing the highest spectral efficiency SR. It is also assumed that RUEs with spectral efficiency less than Smin are not available for activation.

Focusing on the RUE activation problem, let us denote a given BS in the scenario as b. There are a total of R candidate RUEs associated with this BS b in the database. The candidate RUEs are numbered r=1,...,R. The *r-th* candidate RUE has a status mode denoted $a_{b,r}$ ={0,1}, where 0 means that the RUE is deactivated and 1 means it is activated. Therefore, the global status mode configuration associated with BS b can be defined as the R-length vector $C_b=\{a_{b,r}\}$. The objective of the considered problem is to find a policy that optimally activates the RUEs. This means finding the optimum configuration $C_b(t)=\{a_{b,r}(t)\}$ to be applied at every time t when the RAC decides on the activation or deactivation of the existing RUEs. It is assumed that these decisions are made in discrete time instants t with a granularity of ΔT. These discrete times are denoted as *t*, *t+1*,...,*t+k*,...

The criterion to consider a configuration $\mathbf{C_b}(t)$ as optimum is based on the efficiency of each candidate RUE when activated or deactivated. If the r-th RUE is active, i.e., $a_{b,r}(t)$=1, the efficiency $E_{b,r}(a_{b,r}(t))$ accounts for the average number of UEs in outage $N_{b,r}$ served by this RUE until the next decision period. Specifically, $E_{b,r}(a_{b,r}(t))$ =1 if $N_{b,r}\geq x$, where x is a certain threshold, meaning that the activation is worthwhile as the RUE has served a significant number of UE. Instead, the efficiency will be 0 if the RUE has served less than x UEs on average. On the other hand, if the RUE is inactive, i.e., $a_{b,r}(t)$=0, the efficiency is computed based on $P_{b,r}$, which is defined as the average number of UEs that would have been served by the RUE if it had been active during the period [t, t+1]. Then, the efficiency $E_{b,r}(a_{b,r}(t))$ will be equal to 1 if $P_{b,r} < x$, meaning that in this case, it is efficient not to have the RUE active. Similarly, it will be $E_{b,r}(a_{b,r}(t))$ =0 if $P_{b,r} \geq x$, meaning that in this case, keeping the RUE inactive is not efficient, as it could serve a significant number of UEs. In view of the above, the formal problem to solve is to find at every time t the configuration $\mathbf{C_b}(t)$ that maximizes the aggregate global efficiency AGE, which is defined as follows:

$$AGE = \frac{1}{R}\left[\sum_{r=1}^{R} E_{b,r}(a_{b,r}(t))\right] \tag{15}$$

5.4. DQN-based Approach for Optimal Relay UEs Activation

The development of an efficient solution to the RUE activation problem involves a multiplicity of variables, such as the current status mode of the RUEs, the propagation conditions of the RUEs and the nearby UEs, and the traffic dynamics. To address these multiple dimensions, this presented model proposes the use of Deep Reinforcement Learning. DRL techniques combine the use of deep neural networks (DNNs) with reinforcement learning (RL) to assist a software-based agent that makes decisions in relation to a specific problem. This combination is especially interesting because of its capability to handle large state and action spaces. DRL techniques have been applied in many different fields, such as robotics, video processing, and gaming, demonstrating outstanding success, as noted in [98]. Among the different DRL techniques, this work specifically proposes a solution based on the DQN algorithm [58]. In this approach, the learning process is carried out dynamically by a DQN agent that interacts with an environment, and after observing the consequences of its actions measured in terms of a certain reward signal, it learns to modify its own decision-making behaviour.

The DQN algorithm has been selected to address the RUE activation problem mainly for two reasons. The first is that the DQN algorithm has been designed to support high-dimensional states and action spaces. This is convenient for the RUE activation problem since the network dynamics implies a large amount of data that needs to be considered by the agent. The second reason is that with DQN, the policy is progressively updated by considering individual samples of experience as opposed to other methods such as Monte Carlo simulations [98] that update the policy by considering multiple samples collected during an episode. This feature is suitable for the case of the RUE-activation problem since continuous learning of the policy is desired. Moreover, DQN is a useful technique for learning how to select actions from discrete action spaces, as in the problem considered here where the actions involve activations or deactivations of RUEs. A variety of works have approached different RAN-related problems by means of the DQN technique. For example, DQN was used for capacity sharing in [65], while in [66], it was used to address the resource allocation problem in heterogeneous networks. Similarly, in [99], DQN was applied to the multi-connectivity problem, and in [67], it was used for spectrum sharing.

In the proposed solution, the DQN agent is located at the RAC (see Fig. 31) that makes decisions for the RUEs associated with the BSs in the scenario. At time t, the DQN agent of BS b selects an action $\boldsymbol{a(t)}$ that contains the RUE activation configuration $\mathbf{C_b}(t)$ to be applied to the set of RUEs in the next time window of duration ΔT. The selection of a given action is dependent on the state observed at time t denoted as $\boldsymbol{s(t)}$ together with the available policy π at that time. The state is obtained by processing the data from the network monitoring module and the candidate RUE database at the data processing module located in the RAC. This processing is needed to adapt the information to the format required by the DQN agent.

The outcome of applying a certain action that defines a RUE activation configuration is assessed by means of a reward signal $r(t+1)$. This reward is delivered to the DQN agent at the end of the time window ΔT. It essentially measures how effective or ineffective the selected action was. The reward signal obtained over time after selecting the different actions is utilized to progressively enhance the DQN-agent decision-making policy. The main components of the proposed DQN-based solution along with the policy learning process are described in next sections.

5.4.1. State, Action and Reward Definition

The state $\boldsymbol{s}(t)$ is represented as a vector associated with a particular BS b, and it has different components listed in the following:

- $S_{eff}(t)=\{S_{b,1}(t), S_{b,2}(t),\dots,S_{b,R}(t)\}$ represents the spectral efficiency of the RUEs in BS b, computed according to (13).
- $C_b(t)=\{a_{b,1}(t),a_{b,2}(t),\dots,a_{b,R}(t)\}$ denotes the configuration of all RUEs at time t.
- $N_b(t)=\{N_{b,1}(t), N_{b,2}(t),\dots, N_{b,R}(t)\}$ corresponds to the average number of UEs in the outage that have been served by each RUE measured at time t.

The total number of components in the state is $3\cdot R$.

A given action $\boldsymbol{a}(t) \in \mathcal{A}$ can be seen as a vector $C_b(t)=\{a_{b,r}(t)\}$ that contains the RUE activation configuration applied every time window accounting for all the considered RUEs. The so-called action space $\mathcal{A}$ contains all eligible RUE activation configurations. Since an RUE has only 2 possible modes (activated and deactivated), the total number of possible actions in the action space is 2^R.

The reward signal $r(t+1)$ that assesses the efficiency of the action $\boldsymbol{a}(t)$ is selected for state $\boldsymbol{s}(t)$ in relation to the optimization criterion (15). Thus, the reward can be expressed as the obtained value of AGE:

$$r(t+1)=\frac{1}{R}\left[\sum_{r=1}^{R} E_{b,r}(a_{b,r}(t))\right] \tag{16}$$

5.4.2. Policy Learning Process

A stage of major relevance when applying the DQN technique is the training of the DQN agent. By means of the training, the agent actively learns a decision-making policy π that is used for selecting which action to apply under a particular state. The training process considered in this work is based on the algorithm presented in [58], but customized the DQN agent according to the previously defined state, action and reward. The overall training process of the DQN is the same that has been already described in Section 3.4.2. The pseudocode that summarizes the above-mentioned procedure is presented in Algorithm IV, which has been adapted to outline the DQN-agent training procedure for the relay activation problem of BS b. The duration of the training procedure is given by the parameter MaxNumberOfTrainingSteps. The acronyms and notation of the proposed DQN-based solution are summarised in Table IV.

Algorithm IV. DQN training for BS *b*

Initialize DNN counter p=0.
For t=0… MaxNumberOfTrainingSteps
  Collect state $\boldsymbol{s}(t)$ (see section III.A)
  Generate a random number ε' between 0 and 1.
  If $\varepsilon' < \varepsilon$ (where ε=1 if t<=InitialCollectSteps)
    Select a random RUE activation configuration $\boldsymbol{a}(\boldsymbol{t})$.
  Else
    Get a RUE activation configuration $\boldsymbol{a}(t)$ based on π.
  End if
  Compute reward $r(t+1)$ and $s(t+1)$ as a function of action $\boldsymbol{a}(\boldsymbol{t})$.
  If D is full (l samples are stored)
    Delete the oldest experience.
  Store experience < $\boldsymbol{s}(t)$, $\boldsymbol{a}(t)$, $r(t+1)$, $\boldsymbol{s}(t+1)$> in D.
  Randomly sample a minibatch of experiences $U(D)$ from D of length J.
  Compute the loss function $L(\theta)$.
  Compute the mini-batch gradient descent $\nabla L(\theta)$.
  Update weights θ of evaluation DNN using (10).
  If p==P
    Update the weights of target DNN $\theta^{-}=\theta$ and set p=0.
  Else
    $p=p+1$
  End if
End for

5.5. Performance Evaluation

5.5.1. Considered Scenario

The studied scenario is a 200 m × 200 m square area consisting of one 5G NR BS and four UEs able to act as relays. This is illustrated in Fig. 37. The key parameters of both BS and RUEs are shown in Table V. The traffic generation of the different UEs is based on a Poisson session arrival process with an average session generation rate of 0.6 sessions/s and an exponentially distributed session duration with an average of 120 s. A UE remains static for the entire duration of its session.

TABLE IV. BS CONFIGURATION PARAMETERS

Parameter	Value	
Type of RAT	5G NR	Relay UE (RUE)
Position [x, y] m	[100,100]	[164,93] [100,180] [30,130] [100,20]
Frequency	26 GHz	3.5 GHz
Channel bandwidth	100 MHz	100 MHz
Transmitted power	21 dBm	21 dBm
Antenna gain	26 dB	3 dB
Height	10 m	1.5 m
UE antenna gain	10 dB	
UE noise figure	9 dB	
UE height	1.5 m	
Path loss model	Model of Sec 7.4 of [70]	UE-to-UE propagation model of [100]

The traffic spatial distribution assumes that 40% of the UEs are randomly located inside two square hotspots of 15 m × 15 m and 20 m × 20 m, located in the upper and lower centre of the scenario area, respectively (see Fig. 37.). The remaining UEs are randomly distributed in the whole scenario. The model assumes a connectivity model in which a given UE attempts to connect via an RUE just in the case of being in outage with its corresponding BS (i.e., $S_D < S_{min}$). The model has been developed in Python by using the *TF-agents* library [69], which provides tools for the development of DRL models, including DQN. The DQN parameters are detailed in Table VI.

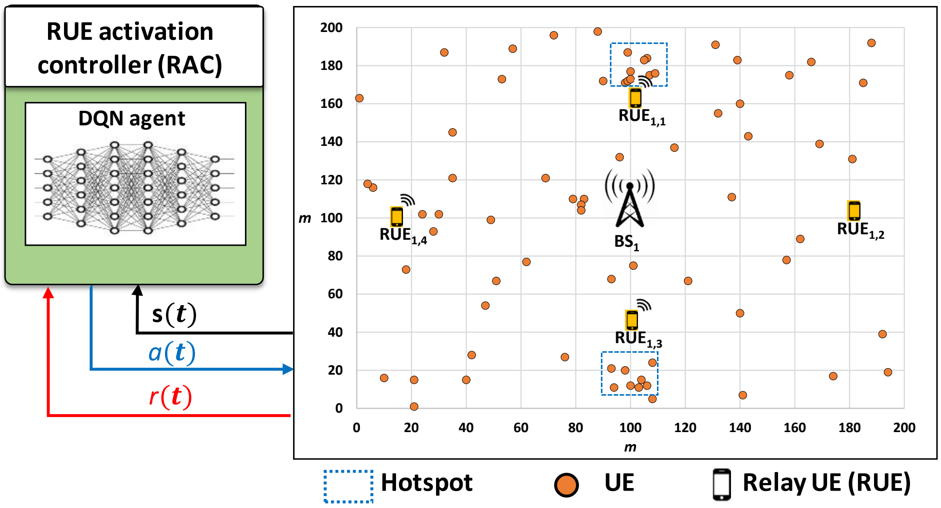

Fig. 37. Graphic representation of the scenario used for training and evaluation

TABLE V. DQN CONFIGURATION PARAMETERS FOR RELAY ACTIVATION ALGORITHM

Parameter	Value
Initial collect steps	500
MaxNumberOfTrainingSteps	200000
Experience Replay buffer maximum length (l)	$100 \cdot 10^3$
Mini-batch size (J)	64
Time window (ΔT)	30 sec
DNN updating period (P)	500 Training Steps
Discount factor (τ)	0.9
Learning rate (α)	0.0003
ε value (ε-Greedy)	0.1
DNN architecture	Input layer: 12 nodes Two hidden layers: 100 and 50 nodes Output layer: 16 nodes

5.5.2. Key Performance Indicators

This section describes the KPIs considered for assessing the performance of the proposed solution:

- **Average reward R_w**: This measures the average of the reward values (i.e., the *AGE* values) obtained for all the actions taken by the DQN agent during the evaluation of the learned policy, which is:

$$R_w = \frac{1}{N}\sum_{n=1}^{N} r(t_n) \tag{17}$$

where N denotes the total number of actions selected during the evaluation and $r(t_n)$ is the reward obtained as a result of the action made at time t_n, n=1,...,N.

- **RUE time in active mode M_T:** This measures the total cumulative time that all RUEs have been active during all evaluations, that is:

$$M_t = \sum_{r=1}^{R}\sum_{n=1}^{N} \Delta T \cdot a_{b,r}(t_n) \tag{18}$$

- **Efficiency of RUEs in active mode F:** This KPI measures how much of the time that an RUE is in active mode is actually efficient, meaning that it has served at least x=1 UE on average. If we define $K(t_n)$ as the total number of activated RUEs given an action $a(t_n)$, the KPI is given by:

$$F = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{K(t_n)}\sum_{k=1}^{K(t_n)} Ae_k \tag{19}$$

- **Average Spectral efficiency S:** This KPI measures the average spectral efficiency obtained during the evaluation process. For this purpose, during each time period (t_n-ΔT, t_n), we measure the spectral efficiency obtained by all the UEs with an active session, taking one sample per UE every second, resulting in a total of $N_{samp}(t_n)$ samples denoted as $s_{eff}(i,t_n)$. Then, the average spectral efficiency is given by:

$$S = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{N_{samp}(t_n)}\sum_{i=1}^{N_{samp}(t_n)} S_{eff}(i,t_n) \tag{20}$$

- **Outage probability O:** This KPI computes the outage probability of a policy during the evaluation process. For this purpose, during each time period (t_n-ΔT, t_n), we take one sample every second for each UE with an active session, where the *i-th* sample is $o(i,t_n)$=1 if the UE experiences a spectral efficiency lower than S_{min}=1 b/s/Hz and $o(i,t_n)$=0 otherwise. Then, denoting the total number of samples for each period as $N_{samp}(t_n)$, the outage probability is given by:

$$S = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{N_{samp}(t_n)}\sum_{i=1}^{N_{samp}(t_n)} o(i,t_n) \tag{21}$$

5.5.3. Assessment of the Training Stage

The training process of the DQN algorithm is assumed to be performed offline in order to learn the policy that will be applied later on in the real or physical network in the so-called inference stage. This approach is aligned with recent initiatives such as O-RAN [101], which consider that ML models are trained on a training host before deploying them in the physical network. On the one hand, the offline training allows that the DQN agent is exposed to a wide variety of network conditions during the training and, on the other hand, it avoids that wrong decisions made during training negatively impact the performance of the physical network. To assess the training stage, this chapter considers a training scenario in which UEs act according to the operation and parameters presented in Section 5.5.1. As explained in Section 5.4., in each training step of ΔT seconds, the DQN agent selects a given action $\boldsymbol{a}(t)$ that results in the activation or deactivation of the available RUEs. After applying an action, the scenario continues its normal operation, and the reward is measured to progressively update the policy. The training was executed until reaching the maximum number of training steps MaxNumberOfTrainingSteps=$2\cdot10^5$.

In the following, some results are presented to study how the policy is progressively improved during the learning process. For this purpose, every 500 training steps, we obtained the current policy of the DQN agent, and we executed an evaluation of this policy by applying it for one hour in a given evaluation scenario that corresponds to a certain realization of the random traffic generation and spatial traffic distribution processes (in this way, the evaluations of all the policies every 500 training steps were performed under the same conditions, so they were comparable).

As a result of this one-hour evaluation, we collected the average reward. Fig. 38. plots the evolution of this average reward obtained with the learned policies every 500 training steps. The results show that during the first $80\cdot10^3$ training steps, the behavior of the learned policy is quite unstable. In fact, at approximately $75\cdot10^3$ steps, there is a significant decrease in the performance, reflecting that the training with this number of steps is still insufficient and the policy has not learnt to select optimal actions yet. However, after this period, the average reward starts to increase and tends to stabilize after approximately $175\cdot10^3$ training steps. It is worth mentioning that, being an offline training, the performance degradation observed at $75\cdot10^3$ does not have any impact on the physical network. Instead, it is the policy obtained at the end of the training process the one that will determine the performance in the physical network, as studied in the next sub-section.

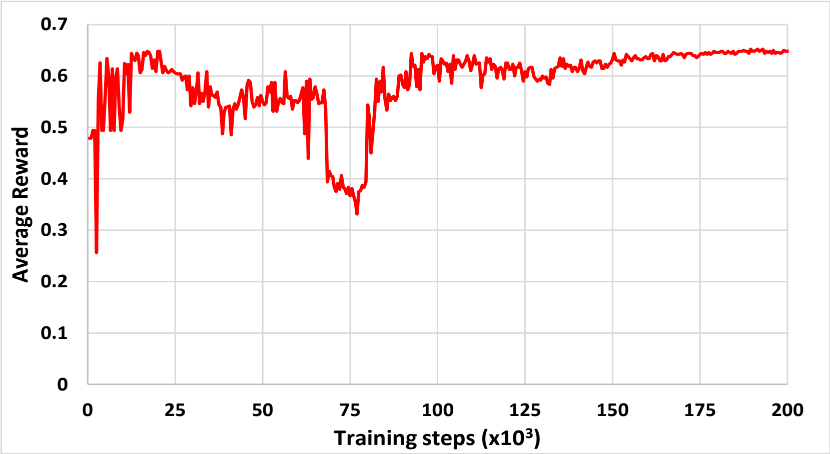

Fig. 38. Evolution of the average reward as a function of the training steps.

5.5.4. Reference Strategies for Benchmarking

This section assesses the performance obtained by the RUE activation in the inference stage using the policy learned by the DQN agent after completing the training process. For benchmarking purposes, the following reference strategies of RUE activation were considered:

1. **Random RUE activation:** this strategy consists of randomly selecting an action $\boldsymbol{a(t)}\in\mathcal{A}$ in every time window ΔT. The actions are selected with equal probability. Therefore, an RUE has a probability of being activated of 50% in every time window.

2. **All RUEs activated:** this strategy maintains all the RUEs activated during the whole time of evaluation. Therefore, this strategy will provide the best possible performance in terms of average spectral efficiency and outage probability, but at the cost of having all RUEs always activated even if they may not be necessary during certain periods of time.

3. **All RUEs deactivated:** this strategy keeps deactivating the RUEs during the entire evaluation time, so it can be considered the classical RAN in which no relay UE capabilities are exploited.

4. **Genetic algorithm:** A genetic algorithm is an optimization technique inspired by the principles of natural selection and genetics [102]. It is used to find solutions to complex problems by mimicking the process of evolution. This strategy is executed in every time window ΔT to decide the combination of relays to be activated/deactivated at that time. The algorithm starts from an initial population of potential solutions that consists of N_p individuals (i.e., actions $\boldsymbol{a}(t) \in \mathcal{A}$ in the problem considered here). These individuals are in turn constituted by a number of genes G, where each gene corresponds to the value of $a_{b,r}(t)$ for the r-th RUEs in BS b. Each individual is evaluated in terms of their fitness, which measures its suitability to solve the given problem. Specifically, the value of fitness considered here is the AGE in (15). Then, through a process of selection, crossover and mutation (modelled with probability of mutation P_{mut}), new generations of individuals are created (details can be found in [102]). This cycle continues for several generations, gradually improving the quality of the solutions. The algorithm terminates when reaching a maximum number of generations N_{gen}. Then, the final solution is the individual with the highest fitness found in all the generations. The considered parameters for the genetic algorithm are N_p=8, G=4, P_{mut}=0.1 and N_{gen}=100.

5. **Algorithm 'G' from [95]:** this algorithm intends to activate the RUEs that minimize the outage while maximizing the spectral efficiency. To achieve the latter, each UE in outage is associated with the candidate RUE that provides the highest value of spectral efficiency and fulfils S_R>$S_{min.}$ After analyzing this for all the UE devices, the strategy activates all the relays that could serve at least one UE device. This strategy assumes perfect instantaneous knowledge of the spectral efficiency conditions for all the active UEs in the links with their serving BSs and in the links with all the candidate RUEs.

6. **Exhaustive search strategy:** this strategy applies at each time step an exhaustive search process among all possible configurations $\mathbf{C_b}(t)$ to find the one that ensures the maximum aggregate global efficiency AGE, thus maximizing the target of the optimization problem formulated in Section 5.3. This exhaustive search strategy also assumes perfect instantaneous knowledge of the spectral

efficiency conditions of all the active UEs in the links with their serving BSs and in the links with all the candidate RUEs.

5.5.5. Evaluation of the Learnt DQN-based policy

The comparison between RUE activation strategies was performed by conducting 100 different policy-evaluation procedures of one-hour duration for each strategy. Each evaluation procedure consisted of applying the assessed strategy under a given realization of traffic generation from the UE in the scenario. Then, by varying the traffic generation realization in each evaluation procedure, the performance of the different strategies was assessed in a wide range of situations in the considered scenario. It is also worth noting that, for a given evaluation procedure, all the strategies were applied with the same traffic realization, so that they were evaluated under exactly the same conditions.

Fig. 39. shows the average reward or, equivalently, the average value of the *AGE* obtained as a result of all the evaluations for the different strategies. The *Exhaustive search strategy* corresponds to the theoretical upper bound of the achievable performance since it assumes perfect instantaneous knowledge of the spectral efficiency conditions in all the different links. *Algorithm 'G'* performs very similarly to the *Exhaustive search strategy* because it also considers the same theoretical assumption of perfect knowledge. The difference can be explained as follows. The *Exhaustive search strategy* is optimized to maximize the *AGE*, while *Algorithm 'G'* targets outage minimization and spectral efficiency maximization.

The results reflect that the *DQN-based strategy* achieves an efficiency equivalent to 84.7% of the *Exhaustive search strategy* and 86% of *Algorithm 'G'*. When compared to the other strategies different from the *Exhaustive search* and the *Algorithm 'G'*, it is observed from Fig. 39. that the *DQN-based strategy* outperforms all of them. Although the result with respect to the *All RUEs deactivated* strategy was expected, when compared to the *random* strategy, the *DQN-based strategy* is 119% more efficient in terms of the reward. Moreover, the comparison with the *All RUEs activated* strategy shows that maintaining all the relays activated during the whole evaluation time does not result in an efficient strategy. In fact, our proposed algorithm outperforms the *All RUEs activated* strategy by 75% since this strategy keeps some RUEs activated even when this is not needed according to the network dynamics.

The comparison between the *Genetic algorithm* and the *DQN-based strategy* in Fig. 39. shows that DQN achieves a bit better performance with an average improvement of 2.6%. This reflects that both algorithms properly solve the optimization problem. However, from an implementation perspective, an importance difference between both approaches is that the *Genetic algorithm* has to conduct an optimization every time that a decision has to be made and this involves an evolutionary search process to find a solution. In contrast, the *DQN-based strategy* benefits from the experience acquired during the training, so that it makes decisions much faster. In fact, using a computer with a Core i5-6400 2.7 GHz processor and 8 GB of RAM, and considering all evaluation procedures, it has been obtained that an execution of the DQN agent in the inference stage lasts on average 0.017 ms, while the *Genetic algorithm* requires on average 104.3 ms.

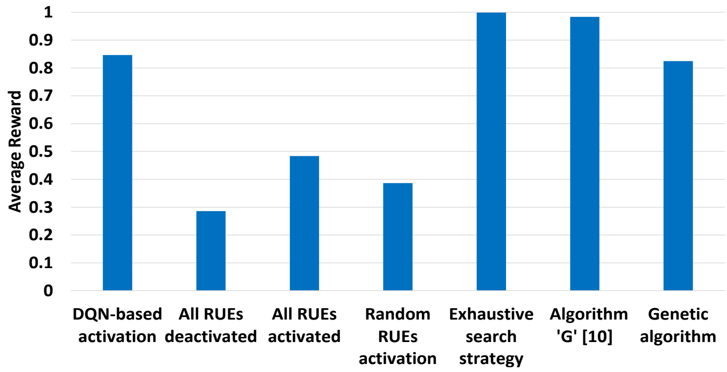

Fig. 39. Average reward for the different strategies.

To deeply analyse the performance of the different strategies in a wide range of situations and to explore how the improvements achieved by the *DQN-based strategy* vary in different policy evaluation procedures, Fig. 40. plots a boxplot of the percentage of reward increase achieved by the DQN-based strategy with respect to all the other strategies considering the 100 policy-evaluation procedures. The boxplot reflects the distribution of this reward improvement and allows us to establish the ranges of best and worst performance. Specifically, for each strategy, the top and bottom lines shown in the plot reflect the maximum and minimum values, while the box represents the range between the 25th and 75th percentiles of the distribution. With respect to the upper bound strategies *Exhaustive search strategy* and *Algorithm 'G'*, the boxplot shows that, although they obtain a higher reward than the *DQN-based strategy*, the differences are small. In the best case, the reward reduction of the *DQN-based strategy* with respect to the *Exhaustive search strategy* is only 4.5%, and it is 2.7% with respect to *Algorithm 'G'*. In the worst case, the reward reduction is 30% with respect to the *Exhaustive search strategy* and 28% with respect to the *'G' algorithm*. The median reward reduction is 14.9% for the *Exhaustive search strategy* and 14.8% for *Algorithm 'G'*.

When compared with the rest of the strategies, the median of the DQN improvement is 1.89% against *Genetic algorithm*, 75.6% against the *All RUEs activated* strategy, 119% against the *Random RUEs activation* strategy and 191% against the *All RUEs deactivated* strategy. In the best situations, the improvements are 20.1%, 105%, 178%, and 373%, respectively, while in the worst cases, they are -16.5%, 42.6%, 61.6%, and 77.7%. These numbers confirm that, thanks to the training process, the DQN-agent is able to learn when the activation of the RUEs will be beneficial, which in turn will translate into a benefit for both network users and MNOs.

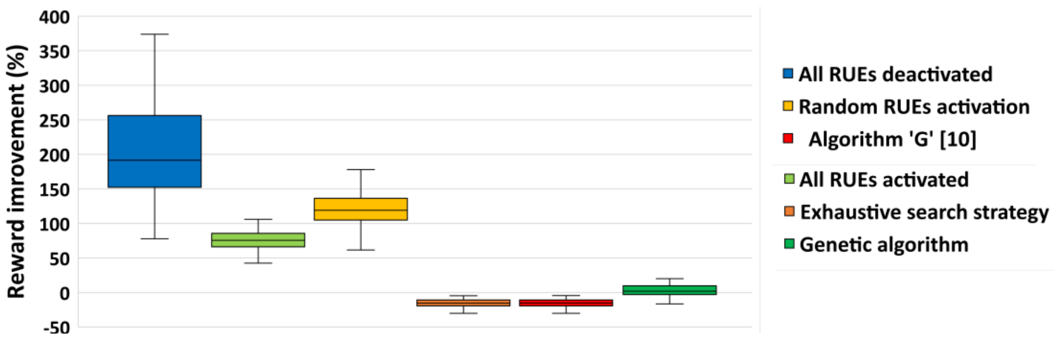

Fig. 40. Boxplot of the reward improvement of the DQN-based strategy against the benchmarking strategies for the 100 policy-evaluation procedures.

The RUE time in active mode is shown in Fig. 41., where we compare the total time for different strategies. For the *All-RUEs activated* strategy, the RUEs are activated for a total of 400 hours (i.e., 4 RUEs activated during 100 hours each), while for the *Exhaustive search strategy* and *Algorithm 'G'*, the RUEs are active for a total of 196.7 hours and 197.7 hours, respectively. These numbers are very close to the *DQN-based strategy*, which has the RUEs active for a total time of 198.75 hours, i.e., a difference of less than 1% with the upper bound strategies. This value is also similar to the one obtained with the random strategy because this strategy tends to activate on average approximately the same number of RUEs, but its random behaviour leads to poorer performance in terms of *AGE*. A similar situation occurs when comparing with the *Genetic algorithm,* since it maintains RUEs active during 191.7 hours but at the cost of a slightly worse performance in terms of *AGE* as seen in the previous results of Fig. 40 and 41. It is clear that by performing a proper activation, the DQN strategy performs closely to the upper bound strategies and clearly reduces by more than 200 hours the time that RUEs remain in active mode with respect to the *All-RUEs activated* strategy. This will translate into significant energy savings.

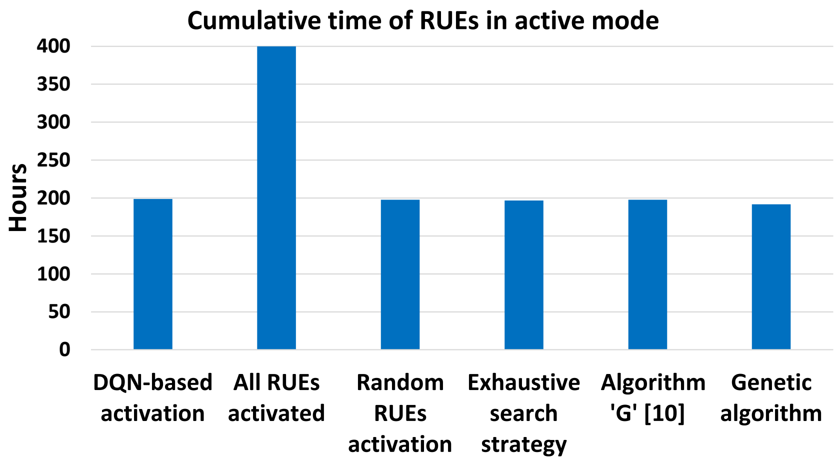

Fig. 41. Time of RUEs in active mode for the different strategies.

Beyond this time reduction, it is also important to measure how much of the time that an RUE is activated is actually useful. For this purpose, Fig. 42. shows the efficiency of RUEs in active mode considering the different activation strategies. As expected, the highest efficiency values correspond to the upper bound strategies, where *Algorithm 'G'* reaches a value of 98.2% while the *Exhaustive search strategy* reaches 100%. The *DQN-based strategy*, on the other hand, has an efficiency of 87.1% and the *Genetic algorithm* reaches an efficiency of 85.5%. In contrast, the efficiency is substantially lower with the *Random RUEs activation* and *All-RUEs activated* strategies, whose values are 46.2% and 48.3%, respectively. Thus, the *DQN-based strategy* is slightly superior to the *Genetic algorithm* and achieves almost twice the performance of the *Random RUEs activation* and *All-RUEs activated* strategies due to the capability of the DQN-based algorithm to consider network context information when deciding which RUEs to activate. Additionally, the comparison against the *All-RUEs activated* strategy reflects that keeping the RUEs activated all the time does not result in an efficient method. The reason is that, given the network dynamics, at certain times, it is not necessary or efficient to use relays; therefore, activation may result in a waste of resources.

To assess the network performance improvement brought by the activated RUEs, Fig. 43. plots the obtained outage probability for the different strategies evaluated in this chapter. A relevant result extracted from Fig. 43. is the significant difference between using or not using relays. For the case when there are no activated relays in the network, the outage probability is 10%, while by using the *DQN-based activation* strategy, the outage is reduced down to approximately 1.1%. In other words, by using and activating relays correctly as the DQN algorithm does, the outage probability is reduced by approximately 89%. This is a stronger reduction than the one obtained with random activation, which provides an outage of approximately 5.1% or a reduction of only 49%.

Note that in the case of *Algorithm 'G'*, *Exhaustive search strategy* and *All-RUEs activated* reduce the achieved outage probability to almost 0. For the case of the *All-RUEs activated* strategy, it was expected because all relays are always active, but as seen previously, this is at the expense of a much longer time in active mode. The *DQN-based activation* is superior to the *Genetic algorithm* at minimizing the outage probability. Indeed, the *DQN-based activation* is able to reach an outage probability similar to the *All-RUEs activated* strategy but with 50% less time spent by relays in active mode. It is also worth remarking that from the perspective of outage, the performance of *Algorithm 'G'* is better than that of the *Exhaustive search strategy* because the former intends to minimize the outage, while the *Exhaustive search strategy* targets the maximization of the *AGE*.

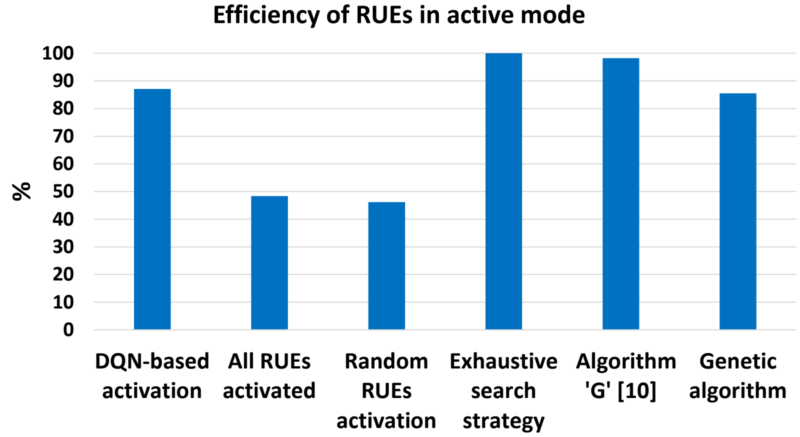

Fig. 42. Efficiency of RUEs in active mode for the different strategies.

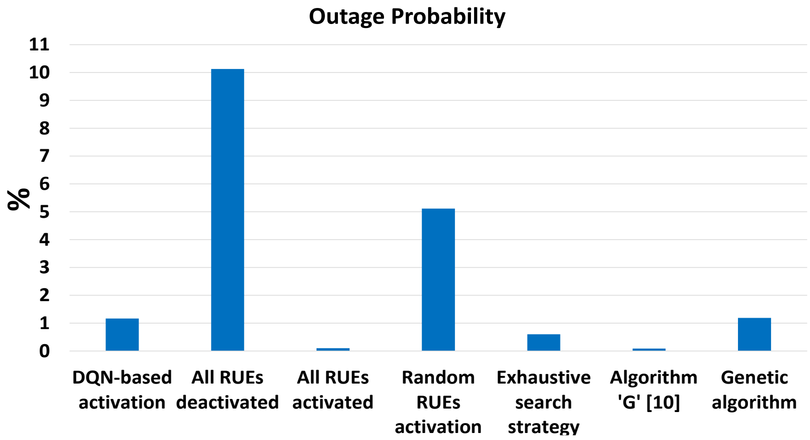

Fig. 43. Outage probability of UEs for different activation strategies.

Fig. 44. plots the average spectral efficiency obtained by applying the different strategies. The *DQN-based strategy* clearly outperforms the *random* and *All RUEs-deactivated* strategies, with improvements of 5.3% and 12.3%, respectively and also showed some improvement against the *Genetic algorithm*. In turn, the *All RUEs Activated* strategy provides the highest spectral efficiency but is only 1.2% higher than with the *DQN-based strategy*, which achieves very similar performance to the *Exhaustive search strategy* and *Algorithm 'G'* strategies. Considering that the *DQN-based strategy* keeps the relays active for less than half of the time than the *All RUEs Activated* strategy did, it is concluded that this strategy achieves a better trade-off between spectral efficiency improvement and RUE activation time.

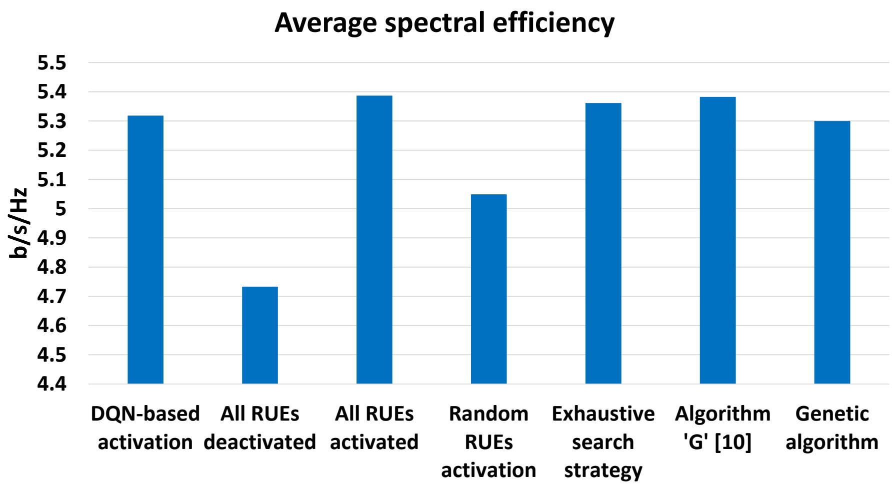

Fig. 44. Average spectral efficiency for different activation strategies.

To assess the range of spectral efficiency values, Fig. 45. plots the cumulative distribution function (CDF) of the spectral efficiency for the different activation strategies. The CDF as a function of a given value of spectral efficiency, s_e, is given by the probability that a spectral efficiency sample $s_{eff}(i,t_n)$ is lower than s_e:

$$CDF(s_e) = P\left(s_{eff}(i,t_n) \leq s_e\right) \tag{22}$$

this computation is performed considering all the spectral efficiency samples taken during the 100 policy-evaluation procedures. When comparing the DQN strategy against both upper bound references, the good performance of our proposed approach is exhibited since a very similar CDF is observed. In fact, if we measure the probability that a given UE experiences a spectral efficiency of at least 5 b/s/Hz, with the *Exhaustive search strategy* and *Algorithm 'G',* this probability is 84.1% and 86%, respectively, while with the *DQN-based strategy*, it is 83%, which is a small difference. It is worth emphasizing the importance of proper activation of the relays that our DQN approach achieves. When performing an activation of the relays based on the *Genetic algorithm*, the probability of experiencing a spectral efficiency of at least 5 b/s/Hz is 79% and takes a value of 56% when performing a random activation. Overall, a significant difference is observed between using or not using RUEs. In fact, it can be observed that when there are no RUEs activated, the probability of having a spectral efficiency higher than 5 b/s/Hz is only 25%, which is much lower than that of the *DQN-based activation strategy*.

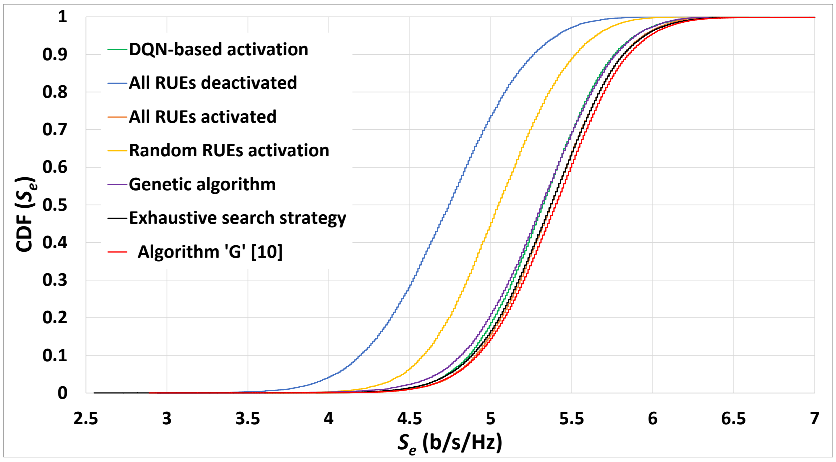

Fig. 45. CDF of the spectral efficiency obtained for different activation strategies.

5.6. Concluding Remarks

In this chapter has been presented a novel approach for exploiting UE's capability to be activated as relays (RUEs) as a way of augmenting the radio access network (RAN) infrastructure and thus increasing the network coverage and improving the network and UE performance. Specifically, in the chapter has been proposed a new strategy for efficiently deciding when to activate or deactivate a certain relay UE. The activation function is based on the deep Q-network algorithm that makes use of the network context information and a training process to learn a policy for activating only those RUEs that are useful under the given the network conditions.

The behaviour of the proposed DQN-based RUEs activation function has been assessed by means of system-level simulations and contrasted against six other reference strategies. The results have shown that, the behaviour of the proposed DQN-based RUE activation function has been assessed through system-level simulations and compared against six other reference strategies. The results demonstrate that the DQN-based activation strategy can learn from network context information and appropriately activate the relays. This is reflected in high performance in terms of aggregate global efficiency. Specifically, the policy learned by the DQN agent substantially reduces the time relays remain active and significantly increases the average global efficiency. Additionally, the proposed approach leads to important reductions in outage probability and increases in spectral efficiency compared to strategies without RUEs and those where RUEs are activated randomly. Notably, the average spectral efficiency improves by 12.3%, and the outage probability is reduced by 89% relative to the case without RUEs. Moreover, the DQN-based activation provides performance very similar to the strategy that keeps the RUEs activated all the time, even though the relays are active less than 50% of the time. When compared to a classical optimization strategy such as a genetic algorithm, the DQN approach exhibits slightly better performance and a significant reduction in decision-making execution time. Furthermore, the results show that the performance of the DQN strategy is quite close to that of two theoretical upper-bound strategies that operate with perfect instantaneous knowledge of all the link conditions.

To summarize the overall performance of the proposed DQN strategy, Fig. 46. plots a radar chart displaying the normalized results for all the studied metrics comparing the DQN-based strategy, against the *All RUEs-deactivated strategy* and the *Exhaustive search strategy*. The performance of the DQN approach has only minimal deviations from the *Exhaustive search strategy*, and it achieves important improvements with respect to the *All RUEs-deactivated* strategy. This highlights two main findings: firstly, it emphasizes the network performance benefits gained through the efficient use of relays; secondly, considering that the exhaustive search approach acts as a theoretical upper bound benchmark, the similarity between the DQN-based approach and this upper bound confirms the effectiveness of our proposed approach for tackling the relay activation problem.

Following the promising results obtained, our future work intends to study the performance of the model based on actual network measurements. Moreover, the identification of the mechanisms required for the practical implementation of the proposed solution, including the implications on current management interfaces and message exchanges between involved nodes, is also envisioned as future work.

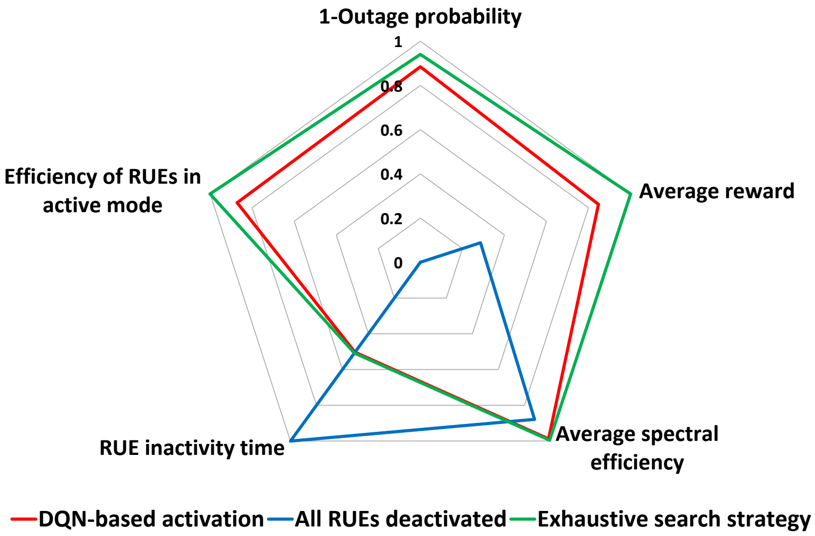

Fig. 46. Radar plot of the performance obtained by different activation strategies for the considered normalized metrics.

Chapter 6.Conclusions and Future Work

6.1. Conclusions

The emergence and deployment of fifth generation mobile networks is expected to facilitate the proliferation of many new services associated with various vertical industries, each having particular requirements in terms of latency, reliability, data rate, etc. To effectively meet these demands, service providers must ensure the efficient operation of their networks. In mobile networks, the coverage is one of the most important aspects, since both network performance and user experience depend entirely on it. Commonly, mobile networks face coverage issues due to a variety of reasons, including suboptimal planning, interference, physical obstacles, etc. In order to meet the stringent requirements of services associated with 5G and future systems such as 6G, operators are called upon formulating and implementing strategies to ensure optimal connectivity conditions for network users. In parallel with the evolution of mobile networks, Artificial Intelligence (AI) is another area of technology that is rapidly evolving and promising for the future. This trend has been embraced by mobile operators, who are considering a variety of AI methods in all the processes involved in running a 5G mobile network. Indeed, in early discussions about the future 6G network (IMT-2030), AI is seen as a central element in the processes of planning, deploying, managing and optimising networks. In this context, this thesis has contributed with different AI-based solutions at a RAN level aimed at enhancing connectivity conditions for network users.

Firstly, in Chapter 3, a multi-connectivity solution aimed at optimising traffic split in a heterogeneous network environment (LTE+5G NR) has been proposed, developed and evaluated. The solution is based on a Deep Reinforcement Learning algorithm known as Deep-Q Network (DQN). The proposed solution intelligently splits the traffic from the UEs at the PDCP layer of the master node to different cells, based on the current traffic loads and radio conditions that the UE is experiencing, with the primary objective of satisfying UE QoS. The solution involves training a DQN agent to determine the multi-connectivity configuration that minimizes the fraction of total bandwidth allocated to the UE, while ensuring its throughput meets or exceeds the required bit rate, and simultaneously avoids congestion in the cells involved in the MC. Furthermore, the solution has been designed in such a way that can be implemented as a third-party application, referred to as xApp, within the framework of the Open-RAN (O-RAN) architecture. The presented solution has been assessed and compared against the optimum case and a classical SINR-based approach in scenarios, involving UEs following trajectories as well as stationary UEs. The obtained results have shown a significant capability of the DQN agent to learn a quasi-optimal policy outperforms the SINR-based approach in up to 50% in terms of reward in certain conditions. In turn, for the case of fixed positions, the proposed strategy achieved substantial throughput gains with respect to the SINR-based strategy. In addition, the statistics obtained highlight the algorithms' ability to learn when to use (or not) MC to maximise UE performance. In this way avoiding unnecessary MC processes. Furthermore, a scenario with multiple MC-capable users coexisting in the same evaluation environment was assessed. The results revealed significantly better performance for all MC-capable users when the solution was applied, reinforcing and confirming the importance of optimizing the MC configuration. Additionally, a study was conducted to analyze the robustness of the learned policy when applied with a required bit rate value different from the one considered during the training stage. It was revealed that the learned policy can function properly with variations in the required bit rate of around 20% to 30% of the value considered in the training. However, by conducting training that considered a wider range of required bit rate values, it was possible to enhance the performance of the obtained policy. Overall, the proposed solution demonstrated adequate performance in all evaluated scenarios. Its implementation would lead to higher performance for network users, thereby improving their connectivity conditions, aligning with the general objective of this thesis.

In Chapter 4, this thesis addresses the issue of coverage optimisation in 5G and beyond deployments as a method to improve service availability in the network, leading to ubiquitous connectivity, thereby improving the user experience. To achieve this goal, a methodology for detecting coverage-constrained areas in mobile network deployments was introduced. A coverage hole detector was proposed and developed, utilizing the DBSCAN clustering algorithm to detect, characterize, and validate the presence of coverage holes with high and sustained traffic using performance reports from network users as the source information. As a solution to mitigate the coverage problem at the identified CHs, the methodology also proposes the use of relays to extend the coverage of the Radio Access Network. This involved incorporating functionality to determine the optimal relay placement coordinates based on the locations of the identified coverage holes and the relay's path loss conditions with the BSs. To evaluate the performance of the proposed methodology in terms of network performance, a metric was formulated to jointly measure the average spectral efficiency and the level of service availability in the network. A comparison was made before and after applying the proposed methodology. In that respect, different system-level simulations replicating real spatio-temporal traffic patterns collected from the Campus Nord of the Universitat Politècnica de Catalunya have been performed. The solution successfully identified and validated five coverage holes on the campus. An updated network deployment with strategically placed relays was analyzed. The obtained results showed remarkable improvements in terms of service availability and spectral efficiency. When analysing performance at the coverage hole level, in terms of spectral efficiency, the improvement ranged from a factor of 6.9 to a factor of 34.5, depending on the coverage hole. In addition, service availability was increased up on average 95.8%. The obtained results of the proposed methodology in Chapter 4 highlight the critical role of coverage optimization in improving connectivity conditions and, consequently, user and network performance. This achievement is consistent with the central objective of this thesis.

Finally, Chapter 5 addressed the problem of how to optimally use the relaying capabilities of some UEs to augment the RAN in 5G and beyond deployments as an alternative for the fixed-relay solution, supported by the powerful computing and communication capabilities of today's UEs. This can be particularly useful in deployments operating in mmWave frequencies, as these deployments are more susceptible to experiencing coverage-constrained regions (e.g., coverage holes). The chapter introduced a DQN-based solution in which a DQN agent was trained to optimize decisions regarding when an RUEs should be activated or deactivated, based on current network context information. This approach was aimed at maximizing the benefits these RUEs can provide for increasing spectral efficiency and decreasing the outage probability of network users. By means of different system-level simulations, the proposed solution was evaluated and compared against six different activation strategies. The obtained results have demonstrated the capability of the proposed activation strategy to learn from the network context, as reflected in its high performance in efficiently activating RUEs. Additionally, the proposed strategy substantially reduced the time that RUEs remain active compared to other strategies, leading to a significant increase in activation efficiency. Moreover, the DQN-based strategy improved average spectral efficiency and substantially reduced outage probability compared to scenarios without RUEs. Notably, the proposed strategy exhibited performance similar to the strategy of keeping RUEs activated all the time, despite the relays being active for less than 50% of the time, thereby also representing energy savings. Overall, the results obtained have demonstrated a performance quite close to the two theoretical upper bound strategies. In particular, the similarity between the DQN-based approach and the exhaustive search approach, which serves as a theoretical upper bound benchmark, emphasises the network performance benefits gained through the efficient use of relays. This confirmation highlights the effectiveness of our proposed approach in tackling the relay activation problem. Overall, considering the solutions presented in this thesis and the results obtained, the thesis has successfully addressed and achieved its central objective of proposing, developing and evaluating various AI-based solutions for the 5G RAN aimed at ensuring optimal connectivity conditions for network users, thereby satisfying their Quality of Service (QoS) requirements.

6.2. Future Directions

This thesis has explored various solutions aimed at enhancing connectivity conditions in 5G and beyond radio access networks. From a general perspective, there are common areas for future work that can be considered for all of them. Firstly, a robustness and reliability assessment of the different solutions against various network conditions and potential anomalies would enforce the trustworthiness of the proposed solutions. This involves testing the solutions' resilience to network failures, fluctuations, and unexpected events. Secondly, it would be interesting to evaluate the solutions in controlled real-world scenarios embracing practical implementation aspects such as the impact on current management interfaces. This would allow performance to be monitored in terms of the key metrics of each solution. By using these metrics, it would be possible to identify practical challenges and areas for improvement.

Appendix 1.Employed DQN Hyperparameters Optimization

This appendix contains information regarding the process followed in the thesis for selecting the hyperparameters for the training stage when using the DQN algorithm. It's worth noting that, in general, the process of hyperparameter optimization is mostly iterative. It involves trying different configurations, such as the learning rate, discount factor, batch size, number of hidden layers in the neural network, number of neurons in the hidden layers, activation functions of the neurons, etc. Each one of these parameters may have an impact on the training stage and the learning process, ultimately affecting the performance of the obtained policy. Each neural network requires particular hyperparameter optimization, as the configuration of hyperparameters will also depend on the complexity of the problem being approached, the structure of the hidden layers, the number of possible outputs, etc. In the case of the relay activation problem approached in Chapter 5, the input layer had 12 nodes and the output layer had 16 values representing the different valid configurations of relay activation. The different configurations in Table VI were tested in separate training sessions, each consisting of 10,000 training steps. The results were compared in terms of reward and were graphically evaluated to observe how the reward and loss plots changed over the training steps.

TABLE VI. DIFFERENT DQN HYPERPARAMETERS CONFIGURATIONS

Configuration	Hidden Layers	Neurons per Layer	Learning Rate	Discount factor	Batch size
1	2	(20, 20)	0.001	0.99	64
2	2	(32, 32)	0.0001	0.95	32
3	1	(40)	0.003	0.98	128
4	3	(16, 16, 16)	0.005	0.99	64
5	2	(100, 75)	0.001	0.97	64
6	2	(50, 50)	0.0001	0.99	32
7	1	(40)	0.002	0.98	64
8	3	(16, 16, 16)	0.0005	0.95	64
9	**2**	**(100, 50)**	**0.0003**	**0.99**	**64**
10	2	(64, 32)	0.0001	0.97	128

When comparing the different configurations in Fig. 47, it is possible to observe how the reward values are distributed, what the smallest and largest values are, and the dispersion of the obtained values, which can reflect how the chosen configuration influences the training stage. Configuration number 9 exhibited the best performance in terms of average reward. Similarly, the variations in reward values were smaller compared to the other configurations such as 2 and 5, for instance, which means that configuration 9 tends to stabilize quickly and therefore converge. This is graphically illustrated in Fig. 47.

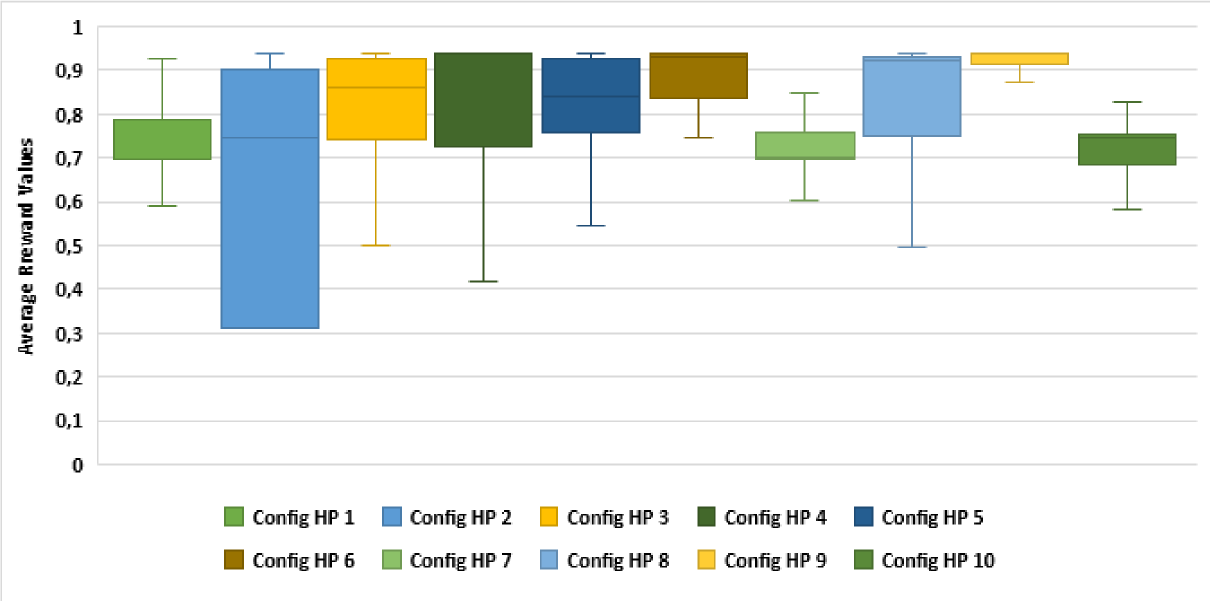

Fig. 47. Boxplot of the average reward values obtained for different DQN hyperparameters configurations

It is important to point out that there is no specific method that confirms 100% that the chosen hyperparameters are the best possible. Although there are methods for optimising hyperparameters, such as Grid Search, Random Search, Bayesian Optimisation, among others, and some libraries such as Hyperopt and Optuna, the configurations obtained still have a certain degree of uncertainty. In the case of this thesis, we carried out several training sessions with different configurations and considered the average reward values and the dispersion of values as the main criterion for deciding the optimal configuration.

It is important to clarify that the DQN parameters used in this thesis were obtained following the methodology described previously, and the selected configurations were the best among those tested. However, we cannot exclude the possibility that other configurations may perform similarly or better, for example in terms of faster convergence. Finally, the final confirmation for us to consider that the selected configuration works is based on the efficiency of the policy obtained, as well as the confirmation that convergence was achieved at some point, as shown, for example in Fig. 38.

References

[1] NGMN Alliance, “5G White Paper 2,” July 2020.

[2] ITU-R, “IMT Vision-Framework and overall objectives of the future development of IMT for 2020 and beyond”. Recommendation ITU-R M.2083, September 2015.

[3] ITU-R, “M Series: Mobile, radiodetermination, amateur and related satellite services. Framework and overall objectives of the future development of IMT for 2030 and beyond”. Recommendation ITU-R M.2160-0, November 2023.

[4] K. Samdanis, X. Costa-Perez and V. Sciancalepore, "From network sharing to multi-tenancy: The 5G network slice broker," in *IEEE Communications Magazine*, vol. 54, no. 7, pp. 32-39, July 2016.

[5] 3GPP TS 38.300 V16.4.0, “NR; NR and NG-RAN overall description; Stage 2 (Release 16),” January 2021.

[6] 3GPP TS 23.501 V15.2.0, “System architecture for the 5G system; Stage 2 (Release 15),” September 2018.

[7] E. Dahlman, S. Parkvall, J. Sköld, “5G NR; The next generation wireless access technology”, Academic Press, Elsevier, 2018.

[8] ETSI GR NFV-MAN 00, “Network Functions Virtualisation (NFV); Management and Orchestration; Report on Management and Orchestration Framework”, V1.2.1,December, 2021.

[9] Differences between an NMS and an EMS. Available Online: https://www.baeldung.com/cs/networking-nms-vs-ems

[10] Ericsson Mobility Report, November, 2023. Available online: https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/november-2023

[11] https://www.ericsson.com/en/network-services/ai-5g-networks.

[12] G. Wang, H. Gu, X. Li, Z. Yu, O. Li, Q. Liu, K. Zeng, J. He, Y. Chen, J. Lu, W. Tong, and D. Wessel, “Terahertz Sensing and Communication Towards Future Intelligence Connected Networks,” Wireless Technology Lab and Ottawa Wireless Advanced System Competency Centre, September 2022.

[13] https://www.3gpp.org/technologies/5g-system-overview

[14] M. Noura and R. Nordin. “A survey on interference management for device-to-device (D22) communication and its challenges in 5g networks”, Journal of Network and Computer Applications, 71:130-150, 2016.

[15] Sylla T, Mendiboure L, Maaloul S, Aniss H, Chalouf MA, Delbruel S. “Multi-Connectivity for 5G Networks and Beyond: A Survey”. *Sensors*. 2022; 22(19):7591. https://doi.org/10.3390/s22197591.

[16] 3GPP TS 37.340 v16.3.0 “Evolved Universal Terrestrial Radio Access (E-UTRA) and NR; Multi-connectivity; Stage 2 (Release 16)”, September, 2020.

[17] 3GPP TR 37.717-41-11 v0.6.0, “Rel-17 Dual Connectivity (DC) of 4 bands LTE inter-band CA

(4DL/1UL) and 1 NR band (1DL/1UL) (Release 17)", August, 2021.

[18] J. Pérez-Romero and O. Sallent, "Leveraging User Equipment for Radio Access Network Augmentation," in Proc. *IEEE Conference on Standards for Communications and Networking* (CSCN), pp. 83-87, Thessaloniki, Greece, 2021.

[19] 3GPP TS 38.174 v18.4.0, "Rel-18 5G; NR; Integrated Access and Backhaul (IAB) radio transmission and reception", May, 2024.

[20] G. Noh, H. Chung and I. Kim, "Mobile Relay Technology for 5G," in IEEE Wireless Communications, vol. 27, no. 3, pp. 6-7, June 2020.

[21] 3GPP Report TR 22.866, "Enhanced Relays for Energy Efficiency and Extensive Coverage; Stage 1 (Release 17)", V17.1.0, , Dec. 2019.

[22] 3GPP Standard TS 22.261, "Service requirements for 5G system; Stage 1 (Release 19)", V18.5.0, Dec. 2023.

[23] Artificial Intelligence. Available online: https://www.britannica.com/technology/artificial-intelligence

[24] What is AI? Available online: https://www.ibm.com/topics/artificial-intelligence.

[25] Mukhamediev, Ravil I., Yelena Popova, Yan Kuchin, Elena Zaitseva, Almas Kalimoldayev, Adilkhan Symagulov, Vitaly Levashenko, Farida Abdoldina, Viktors Gopejenko, Kirill Yakunin, and et al. 2022. "Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges" *Mathematics* 10, no. 15: 2552.

[26] Machine Learning Methods. Available online: https://www.ibm.com/topics/machine-learning

[27] What is Reinforcement Learning, Available online: https://www.ibm.com/topics/reinforcement-learning

[28] David Silver, "Markov Decision Processes", Advanced Topics, 2015. Available Online: https://www.davidsilver.uk/teaching/

[29] What is deep learning? Available online: https://www.ibm.com/topics/deep-learning

[30] Szczepanski, M. Economic Impacts of Artificial Intelligence (AI). 2019. EPRS: European Parliamentary Research Service. Available online: https://policycommons.net/artifacts/1334867/economic-impacts-of-artificial-intelligence-ai/1940719

[31] S. Ardabili, A. Mosavi and I. Felde, "Deep learning for 5G and 6G," *2023 IEEE 17th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, Timisoara, Romania, 2023, pp. 000711-000720

[32] Clustering in Machine Learning. Available online: https://www.geeksforgeeks.org/clustering-in-machine-learning/

[33] What is Clustering? Available Online: https://developers.google.com/machine-learning/clustering/overview

[34] Y. Wu, Y. Shen and H. Fan, "Location Planning of 5G Base Station Based on Immune Algorithm and Clustering Algorithm," 2023 IEEE Smart World Congress (SWC), pp. 1-7, Portsmouth, United

Kingdom, 2023.

[35] L. Jiang, B. Huang, L. Chen and Z. Li, "Research on location planning of 5G base station based on DBSCAN clustering algorithm," 2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), pp. 669-675, Changchun, China, 2023.

[36] "xRAN forum merges with C-RAN alliance to form ORAN alliance." xRAN Forum. 2018. Available Online: https://www.businesswire.com/news/home/20180227005673/en/

[37] O-RAN Working Group 1, "O-RAN architecture description 5.00," O-RAN, Alfter, Germany, document O-RAN.WG1.O-RANArchitecture- Description-v05.00 Technical Specification, Jul. 2021.

[38] "Governance of O-RAN ALLIANCE e.V. in Compliance with WTO Principles" White Paper, July 2023. Available Online: https://www.o-ran.org/about

[39] "The open future of radio access networks." Deloitte—Telecom Engineering Centre of Excellence (TEE). 2021. Available Online: https://www2.deloitte.com/content/dam/Deloitte/pt/Documents/technology-media-telecommunications/TEE/The-Open-Future-of-Radio-Access-Networks.pdf

[40] M. Polese, L. Bonati, S. D'Oro, S. Basagni and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," in *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376-1411, Secondquarter 2023.

[41] L. Bonati, S. D'Oro, M. Polese, S. Basagni, and T. Melodia, "Intelligence and learning in O-RAN for data-driven NextG cellular networks," *IEEE Commun. Mag.*, vol. 59, no. 10, pp. 21–27, Oct. 2021.

[42] O-RAN specifications. Available online: https://www.o-ran.org/specifications

[43] O-RAN map. Available online: https://map.o-ran.org/

[44] S. Niknam *et al.*, "Intelligent O-RAN for Beyond 5G and 6G Wireless Networks," *2022 IEEE Globecom Workshops (GC Wkshps)*, pp. 215-220. Rio de Janeiro, Brazil, 2022.

[45] R. Kaliski, S. -M. Cheng and C. -F. Hung, "Supporting 6G Mission-Critical Services on O-RAN," in *IEEE Internet of Things Magazine*, vol. 6, no. 3, pp. 32-37, September 2023.

[46] O-RAN architecture. Available online: https://www.o-ran.org/about#technical-workgroup

[47] B. B. Haile, E. Mutafungwa, H. Hämäläinen, "A Data-Driven Multiobjective Optimization Framework for Hyperdense 5G Network Plannin", IEEE Access, September, 2020.

[48] Small Cell Forum, "Hyperdense HetNets: Definition, drivers and barriers", February, 2017.

[49] A. Maeder et al. "A Scalable and Flexible Radio Access Network Architecture for Fifth Generation Mobile Networks", IEEE Communications Magazine, November, 2016.

[50] M. Yan, G. Feng, J. Zhou and S. Qin, "Smart Multi-RAT Access Based on Multiagent Reinforcement Learning," in IEEE Transactions on Vehicular Technology, vol. 67, no. 5, pp. 4539-4551, May 2018.

[51] V. F. Monteiro, et al., "Distributed RRM for 5G Multi-RAT Multiconnectivity Networks", IEEE Systems Journal, Vol. 13, No. 1, March, 2019.

[52] V. Poirot, M. Ericson, M. Nordbert, K. Andersson, "Energy efficient multi-connectivity algorithms for ultra-dense 5G networks", Wireless Networks, 2020.

[53] J. Xie, Y. Liang, Y. Pei, J. Fang and L. Wang, “Intelligent Multi-Radio Access Based on Markov Decision Process,” GLOBECOM 2017 - 2017 IEEE Global Communications Conference, Singapore, 2017, pp. 1-6, doi: 10.1109/GLOCOM.2017.8254596.

[54] M. Gerasimenko et al. “Adaptive Resource Management Strategy in Practical Multi-Radio Heterogeneous Networks”, IEEE Access, February, 2017.

[55] P. K. Taksande, A.Roy, A. Karandikar, “Optimal Traffic Splitting Policy in LTE-based Heterogeneous Network”, *IEEE Wireless Communications and Networking Confernece (WCNC)*,2018.

[56] B. Zhang et al. “Goodput-Aware traffic Splitting Scheme with Non-ideal Backhaul for 5G-LTE Multi-Connectivity”, *IEEE Wireless Communications and Networking Confernece (WCNC)*, 2019.

[57] J. Elias, F. Martignon and S. Paris, “Optimal Split Bearer Control and Resource Allocation for Multi-Connectivity in 5G New Radio,” 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), pp. 187-192, 2021.

[58] V. Mnih, et al. “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[59] Zhang, C.; Patras, P.; Haddadi, H. “Deep Learning in Mobile and Wireless Networking: A Survey”. IEEE Comms. Surv. Tutor. 2019.

[60] Zhang, J.; Su, Q.; Tang, B.;Wang, C.; Li, Y. “DPSNet: Multitask Learning Using Geometry Reasoning for Scene Depth and Semantics”, IEEE Trans. Neural Netw. Learn. Syst. 2021.

[61] J.J. Hernández, J. Pérez-Romero, O. Sallent, I. Vilà, F. Casadevall, “Deep Learning-based Multi-Connectivity Optimization in Cellular Networks,” in *IEEE 95th Vehicular Technology Conference (VTC2022-Spring)*, Helsinki, Finland, June 2022.

[62] Hernández-Carlón, J.J.; Pérez-Romero, J.; Sallent, O.; Vilà, I.; Casadevall, F. “A Deep Q Network-Based Multi-connectivity Algorithm for Heterogeneous 4G/5G Cellular Systems”. In Proceedings of the 18th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI 2022), Crete, Greece, 17–20 June 2022.

[63] O-RAN near-RT RAN Intelligent Controller near-RT RIC Architecture 2.00. O-RAN.WG3.RICARCH-v02.00 Technical Specification. 2021. Available online: https://orandownloadsweb.azurewebsites.net/specifications.

[64] O-RAN Use Cases and Requirements. O-RAN.WG3.UCR-v01.00 Technical Specification. 2021. Available online: https://orandownloadsweb.azurewebsites.net/specifications/

[65] I. Vilà, J. Pérez-Romero, O. Sallent and A. Umbert, “A Multi-Agent Reinforcement Learning Approach for Capacity Sharing in Multi-Tenant Scenarios,” in *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 9450-9465, Sept. 2021, doi: 10.1109/TVT.2021.3099557.

[66] Y. Zhang, C. Kang, Y. Teng, S. Li, W. Zheng and J. Fang, "Deep Reinforcement Learning Framework for Joint Resource Allocation in Heterogeneous Networks," *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019, pp. 1-6, doi: 10.1109/VTCFall.2019.8891448.

[67] U. Challita and D. Sandberg, “Deep Reinforcement Learning for Dynamic Spectrum Sharing of LTE

and NR," *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1-6, doi: 10.1109/ICC42927.2021.9500325.

[68] García, V. "Deep reinforcement Learning Based Approaches for Capacity Sharing in Radio Access Network Slicing", Master's Thesis, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, 2020.

[69] S. Guadarrama S.; Korattikara, A.; Ramirez, O.; Castro, P.; Holly, E.; Fishman, S.; Wang, K.; Gonina, E.; Wu, N.; Kokiopoulou, E.; et al. TF-Agents: A Library for Reinforcement Learning in TensorFlow. 2018. [Online]. Available: https://github.com/tensorflow/agents.

[70] 3GPP TS 38.901 v16.1.0 "Study on Channel Model for Frequencies From 0.5 to 100 GHz (Release 17)", Mar. 2022.

[71] 3GPP. TS 38.304 v17.1.0, User Equipment (UE) Procedures in Idle Mode and RRC Inactive State (Release 17). 2022. Available online: https://www.3gpp.org/ftp/Specs/archive/38_series/38.304.

[72] 3GPP. TS 38.331 v17.1.0, "Radio Resource Control (RRC) Protocol Specification (Release 18)", Dec, 2023.

[73] Dhillon, H.S.; Ganti, R.K.; Baccelli, F.; Andrews, J.G, "Modeling and Analysis of K-Tier Downlink Heterogeneous Cellular Networks". IEEE J. Sel. Areas Commun, p.p 550–560, 2012.

[74] Zhu., Z.; Nishimura, T.; Takimoto, E.; Zheng, J, "Performance Evaluation of Heterogeneous Cellular Networks Using Stochastic Petri Nets". In Proceedings of the International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Hvar, Croatia, 23–25 September 2021.

[75] C. Sudhamani, M. Roslee, JJ. Tiang, AU. Rehman. "A Survey on 5G Coverage Improvement Techniques: Issues and Future Challenges", Sensors. 2023; 23(4):2356.

[76] 3GPP TS 37.320 V16.2.0, "Universal Mobile Telecommunications System (UMTS); LTE; Universal Terrestrial Radio Access (UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRA); Radio measurement collection for Minimization of Drive Tests (MDT); Overall description; Stage 2," Dec. 2019.

[77] 3GPP TS 28.313 V17.4.0, "5G; Management and orchestration; Self-Organizing Networks (SON) for 5G networks (Relase 17)", May, 2022.

[78] G. Joshi and A. Karandikar, "Optimal relay placement for cellular coverage extension," *2011 National Conference on Communications (NCC)*, Bangalore, India, 2011.

[79] A. Bou Saleh, S. Redana, J. Hämäläinen and B. Raaf, "On the Coverage Extension and Capacity Enhancement of Inband Relay Deployments in LTE-Advanced Networks". In: Journal of Electrical and Computer Engineering, Article No. 4, 2010.

[80] 3GPP TS 36.216 v17.0.0, "Physical layer for relaying operation; (Release 17)", March, 2022.

[81] J. Gui and J. Deng, "Multi-Hop Relay-Aided Underlay D2D Communications for Improving Cellular Coverage Quality," in IEEE Access, vol. 6, pp. 14318-14338, 2018.

[82] J. Pérez-Romero, O. Sallent, O. Ruiz, "On Relay User Equipment Activation in Beyond 5G Radio Access Networks" 96th Vehicular Technology Conference (VTC2022 Fall), London (UK) / Beijing

(China), September, 2022.

[83] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, 1996.

[84] J. J. Hernández-Carlón, J. Pérez-Romero, O. Sallent, I. Vilà and F. Casadevall, "On the Detection and Solution of Coverage Holes in 5G Networks through Relay User Equipment: a combined DBSCAN and Deep-Q Network Approach," *97th Vehicular Technology Conference (VTC2023-Spring)*, Florence, Italy, June, 2023.

[85] O. Ruiz, J. Sánchez-González, J. Pérez-Romero, O. Sallent, I. Vilà, "Space and time user distribution measurements dataset in a university campus", Computer Networks, Vol. 243, March, 2024, doi:10.1016/ j.comnet.2024.110329

[86] 3GPP TS 38.214 v18.0.0, "NR; Physical layer procedures for data (Release 18)", Sep, 2023.

[87] https://www.statology.org/z-score-normalization/

[88] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-1, no. 2, pp. 224-227, April, 1979.

[89] scikit-learn documentation: https://scikit-learn.org/stable/modules/clustering.html#dbscan

[90] J. Sydir and R. Taori, "An evolved cellular system architecture incorporating relay stations," *IEEE Communications Magazine*, vol. 47, no. 6, pp. 115-121, Jun. 2009.

[91] O. Teyeb, A. Muhammad, G. Mildh, E. Dahlman, F. Barac and B. Makki, "Integrated Access Backhauled Networks," in *Proc. IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, Honolulu, HI, USA, 2019, pp. 1-5.

[92] M. Polese *et al*., "Integrated Access and Backhaul in 5G mmWave Networks: Potential and Challenges," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 62-68, Mar. 2020.

[93] Study on architecture enhancements for vehicle-mounted relays (Release 18), v1.1.0, 3GPP Report TR 23.700-05, Oct. 2022.

[94] P. Mach, and Z. Becvar, "Device-to-Device Relaying: Optimization, Performance Perspectives, and Open Challenges Towards 6G Networks," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1336-1393, third quarter 2022.

[95] J. Pérez-Romero, and O. Sallent, "On the Value of Context Awareness for Relay Activation in Beyond 5G Radio Access Networks," in *Proc. IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, Helsinki, Finland, 2022, pp. 1-6.

[96] A. Abdelreheem, O. A. Omer, H. Esmaiel, and U. S. Mohamed, "Deep Learning-Based Relay Selection In D2D Millimeter Wave Communications," in *Proc. International Conference on Computer and Information Sciences (ICCIS)*, Sakaka, Saudi Arabia, 2019, pp. 1-5.

[97] H. Kim, T. Fujii, and K. Umebayashi, "Relay Nodes Selection Using Reinforcement Learning," in *Proc. International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Jeju Island, Korea (South), 2021, pp. 329-334.

[98] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, “Deep Reinforcement Learning: A Brief Survey,” *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26-38, Nov. 2017.

[99] J.J. Hernández-Carlón, J. Pérez-Romero, O. Sallent, I. Vilà, and F. Casadevall, “A Deep Q-Network-Based Algorithm for Multi-Connectivity Optimization in Heterogeneous Cellular-Networks,” Sensors, vol. 22, no. 16, Aug. 2022.

[100] “TDD UE-UE Interference Simulations”, Siemens AG, R4-030189 document of the 3GPP TSG-RAN Working Group 4 meeting #26, Feb. 2003.

[101] AI/ML Workflow Description and Requirements 1.03, document O-RAN.WG2.AIML-v01.03, O-RAN Alliance, Oct., 2021.

[102] L. Davis, *Handbook of Genetic Algorithm.* New York, NY, USA: Van Nostrand, 1991.