

UNIVERSIDAD DE MÁLAGA
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA DE TELECOMUNICACIÓN
PROGRAMA DE DOCTORADO EN INGENIERÍA DE TELECOMUNICACIÓN



TESIS DOCTORAL POR COMPENDIO

QUALITY-DRIVEN MANAGEMENT OF
NEXT-GENERATION MOBILE NETWORKS FOR
ADVANCED MULTIMEDIA SERVICES

Autor:

OSWALDO SEBASTIÁN PEÑAHERRERA PULLA

Directores:

SERGIO FORTES RODRÍGUEZ

RAQUEL BARCO MORENO

2025



UNIVERSIDAD
DE MÁLAGA

AUTOR: Oswaldo Sebastián Peñaherrera Pulla

 <https://orcid.org/0000-0001-6332-5644>

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga



Esta obra está bajo una licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional:

<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

Cualquier parte de esta obra se puede reproducir sin autorización pero con el reconocimiento y atribución de los autores.

No se puede hacer uso comercial de la obra y no se puede alterar, transformar o hacer obras derivadas.

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de Málaga (RIUMA): riuma.uma.es



DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D. Oswaldo Sebastián Peñaherrera Pulla

Estudiante del programa de doctorado en **Ingeniería de Telecomunicación** de la Universidad de Málaga, autor/a de la tesis presentada para la obtención del título de doctor por la Universidad de Málaga, titulada: **Quality-driven management of next-generation mobile networks for advanced multimedia services**.

Realizada bajo la tutorización de **Raquel Barco Moreno** y dirección de **Sergio Fortes Rodríguez y Raquel Barco Moreno**.

DECLARO QUE:

La tesis presentada es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, conforme al ordenamiento jurídico vigente (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), modificado por la Ley 2/2019, de 1 de marzo.

Igualmente asumo, ante a la Universidad de Málaga y ante cualquier otra instancia, la responsabilidad que pudiera derivarse en caso de plagio de contenidos en la tesis presentada, conforme al ordenamiento jurídico vigente.

En Málaga, a **21 de Febrero de 2025**.

Fdo.: Oswaldo Sebastián Peñaherrera Pulla Doctorado	Fdo.: Raquel Barco Moreno Tutor de tesis
Fdo.: Sergio Fortes Rodríguez y Raquel Barco Moreno Directores de tesis	



UNIVERSIDAD
DE MÁLAGA

AUTORIZACIÓN PARA LA LECTURA DE LA TESIS

Por la presente, la Dra. Raquel Barco Moreno, y el Dr. Sergio Fortes Rodríguez, profesores del Departamento de Ingeniería de Comunicaciones de la Universidad de Málaga,

CERTIFICAN

Que D. Oswaldo Sebastián Peñaherrera Pulla, ha realizado en el Departamento de Ingeniería de Comunicaciones de la Universidad de Málaga bajo su dirección, el trabajo de investigación correspondiente a su TESIS DOCTORAL titulada:

“Quality-driven management of next-generation mobile networks for advanced multimedia services”

En dicho trabajo, se han propuesto aportaciones originales para el desarrollo de mecanismos AI/ML para la gestión de las redes móviles e inalámbricos en el contexto de provisión de servicios multimedia avanzados. Los resultados de dicha tesis han dado lugar a las diversas publicaciones en revista, así como a aportaciones a congresos, superando el requisito de 1 punto ANECA del programa de doctorado regulado por el Real Decreto 99/2011.

Por todo ello, y dada la unidad temática de las distintas contribuciones y la metodología común seguida en todas ellas, el director considera que esta tesis es apta para su presentación al Tribunal que ha de evaluarla y AUTORIZA la presentación de la tesis por COMPENDIO DE PUBLICACIONES en la Universidad de Málaga. Igualmente, certifica que las publicaciones que avalan la tesis no han sido empleadas en trabajos anteriores a la misma.

Málaga, 21 de Febrero de 2025

Fdo.: Raquel Barco Moreno

Fdo.: Sergio Fortes Rodríguez



UNIVERSIDAD
DE MÁLAGA

UNIVERSIDAD DE MÁLAGA
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA DE TELECOMUNICACIÓN
PROGRAMA DE DOCTORADO EN INGENIERÍA DE TELECOMUNICACIÓN

Reunido el tribunal examinador en el día de la fecha, constituido por:

Presidente: Dr. D. _____

Secretario: Dr. D. _____

Vocal: Dr. D. _____

para juzgar la Tesis Doctoral titulada *Quality-driven management of next-generation mobile networks for advanced multimedia services* realizada por D. Oswaldo Sebastián Peñaherrera Pulla, y dirigida por los doctores D. Sergio Fortes Rodríguez y D^a Raquel Barco Moreno, acordó por

_____ otorgar la calificación de

_____ y para que conste,

Málaga, a ____ de _____ de _____.

El Presidente:

Fdo.: _____

El Secretario:

El Vocal:

Fdo.: _____ Fdo.: _____



UNIVERSIDAD
DE MÁLAGA

To my parents, Oswaldo and Marcia, and my incredible sister, Doménica. The only ones who will never disappoint me, the only ones who will make every effort worthwhile.

Mami Chave, Miguel, and Leonor, this is also for you.



UNIVERSIDAD
DE MÁLAGA

“The size of your success is measured by the strength of your desire; the size of your dream; and how you handle disappointment along the way.”

Robert Kiyosaki



UNIVERSIDAD
DE MÁLAGA

Acknowledgments

This great adventure has come to an end, but it left me with invaluable lessons that I will carry with me forever. Nobody said it would be easy, but I'm a different person because of it.

Thank you for being part of my trip. I am grateful to you all for your involvement in this project. Without your input, my dream would not have come true. I know words are never enough, but I will try to remember everyone.

I would like to express my gratitude to my directors, Sergio and Raquel. I want to thank Sergio for recognizing my talent from the very first moment I was his student. This undoubtedly kickstarted the perfect opportunity to make my dream a reality. I am grateful to you, Raquel, for giving me the chance to work on your MobileNet project. Only a select few were fortunate enough to work with you and benefit from your experience. I am certain that without you both, I would have missed the opportunity to work with and meet the spectacular and special people who will remain in my heart forever.

Secondly, I would like to express my gratitude to all my colleagues in the laboratory. You all played a role in helping me achieve this. *Notwithstanding*, I want to highlight my sincere “thank you” to Carlos, my greatest team partner, for his support throughout my research journey, from the early days at the TFM to the challenging times during the pandemic, the tough moments when things didn't go as planned as well as when I only needed a real friend.

I also want to thank the amazing guys I met at that lab, especially Tarri, Truji, and David (Dr. Dasera) for being part of this funny, incredible, sometimes sad story in Spain. I will always remember the amazing times we had in Denmark. I can say with confidence that my time in Málaga/Aalborg/Norway/Germany/Austria/Czechia, everywhere, would have been completely uninteresting without your jokes, smiles, and presence, as well as your unwavering support.

Moreover, I want to say thank you, Hao, for being a reliable, good listener and advisor. I will always remember those lunches where we discussed everything from the latest gossip to the most profound topics concerning society, ethics, and how to make the world a better place for everyone. I must also thank Edu and Javi for their incredible knowledge of the unconventional things that rule the world. Your ethical and professional profiles are crucial to saving our future. You inspire us to be better engineers. I would like to give a special shout-out to my (ex)teammates Pablo, Fran, María, Laura, Juanjo, Manu, Nacho, Pepe, Carlos A., and all the newbies. Without you, my time in the laboratory would have been much more stressful and less enjoyable.

I would also like to thank everyone in Aalborg for creating a fantastic atmosphere that inspired me to learn new things, see professional life from a different perspective, and, most importantly, meet more valuable people in my life. I would like to give a special thank you to Preben, Boyan, Sebastian, Melisa, Carlos, Enric, Jesús, and Kun for sharing your wide experience and for the opportunity to work with such engineers. You are amazing people!

Likewise, I want to say a big thank you to Kevin, Joaquin, Ángel, Ale, Illojuan, and Baity, my roommates, fantastic people, and friends. Your presence made my life easier in Málaga, even though the time was not as long as we would have liked.

Finally, I would like to dedicate a special section to my family. I am really grateful to you, Oswaldo, for staying with me throughout my high school journey. You instilled in me a passion for engineering. Your resilience and desire to make me a great engineer is now bearing fruit. I thank you, Marcia, you taught me how to be a strong man, how to pursue an objective no matter how difficult it would be, and equally important, how to be a good person with ethics within a world full of mistaken people. I would also like to thank Domi for showing your strength throughout the years I was away. Mom and dad would not have made it through without your kindness and unconditional support. I am learning more and more from you, and I am realizing more and more that you are truly great. You must never change. The world needs more people like you.

I would also like to thank my grandparents Miguel and Blanca and Mami

Chave for inspiring my parents. Your knowledge and beliefs have created incredible life teachers. Likewise, a thank you is not enough to express my gratitude to my cousins, aunts, and uncles for their amazing support and for always believing in me.

I am aware that these words may not fully express what I want, but they at least try to materialize my thoughts that have helped me overcome all the obstacles to achieve this. I am forever in your debt.

I hope I have not forgotten to mention anyone. At this point, I am a little worried that this thesis is about gratitude, but I want you to know that if you are reading this, you are part of this accomplishment. So thank you!



UNIVERSIDAD
DE MÁLAGA

This thesis has been partially funded by “Junta de Andalucía. Secretaría General de Universidades, Investigación y Tecnología” under predoctoral Grant PREDOC_01712 and by the following projects:

- Piloto 5G Red.es Andalucía, Case Study OpenRAN: Development of a 5G access network prototype based on GPPs. Project number 8.06/5.59.5630, funded by Ministerio de Asuntos Económicos y Transformación Digital and the European Regional Development Fund (ERDF).
- PENTA: Provisión de servicios PPDR a través de Nuevas Tecnologías de Acceso radio. Project reference P18-FR-4647 funded in the context of “Proyectos de investigación en colaboración con el tejido productivo” and European Regional Development Fund (ERDF).
- AECMA-5G: Advanced E2E Cellular Management for 5G Applications. Project reference: UMA-CEIATECH-14. Funding granted for Junta de Andalucía and the European Regional Development Fund (ERDF).
- MAORI: Massive AI for the OpenRadIo b5G/6G network. Project number TSI-063000-2021-75, receiving funds from Ministerio de Asuntos Económicos y Transformación Digital and European Union - NextGenerationEU within the framework “Recuperación, Transformación, y Resiliencia”.



UNIVERSIDAD
DE MÁLAGA

Summary of contributions

The various publications resulting from this research are listed below, starting with those that support the thesis.

Journal articles

- [1] **O. S. Peñaherrera-Pulla**, C. Baena, S. Fortes, E. Baena and R. Barco, “Measuring key quality indicators in cloud gaming: Framework and assessment over wireless networks”, in *Sensors*, vol. 21, no. 4, 2021, doi: 10.3390/s21041387.
- [2] **O. S. Peñaherrera-Pulla**, C. Baena. S. Fortes, E. Baena and R. Barco, “KQI Assessment of VR Services: A Case Study on 360-Video Over 4G and 5G”, in *IEEE Transactions on Network and Service Management*, vol. 19, num. 4, pp. 5366-5382, Dec. 2022, doi: 10.1109/TNSM.2022.3192762.
- [3] **O. S. Peñaherrera Pulla**, S. B. Damsgaard, B. Yanakiev, P. Mogensen, S. Fortes and R. Barco, “Cloud VR on 5G: A Performance Validation in Industrial Scenarios”, in *IEEE Open Journal of the Communications Society*, vol. 5, pp. 3641-3657, 2024, doi: 10.1109/OJCOMS.2024.3413328. .
- [4] **O. S. Peñaherrera-Pulla**, C. Baena, S. Fortes and R. Barco, “ML-Powered KQI Estimation for XR Services: A Case Study on 360-Video”, in *IEEE Open Journal of the Communications Society*, vol. 5, pp. 4075-4097, 2024, doi: 10.1109/OJCOMS.2024.3422872.
- [5] **O. S. Peñaherrera-Pulla**, C. Baena, H. Q. Luo-Chen, S. Fortes and R. Barco, “Towards KQI/ML-based management: Optimizing radio resource allocation for next-generation networks”, in *TechRxiv*, 2024, doi: 10.36227/techrxiv.173091688.85406078/v1

Additional contributions:

- [6] C. Baena, **O. S. Peñaherrera-Pulla**, L. Camacho, R. Barco and S. Fortes, “Video Streaming and Cloud Gaming Services Over 4G and 5G: A Complete Network and Service Metrics Dataset”, in *IEEE Communications Magazine*, vol. 61, no. 9, pp. 154-160, 2019, doi: 10.1109/MCOM.005.2200719.
- [7] C. Baena, **O. S. Peñaherrera-Pulla**, R. Barco, and S. Fortes, “Measuring and estimating Key Quality Indicators in Cloud Gaming services”, in *Computer Networks*, vol. 231, 2023, doi: 10.1016/j.comnet.2023.109808.
- [8] C. Baena, S. Fortes, **O. S. Peñaherrera-Pulla**, E. Baena and R. Barco, “Gaming in the Cloud: 5G as the Pillar for Future Gaming Approaches”, 2024, doi: 10.1109/MCOM.005.2300256, Early Access in *IEEE Communications Magazine*.

International conferences/workshops

- [9] **O. S. Peñaherrera-Pulla**, C. Baena, S. Fortes, E. Baena and R. Barco, “Experimental Evaluation of Cloud Gaming for Different Communication Technologies”, in *EURACON 2nd Post-IRACON Meeting*, Virtual, 2021.
- [10] **O. S. Peñaherrera-Pulla**, C. Baena, S. Fortes, E. Baena and R. Barco, “KQI Evaluation for 360-video Services Over Mobile Networks”, in *COST ACTION CA20120 - 2nd MC Meeting and 1st Technical Meeting*, Bologna, Italy, 2022.
- [11] **O. S. Peñaherrera-Pulla**, C. Baena, R. Barco and S. Fortes, “ML-Based Network Management Framework for XR Services”, in *2023 European Conference on Networks and Communications (EuCNC) & 6G Summit*, Gothenburg, Sweden, 2023.

Additional contributions

- [12] C. Baena, S. Fortes, **O. S. Peñaherrera-Pulla** and R. Barco, “A Framework to boost the potential of network-in-a-box solutions”, in *12th International Conference on Network of the Future, NoF 2021*, Coimbra, Portugal, 2021, doi: 10.1109/NoF52522.2021.9609861.

National conferences/workshops

- [13] **O. S. Peñaherrera-Pulla**, C. Baena, S. Fortes, E. Baena and R. Barco, “Adquisición de métricas en Cloud Gaming”, in *XXXV Simposium nacional de la Unión Científica Internacional de Radio*, Málaga, 2020.
- [14] **O. S. Peñaherrera-Pulla**, C. Baena, S. Fortes, E. Baena and R. Barco, “Adquisición de métricas para el servicio de video 360/VR”, in *XXXVI Simposium nacional de la Unión Científica Internacional de Radio*, Vigo, 2021.
- [15] **O. S. Peñaherrera-Pulla**, C. Baena, S. Fortes, E. Baena and R. Barco, “Estimación de KQIs para servicio de Video-360”, in *XXXVII Simposium nacional de la Unión Científica Internacional de Radio*, Málaga, 2022.

Additional contributions:

- [16] C. Baena, **O. S. Peñaherrera-Pulla**, S. Fortes, E. Baena and R. Barco, “Evaluación del servicio Cloud Gaming para diferentes tecnologías de acceso”, in *XXXVI Simposium nacional de la Unión Científica Internacional de Radio*, Vigo, 2021.
- [17] J. Pulido, S. Fortes, C. Baena, **O. S. Peñaherrera-Pulla**, A. Fernández-Durán and R. Barco, “Impacto de tráfico XR (eXtended Reality) en entornos Wi-Fi”, in *XXXVII Simposium nacional de la Unión Científica Internacional de Radio*, Cáceres, 2023.
- [18] M. B. Martínez-Rodríguez, J. Pulido, S. Fortes, C. Baena, **O. S. Peñaherrera-Pulla**, A. Fernández-Durán and R. Barco, “Estimación de KQIs de video streaming en entornos Wi-Fi”, in *XXXIX Simposium nacional de la Unión Científica Internacional de Radio*, Cuenca, 2024.

Datasets and Tools

- [19] **O. S. Peñaherrera-Pulla**, C. Baena, S. Fortes and R. Barco, “E2E dataset of 360-degree video service using mobile networks”, in *IEEE Dataport*, 2024. Online. Available in <https://dx.doi.org/10.21227/6gg3-n924>
- [20] C. Baena, **O. S. Peñaherrera-Pulla**, L. Camacho, R. Barco and S. Fortes, “E2E cross-layer dataset of Video Streaming and Cloud Gaming services over 4G and 5G”, in *IEEE Dataport*, 2022. Online. Available in <https://dx.doi.org/10.21227/k0w8-qz67>
- [21] **O. S. Peñaherrera-Pulla**, “KQI prediction ML framework Repository”, in *GitHub Repository*, 2024. Online. Available in https://github.com/sebaspenaherrera/ML_framework_prediction
- [22] C. Baena, **O. S. Peñaherrera-Pulla**, “Service KQI Extraction Repository”, in *GitHub Repository*, 2023. Online. Available in <https://github.com/jcbaglez/ServiceKQIExtraction>



UNIVERSIDAD
DE MÁLAGA

Contents

Summary of contributions	XIX
Abstract	XXVIII
Resumen	XXXI
Acronyms	XXXV
List of Figures	XLV
List of Tables	XLV
I Background	1
1 Introduction	3
1.1 Motivation	4
1.2 Preliminaries	7
1.3 Challenges and objectives	8
1.4 Document structure	12
2 Technical overview	17
2.1 Mobile networks	19
2.1.1 Mobile networks background and architecture	19
2.1.2 LTE	24

2.1.3	5G	30
2.1.4	The Open RAN concept	39
2.2	5G New Radio (NR)	42
2.2.1	NR frequency bands	42
2.2.2	NR Waveforms	44
2.2.3	NR frame structure and numerologies	45
2.3	Machine Learning	49
2.4	Network management	51
2.4.1	Service-oriented management	53
2.5	Extended Reality background	56
2.6	Conclusions of the chapter	58
II	Publications	59
3	Research outline	61
3.1	Research Methodology	62
3.2	Description of the publications	64
3.2.1	C1. Measuring key quality indicators in cloud gaming: Framework and assessment over wireless networks	65
3.2.2	C2. KQI assessment of VR services: A case study on 360-video over 4G and 5G	66
3.2.3	C3. Cloud VR on 5G: A Performance Validation in Industrial Scenarios	67
3.2.4	C4. ML-powered KQI estimation for XR services: A case study on 360-video	69
3.2.5	C5. Towards KQI/ML-based management: Optimizing radio resource allocation for next-generation networks	71
3.3	Conclusions of the chapter	72
4	Experimental new-generation service environments	73

4.1	Measuring key quality indicators in cloud gaming: Framework and assessment over wireless networks	74
4.2	KQI assessment of VR services: A case study on 360-video over 4G and 5G	75
4.3	Cloud VR on 5G: A Performance Validation in Industrial Scenarios .	76
4.4	Conclusions of the chapter	77
5	ML-powered new-generation service monitoring	79
5.1	ML-Powered KQI Estimation for XR Services: A Case Study on 360-Video	80
5.2	Conclusions of the chapter	81
6	KQI-driven E2E network management	83
6.1	Towards KQI/ML-based management: Optimizing radio resource allocation for next-generation networks	84
6.2	Conclusions of the chapter	85
III	Conclusions	87
7	Conclusions	89
7.1	Contributions	90
7.2	Future work	95
7.3	Moving towards mobile industry	96
8	Research activities	99
8.1	Dissemination	100
8.2	Related projects	100
8.3	Research stays	101

IV Appendices 103

Appendix A Summary (Spanish) 105

A.1 Motivación	106
A.2 Preámbulos	109
A.3 Objetivos	111
A.4 Descripción de las publicaciones	115
A.4.1 C1. Measuring key quality indicators in cloud gaming: Framework and assessment over wireless networks	116
A.4.2 C2. KQI assessment of VR services: A case study on 360-video over 4G and 5G	117
A.4.3 C3. Cloud VR on 5G: A Performance Validation in Industrial Scenarios	118
A.4.4 C4. ML-powered KQI estimation for XR services: A case study on 360-video	120
A.4.5 C5. Towards KQI/ML-based management: Optimizing radio resource allocation for next-generation networks	122
A.5 Conclusiones	123
A.6 Actividades de investigación	128
A.6.1 Difusión de los resultados	129
A.6.2 Proyectos relacionados	129
A.6.3 Estancias de investigación	130

Bibliography 131

Abstract

The emergence of advanced next-generation use cases has fundamentally redefined the concepts of services and applications, promising to streamline the management of day-to-day activities through a diverse range of solutions. Notably, metaverse applications, which facilitate immersive interactions across sectors such as industry, education, defense, and entertainment, exemplify these developments. These innovations are reshaping the design and development of next-generation mobile networks. Delivering such services necessitates networks capable of meeting heterogeneous requirements, including higher data rates, reduced latency, enhanced reliability, and substantial computational resources to ensure superior user experiences. This situation can aggregate stress on the infrastructure, hence network management is crucial to address it. Traditionally, this latter has relied on human expertise and decision-making, supplemented by Operations Support Systems (OSS). Efforts to automate these processes led to the introduction of Self-Organizing Networks (SON) in legacy systems to improve network resilience and efficiency.

The advent of 5G has enabled many of these advanced use cases, yet managing such a broad array of services with diverse requirements and data-intensive operations presents substantial challenges. The introduction of architectural flexibility, decentralization, and multi-level intelligence has created a complex environment where Artificial Intelligence (AI) and Machine Learning (ML) offer significant potential. By analyzing network-accessible data, ML can extract valuable insights, recognize patterns, and detect anomalies that might elude human operators. This paradigm shifts from network-centric to user-centric management, enhancing service performance to align with Service-Level Agreement (SLA) commitments. This shift necessitates progress beyond traditional counters and Key Performance Indicators (KPIs) that assess service provision from a network-focused perspective. Instead, Key Quality Indicators (KQIs) have emerged as

critical metrics for evaluating performance across the communication chain.

This thesis leverages the next-generation mobile network paradigm, combined with ML capabilities, to develop mechanisms for managing service quality. It first examines advanced multimedia services, identified as critical applications for 5G and beyond, with a focus on Extended Reality (XR). A comprehensive literature review identifies key KQIs that reflect service performance and addresses the unique requirements and challenges faced by CSPs in delivering high-quality end-to-end (E2E) services under SLA constraints. Service-specific metrics are then developed for Cloud Gaming (CG), 360-degree video, and Cloud Virtual Reality (Cloud VR). Then, empirical research is conducted by integrating these services into testbeds under controlled scenarios, enabling analysis of network impacts by comparing service performance across various wireless network environments (e.g., Wi-Fi, LTE, 5G), different radio conditions, and mobility scenarios.

This work consequently explores the potential of ML in network management, focusing particularly on the 360-degree video service. Using the generated dataset, an ML-based framework is developed to train, optimize, and evaluate models that predict KQIs based on radio conditions and network statistics. The selection of models is guided by the Performance Error and prediction Time score (PET_{score}), a proposed metric that balances error and processing time, ensuring suitability for time-sensitive applications. To illustrate ML's application in E2E network management, a resource optimization approach is proposed. This solution integrates service performance modeling with numerical optimization, where ML-based channel-specific resource models are developed based on KQIs and channel conditions. By optimizing a target function that considers network resources and Quality of Experience (QoE), this approach identifies the most effective resource allocation for SLA compliance.

Concluding, this thesis demonstrates the viability of AI/ML strategies as powerful tools for intelligent management in next-generation networks, highlighting promising avenues for future research and potential contributions to the evolution of future network standards.

Resumen

El surgimiento de casos de uso avanzados de próxima generación ha redefinido fundamentalmente los conceptos de servicios y aplicaciones, con la promesa de simplificar la gestión de actividades diarias mediante una amplia gama de soluciones. En particular, las aplicaciones del metaverso, que facilitarán interacciones inmersivas en sectores como la industria, la educación, la defensa y el entretenimiento, ejemplifican estos desarrollos. Estas innovaciones están transformando el diseño y desarrollo de redes móviles de próxima generación. La provisión de estos servicios exige redes capaces de satisfacer requisitos heterogéneos, que incluyen tasas de datos más altas, menor latencia, mayor fiabilidad y recursos computacionales sustanciales para garantizar experiencias superiores a los usuarios. Esta situación puede incrementar la presión sobre la infraestructura, por lo que la gestión de redes se vuelve crucial para abordarla. Tradicionalmente, esta gestión ha dependido de la experiencia humana y la toma de decisiones, complementadas por Sistemas de Soporte a la Operación (OSS). Los esfuerzos para automatizar estos procesos condujeron a la introducción de Redes Auto-Organizadas (SON) en sistemas heredados para mejorar la resiliencia y eficiencia de la red.

La llegada del 5G ha hecho posible muchos de estos casos de uso avanzados; sin embargo, gestionar una gama tan amplia de servicios con diversos requisitos y operaciones intensivas en datos presenta desafíos considerables. La introducción de flexibilidad arquitectónica, descentralización e inteligencia multinivel ha creado un entorno complejo en el que la Inteligencia Artificial (IA) y el Aprendizaje Automático (ML) ofrecen un potencial significativo. Mediante el análisis de datos accesibles de la red, el ML puede extraer información valiosa, reconocer patrones y detectar anomalías que podrían pasar inadvertidas para los operadores humanos. Este cambio de paradigma se traslada de una gestión centrada en la red a una gestión centrada en el usuario, mejorando el rendimiento del servicio para alinearlos con los

compromisos de Acuerdo de Nivel de Servicio (SLA). Este cambio exige un progreso más allá de los contadores tradicionales e Indicadores Clave de Rendimiento (KPIs) que evalúan la provisión de servicios desde una perspectiva centrada en la red. En su lugar, han surgido los Indicadores Clave de Calidad (KQIs) como métricas críticas para evaluar el rendimiento a lo largo de la cadena de comunicación.

Esta tesis aprovecha el paradigma de redes móviles de próxima generación, en combinación con las capacidades de ML, para desarrollar mecanismos de gestión de la calidad del servicio. Primero, examina servicios multimedia avanzados, identificados como aplicaciones críticas para el 5G y tecnologías posteriores, con un enfoque en la Realidad Extendida (XR). Una revisión exhaustiva de la literatura identifica los principales KQIs que reflejan el rendimiento del servicio y aborda los requisitos y desafíos únicos que enfrentan los Proveedores de Servicios de Comunicaciones (CSPs) al ofrecer servicios E2E de alta calidad bajo las limitaciones de SLA. Posteriormente, se desarrollan métricas específicas del servicio para Juegos en la Nube (CG), video de 360 grados y Realidad Virtual en la Nube (Cloud VR). Luego, se realiza una investigación empírica integrando estos servicios en plataformas de prueba en escenarios controlados, lo que permite analizar el impacto de la red mediante la comparación del rendimiento del servicio en diversos entornos de redes inalámbricas (e.g., Wi-Fi, LTE, 5G), distintas condiciones de radio y escenarios de movilidad.

En consecuencia, la tesis explora el potencial del ML en la gestión de redes, enfocándose particularmente en el servicio de video de 360 grados. Utilizando el conjunto de datos generado, se desarrolla un marco basado en ML para entrenar, optimizar y evaluar modelos que predicen KQIs en función de las condiciones de radio y estadísticas de la red. La selección de modelos se guía mediante el puntaje PET, una métrica propuesta que equilibra el error y el tiempo de procesamiento, asegurando la idoneidad para aplicaciones sensibles al tiempo. Para ilustrar la aplicación del ML en la gestión de redes E2E, se propone un enfoque de optimización de recursos. Esta solución integra el modelado del rendimiento del servicio con la optimización numérica, donde se desarrollan modelos de recursos específicos del canal basados en ML en

función de los KQIs y las condiciones del canal. Al optimizar una función objetivo que considera los recursos de la red y la Calidad de Experiencia (QoE), este enfoque identifica la asignación de recursos más efectiva para el cumplimiento del SLA.

En conclusión, esta tesis demuestra la viabilidad de las estrategias de IA/ML como herramientas poderosas para la gestión inteligente en redes de próxima generación, destacando vías prometedoras para futuras investigaciones y posibles contribuciones a la evolución de los estándares de redes futuras.



UNIVERSIDAD
DE MÁLAGA

Acronyms

1G First-Generation Mobile Networks.

2G Second-Generation Mobile Networks.

3G Third-Generation Mobile Networks.

3GPP 3rd Generation Partnership Project.

4G Fourth-Generation Mobile Networks.

5G Fifth-Generation Mobile Networks.

5GC 5G Core.

5GS Fifth Generation System.

6G Sixth-Generation Mobile Networks.

AAU Aalborg University.

AF Application Function.

AI Artificial Intelligence.

AI/ML Artificial Intelligence / Machine Learning.

AMC Adaptive Modulation and Codification.

AMF Access and Mobility Management Function.

API Application Programming Interface.

AR Augmented Reality.

ARQ Automatic Repeat reQuest.

AS Access Stratum.

ASP Application Service Provider.

AUSF Authentication Server Function.

AVoIP Audio-Video over IP.

B5G Beyond 5G.

BBU Base Band Unit.

BS Base Station.

BWP BandWidth Part.

CA Carrier Aggregation.

CAGR Compound Annual Growth Rate.

CAPEX Capital Expenditure.

CC Component Carrier.

CG Cloud Gaming.

CN Core Network.

CNF Cloud-native Network Function.

CoMP Coordinated Multi-Point transmission/reception.

COTS Commercial-off-the-shelf.

CP Cyclic Prefix.

CP Control Plane.

CQI Channel-Quality Indicator.

CRB Common Resource Block.

CSP Communication Service Provider.

CU Central Unit.

CUPS Control and User Plane Separation.

dApp Distributed Application.

DASH Dynamic Adaptive Streaming over HTTP.

DL Downlink.

DRB Data Radio Bearer.

DU Distributed Unit.

E-UTRAN Evolved Universal Terrestrial Radio Access Network.

E2E End-to-End.

EB Exabytes.

EDGE Enhanced Data rates for GSM Evolution.

eMBB enhanced Mobile BroadBand.

EN-DC Evolved Non-standalone Dual Connectivity.

eNB Evolved Node B.

EPC Evolved Packet Core.

EPS Evolved Packet System.

ETSI European Telecommunications Standards Institute.

FE Feature Extraction.

FoM Figure of Merit.

FOV Field of View.

FS Feature Selection.

FWA Fixed Wireless Access.

GB Gigabytes.

gNB Next-Generation Node B.

GPRS General Packet Radio Service.

GSM Global System for Mobile Communications.

HLS High-Layer Split.

HMD Head-Mounted Device.

HO Handover.

HSPA+ Evolved High Speed Packet.

HSS Home Subscriber Server.

IaaS Infrastructure-as-a-Service.

IIoT Industrial Internet of Things.

IMS IP Multimedia Subsystem.

IoT Internet of Things.

IP Internet Protocol.

ISI Inter-Symbol Interference.

ITU International Telecommunication Union.

JSON JavaScript Object Notation.

KNN K-Near Neighbors.

KPI Key Performance Indicator.

KQI Key Quality Indicator.

LLA Licensed-Assisted Access.

LLS Low-Layer Split.

LTE Long Term Evolution.

LTE-A LTE Advanced.

LTE-A Pro LTE Advanced Pro.

MAC Medium Access Control.

MIMO Multiple Input Multiple Output.

ML Machine Learning.

MLP Multi-Layer Perceptron.

MME Mobility Management Entity.

mMTC massive Machine-Type Communications.

mmWave millimetre Wave.

MNO Mobile Network Operator.

MobileNet Mobile and Aerospace Networks Lab research group.

MOS Mean Opinion Score.

MP Management Plane.

MR Mixed Reality.

NaaS Network-as-a-Service.

NAS Non Access Stratum.

NAT Network Address Translation.

NB NodeB.

NDT Network Digital Twin.

NE Network Element.

Near-RT Near-Real-Time.

NEF Network Exposure Function.

NF Network Function.

NFV Network Function Virtualization.

NG-RAN Next Generation RAN.

NG-SON Next-Generation SON.

NGMN Next Generation Mobile Networks Alliance.

Non-RT Non-Real-Time.

NR New Radio.

NRF NF Repository Function.

ns Network slicing.

NSA Non-Stand-Alone.

NSSAAF Network Slice-Specific Authentication and Authorization Function.

NSSF Network Slice Selection Function.

NWDAF Network Data Analytics Function.

O-CU Open Central Unit.

O-DU Open Distributed Unit.

O-RU Open Radio Unit.

OFDM Orthogonal Frequency Division Multiplexing.

OFDMA Orthogonal Frequency Division Multiple Access.

OPEX Operational Expenditure.

OSS Operations Support System.

P-GW Packet data network GateWay.

PAPR Peak-to-Average Power Ratio.

PCF Policy Control Function.

PCRF Policy and Charging Rules Function.

PDCP Packet Data Convergence Protocol.

PDN Packet Data Network.

PDU Protocol Data Unit.

PHY PHYsical layer.

QCI QoS Class Identifier.

QFI QoS Flow Identifier.

QoE Quality of Experience.

QoS Quality of Service.

RAN Radio Access Network.

rApp RAN Application.

RAT Radio Access Technology.

RB Resource Block.

RCS Rich Communication Service.

RE Resource Element.

REST REpresentational State Transfer.

RF Random Forest.

RF Radio Frequency.

RIC Radio Intelligent Controller.

RL Reinforcement Learning.

RLC Radio Link Control.

RNC Radio Network Controller.

RR Ridge Regression.

RRC Radio Resource Control.

RRE Remote Radio Equipment.

RRU Remote Radio Unit.

RU Radio Unit.

S-GW Serving GateWay.

SA Stand-Alone.

SAE System Architecture Evolution.

SBA Service-based Architecture.

SC-FDMA Single-Carrier Frequency Division Multiple Access.

SDAP Service Data Adaptation Protocol.

SDL Supplemental Downlink.

SDN Software-Defined Network.

SDR Software-Defined Radio.

SDU Service Data Unit.

SIP Session Initiation Protocol.

SLA Service-level Agreement.

SLIV Start and Length Indicator Value.

SMF Session Management Function.

SMO Service Management Orchestration.

SMS Short Messaging Service.

SON Self-Organizing Networks.

SoTA State-of-the-art.

SU-MIMO Single-user Multiple Input Multiple Output.

SUL Supplemental Uplink.

SVM Support Vector Machine.

TDMA Time Division Multiple Access.

UDM Unified Data Management.

UE User Equipment.

UL Uplink.

UMTS Universal Mobile Telecommunications System.

UP User Plane.

UPF User Plane Function.

URLLC Ultra Reliable Low Latency Communications.

VNF Virtual Network Function.

VOD Video-On-Demand.

VoLTE Voice over LTE.

VR Virtual Reality.

vRAN virtualized Radio Access Network.

WCDMA Wideband Code Division Multiple Access.

xApp eXtended Application.

XR Extended Reality.

ZSM Zero-touch network and Service Management.



UNIVERSIDAD
DE MÁLAGA

List of Figures

1.1	Document structure	13
2.1	Mobile network general architecture	21
2.2	3GPP Protocol stack	23
2.3	Features introduced by LTE	25
2.4	Long Term Evolution (LTE) architecture	28
2.5	5G service categories and key capabilities [23]	31
2.6	5G system architecture	33
2.7	5G protocol stack	37
2.8	Functional split comparison	38
2.9	O-RAN architectural vision [24]	39
2.10	Frequency bands in NR	44
2.11	Orthogonality in Orthogonal Frequency Division Multiplexing (OFDM) subcarriers	45
2.12	Resource allocation strategy	46
2.13	New Radio (NR) resource grid for numerology 1 (30 kHz)	48
2.14	Management scope of Quality of Experience (QoE) and Quality of Service (QoS)	54
3.1	Research methodology	62
3.2	Mapping between thesis challenges, objectives and results achieved . .	65

List of Tables

2.1	5G vs 4G capabilities	32
2.2	5G NR time scheme per numerology [25]	47
2.3	Typical network sources of information	54
2.4	Mean Opinion Score evaluation format	55

Part I

Background



UNIVERSIDAD
DE MÁLAGA

Chapter 1

Introduction

Content

1.1	Motivation	4
1.2	Preliminaries	7
1.3	Challenges and objectives	8
1.4	Document structure	12

This initial chapter aims to convey the motivation and rationale behind this thesis, present the stated objectives, and provide an overview of the structure and organization of this document.

1.1 Motivation

The consistent advancement of communication technologies over time has led to the emergence of numerous applications and services designed to enhance daily activities. In essence, the primary objective was to facilitate remote communication. However, this merely paved the way for the design of networks capable of conveying common-purpose data from one point to another, thereby giving rise to the advent of various advanced applications. Nevertheless, with the steady development of high-bandwidth and ubiquitous mobile access networks, the new paradigm is to connect all things and all people, thereby placing control at our fingertips.

This novel mindset has had a significant impact on various segments of society. From the perspective of consumers, the end users of networks are demanding fast, reliable, and secure connectivity in order to access the diverse range of services and applications that are available to them. Similarly, service and communication operators are updating their portfolios with the objective of offering not only the network as a means of connectivity but also as a service to meet the needs of traditional subscribers as well as industrial, and enterprise customers. This situation created an ideal environment for third parties, also known as “verticals” [26, 27], to utilize the communications infrastructure to focus on specific value-added, high-quality, and easily accessible services and applications across a range of sectors, including business, industry, entertainment, military, and emergency. Examples of this trend are emergent technologies such as Artificial Intelligence (AI), metaverse, cloud-based approaches, Web3, Industry 4.0 and 5.0, and connected devices, among others [28]. Likewise, opportunities have been foreseen by device vendors diversifying the market with a plethora of pieces of equipment with different levels of complexity, capabilities, and costs.

Beyond the network’s usage, the data generated by all these heterogeneous sources becomes a critical issue concerning how to manage this volume of information while providing high-quality experiences to customers. In this sense, some of the world’s most important technology providers forecasted in recent years that by 2030 the number of mobile subscriptions will be about 9.3 and 10.3 billion [29, 30]. This prediction points that the traffic Compound Annual Growth Rate (CAGR) will be close to 22% reaching a global volume up to 468 Exabytes (EB) per month that is 46.4 Gigabytes (GB) per user in the same period. In these

studios, it is expected that by 2030 the most critical traffic contributors will be video services with up to 60% of the share, followed by metaverse applications with 13%, social networking adding 10%, gaming 6%, and marketplace 4%. In comparison with current trends where video accounts for 72% of the traffic, this share will be modified by the impact of Extended Reality (XR) and Cloud Gaming (CG) services.

As a consequence of these never-seen needs, Communication Service Providers (CSPs), formerly known as Mobile Network Operators (MNOs), have revolved and redesigned their business models to offer the Infrastructure-as-a-Service (IaaS), thus, it is accessible transparently for customers of different types to deploy high-quality and multi-purpose End-to-End (E2E) experiences. Nevertheless, ensuring such levels of quality is not a trivial task for MNOs considering the variety of service-specific requirements, heterogeneous scenarios demand a degree of network automation that reduces Capital Expenditure (CAPEX) and Operational Expenditure (OPEX).

This need was first addressed by the Next Generation Mobile Networks Alliance (NGMN) by proposing the Self-Organizing Networks (SON) approach use cases [31], later standardized in Release 8 by the 3rd Generation Partnership Project (3GPP) and developed for LTE technology, however, backward compatibility was planned for 3G [32]. In essence, this paradigm was conceived to automate functions that traditionally were managed by the network operators including self-configuration, self-optimization, and self-healing. This automation endowed the infrastructure with a degree of intelligence benefiting from the capabilities to adapt to adverse scenarios and improve overall performance [33].

This paradigm drew the attention of the research community, including R&D from vendors, which presented a plethora of proposals to address the SON use cases [34]. The key aim was to leverage expert knowledge to establish diverse automatic network management algorithms. Nevertheless, as the standards of the network matured towards LTE-A Pro, Machine Learning (ML) was considered the perfect pivotal tool to overcome the complexity introduced by the variety of new services, and devices to ensure network management scalability [35].

With the advent of the Fifth-Generation Mobile Networks (5G), the SON concept has evolved to the Next-Generation SON (NG-SON) concept. This proposition intends to advance the algorithmics to intelligent mechanisms that can improve the network's performance for all situations, differing from the LTE-specific SON

use cases [36]. This strategy leverages the increasing amount of data the network collects, thus, providing the network with context awareness that enables scalability, flexibility, and resilience [37, 38].

Despite the network-centric performance discussion, intelligent networks are envisioned to be more flexible and powerful in the next standards of new-generation 5G and Sixth-Generation Mobile Networks (6G). In this sense, one of the most relevant factors considered is the QoE the customer perceives, advancing the management from network performance to service-based approaches [39]. This consideration aims to transform conventional mobile networks into flexible and extensible infrastructure that can be leveraged by end-users, enterprises, verticals, Application Service Providers (ASPs), etc.

To meet the user expectations, an E2E knowledge of the service is required. However, there is still a gap to be filled in this area since research responds to specific problems but does not create an integral solution for network management. Former research has been performed concerning ML as a technology enabler for network-centric metrics such as Channel-Quality Indicator (CQI) [40], pole capacity [41] as well as the prediction of traffic volume [42, 43]. Likewise, the information contained in the context of the communications (e.g., crowd and social events) has been employed in [44]. Likewise, ML has been used to address load balancing, mobility management, and energy-saving topics [45].

To address the QoE approach objectives and subjective analysis can be carried out regarding the metric used as Figure of Merit (FoM). For instance, previous research has been driven for Mean Opinion Score (MOS) mapping from network Key Performance Indicators (KPIs) [46], Key Quality Indicators (KQIs) from network KPIs [47] or KQI from packet-level statistics [48], KQI estimation from radio metrics [49].

As highlighted above, the need for strategies that involve all the segments of the communication between the customers and the ASPs, the service provision, and the perception of the user experience is mandatory to meet the objectives specified for the new-generation mobile networks. This focus known as E2E management aims to address the challenges introduced by the novel, heterogeneous, stringent, and specific requirements envisioned for the next generation use cases [50].

1.2 Preliminaries

This thesis has been conducted within the *Mobile and Aerospace Networks Lab research group (MobileNet)*, which is part of the Telecommunications Research Institute of the University of Málaga (*TELMA*).

The *MobileNet* group was established in 2000 through a collaboration between the TIC-102 group and Nokia Networks, leading to the creation of the Mobile Communications Research Centre at the Andalusia Technology Park (PTA) in Málaga.

One of the primary research areas of *MobileNet* is the application of Artificial Intelligence / Machine Learning (AI/ML) to mobile communication networks. In collaboration with Nokia Networks - Spain, one of the group's earliest projects involved developing an automated tool for resolving Radio Access Network (RAN) issues. This project laid the groundwork for integrating real-world mobile network data with engineering expertise to create automated troubleshooting solutions.

Since then, the group has been actively involved in developing SON techniques through partnerships with national and international companies as part of various research initiatives. Among these, the *H2020 MONROE* project [51] aimed to create an open-source hardware platform for experimentation and measurement in commercial mobile broadband networks, known as MONROE nodes. Within this project, *MobileNet* introduced the *eSON* proposal as part of the *MONROE* open calls, which utilized radio layer data and network configuration information. The MONROE nodes were employed to collect comprehensive end-to-end data from user devices, while simultaneously accessing and controlling radio network configuration parameters.

Moreover, *MobileNet* contributed to the *H2020 ONE5G* project [52]. This research project was concerned with the development of generic end-to-end (E2E) performance optimization strategies for next-generation mobile networks. The objective was to enhance both the efficiency of network operators and the end-user experience through radio optimizations. In a similar vein, the group has been engaged in the “*Red.es Piloto 5G-Open RAN*” and *AECMA-5G* projects. The aforementioned projects have established the foundation for this thesis in the context of E2E service management for 5G networks.

In these projects, *MobileNet*'s infrastructure played a vital role in the proof-of-

concept experiments. These latter were carried out on a commercial LTE network, known as *UMAHetNet*, which consists of 12 picocells with and Evolved Packet Core (EPC) [53] and a 5G Stand-Alone (SA) network involving 3 macro-cells and 6 indoor picocells [54], and Network-in-a-Box solution that integrates Software-Defined Radio (SDR) and a virtualized core [12]. The interaction with the networks is possible through the use of Application Programming Interfaces (APIs) that exposes some functionalities that enable the use of external intelligent mechanisms developed in the group.

In the context of prior research on service management, MobileNet has performed developments concerning traditional services such as file transfer [55], and web browsing. Similarly, the team has conducted research into the domain of multimedia services, including Video-On-Demand (VOD) through the use of platforms like YouTube or live video on Twitch. [20]. Nevertheless, the advent of sophisticated services (e.g., immersive systems) aims to conceptualize novel scenarios that have yet to be fully examined, thereby opening new avenues for investigation in the context of next-generation networks.

1.3 Challenges and objectives

This thesis is motivated by the aforementioned existing gap concerning service-based network management in the context of a mobile technology revolution that seeks to transform the existing paradigm. As the next generation of mobile networks evolves towards a scalable, flexible, intelligent, resilient, and multi-purpose infrastructure, the advent of cutting-edge use cases and services is imminent. Moreover, the next generation of networks is designed to facilitate the development of advanced services and scenarios, including multiverse use cases, Industrial Internet of Things (IIoT), cloud-supported gaming, storage, and computing, among numerous other proposed applications.

In light of the current technological maturity level and the existing State-of-the-art (SoTA) systems, it is challenging for operators to provide high-quality services relying solely on traditional strategies based on specific rules, which are now considered obsolete in some standardization bodies' debates. This creates an opportunity for the emergence of various technology enablers, including AI/ML, which aim to leverage constantly evolving information in order to identify patterns, behaviors, and data that can be utilized for intelligent automation in

network management.

Accordingly, this thesis's principal objective is to present a quality-driven mechanism for next-generation network management in the context of advanced multimedia services.

To approach this, the main objective will be divided into sub-objectives that address the diverse challenges originated from it. Firstly, it is of paramount importance to analyze these novel use cases with a view to understanding their nature, the requirements they entail, and how to characterize their performance in terms of an E2E user-centric vision. To this end, Objective 1 will examine the prospective integration and deployment of next-generation services within the context of next-generation networks, with a view to ensuring the delivery of high-quality experiences that meet the expectations of both end users and CSPs. However, it should be noted that there is a considerable range of services that fall within this category, and this thesis will therefore place particular emphasis on those that have been derived from CG and XR technology.

The term *XR* is a catch-all designation for a range of technologies, including Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR). The principal distinction between these technologies is the extent to which they facilitate the introduction of elements of immersion and reality modification relative to the physical reality of the user. This situation presents a significant challenge for CSPs, given the need to address a complex set of requirements simultaneously. For example, certain VR applications necessitate high downlink throughputs and stable latencies, whereas AR can prioritize uplink throughput and reliability in communication.

Similarly, CG is a service that is defined by the delegation of game logic to a remote server that acts on behalf of the user. This consideration is intended to facilitate the provision of gaming services on a range of affordable and heterogeneous devices, including mobile phones, computers, and tablets. This system evolution necessitates the fulfillment of particular requirements pertaining to both downlink and uplink, with the objective of ensuring consistent latency and adequate visual quality.

To address this call, Objective 1.1 will focus on the principal metrics that reflect the performance of the service from an E2E perspective. In order to achieve this, the most appropriate KQIs will be selected in accordance with the guidelines

set out by the principal standardization bodies, including 3GPP, European Telecommunications Standards Institute (ETSI) and International Telecommunication Union (ITU). Subsequently, the design, development, and implementation of service-specific clients are essential for achieving this objective. The clients are pieces of software that integrate functionalities to deliver services in end-user equipment. Additionally, they measure and collect high-layer (i.e., application) metrics that reflect the E2E system performance. This allows for the deployment of controlled service exploration and experimentation through the use of experimental testbeds utilizing mobile networks.

Objective 1.2 aims for a holistic approach to examining the influence of diverse network factors on service performance. Although wireless networks offer the potential for enhanced mobility, flexibility, and scalability in service provision, the inherent challenges associated with radio factors (e.g., propagation, obstacles, resource sharing) can result in a reduction in performance. This reality may impede the efforts of various sectors and equipment vendors to expedite the development of services and applications that are compatible with these emerging technological trends. To demonstrate the feasibility of adequate service provision through mobile networks, controlled-scenario measurement campaigns using diverse radio conditions, wireless technologies, and mobility are assessed through the testbeds developed for Objective 1.1.

In light of the heterogeneous requirements of next-generation services, another challenge arises for operators: the need to manage networks in a way that is driven by services and customer expectations, rather than by legacy network-centric performance metrics. Moreover, network architectures are evolving into decentralized schemes, advancing from monolithic equipment to software-based approaches that will bring flexibility and scalability to the infrastructure. Moreover, novel technology drivers are taken into account, including new radio bands, dynamic resource allocation schemes, numerologies, and AI/ML, among other features. These are intended to enable the tailoring of logical networks designed for specific use cases and demands. However, this adds complexity to the network and presents previously unencountered challenges that must be addressed.

To overcome these situations, Objective 2 focuses on the application of machine learning to develop intelligent mechanisms for service-based network management by leveraging the data available on the operator's side.

Previously, Objective 1 emphasized the significance of quantifying service performance through objective, comprehensive metrics that reflect the overall system performance. Nevertheless, in certain circumstances, it is not feasible to quantify this due to concerns regarding user privacy or the absence of mechanisms from the ASPs that facilitate the provision of this information. To address this challenge, the use of ML is imperative for predicting this crucial information, which then feeds into the network management mechanisms. Objective 2.1 will center on the feasibility of training KQI prediction models using network-accessible data, including radio metrics, statistics, and configurations reported by various network segments, except for user-end equipment. This proposal allows for the replication of service behavior and the implementation of necessary actions to ensure compliance with the requirements outlined in the Service-level Agreements (SLAs). The SLA represents a set of rules and measurable indicators that regulate the CSP performance in accordance with the service quality expectations of the customers.

In addition to the ability to model service functionality, the development of an effective strategy for ensuring the provision of high-quality services necessitates the implementation of advanced solutions that capitalize on the advantages offered by ML. Therefore, Objective 2.2 will pursue the development of an E2E, cross-layer service-based mechanism to facilitate the management of networks, concerning service-based resource allocation, while optimizing their performance in terms of KQI compliance. In light of the aforementioned considerations, it is feasible to guarantee the provision of high-quality services in a multitude of circumstances and scenarios, as delineated by the intrinsic characteristics of the network-accessible data. Similarly, the application of ML enables the infrastructure to deploy flexible and scalable solutions tailored to diverse use cases with varying requirements in production. This facilitates the provision of intelligent, transparent, and multi-purpose infrastructure as a service for verticals, ASPs, and end customers.

In summary, the thesis' objectives are as follows:

- **Objective 1:** To explore arising services intended to be deployed in new-generation mobile networks. Especially, this aim is split into two sub-objectives:
 - **Objective 1.1:** To deploy services and measure their performance from an E2E perspective. Service-based management of the networks requires

the collection of metrics that reflect the service performance from a holistic point of view, thus, enabling performance monitoring driven for high-layer data.

- **Objective 1.2:** To assess the network factors that impact the E2E service delivery. Each segment of the communication chain impacts the delivery of the service. Identifying how the network affects service performance under different scenarios is of utmost importance to determine the most adequate manner to tackle negative effects.
- **Objective 2:** To establish data-driven intelligent mechanisms for E2E service-based network management. To address this objective, sub-objectives are proposed to approach these aims:
 - **Objective 2.1:** To leverage rich E2E data and ML mechanisms to predict service performance. Considering that accessing user-end service performance data cannot be ensured because of user privacy, and limitations on the implementation of measurement functions in the service-specific applications/agents, alternative approaches should be used. ML can leverage network-available information to predict KQIs to reflect the degree of satisfaction of the service requirements.
 - **Objective 2.2:** To optimize the E2E service delivery through intelligent network approaches. ML-based approaches are adequate to enable KQI-driven decision-making for service-based network management. Permanent and dynamic knowledge of the service performance can accelerate the adequate resource determination to meet diverse SLA policies for new-generation services.

1.4 Document structure

This thesis is structured through eight chapters involving three main parts: i) Foundational background, ii) Published works and contributions, and iii) Final reflections and Conclusions. This organization is illustrated in Figure 1.1.

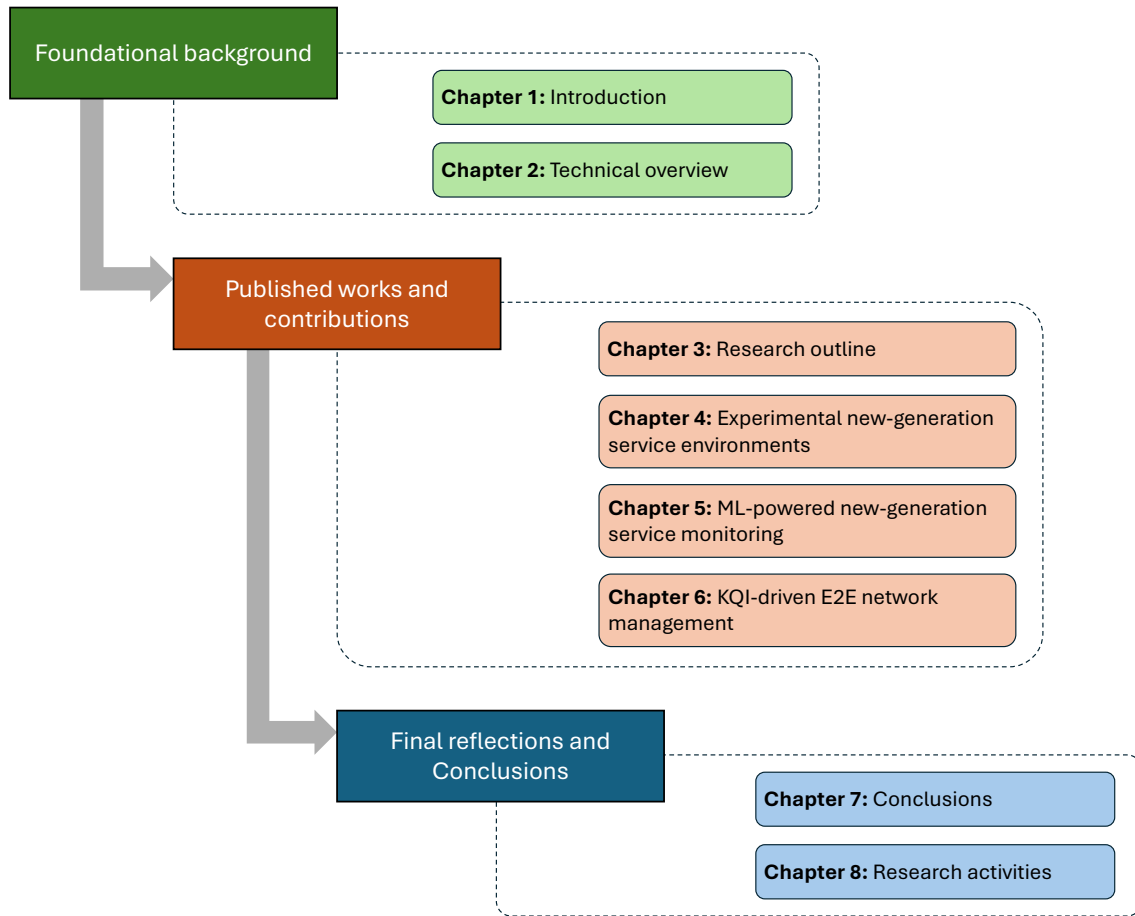


Figure 1.1: Document structure

Part I - Foundational background

The initial section of this document provides the motivation for the thesis and the relevant technical context. It starts with Chapter 1, which outlines the objectives and scope of the thesis. Next, Chapter 2 presents an overview of the theoretical framework, introducing key concepts related to cellular networks, operation and management, and the technologies used for the experimental use cases. This chapter emphasizes the fundamental aspects of new-generation networks that can be considered drivers for the delivery of new-generation services and applications.

Part II - Published works and contributions

This part comprehends the core of the research work developed in this thesis. This set of sections describes the various articles encompassing the challenges and objectives established in the first part. In this manner, Chapter 3 presents the research schema split into two subsections dedicated to conveying the research

methodology and a summary of the articles supporting this predoctoral process, which are mapped into the following chapters:

- Chapter 4 addresses the challenges mapped to Objective 1 by exploring diverse cutting-edge use cases that are expected to be deployed by next-generation networks. In this regard, this chapter develops around three publications that examine the architectural concepts and requirements, as well as the design and development processes of service-specific clients for CG, 360-video, and Cloud VR services. Subsequently, in order to address Objective 1.2, the analysis of the impact of the network on the services is presented in the results section of each article. A range of strategies have been discussed in each article to approach the various network effects.

First, the impact of using wireless networks in CG is analyzed in comparison with Ethernet technology, which is well-known to be the preferred technology among users of this type of service. Regarding the provision of 360-video services via mobile networks, specifically LTE and 5G-SA, the evaluation of these services in varying radio conditions is conducted through the use of KQIs. The final article presents an assessment of Cloud VR employing WiFi and 5G in both static and mobile scenarios, situated within the context of a real-world industry scenario. The results demonstrate the superiority of 5G over WiFi in terms of providing a stable service in challenging environments with significant interference and mobility.

- Chapter 5 employs a framework for automatic KQIs modeling. This framework is designed and implemented with the objective of addressing Objective 2.1. This research work is comprised of a single article that delineates the design procedures of the framework, which aggregates data preparation, training, hyperparameter tuning, evaluation, and model selection through a multitude of algorithms. One of the key functionalities of the training process is the aggregation of multiple feature engineering techniques, such as Feature Selection (FS) and Feature Extraction (FE), that endow the framework with the ability to search the strategy that best fits the service KQIs. Furthermore, a novel metric, designated as PET_{SCORE} , is introduced.

To assess the efficacy of this proposal, the 360-video service was selected as the use case. In this regard, the framework's efficacy is evaluated by inferring

models for an array of service quality indicators, thereby gauging its aptitude for forecasting these metrics. The resultant dataset created for this evaluation was released publicly to the research community in order to strengthen the investigation in this field.

- Chapter 6 presents the final step of the research work performed in this thesis. The article under consideration encompasses Objective 2.2 and integrates the prior developments into an integral approach to perform service-oriented network management. Given the diverse range of utilities that can be derived from this approach, the resource allocation problem has been selected for evaluation of the proposal.

The solution entails an optimization problem, whereby the objective is to maximize the QoE while minimizing the number of assigned KQIs. This is achieved through a target function that is optimized using a dual-annealing approach to determine the combination of service KQIs that satisfies the requisite condition. It is crucial to emphasize that the indicators must meet the minimum values specified in the SLA.

This approach demonstrates the feasibility of leveraging ML and QoE to develop E2E mechanisms with the objective of strengthening the development of intelligent next-generation service management.

Part III - Final reflections and conclusions

This section comprises two chapters that present a synthesis of the findings and outcomes of the thesis. Chapter 7, presents the conclusions derived from the research work, along with prospective avenues for future inquiry and implications pertinent to the industry. Subsequently, Chapter 8 presents the author's participation in various research activities.

Appendices

This section contains an excerpt from the thesis in Spanish, which offers insight into the objectives, conclusions, and research activities. For further details, please refer to Appendix A.



UNIVERSIDAD
DE MÁLAGA

Chapter 2

Technical overview

Content

2.1	Mobile networks	19
2.1.1	Mobile networks background and architecture	19
2.1.2	LTE	24
2.1.3	5G	30
2.1.4	The Open RAN concept	39
2.2	5G New Radio (NR)	42
2.2.1	NR frequency bands	42
2.2.2	NR Waveforms	44
2.2.3	NR frame structure and numerologies	45
2.3	Machine Learning	49
2.4	Network management	51
2.4.1	Service-oriented management	53
2.5	Extended Reality background	56
2.6	Conclusions of the chapter	58

This chapter offers an overview of the fundamental concepts that form the basis of this thesis. In this regard, Section 2.1 elucidates the fundamental elements of mobile networks. It provides a concise overview of the evolution of mobile networks, subsequently focusing on the architectural concepts and features associated with the currently deployed networks, namely LTE, 5G, and its vision for the Open

RAN approach. Section 2.2 describes the principal characteristics of the 5G RAN, which marks the transition from traditional networks to service-based infrastructures designed to deliver next-generation services and applications. Subsequently, Section 2.3 introduces the fundamental concepts of ML, which is expected to drive the evolution of network management. This latter topic is discussed in greater detail in Section 2.4. Finally, Section 2.5 provides an overview of the technology known as XR, including an examination of its various forms and the requirements that define the challenges that next-generation networks will face in order to support the delivery of advanced services.

2.1 Mobile networks

Mobile networks are telecommunication systems designed to provide global wireless connectivity, enabling reliable mobility support and value-added features that other technologies are unable to offer. Although originally conceived as a means of providing voice service, the considerable success and evolution of these networks have demonstrated their suitability as a platform for future communications, including multimedia, data, immersive technologies, and the IIoT, among other potential applications.

These networks, also referred to as “cellular”, derive their nomenclature from the strategy employed to partition the geographical coverage area into discrete units, or “cells”. Each cell is served by an independent base station that is coordinated with the others in order to provide reliable and uninterrupted connectivity across the network’s service area.

2.1.1 Mobile networks background and architecture

The concept of the cellular network was first proposed by AT&T in the 1940s; however, it was not until the 1980s that the first deployment was released by NTT in Japan. This marked the advent of cellular technology in the 1980s with the introduction of the First-Generation Mobile Networks (1G), which provided analog voice services. The advent of 1G networks initiated a transformation that led to the introduction of a digital framework with the potential to enhance the quality of service. This novel approach led to the introduction of the Second-Generation Mobile Networks (2G) standard in the early 1990s, which provided users with digital voice service, enhanced privacy through encryption, and introduced the concept of the Short Messaging Service (SMS), enabling users to communicate using text. A number of competing technologies vied for dominance in this generation, with Global System for Mobile Communications (GSM) ultimately emerging as the selected standard. This latter iteration advanced in its development to provide, for the first time, data service using packet-switching transmission. The General Packet Radio Service (GPRS) standard, which is classified as 2.5G, was released in the late 1990s. Subsequently, further developments were initiated with the advent of upcoming standards, such as Enhanced Data rates for GSM Evolution (EDGE), which sought to enhance the

data rates of GPRS.

In order to mitigate the effects and development delays that are inherent to the presence of diverse competing standards, the 3GPP consortium has formed a coalition with several partners with the objective of defining the Third-Generation Mobile Networks (3G) standard. The organization comprises members from a variety of backgrounds, including global equipment vendors, research institutions, academic entities, and regional standardization bodies. The 3GPP is currently regarded as the primary entity responsible for maintaining and updating the mobile network standards, spanning from 2G to 5G.

In regard to 3G networks, this generation was entirely predicated on the notion of mobile communications. In this regard, the objective of 3G was to emphasize the provision of data as the primary service offered by the infrastructure, rather than voice. This generation was distinguished by the advent of Universal Mobile Telecommunications System (UMTS) as the dominant standard in the early 2000s. The 3G network employed Wideband Code Division Multiple Access (WCDMA) as the mechanism to facilitate simultaneous connectivity among users through the use of different codes, while sharing the entire frequency band. This constituted a suitable advancement in comparison with the preceding Time Division Multiple Access (TDMA) standard in 2G. Similarly, the 3G standard initially introduced the IP Multimedia Subsystem (IMS) as a unified framework to facilitate fixed-mobile convergence in the delivery of multimedia services through IP.

Despite initial expectations that this technology would address future communications, UMTS encountered challenges related to device power consumption, signaling security breaches, and call-dropping when technology switching was needed. Furthermore, the theoretical data speeds could not be met as expected. Hence, the objective of 3GPP was to address these issues through subsequent releases of standards, such as Evolved High Speed Packet (HSPA+), thereby establishing the foundation for the design of the next-generation specifications.

Concerning the generalities of the architecture, mobile networks involve three indispensable parts: i) the Radio Access Network (RAN), ii) the Core Network (CN) representing the operator's side, and iii) the User Equipment (UE) from the customers'. It is important to note that the complexity of this structure varies in terms of the technology generation. The schematic representation of this

arrangement is depicted in Figure 2.1.

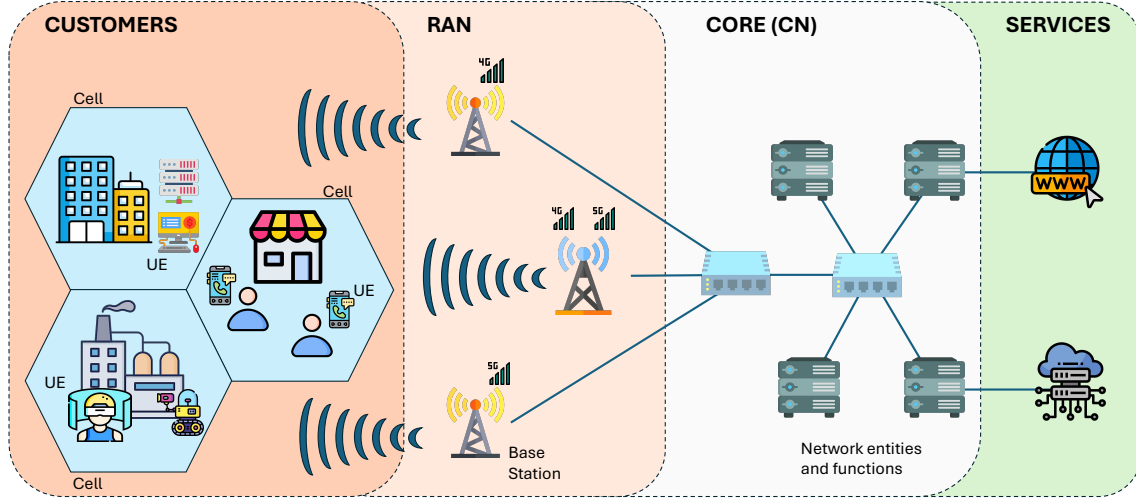


Figure 2.1: Mobile network general architecture

The objective of the RAN is to provide the air interface between the customer and the network core. The RAN is composed of an arrangement of properly located Base Stations (BSs) that dispense ubiquitous connectivity across the service area. The design of the RAN is not a simple undertaking, as it must ensure seamless communication among users (i.e., mobility management for Handover (HO) and service availability for multiple users simultaneously), as well as employing adequate spatial, spectral, and time resource sharing. This approach is of paramount importance, given the necessity of ensuring the coexistence of the numerous technologies that comprise the current complex communication systems.

The CN comprises the network elements and entities that provide the actual mobile service to customers. In this sense, the core can be considered the central component of the network, responsible for routing traffic between users and external networks (e.g., the Internet), managing session initialization and finalization, handling user access and mobility policies, and ensuring the quality of sessions granted. Moreover, the core is tasked with providing supplementary services and exposing the infrastructure's capabilities.

The RAN and CN are generation-specific, meaning that their capabilities are contingent upon the mobile network technology in use. Nonetheless, the efforts of the 3GPP aim to define compatibility among the standards and endow the network with the ability to coexist with other technologies,

In addition to the network, the equipment belonging to the customer is referred

to as the User Equipment (UE). This concept has undergone evolution through the progression of mobile technologies. In the context of 2G, the term "UE" is used to refer to devices that are capable of providing voice and SMS services. However, the evolution of technology from 3G to the current versions of 5G has enabled the UE to encompass any device with the capacity to connect and interact with the network, thereby facilitating the provision of services to the customer (e.g., voice, data, IoT, cloud, and so forth).

The aforementioned components interact through the use of protocols specified by the 3GPP. As the network also necessitates the exchange of control messages, or signaling, the communication is distinguished at varying levels, as defined by a protocol stack. Each layer is responsible for specific functions, which may either control the connection or operate on the data itself.

From the perspective of the network, the CP is responsible for the management of the communication channel between the UE and the network. Examples of these roles include the authentication, establishment, and termination of data sessions; directives from the network pertaining to resource allocation; and the base station that will provide service. In contrast, the UP pertains to the data flow pertaining to customer services and applications. Figure 2.2 depicts the interplay between the protocol stack and the network components, as well as its implications with the latter control or user planes. The interaction of the abovementioned components is performed through the use of protocols specified by the 3GPP. Since the network requires also the exchange of control messages, so-called signaling, the communication is featured at different levels defined by a protocol stack. Each of the layers serves defined functions that can operate on the control of the connection or in the data itself.

On the one side, the Control Plane (CP) comprehends the management functions that maintain the communication channel between the UE and the network. Examples of these roles are the authentication, creation, or finalization of data sessions, directives from the network related to resource assignment, and the base station that will provide service. Conversely, the User Plane (UP) involves the data flow concerning the customer services and applications. Figure 2.2 depicts the interplay between the protocol stack and the network components as well as its implication with the latter control or user planes.

The functions associated with each protocol are the following:

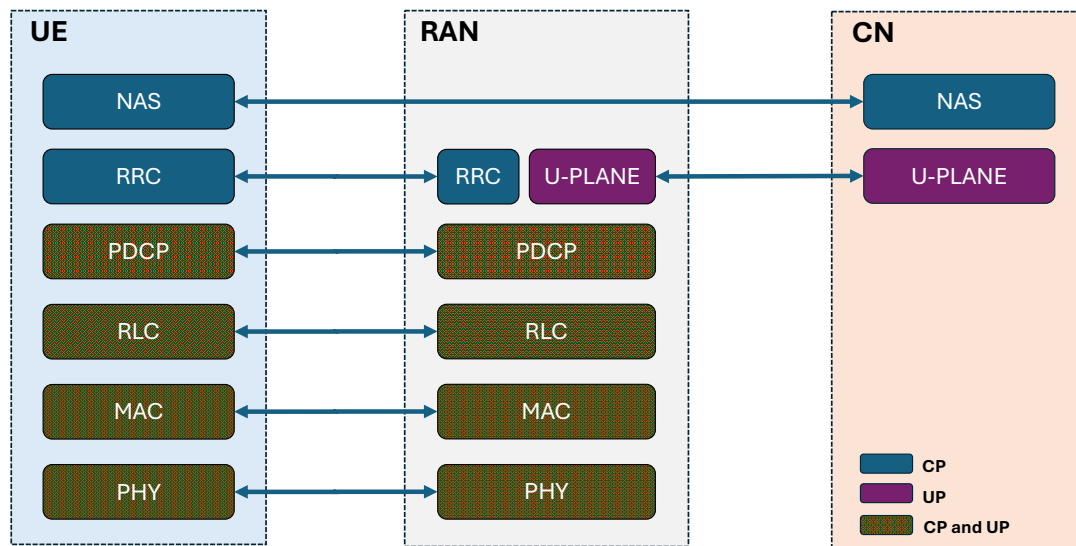


Figure 2.2: 3GPP Protocol stack

- **Non Access Stratum (NAS).** Supports the session management procedures to initialize or finalize an IP session between the UE and the CN entity responsible for managing the user plane. The NAS also conducts mobility management operations.
- **Radio Resource Control (RRC).** Enables broadcasting of the System Information from the NAS. Handles procedures of paging, establishment, maintenance, and finalization of an RRC connection. It is also responsible for some security functionalities.
- **Packet Data Convergence Protocol (PDCP).** This protocol drives IP header compression and decompression, as well as data transfer from both planes. This protocol also handles the PDCP sequence numbers to manage in-order data delivery, cyphering, and control plane data integrity checks.
- **Radio Link Control (RLC).** This layer is responsible for RLC Service Data Unit (SDU) concatenation, segmentation, and reassembly. In addition, RLC performs error correction through Automatic Repeat reQuest (ARQ), re-segmentation, duplicate, and discarding on Protocol Data Units (PDUs).
- **Medium Access Control (MAC).** This layer handles the mapping between logical and transport channels and scheduling information reporting. MAC is responsible for performing multiplexing of SDUs to be delivered to the physical layer in terms of transport blocks. Likewise, it handles

de-multiplexing data coming from the lower layer to be delivered to higher ones.

- **PHYsical layer (PHY).** The lowest layer of the stack is responsible for delivering the MAC-channels data over the air, including modulation, and antenna/physical-resource mapping. The PHY also handles link adaption through Adaptive Modulation and Codification (AMC), power control, and performs measurements on behalf of the RRC layer.

The protocol stack illustrated in Figure 2.2 is a general scheme associated with mobile networks. However, its definition can vary depending on the technology generation. Consequently, the following subsections will focus on the latest mobile technologies that have been the support for the development of the thesis, in particular 4G and 5G.

Beyond the UP and CP, the Management Plane (MP) handles various high-level functionalities for network administration, maintenance and optimization. The MP is tightly integrated with the Operations Support System (OSS), forming the backbone for efficient network operation and service delivery. The OSS enables key functionalities, including configuration management for setting up and maintaining network resources, fault management to detect and resolve network issues, performance management to monitor and optimize KPIs, and security management to enforce access controls and prevent cyber threats [56]. Next-generation OSS are expected to incorporate AI/ML-driven intelligence to facilitate service management to ensure compliance with SLAs and maintain high QoE. Thereby, the OSS ensure cohesive, automated, and scalable network operations, adapting seamlessly to the demands of next-generation networks [57, 58].

2.1.2 LTE

The Fourth-Generation Mobile Networks (4G) were standardized by the 3GPP and first released by late 2008 in Release 8 as an evolution from the predecessor networks. This radio technology named Long Term Evolution (LTE) belongs to the conceptual architecture defined as Evolved Packet System (EPS). In this scheme, a simplified, full-IP core architecture with a new RAN endowed with intelligence were the key innovations of the specifications. The network core is

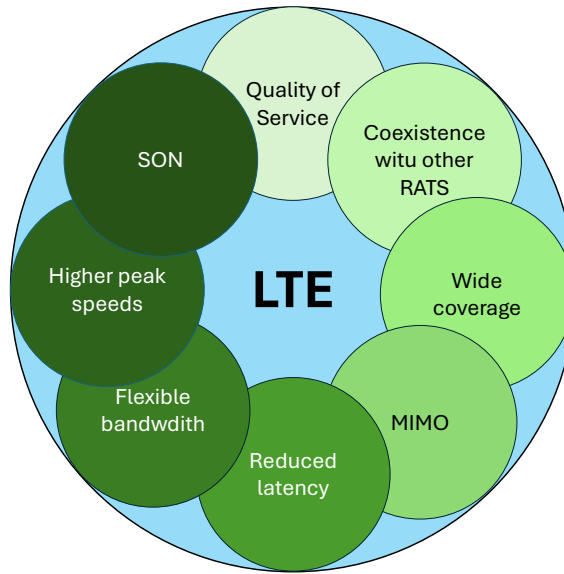


Figure 2.3: Features introduced by LTE

designed according to a System Architecture Evolution (SAE) to reduce the complexity of 3G cores through the EPC. Along these lines, the RAN also advanced to the Evolved Universal Terrestrial Radio Access Network (E-UTRAN) by integrating the functionalities that before were managed by the Radio Network Controllers (RNCs) in 3G (e.g., mobility procedures) into the Evolved Node Bs (eNBs).

LTE presented novel architectural concepts and technology drivers with the dual objective of simplifying the network and enhancing the capabilities of legacy mobile technologies. The main innovation features introduced with LTE are synthesized in Figure 2.3.

The introduction of the LTE standard saw the advent of E2E IP connectivity, whereby all network components are assigned an IP address (i.e., UEs and eNBs). This has the effect of simplifying communications through the implementation of a flat network architecture. Furthermore, LTE demonstrated superior throughputs compared to subsequent generations, reaching up to 100 Mbps in Downlink (DL), while introducing novel use cases. To address this, the utilization of flexible bandwidths was proposed, including 1.4, 3, 5, 10, 15, and 20 MHz. Similarly, the reduction of latency was identified as a significant enabler for time-sensitive services, with the network latency expected to be no more than 10 ms [59].

Moreover, LTE first introduced the use of OFDM as its primary modulation technique, marking a significant evolution in mobile communications by enabling

high data rates, spectral efficiency, and resilience to multipath fading. In addition, the utilization of Multiple Input Multiple Output (MIMO) techniques was considered for spectral efficiency improvement, particularly in the downlink for the initial releases. Furthermore, LTE developed novel mechanisms for coexistence with other 3GPP and non-3GPP Radio Access Technologies (RATs), including WLAN and WiMAX. Similarly, the network coverage was extended to offer up to a theoretical distance of 100 km. With regard to the QoS, LTE guaranteed E2E service quality through the use of specialized bearers (involving core and radio) defined for diverse types of service classes. Additionally, LTE employed the new architectural scheme to integrate network management tasks automation utilizing the SON approach.

The standard LTE has been the subject of continuous evolution over the past decade in order to comply with the current use cases and services. Notwithstanding the innovative approaches delineated by the specifications, numerous sectors classified the Release-8 LTE standard as technology 3.9G. In this regard, the 3GPP developed the LTE Advanced (LTE-A) in Release 10 in early 2011 with the objective of introducing significant improvements for the RAN. This marks the inception of the technology, which is poised to meet the demands of an advanced, evolved technology [60].

The LTE-A standard introduced novel technologies for the radio, including Carrier Aggregation (CA), enhanced uplink multi-antenna transmission, and Coordinated Multi-Point transmission/reception (CoMP). These updates were designed to enhance the network's performance, particularly in the DL direction, with a theoretical throughput of up to 1000 Mbps.

CA defines a methodology for increasing the bandwidth by aggregating Component Carriers (CCs) up to a maximum of 100 MHz. The objective was to repurpose the Release-8 specifications with the aim of enhancing the system's capacity and improving customer experience. Similarly, enhanced UL multi-antenna transmission was proposed as a means of addressing the limitations of the DL MIMO schemes that were available in LTE at the time. The LTE-A standard defined Single-user Multiple Input Multiple Output (SU-MIMO) 8x8 schemes in the downlink and 4x4 schemes in the uplink. Concurrently, CoMP was designed to furnish the UE with multiple connections to geographically disparate independent cells or to Remote Radio Equipments (RREs), acting as centrally controlled relays, that necessitate synchronization and joint decision-making.

However, this can result in an increase in signaling overhead and transmission latency.

Further evolution was provided in the Release-13 standard through the definition of the LTE Advanced Pro (LTE-A Pro) technology enhancements [61]. The objective of this update was to theoretically increase the DL throughput to a value exceeding 3000 Mbps and to reduce latency to a value below 2 ms.

One of the improvements introduced was the increase in the number of CCs from 5 to 32, thereby enabling the attainment of unprecedented wide bandwidths for mobile networks. Similarly, Release-13 announced the utilization of LTE in unlicensed bands, employing Licensed-Assisted Access (LAA). This technology employs a single cell operating in licensed spectrum to ensure control plane integrity for the user, while additional nodes are equipped with aggregated unlicensed spectrum. Nevertheless, the challenge remains in mitigating the negative effects that are inherent to this approach, particularly in the context of the shared spectrum. Furthermore, LTE-A Pro defined enhancements concerning spectral efficiency through the use of Massive MIMO and new use cases, namely LTE-A Pro, in response to the necessity for Internet of Things (IoT) over mobile networks.

Network architecture

The LTE architecture comprises two well-defined parts, the network core, EPC, and the radio network, E-UTRAN, according to the 3GPP specification TS-36.300 [62]. According to the latter, it is signaled that the EPC will be schematically composed of five Network Functions (NFs) that will perform different functionalities [63]. Conversely, the RAN involves the eNBs that interface the customers with the network. Figure 2.4 illustrates the architecture of this technology.

- Home Subscriber Server (HSS). Defined as a central database that stores relevant information of the network subscribers. Key functionalities are the following:
 - Subscriber profile storage including QoS profiles, access restrictions, and credentials.
 - Subscriber authentication and authorization support.
 - Serving MME location tracking.

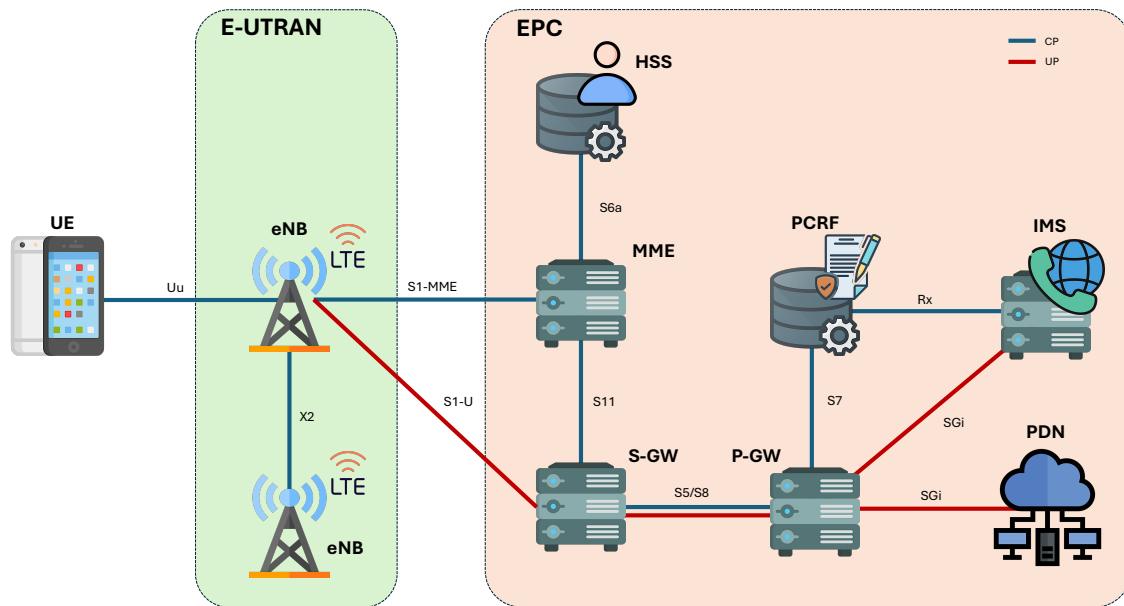


Figure 2.4: LTE architecture

- **Mobility Management Entity (MME).** Handles the control-plane functionalities:
 - Intra-RAT and inter-RAT mobility management.
 - Session management procedures including establishment, maintenance, and release of the bearer between the UE and the network.
 - User authentication and encryption procedures supported by the HSS.
 - Management for users in idle mode, including paging.
- **Serving GateWay (S-GW).** Handles the user-plane functionalities:
 - Packet routing and forwarding between the core network and the RAN.
 - 3GPP intra-RAT and inter-RAT mobility anchor.
 - DL buffer support for subscriber in idle state.
- **Packet data network GateWay (P-GW).** Handles the user-plane traffic between the EPC and external Internet Protocol (IP) networks:
 - Interfaces the LTE network with external Packet Data Network (PDN).
 - IP address assignment to UEs.
 - QoS policy enforcement point.

- Network Address Translation (NAT) and packet filtering.
 - Non-3GPP mobility anchor.
- Policy and Charging Rules Function (PCRF). Defines the policy and charging rules for subscribers:
 - QoS policy control according to subscriber's profile.
 - Charging policy control management based on the subscriber's traffic.

It should be noted that Figure 2.4 depicts the IMS as an external entity that provides additional functionalities to the network. The IMS is a set of standards designed to facilitate the delivery of multimedia services over IP-based networks [64]. The IMS comprises entities that are responsible for session control, media delivery, and service management for Session Initiation Protocol (SIP)-enabled devices. Examples of this approach are the delivery of Voice over LTE (VoLTE), SMS over IP, and Audio-Video over IP (AVoIP), and Rich Communication Services (RCSs).

Concerning the RAN in LTE, specifically the E-UTRAN, the architectural framework has been streamlined to a singular entity, designated as the eNB. The eNB represents an advanced iteration of the NodeB (NB), integrating a multitude of functions initially developed by the RNC and utilized in the preceding generation. This consideration was taken in order to simplify the architecture whilst endowing these entities with functionalities that decentralize the resource management and decision-making processes, thereby reducing latencies and overhead, improving power efficiency, and increasing network scalability [63].

The eNBs integrates two architectural units, i) the Base Band Unit (BBU), and the Remote Radio Unit (RRU). The first acts as the baseband processing part of the eNB and is in charge of resource scheduling, radio link management (i.e., HO, power control), and PHY, MAC and RLC processing. Conversely, the RRU, is responsible for the Radio Frequency (RF) processing tasks including RF amplification, and noise and interference filtering.

In summary, the key functions developed by the eNB are the following:

- Interface between the UE and the network.
- Forwarding control and user plane data to the UE.

- Radio resource management, including resource allocation, QoS, power control, and AMC.
- Control-plane support to the MME and other eNBs.
- Mobility management.
- Session management.
- SON functionalities and load balancing.

2.1.3 5G

The Fifth-Generation Mobile Networks (5G) standard signifies a revolutionary advancement in wireless technology, engineered to address the evolving demands of an increasingly interconnected digital society. In contrast to its predecessors, which were primarily concerned with enhancing data rates and capacity for mobile broadband, 5G introduces a radically new network architecture with the objective of facilitating the emergence of new use cases, including Ultra Reliable Low Latency Communications (URLLC), enhanced Mobile BroadBand (eMBB), and massive Machine-Type Communications (mMTC)[65, 66]. It is anticipated that these innovations will facilitate a diverse array of applications [67], including immersive multimedia experiences, critical infrastructure control, autonomous vehicles, and the proliferation of the IoT and Industry 4.0 and 5.0 [68].

The development of 5G technology has started from 3GPP Release 15 [69] and addresses diverse innovations that cannot be approached by LTE and that respond to the IMT-2020 and beyond objectives [70]. Table 2.1 synthesizes the main differences between 4G and 5G technology capabilities [71]. Likewise, Figure 2.5 illustrates the manner in which the novel use cases can leverage the 5G capabilities and features regarding the level of concern given to each feature.

5G presents new technology drivers that enhance the performance of the networks and pave the way to bring to life the concept of Network-as-a-Service (NaaS) instead of a typical infrastructure scope. In this sense, 5G operates across a wider spectrum of frequencies, including sub-6 GHz bands and millimetre Wave (mmWave), thereby enabling unprecedented data transfer speeds and network capacity. The broader spectrum usage, in conjunction with technologies such as massive MIMO and beamforming, enhances spectral efficiency and ensures a more

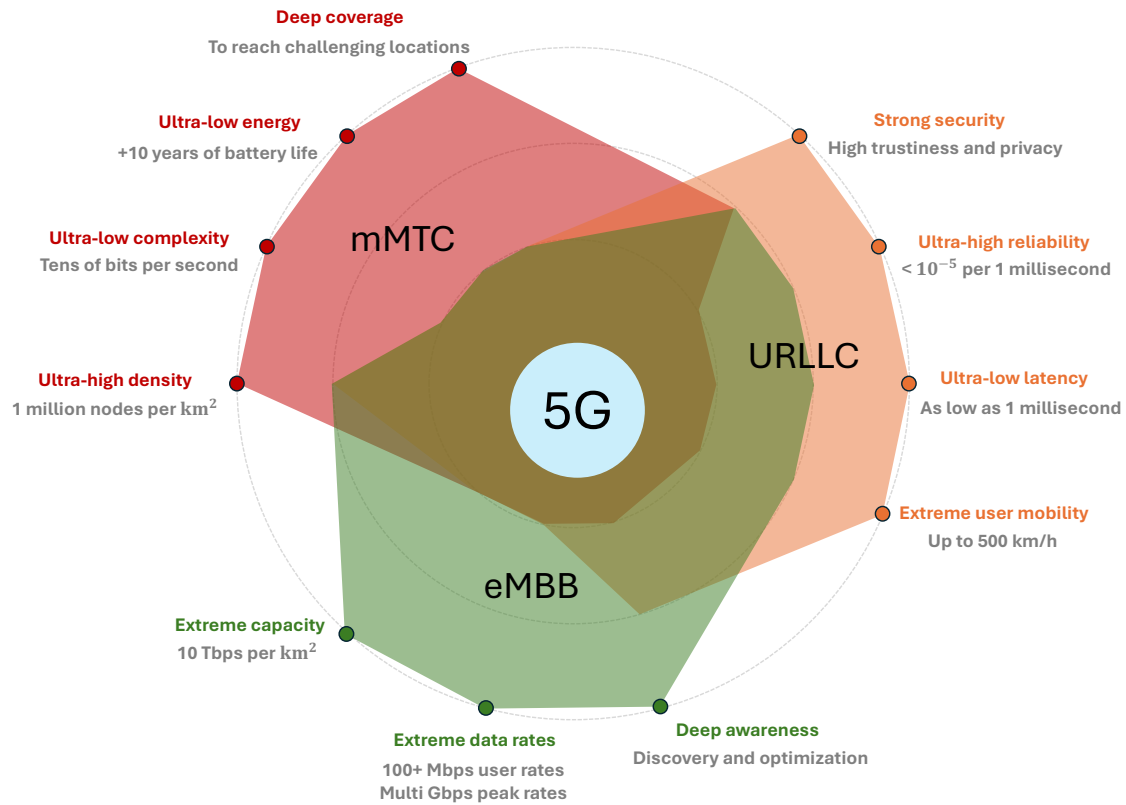


Figure 2.5: 5G service categories and key capabilities [23]

reliable connection, even in densely populated areas. Moreover, the advent of Network slicing (ns) in 5G enables operators to partition the network into multiple virtual segments, each tailored to specific services and performance requirements, thus, facilitating flexibility in service delivery.

Additionally, the 5G architectural framework is evolving towards a more distributed, cloud-native core, leveraging Software-Defined Network (SDN) and Network Function Virtualization (NFV) to enhance scalability, reduce operational costs, and augment the agility of network deployment. These developments render 5G highly adaptable, and capable of evolving in conjunction with emerging technologies such as AI, edge computing, and blockchain, all of which can be integrated into 5G's ecosystem to enable innovative services.

Network architecture

The 5G architecture represents a significant departure from previous network designs, integrating advanced technologies and a flexible architecture to meet the diverse requirements of next-generation mobile networks. The 5G architecture,

Table 2.1: 5G vs 4G capabilities

	4G	5G
Peak download speed	1 Gbps	20 Gbps
Peak upload speed	0.2 Gbps	10 Gbps
Radio latency	10 milliseconds	1 millisecond
Availability	99.99%	99.999%
Reliability	99.9%	99.999%
Mobility	350 Km/h	500 Km/h
Positioning accuracy	45 m	1 m
Device density (per square mile)	250	2.5 million
Energy efficiency	x1	x100

also identified as Fifth Generation System (5GS), is structured to support a wide range of applications agglutinated in the aforementioned use cases. This versatility makes it suitable for both consumer and industrial use cases.

The 5G architecture is comprised of two principal segments: the 5G RAN and the 5G Core (5GC). The collective functionality of these components constitutes a unified system that ensures seamless connectivity, enhanced performance, and the capacity for new services.

On the one side, the 5GC network represents the central component of the 5GS architecture, assuming responsibility for the management of data traffic, the facilitation of user mobility, and the delivery of essential services across the network. The 5GC network is designed to be more flexible, scalable, and efficient than previous generations. It is cloud-native, built on principles such as Service-based Architecture (SBA) [72], NFV, and Control and User Plane Separation (CUPS).

SBA represents a pivotal architectural design wherein NF are implemented as discrete services, named Virtual Network Functions (VNFs), that interact with one another via standardized APIs [73]. In contrast to the conventional, closely integrated network components, the SBA facilitates a more adaptable and modular configuration that sets the basis for the progressive evolution of the 5G standards and their core functionalities. CUPS leverages this flexibility to separate the UP and CP, even physically, thus, creating new opportunities to configure the network based on the upcoming requirements (e.g., closing the UP to the user can significantly reduce latency).

The 5GS architectural approach is illustrated in Figure 2.6, and a detailed

account of its fundamental elements is provided below.

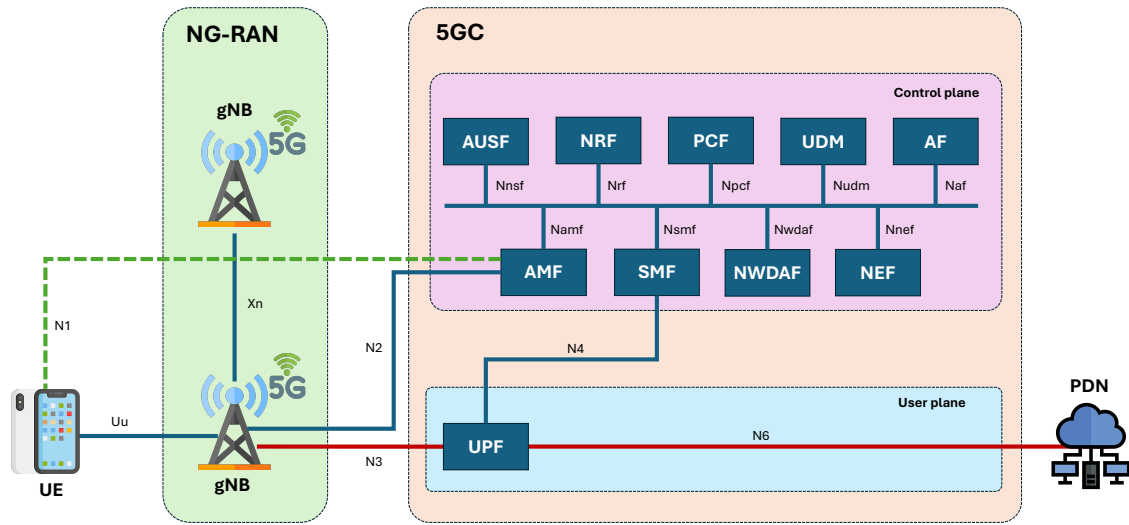


Figure 2.6: 5G system architecture

- User Plane Function (UPF). This entity handles the traffic between the RAN and external networks. Some functionalities are the following:
 - Traffic routing and forwarding.
 - Traffic management, including shaping and prioritization.
 - Packet inspection.
 - Enforcement of QoS policies.
 - Anchor point for inter-RAT and intra-RAT mobility.
 - Facilitate traffic offloading and steering.
- Access and Mobility Management Function (AMF). This VNF manages part of the control panel previously handled by the MME in 4G in terms of access and mobility:
 - User authentication, authorization, and registration management.
 - Mobility management.
 - Handles connection setup and NAS signaling between UEs and the 5GC.
 - NAS ciphering.
- Session Management Function (SMF). This function complements the AMF concerning CP functionalities, and furthermore incorporates functionalities from the LTE P-GW:

- Establishment, maintenance, and release of data sessions.
 - IP address allocation and session management.
 - Supports UPF for data management and QoS enforcement.
- Unified Data Management (UDM). Develop similar functionalities to the HSS in 4G:
 - Storing and management of subscriber profile and data.
 - Handles subscriber authentication credentials.
- Authentication Server Function (AUSF). Handles subscriber authentication:
 - User/device authentications.
 - Leverages UDM for user credential verification and access.
- Policy Control Function (PCF). This entity is constituted as the updated version of the 4G PCF:
 - Enforces network policies concerning QoS, resource allocation and profiles.
 - Manages dynamic policy control based on network data and conditions, and subscriber profiles.
- Network Data Analytics Function (NWDAF). Leverages network data to generate insights for service-based network management.
 - Data analytics provision to the 5GC.
 - Analyze network and user performance and behaviors to adapt resource and quality management.
 - Proactive assistance for network management by identifying anomalies and predicting issues.
- Network Exposure Function (NEF). Exposes network services to external ASPs:
 - Service API provision for network services and capabilities exposure to ASPs and verticals.
 - Facilitation of secure interaction between external services and vertical with the network.

- NF Repository Function (NRF). Supports functionalities related to the network services:
 - Stores NF profiles and register available functions.
 - Support for service discovery and communication.
- Application Function (AF). Develops crucial tasks by communicating external services and applications with the network:
 - Interfaces with the core for service-specific requirements.
 - Coordinates with the PCF for QoS enforcement.
 - Support for the exposition of the Application layer to other functions and applications.
 - Charging and billing.

Additional functionalities have been added concerning network slicing:

- Network Slice Selection Function (NSSF). Network slicing management related to:
 - Selection and allocation of network slices for specific services or customers.
 - Data flow routing control concerning services-specific requirements.
- Network Slice-Specific Authentication and Authorization Function (NSSAAF). Supports the network slice access management:
 - User/device authentication and authorization.
 - Assists NSSF concerning security aspects.

The RAN plays a pivotal role in 5G, facilitating communication between devices and the broader internet. It represents a significant evolution from previous generations, offering enhanced capabilities to support the wide-ranging demands of 5G, such as ultra-high data rates, low latency, massive device connectivity, and improved reliability. The Next Generation RAN (NG-RAN), standardized by the 3GPP in the TS 38.300 [74], specifies the Next-Generation Node B (gNB) as evolved nodes that provide air interface with the users.

In general, the gNB functions are analogous to those provided by the 4G eNB, which are summarized below.

- Control and user plane data forwarding.
- Mobility management.
- Resource scheduling and allocation.
- QoS management.
- Support for UE access and attachment to AMF.
- IP header compression and data encryption.
- Radio measurements for decision making.
- AMC and MIMO schemes.

NG-RAN technological drivers

NG-RAN concept entails the concept of RAN independence from the core. This was initially conceived with the objective of promoting the early development of the 5G-Non-Stand-Alone (NSA) technology, which was introduced in Release-15. Nevertheless, this concept has been employed to provide autonomy to the RAN, irrespective of the core technology capabilities (i.e., EPC or 5GC).

Concerning the protocol stack, 5G introduces the novel Service Data Adaptation Protocol (SDAP) [75] on top of PDCP and below the NAS. This innovation is intended to allow sophisticated and flexible handling of the data flows, with the aim to provide services with different levels of requirements. This differs from LTE, where the QoS management consists of the use of static QoS Class Identifier (QCI) mapped to bearers until service finalization. That QoS scheme was defined to support traditional broadband services. Nonetheless, 5G aims to serve a wide portfolio of services with well-differentiated requirements centered on the use cases (i.e., eMBB, URLLC, and mMTC). To address this, in 5G the QoS flows are fine-grained identified, and mapped to QoS Flow Identifiers (QFIs). Then, several services can share the same Data Radio Bearer (DRB) but receive specific QoS treatment based on their indicators.

The key functions of SDAP are summarized below. Figure 2.7 schematically illustrates the updated protocol stack of 5G.

- Interfaces 5GC QoS scheme with the radio system.

- QoS flow mapping to DRBs.
- Service differentiation in both DL and Uplink (UL).
- QFI management including data prioritization and handling.

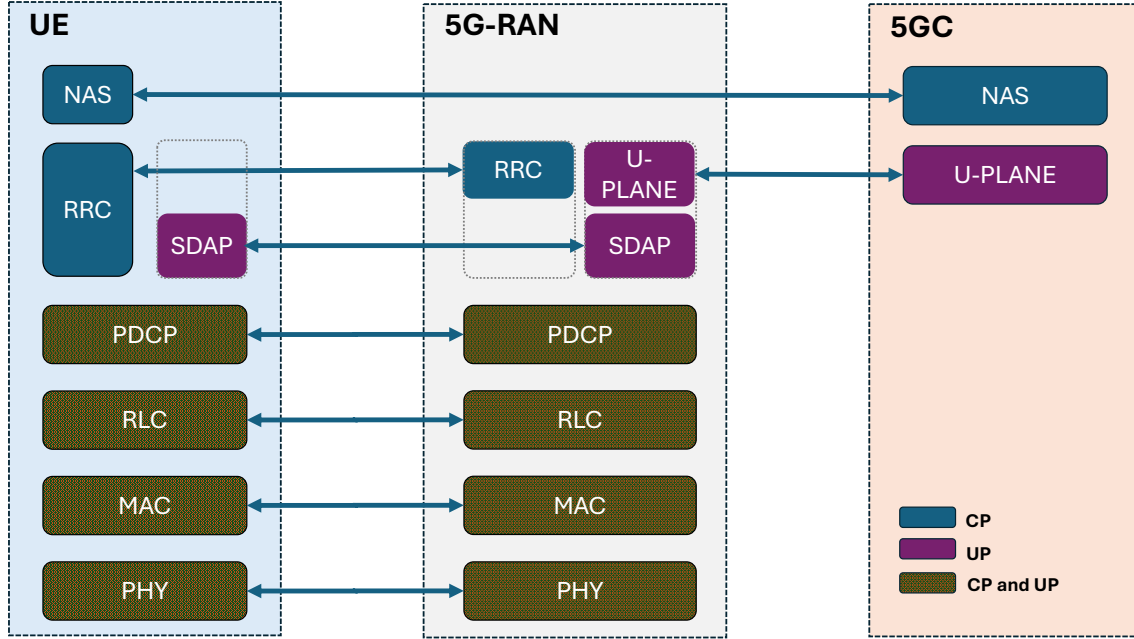


Figure 2.7: 5G protocol stack

Moreover, the NG-RAN proposes a disaggregation of the monolithic architecture by decoupling the hardware from the software components. This strategy permits the extension of the NFV concept through CUPS into the RAN, thereby creating a virtualized environment that facilitates the incremental development of the 5G-RAN standard and the functionalities it can offer. In this sense, the distribution of the RAN network functionalities, described in 3GPP TS 39.801 [76], involves three entities i) the Radio Unit (RU), ii) the Distributed Unit (DU), and iii) the Central Unit (CU).

The terminology “functional splits” responds to the number of layers that agglutinate each RAN entity within the 5G protocol stack. This corresponds to the Access Stratum (AS) protocols (i.e., RRC, PDCP, RLC, MAC, and PHY), where the RLC, MAC, and PHY have been derived into HIGH and LOW sublayers each. This accounts for 8 functional splits, each of them offering a plethora of combinations that provide different benefits and schemes. This feature represents a significant advancement in comparison with the legacy system, where the RAN separated processing and radio functionality through the BBU and RRU entities.

The 3GPP recommends High-Layer Split (HLS) for applications where mobility is not a crucial factor like in Fixed Wireless Access (FWA) service. The HLS considers a partition of the RAN into a CU and DU/RU, delivering most of the functions remotely since the radio coordination is not complex. In the same context, for applications where mobility needs to be leveraged (e.g., eMBB services) a Low-Layer Split (LLS) is suggested, thus, relying on most of the functions in a centralized CU/DU in charge of coordinating RUs [77]. This debate has been put into consideration under the 3GPP in the specification TS 38.816 [78]. Figure 2.8 displays the differences between the HLS and LLS functional splits.

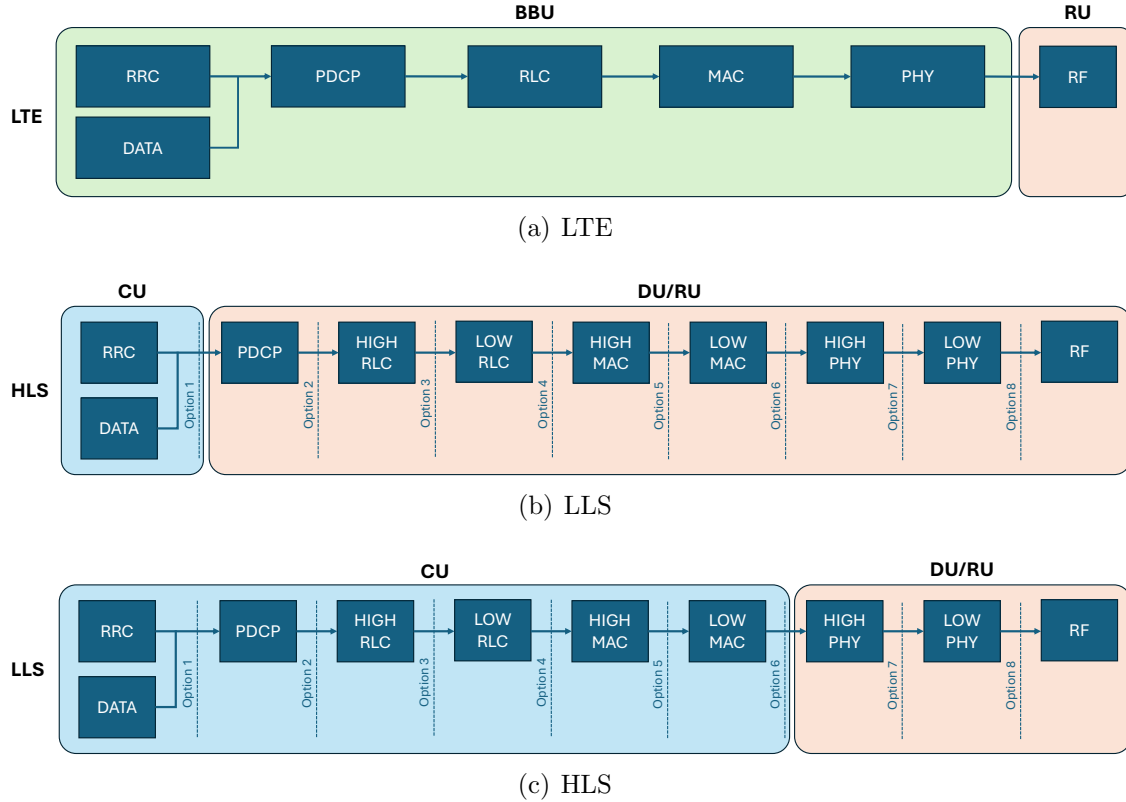


Figure 2.8: Functional split comparison

All these abovementioned technological features are regarded as crucial enablers to enhance network scalability in terms of capacity, and performance. Furthermore, the utilization of software components enables the development of open interfaces, thereby enabling CSPs to deploy infrastructure that is not reliant on a single vendor, reducing costs and opening the way to actual fully-managed heterogeneous networks.

2.1.4 The Open RAN concept

The flexibility introduced by 5G has facilitated the development of novel network paradigms that leverage software-defined approaches, which previously were featured by proprietary monolithic black boxes. Although these new perspectives promise to establish the foundations for network infrastructures that serve as applications, this paradigm necessitates intelligence at every level and adequate coordination to manage the anticipated open interfaces and multi-vendor solutions.

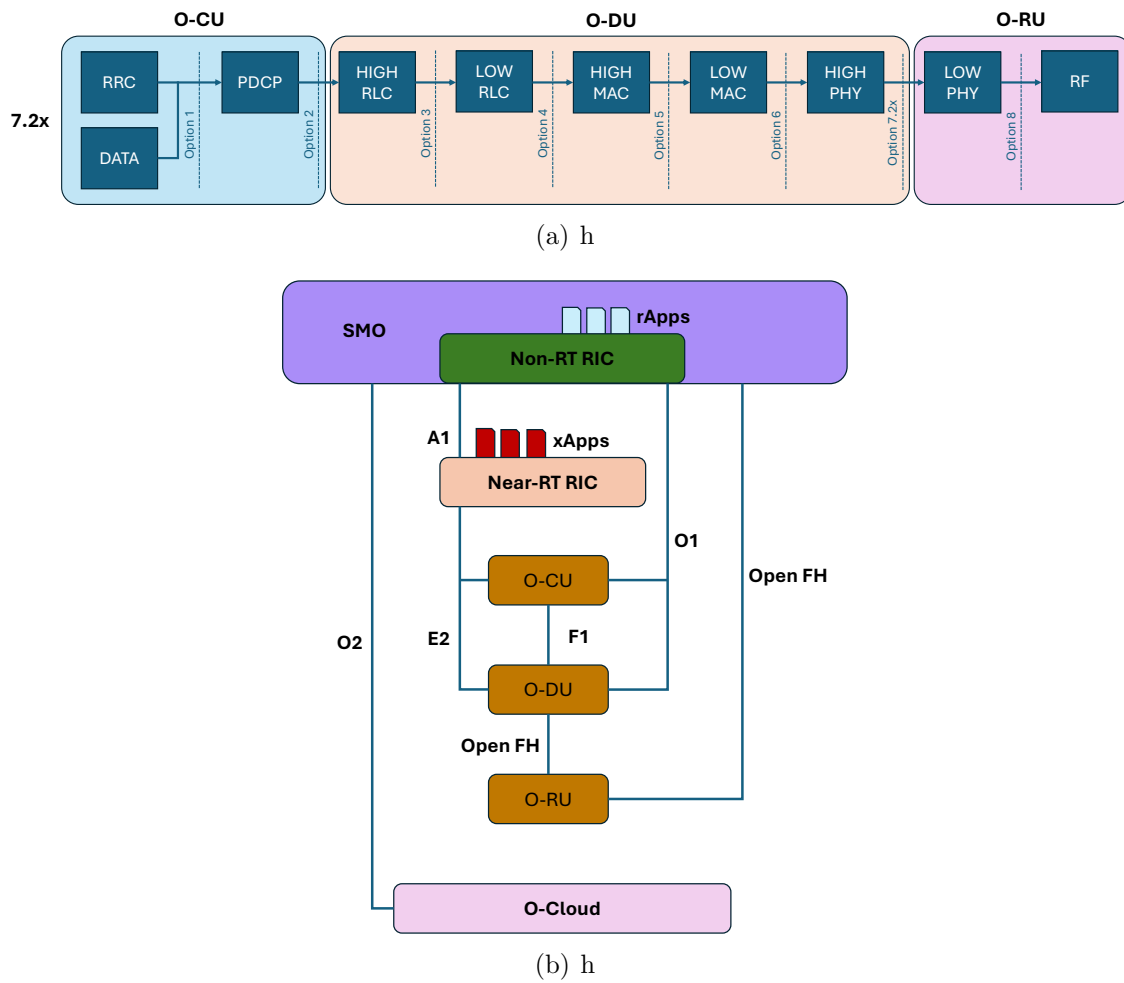


Figure 2.9: O-RAN architectural vision [24]

In response to this trend, the O-RAN Alliance [79] has been established with the objective of spearheading the industry and research community towards a unified criterion. The objective of this organization is to facilitate the development and maturation of the Open RAN architecture through the specification of disaggregated, interoperable, and open-interface-driven virtualized Radio Access Network (vRAN) over Commercial-off-the-shelf (COTS) [80] hardware. In

addition, O-RAN has been working on the definition of use cases, requirements, and verticals [81, 24]. It is crucial to recognize that the Open RAN architecture represents an evolution of the 3GPP specifications, with the objective of advancing the development of the RAN.

The Open RAN architecture aims to aggregate intelligence at various levels through the integration of AI techniques and the exploitation of the RAN disaggregation concept as defined by 3GPP. To address this issue, the O-RAN Alliance employs a Cloud-native Network Functions (CNFs) architecture [82] comprising the Open Radio Unit (O-RU), the Open Distributed Unit (O-DU), and the Open Central Unit (O-CU), in conjunction with innovative architectural entities that facilitate the integration of intelligence into the system, those are the Service Management Orchestration (SMO) and the Radio Intelligent Controller (RIC) [83].

On the one side, the SMO is responsible for the orchestration and management of the whole RAN. This includes handling the lifecycle of the NFs, policy enforcement, and RAN optimization, including configuration and fault management, performance monitoring, and data analytics. Along these lines, the SMO interacts with the diverse RAN elements through the use of interfaces, i) O1 for management, ii) O2 for communications with the O-Cloud, and iii) A1 for policy-based control.

The term O-Cloud refers to a set of hardware and software components that provide cloud computing capabilities to execute the functions related to the RAN, enabling the realization of various network services. It is also important to note that the cloud software is decoupled from the cloud hardware. This means that both the cloud hardware and the software can typically be sourced from different vendors.

On the other side, the RIC is a key component that enables intelligent network optimization and automation (e.g., by leveraging AI). The RIC is a softwarized network service that acts in two modes, Non-Real-Time (Non-RT) RIC and Near-Real-Time (Near-RT) RIC.

The Non-RT RIC component is responsible for the management of non-time-sensitive tasks, defined as those requiring more than one second, and is situated within the SMO. This component engages in interaction with the Near-RT RIC component through the A1 interface, and RAN entities are linked through the O1. In order to incorporate intelligence into its structure, the Non-RT RIC employs RAN Applications (rApps), which are case-specific, modular solutions that describe

policies or strategies that can be leveraged for decision-making, automation, or optimization. These applications interact with the RIC through the use of a standard APIs defined by the O-RAN specifications. For example, an AI-based model can be trained in the Non-RT RIC concerning the analytics provided by the application.

The Non-RT RIC manages non-time-sensitive tasks (i.e., more than 1 second) and it is located in the SMO. This component interacts with the Near-RT RIC by the A1 interface and with RAN entities through the O1. To aggregate intelligence to its structure, the Non-RT RIC utilizes rApps, which are case-specific, modular solutions that describe policies or strategies that can be leveraged for decision-making, automation, or optimization. These applications interact with the RIC through the use of standard APIs defined by the O-RAN specifications.

Conversely, the Near-RT RIC is responsible for time-sensitive functions within a temporal range of 10 milliseconds to one second. In order to address the latency restrictions, the O-RAN recommends the deployment of this element in close proximity to the user domain (e.g., CSP edge/cloud). Similarly, the Near-RT RIC interacts with the network elements via the E2 interface. The functions associated with this entity are near-real-time resource optimization (i.e., spectrum, load balancing) and eXtended Application (xApp) interfacing. The latter are applications that provide control services for the RAN that can be tailored to suit specific requirements.

One illustrative example of the workflow of these entities is the use of AI/ML approaches. In this manner, the data collection and processing, as well as the training of the ML model, which are time-consuming tasks, can be performed in the Non-RT RIC, in regard to the analytics provided by a rApp. Subsequently, the model exploitation is featured in the Near-RT RIC domain based on the functionalities provided by the xApp.

In addition to the aforementioned architectural entities, the Open RAN paradigm is founded upon the 7.2x functional splitting scheme, which separates the PHY layer into high and low sublayers. The 7.2x functional split is designed to assign the majority of time-sensitive tasks to the O-RU for both DL and UL directions, while the more processing-intensive functionalities are handled by the O-DU. This scheme has been defined as performance-balanced and cost-effective due to the generic and simpler O-RU. A graphical representation of the O-RAN architectural vision can be found in Figure 2.9.

The O-RAN architectural framework has been the subject of ongoing development with the objective of extending its scope to encompass both 4G and Beyond 5G (B5G) next-generation networks. In this regard, the development and design of Distributed Applications (dApps) [84] are still in their nascent stages. These are intended to provide real-time intelligence (less than 10 milliseconds) to the O-DU and O-CU. However, this approach is still under study and standardization in the O-RAN work groups [85].

2.2 5G New Radio (NR)

In parallel with the architectural evolution of 5G, which concerns the 5GC approach, and the emergence of new use cases (eMBB, URLLC and mMTC), the NG-RAN technology, designated NR, represents a significant advancement in comparison to legacy RATs. Enhancements include substantial improvements in throughput, capacity, latency, and reliability aimed to support the upcoming new-generation services. The delineation of 5G RAT specifications and structural elements commenced with TS 38.300 [74], and subsequently manifested in TS 38.912 [86].

NR utilizes a diversity of technology enablers that provide more flexibility, and scalability that increase the efficiency of the resource usage and thereby, increasing the system's capacity. Among these features, are involved the use of new frequency bands tailored for different service requirements, advanced waveforms, numerologies, and modulation schemes, massive MIMO and beamforming, flexible resource scheduling, network slicing, and compatibility with LTE through strategies like Evolved Non-standalone Dual Connectivity (EN-DC).

2.2.1 NR frequency bands

The NR technology is specified to be used in different frequency bands, including two new ranges identified as FR1 and FR2. According to the Release-15 3GPP-TR 21.915 document [69], NR can be deployed in four frequency bands regarding:

- LTE “refarming” band, which have correspondence with the current LTE ones.
- FR1. New exclusive NR frequency bands that do not have correspondence with LTE ones.

- FR2. New exclusive NR frequency bands.
- Supplemental Downlink (SDL) / Supplemental Uplink (SUL) band: uplink/downlink only frequencies intended to complement NR bands.

The frequency bands in NR have been selected with the objective of fostering the plethora of services and applications that belong to the 5G use cases. In this regard, each band is distinguished by particular attributes pertaining to resource constraints, transmission power, propagation, and so forth. 5G incorporates these considerations by integrating low and mid-bands into the FR1 frequency range and high-bands into FR2.

The low-band, associated with frequencies less than 1 GHz, is typically preferred by the CSP due to its ability to provide wide coverage through penetration capabilities. However, it is not optimal for high throughput. This situation encourages the utilization of low-band for long-distance or dense scenarios, such as those encountered in mMTC services and applications.

The mid-band, sub-6 GHz, comprising frequencies between 1 and 2.6 GHz and 3.5 to 7.125 GHz [87], is regarded as the optimal choice for 5G implementation, given its capacity and coverage capabilities that are well-suited to initially meeting the requirements of 5G for eMBB and URLLC. In this regard, the 5G standard has extended the mid-band usage to higher frequencies than those utilized by LTE (i.e., 4 GHz) in order to support contiguous larger bandwidths that increase spectrum efficiency by reducing the overhead product of legacy LTE carrier aggregation schemes. In this fashion, FR1 can reach up to 100 MHz of bandwidth.

In contrast, the high-band (greater than 6 GHz) [87] is conducive to the advancement of throughput, latency, and capacity features. Furthermore, this range has the potential to significantly expand the spectrum available for network usage, in contrast to the previous bands where resource availability is constrained or allocating contiguous blocks of resources is not feasible. The primary challenge associated with this band is the difficulty in propagating radio waves. The waveforms utilized in this band necessitate a direct line of sight between the node and the terminal, as well as an increased transmission power. This renders the band suitable for short-distance scenarios with stringent requirements for high data rates (multi-Gbps) and delivery time promised to be met by 5G.

Moreover, the 3GPP standardizes two sub-bands for FR2, FR2-1 (24250 MHz

– 52600 MHz) supporting up to 400 MHz of bandwidth, and FR2-2 (52600 MHz – 71000 MHz) up to 2000 MHz [88]. An illustration of NR can be found in Figure 2.10.

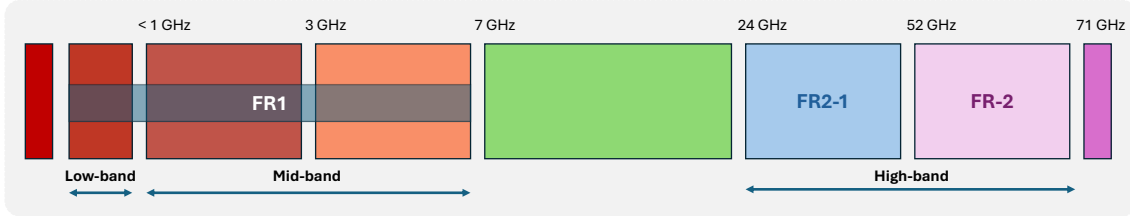


Figure 2.10: Frequency bands in NR

2.2.2 NR Waveforms

5G employs an advanced version of the OFDM technology that was previously utilized in LTE. OFDM is a modulation scheme that divides the available frequency spectrum into narrow subcarriers. The distinguishing feature of this approach is that each subcarrier is orthogonal, meaning that they are independent of one another. As a consequence of this, the complexity of the receivers can be reduced and the information extracted with greater ease.

In order to maintain this condition, the waveforms are spaced in such a way that the peak of one subcarrier's waveform coincides with the zero-crossing points of all the other subcarriers. This advantage allows for the minimization of interference between subcarriers and the enhancement of communication robustness. Similarly, OFDM enhances the spectral efficiency of 5G networks by efficiently packing subcarriers that convey information in parallel, while avoiding interference. Figure 2.11 illustrates the concept of OFDM subcarriers.

Furthermore, OFDM addresses propagation issues. On the one hand, given that certain carriers may be susceptible to frequency-selective channel effects, NR technique disseminates the data across multiple subcarriers, thereby circumventing the transmission of information in a single channel. Similarly, given that orthogonality may be compromised as a result of multipath propagation, NR aggregates Cyclic Prefix (CP) in order to mitigate the effects of Inter-Symbol Interference (ISI). A CP is a portion of a symbol that is replicated at the outset, thereby creating a guard interval that allows the receiver to differentiate overlapped symbols.

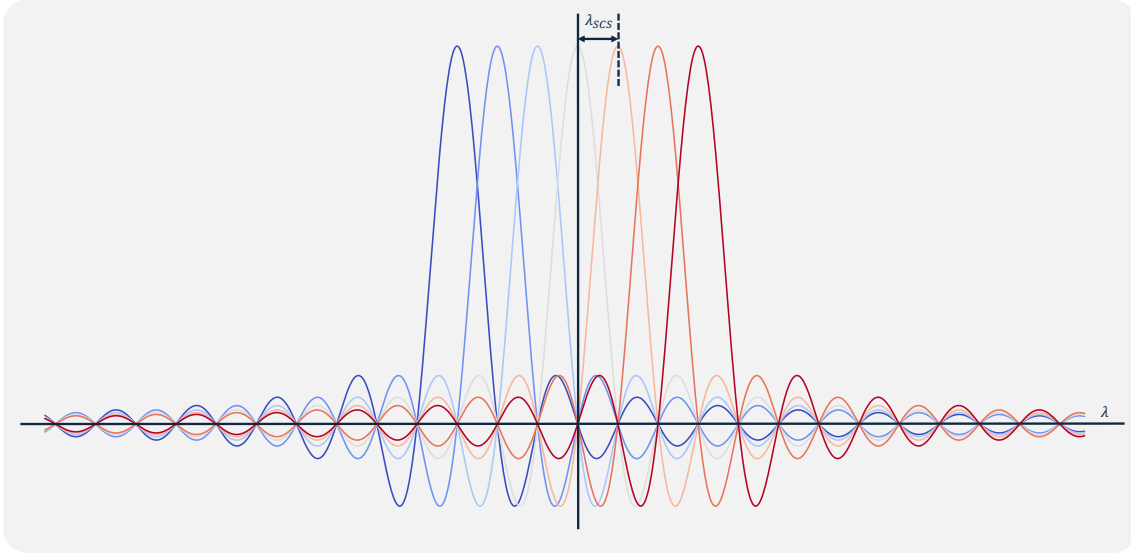


Figure 2.11: Orthogonality in OFDM subcarriers

2.2.3 NR frame structure and numerologies

5G takes into account the wide variety of users that should be served by the mobile network at the same time. To address this, the standard presents some enhancements compared to LTE with flexible frame structures in terms of numerology and resource allocation strategies to meet the requirements of the upcoming 5G use cases.

In the first place, NR leverages the concept of OFDM to serve multiple users simultaneously through Orthogonal Frequency Division Multiple Access (OFDMA). Although LTE implemented OFDMA in its standards, it was only considered in DL since in UL was considered Single-Carrier Frequency Division Multiple Access (SC-FDMA) to approach the Peak-to-Average Power Ratio (PAPR) issue. 5G exploits this technology by applying some modifications that make it feasible to deploy this strategy in both links.

Figure 2.12 illustrates the distinctions between the two methodologies. The OFDM strategy entails the sequential assignment of an entire band, along with its associated subcarriers, to a user in the time domain for the transmission of symbols. Nevertheless, the 5G use cases encompass a multitude of services with disparate requirements pertaining to time constraints, data rates, and other factors. To address this variable scenario, OFDMA offers the possibility of handling multiple users by allocating resources in both the time and frequency domains simultaneously, thereby increasing the capacity of the network while optimizing resource usage.

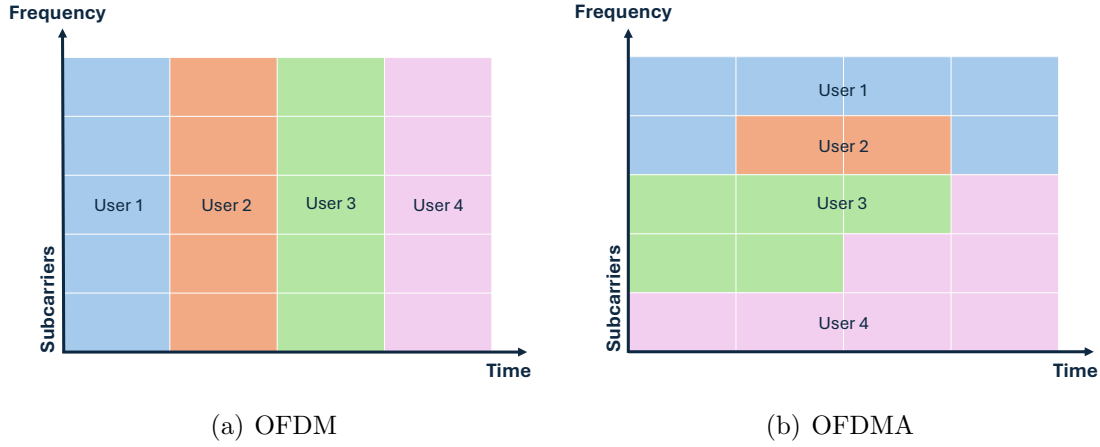


Figure 2.12: Resource allocation strategy

The flexibility introduced by OFDMA can be exploited in conjunction with the flexible numerologies concept introduced in this technology. The term “numerology” is used to describe the management of subcarrier spacing, symbol duration, and other features to handle frequency spectrum and time resources.

The subcarrier spacing Δ_{SCS} is defined as the separation between adjacent subcarriers in the OFDM scheme. 5G has defined multiple subcarrier spacing options defined as $2^u * 15kHz$, where the fixed number is the only spacing used in LTE. The coefficient u indicates the numerology number employed, with a range of 0 to 6. Similarly, the symbol duration is contingent upon the carrier spacing and is defined as the reciprocal of the aforementioned value ($1/\Delta_{SCS}$). Consequently, an increase in subcarrier spacing results in a reduction in symbol time. This allows a robust driver to manage resources for services and applications with diverse requirements, which was previously unfeasible with the fixed LTE carrier spacing (15 kHz).

Beyond the frequency, the organization of the resources in the time domain has also been modified owing to the numerology. In this sense, the frame is the top-level respect to time structure for 5G transmission. The frame establishes the structure of the resources within a fixed window of 10 milliseconds. Each frame can be subdivided into subframes of duration of 1 millisecond. In practical terms, this remains similar to the LTE frame and subframe concept, however, the great differences are handled by the variability in the number of slots per subframe.

This flexibility owes to the multiple subcarrier spacing values that have modified the concept of slot duration in 5G. In this technology, the slot is the minimum unit

of time scheduling and is conformed by 14 OFDM symbols or 12 symbols when an extended cyclic prefix is used. In this regard, the number of slots in a subframe is equal to 2^u . This marks an important difference with LTE where the slot duration is a fixed value of 0.5 milliseconds. The 3GPP in document TS 38.211 [25] defines the time scheme for 5G NR based on the numerology (see Table 2.2).

Table 2.2: 5G NR time scheme per numerology [25]

Numerology (μ)	Symbols per slot	Slots per subframe	Slots per frame	Subcarrier spacing ($2^u * 15$ kHz)	Cyclic prefix
0	14	1	10	15	Normal
1	14	2	20	30	Normal
2	14/12	4	40	60	Normal/Extended
3	14	8	80	120	Normal
4	14	16	160	240	Normal
5	14	32	320	480	Normal
6	14	64	640	960	Normal

The 5G's flexibility introduced in both frequency and time domains drives the opportunity to perform advanced strategies for resource management. In this sense, the physical resources in 5G can be realized as a set of subcarriers in the frequency domain and OFDM symbols in time conforming to a grid. Hence, an OFDM symbol modulated in a single subcarrier is defined as a Resource Element (RE). Then, the set of twelve (12) consecutive arranged REs in frequency is called a Resource Block (RB). The definition of RB in 5G is only attached to the frequency domain since the number of symbols assigned to a slot depends on the numerology.

Figure 2.13 displays the structure of the 5G resources grid for numerology 1 (30 kHz).

One of the main differences between 5G and LTE scheduling is that NR addresses this by adjusting the number of symbols per slot. This approach can be addressed using the Start and Length Indicator Value (SLIV) that allows 5G to enhance service-tailored resource allocation strategies. This concept enables the concept of “mini slots”, which breaks the slot structure into a more flexible scheme by using only a part of the RB time resources. This is in principle designed for prioritized, time-sensitive services that should be served with minimum latency. In addition to the latter technology enabler, 5G introduces the concept of BandWidth Part (BWP) [25]. The BWP can be understood as a subset of contiguous active RBs, belonging to the total carrier bandwidth, named Common Resource Block (CRB), that is dynamically allocated to a user or device [89].

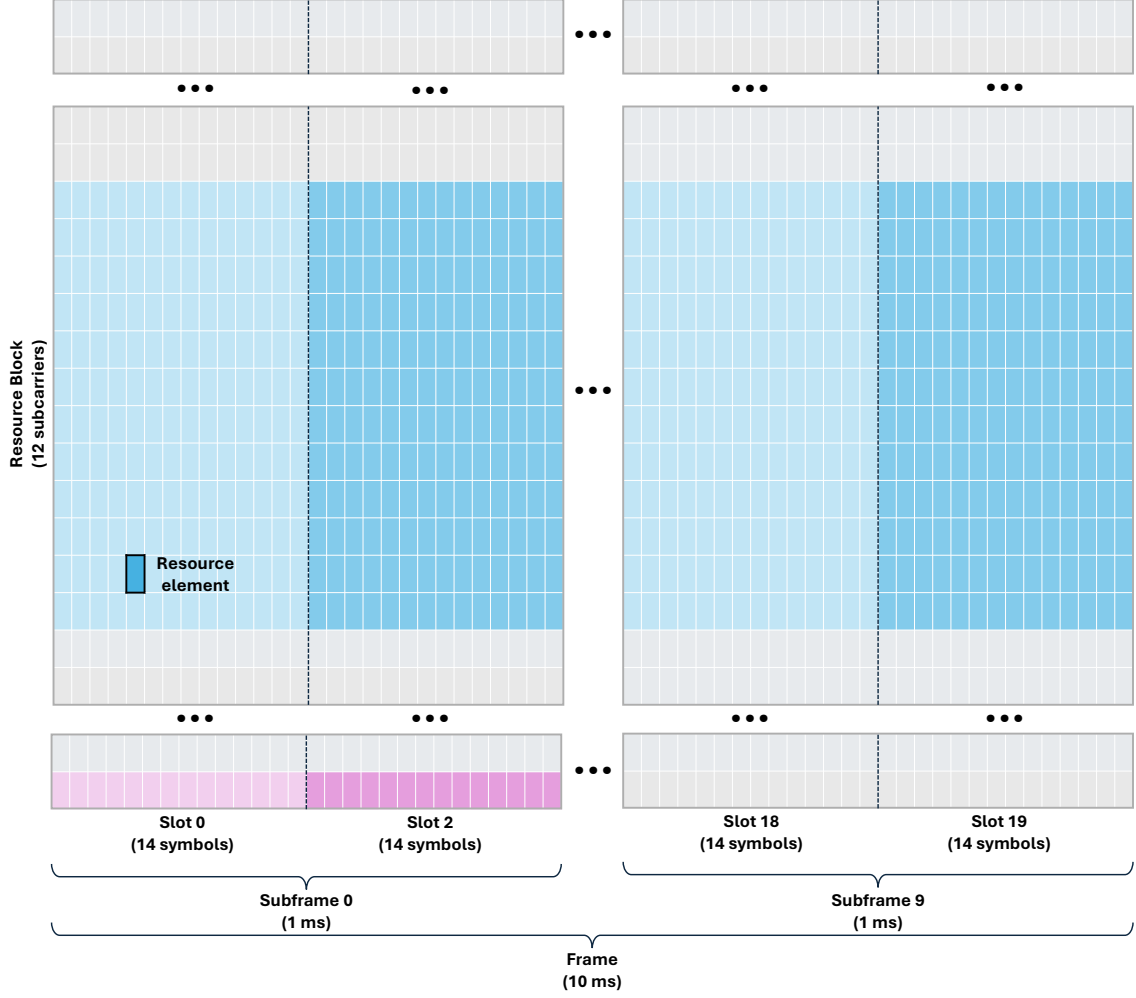


Figure 2.13: NR resource grid for numerology 1 (30 kHz)

The main benefit of this concept is that BWP allows the partitioning of the bandwidth into several subsets supporting various numerologies. The network can allocate up to four BWP, however, only one can be active at a time. Nonetheless, it is possible to switch from/to another BWP through mechanisms defined in the MAC [90]. In addition, this approach has implications for energy saving and efficiency, since the resource-allocated terminal only monitors a portion of the bandwidth instead of scanning the whole CRB.

The abovementioned technology and its enablers open a new way to provide services with high quality that can fulfill the users' expectations. This can be performed in different dimensions, such as modifying the numerology to provide different symbol durations in the function of the service requirements. Similarly, 5G mixes these advantages in service-tailored resource allocation through the BWP. Along the same lines, 5G provides the opportunity to utilize different

frequency bands that can exploit these drivers. Finally, 5G leverages OFDM to mitigate the negative effects of the channel by aggregating CP, and orthogonal subcarriers in both directions of the communication.

2.3 Machine Learning

ML comprehends a set of techniques belonging to the concept of AI that enables systems to automatically learn from data, and capture behaviors that are not explicitly described. ML has become one of the hottest topics in the research field since its applications can be utilized to aggregate added-value features that in the past were developed manually.

The key approach of ML is to leverage algorithms that can process information, generally heterogeneous, to detect patterns, extract meaningful insights, and model behaviors that can be used to make predictions or ease decision-making. In this sense, ML learns from the experience captured in the form of data gathered by means of observations to predict new ones [91]. Therefore, the leveraging of ML offers diverse advantages in terms of performance, development time, and costs, collectively facilitating the acceleration of data-driven processes. This, in turn, relieves the necessity of human expert knowledge, which, in certain circumstances, can have a negative impact on the scalability and flexibility of the data owners.

Conceptually, ML employs a technical lexicon to delineate the heterogeneous elements that compose a model. The inputs, also referred to as features, correspond to a single quantifiable value or characteristic within a data sample. Consequently, the data sample is constituted of a set of multiple features, which are related to one another in a dimension (e.g., time). In contrast, the outputs, or targets/labels, are the values associated with a given data sample. A dataset is an ensemble of a variety of data samples (which may include their respective target values) that share a common feature structure.

In ML, the relationship between the features and the targets is defined through a model that captures their connection by learning on the dataset. These models may be constructed using mathematical structures, probabilities, or rules defined by algorithms. The process of inferring a model based on the data is referred to as training, whereas its application to new data samples that were not included in the training set is termed testing. The dataset employed in the model construction

is designated as the training set, whereas the evaluation is conducted through the testing set.

In some approaches, a validation set is extracted from the training set to perform model evaluation during the training process. This is an effective method for hyperparameter tuning (adjustment of internal algorithm parameters) and generalization assessment, which helps to avoid overfitting. Here, the hyperparameters are internal parameters that the model adjusts during the training process and which express the relationships between the input features. Moreover, overfitting is the situation where an ML model demonstrates strong performance with the training set but is unable to predict adequate outputs on new data.

ML is subdivided into different categories:

- **Supervised Learning.** In this paradigm, the algorithm is trained on a labeled dataset. This means that the training samples are paired with a target. In this situation, the ML model maps the features into the output by using the relationships expressed by the model. According to the type of approached problem, this technique can be subcategorized into:
 - **Regression.** This problem exhibits a numerical continuous output.
 - **Classification.** It performs the prediction of a discrete output, which is referred to as a “class”. If the target disposition is a unique class, the problem is identified as binary classification. Conversely, if the prediction can output multiple labels, the problem is considered to be a multi-class classification problem.
- **Unsupervised Learning.** The model is trained on unlabeled data, hence, the algorithm must identify and capture patterns between the features. The main applications of this type of problem are data exploration, insights information extraction, and data organization. Depending on the usage this ML approach can be split into:
 - **Clustering.** This problem associates data samples concerning metrics or relationships into groups named clusters.
 - **Dimensional reduction.** In this category, the algorithm synthesizes new features based on the information contained in the input features.

- **Reinforcement Learning.** It is a ML technique that mimics the human learning process supported by a trial-and-error strategy. This approach comprehends two knowledge stages, exploration and exploitation. The aim is to maximize the cumulative reward over time throughout episodes where the states of the agent, who interacts with an environment, are observed to take determined actions that can increase or decrease the reward. This learning approach differs from the supervised one since it does not need labeled data [91].

As time has passed, the aforementioned ML paradigms have undergone evolution, resulting in the emergence of new approaches capable of exploring and extracting larger amounts of data (i.e., big data). These approaches are classified as belonging to the Deep Learning category, which encompasses algorithms that are highly computationally intensive. However, their use is effective for complex problems where simpler ML techniques are not performant. Furthermore, the combination of supervised and unsupervised techniques has paved the way for hybrid approaches, including self-supervised learning and semi-supervised learning.

2.4 Network management

The concept of network management has undergone a continuous process of evolution across the various generations of mobile networks. This approach encompasses the full range of processes, techniques, and tools utilized to oversee, regulate, and enhance the infrastructure, and consequently, the services offered by networks. The primary objective is to ensure that the network functions are operating effectively, with optimal reliability, availability, and performance, while facilitating uninterrupted communication for users. From this perspective, effective network management is of paramount importance for the maintenance of service quality, the management of network resources, and the adaptation to the increasing demands of mobile communications, particularly in the context of 4G and next-generation networks.

With regard to this matter, the preliminary approach to network management was devised with the objective of incorporating a certain degree of automation in the context of LTE. The concept of Self-Organizing Networks (SON) was first introduced by the NGMN Alliance [31], and subsequently developed by the 3GPP

[32], was devised with the objective of automating network functions that require human intervention. In essence, the SON technology was designed to address configuration, optimization, and issue mitigation functions, with the objective of reducing CAPEX and OPEX while maximizing network performance in scenarios that are more complex and diverse. In correspondence, the functions defined by the 3GPP as part of SON are the following:

- **Self-configuration.** Designed to provide plug-and-play functionalities to let the network adapt and tune itself according to the deployment environment.
- **Self-optimization.** Covers a set of functions that monitor the network performance and proactively take action to adjust parameters that enhance the RAN capacity and coverage.
- **Self-healing.** Involves functionalities that aim to minimize the service degradation effects caused by network failures caused by internal as well as external factors.

One of the primary constraints associated with the 4G SON is that the functionalities may be dependent on the vendor, which consequently restricts the number of solutions or use cases that can be deployed in the equipment. This is due to the reliance on predefined algorithms and mechanisms that address particular issues. The advent of 5G technology has brought with it the need for more sophisticated mechanisms to accommodate novel use cases, presenting both challenges and opportunities. Therefore, the diversity of services and applications, in conjunction with the deployment of new architectures of heterogeneous networks, gives rise to novel scenarios in which data from disparate sources can be harnessed. This conceptual driver was taken into account during the development of NG-SON, with the aim of endowing networks with intelligence rather than merely limited automation.

In this regard, the Zero-touch network and Service Management (ZSM) [92] paradigm accounts for the need for integral frameworks capable of orchestrating and managing network resources while ensuring adequate service provision to users. From this idea, the main standardization bodies have considered the use AI/ML enablers to exploit the crescent amount of data to manage a wide portfolio of functionalities, such as network automatic configuration, optimization, anomaly

detection, traffic pattern detection, network insight extraction, among others, without the intervention of human operators [93].

Moreover, the utilization of data facilitates a more comprehensive approach to network management, progressing from a network-centric concept to an E2E holistic perspective that capitalizes on the technological enabler of 5G (e.g., NFV, ns). The utilization of data-driven management paradigms expedites the development of network standards, whilst simultaneously furnishing them with intelligent tools that facilitate scalability, flexibility, and adaptability in the context of ever-evolving and challenging environments [94]. Similarly, the implementation of this methodology facilitates the development of management strategies that are tailored to specific services, thereby enhancing the capabilities of the networks and enabling the deployment of advanced applications.

2.4.1 Service-oriented management

The advancement and appearance of new, advanced services have revolutionized the network management paradigm. In principle, the classic legacy services provision management such as voice, SMS, and internet access required simpler strategies based on a few indicators. However, the capabilities, features, and functionalities introduced by 5G have motivated industry, equipment vendors, ASPs to look for integrating and delivering advanced new-generation services (e.g., immersive communications, IIoT, and telemedicine, among diverse). In this sense, the flexibility introduced by 5G technology enablers has prepared the way to introduce modern approaches aiming to provide high-quality experiences for a diversity of upcoming services.

In light of the aforementioned considerations, it becomes evident that a network management approach oriented towards the delivery of superior quality services, subject to comprehensive monitoring, is a necessity. This implies the recognition of all elements involved in the communication chain, from the ASPs to the customers. Figure 2.14 depicts the key aspects that will be discussed in this section.

To address these paradigmatic changes, the network management based on traditional QoS approaches is insufficient to handle the large portfolio of services with heterogeneous requirements and expectations. Conceptually, the QoS is a description or quantitative measure of the overall communications system performance at network and service domains [95, 96]. Although the terminology

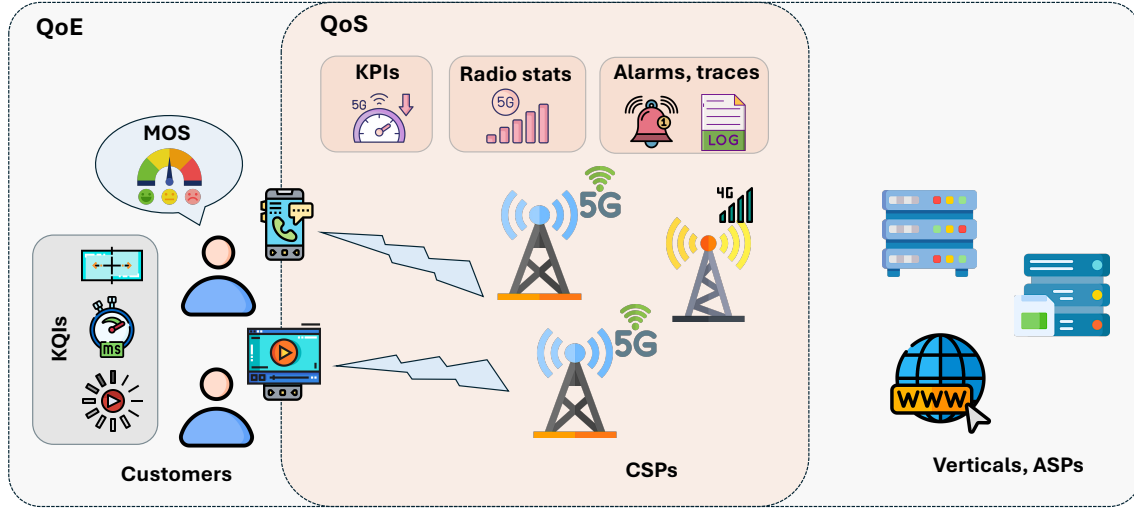


Figure 2.14: Management scope of QoE and QoS

includes the service domain as part of its assessment, the utilized data sources are generally network-centric (i.e., information from Network Elements (NEs)) and do not necessarily consider the customer's point of view. Examples of information contained from different network sources are described in Table 2.3.

Table 2.3: Typical network sources of information

Category	Description	Example
Configuration Management (CM) parameters	Parametric values that reflect the configuration of the network elements	Transmission power
Status monitoring	Periodical/On-demand checking reports that indicate the connectivity status of the network elements	Heartbeat signals
Alarms		eNB down
Counters	Statistics that reflect the number of events related to a category in a period of time	RRC Connection requests
KPIs	Indicators estimated from the combination of counters that reflect network-wide performance	Throughput
Traces	On-demand UE/network element monitoring a time window	Voice session performance
Minimization of Drive Tests (MDT)	Network information collected through dedicated tests oriented to mobility	Coverage measurements in an area
Context information	Provides additional information regarding situational events that can affect network usage	User density in a zone

It is therefore necessary to collect information that encompasses a comprehensive view of the communication chain, including both user and service provider segments, in order to advance the paradigm of QoE. The 3GPP has defined this latter concept as the measurement of the “*degree of delight or annoyance of the user of an application or service*” [97]. In the past, this term has been treated as a subjective assessment of the user's satisfaction. In this regard, a

number of approaches have been defined with the objective of quantifying QoE. Initial approaches consisted of performing a service assessment by collecting the user perception of the impairments employing numerical scales. One of the most frequently employed strategies is the assessment conducted through the MOS metric [98, 99, 100]. For an illustrated example of the MOS scale, please refer to Table 2.4. The key issue with MOS-based assessment is that this metric can aggregate user bias that may lead to erroneous performance evaluation.

Table 2.4: Mean Opinion Score evaluation format

MOS	Quality	Impairment
1	Excellent	Imperceptible
2	Good	Perceptible, but not annoying
3	Fair	Slightly annoying
4	Poor	Annoying
5	Bad	Very annoying

In order to extend the MOS strategy to the variety of new services being delivered by mobile networks, standardization bodies have begun to rely on parametric models that synthesize the user experience from service-specific indicators. This approach was, to some extent, designed to streamline the operator's data collection process and automate certain functionalities. Nevertheless, the necessity for controlled experimentation to obtain data for the models persists. The procedures for model creation, data collection, and user assessment have been delineated with the objective of defining common frameworks among the infrastructure providers, as evidenced by the specifications set forth in [101, 102].

This concept has been widely utilized for conventional services such as voice, file transfer, and video-on-demand. However, with the advent of next-generation networks, the proliferation of advanced services has led to the emergence of numerous models, rendering this strategy obsolete. In order to address this challenge, the mobile network community is moving towards the utilization of objective metrics that reflect the actual service performance from a user-centric perspective, which is defined as KQIs. This approach considers the E2E context, which encompasses the interactions between customers, ASPs, and CSPs. This approach, supported by technology enablers including ML [46, 103, 104, 105, 106], has facilitated the evolution of legacy network management towards scalable and flexible approaches that can adapt to cutting-edge trends.

The KQIs are conceptualized as objective metrics that reflect the performance of the service in consideration of the E2E aspects, as outlined in [97]. In this manner, the KQIs are of interest to various sectors of the communications market, as they facilitate the management of services from the user, ASP, and network perspectives. Therefore, the KQIs can be regarded as a valuable addition to the KPI employed in legacy networks, with the additional benefit of providing a comprehensive understanding of user satisfaction. Moreover, these metrics address the subjectivity introduced by MOS-based techniques, as their values are not dependent on user perception, but rather on an objective estimation. This renders these indicators as powerful drivers, as they can support service-tailored management that aligns with the expectations and objectives of next-generation networks.

2.5 Extended Reality background

Extended Reality (XR) represents a technological paradigm that seeks to integrate disparate real and virtual elements and environments in a unified, immersive experience. This scope encompasses Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR). The degree of interaction with each reality is the determining factor in the selection of the appropriate technology. In this sense, these technologies can benefit from the mobility that wireless networks offer in order to evolve the immersion feeling into real-like experiences. Likewise, the potential that mobile networks offer through their technological enablers, such as the Edge/Cloud computing approaches can ease and accelerate the development of the concept of thin devices. This approach focuses on offloading computational costs at the user devices, so it is possible to deliver high-quality services in more accessible equipment. However, this concept requires a transport network prepared to provide stringent requirements.

On the one hand, MR and AR are distinct categories that superimpose virtual elements onto the real environment, creating a hybrid space that is partially virtual and partially real. MR technology merges elements of the physical world with VR, as illustrated by the incorporation of the player silhouette into a virtual environment. In contrast, AR superimposes some virtual elements and information on top of the physical reality. In contrast, MR is considered an evolution of AR that provides a comprehensive immersive experience.

In the context of AR and MR, the computational capacity is essential for calculating three-dimensional models that facilitate the determination of the optimal position for overlay elements, ensuring that they do not impinge upon the user's field of view or disrupt their sensory experience. In certain instances, the transfer of video may be required to offload local computational resources and conduct model calculations, element recognition, and other processes on a remote server. However, these actions may place undue stress on uplink communications.

The 3GPP has performed theoretical advancement in these both concepts [107]. In terms of latency, motion-to-photon latency of up to 20 ms is expected in scenarios with video, and for audio feedback up to 50 ms, comprehending an E2E latency of about 50-60 ms. Concerning visual quality, Field of View (FOV) of 30 by 20 degrees is the minimum required with image resolutions per eye of 1.5k by 1k for acceptable experiences. The FOV is the area of vision at a given moment.

Conversely, VR requires a significantly greater computational load to generate comprehensive virtual environments and to create a genuinely immersive experience. Typically, VR services necessitate the downloading or streaming of multimedia content to a “VR-ready” device, which is an apparatus prepared for VR rendering (two screens being rendered simultaneously). As stated in [108], a “real-feel” resolution could be achieved with approximately 60 pixels per degree and a 120 Hz frame rate, based on human eye perception at approximately 120 degrees of FOV. In order to achieve these conditions, an approximate bitrate of 1 Gbps is necessary for a theoretical per-eye 9K resolution. Nevertheless, this is not currently attainable with existing technology. For example, a 360° 4K camera streaming at 30 frames per second may generate up to 300 megabits per second of data [109]. Likewise, the 360-video service requires approximately eight times more content than the traditional version (4K per eye in comparison to the typical 1080p ratio) [110]. This supports the conclusion that this service requires adequate network resources.

In terms of latency, the literature suggests the necessity of reduced delays to ensure media fluidity and prevent virtual and motion sickness. In this regard, a RAN latency of 20 ms is recommended for this kind of service [111], however, it is expected to improve this value up to 10 ms. Similarly to the AR case, the 3GPP in [112] remarks that the E2E latency should lie between 50 and 60 ms for non-optimal but minimum quality sessions.

Although these requirements are expected to ensure high-quality XR deployment, the capabilities of contemporary mobile networks are insufficient to meet them. The typical loopback latency of LTE is approximately 25 ms, which is inadequate for the specified requirements. It is anticipated that this will be significantly enhanced with the advent of 5G, which is expected to achieve latencies of up to 10 ms for URLLC use cases. It is important to note that the XR cases additionally require suitable data rates, which place these use cases as a mixture of resources that should be managed dynamically.

This illustrates the necessity for sophisticated mechanisms that transcend static policies and that should capitalize on 5G drivers to attain the requisite levels of quality anticipated by all parties involved in the implementation of this type of service. In light of these considerations, this thesis will address some of these needs by exploring these services, determining the network impact over them, and establishing mechanisms that will support their deployment utilizing next-generation networks.

2.6 Conclusions of the chapter

In summary, an overview of the key technologies and concepts behind today's mobile networks has been provided in this chapter. The architectural concepts of LTE and 5G have been covered, including how 5G's new architecture and the Open RAN concept enable more flexible and powerful networks. The enhancements in flexibility and efficiency offered by 5G New Radio have been highlighted, demonstrating its capability to support a wide range of new applications. Furthermore, the increasing complexity of these networks is being addressed by the growing importance of ML, for which key concepts have been presented. Essential ideas regarding Network Management have also been introduced, along with the evolution toward service-based approaches. Finally, Extended Reality (XR) technology, architecture, and requirements have been introduced, emphasizing how their deployment will push network capabilities to new limits.

These foundational concepts will serve as the basis for the next chapters, where the focus will shift to how new techniques can address the demands of advanced use cases, ensuring that next-generation networks are prepared to meet future needs.

Part II

Publications



UNIVERSIDAD
DE MÁLAGA

Chapter 3

Research outline

Content

3.1	Research Methodology	62
3.2	Description of the publications	64
3.2.1	C1. Measuring key quality indicators in cloud gaming: Framework and assessment over wireless networks	65
3.2.2	C2. KQI assessment of VR services: A case study on 360- video over 4G and 5G	66
3.2.3	C3. Cloud VR on 5G: A Performance Validation in Industrial Scenarios	67
3.2.4	C4. ML-powered KQI estimation for XR services: A case study on 360-video	69
3.2.5	C5. Towards KQI/ML-based management: Optimizing radio resource allocation for next-generation networks . . .	71
3.3	Conclusions of the chapter	72

This chapter presents a description of the research work in two subsections. First, the research methodology applied through the diverse stages of the thesis is described. Then, a holistic overview of the work is presented by mapping the challenges, the objectives arising from them, and the achieved results.

3.1 Research Methodology

The performed research follows a methodology subdivided into multiple steps that contribute to the following. Figure 3.1 depicts the logical process fostered by the thesis:



Figure 3.1: Research methodology

1. Research problem formulation.

The initial stage of the research process involves identifying and defining the mobile network problem that will be addressed by the subsequent studies. In this regard, this task entails a comprehensive search for research gaps in the field. Moreover, to guarantee that the research outcomes are oriented towards the resolution of tangible industry issues, collaboration with stakeholders (i.e., CSPs, vertical or public organizations) is imperative through research projects with the research group's participation. This joint endeavor enables the work to address present and future challenges in alignment with the expectations and trends of the various parties engaged in the development of next-generation networks.

2. SoTA review.

Once the problem has been identified and its respective challenges have been delineated, it is essential to gain an understanding of the technical context in which it is situated. To this end, a comprehensive literature review is conducted to identify the existing research in this field, including the theoretical frameworks that underpin the various technologies contributing to the development of these standards. Furthermore, the SoTA review offers valuable insights and alternatives that may prove beneficial or, conversely, introduce an improved depth to the research approach.

3. Solution design and development.

This phase encompasses the technical design, development, and implementation of solutions that address the research problem. To achieve this, solutions must be developed to address the research gaps identified in the SoTA review and provide advancements beyond existing literature. In this regard, the tasks associated with this step must be thoroughly executed throughout the research process. This meticulous approach ensures that the resulting data are of sufficient quality to be published in high-impact dissemination channels. Typical tasks involved in these steps include solution design according to technical complexities, documentation of challenges and achievements, integration of components, and validation.

4. Solution assessment.

Having completed the design and development of the solution, the next phase is to evaluate its effectiveness. In this regard, the assessment encompasses the fundamental tasks that facilitate the research process, as it yields outcomes that attest to the solution's alignment with the identified problem and challenges.

This stage is of paramount importance, as it provides the essential feedback that guides the direction of the research methodology. Consequently, if the results are positive, they can inform the design of further research lines that can be leveraged for future investigation. Conversely, in the event that the results do not satisfy the objective, it may be necessary to address additional solution tuning or new approaches in order to accomplish the desired outcome. In both instances, significant insights can be obtained in the subsequent phase.

5. Results dissemination.

The final step serves to complement the entire research methodology, as it provides the means by which the work performed can be justified. In this manner, the description encompasses the various stages, as documented in technical literature disseminated in scientific journals or conveyed in popularization forums such as conferences, held at the national or international levels. Moreover, the thesis results, and contributions to parallel projects, have been demonstrated in tangible form in technical venues.

This closed-loop methodology paves the way for new research lines arising from new approaches or alternative ones that were conceived throughout the thesis development.

3.2 Description of the publications

This section presents the contributions, in the form of scientific articles, that respond to the objectives and challenges defined in Section 1.3. As illustrated in Figure 3.2, the logical mapping between the identified challenges, objectives, and outcomes of the thesis. The articles in question have been published in journals deemed relevant and indexed in the Journal Citation Reports (JCR), wherein they summarize the results achieved through the aforementioned methodology. In this regard, the established objectives are reflected in publications that align with the identified challenges and contribute to the attainment of key milestones that address the specified requirements.

Consequently, the resulting articles have been arranged to cover three well-defined chapters. Chapter 4 defines the exploration of new-generation services considering requirements and design challenges, and the development process of environments that enable controlled experimentation. Then, 5 covers the methodology applied to simplify the measurement and monitoring of next-generation services by leveraging the ML approach. To finalize, 6 presents the exploitation of the KQI paradigm for service-based network management.

The following section presents a logical distribution and synthesis of the articles, including a mapping of the challenges, objectives, and milestones of the thesis.

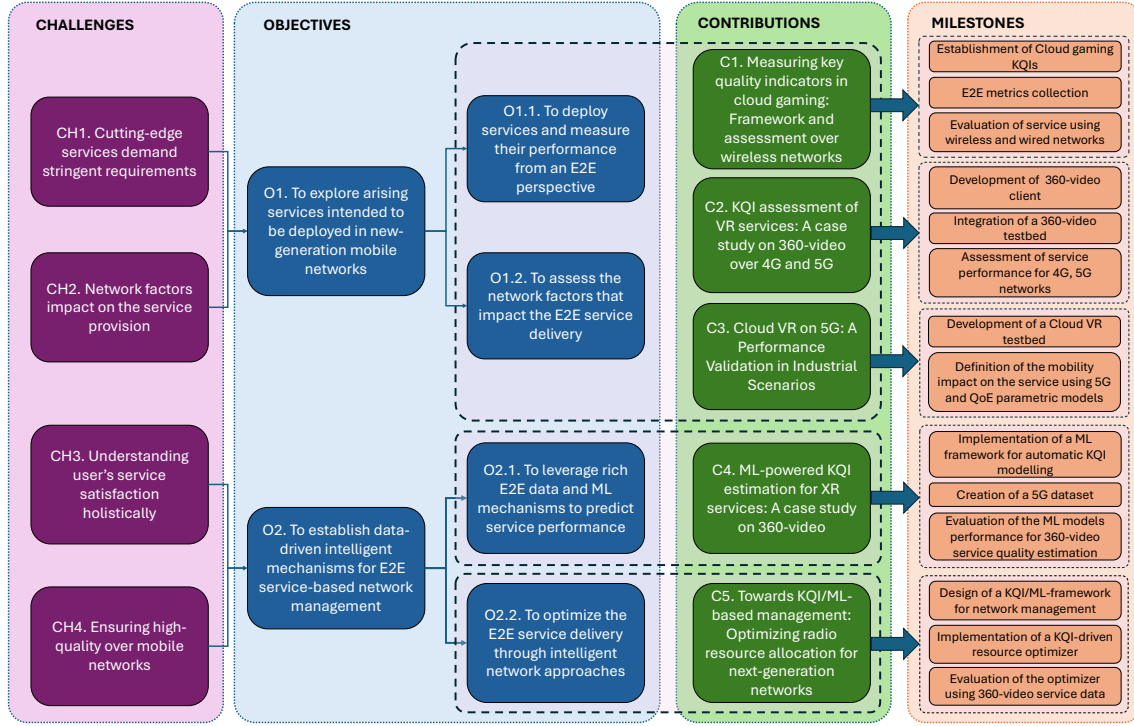


Figure 3.2: Mapping between thesis challenges, objectives and results achieved

3.2.1 C1. Measuring key quality indicators in cloud gaming: Framework and assessment over wireless networks

The article examines cloud gaming as one of the most advanced multimedia services, which is anticipated to be supported by the next generation of mobile networks (Objective 1). This work is framed into Chapter 4.

Firstly, this research work introduces the concept of CG and analyses the various architectural concepts that define the services and their requirements. The fundamental concept underlying this service is the utilization of “thin” clients, which reduce the computational burden on traditional gaming devices (e.g., GPU-enabled computers) through the use of cost-effective devices. Such equipment is capable of capturing user actions and transmitting them to a cloud server, which then executes the necessary game logic on the user’s behalf. Subsequently, the visual feedback is processed on the server side and transmitted back to the user via multimedia data, specifically video and audio.

In order to achieve Objective 1.1, a design, development, and implementation plan for a CG client is presented. This design incorporates the development of a

methodology that facilitates automated user interaction (i.e., coordinates calibration, client execution, and mouse and keyboard actions emulation) with the game software and the collection of E2E metrics. Similarly, the client is capable of automatically configuring various session parameters, including video resolution, frame rate, video encoders/decoders, and audio settings, among other available elements.

In light of the fact that service-based network management is oriented towards the end-to-end performance of the services in question, it is necessary to implement a KQI extraction tool that collects information at the application level from the client while a session is active. The data is then collated and encoded in JavaScript Object Notation (JSON) format and transmitted to a central REpresentational State Transfer (REST)ful server for processing.

Furthermore, this article details the incorporation of the CG client into a testbed that facilitates the assessment of multiple transport networks. This task aims to ascertain the impact of wireless networks on the service (Objective 1.2). To this end, the evaluation assesses LTE, WiFi, and Ethernet technologies through controlled measurement campaigns that modify reference service factors (i.e., resolution and frame rate). The assessment analyzes the degradation associated with the service KQIs generated by the utilization of wireless networks in comparison with Ethernet.

Along these lines, this work provides key insights into the deployment of CG through wireless networks in replacement of the traditional wired technologies (i.e., Ethernet). This is a major enabler considering that mobility in this type of service is important to provide even more advanced experiences in non-static situations, paving the way for evolved approaches such as XR applications and use cases.

3.2.2 C2. KQI assessment of VR services: A case study on 360-video over 4G and 5G

This research outcome is presented in Chapter 4 and focuses on investigating the use of 360-video services as one of the emerging applications of the broader field of XR. This type of service belongs to the category of immersive technology that has recently attracted attention due to the advent of the metaverse concept through wireless networks.

In this context, this article examines the architectural differences between

traditional service and VR technology, which requires specialized equipment, such as Head-Mounted Devices (HMDs), to provide users with an immersive experience. In light of these considerations, new challenges emerge with regard to the delivery of high-quality service levels, with mobility emerging as a particularly crucial factor. Similarly, the requirements associated with VR technology are discussed, as they can be considered stringent in terms of visual quality and latency. Insufficient resources may result in the manifestation of physical discomfort and cybersickness.

This article addresses the theoretical investigation of the 360-video service, including its requirements, architecture, and the current SoTA background. This paper presents the development of a service-specific Dynamic Adaptive Streaming over HTTP (DASH)/360-video client using Unity 3D as a graphic engine. The client is capable of performing a multitude of tasks concurrently, including the rendering of omnidirectional video on a sphere, the detection of user movement, and the collection of KQI data points such as video resolution, latency, frame rate, buffer health, and others. The preliminary development of this work responds to Objective 1.1.

Moreover, the work addresses Objective 1.2 by identifying the network factors (that affect E2E service delivery) that arise as challenges to be faced by mobile networks. To address this, the performance of LTE and 5G networks is evaluated using a 360-video testbed comprising a client, a RESTful server, and a network-in-a-box solution [12] that deploys the service in realistic scenarios. The REST server is responsible for coordinating the data from the client and network (i.e., radio measurements, statistics, and configuration parameters) sides. This framework assesses the service performance under different radio conditions, namely cell-edge and minimum interference cases, and demonstrates the potential of next-generation networks to deliver immersive services.

3.2.3 C3. Cloud VR on 5G: A Performance Validation in Industrial Scenarios

In alignment with the aforementioned research, this article reinforces the necessity for recognizing cutting-edge service contexts in order to adequately prepare upcoming networks for their deployment and integration. This work, which can be found in Chapter 4, explores the Cloud VR service. This service builds upon existing VR approaches by combining the benefits of CG. These enhancements are

characterized by a reduction in user equipment cost, achieved through the use of a cloud server that is responsible for rendering the visual feedback after receiving the user interactions and sensing information. This is made possible by a simple architectural and hardware design.

The primary advantage of the Cloud VR service is its versatility, which makes it appealing to a range of industry verticals and CSPs. Its ability to be readily delivered to a multitude of devices is a significant benefit, particularly given that the data in DL encompasses generic multimedia. Nevertheless, disaggregating the processing from the user device also presents a significant challenge. It is essential to ensure that the latency introduced by the transport of the content is minimal to prevent any degradation in the immersive experience. Given that the consumption of content through the HMD requires even more rigorous considerations than content in CG, this service requires particular attention.

To address this necessity, in this research work the delivery of Cloud VR service is assessed in a real industry environment located at the Aalborg University (AAU). This scenario is selected considering the broad portfolio of XR services that aim to cover diverse use cases of Industry 4.0 and the upcoming 5.0 approach (e.g., remote assistance, remote robot operation, training, and safe inspection, among others). In this sense, the objective is to identify the main issues and challenges that mobile networks will face and the benefits the services can find by using them. This exploration includes the design and implementation of a Cloud VR client/server by integrating Unity 3D, ALVR software, and Steam VR (Objective 1.1).

In order to evaluate the performance, 5G and WiFi technologies are selected to provide connectivity to a VR HMD. The client and server collect KQIs and traces that describe the performance of the service through the experimentation. The results show that mobile networks have important strengths in maintaining a steady latency response but a cost of increasing the median latency in comparison with WiFi. Moreover, the effects of mobility are evaluated along a defined path that provides valuable insights into the degradation of the service performance. This is especially noticeable in WiFi, whose latency and throughput performance are heavily affected whilst the 5G one remains suitable. The results are intended to approach Objective 1.2.

To complement the study, an examination of the QoE is conducted to ascertain the degree of quality associated with the user's perspective. To this end, analytical

models of QoE are constructed that objectively reflect the user experience based on visual quality, latency, and video stalling (i.e., frame loss). The findings demonstrate that 5G offers superior quality cloud VR performance compared to WiFi in scenarios involving mobility. This is a noteworthy outcome of the work, which demonstrates the viability of implementing industry XR use cases and sets a foundation for future research in this field for all parties involved in the development of 5G standards.

3.2.4 C4. ML-powered KQI estimation for XR services: A case study on 360-video

The advent of advanced technologies and services, such as Extended Reality XR, is poised to transform everyday life. Simultaneously, new decentralized architectural approaches have laid the foundation for next-generation mobile networks like 5G, while setting the stage for B5G and future developments. These networks are anticipated to be critical enablers for realizing the metaverse and other forward-looking services. In this context, accurately quantifying service performance is essential for dynamic, adaptive, and proactive network management.

The evaluation of the E2E service performance is a challenging endeavor, particularly given the reliance on subjective user feedback, which necessitates the collection of data that may be biased. Furthermore, in order to ascertain a genuine perception of customers, this methodology necessitates the involvement of multiple individuals per category and service. Consequently, this strategy is not scalable. Conversely, measuring KQIs at the user equipment may result in violations of user privacy or the unavailability of the necessary information. To address these challenges, the estimation of KQI through ML represents a promising solution, as it allows for the exploitation of a vast amount of data available from CSP.

The estimation of KQIs, as objective E2E metrics that reflect the user perception of the service, is one of the key pillars that constitute this thesis. In this context, the presented article covers the goal established in Objective 2, especially with respect to Objective 2.1. Likewise, the methodology and results respond to the thesis structure, in particular to Chapter 5.

This article describes an ML-based framework that leverages network-accessible data including radio measurements, statistics, and configuration parameters, to

identify optimal ML models for predicting various XRs KQIs. The framework incorporates feature engineering techniques, namely FE, and FS, that improve model efficiency and performance by aggregating techniques that can increase the model's ability to capture the most adequate relationships and patterns. Similarly, this approach explores a variety of algorithms looking for the most adequate that better fits the intrinsic nature of the data, including Random Forest (RF), K-Near Neighbors (KNN), Support Vector Machine (SVM), AdaBoost, Multi-Layer Perceptron (MLP), Ridge Regression (RR), among various in their regression and classification schemes for continuous or discrete targets.

Beyond this strategy, the presented framework ensembles multiple ML stages, such as data preprocessing, model definition, training, tuning, and validation, into a pipeline-driven automatic strategy that ensures scalability. In addition, a novel metric, the Performance Error and prediction Time score (PET_{score}), is proposed to assess and select the most suitable algorithm, balancing both error rate and prediction time—crucial for time-sensitive services like XR.

For validation, the framework is applied to address the 360-degree video service, showcasing its effectiveness with a real XR use case. This service has been selected since its performance depends mostly on the network rather than on the processing capacity of the equipment. To evaluate the level of accomplishment of the framework, a dataset has been generated to reflect the service behavior under different channel conditions and resources granted to the user. The dataset originated from the experimentation is publicly available and fully referenced in the article [19], thus, endowing the research community with tools that can accelerate related investigation.

The results demonstrate the capabilities of this proposal by predicting the 360-video KQIs with suitable accuracy. This enables the possibility for the networks to intelligently monitor the service delivery using a holistic perspective. Furthermore, the use of ML approaches ensures scalable deployment and flexibility for service-tailored solutions that strengthen next-generation network management.

3.2.5 C5. Towards KQI/ML-based management: Optimizing radio resource allocation for next-generation networks

Chapter 6 finalizes the research work performed in the thesis by presenting a proposal for service-based network management. The development of this article encompasses Objective 2 with a special focus on optimizing the E2E service delivery (Objective 2.2).

This study proposes an innovative and integral approach, grounded in a ML-enhanced framework, that facilitates quality-based service management. Leveraging ML capabilities, this framework enables comprehensive E2E QoE assessment by utilizing network-accessible data to evaluate and optimize the network and services performances from a holistic perspective, moving beyond traditional, network-centric methods to service and user-centric ones. This mechanism exploits the key functionalities offered by the developments provided in Chapter 4 (i.e., KQI identification and measurement) and Chapter 5 (i.e., ML framework for automatic model provision) to create a versatile tool that able to ease resource management.

The developed approach comprises three fundamental elements: data acquisition, service quality monitoring and acquisition, and optimization. The initial and subsequent components integrate the aforementioned solutions. Subsequently, the optimization component formulates a KQI/ML-based functionality, designed to optimize network performance while ensuring compliance with the SLA in terms of KQI. This is a significant point of consideration, given that legacy SLA management was contingent upon the compliance of data rate and latency indicators. The aforementioned methodology is not feasible for next-generation networks that support advanced services with a wide diversity of requirements, as it lacks the necessary scalability and flexibility.

In this regard, the proposed solution seeks to address the technological gap created by the evolution of network architectural concepts and derive its management through intelligent approaches. To illustrate the applicability of the proposed approach, the optimal resource allocation problem has been selected as the use case. In this context, the optimization mechanism identifies the optimal number of RBs that satisfy the minimum conditions specified in the SLA.

Furthermore, the mechanism combines two key drivers: ML and numerical optimization. The initial approach involves the generation of channel-specific ML models that represent the functions associated with the allocation of resources within the network. The channel-specific models are automatically generated by utilizing the ML framework described in Chapter 5 in combination with a clusterization strategy that identifies the different radio scenarios within the input dataset. Subsequently, these models are integrated with a service-specific QoE function to form a target function. The latter is optimized through a dual-annealing strategy [113] with the objective of defining the combination of KQIs that are mapped into allocated RB, ensuring that the service requirements are satisfied and that resource usage is minimized.

To substantiate the efficacy of this methodology, the 360-degree video service was subjected to evaluation under cell-edge conditions across 4G and 5G networks. The results demonstrate that the optimizer effectively identified the requisite resource allocation to meet service demands while minimizing resource occupation. Furthermore, the evaluation demonstrates the framework's adaptability and scalability, indicating its potential to support a range of emerging services (e.g., XR) and upcoming network architectures, such as Open RAN.

3.3 Conclusions of the chapter

This chapter has outlined the comprehensive methodology employed in this thesis, detailing each stage from defining objectives to disseminating results. It has also systematically presented each contribution in relation to the research objectives, establishing a cohesive framework that underscores their significance and relevance to this line of research.

Chapter 4

Experimental new-generation service environments

Content

4.1	Measuring key quality indicators in cloud gaming: Framework and assessment over wireless networks . . .	74
4.2	KQI assessment of VR services: A case study on 360- video over 4G and 5G	75
4.3	Cloud VR on 5G: A Performance Validation in Industrial Scenarios	76
4.4	Conclusions of the chapter	77

This chapter presents the research work aligned with the thesis's first objective. In particular, it addresses the exploration of diverse advanced multimedia services in terms of their architectural design and the requirements for their implementation. Subsequently, the design and development of service-specific clients are described to enable experimentation that exposes the impact of the network on their performance.

4.1 Measuring key quality indicators in cloud gaming: Framework and assessment over wireless networks

[1] O. S. Peñaherrera-Pulla, C. Baena, S. Fortes, E. Baena and R. Barco, “Measuring key quality indicators in cloud gaming: Framework and assessment over wireless networks”, in *Sensors*, vol. 21, no. 4, 2021, doi: 10.3390/s21041387.

Abstract: The development of telecommunications, particularly mobile communications, has led to the emergence of numerous multimedia services. As these services become increasingly reliant on connectivity, the network has a huge impact on the user experience. However, it is difficult to assess the End-to-End (E2E) performance of the service because of the limited access to the application data and encryption used to secure and integrate the network information exchanged. In this context, this work presents a dataset based on several metrics, namely Key Quality Indicators (KQIs), representing different services’ E2E conditions. In particular, video streaming and Cloud Gaming (CG) services are considered in the dataset. Furthermore, a detailed description of the testbed where the dataset has been generated is given, discussing the rationale behind the choices made for its development.

4.2 KQI assessment of VR services: A case study on 360-video over 4G and 5G

[2] O. S. Peñaherrera-Pulla, C. Baena, S. Fortes, E. Baena and R. Barco, “KQI Assessment of VR Services: A Case Study on 360-Video Over 4G and 5G”, in *IEEE Transactions on Network and Service Management*, vol. 19, num. 4, pp. 5366-5382, Dec. 2022, doi: 10.1109/TNSM.2022.3192762.

Abstract: Extended Reality (XR) arises as one of the current cutting-edge educational and entertainment emergent technologies. This service differs from traditional video streaming approaches due to its immersive experience, which allows the user to enjoy omnidirectional multimedia. However, the service experience must be guaranteed to avoid side effects such as cybersickness or disorientation. This work presents a framework to assess 360-video service streaming performance over mobile networks through Key Quality Indicators (KQIs) using VR (Virtual Reality) HMD (Head Mounted Device). The testbed is composed of a 360-video client for DASH (Dynamic Adaptive Streaming over HTTP), which playbacks multimedia content from a video server located in the cloud while KQI measuring tasks are performed in the user as on the network sides. Various metrics are collected such as resolution, frame rate, initial playback time, throughput, stall events, and round trip time (RTT), among various others. Finally, a performance comparison between LTE and 5G technologies is provided. Results from the KQI measurement highlight the potential of the new generation of mobile networks in the provision of service with high-quality levels of experience.

4.3 Cloud VR on 5G: A Performance Validation in Industrial Scenarios

[3] O. S. Peñaherrera Pulla, S. B. Damsgaard, B. Yanakiev, P. Mogensen, S. Fortes and R. Barco, “Cloud VR on 5G: A Performance Validation in Industrial Scenarios”, in *IEEE Open Journal of the Communications Society*, vol. 5, pp. 3641-3657, 2024, doi: 10.1109/OJCOMS.2024.3413328.

Abstract: The emergence of the novel approach of Industry 5.0 entails the utilization of immersive technologies such as XR (Extended Reality), which promise to transform the conceptualization of industry. The advent of new technology enablers, such as next-generation mobile networks, including 5G and B5G (Beyond 5G), plays a pivotal role in enhancing productivity, optimizing operators’ training programs, and minimizing unnecessary risks. This study presents a performance assessment of Cloud VR services over 5G and WiFi in industrial scenarios. This scenario involves several factors that can degrade radio links, including fading and interference. To assess the suitability of 5G for the Industry 5.0 hot topic of XR, a 5G SA/WiFi-6 Edge Cloud VR setup was developed in the Smart Production Lab at Aalborg University. The results demonstrate that 5G outperforms WiFi in terms of latency and throughput consistency during the experiments in DL (Downlink), which belongs to the streaming channel with UDP (User Datagram Protocol). In the uplink (UL), WiFi generally exhibits lower latency than 5G when employing Transmission Control Protocol (TCP) for the control channel. To provide a comprehensive End-to-End (E2E) analysis, an objective Quality of experience (QoE) metric has been estimated. The outcomes demonstrate that 5G achieves higher QoE values for mobility than WiFi, which shows a decline in performance. The results indicate that 5G performance is within theoretical limits for XR experiences and suggest that it is a promising candidate for enabling Industry 5.0 and its use cases.

4.4 Conclusions of the chapter

This chapter has presented a synthesis of the research work concerning Objective 1, as reflected in multiple scientific publications that explore advanced multimedia services. This study has analyzed the architectural concepts and requirements that characterize the services of CG, 360-video, and Cloud VR.

To address this issue, the design and development of service-specific clients have been undertaken with the objective of identifying and quantifying the diverse KQIs that reflect the quality of service delivery in the E2E context.

Subsequently, with the objective of conducting controlled experimentation, the integration of the clients into the testbed was undertaken. In light of these developments, the design and implementation of measurement campaigns have been introduced with the objective of quantifying the impact of network effects on service performance. In this context, each service is evaluated in terms of different aspects, including radio conditions, mobility, and the feasibility of using wireless networks as replacements for traditional wired networks, with the aim of providing reliable multimedia services.



UNIVERSIDAD
DE MÁLAGA

Chapter 5

ML-powered new-generation service monitoring

Content

5.1	ML-Powered KQI Estimation for XR Services: A Case Study on 360-Video	80
5.2	Conclusions of the chapter	81

This chapter investigates the potential for utilizing ML methodologies to track service performance from an E2E perspective, as evidenced by a recent journal publication. This work presents a ML framework that streamlines the process of KQI modeling and prediction by leveraging network-accessible data, thereby facilitating efficient service monitoring. These developments are situated within the context of Objective 2.1.

5.1 ML-Powered KQI Estimation for XR Services: A Case Study on 360-Video

[4] O. S. Peñaherrera-Pulla, C. Baena, S. Fortes and R. Barco, “ML-Powered KQI Estimation for XR Services: A Case Study on 360-Video”, in *IEEE Open Journal of the Communications Society*, vol. 5, pp. 4075-4097, 2024, doi: 10.1109/OJCOMS.2024.3422872.

Abstract: The emergence of cutting-edge technologies and services such as Extended Reality (XR) promises to change how people approach everyday living. At the same time, the emergence of modern and decentralized architectural approaches has ushered in a new generation of mobile networks, such as 5G, as well as outlining the roadmap for B5G (Beyond-5G) and further advancements. These networks are expected to be the enablers for the realization of the metaverse and other futuristic services. In this context, quantifying the service performance is a key enabler for dynamic, environment-adaptive, and proactive network management. This work presents an ML-based (Machine Learning) framework that uses data from the network, such as radio measurements, statistics, and configuration parameters to infer the best ML models that fit diverse XR Key Quality Indicators (KQIs). The output models integrate feature engineering techniques that enhance model size and performance. The proposed framework comprises data preprocessing, model definition, training, tuning, and validation. Additionally, to select the best combination algorithm this work introduces a metric called PET_{score}, which evaluates algorithm candidates in terms of error performance and prediction time. These are considerations that are needed for time-sensitive services like XR's. To validate our proposal, the 360-video service has been chosen to demonstrate the potential of this ML framework with a real XR use case. In addition, the dataset generated for the use case evaluation is publicly accessible and properly referenced. Furthermore, this work serves as a foundation for future research on end-to-end (E2E) quality of experience (QoE)-based network management in conjunction with other enabling technologies, including network slicing, virtualization, and multi-access edge computing (MEC).

5.2 Conclusions of the chapter

This chapter has introduced a service-based ML approach to accelerate the development of E2E service management by leveraging network-accessible data. The proposed solution addresses the challenges that arise from the necessity of handling service monitoring through KQIs in situations where this data is inaccessible and prevents operators from violating user privacy.

The proposed framework considers a multi-step solution that integrates automatic model training, hyperparameter tuning, evaluation, and selection through a pipeline-based approach. The model training stage incorporates a number of feature engineering techniques, including FE and FS. These additional functionalities enable the creation of models that are more closely aligned with the intrinsic characteristics of each KQI. In a similar vein, these techniques are combined with a variety of algorithms to create a service-oriented approach that can provide flexibility and scalability to networks considering service-oriented management in next-generation technologies.

To assess the efficacy of the proposed approach, a 360-video service has been employed to quantify the level of error and prediction time performance. This is achieved through the utilization of a novel metric, PET_{SCORE} , which weighs and selects the model based on the aforementioned parameters.

Furthermore, the experimental work yielded a dataset that reflects the service behavior under different radio conditions and resource allocation options. The data has been made publicly available for use by the research community.



UNIVERSIDAD
DE MÁLAGA

Chapter 6

KQI-driven E2E network management

Content

6.1	Towards KQI/ML-based management: Optimizing radio resource allocation for next-generation networks .	84
6.2	Conclusions of the chapter	85

This chapter examines the utilization of ML to facilitate the implementation of E2E service-based network management for next-generation mobile technologies. The research presented here is aligned with the aims outlined in Objective 2.2. The developed solution comprises an ML/KQI-driven mechanism that addresses the resource allocation problem whilst guaranteeing SLA compliance with minimum resource usage.

6.1 Towards KQI/ML-based management: Optimizing radio resource allocation for next-generation networks

[5] O. S. Peñaherrera-Pulla, C. Baena, H. Q. Luo-Chen, S. Fortes and R. Barco, “Towards KQI/ML-based management: Optimizing radio resource allocation for next-generation networks”, 2024 - in *TechRxiv*, <https://www.techrxiv.org/users/685237/articles/1236293>

Abstract: This work presents a novel optimization approach for service management in the context of new-generation networks such as B5G (Beyond-5G) and 6G. Our proposal is built on a Machine-Learning-powered (ML) framework designed to support service management based on Quality of Experience (QoE). This framework leverages ML strengths to address the End-to-End (E2E) QoE assessment through the use of network-accessible data to reflect the performance from a holistic perspective instead of legacy network-centric approaches. The framework encompasses three principal functionalities: acquisition, estimation, and optimization. The latter functionality generates a KQI/MLbased target function to be optimized aiming to establish the optimal resource allocation for Service Level Agreement (SLA) compliance. To validate, a 360-degree video service has been used as a use case through 4G and 5G networks in cell-edge conditions. The results show that the optimizer successfully found the optimal resource allocation to satisfy the service requirements while minimizing resource usage. In addition, the evaluation demonstrated the flexibility and scalability capabilities needed to deploy this solution to address the wide variety of upcoming services (e.g., Extended Reality-XR) and network architectures such as OpenRAN.

6.2 Conclusions of the chapter

This chapter has presented the potential of combined ML and quality-based approaches to address the impending complexity regarding network management, with a particular focus on the user and services. The solution described in the research work illustrates the viability of intelligent ML approaches in enhancing network scalability and ensuring the effective delivery of a diverse range of advanced services.

In order to address the challenges associated with Objective 2.2, it is proposed that an integral E2E management strategy be implemented. This approach comprises a three-stage methodology that exploits the advantages conferred by the developments described in Chapters 4 and 5 to enhance network performance. Given the multitude of potential functionalities that can be derived from this perspective, the problem associated with resource allocation was addressed.

The solution examines the potential of an optimization approach to maximize the QoE while minimizing resource usage through a combined target function. This latter component encompasses a channel-specific network component that models the resource allocation in function of the radio conditions, as well as a service-specific quality function. The service function is based on an objective reflection of the user perception with regard to the service aspects.

Subsequently, a 360-video service was selected for the purpose of quantifying the effectiveness of the solution. The results demonstrate the efficacy of the proposed resource allocation strategy in meeting the service requirements with respect to KQI. This illustrates that the implementation of these methodologies can facilitate the administration of intricate and diverse services for next-generation networks.



UNIVERSIDAD
DE MÁLAGA

Part III

Conclusions



UNIVERSIDAD
DE MÁLAGA

Chapter 7

Conclusions

Content

7.1	Contributions	90
7.2	Future work	95
7.3	Moving towards mobile industry	96

This chapter provides a synopsis of the research conducted in this thesis, presented in the preceding chapters. In order to elaborate on the different aspects of this research, the conclusions have been defined from three different perspectives. Firstly, these outcomes result from the contributions made toward achieving the stated objectives are presented in Section 7.1. Secondly, they give rise to new avenues of research in Section 7.2. Thirdly, insights into the mobile networks industry are introduced in Section 7.3.

7.1 Contributions

This thesis has addressed the challenging scenario of deploying advanced multimedia services through the utilization of mobile networks. The provision of these services entails specific requirements that differ significantly from those associated with traditional ones. Therefore, next-generation networks necessitate not only an evolution from existing architectural concepts but also a transition from legacy network-centric management paradigms to service-based ones that ensure the optimal delivery of services to users.

The analysis of the problem has enabled the formulation of two principal objectives, which address the challenges associated with it.

On the one hand, Objective 1 proposed the delivery of new-generation services through mobile networks. In order to address this approach, two sub-objectives have been set forth with the intention of analyzing the scenario and the challenges that must be faced. Subsequently, Objective 1.1 prompted the investigation of advanced services, architectural concepts, their primary performance metrics, and the prerequisites for their effective delivery to users. This led to the formulation of Objective 1.2, which established the necessity of identifying the pivotal impact factors and consequences of mobile network usage in the deployment of these services from an E2E perspective. This, in turn, would facilitate a transition from a network-centric to a user- and service-centric paradigm.

In contrast, Objective 2 focused on developing intelligent methodologies for service-based network management within an E2E architectural framework. This resulted in the pursuit of two sub-objectives. Objective 2.1 investigated the potential of combining network-accessible data with ML algorithms to predict E2E metrics that accurately reflect the actual service performance, without the influence of the user. Subsequently, Objective 2.2 conducted research on the implementation of ML to develop intelligent methodologies for service-based network management. This intended to integrate the findings of Objective 2.1 with other pivotal technology drivers, which are crucial for the advancement and evolution of next-generation networks.

The contributions dealing with the abovementioned objectives are summarized herein.

- **Objective 1: To explore arising services intended to be deployed by new-generation mobile networks.**

- In alignment with Objective 1.1, three advanced multimedia services, which are intended to play an important role in next-generation networks, were subjected to a detailed examination. In this regard, the architectural concepts pertaining to CG, 360-degree video, and Cloud VR were examined in conjunction with the associated requirements and the set of objective performance metrics as outlined by the standardization bodies, including 3GPP, ETSI, and ITU-T. Subsequently, in order to gain further insight into the services beyond literature, the design and development of these were achieved through the creation of service-specific client/service architectures that reflect the E2E performance. In this regard, the deployment of these services entailed the completion of additional preparatory tasks, designed to facilitate experimentation within controlled scenarios. Such tasks include the automation of user actions, the design of visualization mechanisms for VR-related services, the coding of tools for the collection and measurement of metrics, and the development of communication mechanisms between clients and a REST server.
- In regard to Objective 1.2, the aforementioned architectures were integrated with mobile networks to create fully controlled experimentation scenarios. The principal objective of these tasks was to identify the network impact factors and their influence on the quality of service delivery from an E2E perspective. To achieve these experimental scenarios, the infrastructure of MobileNet was employed, including LTE, 5G-SA commercial network deployments, and the network-in-a-box solution [12], as well as the previous developments related to Objective 1.1.

The contributions associated with this objective are aligned with the preceding three services, namely CG, 360-degree video, and Cloud VR. The assessment of these services aimed to identify and analyze different network impact factors. The outcomes intend to establish a baseline for the development of next-generation network standards and use cases.

Firstly, the performance degradation of the CG system was investigated by utilizing wireless networks in lieu of conventional wired technologies.

To this end, experiments were conducted to assess the impact of varying technological configurations, video resolutions, and frame rates on performance. The findings indicated that wireless networks can facilitate effective service delivery in comparison with the wired approach results.

Secondly, it was conducted a comparative analysis of the advantages of utilizing 5G technology in comparison with LTE utilizing a 360-degree video deployment. To quantify this approach, measurement campaigns were conducted in a variety of radio conditions, including cell-edge and minimum-interference scenarios. The findings indicated that 5G exhibited superior performance in comparison to LTE in terms of resolution, frame rate, latency, and stalling time, while also demonstrating reduced frequency resource usage. Furthermore, potential avenues for future research were identified with the aim of enhancing the quality of the service by capitalizing on the capabilities of 5G technology.

Thirdly, the impact of incorporating mobility in service delivery was evaluated using the Cloud VR. To this end, WiFi and 5G-SA were evaluated within a real-world industrial setting, taking into account the effects of indirect visibility and interference due to equipment and personnel. The outcomes demonstrated the advantages of 5G by illustrating consistent performance in both static and mobile scenarios, in contrast to WiFi, which exhibited a notable decline in the latter scenario. To provide additional context, the evaluation of the E2E service performance was conducted using parametric QoE models, which confirmed the feasibility of utilizing 5G for the deployment of such solutions towards Industry 4.0 and 5.0 scenarios.

- **Objective 2: To establish data-driven intelligent mechanisms for E2E service-based network management.**

- Following Objective 2.1, an investigation was undertaken to ascertain the benefits of employing AI/ML methodologies for E2E service management. This approach encompasses the entire communication network, though, in certain instances, data might not be accessible to the CSPs. This is due to constraints in the implementation of user-device services or to prevent potential violations of privacy.

To address this challenge, the potential of ML was investigated as a means of enabling service monitoring through the estimation of KQIs from the data that is available to operators.

Firstly, in order to leverage ML, it was necessary to develop appropriate measurement campaigns in order to conform to a dataset encompassing the service performance under different network conditions. To address this, the 360-degree video setup was selected to characterize the KQI behavior under cell-edge and minimum-interference scenarios with progressive resource allocation. The dataset generated throughout the experimentation has been made publicly available via IEEE DataPort [19]. This release is intended to encourage further research in this field.

Secondly, a machine learning framework was designed and implemented to streamline KQI modeling by automating the process and minimizing the need for manual intervention. This framework includes a pipeline-based procedure that covers model training, hyperparameter tuning, and evaluation. Additionally, various feature engineering techniques, including FE and FS, were incorporated to improve alignment with the characteristics of the KQIs.

Thirdly, to select the optimal model from the plethora of algorithms and techniques that were automated by the framework, the Performance Error and prediction Time score, (PET_{SCORE}) was introduced. This metric allows inferring the most suitable model based on the error and prediction time performance. In this sense, the score offers a flexible framework that allows for the consideration of both aspects in accordance with the specific circumstances. This can be beneficial in circumstances where time-sensitive service monitoring is required.

Finally, the framework was evaluated using the 360-video dataset, and

the results demonstrated that the selected models exhibited a performant level prediction of the service KQIs.

- To conclude the research project, the need for integrated mechanisms that utilize ML to facilitate service-based network management was addressed. The proposed approach comprises a three-step framework, namely metrics acquisition, performance estimation, and optimization. The latter term is of a broad nature, encompassing a number of potential solutions. Consequently, the thesis investigated the problem of resource allocation.

First, the framework acquisition and estimation steps leverage the benefits identified in the developments resulting from Objectives 1.1, 1.2, and 2.1. In this sense, service-based management can be enabled by tracking the service itself through the use of KQIs, which exposes the E2E performance. In instances where these metrics are quantifiable, the CSP is able to retrieve the requisite data and generate conforming datasets for the purpose of feeding ML solutions. Subsequently, the estimation of metrics can be employed to facilitate the generalization of service monitoring through the utilization of intelligent mechanisms.

Secondly, the proposed optimization approach aims at improving the resource allocation problem. To address this, a hybrid ML and numerical optimization approach has been employed to determine the optimal number of RBs to be allocated in order to provide specific KQI requirements. This focus is based on ML-based channel-specific network modeling and service-specific QoE functions. The two components are integrated through a target function that is optimized through a dual-annealing methodology with the objective of minimizing resource usage and maximizing E2E performance. The degree of compliance of the solution is evaluated through an assessment of the SLA requirements in terms of the minimum KQIs.

The results have demonstrated that the proposed tool offers significant advantages in determining the optimal resource allocation for users of the service under diverse radio conditions and SLA requirements. Furthermore, the methodology employed demonstrates advantages with regard to scalability and flexibility, which are inherent to ML paradigms. This integral ML strategy aims to cover and ease different

aspects of service management from acquisition, monitoring, and performance optimization. Likewise, the exploitation of these dynamic mechanisms is intended to facilitate the development of service-specific solutions tailored to the requirements of next-generation networks.

In conclusion, this thesis makes a notable contribution to the transition to next-generation mobile networks through the implementation of advanced services and intelligent service-based E2E management.

7.2 Future work

This thesis has identified multiple further research avenues and potential applications. These encompass enhancements that build on the current research and new developments and broader applications inspired by it.

- Further investigation is required to assess the feasibility of additional advanced multimedia services over cellular networks (e.g., AR and MR), taking into account the diverse requirements of such applications as exemplified by the VR approach. In this regard, it is imperative to assess the viability of the forthcoming array of services with exacting latency requirements, in order to ascertain the capacity of the 5G network to meet these demands. The results of this study can provide valuable insights into the capabilities of the next-generation networks and may also highlight the necessity for the development of additional enablers to address the challenges posed by this technology.
- In alignment with the aforementioned research line, it is of paramount importance to conduct an evaluation of multiple services and users in a simultaneous manner. This encourages the advancement of thesis solutions into multi-instance approaches that can leverage the available data to create service-specific solutions. Such an approach could prove beneficial in providing networks with intelligent, evolved approaches that facilitate the development of service-based mechanisms and generate experimental data sources in the form of datasets, thereby enhancing the research community's available information.
- With regard to the application of AI/ML paradigms in the development of next-generation networks, the use of Deep Learning (DL) has attracted

considerable attention, reflecting the ongoing growth in the number of data sources. Similarly, the exploration of Reinforcement Learning (RL) solutions can be suitable in accordance with the development of the Network Digital Twins (NDTs) concept. In this vein, the deployment of sophisticated algorithms can facilitate the development of advanced mechanisms aiming at enhancing network functionalities with intelligent support, including slicing management, advanced service prediction, proactive QoE degradation detection, and prevention.

- Concerning the advancement of the E2E framework for service-based network management, the incorporation of additional approaches (e.g., time series-based prediction) that capitalize on the drivers provided can augment the functionalities it is capable of offering. Likewise, this thesis explored the potential for optimizing resource allocation based on SLA requirements considering radio conditions. Nevertheless, a multitude of additional factors may be taken into account, including energy efficiency metrics, slice resource availability, and customized traffic policies, among others.
- The advent of the Open RAN paradigm is facilitating the availability of a multitude of options for endowing networks with enhanced flexibility and scalability, contingent upon the intelligence present at each layer and the virtualized functionalities inherent to the software. In this regard, the expansion of the developed solutions can be utilized to provide intelligent mechanisms that leverage data in heterogeneous contexts. This approach can be addressed by adapting the framework to operate within the scope of RIC. For instance, the creation of models can be delegated to the Non-RT RIC, while exploitation of the models should be addressed by solutions instantiated employing rApps and the forthcoming dApps [84]. In this way, time-sensitive procedures can be considered to provide users with high-quality service provision.

7.3 Moving towards mobile industry

In light of the considerable insights offered in this thesis, there exists a compelling rationale for formulating specific recommendations for the mobile industry that can be readily implemented. The following items are key insights derived from the study.

- The incorporation of the KQIs into the E2E service-based network management framework has the potential to facilitate the development of a multitude of cross-layer and multi-party solutions. Such an approach can prove advantageous for CSPs, ASPs, and customers alike.
- The availability of information reflecting the E2E service performance is a prerequisite for efficient management. This will facilitate a new level of service provision, enabling service providers and operators to take proactive actions to ensure quality consistency.
- The implementation of intelligent methodologies based on the exploitation of abundant network-accessible data represents a pivotal factor in the advancement of scalable, flexible, and dependable networks. In this regard, the deployment of AI/ML methodologies is of paramount importance in addressing the complex use cases that will be supported by next-generation architectures.
- The forthcoming generation of mobile architectural concepts (i.e., Open RAN) and technological drivers, exemplified by NFV and the disaggregation of network elements (i.e., Core and RAN), will leverage AI/ML integrated mechanisms to provide indispensable intelligence at every layer. In this sense, network slicing can be effectively supported by this approach, which enables the dynamic E2E quality-ensured handling of a wide variety of services, users, and scenarios.

In conclusion, the insights derived from this thesis provide a valuable foundation for advancing network management practices within the mobile industry. By integrating KQIs within the E2E service-based framework, network operators and service providers can develop more sophisticated, cross-layer solutions that benefit all stakeholders. Additionally, the access to comprehensive E2E performance information empowers proactive quality assurance, while intelligent, data-driven methodologies support scalable, flexible, and resilient networks. With the evolution of mobile network architecture, such as Open RAN and NFV, combined with AI/ML integration, these advancements promise to bring unprecedented intelligence and adaptability to meet the complex demands of next-generation mobile networks.



UNIVERSIDAD
DE MÁLAGA

Chapter 8

Research activities

Content

8.1	Dissemination	100
8.2	Related projects	100
8.3	Research stays	101

This section outlines the research activities conducted as part of this thesis. This encompasses the dissemination of the findings, the research projects in which the author has participated, and the research visits that have contributed to this thesis.

8.1 Dissemination

The findings of this research have been disseminated through various publications in prestigious scientific journals and conferences, as indicated at the outset of the manuscript.

Nevertheless, dissemination activities were not confined to the realm of scientific publications; they also encompassed demonstrations in events within the technological field. Through these demonstrations, the innovative approach and results of the research line of this thesis were shared, facilitating a valuable exchange of knowledge and experience with the academic and professional community. The aforementioned events included:

- Mobile Week Málaga 2021, organized by Palacio de Ferias y Congresos de Málaga, December 2021.
- 12th European Meeting on Science, Technology, and Innovation, organized by TRANSFIERE (Foro Europeo para la Ciencia, Tecnología e Innovación), in Málaga, Spain, February 2023.
- 13th European Meeting on Science, Technology, and Innovation, organized by TRANSFIERE (Foro Europeo para la Ciencia, Tecnología e Innovación), in Málaga, Spain, March 2024.

8.2 Related projects

This thesis was funded by Junta de Andalucía through “Secretaría General de Universidades, Investigación y Tecnología” under predoctoral Grant PREDOC.01712;

The work developed in this thesis has contributed to the following projects:

- **Piloto 5G Red.es Andalucía, Case Study OpenRAN:** Development of a 5G access network prototype based on GPPs. Project number 8.06/5.59.5630, funded by Ministerio de Asuntos Económicos y Transformación Digital and the European Regional Development Fund (ERDF), July 2019 - October 2021.
- **AECMA-5G:** Advanced E2E Cellular Management for 5G Applications. Project number UMA-CEIATECH-14, funded by Junta de Andalucía and

the European Regional Development Fund (ERDF), June 2020 - December 2021.

- **PENTA**: Provisión de servicios PPDR a través de Nuevas Tecnologías de Acceso radio, funded by Junta de Andalucía, February 2020 - February 2023.
- **MAORI**: Massive AI for the OpenRadIo b5G/6G network. Project number TSI-063000-2021-75, receiving funds from Ministerio de Asuntos Económicos y Transformación Digital and European Union - NextGenerationEU within the framework “Recuperación, Transformación, y Resiliencia”, January 2022 - December 2024.

Furthermore, the author has provided inputs to the following projects:

- **SMART**: Sistema de Monitorización y análisis Automático de Redes de Telecomunicación. Project number AT17-5882-UMA, funded by Junta de Andalucía within the framework “Plan Andaluz de Investigación, Desarrollo e Innovación (PAIDI 2020)”, October 2019 - September 2021.
- **MELODIC**: Machine Learning boosted WIFI cloud controller, funded by Ministerio de Economía, Industria y Competitividad via Centro para el Desarrollo Tecnológico Industrial (CDTI), November 2021 - April 2024.

8.3 Research stays

The development of this thesis entails a research stay in the Wireless Communication Networks (WCN) section at Aalborg University (AAU) in Denmark from March to July 2023. The research stay was supervised by Professor Preben Mogensen of the receiving institution.

Oswaldo Sebastián Peñaherrera Pulla

Málaga, Spain, February 2024.



UNIVERSIDAD
DE MÁLAGA

Part IV

Appendices



UNIVERSIDAD
DE MÁLAGA

Appendix A

Summary (Spanish)

Content

A.1 Motivación	106
A.2 Preámbulos	109
A.3 Objetivos	111
A.4 Descripción de las publicaciones	115
A.5 Conclusiones	123
A.6 Actividades de investigación	128

A.1 Motivación

El constante avance de las tecnologías de comunicación a lo largo del tiempo ha llevado a la aparición de numerosas aplicaciones y servicios diseñados para mejorar las actividades diarias. En esencia, el objetivo principal era facilitar la comunicación remota. Sin embargo, esto solo allanó el camino para el diseño de redes capaces de transmitir datos de propósito común de un punto a otro, lo que dio lugar a la aparición de diversas aplicaciones avanzadas. No obstante, con el desarrollo continuo de redes de acceso móvil de alta capacidad y ubicuas, el nuevo discurso es conectar todas las cosas y a todas las personas, colocando así el control al alcance de nuestras manos.

Esta nueva forma de pensar ha tenido un impacto significativo en varios segmentos de la sociedad. Desde la perspectiva de los consumidores, los usuarios finales de las redes demandan conectividad rápida, confiable y segura para acceder a la amplia gama de servicios y aplicaciones que tienen a su disposición. De manera similar, los operadores de servicios y comunicaciones están actualizando sus portafolios con el objetivo de ofrecer no solo la red como medio de conectividad, sino también como un servicio para satisfacer las necesidades de clientes tradicionales, industriales y empresariales. Esta situación creó un entorno ideal para que terceros, conocidos como “verticales”, utilicen la infraestructura de comunicaciones para centrarse en servicios y aplicaciones específicos, de valor añadido, de alta calidad y fácilmente accesibles en una variedad de sectores, incluyendo negocios, industria, entretenimiento, militar y emergencias. Ejemplos de esta tendencia son tecnologías emergentes como Inteligencia Artificial (AI, en inglés), metaverso, los enfoques basados en la nube, Web3, Industria 4.0 y 5.0, y los dispositivos conectados, entre otros [28]. Asimismo, se han vislumbrado oportunidades para los proveedores de dispositivos, diversificando el mercado con una gran variedad de equipos de diferentes niveles de complejidad, capacidades y costos.

Más allá del uso de la red, los datos generados por todos estos factores mencionados se convierten en un tema crítico en cuanto a cómo gestionar este volumen de información mientras se proporciona una experiencia de alta calidad a los clientes. En este sentido, algunos de los proveedores de tecnología más importantes del mundo han pronosticado en los últimos años que para el 2030 el número de suscripciones móviles será de aproximadamente 9,3 a 10,3 mil millones

[29, 30]. Este pronóstico indica que la tasa de crecimiento anual compuesta (CAGR) del tráfico será cercana al 22%, alcanzando un volumen global de hasta 468 EB por mes, lo que representa 46,4 GB por usuario en el mismo periodo. En estos estudios, se espera que para 2030 los principales contribuyentes de tráfico serán los servicios de video con hasta un 60% de la participación, seguidos de las aplicaciones de metaverso con un 13%, las redes sociales con un 10%, los videojuegos con un 6% y el comercio electrónico con un 4%. En comparación con las tendencias actuales, donde el video representa el 72% del tráfico, esta participación se modificará debido al impacto de la Realidad Extendida (XR) y los servicios de Cloud Gaming (CG).

Como consecuencia de estas necesidades nunca antes vistas, los Proveedores de Service de Conectividad (CSPs), anteriormente conocidos como operadores de redes, están rediseñando sus modelos de negocio para ofrecer la *Infraestructura como Servicio* (IaaS), permitiendo así a clientes de diferentes tipos desplegar experiencias de extremo a extremo (E2E) de alta calidad y multipropósito de manera transparente. Sin embargo, asegurar tales niveles de calidad no es una tarea trivial para los Operadores de Redes Móviles (MNOs), considerando la variedad de requisitos específicos de los servicios y los escenarios heterogéneos que demandan un grado de automatización en la red que reduzca los costes de capital (CAPEX) y operación (OPEX).

Esta necesidad fue abordada inicialmente por la Next Generation Mobile Networks Alliance (NGMN) al proponer casos de uso del enfoque de *Self-Organized Networks* (SON) [31], posteriormente estandarizados en la Release-8 por el *3rd Generation Partnership Project (3GPP)* y desarrollados para la tecnología *Long Term Evolution (LTE)*, aunque también se planificó compatibilidad retroactiva para 3G [32]. En esencia, esta tecnología fue concebida para automatizar funciones que tradicionalmente eran gestionadas por los operadores de red, incluyendo la autoconfiguración, la autooptimización y la autorreparación. Esta automatización dotó a la infraestructura de un grado de inteligencia, beneficiándose de capacidades para adaptarse a escenarios adversos y mejorar el rendimiento general [33].

Este paradigma captó la atención de la comunidad de investigación, incluyendo los grupos de I+D del sector tecnológico, quienes presentaron una gran cantidad de propuestas para abordar los casos de uso de SON [34]. El objetivo clave era aprovechar el conocimiento experto para establecer diversos algoritmos de gestión automática de redes. No obstante, a medida que los estándares de la red avanzaron

hacia LTE-A Pro, el Aprendizaje Máquina (ML en inglés) fue considerado como el motor tecnológico perfecto para superar la complejidad introducida por la variedad de nuevos servicios y dispositivos, asegurando la escalabilidad en la gestión de la red [35].

Con la llegada del 5G, el concepto de SON ha evolucionado hacia el concepto de SON de nueva generación (NG-SON). Esta propuesta pretende avanzar en los algoritmos hacia mecanismos inteligentes que puedan mejorar el rendimiento de la red en todas las situaciones, diferenciándose de los casos de uso específicos de SON en LTE [36]. Esta estrategia aprovecha la creciente cantidad de datos que la red recoge, proporcionando así a la red un conocimiento del contexto que permite escalabilidad, flexibilidad y resiliencia [37, 38].

A pesar de la discusión centrada en el rendimiento de la red, se prevé que las redes inteligentes sean más flexibles y potentes en los próximos estándares de las redes de nueva generación 5G y 6G. En este sentido, uno de los factores más relevantes que se considera es la QoE percibida por el cliente, avanzando en la gestión desde el rendimiento de la red hacia enfoques basados en el servicio [39]. Este planteamiento pretende transformar las redes móviles convencionales en infraestructuras flexibles y extensibles que puedan ser aprovechadas por usuarios finales, empresas, verticales, Proveedores de Servicio de Aplicación (ASPs), etc.

Para cumplir con las expectativas de los usuarios, se requiere un conocimiento E2E del servicio; sin embargo, aún existe un vacío en esta área, ya que las investigaciones responden a problemas específicos pero no crean una solución integral para la gestión de redes. Investigaciones anteriores se han llevado a cabo en relación con ML como habilitador tecnológico para métricas centradas en la red, como Indicador de Calidad de canal (CQI) [40], la capacidad polo [41], así como la predicción del volumen de tráfico [42, 43]. Asimismo, la información contenida en el contexto de las comunicaciones (por ejemplo, aglomeraciones y eventos sociales) ha sido empleada en [44]. Igualmente, se ha utilizado ML para abordar temas de balanceo de carga, gestión de movilidad y ahorro de energía [45].

Para abordar el enfoque QoE, se pueden realizar análisis objetivos y subjetivos en función de la métrica utilizada como Figura de Mérito (FoM). Por ejemplo, investigaciones previas se han centrado en el mapeo de *Mean Opinion Score* (MOS) a partir de Indicadores Principales de Rendimiento (KPIs) de la red [46], Indicadores Principales de Calidad (KQIs) a partir de KPIs de la red [47], o la estimación de

KQI a partir de estadísticas a nivel de paquetes [48], y estimación de KQI a partir de métricas de radio [49].

Como se destacó anteriormente, la necesidad de estrategias que involucren todos los segmentos de la comunicación entre los clientes y los ASPs, la provisión de servicios y la percepción de la experiencia del usuario es fundamental para cumplir con los objetivos especificados para las redes móviles de nueva generación. Este enfoque, conocido como gestión E2E, tiene como objetivo abordar los desafíos introducidos por los requisitos novedosos, heterogéneos, estrictos y específicos que se prevén para los casos de uso de próxima generación [50].

A.2 Preámbulos

Esta tesis se ha llevado a cabo dentro del grupo *Mobile and Aerospace Networks Lab research group (MobileNet)*, que forma parte del Instituto de Investigación en Telecomunicaciones de la Universidad de Málaga (*TELMA*).

MobileNet se estableció en el año 2000 a través de una colaboración entre el grupo TIC-102 y Nokia Networks, lo que dio lugar a la creación del Centro de Investigación en Comunicaciones Móviles en el Parque Tecnológico de Andalucía (PTA) en Málaga.

Uno de los principales ámbitos de investigación de *MobileNet* es la aplicación de AI/ML a las redes de comunicaciones móviles. En colaboración con Nokia Networks - España, uno de los primeros proyectos del grupo consistió en desarrollar una herramienta automatizada para resolver problemas relacionados con la RAN. Este proyecto sentó las bases para integrar datos reales de redes móviles con experiencia en ingeniería, con el objetivo de crear soluciones automatizadas para la resolución de problemas.

Desde entonces, el grupo ha participado activamente en el desarrollo de técnicas de SON mediante colaboraciones con empresas nacionales e internacionales en diversos proyectos de investigación. Entre estos, el proyecto *H2020 MONROE* [51] tenía como objetivo crear una plataforma de hardware de código abierto para la experimentación y medición en redes móviles comerciales de banda ancha, conocidas como nodos MONROE. Dentro de este proyecto, *MobileNet* presentó la propuesta *eSON* como parte de las convocatorias abiertas de *MONROE*, utilizando datos de la capa de radio e información de configuración

de la red. Los nodos MONROE se emplearon para recopilar datos completos de extremo a extremo desde dispositivos de usuario, al tiempo que accedían y controlaban parámetros de configuración de la red de radio.

Posteriormente, *MobileNet* también contribuyó al proyecto *H2020 ONE5G* [52]. Esta investigación se centró en desarrollar estrategias genéricas de optimización de rendimiento E2E para redes móviles de próxima generación, con el objetivo de mejorar tanto la eficiencia de los operadores de red como la experiencia de los usuarios finales mediante optimizaciones de radio. En la misma línea, el grupo ha estado involucrado en los proyectos Piloto-5G y AECMA-5G. Ambos proyectos han establecido la base de esta tesis en el contexto de la gestión de servicios E2E para redes 5G.

Además, *MobileNet* contribuyó al proyecto *H2020 ONE5G* [52]. Este proyecto de investigación estaba enfocado en el desarrollo de estrategias genéricas de optimización de rendimiento extremo a extremo (E2E) para redes móviles de próxima generación. El objetivo buscaba mejorar tanto la eficiencia de los operadores de red como la experiencia del usuario final a través de optimizaciones en el sistema de radio. De manera similar, el grupo ha estado involucrado en los proyectos “*Red.es Piloto 5G-Open RAN*” y *AECMA-5G*. Los proyectos mencionados han establecido los cimientos de esta tesis en el contexto de la gestión de servicios desde una perspectiva extremo a extremo (E2E) para redes 5G.

En estos proyectos, la infraestructura de *MobileNet* desempeñó un papel fundamental en los experimentos de prueba de concepto. Estos se llevaron a cabo en una red LTE comercial conocida como *UMAHetNet*, que consta de 12 picoceldas con un EPC [53] y una red 5G-SA que incluye 3 macroceldas y 6 picoceldas interiores [54], además de una solución *Network-in-a-Box* que integra SDR y un núcleo virtualizado [12]. La interacción con las redes es posible mediante el uso de APIs que exponen las funcionalidades de red, permitiendo así, el uso de mecanismos inteligentes externos desarrollados en el grupo.

En el contexto de investigaciones previas sobre la gestión de servicios, *MobileNet* ha llevado a cabo desarrollos relacionados con servicios tradicionales como la transferencia de archivos [55] y la navegación web. De manera similar, el equipo ha realizado investigaciones en el ámbito de los servicios multimedia, incluyendo Vídeo en demanda (VOD) mediante plataformas como YouTube o vídeo en vivo en Twitch [20]. Sin embargo, la aparición de servicios sofisticados

(por ejemplo, sistemas inmersivos) plantea nuevos escenarios que aún no han sido completamente explorados, abriendo así nuevas vías de investigación en el contexto de las redes de próxima generación.

A.3 Objetivos

Esta tesis está motivada en la brecha existente en torno a la gestión de redes basada en servicios en el contexto de una revolución tecnológica móvil que busca transformar el paradigma actual. A medida que la próxima generación de redes móviles evoluciona hacia una infraestructura escalable, flexible, inteligente, resiliente y de uso múltiple, se anticipa la llegada de casos de uso y aplicaciones innovadores. Además, la próxima generación de redes está diseñada para facilitar el desarrollo de servicios y escenarios avanzados, incluyendo casos de uso del multiverso, IIoT, videojuegos, almacenamiento y computación en la nube, entre muchas otras aplicaciones propuestas.

Dado el nivel actual de madurez tecnológica y las aplicaciones del estado del arte (SoTA) existentes, resulta difícil para los operadores proporcionar servicios de alta calidad utilizando únicamente estrategias tradicionales basadas en reglas específicas, algunas consideradas obsoletas en ciertos debates de organismos de estandarización. Esto abre la puerta a la aparición de diversos habilitadores tecnológicos, entre los cuales se destacan el AI/ML, que buscan aprovechar la información en constante evolución para identificar patrones, comportamientos y datos que puedan emplearse para la automatización inteligente en la gestión de redes.

En consecuencia, el objetivo general de esta tesis es presentar un mecanismo orientado a la calidad para la gestión de redes de próxima generación en el contexto de los servicios multimedia avanzados.

Para abordarlo, el objetivo principal se dividirá en subobjetivos que aborden los diversos retos que de él se derivan. En primer lugar, es de vital importancia analizar estos novedosos casos de uso con vistas a comprender su naturaleza, los requisitos que conllevan y cómo caracterizar su rendimiento en términos de una visión centrada en el usuario. Con este fin, el Objetivo 1 examinará la integración y despliegue de servicios de nueva generación dentro del contexto de las redes de próxima generación, con miras a asegurar la entrega de experiencias de alta calidad que cumplan con las expectativas tanto de los usuarios finales como de los CSPs. No obstante, se debe

tener en cuenta que existe una amplia gama de servicios dentro de esta categoría, por lo que esta tesis pondrá un énfasis particular en aquellos derivados de la tecnología XR.

El término *XR* engloba una variedad de tecnologías, tales como VR, AR y MR. La principal distinción entre estas tecnologías radica en la medida en que facilitan la introducción de elementos de inmersión y modificación de la realidad en relación con la realidad física del usuario. Esta situación representa un desafío considerable para los CSPs, dado que deben abordar un conjunto complejo de requisitos simultáneamente. Por ejemplo, ciertas aplicaciones de VR requieren altos niveles de rendimiento en el enlace descendente y latencias estables, mientras que AR puede priorizar el rendimiento en el enlace ascendente y la fiabilidad en la comunicación.

Del mismo modo, CG es un servicio que se define por la delegación de la lógica del juego a un servidor remoto que actúa en nombre del usuario. Con esta consideración se pretende facilitar la prestación de servicios de juego en una serie de dispositivos asequibles y heterogéneos, como teléfonos móviles, ordenadores y tabletas. Esta evolución del sistema requiere el cumplimiento de requisitos particulares relativos tanto al enlace descendente como al ascendente, con el objetivo de garantizar una latencia coherente y una calidad visual adecuada.

Para abordar esta necesidad, el Objetivo 1.1 se centrará en las principales métricas que reflejan el rendimiento del servicio desde una perspectiva E2E. Para ello, se seleccionarán los KQIs más adecuados de acuerdo con las directrices establecidas por los principales organismos de normalización, incluidos 3GPP, ETSI y ITU. Posteriormente, el diseño, desarrollo e implementación de clientes específicos de servicio son esenciales para lograr este objetivo. Los clientes son piezas de software que integran funcionalidades para prestar servicios en equipos de usuario final. Además, miden y recopilan métricas de capa alta (es decir, de aplicación) que reflejan el E2E rendimiento del sistema. Esto permite el despliegue de la exploración y experimentación controlada de servicios mediante el uso de bancos de pruebas experimentales que utilizan redes móviles.

Por otro lado, el Objetivo 1.2 adoptará un enfoque integral para examinar la influencia de diversos factores de red en el rendimiento del servicio. Aunque las redes inalámbricas ofrecen el potencial de mejorar la movilidad, flexibilidad y escalabilidad en la provisión de servicios, los desafíos inherentes asociados a

factores radioeléctricos (e.g., propagación, obstáculos, compartición de recursos) pueden resultar en una disminución en el rendimiento. Esta realidad puede obstaculizar los esfuerzos de diversos sectores y proveedores de equipos para acelerar el desarrollo de servicios y aplicaciones compatibles con estas tendencias tecnológicas emergentes. Para demostrar la viabilidad de una provisión de servicio adecuada a través de redes móviles, se evaluarán campañas de medición en escenarios controlados usando diversas condiciones radioeléctricas, tecnologías inalámbricas y movilidad a través de las plataformas experimentales desarrolladas para el Objetivo 1.1.

Ante la heterogeneidad de los requisitos de los servicios de próxima generación, surge otro desafío para los operadores: la necesidad de gestionar redes de una manera impulsada hacia los servicios y las expectativas de los clientes, en lugar de las métricas de rendimiento centradas en la red. Además, estas arquitecturas están evolucionando hacia esquemas descentralizados, avanzando desde equipos monolíticos hacia enfoques basados en software que aportarán flexibilidad y escalabilidad a la infraestructura. Asimismo, se consideran nuevos motores tecnológicos, incluyendo nuevas bandas de radio, esquemas dinámicos de asignación de recursos, numerologías y AI/ML, entre otras características. Estos elementos están destinados a permitir la adaptación de redes lógicas diseñadas para casos de uso y demandas específicos. Sin embargo, esto añade complejidad a la red y presenta desafíos inéditos que deben abordarse.

Para superar estas situaciones, el Objetivo 2 investiga la aplicación del aprendizaje automático para desarrollar mecanismos inteligentes para la gestión de redes basada en servicios, aprovechando los datos disponibles del lado del operador.

Anteriormente, el Objetivo 1 hizo hincapié en la importancia de cuantificar el rendimiento del servicio a través de métricas objetivas y exhaustivas que reflejen el rendimiento global del sistema. Sin embargo, en ciertas circunstancias, no es factible cuantificar esto debido a cuestiones que afecten la privacidad de los usuarios o la falta de mecanismos por parte de los ASPs que faciliten la provisión de esta información. Para abordar este desafío, el uso de ML es imperativo para predecir esta información crucial, la cual luego alimentará los mecanismos de gestión de red. El Objetivo 2.1 investigará las ventajas de entrenar modelos de predicción de KQIs utilizando datos accesibles en la red, incluyendo métricas de radio, estadísticas y configuraciones reportadas por varios segmentos de la red, con excepción del equipo terminal del

usuario. Esta propuesta permite la replicación del comportamiento del servicio y la implementación de las acciones necesarias para garantizar el cumplimiento de los requisitos establecidos en los acuerdos de nivel de servicio (SLAs en inglés). El SLA representa un conjunto de normas e indicadores medibles que regulan el rendimiento de un CSP de acuerdo con las expectativas de calidad de servicio de los clientes.

Además de la capacidad para modelar la funcionalidad del servicio, el desarrollo de una estrategia eficaz para asegurar la provisión de servicios de alta calidad requiere la implementación de soluciones avanzadas que aprovechen las ventajas ofrecidas por el ML. Por lo tanto, el objetivo 2.2 perseguirá el desarrollo de un mecanismo E2E, basado en servicios entre capas, para facilitar la gestión de las redes, en relación con la asignación de recursos basada en servicios, al tiempo que se optimiza su rendimiento en términos de cumplimiento de los KQIs. A la luz de las consideraciones mencionadas, es factible garantizar la provisión de servicios de alta calidad en una multitud de circunstancias y escenarios, según las características intrínsecas de los datos accesibles en la red. Del mismo modo, la aplicación de ML permite a la infraestructura implementar soluciones flexibles y escalables adaptadas a diversos casos de uso con requisitos variables en producción. Esto facilita la provisión de una infraestructura inteligente, transparente y de usos múltiples como servicio para verticales, ASPs y clientes finales.

En resumen, los objetivos de esta tesis son los siguientes:

- **Objetivo 1:** Explorar los servicios emergentes destinados a ser desplegados en redes móviles de nueva generación. En particular, este objetivo se divide en dos sub-objetivos:
 - **Objetivo 1.1:** Desplegar servicios y medir su rendimiento desde una perspectiva E2E. La gestión de redes basada en servicios requiere la recolección de métricas que reflejen el rendimiento del servicio desde un punto de vista holístico, permitiendo así una monitorización del rendimiento impulsada por datos de capas alta.
 - **Objetivo 1.2:** Evaluar los factores de red que afectan la entrega del servicio E2E. Cada segmento de la cadena de comunicación impacta en la entrega del servicio. Identificar cómo la red afecta el rendimiento del servicio bajo diferentes escenarios es de suma importancia para determinar la forma más adecuada de abordar los efectos negativos.

- **Objetivo 2:** Establecer mecanismos inteligentes basados en datos para la gestión de redes E2E basada en servicios. Para abordar este objetivo, se proponen los siguientes sub-objetivos:
 - **Objetivo 2.1:** Aprovechar los ricos datos E2E y los mecanismos ML para predecir el rendimiento del servicio. Dado que no se puede garantizar el acceso a los datos de rendimiento del servicio del usuario final por razones de privacidad y limitaciones en la implementación de funciones de medición en las aplicaciones/agentes específicos del servicio, deben emplearse enfoques alternativos. ML puede aprovechar la información disponible en la red para predecir los KQIs que reflejan el grado de satisfacción de los requisitos del servicio.
 - **Objetivo 2.2:** Optimizar la entrega del servicio E2E mediante enfoques de red inteligentes. Los enfoques basados en ML son adecuados para habilitar la toma de decisiones orientada a KQI en la gestión de redes basada en servicios. El conocimiento permanente y dinámico del rendimiento del servicio puede acelerar la determinación de los recursos adecuados para cumplir las diversas políticas de acuerdos de nivel de servicio de los servicios de nueva generación.

A.4 Descripción de las publicaciones

Esta sección presenta las contribuciones, en forma de artículos científicos, que responden a los objetivos y desafíos definidos en la Sección 4. Como se ilustra en la Figura 3.2, el mapeo lógico entre los desafíos, objetivos y resultados de la tesis identificados. Los artículos en cuestión han sido publicados en revistas consideradas relevantes e indexadas en el *Journal Citation Reports* (JCR), donde resumen los resultados logrados a través de la metodología mencionada. En este sentido, los objetivos establecidos se reflejan en publicaciones que se alinean con los desafíos identificados y contribuyen al logro de hitos clave que abordan los requisitos especificados.

En consecuencia, los artículos resultantes se han organizado para cubrir tres capítulos bien definidos. El Capítulo 4 define la exploración de servicios de nueva generación considerando los requisitos y desafíos de diseño, y el proceso de desarrollo de entornos que permiten la experimentación controlada. Luego, 5 cubre

la metodología aplicada para simplificar la medición y monitoreo de servicios de nueva generación aprovechando el enfoque de ML. Para finalizar, 6 presenta la explotación del paradigma KQI para la gestión de redes basadas en servicios.

La siguiente sección presenta una distribución lógica y síntesis de los artículos, incluyendo un mapeo de los desafíos, objetivos y hitos de la tesis.

A.4.1 C1. Measuring key quality indicators in cloud gaming: Framework and assessment over wireless networks

El artículo examina los juegos en la nube como uno de los servicios multimedia más avanzados, que se prevé serán soportados por la próxima generación de redes móviles (Objetivo 1). Este trabajo se enmarca en el Capítulo 4.

En primer lugar, esta investigación introduce el concepto de CG y analiza los diversos conceptos arquitectónicos que definen los servicios y sus requisitos. El concepto fundamental que subyace en esta solución es la utilización de clientes “ligeros”, que reducen la carga computacional en dispositivos de juegos tradicionales (por ejemplo, ordenadores con GPU) mediante el uso de dispositivos de menores prestaciones. Dichos equipos son capaces de capturar las acciones del usuario y transmitirlas a un servidor en la nube, que luego ejecuta la lógica del juego en nombre del usuario. Posteriormente, el feedback visual se procesa en el servidor y se transmite al usuario a través de datos multimedia, específicamente video y audio.

Para lograr el Objetivo 1.1, se presenta el diseño, desarrollo e implementación de un cliente de CG. Este diseño incorpora el desarrollo de una metodología que facilita la interacción automatizada del usuario (es decir, calibración de coordenadas, ejecución del cliente y emulación de acciones de ratón y teclado) con el software de juego y la recopilación de métricas de E2E. De manera similar, el cliente es capaz de configurar automáticamente varios parámetros de la sesión, como la resolución de vídeo, tasa de cuadros, codificadores/decodificadores de video y configuraciones de audio, entre otros elementos disponibles.

Dado que la gestión de redes basada en servicios se orienta hacia el rendimiento de extremo a extremo de las soluciones en cuestión, es necesario implementar una herramienta de extracción de KQI que recopile información a nivel de aplicación

desde el cliente mientras una sesión está activa. Luego, los datos se recopilan y codifican en formato *JavaScript Object Notation (JSON)* y se transmiten a un servidor *REpresentational State Transfer (REST)* central para su procesamiento.

Además, este artículo detalla la incorporación del cliente de CG en un banco de pruebas que facilita la evaluación de múltiples redes de transporte. Esta tarea tiene como objetivo determinar el impacto de las redes inalámbricas en el servicio (Objetivo 1.2). Para este fin, la evaluación considera tecnologías LTE, WiFi y Ethernet a través de campañas de medición controladas que modifican factores de referencia del servicio (es decir, resolución y tasa de cuadros). La evaluación analiza la degradación asociada con los KQIs del servicio generada por la utilización de redes inalámbricas en comparación con Ethernet.

En esta línea, este trabajo proporciona información clave sobre el despliegue de CG a través de redes inalámbricas en reemplazo de las tecnologías cableadas tradicionales (es decir, Ethernet). Esto es un habilitador importante considerando que la movilidad en este tipo de servicio es fundamental para proporcionar experiencias más avanzadas en situaciones no estáticas, abriendo el camino para enfoques evolucionados como aplicaciones y casos de uso de XR.

A.4.2 C2. KQI assessment of VR services: A case study on 360-video over 4G and 5G

Este resultado de investigación se presenta en el Capítulo 4 y se centra en investigar el uso de servicios de video de 360 grados como una de las aplicaciones emergentes de XR, que aglutina muchas otras tecnologías. Este tipo de servicio pertenece a la categoría de tecnología inmersiva que ha atraído atención recientemente debido al surgimiento del concepto del metaverso a través de redes inalámbricas.

En este contexto, este artículo examina las diferencias arquitectónicas entre el servicio de vídeo tradicional y la tecnología de VR, que requiere equipo especializado, como HMDs, para proporcionar una experiencia inmersiva a los usuarios. A la luz de estas consideraciones, surgen nuevos desafíos en cuanto a la entrega de niveles de servicio de alta calidad, donde la movilidad emerge como un factor crucial. De manera similar, se discuten los requisitos asociados con la tecnología de VR, considerados exigentes en términos de calidad visual y latencia. La falta de recursos puede resultar en la manifestación de incomodidad física y

ciberenfermedad (cybesickness en inglés).

Este artículo aborda la investigación teórica del servicio de video 360, incluyendo sus requisitos, arquitectura y el actual estado del arte. Se presenta además el desarrollo de un cliente DASH/video 360 específico para el servicio, utilizando Unity 3D como motor gráfico. El cliente es capaz de realizar múltiples tareas simultáneamente, incluyendo el renderizado de video omnidireccional en una esfera, la detección de movimientos del usuario y la recolección de datos, incluyendo KQIs como resolución de vídeo, latencia, tasa de cuadros, estado de búfer y otros. El desarrollo preliminar de este trabajo responde al Objetivo 1.1.

Consecuentemente, el trabajo aborda el Objetivo 1.2 identificando los factores de red (que afectan la entrega del servicio E2E) que representan desafíos para ser enfrentados por las redes móviles. Para abordar esto, se evalúa el rendimiento de las redes LTE y 5G utilizando un banco de pruebas de vídeo 360 que comprende un cliente, un servidor RESTful y una solución de red network-in-a-box [12] que despliega el servicio en escenarios realistas. El servidor REST coordina los datos desde el lado del cliente y la red (es decir, mediciones de radio, estadísticas y parámetros de configuración). Este marco evalúa el rendimiento del servicio bajo diferentes condiciones de radio, considerando casos en el borde de la celda y mínima interferencia, demostrando así el potencial de las redes de próxima generación para ofrecer servicios inmersivos.

A.4.3 C3. Cloud VR on 5G: A Performance Validation in Industrial Scenarios

En concordancia con la investigación mencionada, este artículo refuerza la necesidad de reconocer contextos de servicios de vanguardia para preparar adecuadamente las redes futuras para su despliegue e integración. Este trabajo, que se encuentra enmarcado en el Capítulo 4, explora el servicio de VR en la nube (Cloud VR). Este servicio se basa en los enfoques de VR existentes al aprovechar los beneficios de CG. Estas mejoras se caracterizan por una reducción en el costo del equipo del usuario, lograda mediante el uso de un servidor en la nube responsable de renderizar la retroalimentación visual después de recibir las interacciones del usuario y la información de los sensores, lo cual es posible gracias a una arquitectura y hardware simple.

La principal ventaja del servicio de VR en la nube es su versatilidad, lo que lo hace atractivo para una variedad de sectores industriales y CSPs. Su capacidad para ser distribuido fácilmente a una multitud de dispositivos es un beneficio significativo, particularmente dado que los datos en DL se caracterizan por ser multimedia genérica. Sin embargo, desagregar el procesamiento del dispositivo del usuario también presenta un desafío considerable. Es esencial asegurar que la latencia introducida por el transporte del contenido sea mínima para evitar cualquier degradación en la experiencia inmersiva. Dado que el consumo de contenido a través del HMD requiere consideraciones aún más rigurosas que el contenido en CG, este servicio requiere una atención particular.

Para abordar esta necesidad, en este trabajo de investigación se evalúa la entrega de la solución de VR en la nube en un entorno industrial real ubicado en la Universidad de Aalborg (AAU). Este escenario se selecciona considerando el amplio portafolio de servicios de XR que buscan cubrir diversos casos de uso de la Industria 4.0 y el próximo enfoque 5.0 (por ejemplo, asistencia remota, operación remota de robots, capacitación e inspección segura, entre otros). En este sentido, el objetivo es identificar los principales problemas y desafíos que las redes móviles enfrentarán y los beneficios que los servicios pueden obtener al utilizarlas. Esta exploración incluye el diseño e implementación de un cliente/servidor de VR en la nube integrando Unity 3D, el software ALVR y Steam VR (Objetivo 1.1).

Para evaluar el rendimiento, se seleccionan las tecnologías 5G y WiFi para proporcionar conectividad a un HMD de VR. El cliente y el servidor recogen KQIs y trazas que describen el rendimiento del servicio a través de la experimentación. Los resultados muestran que las redes móviles tienen fortalezas importantes en el mantenimiento de una respuesta de latencia constante, aunque a un costo de aumentar la latencia media en comparación con WiFi. Además, se evalúan los efectos de la movilidad a lo largo de un recorrido definido dentro de las instalaciones que proporciona información valiosa sobre la degradación del rendimiento del servicio. Esto es especialmente notable en WiFi, cuyo rendimiento de latencia y rendimiento se ven fuertemente afectados mientras que el de 5G sigue siendo adecuado. Los resultados están destinados a acercarse al Objetivo 1.2.

Para complementar el estudio, se realiza un examen de la QoE para determinar el grado de calidad asociado con la perspectiva del usuario. Para ello, se construyen modelos analíticos de QoE que reflejan objetivamente la experiencia del usuario basada en la calidad visual, latencia e interrupciones de vídeo (es decir, pérdida de

fotogramas). Los hallazgos demuestran que 5G ofrece un rendimiento de VR en la nube de calidad superior en comparación con WiFi en escenarios con movilidad. Este es un resultado notable del trabajo, que demuestra la viabilidad de implementar casos de uso industrial de XR y establece una base para futuras investigaciones en este campo para todas las partes involucradas en el desarrollo de los estándares 5G.

A.4.4 C4. ML-powered KQI estimation for XR services: A case study on 360-video

La llegada de tecnologías y servicios avanzados, como la Realidad Extendida (XR), está preparada para transformar la vida cotidiana. Simultáneamente, nuevos enfoques arquitectónicos descentralizados han sentado las bases para redes móviles de próxima generación como 5G, mientras establecen el escenario para B5G y desarrollos futuros. Se anticipa que estas redes serán habilitadores clave para concretar el metaverso y otros servicios visionarios. En este contexto, cuantificar con precisión el rendimiento del servicio es esencial para la gestión de redes dinámica, adaptativa y proactiva.

La evaluación del rendimiento de los servicios es una tarea difícil, sobre todo porque depende de la opinión subjetiva de los usuarios, lo que obliga a recopilar datos que pueden estar sesgados. Además, para obtener una percepción genuina de los clientes, esta metodología requiere la participación de múltiples personas por categoría y servicio. En consecuencia, esta estrategia no es escalable. Por otro lado, medir KQIs en el equipo del usuario puede resultar en violaciones de la privacidad del usuario o en la falta de disponibilidad de la información necesaria. Para abordar estos desafíos, la estimación de KQI mediante ML representa una solución prometedora, ya que permite la explotación de una gran cantidad de datos disponibles de CSP.

En este sentido, la estimación de KQIs, como métricas objetivas E2E que reflejan la percepción del usuario sobre el servicio, es uno de los pilares clave que constituyen esta tesis. Consecuentemente, el artículo presentado cubre el desafío establecido en el Objetivo 2, especialmente con respecto al Objetivo 2.1. Asimismo, la metodología y los resultados responden a la estructura de la tesis, en particular al Capítulo 5.

Este artículo describe un marco basado en ML que aprovecha los datos accesibles en la red, incluidas mediciones de radio, estadísticas y parámetros de configuración,

para identificar modelos de ML óptimos para predecir varios KQIs de XR. El marco incorpora técnicas de ingeniería de características, FE y FS, que mejoran la eficiencia y el rendimiento del modelo al agregar técnicas que pueden aumentar la capacidad del modelo para capturar las relaciones y patrones más adecuados. De manera similar, este enfoque explora una variedad de algoritmos en busca del más adecuado que mejor se ajuste a la naturaleza intrínseca de los datos, incluyendo *Random Forest (RF)*, *K-Near Neighbors (KNN)*, *Support Vector Machine (SVM)*, *AdaBoost*, *Multi-Layer Perceptron (MLP)*, *Ridge Regression (RR)*, entre otros, en sus esquemas de regresión y clasificación para variables continuas o discretas.

Más allá de esta estrategia, el marco presentado ensambla múltiples etapas de ML, como preprocesamiento de datos, definición de modelos, entrenamiento, ajuste y validación, en una estrategia automática impulsada por “pipelines” que garantiza la escalabilidad. Además, se propone una métrica novedosa, el Puntaje de Error de Rendimiento y Tiempo de Predicción (PET_{score}), para evaluar y seleccionar el algoritmo más adecuado, equilibrando tanto la tasa de error como el tiempo de predicción, crucial para servicios sensibles al tiempo como XR.

Para su validación, el marco se aplica para abordar el servicio de vídeo 360 grados, demostrando su efectividad con un caso de uso real de XR. Este servicio ha sido seleccionado ya que su rendimiento depende principalmente de la red en lugar de la capacidad de procesamiento del equipo. Para evaluar el nivel de logro del marco, se ha generado un conjunto de datos que refleja el comportamiento del servicio bajo diferentes condiciones de canal y recursos otorgados al usuario. El conjunto de datos resultante de la experimentación está disponible públicamente y completamente referenciado en el artículo [19], dotando así a la comunidad investigadora con herramientas que pueden acelerar investigaciones relacionadas.

Los resultados demuestran las capacidades de esta propuesta al predecir los KQIs de vídeo 360 con una precisión adecuada. Esto permite que las redes monitoreen inteligentemente la entrega del servicio utilizando una perspectiva holística. Además, el uso de enfoques de ML asegura un despliegue escalable y flexibilidad para soluciones adaptadas al servicio que fortalecen la gestión de redes de próxima generación.

A.4.5 C5. Towards KQI/ML-based management: Optimizing radio resource allocation for next-generation networks

El Capítulo 6 finaliza el trabajo de investigación realizado en la tesis al presentar una propuesta para la gestión de redes basada en servicios. El desarrollo de este artículo abarca el Objetivo 2, con un enfoque especial en la optimización de la entrega del servicio E2E (Objetivo 2.2).

Este estudio propone un enfoque innovador e integral, fundamentado en un marco mejorado por ML, que facilita la gestión de servicios basada en la calidad. Aprovechando las capacidades de ML, este marco permite una evaluación exhaustiva de la QoE E2E al utilizar datos accesibles en la red para evaluar y optimizar el rendimiento de la red y los servicios desde una perspectiva holística, yendo más allá de los métodos tradicionales centrados en la red hacia métodos centrados en el servicio y el usuario. Este mecanismo explota las funcionalidades clave proporcionadas por los desarrollos presentados en el Capítulo 4 (es decir, identificación y medición de KQI) y el Capítulo 5 (es decir, el marco de ML para la provisión automática de modelos) para crear una herramienta versátil que facilita la gestión de recursos.

El enfoque desarrollado comprende tres elementos fundamentales: adquisición de datos, monitoreo y adquisición de la calidad del servicio, y optimización. Los primeros dos componentes integran las soluciones mencionadas. Posteriormente, la etapa de optimización formula una funcionalidad basada en KQI/ML, diseñada para mejorar el rendimiento de la red mientras asegura el cumplimiento del SLA en términos de KQI. Este es un punto de consideración significativo, dado que la gestión del SLA tradicional dependía del cumplimiento de los indicadores de tasa de datos y latencia. La metodología mencionada no es viable para las redes de próxima generación que soportan servicios avanzados con una amplia diversidad de requisitos, ya que carece de la escalabilidad y flexibilidad necesarias.

En este sentido, la solución propuesta busca abordar la brecha tecnológica creada por la evolución de los conceptos arquitectónicos de redes y derivar su gestión a través de enfoques inteligentes. Para ilustrar la aplicabilidad del enfoque propuesto, se ha seleccionado el problema de asignación óptima de recursos como caso de estudio. En este contexto, el mecanismo de optimización identifica el

número adecuado de RBs que cumplen las condiciones mínimas especificadas en el SLA. Además, el mecanismo combina dos elementos clave: ML y optimización numérica. El enfoque inicial implica la generación de modelos de ML específicos del canal que representan las funciones asociadas con la asignación de recursos dentro de la red. Los modelos específicos del canal se generan automáticamente mediante el marco de ML descrito en el Capítulo 5 en combinación con una estrategia de clusterización que identifica los diferentes escenarios de radio dentro del conjunto de datos de entrada. Posteriormente, estos modelos se integran con una función de QoE específica del servicio para formar una función objetivo. Esta última se optimiza mediante una estrategia de “dual annealing” [113] con el objetivo de definir la combinación de KQIs que se asignan a los RB asignados, garantizando que se satisfagan los requisitos del servicio y se minimice el uso de recursos.

Para comprobar la eficacia de esta metodología, se evaluó el servicio de vídeo de 360 grados en condiciones de borde de celda en redes 4G y 5G. Los resultados demuestran que el optimizador identificó eficazmente la asignación de recursos necesaria para satisfacer las demandas del servicio al mismo tiempo que minimizaba la ocupación de recursos. Además, la evaluación demuestra la adaptabilidad y escalabilidad del marco, indicando su potencial para soportar una variedad de servicios emergentes (por ejemplo, XR) y arquitecturas de red futuras, como Open RAN.

A.5 Conclusiones

Esta tesis ha abordado el desafiante escenario de la implementación de servicios multimedia avanzados mediante el uso de redes móviles. La provisión de estos servicios conlleva requisitos específicos que difieren significativamente de los asociados a los servicios tradicionales. Por lo tanto, las redes de próxima generación requieren no solo una evolución de los conceptos arquitectónicos existentes, sino también una transición desde paradigmas de gestión centrados en la red hacia modelos basados en el servicio, que aseguren la entrega óptima de servicios a los usuarios.

El análisis del problema ha permitido la formulación de dos objetivos principales, que abordan los desafíos asociados a este contexto.

Por un lado, el Objetivo 1 propuso la provisión de servicios de nueva generación

a través de redes móviles. Para abordar este enfoque, se establecieron dos subobjetivos con la intención de analizar el escenario y los retos que deben enfrentarse. Posteriormente, el Objetivo 1.1 impulsó la investigación de servicios avanzados, conceptos arquitectónicos, sus principales métricas de rendimiento y los requisitos para su entrega efectiva a los usuarios. Esto llevó a la formulación del Objetivo 1.2, que estableció la necesidad de identificar los factores de impacto y consecuencias fundamentales del uso de redes móviles en la implementación de estos servicios desde una perspectiva E2E. Esto, a su vez, facilitará la transición de un paradigma centrado en la red a uno centrado en el usuario y el servicio.

En contraste, el Objetivo 2 se centró en el desarrollo de metodologías inteligentes para la gestión de redes basada en el servicio dentro de un marco arquitectónico E2E. Esto dio lugar a dos subobjetivos. El Objetivo 2.1 investigó el potencial de combinar datos accesibles en la red con algoritmos de ML para predecir métricas E2E que reflejen con precisión el rendimiento real del servicio, sin la influencia del usuario. Posteriormente, el Objetivo 2.2 llevó a cabo una investigación sobre la implementación de ML para desarrollar metodologías inteligentes para la gestión de redes basada en el servicio. Este objetivo buscó integrar los hallazgos del Objetivo 2.1 con otros elementos tecnológicos clave, que son fundamentales para el avance y evolución de las redes de próxima generación.

A continuación, se resumen las contribuciones relacionadas con los objetivos anteriormente mencionados.

- **Objetivo 1: Explorar los nuevos servicios que se pretenden implementar mediante redes móviles de próxima generación.**

- En consonancia con el Objetivo 1.1, se sometieron a un examen detallado tres servicios multimedia avanzados, que se prevé desempeñarán un papel importante en las redes de próxima generación. En este sentido, se examinaron las arquitecturas de servicio relacionados con CG, vídeo en 360 grados y VR en la nube, en conjunto con los requisitos asociados y el conjunto de métricas de rendimiento objetivo establecidas por los organismos de estandarización, incluyendo 3GPP, ETSI y ITU-T.

Posteriormente, para obtener una visión más allá de la literatura sobre estos servicios, se diseñaron y desarrollaron arquitecturas cliente/servidor específicas de cada servicio que reflejan el rendimiento

E2E. La implementación de estos servicios requirió completar tareas preparatorias adicionales, diseñadas para facilitar la experimentación en escenarios controlados. Dichas tareas incluyeron la automatización de acciones de usuario, el diseño de mecanismos de visualización para servicios relacionados con VR, la codificación de herramientas para la recopilación y medición de métricas y el desarrollo de mecanismos de comunicación entre clientes y un servidor REST.

- En relación con el Objetivo 1.2, las arquitecturas mencionadas se integraron con redes móviles para crear escenarios de experimentación completamente controlados. El objetivo principal de estas tareas fue identificar los factores de impacto de la red y su influencia en la calidad de la entrega del servicio desde una perspectiva E2E. Para realizar estos escenarios experimentales, se empleó la infraestructura de MobileNet, incluyendo implementaciones comerciales de redes LTE y 5G-SA y la solución “network-in-a-box” [12], así como los desarrollos previos relacionados con el Objetivo 1.1.

Las contribuciones asociadas a este objetivo se alinean con los tres servicios mencionados anteriormente: CG, vídeo en 360 grados y VR en la nube. La evaluación de estos servicios se realizó con el objetivo de identificar y analizar diferentes factores de impacto de la red. Los resultados obtenidos pretenden establecer una base para el desarrollo de estándares y casos de uso de redes de próxima generación.

En primer lugar, se investigó la degradación del rendimiento del sistema de CG al utilizar redes inalámbricas en lugar de tecnologías cableadas convencionales. Para ello, se realizaron experimentos para evaluar el impacto de diversas configuraciones asociadas al servicio, resoluciones de video y tasas de fotogramas en el rendimiento. Los hallazgos indicaron que las redes inalámbricas pueden facilitar una entrega efectiva del servicio en comparación con los resultados obtenidos con respecto a la tecnología cableada.

En segundo lugar, se realizó un análisis comparativo de las ventajas del uso de la tecnología 5G en comparación con LTE mediante el despliegue de video en 360 grados. Para cuantificar este enfoque, se realizaron campañas de medición en una variedad de condiciones de radio, incluyendo condiciones de borde de célula y escenarios de interferencia

mínima. Los resultados indicaron que 5G mostró un rendimiento superior en comparación con LTE en términos de resolución, tasa de fotogramas, latencia y tiempo de carga, además de demostrar un menor uso de recursos en frecuencia. Además, se identificaron posibles vías de investigación futura con el objetivo de mejorar la calidad del servicio aprovechando las capacidades de la tecnología 5G.

En tercer lugar, se evaluó el impacto de la incorporación de la movilidad en la entrega del servicio utilizando la VR en la nube. Para ello, se evaluaron WiFi y 5G-SA en un entorno industrial real, tomando en cuenta los efectos de la visibilidad indirecta y la interferencia debido a equipos y personal. Los resultados demostraron las ventajas de 5G al mostrar un rendimiento constante tanto en escenarios estáticos como en móviles, en contraste con WiFi, que mostró una notable disminución en este último escenario. Para proporcionar contexto adicional, se realizó una evaluación del rendimiento E2E del servicio mediante modelos paramétricos de QoE, que confirmaron la viabilidad de utilizar 5G para la implementación de estas soluciones en escenarios de la Industria 4.0 y 5.0.

- **Objetivo 2: Establecer mecanismos inteligentes basados en datos para la gestión de redes E2E centrada en servicios.**

- En concordancia con el Objetivo 2.1, se llevó a cabo una investigación para determinar los beneficios de emplear metodologías de AI/ML para la gestión de servicios E2E. Este enfoque abarca toda la red de comunicación, aunque, en ciertos casos, los datos pueden no estar accesibles para los CSPs debido a limitaciones en la implementación de servicios en dispositivos de usuario o para evitar posibles violaciones de privacidad.

Para afrontar este reto, se investigó el potencial de ML como un medio para habilitar el monitoreo de servicios mediante la estimación de KQIs a partir de los datos disponibles para los operadores.

En primer lugar, para aprovechar ML, fue necesario desarrollar campañas de medición apropiadas para conformar un conjunto de datos que abarque el rendimiento del servicio bajo diferentes condiciones de red. Para ello, se seleccionó la solución de vídeo en 360 grados para caracterizar el comportamiento de sus KQIs en escenarios de borde de

célula y de interferencia mínima con asignación progresiva de recursos. El conjunto de datos generado a lo largo de la experimentación se ha hecho público a través de IEEE DataPort [19] con el objetivo de fomentar más investigaciones en este campo.

En segundo lugar, se diseñó e implementó un marco de trabajo de aprendizaje automático para agilizar el modelado de KQIs mediante la automatización del proceso y la minimización de la intervención manual. Este marco incluye un procedimiento basado en pipeline que cubre el entrenamiento de modelos, ajuste de hiperparámetros y evaluación. Además, se incorporaron varias técnicas de ingeniería de características, como FE y FS, para mejorar la alineación con las características de los KQIs.

En tercer lugar, para seleccionar el modelo óptimo entre la variedad de algoritmos y técnicas automatizados por el marco de trabajo, se introdujo el índice de Error de Rendimiento y Tiempo de Predicción, PET_{score} . Este índice permite inferir el modelo más adecuado basado en el rendimiento de error y tiempo de predicción, proporcionando un marco flexible que permite la consideración de ambos aspectos de acuerdo con las circunstancias específicas, útil en monitoreos sensibles al tiempo.

Finalmente, el marco fue evaluado utilizando el conjunto de datos de vídeo en 360 grados, y los resultados demostraron que los modelos seleccionados predicen el nivel de rendimiento de los KQIs del servicio de manera efectiva.

- Para concluir la tesis, se abordó la necesidad de mecanismos integrados que utilicen ML para facilitar la gestión de redes basada en servicios. El enfoque propuesto comprende un marco de tres pasos: adquisición de métricas, estimación de rendimiento y optimización. Este último término tiene un carácter amplio que abarca varias soluciones potenciales, y en consecuencia, se investigó el problema de la asignación de recursos.

Primero, los pasos de adquisición y estimación del marco aprovechan los beneficios identificados en los desarrollos resultantes de los Objetivos 1.1, 1.2 y 2.1. En este sentido, la gestión basada en servicios puede habilitarse al monitorizar el servicio mismo mediante el uso de KQIs, que exponen el rendimiento E2E. En casos en los que estas métricas son

cuantificables, el CSP puede recuperar los datos requeridos y generar conjuntos de datos conformes para alimentar las soluciones de ML. Posteriormente, la estimación de métricas se emplea para facilitar la generalización del monitoreo de servicios mediante el uso de mecanismos inteligentes.

En segundo lugar, el enfoque de optimización propuesto tiene como objetivo mejorar el problema de asignación de recursos. Para abordar este, se emplea un enfoque híbrido de optimización numérica y ML para determinar el número óptimo de RBs a asignar para satisfacer requisitos específicos de KQIs. Este enfoque se basa en el desarrollo de modelos específicos de canal mediante ML y en funciones de QoE específicas del servicio. Los dos componentes se integran a través de una función objetivo que se optimiza mediante una metodología de “dual annealing” con el objetivo de minimizar el uso de recursos y maximizar el rendimiento E2E. El grado de cumplimiento de la solución se evalúa mediante una revisión de los requisitos SLA en términos de los KQIs

Los resultados han demostrado que las herramienta propuestas ofrece ventajas significativas a la hora de determinar la asignación óptima de recursos para los usuarios del servicio bajo diversas condiciones radioeléctricas y requisitos SLA. Además, la metodología empleada demuestra ventajas en cuanto a escalabilidad y flexibilidad, inherentes a los paradigmas ML. Esta estrategia integral ML pretende cubrir y facilitar diferentes aspectos de la gestión de servicios desde la adquisición, monitorización y optimización del rendimiento. Asimismo, la explotación de estos mecanismos dinámicos pretende facilitar el desarrollo de soluciones específicas de servicio adaptadas a los requisitos de las redes de próxima generación.

En conclusión, esta tesis proporciona una contribución significativa en la transición hacia redes móviles de próxima generación mediante la implementación de servicios avanzados y la gestión inteligente E2E basada en servicios.

A.6 Actividades de investigación

Esta sección describe las actividades de investigación realizadas como parte de esta tesis. Esto incluye la difusión de los resultados, los proyectos de investigación en los

que el autor ha participado y las estancias de investigación que han contribuido a esta tesis.

A.6.1 Difusión de los resultados

Los resultados de esta investigación se han divulgado a través de diversas publicaciones en revistas científicas y congresos de prestigio, como se indica al principio del manuscrito.

No obstante, las actividades de difusión no se han limitado al ámbito de las publicaciones científicas, sino que también han abarcado demostraciones en eventos dentro del ámbito tecnológico. A través de estas demostraciones, se compartió el enfoque innovador y los resultados de la línea de investigación de esta tesis, facilitando un valioso intercambio de conocimientos y experiencias con la comunidad académica y profesional. Dichos eventos incluyeron:

- Mobile Week Málaga 2021, organizado por Palacio de Ferias y Congresos de Málaga, Diciembre 2021.
- 12da Reunión Europea en Ciencia, Tecnología, e Innovación, organizado por TRANSFIERE (Foro Europeo para la Ciencia, Tecnología e Innovación), en Málaga, España, Febrero 2023.
- 13ra Reunión Europea en Ciencia, Tecnología, e Innovación, organizado por TRANSFIERE (Foro Europeo para la Ciencia, Tecnología e Innovación), en Málaga, España, Marzo 2024.

A.6.2 Proyectos relacionados

Esta tesis ha sido financiada por la Junta de Andalucía a través de la Secretaría General de Universidades, Investigación y Tecnología mediante la Beca predoctoral PREDOC_01712;

El trabajo desarrollado en esta tesis ha contribuido a los siguientes proyectos:

- **Piloto 5G Red.es Andalucía, Case Study OpenRAN:** Development of a 5G access network prototype based on GPPs. Project number 8.06/5.59.5630, funded by Ministerio de Asuntos Económicos y Transformación Digital and the European Regional Development Fund (ERDF), Julio 2019 - Octubre 2021.

- **AECMA-5G:** Advanced E2E Cellular Management for 5G Applications. Project number UMA-CEIATECH-14, financiado por la Junta de Andalucía y el Fondo Europeo de Desarrollo Regional (FEDER), Junio 2020 - Diciembre 2021.
- **PENTA:** Provisión de servicios PPDR a través de Nuevas Tecnologías de Acceso radio, financiado por la Junta de Andalucía, Febrero 2020 - Febrero 2023.
- **MAORI:** Massive AI for the OpenRadio b5G/6G network. Project number TSI-063000-2021-75, financiado por el Ministerio de Asuntos Económicos y Transformación Digital y la Unión Europea - NextGenerationEU dentro del marco “Recuperación, Transformación, y Resiliencia”, Enero 2022 - Diciembre 2024.

Además, el autor ha aportado contribuciones a los siguientes proyectos:

- **SMART:** Sistema de Monitorización y análisis Automático de Redes de Telecomunicación. Proyecto AT17-5882-UMA, financiado por la Junta de Andalucía dentro del marco “Plan Andaluz de Investigación, Desarrollo e Innovación (PAIDI 2020)”, Octubre 2019 - Septiembre 2021.
- **MELODIC:** Machine Learning boosted WIFI cloud controller, funded by Ministerio de Economía, Industria y Competitividad por medio del Centro para el Desarrollo Tecnológico Industrial (CDTI), Noviembre 2021 - Abril 2024.

A.6.3 Estancias de investigación

El desarrollo de esta tesis incluye una estancia de investigación en la sección de Redes de Comunicación Inalámbrica (WCN) de la Universidad de Aalborg (AAU) en Dinamarca, desde marzo hasta julio de 2023. La estancia fue supervisada por el Profesor Preben Mogensen de la institución mencionada.

Bibliography

- [1] Peñaherrera-Pulla, O.S. and Baena, Carlos and Fortes, Sergio and Baena, Eduardo and Barco Raquel. Measuring key quality indicators in cloud gaming: Framework and assessment over wireless networks. *Sensors*, 21(4), 2021.
- [2] Peñaherrera-Pulla, O.S. and Baena, Carlos and Fortes, Sergio and Baena, Eduardo and Barco Raquel. KQI Assessment of VR Services: A Case Study on 360-Video Over 4G and 5G. *IEEE Transactions on Network and Service Management*, 19(4):5366–5382, 2022.
- [3] O. S. Peñaherrera-Pulla, Sebastian Bro Damsgaard, Boyan Yanakiev, Preben Mogensen, Sergio Fortes, and Raquel Barco. Cloud vr on 5g: A performance validation in industrial scenarios. *IEEE Open Journal of the Communications Society*, 5:3641–3657, 2024.
- [4] O. S. Peñaherrera-Pulla, Carlos Baena, Sergio Fortes, and Raquel Barco. ML-Powered KQI Estimation for XR Services: A Case Study on 360-Video. *IEEE Open Journal of the Communications Society*, 5:4075–4097, 2024.
- [5] O. S. Peñaherrera-Pulla, Carlos Baena, Hao Luo-Chen, Sergio Fortes, and Raquel Barco. Towards kqi/ml-based management: Optimizing radio resource allocation for next-generation networks. *IEEE Communications Magazine*, pages 4075–4097, 2024.
- [6] Carlos Baena, O. S. Peñaherrera-Pulla, Lourdes Camacho, Raquel Barco, and Sergio Fortes. Video Streaming and Cloud Gaming Services Over 4G and 5G: A Complete Network and Service Metrics Dataset. *IEEE Communications Magazine*, 61(9):154–160, 2023.

- [7] Carlos Baena, O.S. Peñaherrera-Pulla, Raquel Barco, and Sergio Fortes. Measuring and estimating Key Quality Indicators in Cloud Gaming services. *Computer Networks*, 231, 2023.
- [8] Carlos Baena, Sergio Fortes, O. S. Penaherrera-Pulla, Eduardo Baena, and Raquel Barco. Gaming in the Cloud: 5G as the Pillar for Future Gaming Approaches. *IEEE Communications Magazine*, 2024.
- [9] O.S. Peñaherrera-Pulla, C. Baena, S. Fortes, E. Baena, and R. Barco. Experimental Evaluation of Cloud Gaming for Different Communication Technologies. In *2nd IRACON Meeting*, 2021.
- [10] O.S. Peñaherrera-Pulla, C. Baena, S. Fortes, E. Baena, and R. Barco. KQI Evaluation for 360-video Services Over Mobile Networks. In *2nd MC Meeting and 1st Technical Meeting*, Bologna, Italy, 2022.
- [11] O.S. Peñaherrera-Pulla, C. Baena, R. Barco, and S. Fortes. ML-Based Network Management Framework for XR Services. In *European Conference on Networks and Communications (EuCNC)*, Gothenburg, Sweden, 2023.
- [12] Carlos Baena, Sergio Fortes, O.S. Peñaherrera-Pulla, and Raquel Barco. A Framework to boost the potential of network-in-a-box solutions. In *2021 12th International Conference on Network of the Future (NoF)*, pages 1–3, 2021.
- [13] O.S. Peñaherrera-Pulla, C. Baena, S. Fortes, E. Baena, and R. Barco. Adquisición de métricas en Cloud Gaming. In *XXXV Simposium nacional de la Unión Científica Internacional de Radio*, Málaga, Spain, 2020.
- [14] O.S. Peñaherrera-Pulla, C. Baena, S. Fortes, E. Baena, and R. Barco. Adquisición de métricas para el servicio de video 360/VR. In *XXXVI Simposium nacional de la Unión Científica Internacional de Radio*, Vigo, Spain, 2021.
- [15] O.S. Peñaherrera-Pulla, C. Baena, S. Fortes, E. Baena, and R. Barco. Estimación de KQIs para servicio de Video-360. In *XXXVII Simposium nacional de la Unión Científica Internacional de Radio*, Málaga, Spain, 2022.
- [16] C. Baena, O.S. Peñaherrera-Pulla, S. Fortes, E. Baena, and R. Barco. Evaluación del servicio Cloud Gaming para diferentes tecnologías de acceso. In *XXXVI Simposium nacional de la Unión Científica Internacional de Radio*, Vigo, Spain, 2024.

- [17] J. Pulido, S. Fortes, E. Baena, O. S. Peñaherrera-Pulla, A. Fernández-Durán, and R. Barco. Impacto de tráfico XR (eXtended Reality) en entornos Wi-Fi. In *XXXVIII Simposio Nacional de la Unión Científica Internacional de Radio*, Cáceres, Spain, 2023.
- [18] M. B. Martínez, J. Pulido, S. Fortes, C. Baena, O. S. Peñaherrera-Pulla, A. Fernández-Durán, and R. Barco. Estimación de KQIs de video streaming en entornos Wi-Fi. In *XXXIX Simposio Nacional de la Unión Científica Internacional de Radio*, Cuenca, Spain, 2024.
- [19] O. S. Peñaherrera-Pulla, Carlos Baena, Sergio Fortes, and Raquel Barco. E2E dataset of 360-degree video service using mobile networks, 2024. IEEE Dataport.
- [20] Carlos Baena, O.S Peñaherrera-Pulla, Lourdes Camacho, Raquel Barco, and Sergio Fortes. E2E dataset of Video Streaming and Cloud Gaming services over 4G and 5G, 2022.
- [21] O. S. Peñaherrera-Pulla. KQI prediction ML framework Repository. https://github.com/sebaspenaherrera/ML_framework_prediction, 2024. GitHub repository.
- [22] Carlos Baena. Service KQI Extraction Repository. <https://github.com/jcbaglez/ServiceKQIExtraction>, 2023. GitHub repository.
- [23] Qualcomm. 5G-Vision for the next generation of connectivity. <https://www.qualcomm.com/media/documents/files/whitepaper-5g-vision-for-the-next-generation-of-connectivity.pdf>, 2015. (Online, Accessed October 2024).
- [24] O-RAN Alliance. O-RAN Empowering Vertical Industry: Scenarios, Solutions and Best Practice. *Whitepaper*, 2023. Online. Accessed on October, 2024.
- [25] 3GPP TS 38.211. NR; Physical channels and modulation (Release 18). Tech. Spec. V18.4.0, 3rd Generation Partnership Project (3GPP), 2024.
- [26] GSMA. The Mobile Economy 2024. <https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-economy/wp-content/uploads/2024/02/260224-The-Mobile-Economy-2024.pdf>, 2024. (Online, Accessed September 2024).

- [27] Albert Banchs, David M. Gutierrez-Estevez, Manuel Fuentes, Mauro Boldi, and Silvia Provvedi. A 5G Mobile Network Architecture to Support Vertical Industries. *IEEE Communications Magazine*, 57(12):38–44, 2019.
- [28] Nokia. “Discover 2030. A time of unprecedented change and opportunity”. <https://www.nokia.com/innovation/technology-vision-2030/>, 2023. (Online, Accessed September 2024).
- [29] Nokia. “Global Network Traffic 2030 Report”. <https://onestore.nokia.com/asset/213660>, 2023. (Online, Accessed September 2024).
- [30] Ericsson. “Ericsson Mobility Report”. <https://www.ericsson.com/49ed78/assets/local/reports-papers/mobility-report/documents/2024/ericsson-mobility-report-june-2024.pdf>, June 2024. (Online, Accessed September 2024).
- [31] NGMN Alliance. Next generation mobile networks use cases related to self-organising network. Overall description, NGMN Alliance, 2007.
- [32] 3GPP TS 32.500. Telecommunication management; Self-Organizing Networks (SON); Concepts and requirements. Tech. Rep. V18.0.0, 3rd Generation Partnership Project (3GPP), 2024.
- [33] S. Hämmäläinen, H. Sanneck, and C. Sartori. *LTE Self-Organising Networks (SON): Network Management Automation for Operational Efficiency*. Wiley, 2011.
- [34] O.G. Aliu, A. Imran, M.A. Imran, and B. Evans. A survey of self organisation in future cellular networks. *IEEE Communications Surveys & Tutorials*, 15(1):336–361, 2013.
- [35] P.V. Klaine, M.A. Imran, O. Onireti, and R.D. Souza. A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks. *IEEE Communications Surveys & Tutorials*, 19(4):2392–2431, 2017.
- [36] Hasna Fourati, Rihab Maaloul, Lamia Chaari, and Mohamed Jmaiel. Comprehensive survey on self-organizing cellular network approaches applied to 5g networks. *Computer Networks*, 199:108435, 11 2021.
- [37] Huawei and TELUS. New generation son for 5g networks, 2016.

- [38] Ali Imran and Ahmed Zoha. Challenges in 5g: how to empower son with big data for enabling 5g. *IEEE Network*, 28:27–33, 11 2014.
- [39] E. Müller, M. Svensson, M. Walthier, C. Gustafsson, and A. Báder. “Service quality monitoring in the digital economy”. <https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/service-quality-monitoring-in-the-digital-economy>, 1 2024. (Online, Accessed September 2024).
- [40] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li. Channel state information prediction for 5G wireless communications: A deep learning approach. *IEEE Transactions on Network Science and Engineering*, 7(1):227–236, 2020.
- [41] Carolina Gijón, Matías Toril, Salvador Luna-Ramírez, Juan L. Bejarano-Luque, and María Luisa Marí-Altozano. Estimating pole capacity from radio network performance statistics by supervised learning. *IEEE Transactions on Network and Service Management*, 17(4):2090–2101, 2020.
- [42] L.V. Le, D. Sinh, L.P. Tung, and B.S.P. Lin. A practical model for traffic forecasting based on big data, machine-learning, and network KPIs. In *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pages 1–4. IEEE, 2018.
- [43] C.J. Sagar, H. Gupta, J.S. Jadon, N. Arora, and S. Kumar. Prediction of traffic volume in BTS sites using deep learning. In *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*, pages 587–590. IEEE, 2020.
- [44] J. L. Bejarano-Luque, M. Toril, M. Fernández-Navarro, C. Gijón, and S. Luna-Ramírez. A deep-learning model for estimating the impact of social events on traffic demand on a cell basis. *IEEE Access*, 9:71 673–71 686, 2021.
- [45] Young Ik Jo and Joern Krause. TR 37.817 - Study on enhancement for data collection for NR and ENDC, 9 2022.
- [46] Christian Esteve Rothenberg, Danny A. Lachos Perez, Nathan F. Saraiva De Sousa, Raphael V. Rosa, Raza Ul Mustafa, Md Tariqul Islam, and Pedro Henrique Gomes. Intent-based control loop for dash video service assurance using ml-based edge qoe estimation. *Proceedings of the 2020 IEEE*

- Conference on Network Softwarization: Bridging the Gap Between AI and Network Softwarization, NetSoft 2020*, pages 353–355, 6 2020.
- [47] Kaisa Zhang, Gang Chuai, Saidiwaerdi Maimaiti, and Qian Liu. User experience estimation in multi-service scenario of cellular network. *Sensors 2022, Vol. 22, Page 89*, 22:89, 12 2021.
- [48] Sarah Wassermann, Michael Seufert, Pedro Casas, Li Gang, and Kuang Li. Vicrypt to the rescue: Real-time, machine-learning-driven video-qoe monitoring for encrypted streaming traffic. *IEEE Transactions on Network and Service Management*, 17:2007–2023, 12 2020.
- [49] Carlos Baena, O.S. Peñaherrera-Pulla, Raquel Barco, and Sergio Fortes. Measuring and estimating key quality indicators in cloud gaming services. *Computer Networks*, 231:109808, 7 2023.
- [50] Zhengquan Zhang, Yue Xiao, Zheng Ma, Ming Xiao, Zhiguo Ding, Xianfu Lei, George K. Karagiannidis, and Pingzhi Fan. 6g wireless networks: Vision, requirements, architecture, and key technologies. *IEEE Vehicular Technology Magazine*, 14(3):28–41, 2019.
- [51] MONROE Project. MONROE: Measuring Mobile Broadband Networks in Europe. <https://www.monroe-project.eu/>. Accessed: 2023-06-28.
- [52] One5G. E2E-aware Optimizations and advancements for the Network Edge of 5G New Radio. <https://one5g.eu/>. Accessed: 2023-06-28.
- [53] S. Fortes, J. A. Santoyo-Ramón, D. Palacios, E. Baena, R. Mora-García, M. Medina, P. Mora, and R. Barco. The campus as a smart city: University of Málaga environmental, learning, and research approaches. *Sensors*, 19(6), 2019.
- [54] Antonio Tarrías, Eduardo Baena, Sergio Fortes, and Raquel Barco. Leveraging 5G SA for R&D: Capabilities and Beam-Based Empirical Analysis. *IEEE Open Journal of the Communications Society*, 5:5608–5618, 2024.
- [55] Ana Herrera-Garcia, Sergio Fortes, Eduardo Baena, Jessica Mendoza, Carlos Baena, and Raquel Barco. Modeling of Key Quality Indicators for End-to-End Network Management: Preparing for 5G. *IEEE Vehicular Technology Magazine*, 14(4):76–84, 2019.

- [56] 3GPP TS 32.101. Telecommunication management; Principles and high level requirements. Tech. Spec. V18.0.0, 3rd Generation Partnership Project (3GPP), 2024.
- [57] Huawei. Digital transformation of industries summit. <https://carrier.huawei.com/en/technical-topics/service/SolutionTopic/OSS/OSS1>. Online. Accessed on November 2024.
- [58] Guy Saadon, Yoram Haddad, and Noemie Simoni. A survey of application orchestration and oss in next-generation network management. *Computer Standards & Interfaces*, 62:17–31, 2019.
- [59] 3GPP TR 25.913. Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN). Tech. Report V7.3.0, 3rd Generation Partnership Project (3GPP), 2006.
- [60] 3GPP TR 36.913. Requirements for further advancements for evolved universal terrestrial radio access (e-utra) (lte-advanced). Tech. Report V18.0.0, 3rd Generation Partnership Project (3GPP), 2024.
- [61] 3GPP TR 21.914. Release 14 description. summary of rel-14 work items. Tech. Report V14.0.0, 3rd Generation Partnership Project (3GPP), 2018.
- [62] 3GPP TS 36.300. Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 17). Technical Report V17.5.0, 3rd Generation Partnership Project (3GPP), 2023.
- [63] 3GPP TS 23.401. General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 18). Technical Report V18.2.0, 3rd Generation Partnership Project (3GPP), 2023.
- [64] 3GPP TS 23.228. IP Multimedia Subsystem (IMS); Stage 2 (Release 18). Technical Report V18.2.0, 3rd Generation Partnership Project (3GPP), 2023.
- [65] Wenqiang Tian and Kevin Lin. Chapter 2 - Requirements and scenarios of 5G system. In Jia Shen, Zhongda Du, Zhi Zhang, Ning Yang, and Hai Tang, editors, *5G NR and Enhancements*, pages 41–52. Elsevier, 2022.

- [66] Dajie Jiang and Guangyi Liu. *An Overview of 5G Requirements*, pages 3–26. Springer International Publishing, Cham, 2017.
- [67] 3GPP TS 22.261. Service requirements for the 5G system (Release 19). Tech. Rep. V19.3.0, 3rd Generation Partnership Project (3GPP), 2023.
- [68] Maija Breque, Lars De Nul, Athanasios Petridis, et al. Industry 5.0: towards a sustainable, human-centric and resilient european industry. *Luxembourg, LU: European Commission, Directorate-General for Research and Innovation*, 46, 2021.
- [69] 3GPP TS 21.915. Release 15 Description; Summary of Rel-15 Work Items (Release 15). Tech. Rep. V15.0.0, 3rd Generation Partnership Project (3GPP), 2019.
- [70] International Telecommunication Union (ITU). Recommendation ITU-R M.2083-0: IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond. Recommendation M.2083-0, ITU-R, 2015.
- [71] Ericsson. What is 5g? how will it transform our world? <https://www.ericsson.com/en/5g>, 2024. (Online, Accessed September 2024).
- [72] Gabriel Brown. Service-based architecture for 5G core networks. *Huawei White Paper*, 1, 2017.
- [73] 3GPP TS 23.501. System Architecture for the 5G System (5GS) (Release 18). Tech. Rep. V18.2.2, 3rd Generation Partnership Project (3GPP), 2023.
- [74] 3GPP TS 38.300. NR and NG-RAN Overall description; Stage-2 (Release 17). Tech. Spec. V18.3.0, 3rd Generation Partnership Project (3GPP), 2024.
- [75] 3GPP TS 37.324. Evolved Universal Terrestrial Radio Access (E-UTRA) and NR; Service Data Adaptation Protocol (SDAP) specification. Tech. Spec. V18.0.0, 3rd Generation Partnership Project (3GPP), 2024.
- [76] 3GPP TS 38.801. Study on new radio access technology: Radio access architecture and interfaces (Release 14). Tech. Rep. V14.0.0, 3rd Generation Partnership Project (3GPP), 2017.

- [77] Luis Diez, Cristina Hervella, and Ramón Agüero. Understanding the Performance of Flexible Functional Split in 5G vRAN Controllers: A Markov Chain-Based Model. *IEEE Transactions on Network and Service Management*, 18(1):456–468, 2021.
- [78] 3GPP TS 38.816. Study on Central Unit (CU) - Distributed Unit (DU) lower layer split for NR (Release 15). Tech. Spec. V15.0.0, 3rd Generation Partnership Project (3GPP), 2018.
- [79] O-RAN Alliance. <https://www.o-ran.org>. Visited in October 2024.
- [80] Engin Zeydan, Josep Mangues-Bafalluy, Jorge Baranda, Manuel Requena, and Yekta Turk. Service Based Virtual RAN Architecture for Next Generation Cellular Systems. *IEEE Access*, 10:9455–9470, 2022.
- [81] O-RAN Alliance. O-RAN Use Cases and Deployment Scenarios. *Whitepaper*, 2020. Online. Accessed on September, 2024.
- [82] Michele Polese, Leonardo Bonati, Salvatore D’Oro, Stefano Basagni, and Tommaso Melodia. Understanding o-ran: Architecture, interfaces, algorithms, security, and research challenges. *IEEE Communications Surveys & Tutorials*, 25(2):1376–1411, 2023.
- [83] O-RAN WG1. O-RAN Architecture Description. Tech. Report R003-v12.00, O-RAN Alliance, 2024.
- [84] Salvatore D’Oro, Michele Polese, Leonardo Bonati, Hai Cheng, and Tommaso Melodia. dapps: Distributed applications for real-time inference and control in o-ran. *IEEE Communications Magazine*, 60(11):52–58, 2022.
- [85] O-RAN RR-2024-10. dapps for real-time ran control: Use cases and requirements. Research Report V1.0.0, O-RAN next Generation Research Group (nGRG), 2024.
- [86] 3GPP TR 38.912. Study on New Radio (NR) access technology. Tech. Report V18.0.0, 3rd Generation Partnership Project (3GPP), 2024.
- [87] Nokia. “5G spectrum bands explained— low, mid and high band”. <https://www.nokia.com/thought-leadership/articles/spectrum-bands-5g-world/>, 2023. (Online, Accessed September 2024).

- [88] 3GPP TS 38.104. NR; Base Station (BS) radio transmission and reception. Tech. Report V18.7.0, 3rd Generation Partnership Project (3GPP), 2024.
- [89] 3GPP TS 38.213. NR; Physical Layer Procedures for Control (Release 18). Tech. Spec. V18.4.0, 3rd Generation Partnership Project (3GPP), 2024.
- [90] 3GPP TS 38.321. NR; Medium Access Control (MAC) protocol specification. Tech. Spec. V18.3.0, 3rd Generation Partnership Project (3GPP), 2024.
- [91] Z.H. Zhou and S. Liu. *Machine Learning*. Springer Nature Singapore, 2021.
- [92] Estefanía Coronado, Rasoul Behraves, Tejas Subramanya, Adriana Fernández-Fernández, Muhammad Shuaib Siddiqui, Xavier Costa-Pérez, and Roberto Riggio. Zero Touch Management: A Survey of Network Automation Solutions for 5G and 6G Networks. *IEEE Communications Surveys & Tutorials*, 24(4):2535–2578, 2022.
- [93] Jorge Gallego-Madrid, Ramon Sanchez-Iborra, Pedro M. Ruiz, and Antonio F. Skarmeta. Machine learning-based zero-touch network and service management: a survey. *Digital Communications and Networks*, 8(2):105–123, 2022.
- [94] Chafika Benzaid and Tarik Taleb. AI-Driven Zero Touch Network and Service Management in 5G and Beyond: Challenges and Research Directions. *IEEE Network*, 34(2):186–194, 2020.
- [95] ITU-T Telecommunication Standardization Sector. Definitions of terms related to quality of service. ITU-T Recommendation E.800, International Telecommunication Union, 2008.
- [96] ETSI. TR 102 643. Human Factors (HF); Quality of Experience (QoE) requirements for real-time communication services. Tech. report, ETSI, 2010.
- [97] 3GPP TR 28.863. Study on Key Quality Indicators (KQIs) for 5G service experience (Release 18). Tech. Report. V18.8.0, 3rd Generation Partnership Project (3GPP), 2024.
- [98] Pablo Oliver-Balsalobre, Matías Toril, Salvador Luna-Ramírez, and Jose Maria Ruiz Aviles. Self-tuning of scheduling parameters for balancing the quality of experience among services in LTE. *EURASIP Journal on Wireless Communications and Networking*, 2016(1):1–12, 2016.

- [99] María Luisa Marí Altozano, Matías Toril, Salvador Luna-Ramírez, and Carolina Gijón. A Self-Tuning Algorithm for Optimal QoE-Driven Traffic Steering in LTE. *IEEE Access*, 8:156707–156717, 2020.
- [100] Jessica Mendoza, Isabel de-la Bandera, David Palacios, and Raquel Barco. QoE Optimization in a Live Cellular Network through RLC Parameter Tuning. *Sensors*, 21(16), 2021.
- [101] 3GPP SA 4. Specification # 26.247 - Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH). Technical report, 3GPP, 2023.
- [102] Series P: Telephone transmission quality, telephone installations, local line networks. P.1203: Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport. Technical report, ITU-T, 2017.
- [103] U. S. Hashmi, A. Rudrapatna, Z. Zhao, M. Rozwadowski, J. Kang, R. Wuppalapati, and A. Imran. Towards real-time user QoE assessment via machine learning on LTE network data. In *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, pages 1–7. IEEE, 2019.
- [104] S. Pezzulli, M. G. Martini, and N. Barman. Estimation of Quality Scores From Subjective Tests-Beyond Subjects’ MOS. *IEEE Transactions on Multimedia*, 23:2505–2519, 2021.
- [105] Craig Gutterman, Katherine Guo, Sarthak Arora, Xiaoyang Wang, Les Wu, Ethan Katz-Bassett, and Gil Zussman. Requet: Real-time QoE detection for encrypted YouTube traffic. *Proceedings of the 10th ACM Multimedia Systems Conference, MMSys 2019*, pages 48–59, 6 2019.
- [106] Md Tariqul Islam, Christian Esteve Rothenberg, and Pedro Henrique Gomes. Predicting XR Services QoE with ML: Insights from In-band Encrypted QoS Features in 360-VR. *2023 IEEE 9th International Conference on Network Softwarization: Boosting Future Networks through Advanced Softwarization, NetSoft 2023 - Proceedings*, pages 80–88, 2023.
- [107] 3GPP TR 26.998. Support of 5G glass-type Augmented Reality / Mixed Reality (AR/MR) devices. Tech. Report V18.1.0, 3rd Generation Partnership Project (3GPP), 2024.

- [108] Mohammed S. Elbamby, Cristina Perfecto, Mehdi Bennis, and Klaus Doppler. Toward Low-Latency and Ultra-Reliable Virtual Reality. *IEEE Network*, 32(2):78–84, 3 2018.
- [109] Ejder Bastug, Mehdi Bennis, Muriel Medard, and Merouane Debbah. Toward Interconnected Virtual Reality: Opportunities, Challenges, and Enablers. *IEEE Communications Magazine*, 55(6):110–117, 2017.
- [110] Mallesham Dasari, Arani Bhattacharya, Santiago Vargas, Pranjal Sahu, Aruna Balasubramanian, and Samir R. Das. Streaming 360-Degree Videos Using Super-Resolution. *Proceedings - IEEE INFOCOM*, 2020-July:1977–1986, 7 2020.
- [111] Polona Caserman, Michelle Martinussen, and Stefan Göbel. Effects of End-to-end Latency on User Experience and Performance in Immersive Virtual Reality Applications. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11863 LNCS, pages 57–69. Springer, 11 2019.
- [112] 3GPP TR 26.918. Virtual Reality (VR) media services over 3GPP. Tech. Report V18.0.0, 3rd Generation Partnership Project (3GPP), 2024.
- [113] A. Ryczkowski, T Piotrowski, M Staszczak, M Wiktorowicz, and P Adrich. Optimization of the regularization parameter in the Dual Annealing method used for the reconstruction of energy spectrum of electron beam generated by the AQURE mobile accelerator. *Zeitschrift für Medizinische Physik*, 2023.