

UNIVERSIDADE FEDERAL DE UBERLÂNDIA (UFU)
FACULDADE DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

USO DE *DEEP Q-LEARNING* PARA ALOCAÇÃO DE RECURSOS *UPLINK* EM
REDES HETEROGÊNEAS LTE-A/5G

Itagildo Edmar Garbazza

Uberlândia – MG
2025

ITAGILDO EDMAR GARBAZZA

**USO DE *DEEP Q-LEARNING* PARA ALOCAÇÃO DE RECURSOS *UPLINK* EM
REDES HETEROGÊNEAS LTE-A/5G**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Faculdade de Engenharia Elétrica da Universidade Federal de Uberlândia, como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica.

Área de concentração: Engenharia Elétrica

Orientador: Prof. Dr. Éderson Rosa da Silva

Coorientador: Prof. Dr. Paulo Roberto Guardieiro

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

G213 Garbazza, Itagildo Edmar, 1977-
2025 Uso de Deep Q-Learning para alocação de recursos uplink em
redes heterogêneas LTE-A/5G [recurso eletrônico] / Itagildo Edmar
Garbazza. - 2025.

Orientador: Éderson Rosa Silva.

Coorientador: Paulo Roberto Guardieiro.

Tese (Doutorado) - Universidade Federal de Uberlândia, Pós-
graduação em Engenharia Elétrica.

Modo de acesso: Internet.

DOI <http://doi.org/10.14393/ufu.te.2025.374>

Inclui bibliografia.

1. Engenharia elétrica. I. Silva, Éderson Rosa, 1984-, (Orient.). II.
Guardieiro, Paulo Roberto, 1952-, (Coorient.). III. Universidade
Federal de Uberlândia. Pós-graduação em Engenharia Elétrica. IV.
Título.

CDU: 621.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091

Nelson Marcos Ferreira - CRB6/3074

ITAGILDO EDMAR GARBAZZA

**USO DE *DEEP Q-LEARNING* PARA ALOCAÇÃO DE RECURSOS *UPLINK* EM
REDES HETEROGÊNEAS LTE-A/5G**

Presidente da banca: Prof. Dr. Éderson Rosa da Silva

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Faculdade de Engenharia Elétrica da Universidade Federal de Uberlândia, como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica.

Área de concentração: Engenharia Elétrica

Banca examinadora

Prof. Dr. Éderson Rosa da Silva (Orientador) – UFU

Prof. Dr. Alexandre Coutinho Mateus – UFU

Prof. Dr. André Luiz Aguiar da Costa – UFU

Prof. Dr. Márcio Andrey Teixeira – IFSP

Prof. Dr. Saulo Henrique da Mata – IFTM



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Coordenação do Programa de Pós-Graduação em Engenharia Elétrica
Av. João Naves de Ávila, 2121, Bloco 3N - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3239-4707 - www.posgrad.feelt.ufu.br - copel@ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Engenharia Elétrica				
Defesa de:	Tese de Doutorado, 348 , PPGEELT				
Data:	Quatro de Junho de Dois mil e vinte e cinco	Hora de início:	14:00	Hora de encerramento:	16:30
Matrícula do Discente:	11723EEL003				
Nome do Discente:	Itagildo Edmar Garbazza				
Título do Trabalho:	Uso de Deep Q-learning para alocação de recursos uplink em redes heterogêneas LTE-A/5G				
Área de concentração:	Processamento da Informação				
Linha de pesquisa:	Processamento Digital de Sinais e Redes de Comunicação				
Projeto de Pesquisa de vinculação:	Coordenador do projeto: Éderson Rosa da Silva . Título do projeto: Desenvolvimento e simulação de técnicas de alocação de recursos em redes de comunicação. Agência financiadora: Não se aplica. Número do processo na agência financiadora: Não se aplica. Vigência do projeto: 2018 - atual.				

Reuniu-se através de videoconferência, a Banca Examinadora designada pelo Colegiado do Programa de Pós-graduação em Engenharia Elétrica, assim composta:

Doutores: Alexandre Coutinho Mateus (UFU), André Luiz Aguiar da Costa (UFU), Márcio Andrey Teixeira (IFSP), Saulo Henrique da Mata (IFTM) e Éderson Rosa da Silva, orientador do discente.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. Éderson Rosa da Silva, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao discente a palavra para a exposição do seu trabalho. A duração da apresentação do discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir, o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

APROVADO.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor. O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme, foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Saulo Henrique da Mata, Usuário Externo**, em 04/06/2025, às 16:36, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Ederson Rosa da Silva, Professor(a) do Magistério Superior**, em 04/06/2025, às 16:37, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Alexandre Coutinho Mateus, Professor(a) do Magistério Superior**, em 04/06/2025, às 16:37, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Andre Luiz Aguiar da Costa, Professor(a) do Magistério Superior**, em 04/06/2025, às 16:37, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **MARCIO ANDREY TEIXEIRA, Usuário Externo**, em 04/06/2025, às 16:38, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6295652** e o código CRC **C1E6E078**.

Dedico este trabalho à minha esposa, Renata, aos meus filhos, Gabriel e Sara, e aos meus pais e irmãos, por sempre me incentivarem, auxiliarem e tornarem possível a concretização deste sonho.

AGRADECIMENTOS

Expresso minha gratidão a Deus, a quem reconheço como o Senhor de minha vida, por conceder-me a oportunidade de realizar este sonho acadêmico. Agradeço-lhe pela vida abundantemente abençoada, pela constante orientação e por jamais permitir que eu desfalecesse. Toda a glória seja dada a Ti, Senhor!

À minha querida esposa, Renata. Expresso minha gratidão pelo amor, pelas palavras de conforto e incentivo, pelo apoio constante, pelas orações e pelo compaheirismo nos momentos difíceis. Celebro os frutos que juntos cultivamos. Agradeço por sua crença e apoio incondicional. Meu amor por você é imenso!

Ao meu orientador, Prof. Dr. Éderson Rosa da Silva, manifesto meu profundo agradecimento pela orientação durante minha trajetória acadêmica. Sua receptividade, apoio ao desenvolvimento de minhas ideias e as oportunidades concedidas foram fundamentais. Sou muito grato.

Ao meu coorientador, Prof. Dr. Paulo Roberto Guardieiro, expresso minha gratidão pelo conhecimento transmitido, pela dedicação, paciência, compreensão e pelo estímulo essencial à conclusão deste trabalho.

Gostaria de agradecer aos professores membros da banca pelo tempo dedicado à avaliação desta tese.

Aos professores do PPGEELT e aos demais que contribuíram para minha formação, agradeço não apenas pela instrução técnica, mas também pelos princípios que se mostraram e permanecem cruciais em meu desenvolvimento.

Aos professores do Núcleo de Computação do IFMG – Campus Bambuí, meu sincero agradecimento por todo o apoio nesta jornada.

Aos amigos e colegas que encontrei durante o doutorado, que compartilharam direta ou indiretamente este período de intenso trabalho, em especial André, Arthur, Flávio, Jaqueline, Júlia, Júnio Moreira e Thales, registro minha gratidão.

A todos aqueles que, de alguma forma, colaboraram para a realização desta etapa desafiadora, porém significativa, em minha vida, estendo meus sinceros agradecimentos!

“Não se turbe o vosso coração; credes em Deus, crede também em mim.”
João 14:1

RESUMO

Garbazza, Itagildo Edmar, Uso de *Deep Q-Learning* para alocação de recursos *uplink* em redes heterogêneas LTE-A/5G, UFU, Uberlândia, Brasil, 2025. DOI <http://doi.org/10.14393/ufu.te.2025.374>

A crescente demanda por conectividade móvel, impulsionada pela diversidade de aplicações com requisitos rigorosos de qualidade de serviço (QoS), impõe desafios à alocação eficiente de recursos em redes heterogêneas LTE-A/5G. Em vista disso, este trabalho propõe uma solução integrada de escalonamento e alocação de recursos no *uplink*, com o objetivo de garantir a QoS, reduzir o consumo energético e mitigar interferências intercelulares em cenários densos e de alta mobilidade. A metodologia desenvolvida combina técnicas de *Deep Q-Learning* (DQL), capazes de aprender políticas ótimas de alocação de recursos por meio da interação com o ambiente, e o método multicritério *Weighted Aggregated Sum Product Assessment* (WASPAS), aplicado à tomada de decisão de *handover* com base em múltiplos critérios, como *Signal to Interference plus Noise Ratio* (SINR), carga de tráfego celular e mobilidade do usuário. A proposta contempla três componentes principais: (i) um escalonador de objetivo duplo que prioriza, de forma diferenciada, o tráfego em tempo real (RT) e não em tempo real (NRT), considerando os requisitos específicos de cada classe de serviço; (ii) um mecanismo de alocação de recursos com restrição de energia, adaptativo às variações do canal e à demanda de QoS; e (iii) uma estratégia inteligente de *handover*, que promove continuidade de serviço e redução de quedas de conexão. O desempenho do sistema foi avaliado por meio de simulações no ambiente *Network Simulator 3* (ns-3), utilizando diferentes densidades de usuários e cenários dinâmicos. Os resultados demonstram ganhos significativos em relação a algoritmos tradicionais, como *Proportional Fair* (PF) e *New Courteous Algorithm* (NCA), evidenciando redução do atraso, aumento do *throughput*, melhoria do índice de justiça na distribuição de recursos e diminuição das taxas de perda de pacotes e de queda de chamadas. Como consequência, evidencia-se o potencial das abordagens baseadas em inteligência artificial e decisão multicritério para a otimização da alocação de recursos em redes móveis celulares modernas.

Palavras-chave: Alocação de Recursos, *Deep Q-Learning*, Escalonamento *Uplink*, Qualidade de Serviço, Redes Heterogêneas.

ABSTRACT

Garbazza, Itagildo Edmar, Deep Q-Learning Application to Uplink Resource Allocation in LTE-A/5G Heterogeneous Networks, UFU, Uberlândia, Brazil, 2025. DOI <http://doi.org/10.14393/ufu.te.2025.374>

The growing demand for mobile connectivity, driven by diverse applications with stringent Quality of Service (QoS) requirements, poses significant challenges to the efficient allocation of resources in heterogeneous LTE-A/5G networks. This work proposes an integrated solution for uplink scheduling and resource allocation, aiming to ensure QoS, reduce energy consumption, and mitigate intercell interference in dense and high-mobility scenarios. The proposed methodology combines Deep Q-Learning (DQL), capable of learning optimal resources allocation policies through environment interaction, and the multi-criteria decision-making method Weighted Aggregated Sum Product Assessment (WASPAS), applied to intelligent handover decisions based on multiple criteria, such as Signal to Interference plus Noise Ratio (SINR), cell load, and user mobility. The solution is structured into three main components: (i) a dual-objective scheduler that prioritizes real-time (RT) and non-real-time (NRT) traffic based on the specific requirements of each service class; (ii) an energy-constrained resource allocation mechanism that adapts to channel variations and QoS demands; and (iii) an intelligent handover strategy that promotes service continuity and reduces connection drops. System performance was evaluated through simulations using the Network Simulator 3 (ns-3) environment under various user densities and dynamic scenarios. The results show significant improvements compared to traditional algorithms such as Proportional Fair (PF) and New Courteous Algorithm (NCA), including delay reduction, increased throughput, improved fairness index in resource distribution, and lower packet loss and call drop rates. As a result, the potential of artificial intelligence and multi-criteria decision-making based approaches for optimizing resource allocation in modern cellular mobile networks is highlighted.

Keywords: Deep Q-Learning, Heterogeneous Networks, Quality of Service, Resource Allocation, Uplink Scheduling.

LISTA DE FIGURAS

Figura 1 - Principais aplicações dos sistemas 5G (Castro Alves, 2020)	23
Figura 2 - Arquitetura da Rede LTE (Amitabha Ghosh, 2011).	33
Figura 3 - Evolução da Rede LTE, adaptado de (Dahlman, 2023).	34
Figura 4 - Casos de uso do IMT-2020 e mapeamento para cenários de utilização, adaptado de (Dahlman, 2023).	35
Figura 5 - Arquitetura Geral (Semov <i>et al.</i> , 2020).	36
Figura 6 - Exemplo de fatiamento dinâmico de rede para serviço de carro autônomo (Spadinger, 2021).	38
Figura 7 - Diferentes requisitos suportados pela rede 5G, adaptado de (Dahlman, 2023).	41
Figura 8 - Logo 5G Advanced (3GPP, 2024b).	43
Figura 9 - Arquitetura de Rede Móvel Celular (Peterson; Sunay; Davie, 2023). .	45
Figura 10 -Arquitetura de Rede 5G NSA (Kumar; AR, 2020).	47
Figura 11 -Arquitetura de Rede 5G SA (Kumar; AR, 2020).	48
Figura 12 -Formato Tradicional de Rede.	49
Figura 13 -Rede Heterogênea com diferentes tipos de nós, adaptado de (Sanchez; Gerardino, 2014) e (Ali, M. <i>et al.</i> , 2017).	50
Figura 14 -Processo de <i>Handover</i> em redes 5G, adaptado de (Ullah <i>et al.</i> , 2023). .	53
Figura 15 -Fluxograma esquemático do processo de escalonamento <i>uplink</i>	56
Figura 16 -Funcionalidade e Estrutura do Escalonador de Pacotes (Singh <i>et al.</i> , 2022).	58
Figura 17 -Estrutura de <i>frame</i> LTE no domínio do tempo e frequência, adaptado de (Dahlman; Parkvall; Skold, 2016).	60
Figura 18 -Arquitetura do escalonador na camada MAC do LTE.	61
Figura 19 -SDFs e EPS <i>bearers</i> , adaptado de (Carvalho Ferreira, 2020).	65
Figura 20 -Arquitetura da metodologia proposta.	72
Figura 21 -Topologia de Rede, adaptado de (Abubakar <i>et al.</i> , 2022).	81
Figura 22 -Organização do software ns-3, adaptado de (Nsnam, 2024a).	83
Figura 23 -Atraso para tráfego RT em função do número de UEs.	88
Figura 24 -Atraso para tráfego NRT em função do número de UEs.	89
Figura 25 - <i>Throughput</i> para tráfego RT em função do número de UEs.	89
Figura 26 - <i>Throughput</i> para tráfego NRT em função do número de UEs.	90
Figura 27 -Índice de justiça para tráfego RT em função do número de UEs. . . .	90
Figura 28 -Índice de justiça para tráfego NRT em função do número de UEs. . .	91
Figura 29 -Taxa de queda de chamadas para tráfego RT em função do número de UEs.	91

Figura 30 -Taxa de queda de chamadas para tráfego NRT em função do número de UEs.	92
Figura 31 -Taxa de perda de pacotes para tráfego RT em função do número de UEs.	92
Figura 32 -Taxa de perda de pacotes para tráfego NRT em função do número de UEs.	93
Figura 33 -Análise comparativa da complexidade computacional do esquema proposto em relação aos demais métodos avaliados.	93

LISTA DE TABELAS

Tabela 1 - Comparativo entre tipos de células, adaptado de (Yuan, 2015) . . .	27
Tabela 2 - Mapeamento CQI e esquemas de modulação e codificação, adaptado de (Nwogu, 2021)	63
Tabela 3 - Comparativo entre SC-FDMA (LTE) e OFDMA (5G/NR) no Uplink	66
Tabela 4 - Parâmetros de Simulação	82

LISTA DE SIGLAS

- 1G – Primeira Geração de Rede Móvel
2G – Segunda Geração de Rede Móvel
3G – Terceira Geração de Rede Móvel
3GPP – 3rd Generation Partnership Project
4G – Quarta Geração de Rede Móvel
5G – Quinta Geração de Rede Móvel
5GCN – 5G Core Network
5G NR – 5G New Radio
AMC – Adaptive Modulation and Coding
AR – Augmented Reality
BS – Base Station
BSR – Buffer Status Report
CDMA – Code Division Multiple Access
CIoT – Cellular Internet of Things
CoMP – Coordinated Multi-Point
CQI – Channel Quality Indicator
CRC – Cyclic Redundancy Check
DL-SCH – Downlink Shared Channel
DQL – Deep Q-Learning
DQN – Deep Q-Network
DRX – Discontinuous reception
DSS – Dynamic Spectrum Sharing
eDRX – enhanced DRX
EDGE – Enhanced Data Rates for GSM Evolution
eICIC – enhanced Inter-Cell Interference Coordination
eMBB – enhanced Mobile Broadband
eNodeB – evolved Node B
EPC – Evolved Packet Core
ePDCCH – enhanced Physical Downlink Control Channel
EPS – Evolved Packet System
ETSI – European Telecommunications Standards Institute
E-UTRAN – Evolved UMTS Terrestrial Radio Access Network
FDPS – Frequency Domain Packet Scheduling
FR1 – Frequency Range 1
FR2 – Frequency Range 2
GPRS – General Packet Radio Service
GSM – Global System for Mobile Communication

HARQ – Hybrid ARQ
HetNets – Heterogeneous Networks
HSPA+ – High Speed Packet Access plus
IA – Inteligência Artificial
IMS – IP Multimedia Sub-system
IMT – International Mobile Telecommunications
IoT – Internet of Things
ITU – International Telecommunication Union
ITU-R – ITU Radiocommunication Sector
KPI – Key Performance Indicators
LAA – License Assisted Access
LPN – Low Power Nodes
LTE – Long Term Evolution
LTE-A – LTE-Advanced
MAC – Media Access Control
MCDM – Multi-Criteria Decision Making
MCN – Mobile Core Network
MCS – Modulation and Coding Scheme
MDP – Markov Decision Process
MIMO – Multiple-Input Multiple-Output
ML – Machine Learning
MR – Measurement Report
MTC – Machine Type Communication
mMTC – massive Machine Type Communication
NCA – New Courteous Algorithm
NeNBs – Neighboring eNodeB
NG-Core – Next Generation Core
NPN – Non-Public Networks
NR – New Radio
NR-U – New Radio-Unlicensed
NRT – Non Real Time
ns-3 – Network Simulator 3
NSA – Non-StandAlone
OFDM – Orthogonal Frequency-Division Multiplexing
OFDMA – Orthogonal Frequency Division Multiple Access
PF – Proportional Fair
PDN – Packet Data Network
PUCCH – Physical Uplink Control Channel
PUSCH – Physical Uplink Shared Channel

QAL – Qos Aware LTE-A/5G
QoE – Quality of Experience
QoS – Quality of Service
QPSK – Quadrature Phase Shift Keying
RA – Realidade Aumentada
RAN – Radio Access Network
RB – Resource Block
RE – Resource Element
RL – Reinforcement Learning
RR – Round Robin
RRH – Remote Radio Head
RRM – Radio Resource Management
RSRP – Received Signal Reference Power
RT – Real Time
RV – Realidade Virtual
SA – StandAlone
SB – Scheduling Block
SC-FDMA – Single Carrier Frequency Division Multiple Access
SC-PTM – Single Cell Point to Multipoint
SDF – Service Data Flows
SeNB – Serving eNodeB
SG – Scheduling Grant
SINR – Signal to Interference plus Noise Ratio
SMS – Short Message Service
SNR – Signal to Noise
SON – Self Organizing Networks
SPT – Semi-Persistent-Transmission
SR – Scheduling Request
SRR – Smoothed Round-Robin
SRS – Sounding Reference Signals
TB – Transport Block
TBS – Transport Block Size
TDMA – Time Division Multiple Access
TDPS – Time Domain Packet Scheduling
TIA – Telecommunications Industry Association
TTI – Transmission Time Interval
UAS – Unmanned Aerial Systems
UDN – Ultra Dense Networks
UE – User Equipment

UL-SCH – Uplink Shared Channel
UMTS – Universal Mobile Telecommunication System
URLLC – Ultra-Reliable and Low-Latency Communications
USIM – Universal Subscriber Identity Module
UTRAN – UMTS Terrestrial Radio Access Network
V2V – Vehicle-to-Vehicle
V2X – Vehicle-to-Everything
VoLTE – Voice over LTE
VR – Virtual Reality
WASPAS – Weighted Aggregated Sum Product Assessment
WPM – Weighted Product Model
WSM – Weighted Sum Model

SUMÁRIO

1	Introdução	20
1.1	Estado da Arte	25
1.2	Definição do Problema	26
1.3	Objetivo	28
1.4	Solução Proposta	29
1.5	Contribuições	30
1.6	Sinopse dos Capítulos	31
2	Redes LTE/LTE-A e 5G NR	32
2.1	A Evolução das Redes Móveis: do LTE ao 5G-Advanced	33
2.1.1	<i>Release 8</i>	37
2.1.2	<i>Release 9</i>	38
2.1.3	<i>Release 10</i>	38
2.1.4	<i>Release 11</i>	39
2.1.5	<i>Release 12</i>	39
2.1.6	<i>Release 13</i>	39
2.1.7	<i>Release 14</i>	40
2.1.8	<i>Release 15</i>	40
2.1.9	<i>Release 16</i>	42
2.1.10	<i>Release 17</i>	43
2.1.11	<i>Release 18</i>	43
2.2	Arquitetura	44
2.2.1	<i>Rede de Acesso por Rádio</i>	44

2.2.2 Núcleo da Rede Móvel	46
2.2.2.1 Rede Non-Standalone	46
2.2.2.2 Rede Standalone	47
2.3 Tipos de Implantação de Estações Base	47
2.3.1 Redes Homogêneas	48
2.3.2 Redes Heterogêneas	48
2.3.3 Redes Heterogêneas Ultra Densas	49
2.4 Handover	51
2.5 Alocação de Recursos e Escalonamento em Redes Móveis Celulares	53
2.6 Considerações Finais	54
3 Escalonamento e Alocação de Recursos <i>Uplink</i> em Redes LTE-A/5G	55
3.1 Desafios na Implementação de Algoritmos de Escalonamento Uplink em Sistemas LTE-A/5G	57
3.2 Escalonador de Pacotes	58
3.3 Alocação de Recursos	60
3.4 Comparação entre SC-FDMA e OFDMA no <i>Uplink</i>	65
3.5 Escalonador <i>Smoothed Round-Robin</i> (SRR)	66
3.6 Escalonador <i>Proportional Fair</i> (PF)	67
3.7 Escalonador <i>New Courteous Algorithm</i> (NCA)	68
3.8 Considerações Finais	70
4 Esquema de Escalonamento e Alocação de Recursos <i>Uplink</i> em Het-Nets LTE-A/5G baseado em DQL e WASPAS	71
4.1 Solução Proposta	71
4.1.1 Escalonamento de Objetivo Duplo	71
4.1.2 Alocação de Recursos com Deep Q-Learning	73

4.1.3 Tomada de Decisão de Handover	76
4.2 Estrutura do Sistema	80
4.3 Simulações e Parâmetros de Desempenho	81
4.4 Métricas de Avaliação	83
 4.4.1 Atraso médio	83
 4.4.2 Vazão	84
 4.4.3 Índice de justiça	84
 4.4.4 Taxa de queda de chamadas	84
 4.4.5 Taxa de perda de pacotes	84
 4.4.6 Complexidade Assintótica	85
4.5 Metodologia de Análise	85
4.6 Considerações Finais	86
5 Apresentação e Análise de Resultados	87
5.1 Análise dos Resultados	88
 5.1.1 Atraso Médio	88
 5.1.2 Throughput	88
 5.1.3 Índice de Justiça	89
 5.1.4 Taxa de Queda de Chamadas	90
 5.1.5 Taxa de Perda de Pacotes	91
 5.1.6 Complexidade	93
5.2 Discussão dos Resultados e Implicações	94
5.3 Considerações Finais	95
6 Conclusão	96
REFERÊNCIAS	98

1 INTRODUÇÃO

As redes móveis celulares têm desempenhado um papel fundamental na conectividade global, permitindo a comunicação sem fio em qualquer lugar e evoluindo rapidamente para atender às crescentes demandas por taxas de transmissão elevadas, baixa latência e alta confiabilidade. O tráfego de dados móveis tem crescido exponencialmente, impulsionado pela comunicação humana e por novas aplicações, como Internet das Coisas (IoT, do inglês *Internet of Things*), veículos autônomos e automação industrial. Projeções indicam que esse volume poderá ultrapassar 218 Exabytes por mês em 2025 (Ericsson, 2024), impondo uma demanda crescente sobre a capacidade das redes.

A evolução das gerações de redes móveis teve início com a tecnologia de primeira geração (1G) (década de 1980) utilizando sistema de comunicação analógico com divisão de frequência para recepção e transmissão de sinal, oferecendo suporte a apenas uma única chamada por canal (Carvalho Ferreira, 2020). Existiram diversos projetos independentes, como destacado à seguir:

- Sistema Avançado de Telefonia Móvel (AMPS, do inglês *Advanced Mobile Phone System*), adotado nos EUA;
- Telecomunicações Móveis Nórdicas (NMT, do inglês *Nordic Mobile Telecommunications*)/Sistema de Comunicações de Acesso Total (TACS, do inglês *Total Access Communications System*), adotado na Europa;
- Sistema de Comunicação de Acesso Total Japonês (JTACS, do inglês *Japanese Total Access Communication System*), adotado no Japão e em Hong Kong.

Todos esses sistemas estavam limitados apenas a serviços de voz e os dispositivos móveis eram grandes e caros (Dahlman, 2016). No final da década de 1980 e começo da década de 1990, teve início a segunda geração (2G) das comunicações móveis, marcada pela adoção da tecnologia digital para a transmissão sem fio. Além dos serviços de voz esses sistemas também permitiram o serviço de dados comutados por circuito, sendo o serviço de mensagens curtas (SMS, do inglês *Short Message Service*) o serviço mais popular (Dahlman; Parkvall; Skold, 2016). O cdmaOne, nome comercial do padrão IS-95 desenvolvido pela Qualcomm, utilizava o Acesso Múltiplo por Divisão de Código (CDMA, do inglês *Code Division Multiple Access*) para acesso ao meio, o que possibilitava melhor aproveitamento da frequência com a alocação dos usuários em códigos de espalhamento espectral diferentes. O cdmaOne foi escolhido pela *Telecommunications Industry Association* (TIA) como o padrão 2G nos EUA, sendo inclusive adotado por operadores na Coréia. O *Global System for Mobile Communication* (GSM) tornou-se o padrão 2G na Europa. O GSM foi desenvolvido pelo *European Telecommunications Standards Institute* (ETSI) e utili-

zava o Acesso Múltiplo por Divisão de Tempo (TDMA, do inglês *Time Division Multiple Access*) para acesso ao meio. Inicialmente, o GSM permitia taxas de transmissão de até 9,6 Kbps. O *General Packet Radio Service* (GPRS) acrescentou ao GSM a comunicação por pacotes, oferecendo suporte transparente ao protocolo TCP/IP na rede de núcleo e uma interface aérea modificada para lidar com a transmissão de voz e dados (Dahlman, 2016), atingindo taxas de dados de até 114 Kbps. A rede GSM foi aprimorada com a inclusão da tecnologia *Enhanced Data Rates for GSM Evolution* (EDGE), que possibilitou taxas de dados máximas aproximadamente três vezes superiores às do GPRS (Santos, 2019).

Os requisitos básicos necessários para a terceira geração de telefonia móvel (3G) foram estabelecidos pela União Internacional de Telecomunicações (ITU, do inglês *International Telecommunication Union*) no final da década de 1990, quando foi criado o programa IMT-2000; oferecendo banda larga móvel de alta velocidade, sistema baseado em IP, *roaming* global, dentre outros recursos (ITU-R, 1997). Nesse período dois grupos foram formados: o *3rd Generation Partnership Project* (3GPP), com foco na família GSM; e o 3GPP2 com foco em tecnologias IS-95. A partir de então dois novos padrões passaram a competir pelo mercado das operadoras de celular: o WCDMA e o CDMA2000. A tecnologia de rede de acesso via rádio conhecida como *UMTS Terrestrial Radio Access Network* (UTRAN) é baseada no WCDMA. São três as principais diferenças entre os sistemas UTRAN e CDMA2000 (Christopher Cox, 2014):

- Largura de banda de 5 MHz pelo UTRAN e de 1,25 MHz pelo CDMA2000;
- Compatibilidade entre sistemas CDMA2000 e cdmaOne, não existindo compatibilidade entre os sistemas UMTS e GSM;
- Compartilhamento de tráfego de voz e dados numa mesma portadora pelo sistema UTRAN, enquanto o CDMA2000 utiliza portadoras diferentes para voz e dados.

A partir da 3GPP *Release 8* foi criada a especificação *Long Term Evolution* (LTE), que mesmo não atendendo todos os requisitos estabelecidos pelo *International Mobile Telecommunications* (IMT) da ITU – Setor de Radiocomunicação (ITU-R, do inglês *ITU - Radiocommunication Sector*) era a escolha para assumir o posto de quarta geração de rede móvel (4G) pelo 3GPP (Gupta; Jha, 2015). Os requisitos estabelecidos pelo IMT-Advanced apenas foram completamente atendidos pelo 3GPP nas próximas *Releases*, com a chegada do *LTE-Advanced* (LTE-A).

O LTE é uma especificação que oferece muita flexibilidade em termos de formas de implantação e ofertas de serviços. As principais características do LTE são o custo reduzido da rede (custo por *bit*), melhor provisionamento de serviços, baixa latência, maior eficiência na utilização do espectro sem fio e compatibilidade com sistemas 3GPP (Dahlman; Parkvall; Skold, 2016). O LTE-A fornece altas taxas

de dados e maior capacidade da rede em comparação ao LTE. Além disso, pode-se destacar algumas melhorias (Amitabha Ghosh, 2011):

- Mobilidade: suporte em toda a rede celular para velocidades móveis até 350 Km/h;
- *Orthogonal Frequency-Division Multiple Access* (OFDMA): utilizado no *downlink* permite alta eficiência espectral, oferece robustez contra interferência de múltiplos caminhos, além de suportar técnicas avançadas como escalonamento *channel-dependent* no domínio da frequência e técnicas *Multiple-Input Multiple-Output* (MIMO);
- TDD: transmissão e recepção na mesma frequência da portadora em intervalos de tempo diferentes;
- FDD: transmissão e recepção em frequências de portadora diferentes;
- Taxa de transmissão: nas *Releases 8/9* a taxa de transmissão pode alcançar até 100 Mbps no *downlink* e 50 Mbps no *uplink*. Na *Release 10* com a estratégia de agregação de portadoras pode-se alcançar até 3 Gbps no *downlink* e 1,5 Gbps no *uplink*, com 100 MHz de largura de banda. Já a *Release 13* elevou essas taxas para 25 Gbps no *downlink* e 12,5 Gbps no *uplink*, considerando 640 MHz de largura de banda;
- Aprimoramento de técnicas de múltiplas antenas: o esquema MIMO melhora a eficiência espectral, a vazão média da célula e o desempenho na borda da célula. A introdução do *Coordinated Multi-Point* (CoMP) permite que as antenas de transmissão/recepção da célula servidora e vizinhas sejam utilizadas para melhorar a qualidade do sinal e reduzir interferências co-canal das células vizinhas.

Para resolver os problemas de demanda de largura de banda e latência o 3GPP e o ETSI propuseram uma arquitetura renovada para o padrão de quinta geração de rede móvel (5G), chamado de *New Radio* (NR). Esta nova arquitetura é baseada no conceito de nuvem, que substitui grande parte do hardware em favor de funções de rede virtualizadas conectadas a diferentes células e à Internet (Ordonez-Lucena *et al.*, 2017).

O 4G LTE/LTE-A e o 5G NR são padrões de rede móvel que compartilham uma grande quantidade de semelhanças. As principais diferenças entre eles vêm da natureza virtualizada da infraestrutura NR, com foco na redução de custos, facilidade de manutenção e atualização, enquanto o LTE/LTE-A depende de hardware dedicado para a maioria de seus componentes.

O principal impulsionador do 4G foi a banda larga móvel, enquanto que para o 5G tem-se a baixa latência, a alta taxa de transferência e a mobilidade. Casos de uso de redes 5G são, frequentemente, divididos nas três categorias descritas a seguir (Dahlman, 2021) e ilustradas na Figura 1:

- Comunicação entre Máquinas (MTC, do inglês *Machine Type Communication*): tem como principais aplicações edifícios e cidades inteligentes, transporte, logística e agricultura inteligente. Para essas aplicações, os principais requisitos são alta cobertura, densidade de usuários e economia de energia;
- Banda Larga Móvel Aprimorada (eMBB, do inglês *enhanced Mobile Broadband*): principais serviços são casas inteligentes, Realidade Virtual e Aumentada, *streaming* de alta resolução e trabalho/jogos em nuvem. Os principais requisitos para atender os serviços da eMBB são capacidade de banda e inteligência da rede;
- Comunicação de Baixa Latência e Alta Confiabilidade (URLLC, do inglês *Ultra-Reliable and Low-Latency Communications*): tecnologias que operam sobre baixa latência, grande área de cobertura e forte segurança, tais como robôs, drones, veículos autônomos, saúde digital e automação industrial.

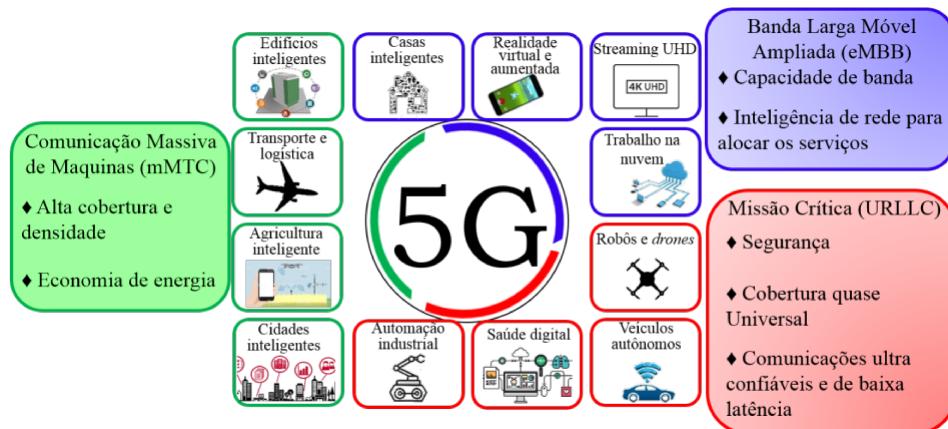


Figura 1 – Principais aplicações dos sistemas 5G (Castro Alves, 2020)

O ITU-R iniciou os trabalhos de desenvolvimento do 5G NR em 2012, sendo identificada por IMT-2020. Como parte do desenvolvimento, o ITU-R definiu um conjunto de requisitos para a tecnologia IMT-2020 suportar os casos de uso do 5G e cenários de uso identificados através de órgãos regionais, projetos de pesquisa, operadores e outras organizações (Dahlman, 2021). As tendências que impulsionam a evolução tecnológica 5G, são (Zaidi *et al.*, 2018):

- Baixa latência;
- Alta densidade de usuários;
- Alta mobilidade;
- Internet das coisas;
- Localização e rastreamento de alta precisão.

Além da comunicação convencional entre pessoas (de humano para humano), será ampliado o paradigma de comunicação de coisa para coisa e de humano para coisa. Como exemplo tem-se a comunicação instantânea entre veículos, infraestrutura rodoviária e centros de controle viabilizada, permitindo a coordenação do tráfego, a prevenção de acidentes e a otimização das rotas. Além disso, as redes móveis desempenham um papel crucial na computação de borda, fornecendo o processamento de dados e a inteligência necessária para as funcionalidades avançadas dos veículos autônomos, como a condução autônoma e a experiência de entretenimento dos passageiros. Como resultado, espera-se que o número de dispositivos de comunicação e a quantidade de dados transmitidos aumentem drasticamente (Xu *et al.*, 2021).

Assim, torna-se necessário que as redes móveis celulares tenham que lidar com o aumento cada vez maior de dispositivos conectados e com a alta demanda dos usuários por conteúdos de vídeo e jogos online. Inicialmente, a adição de estação base (BS, do inglês *Base Station*) de longo alcance foi uma estratégia adotada para aumentar a capacidade das redes. No entanto, essa abordagem apresentou limitações, como os altos custos de implementação e a necessidade de grandes áreas de instalação (Zhang; Chu; Wen, 2013). Embora a melhoria da eficiência espectral continue sendo essencial, ela se aproxima dos limites teóricos, o que torna insuficiente utilizá-la como única solução para lidar com a crescente demanda. A escassez do espectro de rádio tem incentivado a busca por alternativas, como as Redes Heterogêneas (HetNets, do inglês *Heterogeneous Networks*), que combinam macrocélulas com micro, pico e femtocélulas, aumentando a eficiência espectral e ampliando a cobertura em áreas densamente povoadas (Yuan, 2015). No entanto, HetNets exigem técnicas eficazes de gerenciamento de interferência, pois células menores tendem a sofrer interferências significativas das macrocélulas. Redes ultradensas (UDN, do inglês *Ultra Dense Networks*), compostas por um grande número de nós de baixa potência, também são uma solução promissora para suportar o tráfego crescente do 5G (Kazi; Wainer, 2018).

O uso eficiente dos recursos de rádio é essencial para atender às metas de desempenho do sistema e satisfazer as necessidades dos usuários de acordo com requisitos específicos de qualidade de serviço (QoS, do inglês *Quality of Service*). Assim, estratégias efetivas para alocação de recursos tornam-se cruciais, e os esquemas de escalonamento são responsáveis por estas tarefas, além de decidir a ordem do fluxo de aplicações (Dahlman, 2023). Os esquemas de escalonamento devem proporcionar garantias de desempenho das aplicações críticas e promover critérios de justiça no compartilhamento de recursos.

1.1 Estado da Arte

A crescente demanda por aplicações multimídia, comunicação em tempo real e serviços baseados em nuvem tem ampliado significativamente o tráfego de dados em redes móveis celulares. Em HetNets LTE-A/5G, a integração de diferentes tipos de células — como macrocélulas, microcélulas, picocélulas e femtocélulas — aumenta a capacidade e a cobertura da rede, mas introduz desafios complexos no gerenciamento de recursos de rádio, especialmente no *uplink*, onde limitações de potência e mobilidade dos usuários são críticas (Arafat; Gregory, 2020; Gatti *et al.*, 2022).

Garantir a QoS em cenários com tráfego misto, Tempo Real (RT, do inglês *Real Time*) e Não em Tempo Real (NRT, do inglês *Non Real Time*), e usuários heterogêneos exige soluções adaptativas que equilibrem latência, vazão, justiça e eficiência energética. Abordagens tradicionais, como *Proportional Fair* (PF) e *Maximum Throughput* (MT), embora eficazes em ambientes homogêneos, mostram-se limitadas em contextos dinâmicos com múltiplos requisitos de QoS. Soluções mais recentes, como algoritmos baseados em pesos (Arafat; Gregory, 2020), tesselação de Voronoi (Gatti *et al.*, 2022) e alocação por tempo médio de espera (Tata; Fellag; Kadoch, 2020), avançam nessa direção, mas ainda negligenciam aspectos essenciais, como restrições de potência, subpriorização do tráfego RT e critérios abrangentes para *handover*, como relação sinal-interferência mais ruído (SINR, do inglês *Signal to Interference plus Noise Ratio*) e carga de tráfego celular (Nomeir; Gadallah; Seddik, 2023; Kasim *et al.*, 2020; Shaddad *et al.*, 2022).

Diante desses desafios, técnicas de aprendizado por reforço profundo (DRL, do inglês *Deep Reinforcement Learning*) emergem como alternativas promissoras para otimizar a alocação dinâmica de recursos. O Aprendizado por Reforço Q Profundo (DQL, do inglês *Deep Q-Learning*), em particular, permite que agentes aprendam políticas de escalonamento e alocação eficientes por meio da interação contínua com o ambiente, mesmo sem um modelo explícito da rede (Liang; Lyu; Fan, 2023; Tang; Zhou; Kato, 2020). Aplicações recentes demonstram ganhos significativos em *throughput*, latência e eficiência energética, além de possibilitar um controle mais preciso de potência e interferência. Um exemplo é a combinação de DQL com métodos multicritério, como a Avaliação Ponderada Agregada por Soma e Produto (WASPAS, do inglês *Weighted Aggregated Sum Product Assessment*), para decisões de *handover* em ambientes densos e de alta mobilidade (Mukhopadhyay; Das, 2021; Ferng *et al.*, 2019).

A falta de priorização do tráfego RT pode prejudicar aplicações críticas, como VoIP e videoconferência (Nomeir; Gadallah; Seddik, 2023). Além disso, a otimização de potência para usuários de borda é frequentemente negligenciada, impac-

tando o equilíbrio na alocação de recursos e a QoS geral (Kasim *et al.*, 2020). Portanto, o desenvolvimento de soluções integradas — que combinem escalonamento inteligente, alocação energética eficiente e decisões de mobilidade baseadas em múltiplos parâmetros — é essencial para aprimorar continuamente a QoS em HetNets LTE-A/5G.

1.2 Definição do Problema

O surgimento de novos serviços e aplicativos tem contribuído para o aumento do acesso à Internet através das redes móveis celulares e impulsionado as pesquisas para desenvolvimento de novas técnicas que maximizem as taxas de transmissão e minimizem o atraso na rede. A ampliação da capacidade e a garantia dos parâmetros de QoS em redes móveis celulares, com recursos limitados e compartilhados, é uma tarefa complexa que requer informações sobre a variação do canal, a propagação do sinal, o comportamento do tráfego e a alocação de recursos.

Tecnicamente, existem três maneiras possíveis para aumento de capacidade em redes: ampliação da largura de banda; otimização da eficiência espectral e densificação da rede (Trabelsi, 2016). Dada a limitação física, bem como alguns aspectos reguladores, a ampliação da largura de banda encontra-se restrita em determinadas situações. A otimização da eficiência espectral requer técnicas para melhoramento da relação sinal-ruído no canal e/ou aumento na quantidade de canais disponíveis, enquanto a densificação da rede é um importante recurso para compensar a limitação da banda disponível, da distância e das barreiras físicas.

A técnica mais promissora para atender o aumento de demanda dos usuários móveis tem sido a densificação celular (Edet; Ekpe; Ukommi, 2025; Alam *et al.*, 2024). A Tabela 1 diferencia os tipos de células quanto à área de cobertura e potência máxima de transmissão. As macrocélulas (M-BS) fornecem ampla cobertura e são responsáveis pelo atendimento em grandes áreas geográficas. As microcélulas (Mi-BS) operam com menor potência e são utilizadas para aumentar a capacidade em regiões com alta densidade de usuários. As picocélulas (P-BS) são implantadas em ambientes internos ou áreas restritas, como centros comerciais e estações de transporte. As femtocélulas (F-BS), por sua vez, são voltadas para uso residencial ou em pequenos escritórios, proporcionando cobertura local com baixo custo e interferência reduzida (Yuan, 2015). O uso de células menores contribui para o aumento da complexidade na alocação de recursos e gerenciamento de interferência, especialmente em redes ultra densas (Kazi; Wainer, 2018).

A utilização de esquemas de escalonamento responsável pela alocação dos bloco de recursos (RB, do inglês *Resource Block*) para diversos usuários é uma das principais características do sistema de redes móveis celulares, tendo como ob-

Tabela 1 – Comparativo entre tipos de células, adaptado de (Yuan, 2015)

	Macrocélula	Microcélula	Picocélula	Femtocélula
Cobertura	2-35 Km	\leq 2 Km	\leq 200 m	\leq 50 m
Potência	5-40 W	\leq 5W	\leq 2W	\leq 200mW

jetivo indicar a forma de distribuição dos recursos disponíveis entre os usuários ativos no sistema (Piro *et al.*, 2011). Para realizar a distribuição dos blocos de recursos de rádio, o escalonador pode levar em consideração informações da camada física, configurações da BS, parâmetros de QoS das aplicações, dentre outros (Ragaleux; Baey; Karaca, 2017). Um mapa de alocação de recursos de rádio é construído pelo escalonador com base nestas informações para que seja efetivada a distribuição dos blocos de recurso aos usuários (Dahlman, 2023). Um bom esquema de escalonamento permite maximizar a eficiência espectral, reduz os impactos negativos provocados pelas especificidades do ambiente de rede sem fio, além de promover a justiça na distribuição dos blocos de recursos e atender os requisitos de QoS das aplicações (Taha; Ali, N. A.; Hassanein, 2011). O esquema de escalonamento determinará quais usuários serão priorizados e para isso um conjunto de requisitos estabelecidos pelo projetista são considerados.

O 3GPP não especifica um padrão de implementação para esquemas de escalonamento. Por esse motivo fabricantes de produtos cuja tecnologia contenha especificação do 3GPP possuem a liberdade para desenvolvimento e implementação de seus próprios esquemas (Dahlman; Parkvall; Skold, 2016). Isso justifica a quantidade de abordagens diferentes de esquemas de escalonamento presente na literatura. O escalonamento *downlink* e *uplink* representam duas entidades diferentes presentes na camada de Controle de Acesso ao Meio (MAC, do inglês *Media Access Control*) da BS (Abu-Ali *et al.*, 2014). Com o crescimento do tráfego *uplink* impulsionado por aplicações em tempo real, dispositivos conectados e serviços em nuvem, aumentam as exigências sobre o desempenho das redes móveis celulares. O escalonamento eficiente no *uplink* é essencial para garantir qualidade de serviço, especialmente em cenários heterogêneos com usuários móveis e requisitos diversos. A alocação de recursos inadequada pode comprometer a equidade, elevar a latência e aumentar a taxa de perda de pacotes. Além disso, o consumo energético e as restrições de potência dos dispositivos tornam-se fatores críticos. Diante desses desafios, torna-se necessário investigar soluções de escalonamento *uplink* mais adaptativas, eficientes e conscientes da QoS. O processo de escalonamento visa resolver o problema da alocação justa e eficiente de recursos escassos de tempo e frequência entre múltiplos UEs em um ambiente dinâmico e heterogêneo, marcado por mobilidade, variações de canal e interferência intercelular. Esse processo busca conciliar três objetivos principais: maximizar a taxa de transmissão agregada da rede, garantir os requisitos de

QoS e assegurar a justiça na distribuição dos RBs (Piro *et al.*, 2011). Como a grade de alocação de recursos é dividida em tempo, frequência e espaço, o processo de alocação torna-se ainda mais complexo (Dahlman, 2023). Outra restrição adotada nesse trabalho é a alocação contígua dos blocos de recursos aos usuários. Tal restrição amplia consideravelmente a complexidade do projeto do esquema de escalonamento (Mata; Guardieiro, 2017).

A alocação ótima de recursos aos usuários é um problema NP-Completo, por se tratar de um problema de alocação combinatória de recursos de alta complexidade (Huang *et al.*, 2010). Dessa forma, num cenário de rede ultra densa, um esquema de escalonamento e alocação de recursos que atenda as demandas de aplicações em tempo real no *uplink* e que maximize a capacidade total da rede, além de promover a justiça na distribuição dos blocos de recursos exige o uso de técnicas inovadoras, que minimize o custo computacional despendido pela solução.

1.3 Objetivo

O principal objetivo deste trabalho é desenvolver um esquema de alocação de recursos e gerenciamento de QoS em HetNets LTE-A/5G, visando garantir uma experiência de comunicação eficiente e confiável para os usuários finais. Este esquema deve ser capaz de otimizar a alocação de recursos de forma dinâmica, atendendo às características específicas das redes heterogêneas.

Para alcançar o objetivo principal, a pesquisa se desdobra nos seguintes objetivos específicos:

- Analisar os desafios e oportunidades na alocação de recursos e gerenciamento de QoS em HetNets LTE-A/5G. Essa análise abrange a revisão da literatura e a investigação de abordagens promissoras para otimizar a QoS em HetNets;
- Propor um novo esquema de alocação de recursos que leve em consideração as características específicas das HetNets LTE-A/5G. O esquema proposto deve ser capaz de lidar com a interferência intercelular, a mobilidade dos usuários e a diversidade de aplicações e serviços, com o objetivo de garantir uma QoS diferenciada para diferentes classes de tráfego;
- Desenvolver um mecanismo de *handover* inteligente que utilize técnicas de inteligência artificial (IA)/aprendizado de máquina (ML, do inglês *Machine Learning*) para otimizar as decisões de *handover* em HetNets. O mecanismo proposto deve ser capaz de prever a movimentação dos usuários e as condições de tráfego, e tomar decisões de *handover* de forma proativa, minimizando a interrupção do serviço e garantindo uma experi-

ência de comunicação contínua;

- Avaliar o desempenho do esquema de alocação de recursos através de simulações computacionais. As simulações são realizadas com diferentes densidades de usuários e o desempenho do sistema é avaliado com base em métricas de QoS relevantes, como atraso, taxa de transferência, *jitter* e taxa de perda de pacotes;
- Comparar o desempenho do esquema proposto com outras abordagens existentes na literatura. Essa comparação permite avaliar a eficiência e a efetividade do esquema proposto em relação às técnicas tradicionais de alocação de recursos.

1.4 Solução Proposta

O presente trabalho propõe uma solução integrada para a alocação eficiente de recursos no *uplink* em redes heterogêneas LTE-A/5G, combinando técnicas de aprendizado por reforço profundo e métodos de otimização multicritério. O objetivo central da proposta é garantir a QoS em cenários com elevada mobilidade e requisitos diversos de tráfego, especialmente em ambientes ultra densos.

A metodologia desenvolvida é composta por três componentes principais. O primeiro é o escalonamento de objetivo duplo, baseado em um algoritmo adaptativo que prioriza, de forma diferenciada, o tráfego RT e NRT, considerando subclasses de aplicações críticas, como videoconferência, *streaming* e VoIP. Esse mecanismo visa garantir a alocação justa e eficiente dos blocos de recursos entre os usuários ativos.

O segundo componente trata da alocação de recursos com restrição de energia, empregando a técnica DQL para otimizar a distribuição dos blocos de recursos. Essa abordagem leva em conta parâmetros de QoS, consumo energético dos dispositivos e variações dinâmicas nas condições do canal. O DQL permite a aprendizagem de políticas de alocação eficientes por meio da interação contínua com o ambiente de rede, promovendo o aumento do *throughput* e a mitigação da interferência intercelular.

O terceiro componente corresponde à tomada de decisão de *handover* inteligente, utilizando o método WASPAS para avaliar múltiplos critérios, como a SINR, a carga de tráfego celular e a mobilidade do usuário. Essa técnica permite decisões mais precisas na troca de células, contribuindo para a redução das quedas de conexão e para a continuidade do serviço.

A proposta foi avaliada por meio de simulações realizadas no simulador ns-3, nas quais se observaram melhorias significativas em métricas como atraso, eficiência espectral e taxa de perda de pacotes. O Capítulo 4 apresenta em detalhes a arquitetura da solução, os aspectos de implementação do algoritmo DQL — incluindo o

espaço de estados, as ações e a função de recompensa — bem como a configuração dos cenários de simulação utilizados para validação da proposta.

1.5 Contribuições

Esta tese apresenta contribuições relevantes para o escalonamento e a alocação de recursos no enlace *uplink* em redes heterogêneas LTE-A/5G, por meio da integração de técnicas de inteligência artificial com métodos de apoio à decisão multicritério. As principais contribuições podem ser destacadas da seguinte forma:

- Desenvolvimento de um escalonador de pacotes baseado no algoritmo *Smoothed Round-Robin* (SRR), com dupla priorização, destinado à classificação e ao tratamento diferenciado dos fluxos de tráfego RT e NRT, de acordo com os requisitos de QoS de cada classe de serviço;
- Proposição de um mecanismo de alocação de recursos com restrição de potência, utilizando a técnica DQL, capaz de aprender políticas ótimas por meio da interação com o ambiente, adaptando-se dinamicamente às condições do canal e às exigências de QoS dos usuários;
- Formulação de uma estratégia de *handover* inteligente, baseada no método multicritério WASPAS, a qual considera simultaneamente múltiplos critérios – como SINR, mobilidade dos usuários e carga de tráfego celular – visando à continuidade do serviço e à redução de quedas de chamadas;
- Integração dos módulos de escalonamento, alocação de recursos e decisão de *handover* em um sistema unificado, avaliado por meio de simulações conduzidas no simulador *Network Simulator 3* (ns-3);
- Demonstração de melhorias significativas em métricas de desempenho, como a redução do atraso médio, o aumento da vazão (*throughput*), a elevação do índice de justiça na distribuição de recursos, bem como a diminuição das taxas de perda de pacotes e de queda de chamadas, quando comparadas aos algoritmos *Proportional Fair* (PF) e *New Courteous Algorithm* (NCA).

Os resultados obtidos nesta pesquisa foram validados e divulgados na comunidade científica por meio da publicação em periódico especializado, conforme o artigo:

GARBAZZA, I. E.; SILVA, E. R.; GUARDIEIRO, P. R. Optimizing QoS in LTE-A/5G HetNets: a deep Q-learning approach to uplink resource allocation. **Telecommunication Systems**, v. 88, p. 49, 2025. DOI: <https://doi.org/10.1007/s11235-025-01277-z>.

Essas contribuições evidenciam o potencial da combinação entre aprendizado por reforço profundo e métodos de decisão multicritério para a otimização da

alocação de recursos em redes móveis celulares modernas, especialmente em contextos caracterizados por elevada heterogeneidade, mobilidade e demanda por QoS.

1.6 Sinopse dos Capítulos

Esta tese de doutorado está estruturada em capítulos, organizados de forma a apresentar a pesquisa de maneira clara, concisa e logicamente estruturada. Os capítulos abordam os seguintes tópicos:

O Capítulo 2 apresenta a fundamentação teórica, abordando os conceitos fundamentais de redes móveis celulares, redes heterogêneas, alocação de recursos, gerenciamento de QoS e técnicas de IA/ML.

O Capítulo 3 continua com a fundamentação teórica detalhando o processo de escalonamento e alocação de recursos em redes LTE-A/5G. Além disso, apresenta também o funcionamento do escalonador de pacotes e a distribuição dos recursos para os usuários.

O Capítulo 4 apresenta a solução proposta, o cenário de simulação, as métricas de avaliação e a metodologia de análise dos resultados.

No Capítulo 5 é apresentada a avaliação do esquema de escalonamento e alocação de recursos, apresentando o resultado das simulações e uma análise comparativa com outras abordagens.

Por fim, o Capítulo 6 discute as conclusões da pesquisa e as direções para trabalhos futuros.

2 REDES LTE/LTE-A E 5G NR

O aumento da demanda por conectividade, altas taxas de transmissão e baixa latência tem impulsionado a evolução das redes móveis celulares. A transição das redes 4G para as redes 5G representa um marco significativo, introduzindo novas arquiteturas, técnicas e desafios para atender às exigências de aplicações como IoT, veículos autônomos, realidade aumentada (RA) e realidade virtual (VR). Este capítulo apresenta os fundamentos teóricos essenciais para compreender o funcionamento das redes LTE-A e 5G NR.

A evolução das redes móveis, desde as primeiras gerações (1G e 2G) até as redes 5G, foi caracterizada por avanços significativos em capacidade, eficiência espectral e suporte a novos serviços. O LTE, introduzido no 4G, trouxe uma arquitetura mais simplificada e flexível, proporcionando altas taxas de transmissão e baixa latência. Com o advento do LTE-A, foram incorporadas técnicas avançadas, como agregação de portadoras, MIMO massivo e coordenação aprimorada de interferência entre células (eICIC, do inglês *Enhanced Inter-Cell Interference Coordination*), otimizando a capacidade e a eficiência da rede.

A introdução do 5G NR representa uma revolução na comunicação sem fio, adotando uma arquitetura virtualizada e baseada em nuvem para oferecer maior flexibilidade e escalabilidade. A utilização de novas bandas de frequência, incluindo ondas milimétricas, aliada a técnicas como *beamforming* e *network slicing*, possibilita o atendimento de casos de usos diversos, como mMTC e URLLC. No entanto, a complexidade das redes modernas também introduz desafios, como interferência intercelular, variação dinâmica das condições do canal e demandas conflitantes de tráfego RT e NRT.

A heterogeneidade dessas redes, marcada pela coexistência de macro, micro, pico e femtocélulas, amplia a capacidade e a cobertura, mas aumenta a complexidade no gerenciamento de recursos. A alocação de espectro precisa equilibrar eficiência espectral, distribuição justa de recursos e consumo energético, atendendo às necessidades específicas de cada classe de serviço. A QoS torna-se um fator essencial, exigindo técnicas de escalonamento que priorizem aplicações sensíveis a atrasos, como videoconferência e VoIP, sem comprometer a eficiência de alocação de banda para aplicações com tráfego elástico.

Diante da dinamicidade das redes densas, estratégias baseadas em inteligência artificial têm sido aplicadas para otimização de recursos. Em particular, técnicas de aprendizado por reforço, como o DQL, destacam-se por sua capacidade de adaptação a ambientes dinâmicos, permitindo decisões inteligentes de alocação com base nas condições do canal, restrições energéticas e padrões de tráfego.

Este capítulo discute a evolução tecnológica das redes LTE-A/5G, a arqui-

tetura de sistemas heterogêneos, os desafios do *handover* e os fundamentos dos algoritmos de aprendizado de máquina. Esses conceitos estabelecem a base teórica para a proposta metodológica apresentada nos capítulos subsequentes, que integra técnicas híbridas para otimização de redes densas.

2.1 A Evolução das Redes Móveis: do LTE ao 5G-Advanced

A introdução do LTE trouxe mudanças significativas à arquitetura das redes celulares. A Figura 2 apresenta a arquitetura de rede do LTE, dividida em quatro domínios principais: Equipamento de Usuário (UE, do inglês *User Equipment*), Rede de Acesso de Rádio Terrestre UMTS Evoluída (E-UTRAN, do inglês *Evolved UMTS Terrestrial Radio Access Network*), Rede de Núcleo de Pacotes Evoluída (EPC, do inglês *Evolved Packet Core*) e o domínio de Serviços. A camada de conectividade do protocolo de Internet (IP), composta por UE, E-UTRAN e EPC, é denominada Sistema de Pacotes Evoluído (EPS, do inglês *Evolved Packet System*).

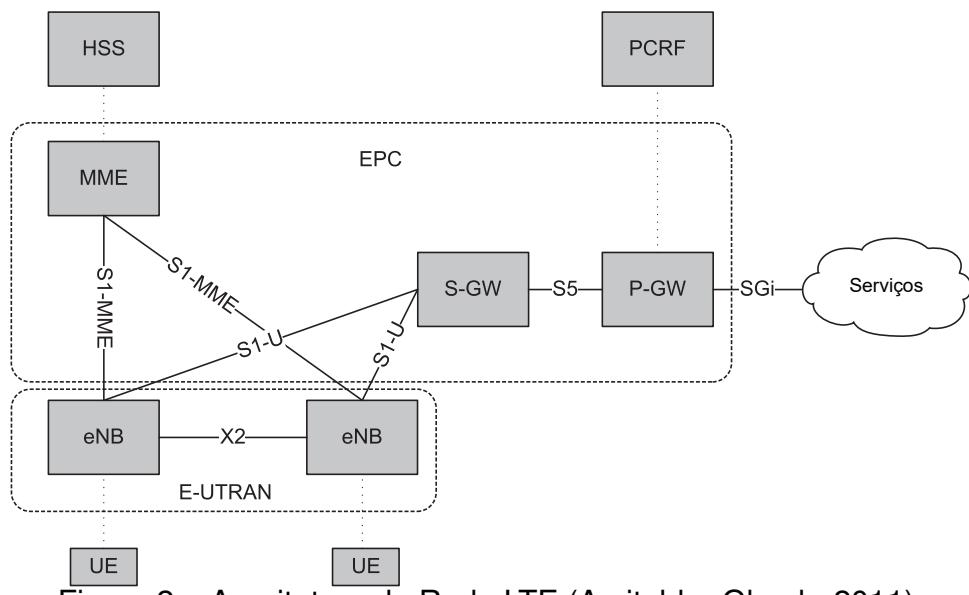


Figura 2 – Arquitetura da Rede LTE (Amitabha Ghosh, 2011).

O UE é o dispositivo utilizado pelos usuários para comunicação, contendo o Módulo de Identidade Universal do Assinante, (USIM, do inglês *Universal Subscriber Identity Module*), que permite identificar e autenticar o usuário, além de proteger a transmissão na interface de rádio com chaves de segurança. O UE também atua como plataforma para aplicativos de comunicação.

A E-UTRAN possui como principal elemento a *E-UTRAN NodeB* (eNodeB), uma estação base que controla as funções de rádio na rede fixa. As eNodeB conectam o UE ao EPC e são interligadas por meio da interface X2, desempenhando funções como compactação de cabeçalho IP e cifragem de dados do plano do usuário.

O EPC é responsável pelo encaminhamento de pacotes de dados, auten-

ticação e controle de mobilidade. Além disso, lida com funções de gerenciamento de mobilidade e autenticação. No EPC, destacam-se os seguintes elementos:

- *Mobility Management Entity* (MME): elemento de controle responsável pela autenticação, segurança, gerenciamento de mobilidade e conectividade de serviço;
- *System Architecture Evolution Gateway* (SAE-GW): composto pelo *Serving Gateway* (S-GW), responsável pelo gerenciamento do plano do usuário, e pelo *Packet Data Network Gateway* (P-GW), que atua como roteador entre o EPS e as redes externas;
- *Policy and Charging Rule Function* (PCRF): responsável por prover QoS no LTE, determinar as políticas da rede e tarifar as chamadas de dados;
- *Home Subscription Server* (HSS): armazena identificação e autenticação dos usuários da rede LTE.

O Domínio de Serviços abrange serviços de alto nível, como autenticação de usuário, qualidade de serviço (QoS) e funções de aplicativos. Ele inclui tanto os serviços fornecidos pela operadora de rede móvel quanto os serviços disponíveis na Internet, suportados pelo *IP Multimedia Sub-system* (IMS). O IMS, parte integrante dos padrões 3GPP, fornece uma interface IP comum para sinalização, tráfego e aplicativos, como o VoLTE (Voice over LTE), que permite comunicação de voz sobre redes LTE.

A padronização do LTE iniciou-se em 2004 pelo 3GPP, durante o desenvolvimento do HSPA+ (*High Speed Packet Access plus*). O primeiro lançamento do LTE, denominada *Release 8* do 3GPP, foi introduzido em dezembro de 2008. A *Release 9*, finalizada em dezembro de 2009, trouxe aprimoramentos à tecnologia. Em 2011, a *Release 10* introduziu o *LTE-Advanced*, com recursos avançados que marcaram um salto na evolução da tecnologia. As versões posteriores a *Release 10* são consideradas *LTE-Advanced*. A Figura 3 ilustra essa evolução.



Figura 3 – Evolução da Rede LTE, adaptado de (Dahlman, 2023).

O desenvolvimento do 5G marca uma transformação significativa na comunicação sem fio, atendendo à crescente demanda por conectividade onipresente e serviços avançados. Alinhado ao IMT-2020, estabelecido pela ITU-R, o 5G introduz melhorias em taxa de dados, latência, eficiência espectral e energética, além de maior densidade de conexões, conforme pode ser observado na Figura 4. Essas inovações possibilitam avanços em setores como transporte, saúde e logística, suportando bilhões de dispositivos simultaneamente em ambientes de alta densidade.

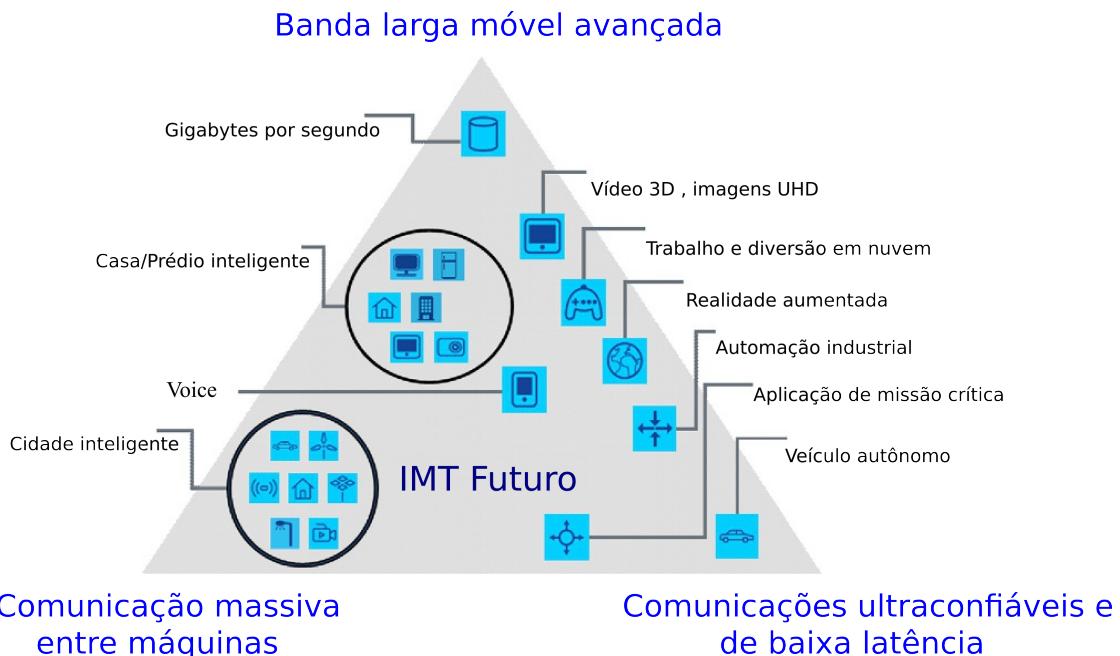


Figura 4 – Casos de uso do IMT-2020 e mapeamento para cenários de utilização, adaptado de (Dahlman, 2023).

O 5G baseia-se em uma arquitetura plana, herdada do LTE, que eliminou componentes intermediários como controladores de sites (*Radio Network Controller* e *Base Station Controller*), simplificando conexões e reduzindo a latência. Nessa arquitetura, o EPC conecta a rede a sistemas externos, como a Internet, enquanto interfaces como Uu, S1 e SGi garantem a comunicação eficiente entre os componentes, conforme apresentado na Figura 5. A interface Uu é a interface de rádio, ou seja, a interface *over-the-air*, entre o UE e a BS da rede de acesso de rádio, e é responsável pela transmissão e recepção de sinais de rádio que carregam tanto os dados do UE quanto os sinais de controle. No LTE/LTE-A, a interface Uu é baseada na tecnologia E-UTRAN. Enquanto que no 5G a interface Uu é baseada na tecnologia NR. A NG-Uu também lida com a sinalização de controle entre o UE e o gNodeB, utilizando protocolos NR específicos. A interface X2 é uma interface entre BSs vizinhos no LTE/LTE-A, que permite a comunicação direta entre as BSs sem a necessidade de passar pela rede central para funções relacionadas à mobilidade do UE, dentre outras. No 5G a interface equivalente à X2 é chamada de Xn. Na rede 5G ainda existem a funcionalidade de gerenciamento de mobilidade e acesso, (AMF, do inglês *Access and Mobility Management Function*) e a função do plano de usuário (UPF, do inglês *User Plane Function*). A AMF é responsável por funções relacionadas ao acesso do UE à rede e ao gerenciamento de mobilidade. Já a UPF trata o tráfego de dados do UE, incluindo roteamento e encaminhamento de pacotes entre a rede de acesso e as redes externas, como a Internet. Tecnologias avançadas, como MIMO e OFDM, fortalecem a capacidade do 5G de atender aos rigorosos requisitos do IMT-2020. A alta velocidade de transmissão e a latência quase desprezível (Rischke et al., 2022) tor-

nam o 5G ideal para aplicações críticas, como direção autônoma e telemedicina, além de serviços que demandam conexões rápidas, como *streaming* de vídeo em ultra-alta definição e realidade virtual.

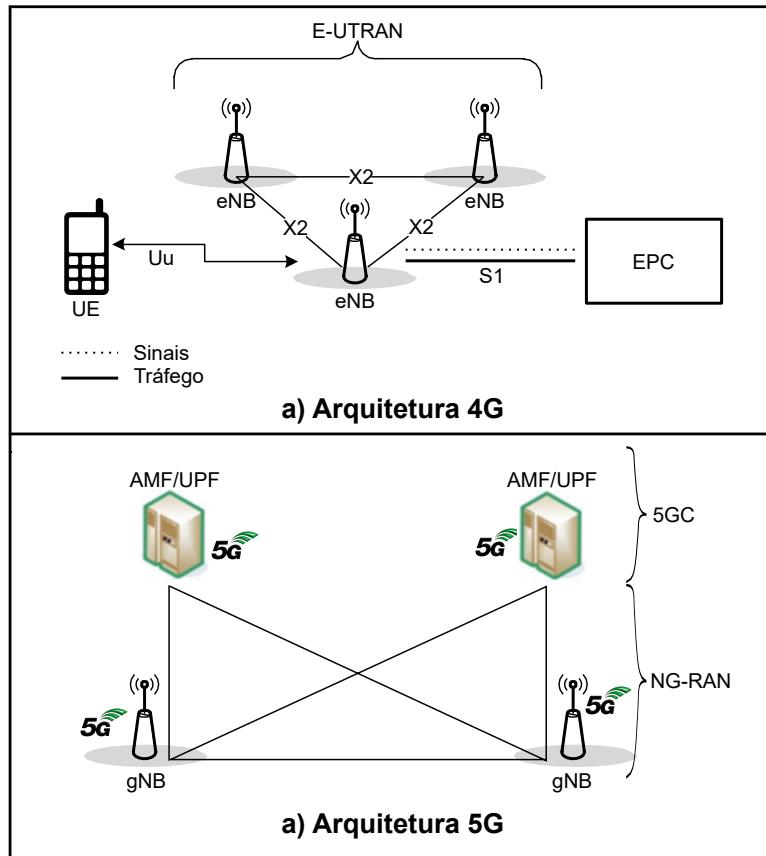


Figura 5 – Arquitetura Geral (Semov *et al.*, 2020).

A introdução do 5G representa uma evolução significativa em relação às gerações anteriores de redes móveis, atendendo às demandas por maior capacidade, menor latência e suporte a uma densidade elevada de dispositivos conectados (Dangi *et al.*, 2021). Seus cenários de uso podem ser classificados em três pilares principais (Dangi *et al.*, 2021; Ullah *et al.*, 2023; Kamal *et al.*, 2021):

- eMBB: focada em fornecer altas taxas de transmissão e conectividade robusta em ambientes densamente povoados, a eMBB suporta aplicações como streaming de vídeo em 4K/8K, realidade virtual (VR, do inglês *Virtual Reality*) e realidade aumentada (AR, do inglês *Augmented Reality*). Com velocidades que podem alcançar Gbps, esta tecnologia viabiliza também o acesso à banda larga fixa por meio de dispositivos conectados ao 5G, otimizando a experiência do usuário mesmo em condições desafiadoras (Souza Lopes; Coelho Rocha; Teles Vieira, 2023; Dangi *et al.*, 2021);
- Comunicação Massiva do Tipo Máquina (mMTC, do inglês *massive Ma-*

chine Type Communication): projetadas para atender ao crescente número de dispositivos para Internet das Coisas (IoT, do inglês *Internet of Things*), as mMTC permitem a troca de pequenos pacotes de dados de maneira eficiente. Aplicações típicas incluem sensores em redes agrícolas, cidades inteligentes e redes de energia. O foco está na economia de energia e na capacidade de suportar altas densidades de dispositivos conectados, essenciais para a sustentabilidade e a longevidade das baterias em dispositivos IoT, em cenários como agricultura inteligente e gerenciamento automatizado de cidades inteligentes (Dangi *et al.*, 2021; Ullah *et al.*, 2023; D, R. S. V.; P; D, G. R. M., 2023);

- URLLC: as URLLC são destinadas a aplicações críticas que requerem alta confiabilidade e latências inferiores a 1 ms, como direção autônoma, cirurgias remotas e automação industrial. Essa categoria permite trocas de dados em tempo real, oferecendo maior precisão e segurança em cenários onde a resposta imediata é essencial (Siddiqui *et al.*, 2023; Haque *et al.*, 2023).

O 5G ainda apresenta o conceito de *network slicing*, que permite a criação de redes virtuais personalizadas. O *network slicing* possibilita que diferentes serviços tenham redes configuradas de acordo com suas necessidades específicas, como taxa de dados, QoS, confiabilidade e segurança (Figura 6). Essa funcionalidade é habilitada pela *Release 15 SA* e aprimorada na *Release 16*, permitindo às operadoras atender demandas de diferentes indústrias com maior flexibilidade (Dangi *et al.*, 2021). Além disso, a localização precisa no espaço tridimensional e o suporte a redes em malha aumentam a conectividade e criam novas oportunidades de aplicação.

Este capítulo explora a evolução das redes móveis IP, desde o LTE até o 5G NR, destacando suas principais características. Os conceitos fundamentais são apresentados com terminologia genérica para descrever os elementos das redes móveis, enquanto termos específicos do 3GPP são utilizados, quando necessário, para diferenciar as gerações, como a mudança na nomenclatura das estações base (eNB no 4G e gNB no 5G). A seguir é apresentado um resumo da evolução do LTE até o 5G *Advanced*.

2.1.1 Release 8

O LTE, inicialmente implementado com macro/microcélulas, superou o HSPA+ em termos de taxa de dados, capacidade e cobertura. A *Release 8* introduziu baixa latência, redução de custos, suporte a múltiplas antenas e largura de banda flexível. O LTE utiliza OFDMA para acesso de rádio, garantindo ortogonalidade entre usuários, controle de potência e coordenação de interferência.



Figura 6 – Exemplo de fatiamento dinâmico de rede para serviço de carro autônomo (Spadinger, 2021).

2.1.2 Release 9

A *Release 9* é a versão completa do LTE, que amplia as funcionalidades introduzidas na *Release 8*. Esta versão abrange serviços de *broadcast/multicast*, serviços de posicionamento avançados e funcionalidades aprimoradas para chamadas de emergência. Os aperfeiçoamentos incluem maior velocidade de conexão, redução da latência e aumento da eficiência espectral. Com essas melhorias incluídas na *Release 9* proporciona-se uma experiência de comunicação mais avançada e confiável aos usuários.

2.1.3 Release 10

A *Release 10*, concluída no final de 2010, introduziu melhorias significativas em desempenho e capacidade em comparação com as versões 8 e 9. Conhecido como *LTE-Advanced*, a *Release 10* e suas versões posteriores atendem plenamente aos requisitos do IMT-Advanced, consolidando-se como um sistema de comunicação móvel 4G. O *LTE-Advanced* incorpora recursos adicionais em relação às versões anteriores, incluindo:

- Agregação de portadoras;
- Técnicas avançadas de MIMO, 8x8 para o *downlink* e 4x4 para o *uplink*;

- Redes Heterogêneas;
- Coordenação aprimorada de interferência entre células.

2.1.4 Release 11

A *Release 11* adicionou melhorias na Agregação de Portadoras, MIMO, nós de retransmissão, integração com Wi-Fi e eICIC. Também introduziu novas bandas de frequência, o Canal de Controle Físico de Downlink Aprimorado (ePDCCH, do inglês *Enhanced Physical Downlink Control Channel*) e um novo recurso chamado Multi-Ponto Coordenado (CoMP), permitindo a formação de feixes (*beamforming*) e o escalonamento de múltiplas portadoras. O ePDCCH adicionou suporte para aumento da capacidade de *downlink* e coordenação de interferência intercelular.

2.1.5 Release 12

O rápido aumento no uso de dados móveis estabeleceu altos requisitos de capacidade para a *Release 12*, além da demanda por uma experiência de usuário de alta qualidade e suporte a novas aplicações. As principais funcionalidades da *Release 12* incluem:

- Novas técnicas de antena e receptores avançados para maximizar a capacidade do sistema;
- Novos recursos para melhorar o suporte a HetNets, como conectividade dupla e recurso de ativação/desativação de *small cells*;
- Agregação de portadora intra-banda não contígua;
- Aprimoramento da interoperabilidade entre LTE e Wi-Fi;
- Modulação por Quadratura de Amplitude 256-QAM para downlink;
- Recursos de Comunicação Máquina-a-Máquina (MTC) que permitem que dispositivos usufruam dos serviços LTE; novas categorias de UE, suporte para taxas de dados baixas por meio de antena de recepção única, canal de dados *narrowband* e operação *half duplex*.

2.1.6 Release 13

A *Release 13* foi a primeira *Release* para o LTE-Advanced Pro. A *Release 13* trouxe vários novos recursos e melhorias para funcionalidades já existentes, introduzidas em *Releases* anteriores, algumas das principais funcionalidades incluem:

- Recursos para segurança pública, como D2D e *Proximity-Services* (ProSe);
- Aprimoramento adicional dos recursos de MTC e posicionamento em am-

- bientes internos;
- Aprimoramentos de Agregação de Portadoras;
 - Conectividade dupla em *small cells* e interoperabilidade com Wi-Fi;
 - Acesso Assistido Licenciado (LAA, do inglês *License Assisted Access*) em 5 GHz;
 - Beamforming e MIMO massivo;
 - Ponto-a-Multiponto em célula única (SC-PTM, do inglês *Single Cell Point to Multipoint*);
 - Redução de latência.

2.1.7 Release 14

A *Release 14* concentra-se em aprimorar elementos de missão crítica, com ênfase na implementação de serviços de vídeo e dados; introduzir funcionalidades de Veículo para Tudo (V2X, do inglês *Vehicle-to-Everything*), especialmente Veículo para Veículo (V2V, do inglês *Vehicle-to-Vehicle*); otimizar a Internet das Coisas Celular (CIoT, do inglês *Cellular Internet of Things*), oferecendo suporte para MTC nas redes 2G, 3G e 4G; e aprimorar a interface de rádio, particularmente na coordenação com WLAN e espectro não licenciado. Além disso, um conjunto de melhorias diversas, como Voz sobre LTE (VoLTE, do inglês *Voice over LTE*), IMS e relatórios de localização, também foram considerados.

2.1.8 Release 15

O 3GPP dividiu o padrão 5G em duas etapas: a *Release 15*, correspondente à Fase 1 do NR, e a *Release 16*, correspondente à Fase 2 do NR. Na Fase 1, elementos comuns ao LTE e NR, como a multiplexação por divisão de frequência ortogonal (OFDM, do inglês *Orthogonal Frequency Division Multiplexing*), são mantidos. A implementação completa do NR exige a instalação de um volume significativo de novo hardware. Para otimizar o uso do hardware existente, foi adotada uma abordagem em fases, com uma versão não autônoma (NSA, do inglês *Non-StandAlone*), que permite utilizar a rede de núcleo EPC do LTE ligada à eNodeBs e gNBs, e uma versão autônoma (SA, do inglês *StandAlone*), em que a gNB é conectada diretamente ao 5G Core Network (5GCN).

A Figura 7 apresenta os principais recursos e seus respectivos valores de referência, conforme definido pelo IMT-2020. Apresenta-se um conjunto de métricas críticas que caracterizam o desempenho esperado das redes 5G em comparação com gerações anteriores. As métricas incluem taxa de dados percebida pelo usuário (até 100 Gbit/s), eficiência espectral (até 30 vezes maior que o IMT-Advanced), mobilidade

suportada (até 500 km/h), latência (inferior a 1 ms), densidade de conexão (até 1 milhão de dispositivos por km²) e eficiência energética da rede (aproximadamente 100 vezes maior que o IMT-Advanced). Além disso, destaca a capacidade de tráfego por área (até 10 Tbit/s por m²) e a taxa máxima de dados (até 20 Gbit/s). Esses parâmetros definem os limites de desempenho necessário para aplicações críticas em cenários de alta densidade e baixa latência.

A diversidade de frequências presentes no NR gera variações significativas. Nas bandas de ondas milimétricas acima de 24 GHz, dispositivos e estações base requerem atualizações com tecnologia parcialmente nova. A utilização de MIMO massivo, *beamforming* e sistemas avançados de antenas integradas é essencial para a operação do NR, o que impõe novos requisitos de radiofrequênci a e critérios de avaliação de desempenho.

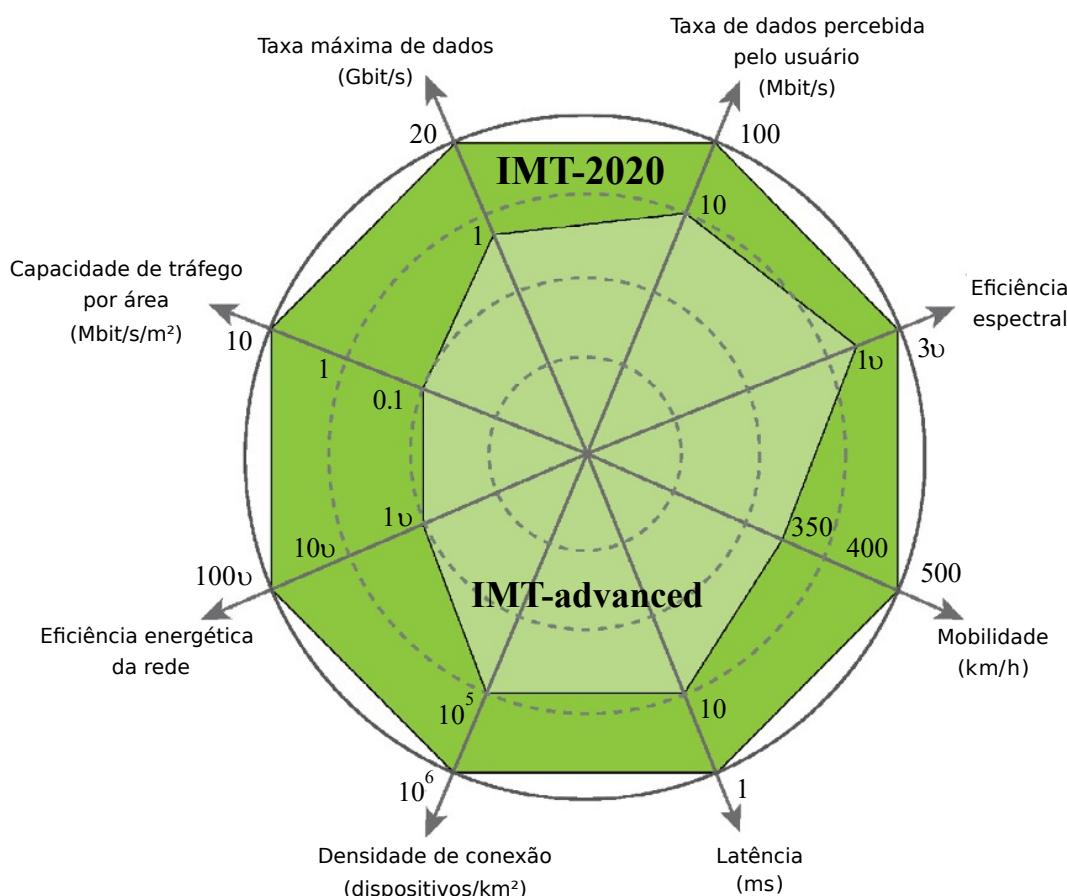


Figura 7 – Diferentes requisitos suportados pela rede 5G, adaptado de (Dahlman, 2023).

Na *Release 15*, as frequências para o NR são divididas em duas bandas:

- Faixa de Frequência 1 (FR1, do inglês *Frequency Range 1*): inclui todas as bandas existentes e novas abaixo de 6 GHz;
- Faixa de Frequência 2 (FR2, do inglês *Frequency Range 2*): abrange bandas de 24,25 a 52,6 GHz.

As bandas de operação do NR incluem espectros pareados e não pareados, exigindo flexibilidade no modo duplex. O NR apresenta várias vantagens em relação ao LTE:

- A introdução do conceito de eMBB, um dos principais requisitos do 5G;
- A adoção de uma nova estrutura de quadro e espectro;
- A exploração de frequências mais altas, disponibilizando espectro adicional e suportando larguras de banda amplas para altas taxas de dados;
- Design ultra eficiente para melhorar o desempenho energético e reduzir a interferência;
- Compatibilidade com versões futuras, antecipando novos casos de uso e tecnologias;
- Baixa latência, possibilitando melhor desempenho e novos casos de uso;
- O design focado em feixes permite a ampla utilização de *beamforming* e de múltiplos elementos de antena, tanto para transmissão de dados (como no LTE, em menor grau) quanto para procedimentos de controle, como o acesso inicial;
- Comunicação V2X fase 2;
- Fatiamento de redes lógicas fim-a-fim (*network slicing*).

2.1.9 Release 16

A *Release 16* do 3GPP desempenhou um papel fundamental na padronização de IoT no contexto das redes 5G (3GPP, 2018). Além disso, estabeleceu requisitos de latência para URLLC (3GPP, 2018). Finalizada em meados de 2020, a *Release 16* apresenta as seguintes características (3GPP, 2018):

- 5G fase 2;
- V2X fase 3;
- IoT Industrial;
- Confiabilidade de *link* - aprimoramentos no URLLC;
- Acesso a espectro não licenciado (NR-U, do inglês *New Radio-Unlicensed*) em 5GHz e 6GHz;
- *Backhaul* flexível e escalável, permitindo instalação de estações base de forma dinâmica;
- Acesso satelital em 5G;
- Redes sensíveis ao tempo, permitindo a entrega de pacotes de dados com tempo determinado.

2.1.10 Release 17

A *Release 17* do 3GPP introduz aprimoramentos significativos em diversos aspectos da rede de acesso, incluindo MIMO/*Beamforming*, *Dynamic Spectrum Sharing* (DSS), cobertura, transmissão de pequenos volumes de dados e posicionamento (3GPP, 2021). No que se refere à rede de acesso, destacam-se os seguintes aprimoramentos:

- Aprimoramento do NR *Sidelink*;
- Extensão da operação do NR para 71 GHz;
- Aprimoramento do RAN *Slicing* para NR;
- Aprimoramento das Redes Auto-organizáveis (SON, do inglês *Self Organizing Networks*) para redes 5G;
- Gerenciamento de Redes Não Públicas (NPN, do inglês *Non-Public Networks*) e suporte aprimorado para Redes Não Públicas;
- Aprimoramento do *Network Slicing* - Fase 2;
- Suporte à Conectividade, Identificação e Rastreamento de Sistemas Aéreos Não Tripulados (UAS, do inglês *Unmanned Aerial Systems*);
- Aprimoramento do suporte à Computação de Borda.

2.1.11 Release 18

A *Release 18* do 3GPP, definida como "5G-Advanced", teve suas discussões de escopo tecnológico iniciadas em junho de 2021, com o objetivo de iniciar os trabalhos de especificação em 2022 (3GPP, 2024b). O 3GPP padronizou o nome e o logo para 5G Advanced, conforme ilustrado na Figura 8.

Esta nova versão inclui aprimoramentos como melhoria no desempenho do *uplink* (taxa de dados, capacidade, cobertura), aprimoramento de funções para *eXtended Reality* (XR), economia de energia de rede e utilização de inteligência artificial e aprendizado de máquina para RAN (Bojović *et al.*, 2023). Além disso, a *Release 18* aborda extensão de funções voltadas para o 6G (3GPP, 2023).



Figura 8 – Logo 5G Advanced (3GPP, 2024b).

2.2 Arquitetura

A conectividade sem fio, proporcionada pela rede móvel celular, viabiliza a comunicação de dispositivos, potencialmente em movimento, denominados UEs. Inicialmente limitados a telefones celulares e *tablets*, os UEs atualmente abrangem uma ampla gama de equipamentos, como automóveis, drones, máquinas industriais, robôs, e dispositivos médicos. Em determinados cenários, os UEs podem ser dispositivos estacionários, como roteadores que utilizam conectividade celular para fornecer acesso à internet em áreas remotas.

Conforme ilustrado na Figura 9, a rede móvel celular é composta por dois subsistemas principais: a Rede de Acesso por Rádio (RAN, do inglês *Radio Access Network*) e o Núcleo da Rede Móvel (MCN, do inglês *Mobile Core Network*) (Peterson; Sunay; Davie, 2023). A RAN conecta os UEs ao MCN, sendo responsável pela gerência dos recursos de rádio (espectro). Ainda, deve garantir sua utilização eficiente e a conformidade com os critérios de QoS para cada usuário, bem como pelo roteamento de dados e voz entre os dispositivos e a rede central. Essa rede inclui uma matriz descentralizada de estações rádio base, denominadas *eNodeB* ou eNB (*evolved Node B*) no 4G. No 5G, essas estações são chamadas gNB, em que o “g” se refere à próxima Geração (do inglês, *Generation*).

As funcionalidades do MCN podem ser categorizadas em diversos domínios, abrangendo desde o estabelecimento da conexão inicial até a entrega de serviços complexos e a gestão da mobilidade. Dentre estas funcionalidades destaca-se a autenticação e autorização de UEs na rede, o gerenciamento de mobilidade, o roteamento e encaminhamento de dados, aplicação de políticas e QoS, suporte a serviços de valor agregado e gerenciamento de configuração. No 4G o Núcleo Móvel é identificado como EPC, e no 5G como 5GCN (Dahlman, 2023). Ambos os componentes apresentam uma estrutura em camadas, composta por diversos blocos funcionais que asseguram a eficiência e a escalabilidade da rede. No 5G a arquitetura do MCN passou por uma grande evolução com a introdução do 5GCN, como a separação do plano de controle e do plano de usuário, virtualização das funções de rede, arquitetura orientada a serviços, dentre outras.

2.2.1 Rede de Acesso por Rádio

A RAN é um componente essencial das redes móveis celulares, responsável por fornecer a interface entre os UEs e o núcleo da rede. Composta por, dentre outros itens, BS e antenas, a RAN gerencia os recursos de rádio e viabiliza a comunicação entre os dispositivos dos usuários e a rede central. As BS implantadas são responsáveis por processar os sinais provenientes dos UEs, extrair os dados transmi-

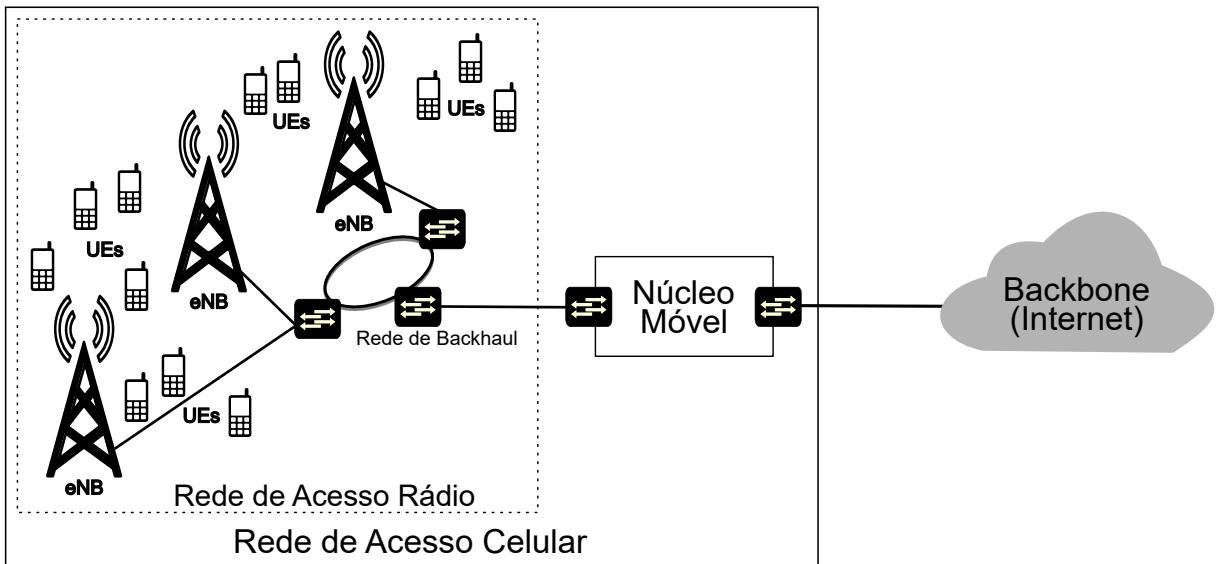


Figura 9 – Arquitetura de Rede Móvel Celular (Peterson; Sunay; Davie, 2023).

tidos e encaminhá-los ao MCN para posterior roteamento (Ullah *et al.*, 2023).

As antenas utilizadas na RAN podem ser setoriais ou omnidirecionais (com abertura de 360 graus), adaptando-se às características das células, cujo tamanho e forma variam de acordo com as necessidades da rede (Sanij; Casetti, 2016). O *beamforming* e as técnicas de modulação multiportadora são aplicados para otimizar a comunicação, reduzindo a interferência e permitindo que múltiplos UEs compartilhem eficientemente os recursos de rádio por meio da reutilização espacial e da multiplexação espectral (Nwogu, 2021; Semov *et al.*, 2020). Dessa forma, a RAN desempenha um papel crucial na otimização de Indicadores-Chave de Desempenho (KPI, do inglês *Key Performance Indicators*), como cobertura, capacidade, eficiência espectral e qualidade da experiência (QoE, do inglês *Quality of Experience*) dos usuários, impactando diretamente o desempenho geral da rede (Agiwal; Roy; Saxena, 2016).

A RAN é composta por BS, controladores de rede e enlaces de *backhaul*. As BS incluem unidades em processamento em banda base, antenas e transceptores, que trabalham em conjunto para estabelecer e gerenciar os enlaces de comunicação com os UEs. Os controladores de rede de rádio coordenam as operações da RAN e se conectam às BS por meio de enlaces de *backhaul* tipicamente cabeados, podendo ter ou não a topologia em anel como mostrado na Figura 9. Além de tecnologia cabeadas, normalmente a fibra óptica, outras tecnologias podem ser utilizadas para o enlace de *backhaul*, como micro-ondas ou satélite. A rede *backhaul* é uma parte importante da RAN, entretanto sua implementação não está descrita nos padrões 3GPP (Agiwal; Roy; Saxena, 2016).

O Gerenciamento de Recursos de Rádio (RRM, do inglês *Radio Resource Management*) é responsável por garantir a alocação eficiente dos recursos de rádio, otimizando o desempenho da rede e assegurando a manutenção de uma QoS mí-

nima. A RAN oferece funcionalidades básicas como alocação e estimativa de canais, formação de feixe, recepção e transmissão multiantena, diversidade de transmissão temporal e execução de *handovers* suaves e rígidos (Abu-Ali *et al.*, 2014).

A arquitetura da RAN foi projetada para ser adaptável a novas tecnologias e padrões, possibilitando sua evolução contínua. Funcionalidades avançadas, como gerenciamento de mobilidade, controle de potência em malha aberta e algoritmos de otimização de QoS, são implementadas para atender à crescente demanda por eficiência e qualidade de comunicação (Agiwal; Roy; Saxena, 2016).

A necessidade crescente por conectividade móvel tem levado à implantação de redes densas e heterogêneas, caracterizadas pela coexistência de células macro, femto e pico. Contudo, essa complexidade traz desafios relacionados à aquisição de locais, disponibilidade de infraestrutura e conformidade com requisitos regulatórios e ambientais, que devem ser abordados para garantir o sucesso da implementação e operação dessas redes.

2.2.2 Núcleo da Rede Móvel

O MCN integra funcionalidades essenciais, como autenticação, gerenciamento de mobilidade e suporte a QoS. No LTE, o núcleo é conhecido como EPC, enquanto no 5G é denominado 5GCN. Este último introduz suporte avançado para virtualização e computação em nuvem, otimizando a operação e reduzindo a latência.

A interoperabilidade entre LTE e 5G é viabilizada pelo suporte a redes Non-Standalone (NSA, do inglês *Non-Standalone*), onde o EPC gerencia o plano de controle, e Standalone (SA, do inglês *Standalone*), com uma implementação independente baseada no 5GCN.

2.2.2.1 Rede Non-Standalone

Diferentemente das gerações anteriores, as redes 5G permitem a integração de elementos de diferentes gerações em configurações diversas. A rede NSA utiliza conectividade dupla (*dual connectivity*) para integrar as redes de acesso 5G NR e LTE, permitindo que os dispositivos móveis se conectem a ambas as tecnologias (3GPP, 2022).

Na rede NSA, o 5G depende do núcleo das redes 4G para atividades de plano de controle, como alocação de recursos, gerenciamento de sessão e *handover* (3GPP, 2022). Essa configuração facilita a transição para o 5G *Standalone*, combinando recursos de rádio 4G e 5G NR com o núcleo EPC existente (Gures *et al.*, 2020).

A Figura 10 ilustra a comunicação na rede NSA, onde o plano de usuário da

gNB se comunica diretamente com o EPC, enquanto o plano de controle se comunica com o núcleo da rede através da eNodeB.

Usuários com dispositivos compatíveis podem utilizar o 5G NSA com velocidades superiores ao 4G, experimentando maior *throughput* devido ao eMBB (Gures *et al.*, 2020).

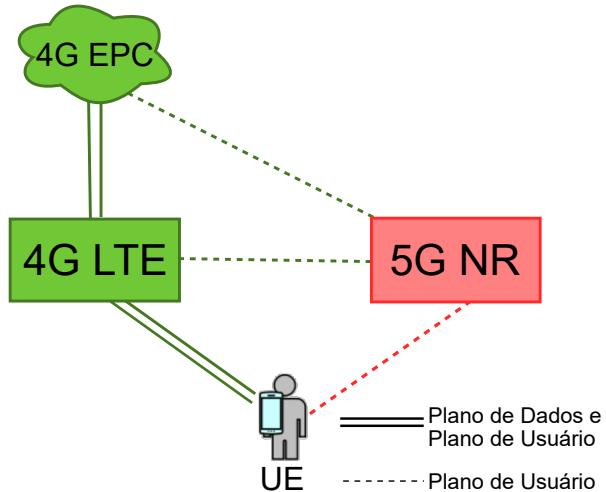


Figura 10 – Arquitetura de Rede 5G NSA (Kumar; AR, 2020).

2.2.2.2 Rede *Standalone*

A rede *Standalone* é uma rede 5G independente, com uma nova interface aérea 5G NR e um novo núcleo de rede 5GCN (3GPP, 2022). A arquitetura 5G SA oferece uma experiência 5G completa, com interoperabilidade com a rede 4G para garantir a continuidade do serviço (Gures *et al.*, 2020).

Na arquitetura 5G SA, a gNB se comunica diretamente com o 5GCN para o plano de controle e usuário. A Figura 11 ilustra a arquitetura de rede 5G SA.

O 5GCN, projetado para virtualização e computação em nuvem, permite que as operadoras ofereçam maior *throughput*, serviços de comunicação massiva de IoT, *Machine-to-Machine* (M2M) e comunicações de baixa latência, como jogos em nuvem e aplicações críticas em saúde. A estrutura do 5GCN permite a integração de elementos de diferentes gerações e configurações, oferecendo flexibilidade para as operadoras escolherem o *roadmap* mais adequado para cada região de operação (Dangi *et al.*, 2021).

2.3 Tipos de Implantação de Estações Base

As estratégias de implantação de BS na RAN evoluíram desde cenários homogêneos, onde há um único tipo e poucas BSs, para cenários heterogêneos e, pos-

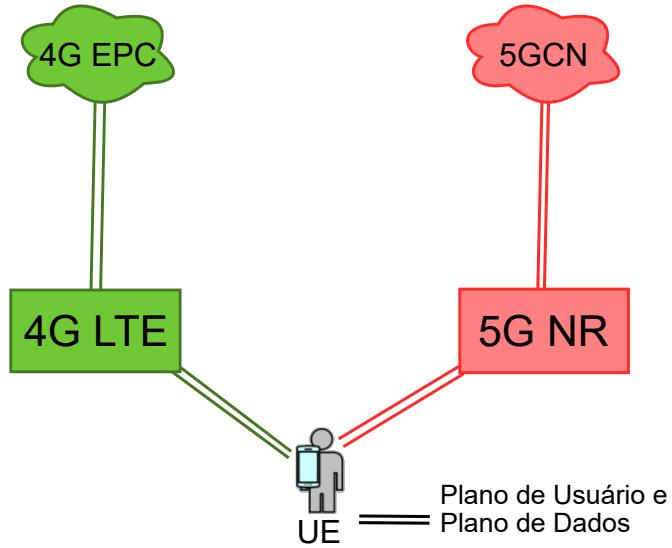


Figura 11 – Arquitetura de Rede 5G SA (Kumar; AR, 2020).

teriormente, para UDN, com um grande número de BSs de diferentes tipos (Mbulwa *et al.*, 2024). A seguir, são detalhados os principais tipos de implantação.

2.3.1 Redes Homogêneas

As redes homogêneas utilizam apenas um tipo de BS, geralmente BS Macro, que apresenta características uniformes, como potência de transmissão, área de cobertura e padrão de radiação das antenas. Essas BSs são comumente dispostas em uma grade hexagonal uniforme, possuem alta potência de transmissão (5W a 40W) e grande alcance (até 35 km), sendo necessárias poucas unidades para cobrir áreas urbanas ou rurais. Contudo, possuem baixa eficiência espectral por unidade de área, resultando em baixas taxas de dados, adequando-se apenas a chamadas de voz e serviços de baixa demanda, como nas redes 2G (Yuan, 2015). Com a chegada das redes 3G, que trouxeram maior demanda por dados e velocidades superiores, além de maior cobertura interna, as limitações desse modelo tornaram necessária a transição para redes heterogêneas (Kapoor; Grace; Clarke, 2019). A Figura 12 ilustra uma rede homogênea composta exclusivamente por BS macro.

2.3.2 Redes Heterogêneas

A crescente demanda por tráfego de dados em redes móveis, impulsionada pela proliferação de dispositivos conectados à Internet, exige novas soluções para aumentar a capacidade e a eficiência espectral do sistema (Xu *et al.*, 2021). As HetNets, introduzidas a partir da *Release 9*, surgem como uma alternativa promissora para lidar com esse desafio (Ali, M. *et al.*, 2017). Ao combinar nós de baixa potência (LPN,

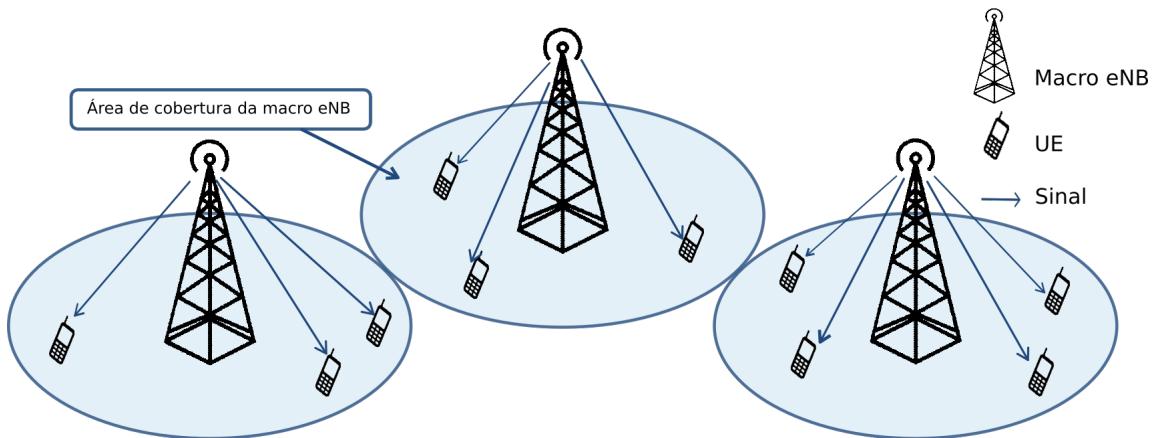


Figura 12 – Formato Tradicional de Rede.

do inglês *Low Power Nodes*) com a infraestrutura de macrocélulas, como ilustrado na Figura 13, as HetNets permitem otimizar o uso dos recursos e melhorar a qualidade do serviço.

Os LPN, operando com potência reduzida, oferecem vantagens significativas, como o aumento da eficiência espectral por unidade de área, a correção de falhas de cobertura e a ampliação da capacidade em áreas de alto tráfego (Agiwal; Roy; Saxena, 2016). Além disso, contribuem para a eficiência energética do sistema. As HetNets empregam diferentes tipos de estações base, cada qual com características específicas de potência, cobertura, tamanho e conectividade. Os principais tipos são (Yuan, 2015; Ali, M. et al., 2017):

- Cabeça de Rádio Remota (RRH, do inglês *Remote Radio Head*): estendem a cobertura de uma BS macro para áreas remotas. Conectadas via fibra óptica à BS macro, oferecem densificação centralizada de rede;
- BS Micro: com potência de até 5W, são implantadas de forma planejada para fornecer cobertura em áreas de alta demanda. Podem ser equipadas com antenas omnidirecionais ou direcionais;
- BS Pico: implantadas interna ou externamente, com potência variando entre 250mW a 2W, com alcance de até 200 metros, geralmente equipadas com antenas omnidirecionais;
- BS Femto: possuem alcance inferior a 50 metros e potência de transmissão menor que 200mW. Podem ser conectadas à rede através de uma conexão banda larga fixa.

2.3.3 Redes Heterogêneas Ultra Densas

A demanda crescente por tráfego e os rigorosos requisitos de QoS tornou a ultra-densificação uma abordagem central para redes 4G/5G. Este conceito visa

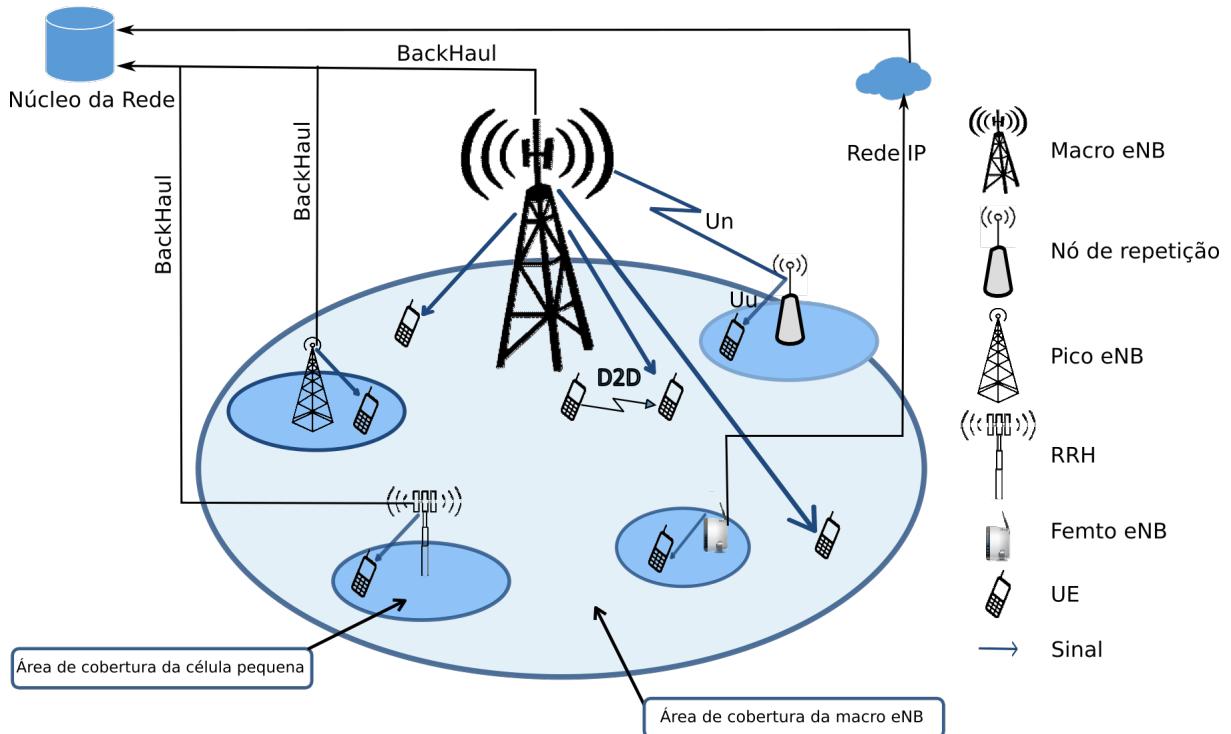


Figura 13 – Rede Heterogênea com diferentes tipos de nós, adaptado de (Sanchez; Gerardino, 2014) e (Ali, M. et al., 2017).

atender às necessidades de capacidade das redes modernas, caracterizadas por novos casos de uso que exigem maior eficiência espectral e flexibilidade. Em UDN, a implantação massiva de pequenas estações base possibilita maior reutilização de frequência, resultando em aumento significativo da capacidade da rede (Iqbal *et al.*, 2022).

As UDNs apresentam características específicas que as diferenciam de outros cenários de implantação. A densidade elevada e o uso de múltiplos níveis de interconexão entre BSs podem, por um lado, melhorar a eficiência espectral, mas, por outro, aumentam significativamente a complexidade da rede. Adicionalmente, a redução na área de cobertura de cada célula implica em um aumento de *handovers*, exigindo mecanismos eficientes para garantir uma conexão contínua e estável ao UE.

Apesar das vantagens, como o aprimoramento de capacidade,平衡amento de carga e maior velocidade de conexão, o uso intensivo de BSs pequenas também apresenta desafios. O consumo energético cresce proporcionalmente à densidade da rede, especialmente em cenários onde BSs permanecem ativas sem atender a usuários. Esse problema, se não tratado de maneira eficiente, pode comprometer os benefícios esperados das UDNs (Alamu *et al.*, 2020).

Portanto, a ultra-densificação emerge como uma solução promissora, mas exige um gerenciamento criterioso de recursos, consumo energético e interferências, especialmente em redes 4G/5G.

2.4 Handover

O procedimento de *handover* em redes móveis celulares apresenta desafios significativos, uma vez que envolve diversas tarefas que podem resultar em interrupções nos serviços e degradação do desempenho das aplicações (Ullah *et al.*, 2023). O aumento na frequência de migrações e transições intensifica esse desafio, elevando a ocorrência de *handover*.

O conceito de *handover* evoluiu, deixando de se restringir à continuidade de chamadas telefônicas para abranger a continuidade de sessões de *streaming*, a manutenção da QoS e o acesso à Internet. Nesse contexto, a pesquisa em sistemas celulares busca aprimorar o sistema de admissão de chamadas, visando controlar e reduzir a probabilidade de bloqueio e melhorar a qualidade do serviço (Alam *et al.*, 2024).

Essa expansão do conceito de *handover* é impulsionada pela popularização de dispositivos como *tablets* e *smartphones*, que permitem a experiência coletiva de usuários compartilhando a mesma área de cobertura. Atualmente, cenários de mobilidade com usuários que se deslocam em diferentes velocidades e utilizam aplicativos diversos são cada vez mais comuns.

A decisão de realizar o *handover* entre duas BSs é tomada pela própria BS de origem, sem intervenção do MME (Tayyab; Gelabert; Jantti, 2019; Khan *et al.*, 2022). Essa decisão baseia-se nos relatórios de medição enviados pelo UE. A BS de destino, por sua vez, prepara os recursos de rádio antes de aceitar o *handover*.

No LTE/LTE-A, após a conclusão do *handover*, a BS de destino instrui a BS de origem a liberar os recursos e envia uma mensagem de comutação de caminho para o MME. O MME, então, transmite um plano de atualização para o S-GW, garantindo que os pacotes do UE sejam direcionados pela nova rota. Para o 5G NR, os procedimentos equivalentes são detalhados e ilustrados na sequência.

As mensagens de controle são trocadas através da interface X2 entre as BSs, e os dados de pacotes do *downlink* também são encaminhados pela mesma interface (Tayyab; Gelabert; Jantti, 2019; Khan *et al.*, 2022; Ullah *et al.*, 2023).

O processo de *handover* em redes móveis celulares consiste em diversas etapas. Inicialmente, o *handover* é acionado quando o UE atinge um limite predefinido, com base em critérios como qualidade do sinal e força do sinal recebido da célula atual (Alam *et al.*, 2024). Caso a decisão seja positiva, o dispositivo móvel é comutado para a célula vizinha, mantendo a continuidade da sessão de comunicação.

Antes da execução do *handover*, o dispositivo móvel e a rede coletam informações relevantes, como medições do sinal das células vizinhas e informações sobre a carga de tráfego (Khan *et al.*, 2022). Com base nesses dados, a rede decide se o *handover* é necessário e para qual célula o dispositivo deve ser transferido.

Após o *handover*, o UE e a rede monitoram o desempenho da nova conexão. Caso a qualidade do serviço seja não seja satisfatória, a rede pode iniciar um novo *handover* ou tomar medidas para melhorar a conexão.

Durante o deslocamento na rede, o UE pode encontrar diferentes condições de propagação e interferência. Para garantir transições entre células, o UE realiza medições periódicas nos canais de rádio *downlink*, incluindo medições da potência de sinal de referência recebida (RSRP, do inglês *Received Signal Reference Power*) da célula atual (SeNB, do inglês *Serving eNodeB*) e das células vizinhas (NeNBs, do inglês *Neighboring eNodeB*) (Costa Silva, 2018). Para o 5G NR, os procedimentos equivalentes são detalhados e ilustrados na sequência.

O procedimento de *handover* em redes móveis celulares é iniciado quando a seguinte condição é satisfeita por um período específico:

$$RSRP_{NeNB} > RSRP_{SeNB+\delta_{HM}}, \quad (2.1)$$

onde $RSRP_{NeNB}$ indica o RSRP da célula vizinha, $RSRP_{SeNB}$ indica o RSRP da célula atual, e HM é a margem de histerese.

Na formulação do *handover*, a margem de histerese atua como uma margem de segurança para minimizar as chances de falha no *handover* (Ahmad *et al.*, 2017). Essa margem é configurável e pode variar de 0 a 10 dB, com intervalos de 0,5 dB.

Outro parâmetro crucial no processo de *handover* é o tempo para ativação (TTT, do inglês *Time-to-Trigger*), que define o período durante o qual a condição deve ser mantida antes de iniciar o *handover* (Alraih *et al.*, 2023). O processo de *handover* em redes 5G é ilustrado na Figura 14 e envolve as etapas:

1. O UE recebe sinais tanto da BS de origem (s-gNB) quanto da BS alvo (T-gNB) no *downlink*, onde a s-gNB é a célula servidora e a T-gNB é a célula vizinha.
2. O UE mede a intensidade desses sinais e gera um relatório de medição (MR, do inglês *Measurement Report*).
3. O UE envia o MR para a s-gNB, e se a intensidade do sinal da T-gNB for superior à da s-gNB por um determinado limiar durante um período predefinido, um evento de *handover* é iniciado.
4. Com base no MR, a s-gNB envia uma requisição de *handover* (HO-req) para a T-gNB para iniciar o processo de HO.
5. A T-gNB responde com um comando de confirmação da requisição de *handover* (HO request Ack-command) para indicar a disponibilidade de recursos para a execução do *handover*.
6. Após receber o comando de confirmação (Ack-command), a s-gNB in-

terrompe a transmissão de dados no *downlink* e envia o comando de *handover* (HO-cmd) para o UE, iniciando o processo de execução do *handover* para a T-gNB.

7. Na fase de execução, o UE desconecta-se da s-gNB e estabelece uma conexão com a T-gNB para sincronizar.
8. Finalmente, na fase de conclusão do *handover*, a T-gNB envia um comando de HO completo para a s-gNB para liberar os recursos, e o caminho do *downlink* é alterado da s-gNB para a T-gNB.

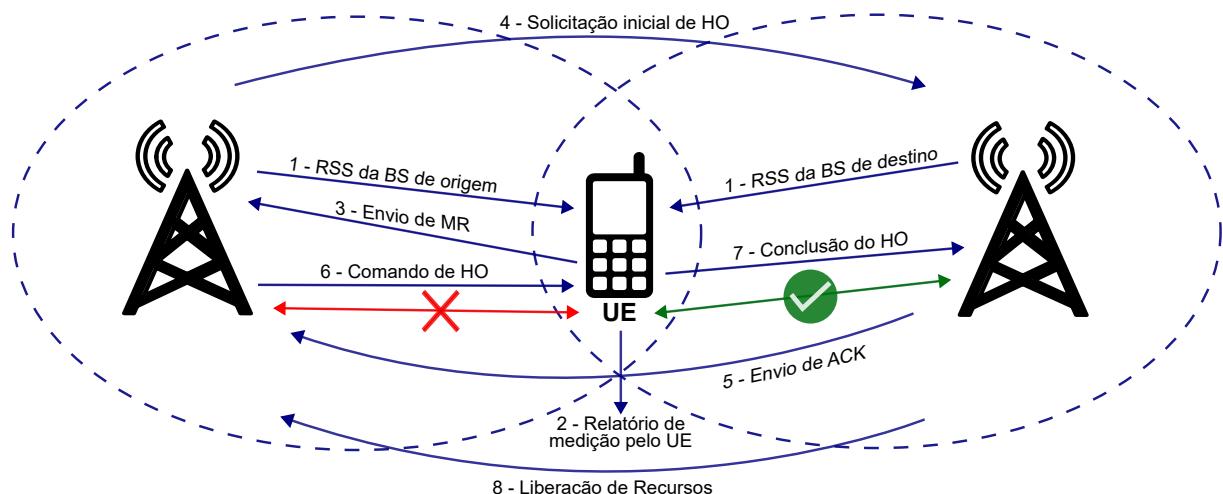


Figura 14 – Processo de *Handover* em redes 5G, adaptado de (Ullah *et al.*, 2023).

2.5 Alocação de Recursos e Escalonamento em Redes Móveis Celulares

A crescente demanda por serviços com requisitos rigorosos de latência, confiabilidade e eficiência espectral tem motivado a evolução dos algoritmos de alocação de recursos e escalonamento em redes móveis celulares, particularmente em cenários densamente povoados e heterogêneos.

Na literatura, diversas abordagens têm sido propostas para enfrentar os desafios impostos pela alocação eficiente dos RB e pela priorização de usuários ativos no sistema. Algoritmos tradicionais de escalonamento, como PF, MT e *Round Robin* (RR), apresentam diferentes compromissos entre eficiência espectral e justiça (Guo, 2004; Siddiqui *et al.*, 2023). Avanços recentes têm incorporado métodos baseados em aprendizado por reforço (RL, do inglês *Reinforcement Learning*) e aprendizado profundo (*Deep RL*), promovendo decisões mais adaptativas frente às condições dinâmicas da rede e à mobilidade dos usuários (Yuan, 2015).

No que se refere à alocação de recursos, técnicas baseadas em teoria dos jogos, otimização matemática (convexa e combinatória), além de métodos heurísticos e metaheurísticos — como algoritmos genéticos e *Particle Swarm Optimization* — têm

demonstrado viabilidade e desempenho satisfatório em contextos complexos (Guo, 2004; Abu-Ali *et al.*, 2014).

Essas estratégias ganham ainda mais relevância no contexto das redes 5G, onde casos de uso com características diferentes (eMBB, URLLC, mMTC), exigindo soluções que atendam simultaneamente a critérios rigorosos de latência, confiabilidade e eficiência energética (Yuan, 2015; Kapoor; Grace; Clarke, 2019).

A importância do tema também se revela na sua aplicação direta ao planejamento de mobilidade, controle de *handover*, gerenciamento de interferência e fatiamento de rede (*network slicing*), sendo, portanto, transversal a várias funções da arquitetura de redes móveis modernas (Kapoor; Grace; Clarke, 2019; Siddiqui *et al.*, 2023).

Dado o seu impacto direto na performance e na experiência do usuário, o tema de alocação de recursos e escalonamento é tratado no Capítulo 3 deste trabalho, onde são discutidas as principais abordagens, modelos e algoritmos aplicáveis às redes celulares densificadas e heterogêneas.

2.6 Considerações Finais

Este capítulo apresentou a evolução das redes móveis celulares, do LTE ao 5G NR, com ênfase nos avanços em arquitetura, eficiência espectral, escalabilidade e suporte a novos serviços. Foram discutidas as inovações introduzidas em diferentes *releases* do 3GPP, bem como os desafios decorrentes da heterogeneidade e densificação das redes. Também foram abordados o procedimento de *handover*, essencial em cenários de alta mobilidade, e os fundamentos da alocação de recursos e do escalonamento. Esses conceitos constituem a base teórica para o capítulo seguinte, que aprofunda a análise dos mecanismos de escalonamento e alocação de recursos no *uplink* de redes LTE-A/5G.

3 ESCALONAMENTO E ALOCAÇÃO DE RECURSOS *UPLINK* EM REDES LTE-A/5G

O escalonamento de pacotes constitui uma funcionalidade crítica em redes móveis celulares, sendo responsável por organizar a transmissão dos dados dos UEs de modo a otimizar o uso dos recursos de rede, atender aos requisitos de QoS, garantir equidade entre UEs e maximizar a eficiência espectral do sistema.

Nas redes LTE e LTE-Advanced, o escalonamento no enlace *uplink* apresenta desafios específicos em comparação ao *downlink*, notadamente pela restrição de alocação contígua de blocos de recursos imposta pelo uso do Acesso Múltiplo por Divisão de Frequência de Portadora Única (SC-FDMA, do inglês *Single Carrier Frequency Division Multiple Access*), e pela ausência de visibilidade total do estado dos *buffers* dos UEs por parte da eNodeB (Lei; Li, 2009). O processo inicia-se com o envio de uma solicitação de escalonamento (SR, do inglês *Scheduling Request*) de 1 bit pelo terminal móvel através do canal de controle físico *uplink* (PUCCH, do inglês *Physical Uplink Control Channel*), indicando a necessidade de transmissão. Essa solicitação pode ser complementada por um relatório de status do *buffer* (BSR, do inglês *Buffer Status Report*), informando a quantidade aproximada de dados a serem transmitidos. Com base nessas informações, além das métricas de QoS, qualidade do canal (SINR), potência disponível e restrições físicas, a BS decide sobre a alocação dos recursos e emite a concessão de escalonamento (SG, do inglês *Scheduling Grant*) com número determinado de RBs (Lei; Li, 2009; Abu-Ali *et al.*, 2014).

Em 5G NR, o processo de escalonamento *uplink* torna-se ainda mais complexo e flexível. A introdução de numerologias escaláveis, múltiplas bandas de operação, modos *grant-free*, alocação semi-persistente e serviços com requisitos extremos, como URLLC e mMTC, exige da BS mecanismos de decisão mais inteligentes e adaptativos (Gautam; Tamma, 2024). O 5G NR suporta a criação de múltiplos fluxos de QoS por *bearer*, gerenciamento energético com eDRX, multi-conectividade e técnicas como a agregação de portadoras. Tais avanços permitem uma alocação mais granular de recursos e menor latência, essencial para aplicações críticas.

O escalonamento de pacotes pode ser classificado de diversas maneiras, sendo que a seguir é apresentado em duas categorias (Mamane *et al.*, 2022):

- a) Escalonamento dependente do canal (*channel-dependent*): também conhecido como escalonamento dinâmico, utiliza o Indicador de Qualidade do Canal (CQI, do inglês *Channel Quality Indicator*), estimado com base nos Sinais de Referência de Sondagem (SRS, do inglês *Sounder Reference Signals*), para decidir o MCS e o tamanho do bloco de transporte (TBS, do inglês *Transport Block Size*). Esses sinais são tipicamente transmitidos no último símbolo SC-FDMA de cada subquadro e

fornecem dados sobre a qualidade do canal *uplink*, inclusive para controle de potência;

- b) Escalonamento independente do canal: ignora os valores de CQI e baseia suas decisões em parâmetros adicionais, como o BSR, o tipo de tráfego, a existência Solicitação de Retransmissões Automáticas Híbrida (HARQ, do inglês *Hybrid Automatic Repeat reQuest*), e os requisitos de QoS da *bearer*. Essa abordagem é útil em contextos nos quais a qualidade do canal é menos determinante ou está sujeita a rápidas variações.

No contexto do 5G, destaca-se a crescente adoção de técnicas de aprendizado de máquina para prever padrões de tráfego, ajustar políticas de escalonamento em tempo real e lidar com a complexidade resultante da densificação da rede e da diversidade de serviços. Além disso, a flexibilidade do 5G NR permite ajustar o intervalo de escalonamento com granularidade inferior a 1 ms, ampliando a capacidade de resposta da rede às solicitações de transmissão emergentes.

A Figura 15 apresenta um fluxograma esquemático do processo de escalonamento *uplink* para as redes LTE/LTE-A e 5G NR, contemplando as etapas de solicitação, avaliação, concessão e eventual retransmissão de pacotes.

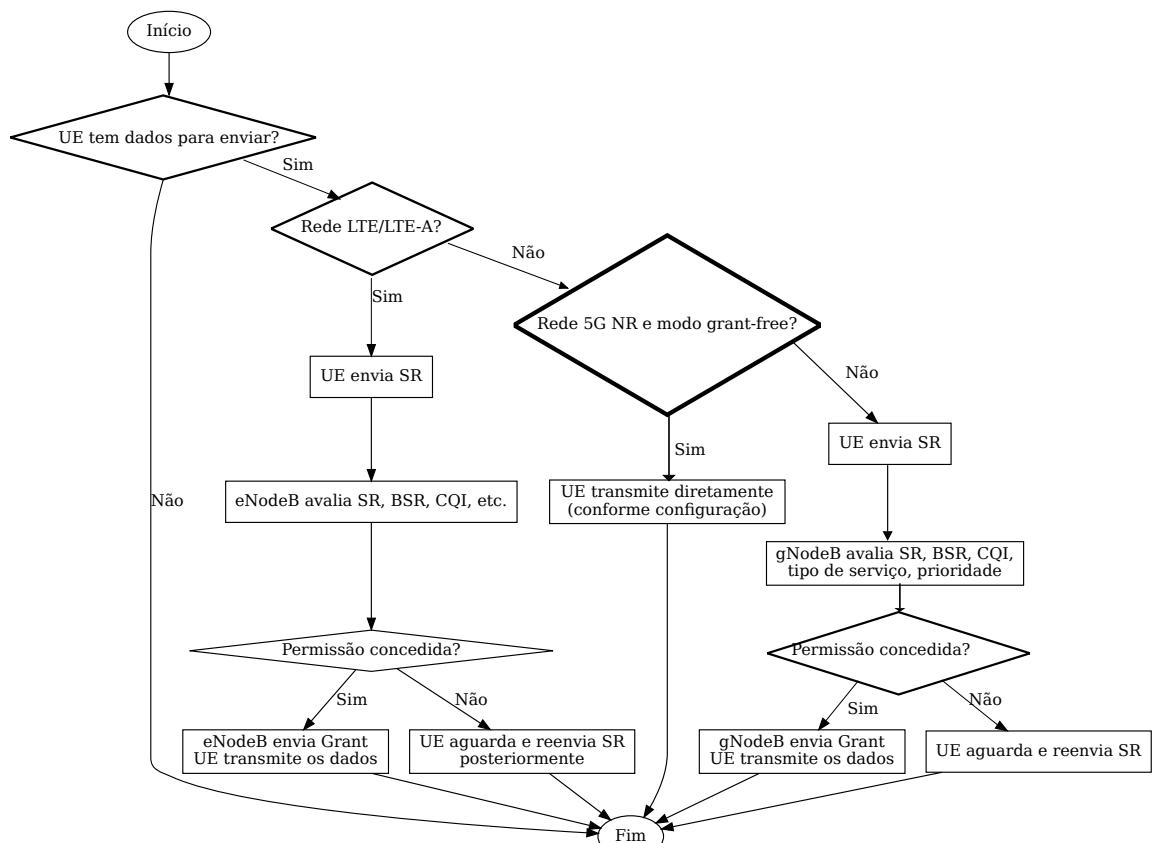


Figura 15 – Fluxograma esquemático do processo de escalonamento *uplink*.

3.1 Desafios na Implementação de Algoritmos de Escalonamento Uplink em Sistemas LTE-A/5G

A implementação prática de algoritmos de escalonamento no *uplink* em sistemas móveis LTE-A/5G enfrenta diversos desafios significativos, relacionados especialmente à complexidade computacional e aos requisitos de hardware envolvidos.

Um dos principais desafios na aplicação de algoritmos de escalonamento *uplink* reside na necessidade de realizar decisões de alocação em intervalos de tempo extremamente reduzidos, denominados Intervalo de Tempo de Transmissão (TTI, do inglês *Transmission Time Interval*), tipicamente na ordem de 1 ms ou menos (Gautam; Tamma, 2024). Neste curto espaço de tempo, o algoritmo precisa processar informações detalhadas e atualizadas sobre o estado do canal de múltiplos usuários, exigindo capacidade computacional considerável para garantir eficiência e justiça na distribuição dos recursos.

Além disso, algoritmos sofisticados, como PF, MT ou algoritmos conscientes de QoS, demandam cálculos complexos que aumentam substancialmente o custo computacional (Gatti *et al.*, 2022). A execução desses algoritmos em tempo real torna necessário o uso de hardware de alto desempenho e técnicas avançadas de processamento paralelo ou distribuído, o que pode representar uma limitação significativa, especialmente em estações rádio base de pequeno porte e com restrições energéticas, como é comum em pequenas células e femtocells (Jon *et al.*, 2024).

Outro desafio crítico é a escalabilidade desses algoritmos em redes densas. À medida que o número de usuários simultâneos aumenta, a complexidade do escalonamento cresce exponencialmente, impactando diretamente a latência das decisões tomadas e a eficiência da rede como um todo (Mamane *et al.*, 2022). Este problema pode ser especialmente crítico em ambientes de alta densidade de usuários, onde centenas ou até milhares de dispositivos podem requisitar recursos simultaneamente.

Além disso, há uma sobrecarga significativa na sinalização necessária para coletar as informações de estado do canal (CSI, do inglês *Channel State Information*) e os relatórios periódicos enviados pelos UEs, necessários para o funcionamento eficiente dos algoritmos de escalonamento. Esta sinalização excessiva pode consumir uma porção substancial dos recursos espectrais disponíveis, reduzindo a eficiência espectral global da rede (Gautam; Tamma, 2024).

Finalmente, o gerenciamento eficiente dos recursos computacionais e energéticos disponíveis no hardware das estações rádio base impõe limitações adicionais. A necessidade de hardware especializado e técnicas de otimização energética é crítica para garantir não apenas o desempenho dos algoritmos de escalonamento, mas também a sustentabilidade econômica e ambiental da infraestrutura de rede móvel LTE-A e 5G (Jon *et al.*, 2024).

Portanto, os desafios relacionados à complexidade computacional, escalabilidade e requisitos de hardware exigem atenção cuidadosa no desenvolvimento e implantação prática de algoritmos de escalonamento *uplink* eficientes em sistemas LTE-A e 5G.

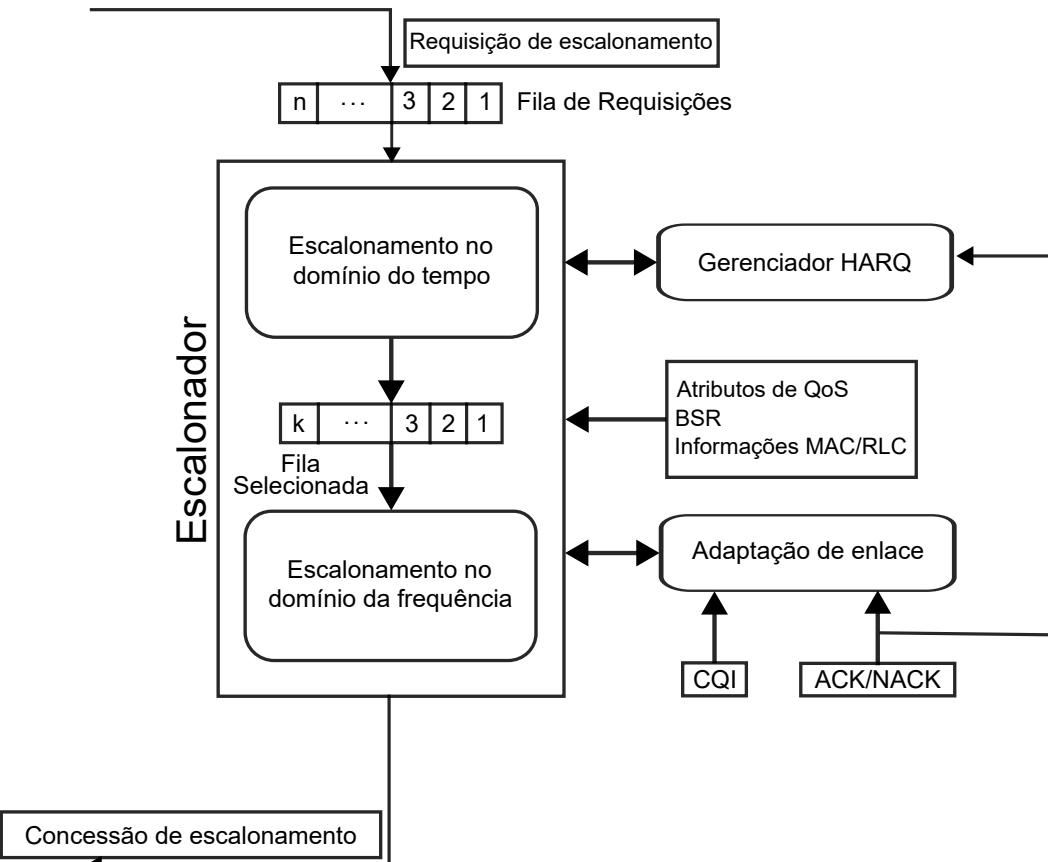


Figura 16 – Funcionalidade e Estrutura do Escalonador de Pacotes (Singh *et al.*, 2022).

3.2 Escalonador de Pacotes

O Escalonador de Pacotes (PS, do inglês *Packet Scheduler*) é uma entidade da camada MAC situada na BS, responsável pela atribuição dos RBs no 4G, e dos blocos de transporte (TB, do inglês *Transport Block*) no 5G, aos UEs com base em um algoritmo específico. O padrão LTE do 3GPP não define este algoritmo, permitindo diferentes abordagens propostas pelos fornecedores. Como resultado, diversas soluções podem ser aplicadas ao problema de escalonamento de recursos. A funcionalidade e a estrutura do escalonador de pacotes são mostradas na Figura 16, onde Solicitações de Escalonamento referem-se às requisições de transmissão enviadas pelos UEs. Essas solicitações são organizadas em uma fila de requisições de escalonamento, que pode ser priorizada com base em critérios como QoS, urgência ou métricas de canal. O escalonamento no domínio do tempo (TDPS, do inglês *Time Domain Packet Scheduling*) seleciona os UEs com base em parâmetros como atraso

e prioridade, enquanto o escalonamento no domínio da frequência (FDPS, do inglês *Frequency Domain Packet Scheduling*) aloca os recursos físicos de acordo com as condições do canal e a eficiência espectral. O gerenciador de retransmissões, baseado no protocolo *Hybrid ARQ* (HARQ), assegura a confiabilidade das transmissões. Os UEs pré-selecionados pelo TDPS são armazenados na fila de selecionados para processamento no domínio da frequência. Por fim, a concessão de recursos é enviada aos UEs, finalizando o ciclo de escalonamento. No entanto, o desenvolvimento de algoritmos de escalonamento no *uplink* geralmente considera alguns fatores essenciais, entre os quais destacam-se:

- Solicitações de escalonamento (SR, do inglês *Scheduling Requests*): usado para distinguir usuários ativos com dados em seus *buffers* de usuários inativos. Cada usuário com dados pendentes na sua memória intermediária envia um SR de um bit para a BS, informando da sua necessidade de uma concessão de *uplink* ser escalonada;
- BSR: cada UE envia à BS informações sobre o estado de seu *buffer*. Embora o relatório forneça apenas um valor aproximado e não a quantidade exata de dados no *buffer*, ele permite à BS inferir o tempo de espera dos pacotes armazenados;
- Parâmetros de QoS da *Bearer*: informações como GBR, atraso e taxa de perda de pacotes são utilizadas pelo escalonador para determinar a melhor forma de alocar os recursos;
- Retransmissões HARQ: é necessário que o escalonador identifique os RB reservados para retransmissões e aqueles disponíveis para novas alocações;
- SRS: introduzido no *uplink* LTE como um sinal de referência de banda larga. Tipicamente, ele é transmitido no último símbolo SC-FDMA de um subquadro de 1 ms e é usado para fornecer informações sobre a qualidade do canal *uplink*. A transmissão de dados do usuário não é permitida neste bloco, o que resulta em uma redução de até 7% na capacidade de *uplink* (Lei; Li, 2009);
- Relatórios de Margem de Potência (PHR, do inglês *Power Headroom Reports*): a potência limitada dos UEs, devido à restrição da vida útil da bateria, influencia diretamente o processo de escalonamento;
- Restrição de Contiguidade: o esquema SC-FDMA, obrigatório apenas no LTE, mas relevante em alguns cenários 5G, exige que os RBs alocados a um UE sejam contíguos. No entanto, a partir da *Release* 10, tornou-se possível alocar os RBs em até dois *clusters* para um mesmo usuário (Abu-Ali *et al.*, 2014; Gautam; Tamma, 2024).

Esses fatores são considerados fundamentais na construção de algoritmos

de escalonamento, permitindo o uso eficiente dos recursos de rede enquanto atendem às necessidades específicas dos UEs.

3.3 Alocação de Recursos

Um dos grandes desafios em redes sem fio é a comunicação multiusuário, onde múltiplos UEs, localizados em uma mesma região, demandam altas taxas de transmissão, baixo atraso e compartilham uma largura de banda finita. Para suprir tais necessidades utiliza-se o compartilhamento da largura de banda, destinando uma fração dos recursos totais a cada usuário, acomodando assim diferentes requisitos de QoS e taxas de dados.

A transmissão de dados se baseia no compartilhamento de um canal físico, estruturado em uma malha tempo-frequência, cujos recursos são dinamicamente alocados aos usuários, como ilustrado na Figura 17, para LTE/LTE-A, sendo idêntico no 5G NR para espaçamento de subportadora de 15 KHz (Mouadh; Sofiane, 2020).

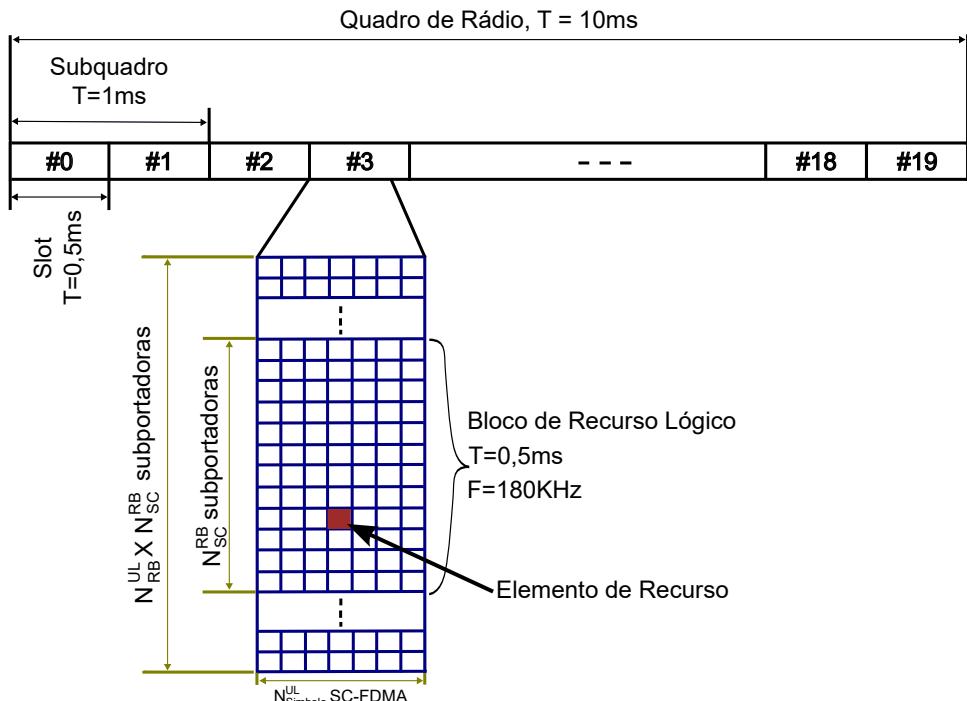


Figura 17 – Estrutura de *frame* LTE no domínio do tempo e frequência, adaptado de (Dahlman; Parkvall; Skold, 2016).

No domínio do tempo, os elementos de recurso são agrupados em *slots* de tempo de 0,5 ms. No domínio da frequência, os recursos são agrupados em RBs, cada um composto por 12 subportadoras de 15 KHz, totalizando 180 KHz. Os RBs são escalonados em pares, denominados Blocos de Escalonamento (SB, do inglês *Scheduling Block*), com duração de 1 ms (Oussakel, 2020).

A transmissão pode ser realizada de forma localizada ou distribuída. Na transmissão localizada, os RBs são agrupados em blocos contíguos na grade tempo-

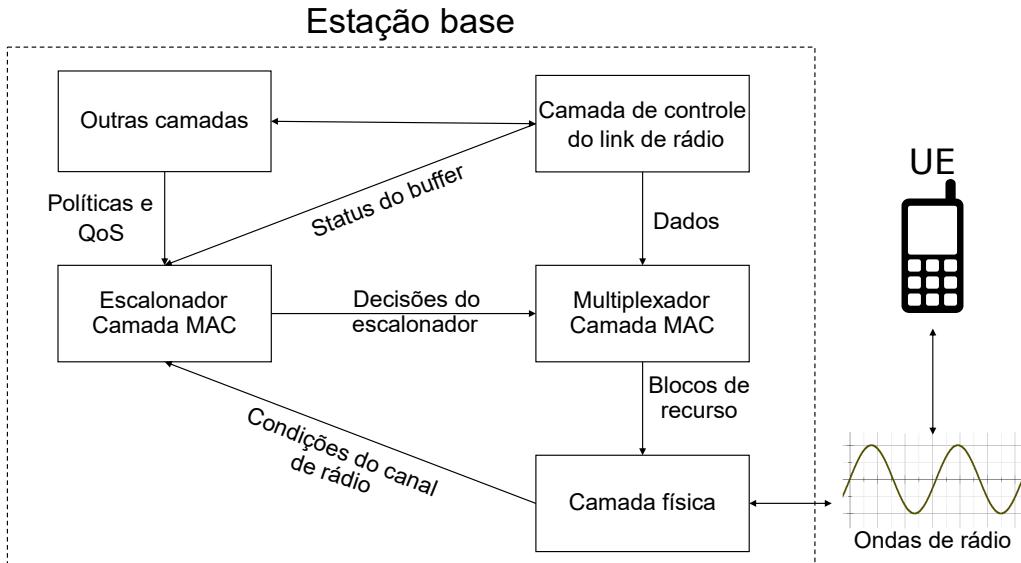


Figura 18 – Arquitetura do escalonador na camada MAC do LTE.

frequência. Este método requer *feedback* do UE sobre a qualidade das subportadoras, permitindo a BS selecionar os RBs mais adequados. Na transmissão distribuída, os símbolos são espalhados pela banda, e o *feedback* de canal é mais esparso ou até inexistente, sendo usada em cenários onde robustez à interferência e simplicidade de sinalização são desejadas (Yaacoub; Dawy, 2010).

A otimização do escalonamento e da alocação de recursos de rádio é crucial para melhorar o desempenho do sistema e aumentar a eficiência espectral da interface sem fio. O escalonador, componente da camada MAC, gerencia a alocação dessas unidades de recursos tanto no *downlink* quanto no *uplink* (Figura 18). O escalonador de pacotes na Camada MAC é responsável por coordenar a alocação de recursos com base nas condições do canal, nos requisitos de QoS e no status do *buffer*, considerando informações provenientes da Camada Física, da Camada de Controle do Link de Rádio (RLC, do inglês *Radio Link Control*) e de outras camadas superiores. Com base nas informações recebidas, o escalonador toma decisões dinâmicas e instrui o multiplexador MAC sobre a forma de combinar os dados para transmissão na Camada Física. Sua atuação é essencial para garantir o cumprimento das políticas de QoS, maximizar a taxa de transferência da rede, reduzir a latência e a perda de pacotes, contribuindo para o desempenho global da rede em ambientes de redes móveis celulares. No *downlink*, o escalonador determina dinamicamente quais terminais receberão dados e a quantidade de recursos alocada a cada um. A cada *subframe*, o escalonador aloca blocos de recursos para transmissão pelo Canal Compartilhado de *Downlink* (DL-SCH, do inglês *Downlink Shared Channel*) (Monikandan B; Sivasubramanian; Babu, 2017).

O escalonamento para *uplink* é realizado de forma independente do *downlink*. Nessa etapa, define-se quais UEs transmitirão nos Canais Compartilhados de

Uplink (UL-SCH, do inglês Uplink Shared Channel) e como a grade de recursos será distribuída entre eles. No *uplink*, a eNodeB (ou a gNodeB, no 5G NR) determina a carga útil, o formato de transporte e os recursos a serem alocados, com base em parâmetros como o BSR, CQI, PHR, parâmetros de QoS (como GBR e atraso máximo tolerável) e presença de retransmissões HARQ (Abu-Ali *et al.*, 2014; Singh *et al.*, 2022).

Em cenários com grandes larguras de banda e elevada seletividade em frequência, especialmente para usuários com baixas taxas de transmissão, a adaptação da transmissão às condições do canal é crucial para garantir a qualidade da comunicação (Mukhopadhyay; Das, 2021). A variabilidade do canal, no entanto, impede o uso de uma modulação fixa, o que motiva a adoção de técnicas de Codificação e Modulação Adaptativa (AMC, do inglês *Adaptive Modulation and Coding*) (Sanchez; Gerardino, 2014).

Os principais fatores que limitam a taxa de dados são a largura de banda e a relação sinal-ruído (SNR, do inglês *Signal to Noise Ratio*). A estratégia AMC consiste em transmitir a taxas mais altas quando o canal apresenta SINR alta, e a taxas menores quando a SINR é baixa, evitando perdas excessivas de pacotes (Nwogu, 2021). Taxas de transmissão baixas são obtidas com o uso de técnicas de modulação mais simples, como a Modulação por Deslocamento de Fase em Quadratura (QPSK, do inglês *Quadrature Phase Shift Keying*), combinadas com técnicas de codificação mais robustas. Altas taxas de transmissão, por outro lado, são alcançadas com técnicas de modulação mais complexas, como 64-QAM ou 256-QAM, e códigos de correção de erros menos robustos (Souza Lopes; Coelho Rocha; Teles Vieira, 2023).

O escalonamento *Channel-dependent* é amplamente utilizado no *downlink*. Para suportá-lo, o UE reporta periodicamente à BS a qualidade do canal através do CQI. O CQI, um indicador de 4 bits, informa a taxa máxima de transmissão que o dispositivo móvel pode alcançar com uma taxa de erro de bloco de até 10%. O CQI está diretamente relacionado à SINR, mas também é influenciado pela capacidade do hardware do receptor (Ferreira, 2021). A Tabela 2 apresenta os valores de CQI definidos pela especificação 3GPP TS38.214 para procedimentos de dados na camada física.

A eficiência espectral, medida em bps/Hz, representa a quantidade de bits de informação transmitidos por símbolo OFDM. A taxa de codificação, por sua vez, é definida pela Equação 3.1 (Wu, 2022):

$$\text{Taxa de codificação} = \frac{TBS + CRC}{RE \times m}, \quad (3.1)$$

onde *TBS* é o Tamanho do Bloco de Transporte (TBS, do inglês *Transport Block Size*) em bits, *CRC* é o número de bits de Verificação de Redundância Cíclica (CRC, do

Tabela 2 – Mapeamento CQI e esquemas de modulação e codificação, adaptado de (Nwogu, 2021)

Índice CQI	Modulação	Taxa de Codificação × 1024	Eficiência Espectral
0	-	-	-
1	QPSK	78	0,1523
2	QPSK	193	0,3770
3	QPSK	449	0,8770
4	16-QAM	378	1,4766
5	16-QAM	490	1,9141
6	16-QAM	616	2,4063
7	64-QAM	466	2,7305
8	64-QAM	567	3,3223
9	64-QAM	666	3,9023
10	64-QAM	772	4,5234
11	256-QAM	873	5,1152
12	256-QAM	711	5,5547
13	256-QAM	797	6,2266
14	256-QAM	885	6,9141
15	256-QAM	948	7,4063

inglês *Cyclic Redundancy Check*), RE é a quantidade de Elementos de Recurso aloca-
dos (RE, do inglês *Resource Element*) e m é a quantidade de bits por símbolo de
modulação.

Modulações de baixa ordem, como QPSK, são mais robustas a interferências, mas oferecem menor taxa de transmissão. Modulações de alta ordem, como 256-QAM, atingem taxas de transmissão mais elevadas, porém são mais suscetíveis a erros. A escolha da modulação e da taxa de codificação depende das condições do canal, buscando o equilíbrio entre taxa de transmissão e confiabilidade (Abu-Ali *et al.*, 2014).

A alocação de recursos que considera as variações do canal no tempo e na frequência oferece vantagens significativas. Em serviços sensíveis a atrasos, a alocação de recursos pode priorizar o tempo de entrega, independentemente da qualidade do canal (Haque *et al.*, 2023). Em outros casos, a exploração das variações na frequência pode otimizar o desempenho global do sistema (Myung; Lim; Goodman, 2006). O esquema de alocação do LTE, LTE Advanced e LTE Advanced Pro explora as variações no tempo e na frequência com blocos (*subframes*) de 1 ms e 180 KHz, respectivamente (Figura 17).

Apesar da estratégia de alocação de recursos não ser definida pela 3GPP, o principal objetivo é explorar a variabilidade do canal entre os dispositivos móveis, bus-

cando atender aos requisitos de eficiência espectral, justiça e QoS. Atender a esses requisitos simultaneamente é desafiador. Esquemas que visam maximizar a taxa de dados podem ser injustos, favorecendo usuários com melhores condições de canal, enquanto esquemas que priorizam a justiça podem comprometer a eficiência espectral (Sanchez; Gerardino, 2014).

Soluções ótimas podem ser obtidas alocando os subcanais com melhores condições a cada usuário. No entanto, a complexidade aumenta quando o melhor subcanal de um usuário é também o melhor para outros, especialmente se esses usuários não possuem alternativas de boa qualidade. Nas estratégias tradicionais, o desvanecimento do canal atua como um seletor aleatório, contribuindo para a diversidade multiusuário (Yaacoub; Dawy, 2010).

Uma BS serve múltiplos UEs através de múltiplos canais compartilhados, gerenciando a transmissão de dados em ambas as direções. Os pacotes das camadas superiores são classificados em fluxos de dados com diferentes requisitos de QoS (Abu-Ali *et al.*, 2014). Cada fluxo é associado a uma *bearer*, que representa uma conexão lógica entre o *gateway* da Rede de Pacotes de Dados (PDN, do inglês *Packet Data Network*) e o UE. A Figura 19 apresenta os fluxos de dados de serviço (SDF, do inglês *Service Data Flows*) e os EPS *bearers*, bem como a relação entre esses elementos. Em redes LTE/LTE-A, o tráfego gerado pelos usuários é categorizado em dois tipos principais: tráfego de SDF e tráfego de EPS *bearer*. Um SDF compreende um conjunto de fluxos IP associados a um determinado serviço utilizado pelo assinante, enquanto um EPS *bearer* corresponde à agregação de múltiplos SDFs que compartilham a mesma classe de QoS.

Do ponto de vista técnico, os SDFs representam fluxos de dados específicos relacionados a diferentes serviços consumidos pelo UE, como vídeo, voz ou navegação web. Por outro lado, os EPS *bearers* funcionam como túneis lógicos no plano de usuário, sendo responsáveis pelo transporte de múltiplos SDFs com requisitos semelhantes de QoS. Essa abordagem permite que a rede gerencie de forma mais eficiente diferentes tipos de tráfego, garantindo que cada fluxo seja tratado conforme suas necessidades específicas de desempenho (Mbale, 2021).

Após a classificação pela camada MAC, os fluxos são mapeados em *slots* da matriz tempo-frequência, segmentando os dados em blocos que se encaixam nos *slots* OFDMA. Cada usuário possui uma fila de pacotes a serem recebidos e a BS gerencia as demandas simultâneas de múltiplos UEs, alocando os *slots* disponíveis na matriz de recursos. Essa estratégia suporta uma ampla gama de taxas de dados, explorando o conhecimento das condições do canal e as demandas individuais de cada UE (Mouadh; Sofiane, 2020).

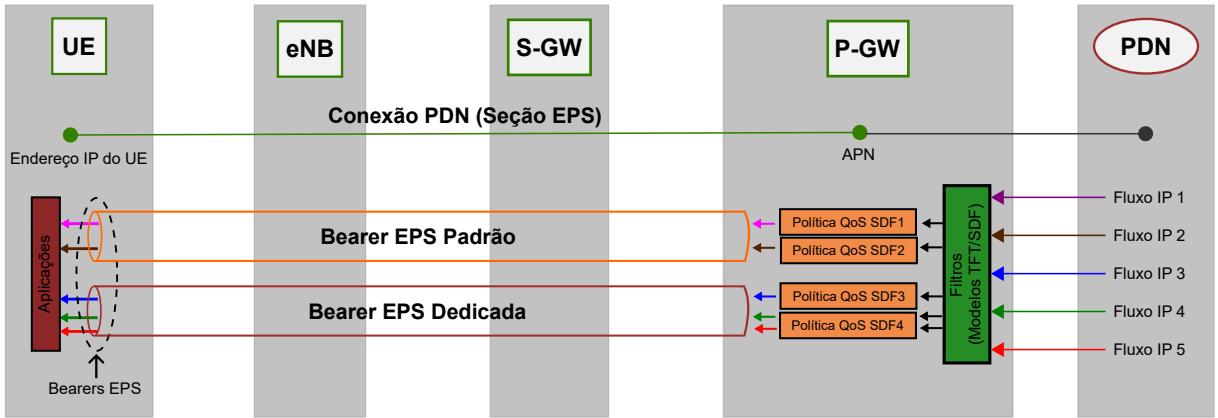


Figura 19 – SDFs e EPS bearers, adaptado de (Carvalho Ferreira, 2020).

3.4 Comparação entre SC-FDMA e OFDMA no *Uplink*

A técnica de acesso múltiplo utilizada no *uplink* das redes LTE e LTE-Advanced é o SC-FDMA, enquanto no *downlink* é utilizado o OFDMA. Ambos os esquemas são baseados na multiplexação ortogonal em frequência, mas apresentam características distintas que justificam suas aplicações específicas em cada enlace.

O SC-FDMA pode ser compreendido como uma variação do OFDMA, com a introdução de um estágio de pré-codificação via Transformada Discreta de Fourier (DFT) antes da alocação das subportadoras. Essa modulação DFT-spread reduz significativamente a Relação entre Potência de Pico e Média (PAPR, do inglês Peak-to-Average Power Ratio), tornando o SC-FDMA mais apropriado para o *uplink*, onde os UEs possuem restrições energéticas. Com uma menor PAPR, é possível utilizar amplificadores de potência mais eficientes, reduzindo o consumo de bateria e o aquecimento do equipamento terminal (Dahlman, 2023).

Por outro lado, o OFDMA proporciona uma maior flexibilidade na alocação de recursos, pois permite atribuição não contígua de subportadoras para diferentes UEs. Essa característica é vantajosa para aproveitar eficientemente as condições variáveis do canal em diferentes bandas de frequência. No entanto, essa flexibilidade é sacrificada no SC-FDMA, que restringe a alocação de RB a segmentos contíguos em frequência para manter o caráter de portadora única, condição necessária para reduzir a PAPR (Myung; Lim; Goodman, 2006).

Em termos de eficiência espectral, tanto SC-FDMA quanto OFDMA oferecem desempenho semelhante, com taxas de transmissão compatíveis, especialmente quando utilizados com esquemas de modulação adaptativa (AMC). Todavia, o SC-FDMA apresenta desempenho inferior sob condições de desvanecimento seletivo em frequência, devido à sua menor diversidade de subportadoras. Esse fator pode afetar negativamente a robustez do enlace *uplink* em cenários com mobilidade elevada ou canais altamente dispersivos (Sesia; Toufik; Baker, 2011).

Com a introdução do 5G NR, o *uplink* passou a permitir o uso de ambas as técnicas, sendo o DFT-s-OFDM (evolução do SC-FDMA) utilizado como padrão básico para o *uplink*, e o OFDMA reservado para situações onde alta taxa de pico ou baixa latência é desejada, como em aplicações eMBB ou URLLC. Essa coexistência visa equilibrar a eficiência energética com os requisitos crescentes de desempenho do 5G (3GPP, 2024a).

Dessa forma, a escolha entre SC-FDMA e OFDMA no *uplink* envolve um *trade-off* entre eficiência de potência e flexibilidade de alocação de recursos. A estratégia adotada pelas redes móveis contemporâneas busca otimizar esse equilíbrio, combinando o uso adaptativo das tecnologias conforme o perfil de tráfego, o tipo de dispositivo e as condições de canal. A Tabela 3 resume a comparação entre os dois esquemas.

Tabela 3 – Comparativo entre SC-FDMA (LTE) e OFDMA (5G/NR) no Uplink

Característica	SC-FDMA (LTE)	OFDMA (5G)
PAPR	Baixa	Alta
Alocação de RBs	Contígua	Contígua ou dispersa
Eficiência espectral	Moderada	Alta
Flexibilidade de escalonamento	Baixa	Alta
Adequado para URLLC	Limitado	Sim

3.5 Escalonador *Smoothed Round-Robin* (SRR)

O algoritmo SRR é uma variação aprimorada do RR, projetado para melhorar a equidade e suavizar a distribuição do tráfego. O RR clássico distribui as tarefas de forma uniforme, mas não considera pesos associados às entidades atendidas. O SRR resolve essa limitação introduzindo uma distribuição mais suave, baseada em pesos atribuídos a cada entidade (Wangdong *et al.*, 2003). O SRR utiliza duas estruturas principais para realizar a distribuição balanceada dos pacotes entre os fluxos: a Matriz de Pesos (WM, do inglês *Weight Matrix*) e a Sequência de Distribuição de Pesos (WSS, do inglês *Weight Spread Sequence*) (Guo, 2008). A Matriz de Pesos é responsável por determinar os pesos atribuídos a cada fluxo, enquanto a Sequência de Distribuição de Pesos define a ordem em que os pacotes são distribuídos. Essas estruturas trabalham em conjunto para garantir um equilíbrio na alocação de carga, evitando congestionamentos e garantindo melhor desempenho e eficiência na transmissão de dados. O SRR é uma técnica que proporciona gerenciamento do tráfego de forma a garantir uma distribuição equitativa e otimizada dos pacotes nos diferentes fluxos de dados.

O funcionamento do SRR é fundamentado em duas estruturas principais:

a *Weight Spread Sequence* (WSS) e a *Weight Matrix* (WM). A WSS consiste em uma sequência temporal construída com base nos pesos atribuídos a cada fluxo. Essa sequência determina a ordem em que os fluxos serão atendidos durante um ciclo de alocação, assegurando que os fluxos com maior peso sejam distribuídos uniformemente ao longo do tempo. Por exemplo, para três fluxos com pesos proporcionais a 3:2:1, a sequência WSS pode assumir a forma: $[f_1, f_2, f_1, f_3, f_1, f_2]$, evitando a concentração de alocações para um mesmo fluxo em janelas curtas.

A WM, por sua vez, é uma matriz bidimensional $WM[i][j]$ onde cada linha i representa a sequência de alocação do fluxo f_i e cada coluna j corresponde a uma unidade de tempo ou recurso (por exemplo, um RB no LTE *uplink*). Essa matriz é utilizada para guiar a decisão do escalonador em cada ciclo, verificando quais fluxos devem ser priorizados com base na suavização temporal.

Seja $\mathcal{F} = \{f_1, f_2, \dots, f_n\}$ o conjunto de fluxos ativos, com pesos associados $w_1, w_2, \dots, w_n \in \mathbb{N}$, e R o número total de recursos disponíveis (por exemplo, blocos de recurso por subframe). O número ideal de alocações para cada fluxo f_i por ciclo é dado por:

$$q_i = \left\lfloor \frac{w_i}{\sum_{j=1}^n w_j} \cdot R \right\rfloor \quad (3.2)$$

Durante o processo de escalonamento, o próximo fluxo a ser atendido é selecionado conforme o critério:

$$f_k = \arg \min_{f_i \in \mathcal{F}} \left(\frac{C_i(t)}{w_i} \right), \quad (3.3)$$

onde $C_i(t)$ representa o número acumulado de alocações feitas ao fluxo f_i até o tempo t . Esse critério garante proporcionalidade em relação ao peso e suavidade na distribuição temporal.

O SRR é aplicável às redes LTE-A e 5G, especialmente no *uplink*. A cada *subframe*, o escalonador percorre a WSS e consulta a matriz WM para decidir qual fluxo receberá os recursos, respeitando os pesos e a suavização temporal. Essa abordagem é útil em cenários com múltiplos serviços simultâneos, permitindo conciliar equidade e eficiência computacional.

3.6 Escalonador *Proportional Fair* (PF)

O escalonador PF é um algoritmo amplamente utilizado em redes móveis celulares modernas, como LTE/LTE-A e 5G NR, projetado para otimizar a alocação de recursos de rádio entre múltiplos UEs. Sua principal característica é a busca por um equilíbrio entre a eficiência espectral e a justiça na distribuição de recursos, garantindo

que todos os UEs recebam um nível mínimo de serviço, ao mesmo tempo em que se maximiza o desempenho total do sistema.

O PF opera com base em uma métrica de prioridade que considera tanto a taxa de transmissão instantânea quanto a taxa média histórica de cada UE. A taxa de transmissão instantânea é influenciada pelas condições dinâmicas do canal, como a distância a BS, interferências, mobilidade e outros fatores ambientais, e é determinada pela CQI e pela modulação utilizada. A taxa média histórica é calculada por meio de uma média móvel, permitindo que o escalonador se adapte às variações nas condições do canal ao longo do tempo. A métrica de seleção do PF é calculada como a razão entre a taxa de transmissão instantânea e a média móvel da taxa de dados, conforme expresso na Equação 3.4:

$$M_{(i)}(t) = \frac{r_i(t)}{R_i}, \quad (3.4)$$

onde $r_i(t)$ representa a taxa de dados instantânea do canal para o UE i , enquanto R_i representa a taxa média histórica do UE i .

Essa métrica garante que UEs com melhores condições de canal tenham prioridade na alocação de recursos, mas também evita que monopolizem a capacidade do sistema, garantindo oportunidades de transmissão para UEs em condições desfavoráveis e prevenindo a “inanição” de dispositivos em borda de célula (Dahlman; Parkvall; Skold, 2016).

3.7 Escalonador *New Courteous Algorithm* (NCA)

O crescente aumento da diversidade de serviços e dos requisitos de QoS nas redes móveis de próxima geração tem demandado soluções mais inteligentes e adaptativas para o escalonamento de pacotes, especialmente em cenários de alta carga e heterogeneidade de tráfego. Nesse contexto, o *New Courteous Algorithm* (NCA), proposto por (Tata; Fellag; Kadoch, 2020), surge como uma proposta avançada voltada à promoção de justiça no atendimento entre diferentes classes de tráfego, conciliando eficiência na alocação de recursos com garantia de desempenho.

O NCA foi projetado para operar em redes LTE/LTE-A e 5G, considerando três classes principais de tráfego: *handoff* (HO), tempo real (RT, do inglês *Real Time*) e não tempo real (NRT, do inglês *Non Real Time*). A classe HO, associada à mobilidade do usuário, possui a maior prioridade, sendo tratada por uma disciplina de enfileiramento por prioridade estrita (PQ, do inglês *Priority Queuing*) com o intuito de evitar a queda de chamadas durante o processo de mobilidade. A classe RT, que engloba aplicações sensíveis à latência, como voz e vídeo, recebe prioridade intermédia, enquanto a classe NRT, correspondente a serviços tolerantes a atraso, como

transferência de arquivos, possui menor prioridade no tratamento inicial.

A principal inovação do NCA consiste na introdução de um mecanismo de abordagem "cortês", onde os fluxos colaboram para um compartilhamento equilibrado dos recursos, o que permite às classes RT e NRT assumirem, temporariamente, maior prioridade durante um intervalo conhecido como *tolerance time*. Esse tempo é delimitado de forma a assegurar que os parâmetros de QoS da classe HO sejam respeitados, evitando degradação em sua latência ou aumento na taxa de perda de pacotes. Para isso, o algoritmo adota uma abordagem híbrida que combina a política de enfileiramento ponderado baseado em classes (CBWFQ, do inglês *Class-Based Weighted Fair Queuing*) com o monitoramento dinâmico das condições da rede, permitindo uma realocação adaptativa das prioridades.

O NCA estabelece uma hierarquia de prioridades entre as classes de tráfego HO, RT e NRT, com prioridade inicial atribuída segundo a ordem HO > RT > NRT, com pesos w_1 para o HO e w_2 para RT e NRT, tal que $w_1 > w_2$.

Com o objetivo de adaptar-se às condições dinâmicas da rede, o NCA realiza o monitoramento contínuo do tempo médio de espera nas filas das classes RT e NRT. Caso o tempo de espera da classe RT ultrapasse um limiar definido (TH_{RT}), é acionado um mecanismo de priorização temporária, no qual essa classe passa a ser atendida com prioridade máxima por meio da PQ. Durante esse período, as classes HO e NRT são redistribuídas sob a política CBWFQ, com novos pesos (w'_0 para HO e w'_2 para NRT), garantindo a continuidade do atendimento, porém com prioridade relativa inferior. Esse período de priorização é limitado por um tempo T_{r1} , de forma que a QoS da classe HO não seja comprometida.

De forma análoga, quando o tempo médio de espera da fila NRT atinge um limiar ajustável (TH_{NRT}), definido como apresentado na Equação 3.5, o algoritmo transfere temporariamente a prioridade máxima para essa classe. Nesse cenário, o tráfego NRT passa a ser processado via PQ por um intervalo limitado (T_{r4}), enquanto os fluxos HO e RT são atendidos por CBWFQ com pesos ajustados (w''_0 para HO e w''_1 para RT), mantendo a relação $w''_0 > w''_1$.

$$Th_{NRT} = \frac{\text{tempo médio de espera}}{x}, \quad (3.5)$$

onde x é um parâmetro de ajuste.

Encerrado o período de priorização temporária, o sistema retorna ao seu estado inicial, restabelecendo as prioridades e pesos originais. Dessa forma, o algoritmo assegura a estabilidade do sistema e promove uma melhor utilização dos recursos disponíveis, ao mesmo tempo em que respeita os requisitos de QoS específicos de cada classe de tráfego. Tal abordagem torna-se especialmente eficaz em cenários de congestionamento e variações abruptas nas condições da rede, contribuindo para

um desempenho mais robusto e adaptável.

Tal abordagem permite ao NCA melhorar significativamente o desempenho das classes de menor prioridade sem comprometer o atendimento às classes críticas, promovendo a redução de atrasos e perdas de pacotes e contribuindo para o aumento do vazão global do sistema. Sua formulação matemática robusta, aliada à flexibilidade para adaptação em cenários densamente carregados e heterogêneos, consolida o NCA como uma alternativa eficaz para o escalonamento de pacotes nas redes móveis de última geração.

3.8 Considerações Finais

Este capítulo apresentou os principais desafios do escalonamento e da alocação de recursos *uplink* em redes LTE-A e 5G, destacando as restrições do SC-FDMA, a limitação de informações de *buffer* e a necessidade de decisões rápidas e eficientes. Foram discutidas as categorias de escalonamento, os fatores críticos como CQI, BSR e SINR. Foram apresentados os algoritmos SRR, PF e NCA, evidenciando suas limitações em cenários heterogêneos e dinâmicos.

Diante dessas limitações, o próximo capítulo propõe um esquema de escalonamento e alocação de recursos baseado em DQL e no método multicritério WAS-PAS, visando otimizar a QoS e a eficiência em HetNets LTE-A/5G.

4 ESQUEMA DE ESCALONAMENTO E ALOCAÇÃO DE RECURSOS *UPLINK* EM HETNETS LTE-A/5G BASEADO EM DQL E WASPAS

Este capítulo apresenta uma proposta metodológica integrada para o escalonamento e a alocação eficiente de recursos no enlace *uplink* de redes heterogêneas (HetNets) LTE-A/5G. A solução é fundamentada na combinação de técnicas de aprendizado por reforço profundo (DQL, do inglês *Deep Q-Learning*) e métodos de decisão multicritério (WASPAS, do inglês *Weighted Aggregated Sum Product Assessment*), com o objetivo de otimizar a QoS, reduzir o consumo energético e lidar com os desafios impostos por cenários densamente povoados e altamente dinâmicos.

Inicialmente, destaca-se o escalonamento de objetivo duplo, baseado em uma versão adaptativa do algoritmo *Smoothed Round-Robin* (SRR), que diferencia os fluxos RT e NRT, atribuindo prioridade a subclasses críticas. Essa estratégia assegura a distribuição justa e eficiente dos blocos de recursos entre os usuários, respeitando requisitos específicos de latência e taxa de transferência.

Na etapa seguinte, a alocação de recursos é realizada com restrições energéticas, por meio do algoritmo DQL, que aprende continuamente políticas ótimas com base nas condições variáveis do canal, no consumo de energia dos dispositivos e nos parâmetros de QoS. A utilização do DQL possibilita a adaptação ao ambiente sem depender de modelos explícitos da rede, maximizando o *throughput* e mitigando interferências intercelulares.

Complementarmente, é implementado um mecanismo de *handover*, baseado no método WASPAS, que permite decisões mais precisas ao considerar múltiplos critérios simultaneamente, como SINR, carga da célula e mobilidade do usuário. Essa abordagem contribui para a continuidade do serviço e a redução de quedas de conexão em cenários de alta mobilidade.

4.1 Solução Proposta

A solução proposta visa otimizar a alocação de recursos em redes heterogêneas LTE-A/5G, considerando os requisitos de QoS de diferentes classes de tráfego. O algoritmo proposto é baseado em um esquema de escalonamento de objetivo duplo e um método de alocação de recursos com restrição de energia.

4.1.1 Escalonamento de Objetivo Duplo

O escalonamento de tráfego é realizado utilizando um algoritmo SRR que classifica o tráfego em duas classes: RT e NRT. O tráfego RT é subdividido em três subclasses: videoconferência, *streaming* de vídeo e VoIP, com prioridades decres-

centes. O algoritmo SRR garante que o tráfego de maior prioridade seja escalonado primeiro e que os recursos alocados sejam utilizados de forma eficiente.

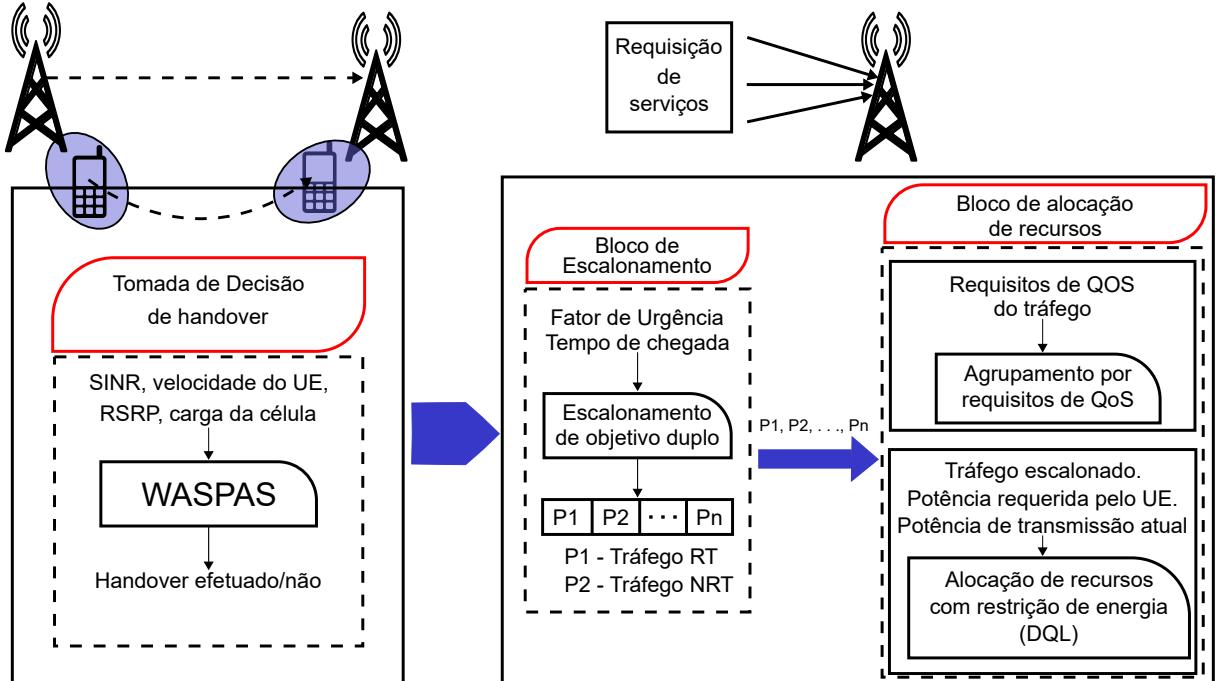


Figura 20 – Arquitetura da metodologia proposta.

O escalonamento de objetivo duplo utilizado na arquitetura da metodologia proposta é mostrado na Figura 20. Todos os UEs transmitem solicitações de alocação de recursos *uplink* para as respectivas BSs e o mecanismo de escalonamento de objetivo duplo clássico prioriza o tráfego da rede com base em fator de urgência e tempo de chegada. O tráfego na rede é classificado em RT e NRT. O fator de urgência garante a priorização do processamento de pacotes críticos, mitigando atrasos e perdas. Para pacotes com fator de urgência igual, o critério de desempate adotado é o tempo de chegada. A aplicação desses critérios no SRR permite garantir QoS para tráfego RT, enquanto preserva a eficiência no processamento de tráfego NRT. Para lidar com essa classificação de forma eficiente, optou-se pelo algoritmo SRR, considerando sua baixa complexidade computacional e adequação a cenários de alto tráfego.

Os pesos utilizados para o tráfego RT e NRT foram definidos como $w_{RT} = 0,7$ e $w_{NRT} = 0,3$, respectivamente. A taxa de serviço (ϵ_0) é calculada como:

$$\epsilon_0 = \iota_1 * w_{RT} + \iota_2 * w_{NRT}, \quad (4.1)$$

onde ι_1 e ι_2 são calculados conforme abaixo:

$$\iota_1 = \frac{\delta_r(\text{bits/s})}{P_l RT(\text{bits})}, \quad (4.2)$$

$$\iota_2 = \frac{\delta_r(\text{bits/s})}{P_l NRT(\text{bits})}, \quad (4.3)$$

onde δ_r indica a largura de banda disponível na rede, enquanto $P_l RT$ e $P_l NRT$ indicam a quantidade de dados pendentes para os tráfegos RT e NRT, respectivamente. Ao aproveitar melhor a largura de banda, o sistema minimiza as taxas de queda de chamadas e as probabilidades de bloqueio, buscando garantir um desempenho ideal em ambientes de rede heterogênea.

Os valores de w_{RT} e w_{NRT} representam a política de priorização da rede e são responsáveis por garantir o balanceamento adequado entre os requisitos de QoS dos diferentes tipos de tráfego. O tráfego RT recebe 70% da prioridade no processo de escalonamento, em razão de sua natureza crítica, que exige baixa latência, largura de banda rigorosa e taxas reduzidas de perda de pacotes (Monikandan B; Sivasubramanian; Babu, 2017). Por sua vez, para o tráfego NRT, mais tolerante a atrasos — como downloads ou *e-mails* — (Tata; Fellag; Kadoch, 2020), é atribuído 30% da prioridade. Essa proporção baseia-se em três fatores principais: (i) os requisitos de QoS, dado que os serviços RT possuem limites mais restritivos de latência e perda; (ii) a otimização de recursos, priorizando o tráfego RT para maximizar a eficiência global da rede, evitando desperdício com tráfego menos crítico; e (iii) flexibilidade dinâmica, que permite o ajuste automático dos parâmetros (ι_1 e ι_2) conforme as condições variáveis da rede, assegurando a manutenção da política de priorização sob diferentes cenários de carga. Um pseudocódigo que descreve o funcionamento do SRR é apresentado no Algoritmo 1.

4.1.2 Alocação de Recursos com Deep Q-Learning

A alocação eficiente de recursos em redes móveis celulares modernas, particularmente no enlace *uplink* de ambientes densos e heterogêneos como as redes LTE-A/5G, apresenta um problema inherentemente complexo. Desafios dinâmicos, como a variabilidade das condições de canal, a mobilidade dos usuários, a interferência intercelular, as restrições de energia dos dispositivos e a diversidade de requisitos de QoS para diferentes tipos de tráfego (RT e NRT), limitam a eficácia de mecanismos tradicionais de alocação baseados em regras fixas ou algoritmos estáticos. Nesse contexto, a adoção do aprendizado de máquina justifica-se pela sua capacidade de adaptação dinâmica e otimização contínua em ambientes complexos e não estacionários.

O aprendizado por reforço, ramo do aprendizado de máquina, investiga como agentes aprendem a tomar ações em ambientes dinâmicos, com o objetivo de maximizar a recompensa acumulada ao longo do tempo. Essa área é fundamental

Algoritmo 1: Algoritmo SRR com dupla priorização de atendimento considerando tipo de tráfego e requisitos de QoS

Entrada: Lista de UEs conectados
 Para cada UE:
 tipo_tráfego, requisitos_QoS, buffer_UE, tempo_espera
 JanelaSRR
 TTI
Saída: Lista de UEs ordenados para transmissão *uplink* no intervalo JanelaSRR

```

1 início
2   FilaRT ← lista circular de UEs com tráfego RT
   FilaNRT ← lista circular de UEs com tráfego NRT
   para cada UE em Lista_UEs
   faz
3     se UE.tipo_tráfego == "RT" então
4       INSERIR UE em FilaRT senão
5         INSERIR UE em FilaNRT
6     fim
7   fim
8
9   IndiceRT ← 0
10  IndiceNRT ← 0
11  Tempo_Janela ← 0
12  Lista_Alocados ← lista_vazia
13  enquanto Tempo_Janela < JanelaSRR faz
14    se FilaRT NÃO está vazia então
15      UE ← FilaRT[IndiceRT]
16      se UE.buffer_UE > 0 então
17        ADICIONAR UE à Lista_Alocados
18        Tempo_Janela ← Tempo_Janela + TTI
19      fim
20      IndiceRT ← (IndiceRT + 1) MOD tamanho(FilaRT)
21    fim
22    se FilaNRT NÃO está vazia E Tempo_Janela < JanelaSRR
        então
23      UE ← FilaNRT[IndiceNRT]
24      se UE.buffer_UE > 0 então
25        ADICIONAR UE à Lista_Alocados
26        Tempo_Janela ← Tempo_Janela + TTI
27      fim
28      IndiceNRT ← (IndiceNRT + 1) MOD tamanho(FilaNRT)
29  fim
```

para o desenvolvimento de sistemas autônomos e para a tomada de decisões inteligentes. Algoritmos de aprendizado por reforço baseiam-se em conceitos como políticas, recompensas e valores. Entre os algoritmos clássicos desse paradigma, o *Q*-

Learning destaca-se por modelar o problema como um Processo de Decisão de Markov (MDP, do inglês *Markov Decision Process*), representado pela tupla (S, A, R, P, γ) :

- S é o conjunto de estados;
- A o conjunto de ações;
- R a função de recompensa, que define a recompensa recebida ao tomar uma ação a em um estado s ;
- P é a função de transição, que define a probabilidade de transitar para um novo estado s' após tomar uma ação a no estado s ;
- γ é o fator de desconto, que determina a importância das recompensas futuras em relação às recompensas imediatas.

A função valor-ação $Q(s, a)$ estima o valor esperado de executar uma ação a em um estado s . A função Q é normalmente representada como uma tabela, onde cada entrada corresponde a um par estado-ação. No entanto, em ambientes complexos com grandes espaços de estados e ações, manter essa tabela se torna computacionalmente inviável. Para solucionar essa limitação, o *Deep Q-learning* combina o *Q-learning* tradicional com redes neurais profundas. As redes neurais aproximam a função $Q(s, a)$, permitindo a aplicação do algoritmo em problemas complexos. A Rede Q Profunda (DQN, do inglês *Deep Q-Network*) recebe o estado como entrada e retorna os valores de Q para todas as ações possíveis. A atualização dos parâmetros da rede, denotados por θ , segue uma versão adaptada da Equação de Bellman (Alavizadeh; Alavizadeh; Jang-Jaccard, 2022):

$$Q(s, a; \theta) \leftarrow Q(s, a; \theta) + \alpha \left(r + \gamma \max_{a'} Q(s', a'; \theta) - Q(s, a; \theta) \right) \quad (4.4)$$

onde $Q(s, a; \theta)$ representa o valor atual da função Q para o par estado-ação (s, a) , α é a taxa de aprendizado ($0 < \alpha \leq 1$), r é a recompensa recebida ao tomar a ação a , γ é o fator de desconto que pondera a importância das recompensas futuras, s' é o próximo estado após execução da ação a , $\max_{a'} Q(s', a'; \theta)$ é a melhor estimativa do valor futuro esperado.

Para lidar com cenários dinâmicos, como variações no número de UEs conectados, demandas de QoS e níveis de interferência intercelular, o DQL adapta continuamente suas políticas de alocação. Isso garante uma resposta mais eficiente às mudanças repentinas na rede, minimizando atrasos e reduzindo as taxas de perda de pacotes. Além disso, técnicas como *replay buffer* são empregadas para armazenar experiências passadas (transições (s, a, r, s')) e amostrar mini-lotes aleatórios, melhorando a generalização e estabilidade do modelo. Um pseudocódigo que descreve o funcionamento do DQL é apresentado no Algoritmo 2.

Aplicado à alocação de recursos em redes LTE-A/5G, o DQL otimiza a dis-

tribuição de RBs considerando QoS, consumo energético e *throughput*. Conforme detalhado no Algoritmo 2, o sistema inicializa a DQN com pesos aleatórios e utiliza uma política ϵ -greedy para selecionar ações (alocação de canal e potência). A recompensa R é calculada com base em métricas de QoS, eficiência energética e taxa de transferência, permitindo que o modelo aprenda políticas adaptativas mesmo sob condições não estacionárias.

Esta técnica se adapta às condições dinâmicas das redes heterogêneas LTE-A/5G conforme o tráfego da rede e as condições do canal se alteram. Isso permite que o sistema responda de forma mais eficiente a mudanças repentinas na rede, como variações no número de UEs conectados, suas demandas de QoS e níveis de interferência intercelular. Ao aprender com as interações observadas na rede, o DQL desenvolve políticas de alocação mais eficientes, melhorando a QoS, minimizando atrasos e reduzindo as taxas de perda de pacotes em cenários complexos e dinâmicos.

A adoção da técnica DQL para auxiliar na solução do problema de alocação de recursos justifica-se por sua capacidade de operar sob incerteza e aprender políticas ótimas diretamente a partir da interação com o ambiente, mesmo sem conhecimento explícito do modelo de transição de estados (Nomeir; Gadallah; Seddik, 2023). O DQL é uma extensão do Q-Learning tradicional, que utiliza redes neurais profundas como aproximadores de função de valor, sendo, portanto, capaz de lidar com espaços de estado e ação de alta dimensionalidade, como os observados em cenários com múltiplos usuários, múltiplos canais e múltiplas restrições de QoS. Embora outras técnicas baseadas em aprendizado por reforço, como *Deep Deterministic Policy Gradient*, *Proximal Policy Optimization*, *Asynchronous Advantage Actor-Critic* ou *Soft Actor-Critic*, também sejam aplicáveis à alocação de recursos, elas envolvem maior complexidade computacional, sensibilidade elevada a hiperparâmetros e necessidade de arquiteturas mais sofisticadas de aprendizado contínuo (Iqbal *et al.*, 2022). O DQL, por sua vez, oferece um bom equilíbrio entre desempenho, simplicidade de implementação e robustez, sendo particularmente adequado para aplicações em redes móveis celulares (Tang; Zhou; Kato, 2020). Dessa forma, a escolha do DQL neste trabalho dá-se pela sua capacidade de fornecer decisões de alocação de recursos eficientes, mesmo sob condições não estacionárias e imprevisíveis, representando uma alternativa superior às abordagens convencionais e aos algoritmos clássicos de aprendizado supervisionado e não supervisionado.

4.1.3 Tomada de Decisão de Handover

Para mitigar a interferência intercelular, um processo robusto de tomada de decisão de *handover* é implementado, utilizando o algoritmo Avaliação do Produto de

Algoritmo 2: Alocação de Recursos com Restrição de Energia e Garantia de QoS

Entrada: Para cada UE em movimento:
 Lista de UEs conectados, Parâmetros de QoS de cada fluxo, Níveis de consumo de energia dos UEs, Potência de transmissão solicitada, Condições do canal e Total de RBs disponíveis

Saída: Alocação otimizada de RBs por UE

```

1 início
2   Inicializar rede neural  $Q(S, A)$  com pesos aleatórios
3   Definir parâmetros do DQL: taxa de aprendizado  $\alpha$ , fator de desconto  $\gamma$ ,  $\epsilon$  para política  $\epsilon$ -greedy e buffer de experiência
4   para cada UE em movimento faz
5     Observar estado  $S$ 
6      $S \leftarrow$  (tipo de tráfego, consumo energético, potência solicitada)
7     Selecionar ação  $A$  (canal e potência) com política  $\epsilon$ -greedy:
8       Com probabilidade  $\epsilon \rightarrow$  selecionar  $A$  aleatoriamente
9       Com probabilidade  $(1 - \epsilon) \rightarrow$  selecionar  $A = argmaxQ(S, A)$ 
10    Executar  $A$ :
11      Alocar canal do cluster correspondente
12      Definir potência de transmissão
13      Alocar quantidade necessária de RBs
14    Observar nova recompensa
15     $R \leftarrow$  função(QoS, energia, throughput)
16  fim
17 fim

```

Soma Agregada e Ponderada (WASPAS, do inglês *Weighted Aggregated Sum Product Assessment*). Este método, proposto em 2012, combina o Modelo de Produto Ponderado (WPM, do inglês *Weighted Product Model*) e o Modelo de Soma Ponderada (WSM, do inglês *Weighted Sum Model*), sendo uma abordagem robusta para determinar a utilidade em Tomada de Decisão Multicritério (MCDM, do inglês *Multi-Criteria Decision Making*) e classificar as alternativas da melhor para a pior (Mardani *et al.*, 2017). Segundo E. K. Zavadskas, Turskis e Antucheviciene (2012), o método WASPAS oferece maior precisão em comparação ao uso isolado dos métodos WPM e WSM, sendo amplamente aplicado em diferentes contextos de tomada de decisão.

Como outras técnicas MCDM, o método WASPAS utiliza uma matriz de decisão como entrada. Essa matriz, $X = (x_{ij})_{m \times n}$, representa o desempenho de m alternativas em relação a n critérios de avaliação (Mardani *et al.*, 2017). Para assegurar uma análise consistente e imparcial, os critérios devem ser objetivos e mensuráveis. A estrutura da matriz de decisão está apresentada na Equação 4.5.

$$WM = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,n} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m,1} & w_{m,2} & \cdots & w_{m,n} \end{bmatrix} \quad (4.5)$$

A matriz de decisão contém valores associados a diferentes escalas e unidades, o que exige um processo de normalização para torná-los comparáveis. Para critérios de máximo benefício, utiliza-se a Equação 4.6, enquanto para critérios de mínimo benefício aplica-se a Equação 4.7. Esse processo de normalização é essencial para que todas as métricas sejam convertidas em uma forma adimensional, permitindo sua combinação e análise apropriada.

$$\bar{X}_{ij} = \frac{x_{ij}}{\max_i(X_{ij})} \quad (4.6)$$

$$\bar{X}_{ij} = \frac{\min_i(X_{ij})}{x_{ij}} \quad (4.7)$$

Após a normalização, as pontuações das alternativas são calculadas separadamente pelos métodos WSM e WPM. No método WSM, a importância relativa de cada alternativa (IR_i) é obtida pela soma ponderada dos valores normalizados, em que W_j representa o peso do j -ésimo critério, conforme descrito na Equação 4.8.

$$IR_i^1 = \sum_n^{j=1} \bar{X}_{ij} W_j \quad (4.8)$$

No método WPM, utiliza-se o produtório dos valores normalizados (x_{ij}) elevados aos seus respectivos pesos (w_j), como apresentado na Equação 4.9.

$$IR_i^2 = \prod_n^{j=1} (\bar{X}_{ij})^{W_j} \quad (4.9)$$

O método WASPAS combina as pontuações dos modelos WSM e WPM por meio de um parâmetro λ , que varia entre 0 e 1, ou seja, a soma dos modelos de soma ponderada e do produto ponderado é dada pela Equação 4.10:

$$IRT_i = \lambda \times IR_i^1 + (1 - \lambda) \times IR_i^2, \quad (4.10)$$

onde IRT_i é a importância total referente a alternativa i ; IR_i^1 é a importância relativa obtida pelo método WSM; e IR_i^2 é a importância relativa obtida pelo método WPM.

Essa combinação pode ser interpretada como uma média ponderada entre

os dois métodos, em que λ ajusta a contribuição relativa de cada um. Quando ($\lambda = 0$), o método se comporta como o WPM, enquanto, quando ($\lambda = 1$), equivale ao WSM. O comportamento de classificação do WASPAS depende fortemente dos pesos atribuídos aos critérios considerados e do valor da constante λ . Edmundas Kazimieras Zavadskas, Antucheviciene *et al.* (2013) sugeriram que a utilização de valores de λ entre 0,32 e 0,49 resulta em soluções mais confiáveis.

A classificação final das alternativas é realizada em ordem crescente dos valores de IRT_i , calculados pela Equação 4.10. Nesse contexto, a alternativa com maior valor de IRT_i é considerada a melhor, permitindo uma hierarquização clara e objetiva das opções disponíveis.

A metodologia WASPAS permite a análise simultânea de variáveis quantitativas e qualitativas, facilitando a ponderação de fatores fundamentais para decisões de *handover* mais confiáveis e adaptativas (Chakraborty; Zavadskas, E.; Antucheviciene, 2015). Para o *handover*, o WASPAS avalia simultaneamente múltiplos critérios, como SINR, RSRP, Qualidade de Sinal Recebido de Referência (RSRQ, do inglês *Reference Signal Received Quality*), carga da célula (categorizada como subcarregada, sobre-carregada ou normal, com base em métricas de carga) e velocidade do usuário — fundamentais para decisões de *handover* mais confiáveis e adaptativas (Chakraborty; Zavadskas, E.; Antucheviciene, 2015). O algoritmo WASPAS utiliza esses parâmetros para otimizar as decisões de *handover*, com o objetivo de minimizar a interferência intercelular e melhorar o desempenho geral da rede. O Algoritmo 3 detalha sua implementação, mostrando que quando a SINR ou RSRP está abaixo de um limiar, as células vizinhas são classificadas com base em seus escores WASPAS. O *handover* é executado apenas se o escore da célula alvo for maior que o da célula atual.

Em comparação com outras abordagens multicritério, como AHP (*Analytic Hierarchy Process*), TOPSIS (*Technique for Order Preference by Similarity to Ideal Solution*) ou ELECTRE (*ELimination Et Choix Traduisant la REalité - ELimination and Choice Expressing the REality*), o algoritmo WASPAS destaca-se por seu baixo custo computacional e facilidade de implementação em sistemas em tempo real, características essenciais para ambientes móveis onde as decisões precisam ser tomadas com alta frequência e baixa latência (Souza *et al.*, 2020). Ao permitir a combinação linear e multiplicativa dos critérios avaliados, o WASPAS também oferece uma forma mais estável de classificação das alternativas, reduzindo a sensibilidade a flutuações pontuais nos parâmetros de rede (Barbara *et al.*, 2023). Características que tornam o WASPAS especialmente adequado ao contexto das redes móveis celulares, reforçando sua escolha neste trabalho em detrimento de outras técnicas de decisão multicritério.

Algoritmo 3: Mecanismo de *handover* utilizando o algoritmo WAS-PAS

Entrada: Para cada UE em movimento:
 SINR atual,
 RSRP atual,
 RSRQ atual,
 Velocidade do usuário,
 Carga da célula atual,
 Lista de células vizinhas com respectivos parâmetros
Saída: Decisão de handover (executar ou manter conexão atual)

```

1 início
2   para cada UE em movimento faz
3     Calcular SINR, RSRP, RSRQ, velocidade e carga atual da célula
4     se SINR ou RSRP for baixo então
5       para cada célulaVizinha em listaCélulasVizinhas faz
6         Obter parâmetros da célulaVizinha: SINR_v, RSRP_v,
7           RSRQ_v, carga_v
8         Calcular escore_WASPAS[célulaVizinha] com base em:
9           SINR_v, RSRP_v, RSRQ_v, carga_v e velocidade do UE
10        fim
11        célulaDestino ← célulaVizinha com maior escore_WASPAS
12        se escore_WASPAS[célulaDestino] >
13          escore_WASPAS[célulaAtual] então
14            Executar handover para célulaDestino
15            Atualizar conexão do UE
16            Recalcular parâmetros de QoS
17            senão
18              Manter conexão na célula atual
19            fim
20        fim
21      fim
22    fim
23  fim
24
```

4.2 Estrutura do Sistema

O ambiente de rede considerado neste estudo, ilustrado na Figura 21, utiliza uma arquitetura HetNet composta por macrocélulas, microcélulas, picocélulas e femtocélulas, cada uma com sua eNB associada. Até 100 UEs são distribuídos aleatoriamente na área de cobertura. O sistema permite a classificação do tráfego em RT e NRT, com pesos de prioridade ajustáveis para otimizar a alocação de recursos de acordo com a urgência e o tipo de serviço.

A Tabela 4 apresenta os principais parâmetros de simulação, conforme os trabalhos de (Abubakar *et al.*, 2022; ITU-T, 2023; Yamagishi *et al.*, 2021). Em Tru-

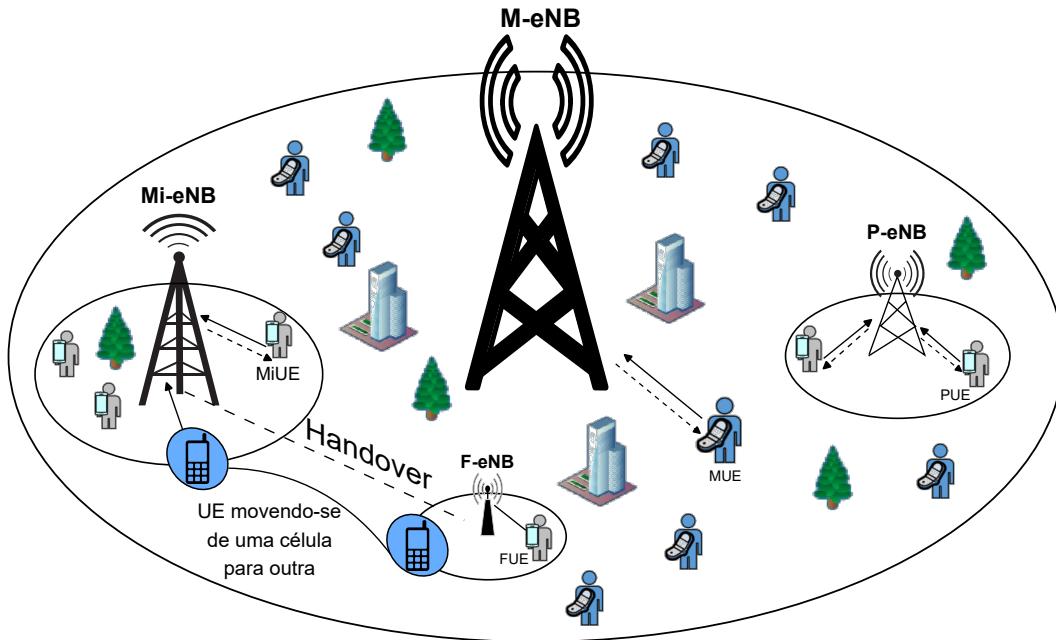


Figura 21 – Topologia de Rede, adaptado de (Abubakar *et al.*, 2022).

jillo (2015) tem-se os valores para as taxas de dados de VoIP, Video e Web utilizados nesse trabalho. Para as *smallcells* foram definidos raio de atuação como se segue: 50 metros para femtocélula, 100 metros para picocélula, 500 metros para a microcélula e 1000 metros para a macrocélula. O cenário considerado tem como foco exclusivo o enlace *uplink*, no qual a carga de tráfego varia de forma proporcional ao número de UEs conectados ao sistema. Para cada UE, é executada uma única instância de aplicação. As instâncias de aplicação foram distribuídas de forma aleatória entre os UEs, permitindo avaliar o desempenho do esquema de escalonamento e alocação de recursos proposto. O protocolo desenvolvido contempla mecanismos capazes de atender a diferentes requisitos de QoS, considerando a heterogeneidade das aplicações simuladas.

4.3 Simulações e Parâmetros de Desempenho

O simulador de eventos discretos *Network Simulator 3* (ns-3) foi selecionado para a realização dos experimentos deste trabalho devido à sua arquitetura modular e flexível, que oferece suporte a múltiplas Tecnologias de Acesso de Rádio (RAT, do inglês *Radio Access Technology*) e multi-banda, incluindo Wi-Fi, LTE (LTE-A, Acesso Assistido Licenciado e LTE-Não Licenciado), NR e NR Não Licenciado. Trata-se de um simulador de código aberto, escrito em C++, que garante modularidade, polimorfismo e flexibilidade (Nsnam, 2024b).

O projeto ns-3 visa construir um núcleo de simulação sólido, bem documentado, de fácil uso e depuração, que atenda às demandas de todo o fluxo de trabalho

Tabela 4 – Parâmetros de Simulação

Parâmetros de Simulação	Valores
Raio da Macrocelula	1000m
Duração da simulação	300s
Quantidade de UEs	[20,40,60,80,100]
Quantidade de M-BS	1
Quantidade de Mi-BS	1
Quantidade de P-BS	1
Quantidade de F-BS	1
Layout da Célula	Circular
UL Bandwidth	10MHz
Potência Máx. dos UEs	10 dBm
Ganho Antena dos UEs	0 dBi
Modelo de Canal	Modelo de Friss
Modelo de Mobilidade	<i>Random Direction</i>
TTI	1ms
Potência Macro eNB	40 dBm
Velocidade dos UEs	3 km/h
VoIP Bit Rate	8.4 Kbps
Video Bit Rate	242 Kbps
Web Bit Rate	52 Kbps
Quantidade de RBs	50

de simulação, desde a configuração inicial da simulação até a análise de rastreamento (Koutlia *et al.*, 2022). O núcleo do simulador, representado na Figura 22, oferece funcionalidades comuns a diversos tipos de redes, como gerenciamento de eventos, aritmética temporal e variáveis aleatórias. O módulo de rede (*network*) define a estrutura de pacotes, enquanto o módulo auxiliar (*helper*) simplifica a criação de *scripts* de simulação em C++ (ou em Python) (Nsnam, 2024b), permitindo a especificação de parâmetros como topologia, interface aérea e tempo de simulação.

As simulações consideraram a configuração de uma rede LTE-A, abrangendo cenários com macro, micro, pico e femto eNBs, largura de banda de 10 MHz e diferentes densidades de usuários. Embora o estudo tenha utilizado parâmetros de redes LTE-A, a metodologia e a abordagem propostas são compatíveis com ambientes 5G NSA da *Release 15* do 3GPP. Esta compatibilidade justifica-se pela adoção de práticas alinhadas aos requisitos de QoS, eficiência espectral e baixa latência, essenciais tanto para aplicações avançadas em LTE-A quanto para os serviços críticos introduzidos pelo 5G NR, como a gestão de mobilidade baseada no método WASPAS

e a mitigação de interferência intercelular.

Para a obtenção de métricas de desempenho como SINR, perda de pacotes, taxa de *handover*, atraso e latência foram utilizadas as funcionalidades integradas de rastreamento e registro do ns-3. Por meio da instrumentação sistemática do código de simulação, essas métricas foram capturadas durante a execução, possibilitando uma análise detalhada do desempenho da rede.

Com o objetivo de garantir significância estatística e confiabilidade nos resultados obtidos, foi adotada a abordagem de simulação de Monte Carlo. Ao conduzir 10 execuções de simulação com parâmetros variados, o estudo abordou as incertezas e variações inerentes aos cenários de comunicação sem fio, contribuindo para a robustez geral dos resultados da pesquisa.

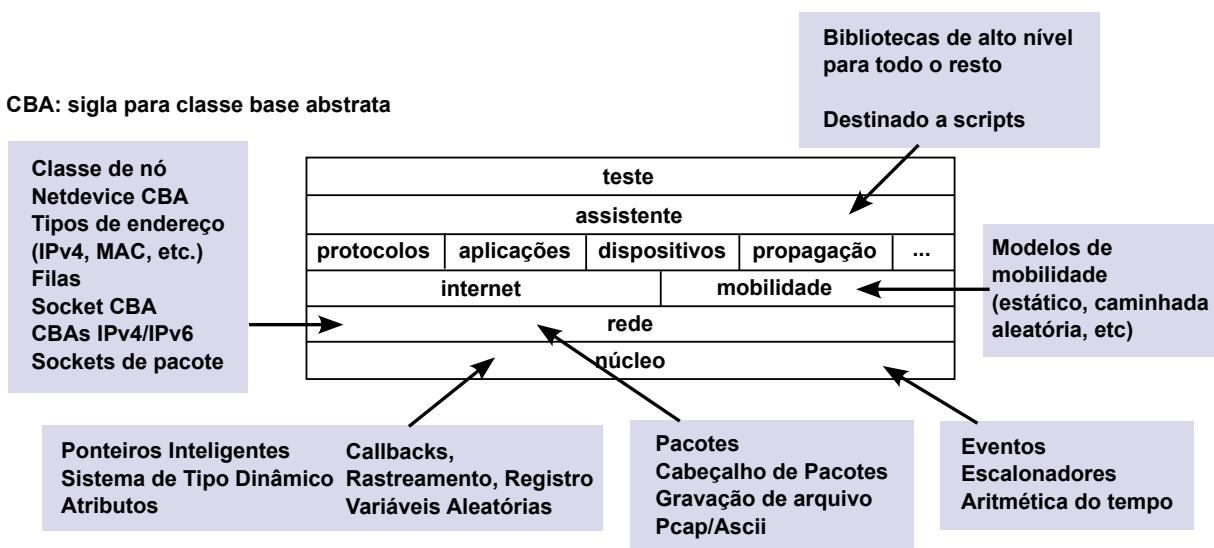


Figura 22 – Organização do software ns-3, adaptado de (Nsnam, 2024a).

4.4 Métricas de Avaliação

Para avaliar o desempenho do esquema de escalonamento de alocação de recursos proposto, foram utilizadas as seguintes métricas: Atraso, Taxa de transferência, Índice de justiça, Taxa de queda de chamadas e Taxa de perda de pacotes.

4.4.1 Atraso médio

O atraso médio, expresso pela soma dos atrasos de transmissão, propagação, enfileiramento e retransmissão (Equação 4.11), é uma métrica crucial para avaliar o impacto do escalonamento e da alocação de recursos na QoS.

$$D = T_d + P_d + Q_d + RT_d, \quad (4.11)$$

onde T_d é o atraso de transmissão, P_d é o atraso de propagação, Q_d é o atraso de enfileiramento, e RT_d é o atraso de retransmissão.

4.4.2 Vazão

A vazão (*throughput*), medida em *bits* por segundo, indica a eficiência do sistema na transmissão de dados, sendo um fator fundamental na experiência do usuário. Neste trabalho a taxa de transferência é calculada considerando a quantidade de bits transferidos com sucesso e o tempo necessário para transferência.

4.4.3 Índice de justiça

O índice de justiça mede a equidade da alocação de recursos entre as classes de tráfego, garantindo que diferentes tipos de aplicações tenham acesso proporcional aos recursos (Xu *et al.*, 2021). O mais usado é o Índice de Jain (Mamane *et al.*, 2022), que calcula a justiça conforme expresso na Equação 4.12:

$$I_j = \frac{(\sum x_i)^2}{n \times \sum x_i^2}, \quad (4.12)$$

onde x_i são os dados do i -ésimo UE e n é o número total de UEs.

4.4.4 Taxa de queda de chamadas

A taxa de queda de chamadas (CDR, do inglês, *Call Drop Rate*) mede a porcentagem de chamadas interrompidas devido à insuficiência de recursos ou problemas na conexão, conforme descrito na Equação 4.13:

$$CDR = \frac{N_{dropped}}{N_{dropped} + N_{success}}, \quad (4.13)$$

onde $N_{dropped}$ representa o número de chamadas interrompidas e $N_{success}$ representa o número de chamadas bem-sucedidas.

4.4.5 Taxa de perda de pacotes

Taxa de perda de pacotes (PLR, do inglês, *Packet Loss Rate*) quantifica a porcentagem de pacotes não entregues, impactando diretamente a QoS e a confiabilidade da rede (Equação 4.14).

$$PLR = \frac{N_{PS}}{N_{PR}} * 100\%, \quad (4.14)$$

onde N_{PS} representa o número de pacotes enviados e N_{PR} representa o número de pacotes recebidos.

4.4.6 Complexidade Assintótica

A complexidade assintótica, representada pela notação Big O, é uma métrica essencial para avaliar o desempenho de algoritmos em relação ao tempo de execução e ao consumo de memória conforme o tamanho da entrada aumenta indefinidamente. Essa notação descreve o comportamento da função de custo do algoritmo no pior caso, ignorando constantes e termos de menor ordem, o que possibilita a comparação padronizada entre diferentes soluções, independentemente de implementações específicas.

O esquema proposto adota duas etapas principais para o escalonamento e a alocação de recursos: o SRR, responsável pela classificação do tráfego em fluxos RT e NRT com base na urgência e no tempo de chegada; e o DQL, utilizado para a alocação de recursos sob restrições de potência. A complexidade computacional do escalonamento SRR é $O(N \log N)$, em que N representa o número de UEs ativos, indicando um crescimento logarítmico do tempo de execução em relação a N inferior ao da complexidade linear $O(N)$. Para a alocação de recursos via DQL, a complexidade é $O(NC^2)$, sendo C a quantidade de RBs disponíveis. Como C é constante, a complexidade assintótica do DQL é reduzida para $O(N)$. Dessa forma, a complexidade total do esquema proposto é dominada pelo estágio de escalonamento, mantendo-se em $O(N \log N)$. O WASPAS opera em paralelo com o processo de escalonamento e alocação de recursos como parte de um mecanismo auxiliar de mobilidade (*handover*), acionado apenas sob condições específicas, como degradação de SINR ou sobrecarga da célula. Assim, sua complexidade não está embutida no laço principal de escalonamento executado a cada TTI.

O NCA apresenta complexidade computacional de $O(K^3)$ no cenário otimizado, sendo K a capacidade total do *buffer*. Essa complexidade decorre principalmente da resolução do modelo de Markov tridimensional que descreve as filas de tráfego. Em comparação, o algoritmo PF, em sua implementação clássica para escalonamento em redes móveis, exibe complexidade assintótica de $O(N)$, onde N representa o número total de UEs ativos elegíveis para escalonamento em um dado instante de tempo.

4.5 Metodologia de Análise

Os resultados das simulações foram comparados a dois esquemas de alocação de recursos (PF e NCA), considerados relevantes na literatura, a fim de avaliar

a eficácia do esquema proposto na melhoria da QoS e da eficiência de utilização de recursos.

4.6 Considerações Finais

Este capítulo abordou os principais desafios do escalonamento e da alocação de recursos no *uplink* em redes LTE-A e 5G, destacando as limitações de algoritmos tradicionais como SRR, PF e NCA frente à complexidade computacional e às restrições de QoS. Foram discutidos os fatores que influenciam o processo de alocação, como BSR, CQI e parâmetros energéticos, além das implicações da alocação contígua exigida pelo SC-FDMA. A análise evidenciou a necessidade de soluções mais adaptativas e eficientes para cenários heterogêneos e dinâmicos. Em vista disso foi apresentada uma proposta de esquema de escalonamento e alocação de recursos baseada na combinação do DQL com o método multicritério WASPAS visando superar as limitações identificadas. Os resultados obtidos via simulação e a avaliação da proposta são discutidos no próximo capítulo.

5 APRESENTAÇÃO E ANÁLISE DE RESULTADOS

O presente estudo apresenta a proposta de um novo esquema de alocação de recursos QoS-aware para redes heterogêneas LTE-A/5G, com o objetivo de garantir a QoS, reduzir o consumo energético e mitigar interferências intercelulares em cenários densos e de alta mobilidade. A crescente demanda por largura de banda e a complexidade do tráfego de dados tornam essencial a implementação de estratégias eficientes para lidar com os requisitos de QoS diferenciados das aplicações modernas. O esquema proposto combina algoritmos inteligentes e mecanismos otimizados de *handover* para alcançar uma alocação eficiente de recursos.

A crescente popularidade de serviços multimídia, como *streaming* de vídeo em alta definição e videoconferências, apresenta desafios significativos às redes móveis celulares. Esses serviços, classificados como tráfego RT, exigem baixa latência e alta taxa de transferência para proporcionar uma experiência satisfatória ao usuário. Por outro lado, serviços como navegação na web e e-mails, classificados como tráfego NRT, possuem requisitos de QoS menos rigorosos.

Para atender às necessidades desses diferentes tipos de tráfego, o esquema proposto emprega o algoritmo SRR, responsável por classificar o tráfego em RT e NRT. Esse algoritmo de escalonamento assegura uma alocação justa de recursos entre as classes de tráfego, evitando que o tráfego RT seja prejudicado pelo tráfego NRT.

Após a classificação do tráfego, a alocação de recursos é realizada utilizando o algoritmo DQL, uma técnica de aprendizado de máquina capaz de adaptar-se dinamicamente às condições da rede. O DQL considera a restrição de energia como um fator crítico, especialmente em dispositivos móveis, que possuem capacidade de bateria limitada.

Para assegurar a continuidade da comunicação durante a mobilidade do usuário, o esquema proposto incorpora um mecanismo de *handover* baseado no método WASPAS. Este método multicritério avalia diferentes alternativas de *handover* considerando fatores como intensidade do sinal, carga de tráfego da célula e qualidade do canal. O *handover* baseado no WASPAS minimiza a interferência intercelular e garante transições suaves entre células, evitando interrupções na comunicação.

O desempenho do esquema, identificado como *Qos Aware LTE-A/5G* (QAL) foi avaliado por meio de simulações, considerando diversos cenários, com diferentes quantidades de UEs.

5.1 Análise dos Resultados

5.1.1 Atraso Médio

As Figuras 23 e 24 apresentam o desempenho de atraso médio para tráfego RT e NRT considerando diferentes números de UEs. Conforme pode ser observado, o esquema proposto reduz significativamente os atrasos do tráfego RT em comparação aos algoritmos PF e NCA, atendendo aos requisitos de QoS de aplicações sensíveis à latência, como videoconferências e jogos online. Essa melhoria é obtida por meio da priorização eficiente do tráfego RT, realizada pelo SRR, e da alocação otimizada de recursos, conduzida pelo DQL. O algoritmo PF, baseado no princípio da justiça proporcional, favorece usuários com melhores condições de canal, o que resulta em maiores atrasos na borda da célula. Já o NCA não prioriza adequadamente o tráfego urgente. O esquema proposto supera essas limitações, apresentando desempenho superior em termos de latência.

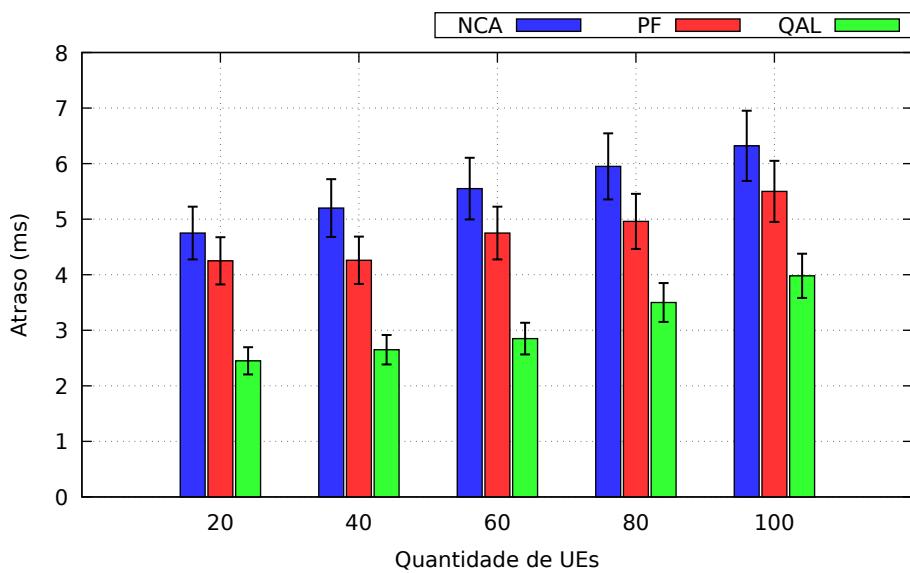


Figura 23 – Atraso para tráfego RT em função do número de UEs.

5.1.2 Throughput

As Figuras 25 e 26 mostram que o esquema proposto proporciona um desempenho superior com relação ao *throughput* à medida que o número de UEs aumenta. Este incremento possibilita aos usuários usufruírem de taxas de transferência mais altas, proporcionando uma experiência mais fluida em aplicações multimídia e maior agilidade em atividades como navegação web.

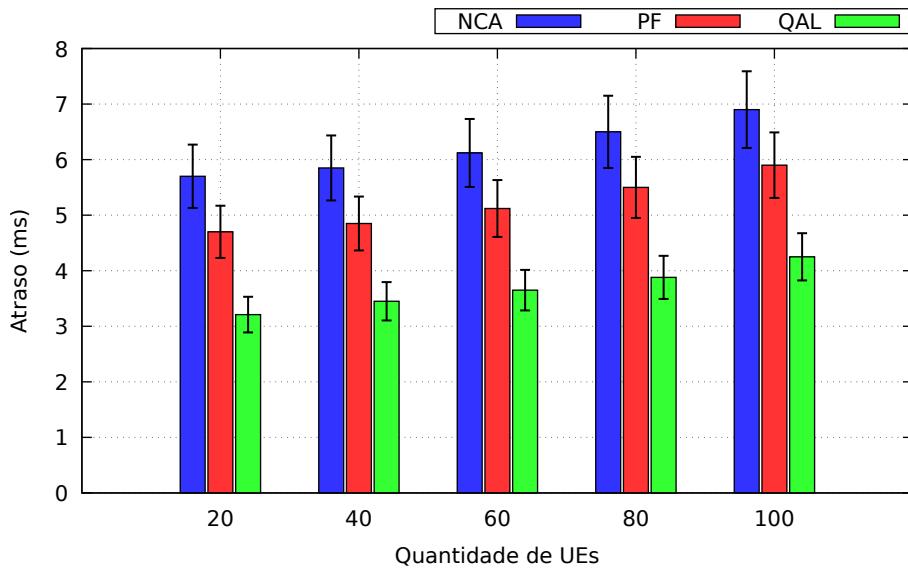


Figura 24 – Atraso para tráfego NRT em função do número de UEs.

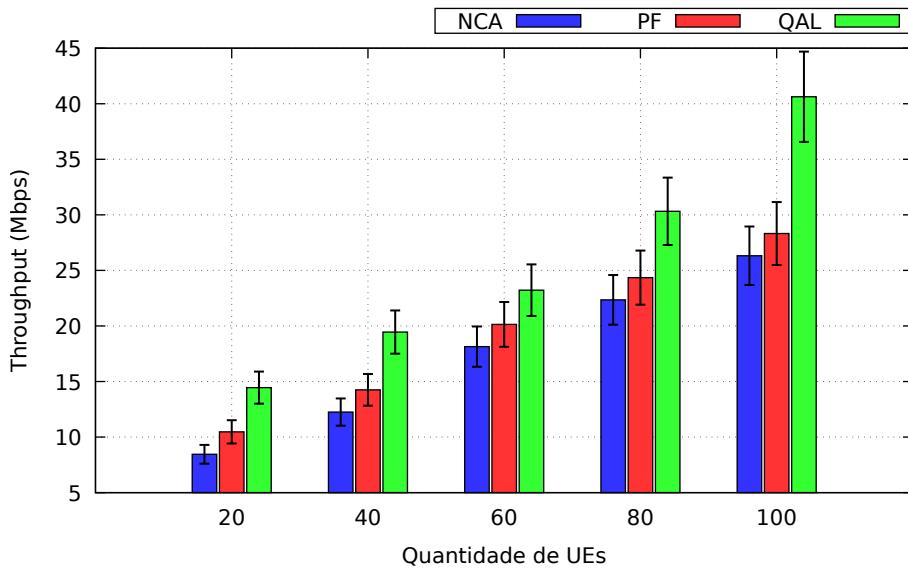


Figura 25 – Throughput para tráfego RT em função do número de UEs.

5.1.3 Índice de Justiça

O índice de justiça, que avalia a equidade na distribuição de recursos entre os usuários, também foi significativamente aprimorado. Essa melhoria assegura que os recursos da rede sejam compartilhados de maneira justa, evitando a monopolização por parte de determinados usuários e garantindo uma experiência uniforme para todos.

As Figuras 27 e 28 destacam que o esquema proposto apresenta índice de justiça mais elevado em comparação aos escalonadores avaliados, tanto para tráfego RT quanto para NRT. O QAL, por meio de sua estrutura de escalonamento de objetivo duplo e mecanismo de alocação de recursos baseado em DQL, prioriza o atendimento

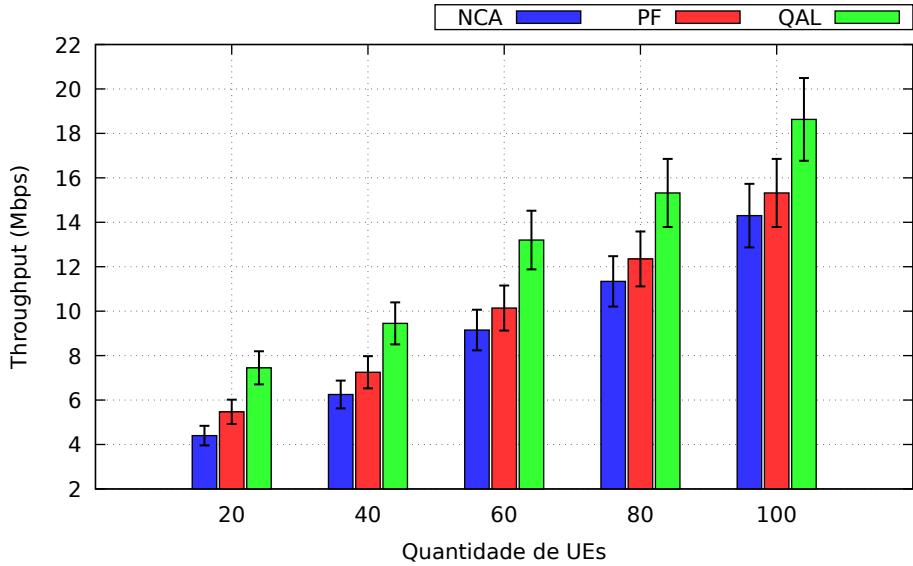


Figura 26 – *Throughput* para tráfego NRT em função do número de UEs.

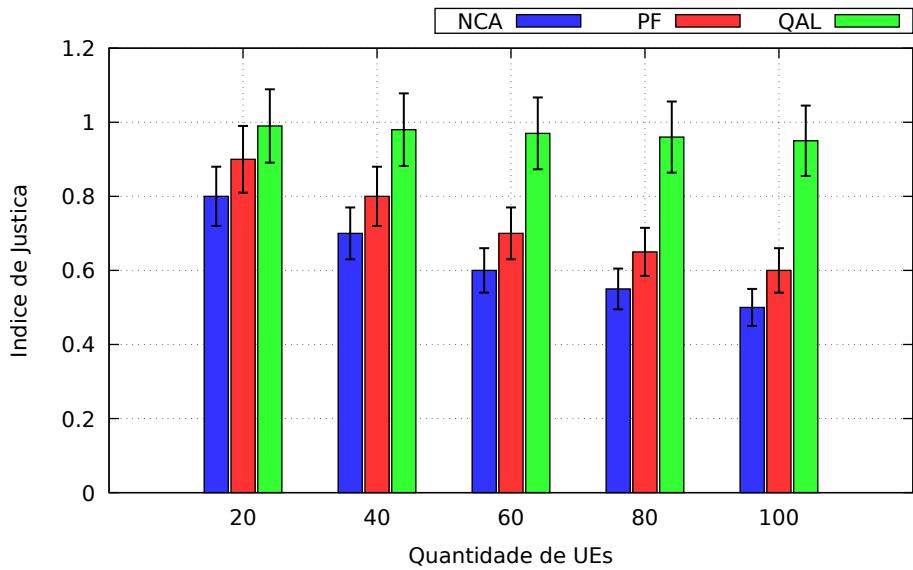


Figura 27 – Índice de justiça para tráfego RT em função do número de UEs.

aos requisitos de QoS do tráfego RT, que exige baixa latência e baixa taxa de perda de pacotes. Simultaneamente, o tráfego NRT, que é mais tolerante a atrasos, recebe alocação equitativa de recursos.

5.1.4 Taxa de Queda de Chamadas

A taxa de queda de chamadas, que mede a probabilidade de interrupção de chamadas devido a falhas na rede, foi reduzida consideravelmente pelo esquema proposto. Essa diminuição deve-se ao mecanismo de *handover* baseado no método WASPAS, que proporciona transições suaves entre células, mitigando interrupções na comunicação.

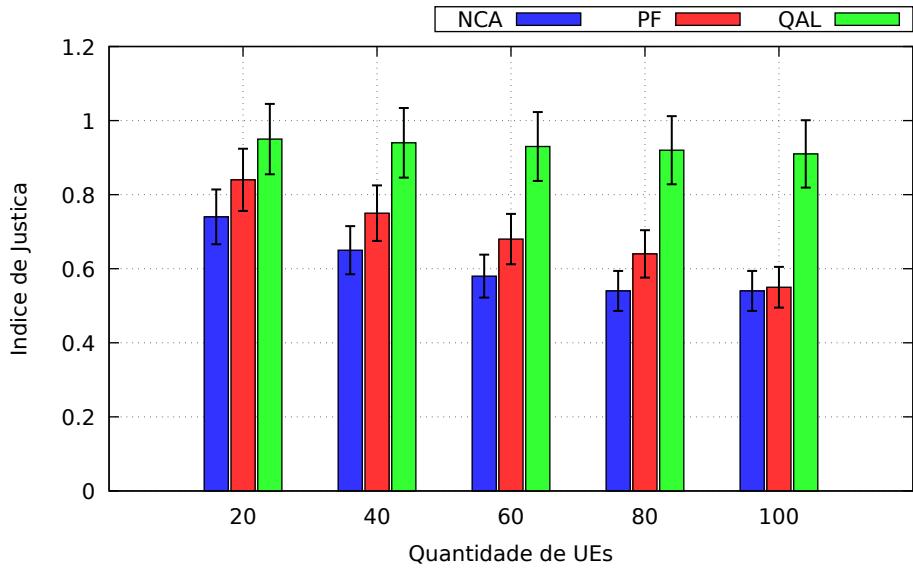


Figura 28 – Índice de justiça para tráfego NRT em função do número de UEs.

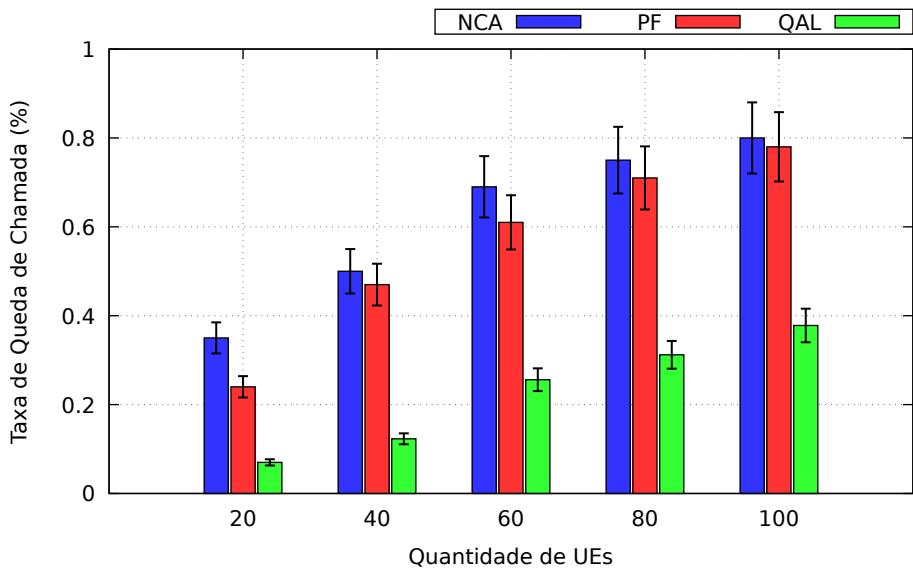


Figura 29 – Taxa de queda de chamadas para tráfego RT em função do número de UEs.

As Figuras 29 e 30 indicam que o esquema proposto reduz a taxa de queda de chamadas, o que pode ser atribuído à precisão das estratégias de *handover* e à eficiência na alocação de recursos.

5.1.5 Taxa de Perda de Pacotes

A taxa de perda de pacotes, indicador da quantidade de pacotes de dados perdidos durante a transmissão, também apresentou uma redução significativa. Essa melhora contribui para uma comunicação mais confiável e eficiente, minimizando a perda de informações e reduzindo a necessidade de retransmissões.

As Figuras 31 e 32 evidenciam que o esquema proposto diminui a taxa de

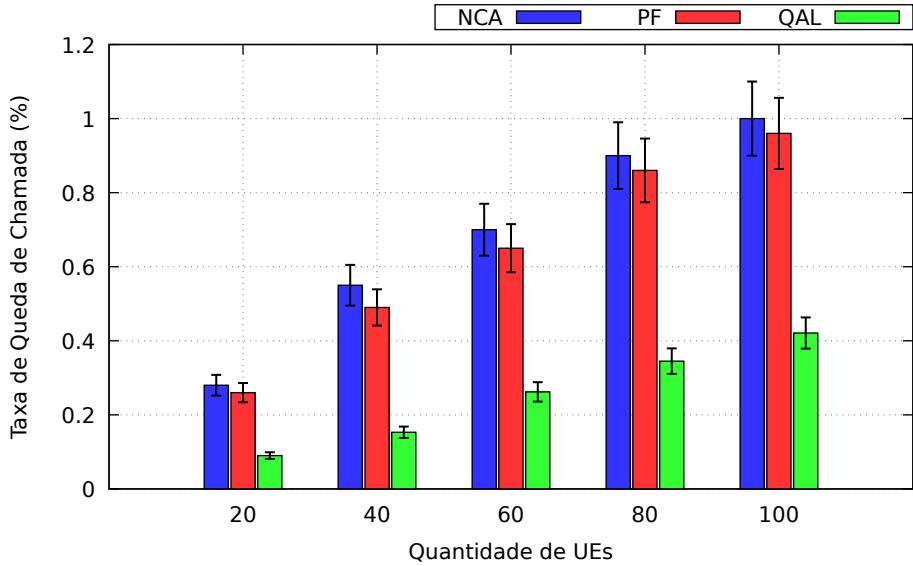


Figura 30 – Taxa de queda de chamadas para tráfego NRT em função do número de UEs.

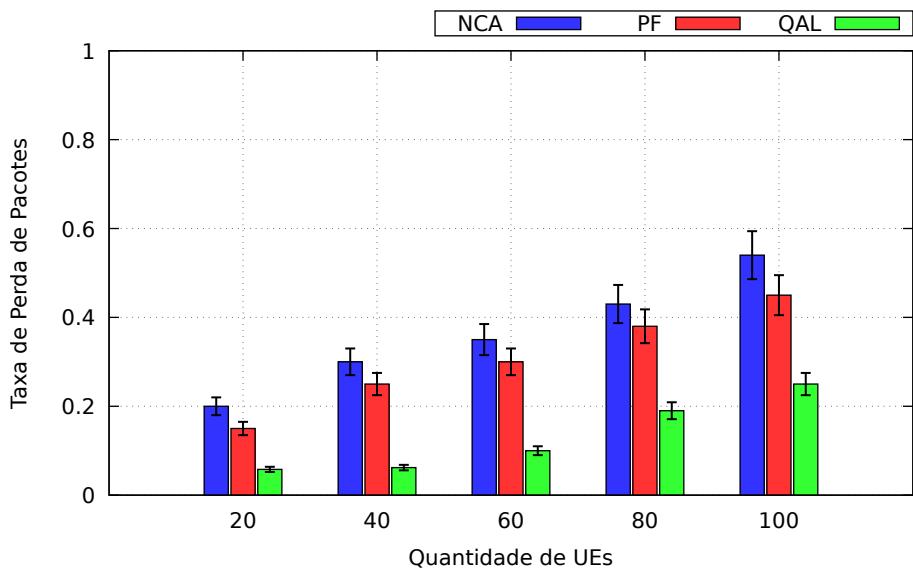


Figura 31 – Taxa de perda de pacotes para tráfego RT em função do número de UEs.

perda de pacotes em aproximadamente 50% quando comparado aos métodos avaliados. Esse desempenho é garantido pelo gerenciamento eficiente de tráfego, que reduz a perda de pacotes e minimiza as solicitações de retransmissão. O esquema proposto otimiza a alocação de recursos por meio do DQL, que aprende dinamicamente com o ambiente da rede para atribuir canais e potências de transmissão de forma eficiente. Ao considerar fatores como o consumo de energia do usuário, a potência de transmissão atual e os requisitos de QoS do tráfego, o DQL garante que os recursos sejam alocados de maneira a minimizar o congestionamento da rede e, consequentemente, a perda de pacotes.

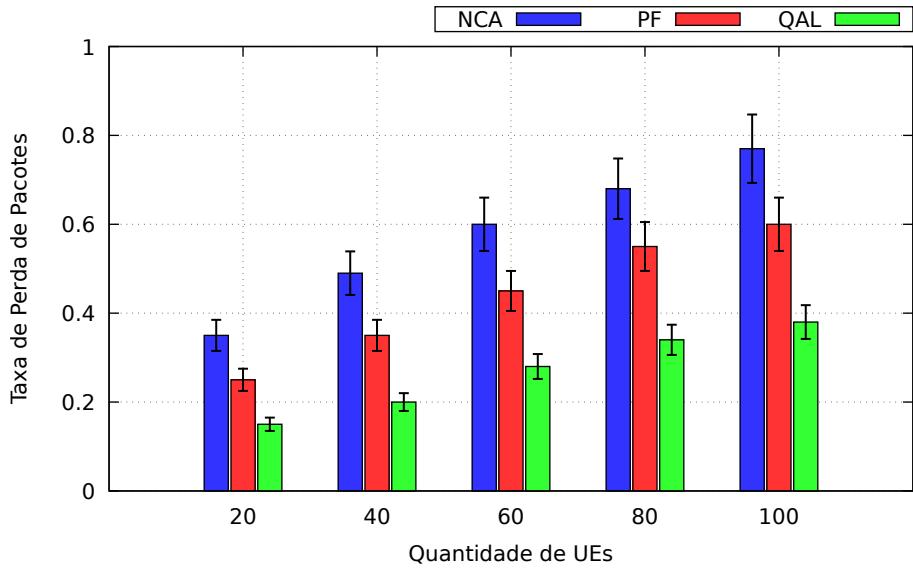


Figura 32 – Taxa de perda de pacotes para tráfego NRT em função do número de UEs.

5.1.6 Complexidade

A Figura 33 apresenta a análise comparativa da complexidade computacional do esquema proposto em relação aos demais métodos avaliados. Em cenários com poucos UEs, a complexidade é reduzida; contudo, a diminuição do número de UEs é limitada em ambientes de larga escala. O esquema proposto apresenta complexidade computacional semelhante à do PF, com maior adaptabilidade e melhor atendimento aos requisitos de QoS.

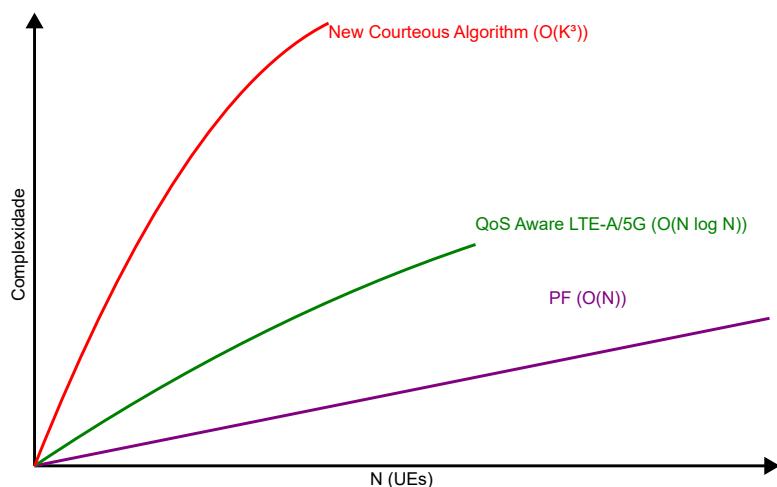


Figura 33 – Análise comparativa da complexidade computacional do esquema proposto em relação aos demais métodos avaliados.

5.2 Discussão dos Resultados e Implicações

Os resultados obtidos por meio de simulações indicam que o esquema proposto apresenta desempenho superior em relação ao PF e ao NCA nas métricas analisadas. A redução no atraso e o aumento no *throughput* destacam a eficácia do esquema em garantir QoS para aplicações sensíveis à latência, como o tráfego RT.

A melhoria no índice de justiça demonstra a capacidade do esquema em distribuir os recursos de forma equitativa entre os usuários, promovendo um ambiente de rede mais equilibrado e justo. Além disso, a redução significativa nas taxas de queda de chamadas e de perda de pacotes reforça a robustez do esquema em cenários caracterizados por alta mobilidade e tráfego intenso.

Embora o esquema tenha sido avaliado em um ambiente de simulação, é importante considerar que a implementação em cenários reais pode envolver desafios adicionais, como a complexidade do algoritmo DQL. Embora o DQL demonstre alta capacidade de aprendizado adaptativo, sua aplicação pode ser computacionalmente onerosa, especialmente em redes heterogêneas de alta complexidade. A implementação em tempo real em BS representa um desafio, exigindo o uso de hardware e software especializados para lidar com a alta carga computacional. No entanto, os resultados obtidos em simulações sugerem que o esquema tem potencial para ser explorado em estudos futuros voltados para a validação em redes heterogêneas LTE-A/5G reais.

A classificação de tráfego em apenas duas categorias, RT e NRT, pode ser ampliada para atender às características diversificadas dos tipos de tráfego em redes 5G e futuras. Essa expansão é essencial para capturar as demandas específicas de aplicações emergentes e garantir maior granularidade na alocação de recursos.

A decisão de *handover* no esquema proposto considera métricas como SINR, RSRP, velocidade do usuário e a carga de tráfego da célula. No entanto, a inclusão de métricas adicionais, como a capacidade do *buffer* do UE e capacidade da célula alvo (além da carga de tráfego), poderia melhorar ainda mais a precisão das decisões de *handover*. Esse aspecto constitui uma oportunidade para pesquisas futuras voltadas para o aprimoramento do processo de mobilidade em redes móveis.

As contribuições deste trabalho possuem implicações significativas para o desenvolvimento de redes móveis celulares mais eficientes e adaptadas às demandas do tráfego contemporâneo. A aplicação do esquema proposto em cenários reais pode proporcionar melhorias substanciais na QoS e na experiência do usuário em redes heterogêneas LTE-A/5G, impulsionando a evolução das redes móveis e a ampliação de serviços multimídia avançados.

5.3 Considerações Finais

Este capítulo apresentou uma avaliação via simulação da proposta de algoritmo de escalonamento e alocação de recursos no enlace *uplink* de HetNets LTE-A/5G, baseado na combinação de DQL e do método multicritério WASPAS. A metodologia foi estruturada em três componentes principais: escalonamento com objetivo duplo para tráfego RT e NRT, alocação de recursos com restrição energética e tomada de decisão de *handover* orientada por múltiplos critérios. As simulações demonstraram a eficácia do esquema proposto quando comparado a algoritmos tradicionais como PF e NCA.

6 CONCLUSÃO

A otimização da alocação de recursos em ambientes de redes heterogêneas LTE-A/5G surge como um fator crucial para assegurar a QoS em contextos marcados pela densidade de usuários e diversidade de tráfego. No presente estudo, introduz-se um esquema que une técnicas de aprendizado por reforço profundo com métodos multicritério de decisão, buscando um equilíbrio entre eficiência espectral, consumo energético e uma distribuição equitativa dos recursos. A proposta abrange um algoritmo de escalonamento que busca conceder prioridade ao tráfego RT em detrimento do tráfego NRT, em conjunto com um mecanismo de alocação baseado em DQL, para maximizar o *throughput* e a eficiência energética. Ainda, um processo de *handover* inteligente utilizando o WASPAS faz parte da técnica desenvolvida neste trabalho.

Os resultados obtidos nas simulações conduzidas no simulador ns-3 demonstram que a abordagem proposta supera paradigmas tradicionais, ao proporcionar uma redução da latência para o tráfego em tempo real, uma ampliação do *throughput* total e um aumento no índice de justiça. Adicionalmente, ocorreu redução na taxa de perda de pacotes em cenários de alta densidade, evidenciando a robustez da solução mesmo em condições adversas. A adoção do DQL possibilitou uma adaptação dinâmica às variações de canal e às demandas dos usuários, otimizando assim a alocação de recursos em tempo real. O método WASPAS, permitiu aprimorar a eficiência do *handover* ao considerar variáveis como SINR, RSRP e carga da célula, resultando em uma diminuição da taxa de queda de chamadas.

Esses avanços são fundamentais para aumentar a confiabilidade e a eficiência espectral das redes heterogêneas, assegurando desempenho superior aos serviços disponibilizados. Apesar dos resultados encorajadores, há ainda desafios a serem enfrentados. A complexidade computacional associada ao DQL pode inviabilizar sua implementação em BS com recursos limitados, exigindo o desenvolvimento de soluções otimizadas para operações em tempo real. A classificação binária adotada para os fluxos de tráfego, restrita às categorias RT e NRT, revela-se limitada diante da diversidade de serviços nas redes modernas. A inclusão de classes mais granulares, como URLLC e mMTC, pode conferir maior precisão à alocação de recursos.

Pesquisas futuras podem direcionar-se à exploração da integração de novas métricas no processo de *handover*, como a capacidade do *buffer* do UE e capacidade da célula alvo (além da carga de tráfego), visando uma tomada de decisão ainda mais eficiente. A adaptação da proposta para redes 6G também se revela como uma via promissora, especialmente em cenários de comunicação ultra-massiva e em aplicações fundamentadas em inteligência artificial federada. Ainda, a combinação do aprendizado por reforço com técnicas de *transfer learning* pode contribuir significativa-

mente para a redução do consumo energético em dispositivos IoT. O presente estudo reafirma a viabilidade do emprego de aprendizado por reforço no gerenciamento dinâmico de recursos, promovendo avanços substanciais nas redes móveis em direção a uma maior eficiência e automação. A abordagem delineada demonstra um potencial considerável para aplicação em redes 5G e nas futuras gerações, viabilizando soluções escaláveis e eficazes para a gestão de recursos em cenários dinâmicos e densos.

REFERÊNCIAS

3GPP. 5G System Overview. 2022. Disponível em: https://www.3gpp.org/technologies/5g-system-overview?_x_tr_sl=en&_x_tr_tl=pt&_x_tr_hl=pt&_x_tr_pto=tc. Acesso em: 17/12/2024.

3GPP. NR; Physical channels and modulation (Release 18). 2024. url<https://www.3gpp.org/DynaReport/38211.htm>.

3GPP. Release 16 Description; Summary of Rel-16 Work Items. 2018. Disponível em: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3493>. Acesso em: 17/12/2024.

3GPP. Release 17 Description; Summary of Rel-17 Work Items. 2021. Disponível em: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3937>. Acesso em: 17/12/2024.

3GPP. Release 18 Description; Summary of Rel-18 Work Items. 2023. Disponível em: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=4229>. Acesso em: 17/12/2024.

3GPP. Specifications and Technologies/Rleases/Release 18. 2024. Disponível em: <https://www.3gpp.org/specifications-technologies/releases/release-18>. Acesso em: 17/12/2024.

ABU-ALI, N. *et al.* Uplink Scheduling in LTE and LTE-Advanced: Tutorial, Survey and Evaluation Framework. **IEEE Communications Surveys and Tutorials**, Institute of Electrical e Electronics Engineers (IEEE), v. 16, n. 3, p. 1239–1265, 2014. DOI: 10.1109/surv.2013.1127.00161.

ABUBAKAR, A. I. *et al.* Revenue Maximization Through Cell Switching and Spectrum Leasing in 5G HetNets. **IEEE Access**, Institute of Electrical e Electronics Engineers (IEEE), v. 10, p. 48301–48317, 2022. DOI: 10.1109/access.2022.3172280.

AGIWAL, M.; ROY, A.; SAXENA, N. Next Generation 5G Wireless Networks: A Comprehensive Survey. **IEEE Communications Surveys & Tutorials**, Institute of Electrical e Electronics Engineers (IEEE), v. 18, n. 3, p. 1617–1655, 2016. ISSN 1553-877X. DOI: 10.1109/comst.2016.2532458.

AHMAD, R. *et al.* Handover in LTE-advanced wireless networks: state of art and survey of decision algorithm. **Telecommunication Systems**, Springer Science e

Business Media LLC, v. 66, n. 3, p. 533–558, 2017. ISSN 1572-9451. DOI: 10.1007/s11235-017-0303-6.

ALAM, M. J. *et al.* Optimizing cell association in 5G and beyond networks: a modified load-aware biased technique. **Telecommunication Systems**, Springer Science e Business Media LLC, v. 87, n. 3, p. 731–742, 2024. ISSN 1572-9451. DOI: 10.1007/s11235-024-01202-w.

ALAMU, O. *et al.* Energy efficiency techniques in ultra-dense wireless heterogeneous networks: An overview and outlook. **Engineering Science and Technology, an International Journal**, Elsevier BV, v. 23, p. 1308–1326, 2020. DOI: 10.1016/j.estch.2020.05.001.

ALAVIZADEH, H.; ALAVIZADEH, H.; JANG-JACCARD, J. Deep Q-Learning Based Reinforcement Learning Approach for Network Intrusion Detection. **Computers**, MDPI AG, v. 11, n. 3, p. 41, 2022. ISSN 2073-431X. DOI: 10.3390/computers11030041.

ALI, M. *et al.* Smart heterogeneous networks: a 5G paradigm. **Telecommunication Systems**, Springer Science e Business Media LLC, v. 66, n. 2, p. 311–330, 2017. DOI: 10.1007/s11235-017-0291-6.

ALRAIH, S. *et al.* A Survey on Handover Optimization in Beyond 5G Mobile Networks: Challenges and Solutions. **IEEE Access**, Institute of Electrical e Electronics Engineers (IEEE), v. 11, p. 59317–59345, 2023. ISSN 2169-3536. DOI: 10.1109/access.2023.3284905.

AMITABHA GHOSH, R. R. **Essentials of Lte and Lte-A**. CAMBRIDGE, 2011. 264 p. ISBN 9780521768702. DOI: 10.1017/CBO9780511997082. Disponível em: https://www.ebook.de/de/product/14736348/amitabha_ghosh_rapeepat_ratasuk_essentials_of_lte_and_lte_a.html.

ARAFAT, A. O.; GREGORY, M. A. A Weight-Based Resource Scheduling Algorithm for Uplink LTE-A Femtocell Network. **Wireless Personal Communications**, Springer Science e Business Media LLC, v. 114, n. 3, p. 1945–1961, 2020. ISSN 1572-834X. DOI: 10.1007/s11277-020-07456-2.

BARBARA, F. *et al.* Interactive Internet Framework Proposal of WASPAS Method: A Computational Contribution for Decision-Making Analysis. **Mathematics**, MDPI AG, v. 11, n. 15, p. 3375, 2023. ISSN 2227-7390. DOI: 10.3390/math11153375.

BOJOVIĆ, B. *et al.* Enhancing 5G QoS Management for XR Traffic Through XR Loopback Mechanism. **IEEE Journal on Selected Areas in Communications**,

Institute of Electrical e Electronics Engineers (IEEE), v. 41, n. 6, p. 1772–1786, 2023.
ISSN 1558-0008. DOI: 10.1109/jsac.2023.3273701.

CARVALHO FERREIRA, G. de. **Técnicas para Melhorar a Eficiência do Sensoriamento Colaborativo em Redes 5G para Áreas Remotas**. 2020. p. 86.
Diss. (Mestrado) – Universidade de Brasília.

CASTRO ALVES, A. A. de. **Arranjo de antenas reconfigurável para 5G utilizando metamateriais**. 2020. Tese (Doutorado) – Universidade Federal de Itajubá.

CHAKRABORTY, S.; ZAVADSKAS, E.; ANTUCHEVICIENE, J. Applications of waspas method as a multi-criteria decision-making tool. **Economic computation and economic cybernetics studies and research**, v. 49, p. 5–22, 2015.

CHRISTOPHER COX, B. C. **An introduction to LTE LTE, LTE-advanced, SAE, VoLTE and 4G mobile communications**. John Wiley e Sons, 2014. 486 p. ISBN 9781118818039. Disponível em: https://www.ebook.de/de/product/22341758/christopher_cox_baggy_cox_an_introduction_to_lte.html.

COSTA SILVA, K. da. **Estratégia de Planejamento e Otimização do Handover em Redes Móveis Densificadas**. 2018. p. 88. Tese (Doutorado) – Universidade Federal do Pará.

D, R. S. V.; P, P. M.; D, G. R. M. Intelligent resource allocation decision using deep learning and optimization techniques for HetNets. **Wireless Networks**, Springer Science e Business Media LLC, v. 29, n. 7, p. 3105–3119, 2023. ISSN 1572-8196.
DOI: 10.1007/s11276-023-03360-2.

DAHLMAN, E. **4G, LTE-advanced pro and the road to 5G**. Edição: Johan Sköld e Stefan Parkvall. Third edition. London: Academic Press, 2016. 11 p. Includes bibliographical references and index. ISBN 9780128046111.

DAHLMAN, E. **5G NR : the next generation wireless access technology**. Amsterdam: Academic Press, 2021. ISBN 9780128223208.

DAHLMAN, E. **5G/5G-ADVANCED**: The next generation wireless access technology. Edição: Stefan Parkvall e Johan Sköld. [S.I.]: ELSEVIER ACADEMIC PRESS, 2023. 1 p. ISBN 0443131740.

DAHLMAN, E.; PARKVALL, S.; SKOLD, J. **4G, LTE-Advanced Pro and The Road to 5G, Third Edition**. 3rd. USA: Academic Press, Inc., 2016. 616 p. ISBN 0128045752.

DANGI, R. *et al.* Study and Investigation on 5G Technology: A Systematic Review. **Sensors**, MDPI AG, v. 22, n. 1, p. 26, 2021. ISSN 1424-8220. DOI: 10.3390/s22010026.

EDET, M. V.; EKPE, U. M.; UKOMMI, U. S. Advancing beyond 5G: A review of network densification strategies. **World Journal of Advanced Engineering Technology and Sciences**, GSC Online Press, v. 15, n. 1, p. 774–779, 2025. ISSN 2582-8266. DOI: 10.30574/wjaets.2025.15.1.0185.

ERICSSON. **Mobile data traffic outlook**. 2024. Disponível em: <https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/mobile-traffic-forecast>. Acesso em: 24/10/2024.

FERNG, H.-W. *et al.* Urgency-Based Fair Scheduling for LTE to Improve Packet Loss and Fairness: Design and Evaluation. **IEEE Transactions on Vehicular Technology**, Institute of Electrical e Electronics Engineers (IEEE), v. 68, n. 3, p. 2825–2836, 2019. ISSN 1939-9359. DOI: 10.1109/tvt.2019.2893977.

FERREIRA, M. V. G. **Escalonamento de Recursos em Redes Sem Fio 5G Baseado em Otimização de Retardo e de Alocação de Potência Considerando Comunicação Dispositivo a Dispositivo**. 2021. p. 202. Tese de Doutorado – Universidade Federal de Goiás, Goiânia.

GATTI, R. *et al.* Optimal resource scheduling algorithm for cell boundaries users in heterogenous 5G networks. **Physical Communication**, v. 55, p. 101915, 2022. DOI: 10.1016/j.phycom.2022.101915.

GAUTAM, V. K.; TAMMA, B. R. RETALIN: A Queue Aware Uplink Scheduling Scheme for Reducing Scheduling Signaling Overhead in 5G NR. **IEEE Access**, v. 12, p. 16632–16649, 2024. DOI: 10.1109/ACCESS.2024.3359028.

GUO, C. Improved Smoothed Round Robin Schedulers for High-Speed Packet Networks. In: IEEE INFOCOM 2008 - The 27th Conference on Computer Communications. IEEE, 2008. p. 906–914. DOI: 10.1109/infocom.2008.143.

GUO, C. SRR: An O(1) Time-Complexity Packet Scheduler for Flows in Multiservice Packet Networks. **IEEE/ACM Transactions on Networking**, Institute of Electrical e Electronics Engineers (IEEE), v. 12, n. 6, p. 1144–1155, 2004. ISSN 1063-6692. DOI: 10.1109/tnet.2004.838601.

GUPTA, A.; JHA, R. K. A Survey of 5G Network: Architecture and Emerging Technologies. **IEEE Access**, Institute of Electrical e Electronics Engineers (IEEE), v. 3, p. 1206–1232, 2015. ISSN 2169-3536. DOI: 10.1109/access.2015.2461602.

GURES, E. *et al.* A Comprehensive Survey on Mobility Management in 5G Heterogeneous Networks: Architectures, Challenges and Solutions. **IEEE Access**, Institute of Electrical e Electronics Engineers (IEEE), v. 8, p. 195883–195913, 2020. ISSN 2169-3536. DOI: 10.1109/access.2020.3030762.

HAQUE, M. E. *et al.* A Survey of Scheduling in 5G URLLC and Outlook for Emerging 6G Systems. **IEEE Access**, Institute of Electrical e Electronics Engineers (IEEE), v. 11, p. 34372–34396, 2023. ISSN 2169-3536. DOI: 10.1109/access.2023.3264592.

HUANG, P.-H. *et al.* Subcarrier Allocation in Multiuser OFDM Systems: Complexity and Approximability. In: 2010 IEEE Wireless Communication and Networking Conference. IEEE, 2010. DOI: 10.1109/wcnc.2010.5506244.

IQBAL, M. U. *et al.* Improving the QoS in 5G HetNets Through Cooperative Q-Learning. **IEEE Access**, Institute of Electrical e Electronics Engineers (IEEE), v. 10, p. 19654–19676, 2022. ISSN 2169-3536. DOI: 10.1109/access.2022.3151090.

ITU-R. Recommendation M.687-2 International Mobile Telecommunications-2000 (IMT-2000). v. 2000. 1997. p. 1–20.

ITU-T. Recommendation G.729. 2023. <https://www.itu.int/rec/T-REC-G.729/>.

JON, J.-H. *et al.* Enhanced uplink handover scheme for improvement of energy efficiency and QoS in LTE-A/5G HetNet with ultra-dense small cells. **Wireless Networks**, v. 30, p. 1321–1338, 2024. DOI: 10.1007/s11276-023-03547-7.

KAMAL, M. A. *et al.* Resource Allocation Schemes for 5G Network: A Systematic Review. **Sensors**, MDPI AG, v. 21, n. 19, p. 6588, 2021. ISSN 1424-8220. DOI: 10.3390/s21196588.

KAPOOR, S.; GRACE, D.; CLARKE, T. User Association in Ultra-Dense Small Cell Dynamic Vehicular Networks: A Reinforcement Learning Approach. **Journal of Communications and Information Networks**, v. 4, n. 1, p. 1–12, 2019. DOI: 10.23919/JCIN.2019.8916641.

KASIM, A. N. *et al.* Evolutionary paths towards mobility management in 5G Heterogeneous Networks. In: 2020 IEEE Microwave Theory and Techniques in

Wireless Communications (MTTW). IEEE, 2020. p. 24–29. DOI: 10.1109/mttw51045.2020.9245046.

KAZI, B. U.; WAINER, G. A. Next generation wireless cellular networks: ultra-dense multi-tier and multi-cell cooperation perspective. **Wireless Networks**, Springer Science e Business Media LLC, v. 25, n. 4, p. 2041–2064, 2018. DOI: 10.1007/s11276-018-1796-y.

KHAN, S. A. *et al.* Handover management over dual connectivity in 5G technology with future ultra-dense mobile heterogeneous networks: A review. **Engineering Science and Technology, an International Journal**, Elsevier BV, v. 35, p. 101172, 2022. ISSN 2215-0986. DOI: 10.1016/j.estch.2022.101172.

KOUTLIA, K. *et al.* Calibration of the 5G-LENA system level simulator in 3GPP reference scenarios. **Simulation Modelling Practice and Theory**, Elsevier BV, v. 119, p. 102580, 2022. ISSN 1569-190X. DOI: 10.1016/j.simpat.2022.102580.

KUMAR, S.; AR, A. K. 5G New Radio Deployment Modes. **International Research Journal of Engineering and Technology**, 2020.

LEI, H.; LI, X. System level study of LTE uplink employing SC-FDMA and virtual MU-MIMO. In: 2009 IEEE International Conference on Communications Technology and Applications. IEEE, 2009. p. 152–156. DOI: 10.1109/iccomta.2009.5349218.

LIANG, R.; LYU, H.; FAN, J. A deep reinforcement learning-based power control scheme for the 5G wireless systems. **China Communications**, Institute of Electrical e Electronics Engineers (IEEE), v. 20, n. 10, p. 109–119, 2023. ISSN 1673-5447. DOI: 10.23919/jcc.ea.2021-0523.202302.

MAMANE, A. *et al.* Scheduling Algorithms for 5G Networks and Beyond: Classification and Survey. **IEEE Access**, Institute of Electrical e Electronics Engineers (IEEE), v. 10, p. 51643–51661, 2022. ISSN 2169-3536. DOI: 10.1109/access.2022.3174579.

MARDANI, A. *et al.* A systematic review and meta-Analysis of SWARA and WASPAS methods: Theory and applications with recent fuzzy developments. **Applied Soft Computing**, Elsevier BV, v. 57, p. 265–292, 2017. ISSN 1568-4946. DOI: 10.1016/j.asoc.2017.03.045.

MATA, S. H. d.; GUARDIEIRO, P. R. Resource allocation for the LTE uplink based on Genetic Algorithms in mixed traffic environments. **Computer Communications**, Elsevier BV, v. 107, p. 125–137, 2017. DOI: 10.1016/j.comcom.2017.04.004.

MBALE, T. F. **PERFORMANCE ANALYSIS OF RADIO RESOURCE ALLOCATION IN MIMO LTE SYSTEMS.** 2021. Diss. (Mestrado) – University of Zambia.

MBULWA, A. I. *et al.* Self-Optimization of Handover Control Parameters for 5G Wireless Networks and Beyond. **IEEE Access**, Institute of Electrical e Electronics Engineers (IEEE), v. 12, p. 6117–6135, 2024. ISSN 2169-3536. DOI: 10.1109/access.2023.3346039.

MONIKANDAN B, S.; SIVASUBRAMANIAN, A.; BABU, S. A Review of MAC Scheduling Algorithms in LTE System. **International Journal on Advanced Science, Engineering and Information Technology**, Insight Society, v. 7, n. 3, p. 1056, 2017. ISSN 2088-5334. DOI: 10.18517/ijaseit.7.3.2059.

MOUADH, R.; SOFIANE, E. **Reducing Latency In 5G networks.** 2020. Diss. (Mestrado) – Université Amar Telidji - LAGHOUAT.

MUKHOPADHYAY, A.; DAS, G. Low Complexity Fair Scheduling in LTE/LTE-A Uplink Involving Multiple Traffic Classes. **IEEE Systems Journal**, Institute of Electrical e Electronics Engineers (IEEE), v. 15, n. 2, p. 1616–1627, 2021. ISSN 2373-7816. DOI: 10.1109/jsyst.2020.2991325.

MYUNG, H. G.; LIM, J.; GOODMAN, D. J. Single Carrier FDMA for Uplink Wireless Transmission. **IEEE Vehicular Technology Magazine**, v. 1, n. 3, p. 30–38, 2006. DOI: 10.1109/MVT.2006.307304.

NOMEIR, M. W.; GADALLAH, Y.; SEDDIK, K. G. Machine Learning-Based Uplink Scheduling Approaches for Mixed Traffic in Cellular Systems. **IEEE Access**, Institute of Electrical e Electronics Engineers (IEEE), v. 11, p. 10238–10253, 2023. ISSN 2169-3536. DOI: 10.1109/access.2023.3240569.

NSNAM. **ns-3 Manual.** 2024. Disponível em:
<https://www.nsnam.org/docs/release/3.40/manual/html/index.html>. Acesso em: 01/10/2024.

NSNAM. **ns-3 Network Simulator.** 2024. Disponível em: <https://www.nsnam.org/>. Acesso em: 01/10/2024.

NWOGU, O. A. **Resource Allocation Enhancements for 5G New Radio Architecture.** 2021. Tese (Doutorado) – Université Paris-Nord - Paris XIII. NNT: 2021PA131075, tel-03886037. hal: tel-03886037.

ORDONEZ-LUCENA, J. *et al.* Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges. **IEEE Communications Magazine**, Institute of Electrical e Electronics Engineers (IEEE), v. 55, n. 5, p. 80–87, 2017. DOI: 10.1109/mcom.2017.1600935.

OUSSAKEL, I. **4G/5G Cellular Networks Metrology and Management**. 2020. Tese (Doutorado) – Université Paul Sabatier - Toulouse III. NNT: 2020TOU30261, tel-03236078. hal: tel-03236078.

PETERSON, L. L.; SUNAY, O.; DAVIE, B. S. **Private 5G: A Systems Approach**. Systems Approach LLC, 2023. ISBN 9781736472163.

PIRO, G. *et al.* Simulating LTE Cellular Systems: An Open-Source Framework. **IEEE Transactions on Vehicular Technology**, Institute of Electrical e Electronics Engineers (IEEE), v. 60, n. 2, p. 498–513, 2011. DOI: 10.1109/tvt.2010.2091660.

RAGALEUX, A.; BAEY, S.; KARACA, M. Standard-Compliant LTE-A Uplink Scheduling Scheme With Quality of Service. **IEEE Transactions on Vehicular Technology**, Institute of Electrical e Electronics Engineers (IEEE), v. 66, n. 8, p. 7207–7222, 2017. DOI: 10.1109/tvt.2017.2654299.

RISCHKE, J. *et al.* Empirical Study of 5G Downlink & Uplink Scheduling and its Effects on Latency. In: 2022 IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM). IEEE, 2022. p. 11–19. DOI: 10.1109/wowmom54355.2022.00017.

SANCHEZ, S. B.; GERARDINO, G. A. P. **LTE-Advanced HetNet Investigations Under Realistic Conditions**. 2014. Diss. (Mestrado) – Aalborg University.

SANIJ, M. T.; CASETTI, C. A system-level assessment of Uplink CoMP in LTE-A Heterogeneous Networks. In: 2016 13th IEEE Annual Consumer Communications Networking Conference (CCNC). IEEE, 2016. p. 188–193. DOI: 10.1109/ccnc.2016.7444754.

SANTOS, E. C. **Alocação de Recursos baseada em Clustering com Aprendizado de Características e Orientação a QoS em Redes LTE-Advanced**. 2019. Tese (Doutorado) – Universidade Federal de Uberlândia. DOI: 10.14393/ufu.te.2019.2060.

SEMOV, P. *et al.* Evolution of Mobile Networks and C-RAN on the Road Beyond 5G. In: 2020 43rd International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2020. p. 392–398. DOI: 10.1109/tsp49548.2020.9163513.

SESSIA, S.; TOUFIK, I.; BAKER, M. **LTE – The UMTS Long Term Evolution: From Theory to Practice**. 2nd: John Wiley e Sons, 2011. ISBN 978-0470660256. DOI: 10.1002/9780470978504.

SHADDAD, R. Q. *et al.* Mobility Management for Small Cells in 5G Ultra-Dense Wireless Network. In: 2022 2nd International Conference on Emerging Smart Technologies and Applications (eSmarTA). IEEE, 2022. p. 1–6. DOI: 10.1109/esmarta56775.2022.9935394.

SIDDQUI, M. U. A. *et al.* URLLC in Beyond 5G and 6G Networks: An Interference Management Perspective. **IEEE Access**, Institute of Electrical e Electronics Engineers (IEEE), v. 11, p. 54639–54663, 2023. ISSN 2169-3536. DOI: 10.1109/access.2023.3282363.

SINGH, U. *et al.* QoS Aware Uplink Scheduling for M2M Communication in LTE/LTE-A Network: A Game Theoretic Approach. **IEEE Transactions on Vehicular Technology**, v. 71, n. 4, p. 4156–4170, 2022. DOI: 10.1109/TVT.2021.3132535.

SOUZA, D. D. S. *et al.* A Novel Heuristic for Handover Priority in Mobile Heterogeneous Networks. **IEEE Access**, Institute of Electrical e Electronics Engineers (IEEE), v. 8, p. 4043–4050, 2020. ISSN 2169-3536. DOI: 10.1109/access.2019.2963069.

SOUZA LOPES, H. H. de; COELHO ROCHA, F. G.; TELES VIEIRA, F. H. Deep Reinforcement Learning Based Resource Allocation Approach for Wireless Networks Considering Network Slicing Paradigm. **Journal of Communication and Information Systems**, Sociedad Brasileira de Telecomunicacoes, v. 38, n. 1, p. 21–33, 2023. ISSN 1980-6604. DOI: 10.14209/jcis.2023.4.

SPADINGER, R. Nota Técnica - 2021 - Janeiro - Número 79 - Diset - Implementação da Tecnologia 5G no Contexto da Transformação Digital e Indústria 4.0. **Notas Técnicas**, Instituto de Pesquisa Economica Aplicada - IPEA, p. 1–29, 2021. DOI: 10.38116/ntdiset79.

TAHA, A.-E. M.; ALI, N. A.; HASSANEIN, H. S. **Lte, Lte-Advanced and Wimax: Towards Imt-Advanced Networks**. WILEY, 2011. 304 p. ISBN 0470745681. DOI: 10.1002/9781119970446. Disponível em: https://www.ebook.de/de/product/14532441/abd_elhamid_m_taha_najah_abu_ali_hossam_s_hassanein_lte_lte_advanced_and_wimax_towards_imt_advanced_networks.html.

TANG, F.; ZHOU, Y.; KATO, N. Deep Reinforcement Learning for Dynamic Uplink/Downlink Resource Allocation in High Mobility 5G HetNet. **IEEE Journal on Selected Areas in Communications**, Institute of Electrical e Electronics Engineers

(IEEE), v. 38, n. 12, p. 2773–2782, 2020. ISSN 1558-0008. DOI: 10.1109/jsac.2020.3005495.

TATA, C.; FELLAG, N.; KADOCH, M. New Courteous Algorithm for Uplink Scheduling in LTE-Advanced and 5G Networks. **Journal of Computer Networks and Communications**, Wiley, v. 2020, p. 1–15, 2020. ISSN 2090-715X. DOI: 10.1155/2020/4189789.

TAYYAB, M.; GELABERT, X.; JANTTI, R. A Survey on Handover Management: From LTE to NR. **IEEE Access**, Institute of Electrical e Electronics Engineers (IEEE), v. 7, p. 118907–118930, 2019. ISSN 2169-3536. DOI: 10.1109/access.2019.2937405.

TRABELSI, N. **A Game Theoretic Framework for User Association and Inter-cell Interference Management in LTE Cellular Networks**. 2016. Tese (Doutorado) – University of Avignon.

TRUJILLO, C. A. A. **Provisão de Qualidade de Serviço em Redes Integradas LTE-EAPON**. 2015. p. 89. Diss. (Mestrado) – Universidade Estadual de Campinas.

ULLAH, Y. *et al.* A Survey on Handover and Mobility Management in 5G HetNets: Current State, Challenges, and Future Directions. **Sensors**, MDPI AG, v. 23, n. 11, p. 5081, 2023. ISSN 1424-8220. DOI: 10.3390/s23115081.

WANGDONG, Q. *et al.* How smooth is smoothed round robin? In: INTERNATIONAL Conference on Communication Technology Proceedings, 2003. ICCT 2003. Beijing Univ. Posts & Telecommun. Press, 2003. v. 1. (ICCT-03), p. 421–428. DOI: 10.1109/icct.2003.1209112.

WU, H. Efficient Allocation of the Amount of Radio Resources in 5G NR to Efficient Allocation of Radio Resources in 5G NR. **IEEE Transactions on Wireless Communications**, Institute of Electrical e Electronics Engineers (IEEE), v. 21, n. 5, p. 3321–3332, 2022. ISSN 1558-2248. DOI: 10.1109/twc.2021.3120271.

XU, Y. *et al.* A Survey on Resource Allocation for 5G Heterogeneous Networks: Current Research, Future Trends, and Challenges. **IEEE Communications Surveys; Tutorials**, Institute of Electrical e Electronics Engineers (IEEE), v. 23, n. 2, p. 668–695, 2021. ISSN 2373-745X. DOI: 10.1109/comst.2021.3059896.

YAACOUB, E.; DAWY, Z. A comparison of uplink scheduling in OFDMA and SCFDMA. In: 2010 17th International Conference on Telecommunications. IEEE, 2010. p. 466–470. DOI: 10.1109/ictel.2010.5478799.

YAMAGISHI, K. *et al.* Derivation Procedure of Coefficients of Metadata-based Model for Adaptive Bitrate Streaming Services. **IEICE Transactions on Communications**, Institute of Electronics, Information e Communications Engineers (IEICE), 2021. DOI: 10.1587/transcom.2020cqp0002.

YUAN, P. **Resource allocation in heterogeneous networks using game theory**. 2015. Tese (Doutorado) – Nanyang Technological University. DOI: 10.32657/10356/65537.

ZAIDI, A. *et al.* Introduction: 5G Radio Access. In: 5G Physical Layer. Elsevier, 2018. p. 1–19. DOI: 10.1016/b978-0-12-814578-4.00006-0.

ZAVADSKAS, E. K.; TURSKIS, Z.; ANTUCHEVICIENE, J. Optimization of Weighted Aggregated Sum Product Assessment. **Electronics and Electrical Engineering**, Kaunas University of Technology (KTU), v. 122, n. 6, 2012. ISSN 1392-1215. DOI: 10.5755/j01.eee.122.6.1810.

ZAVADSKAS, E. K.; ANTUCHEVICIENE, J. *et al.* MCDM methods WASPAS and MULTIMOORA: Verification of robustness of methods when assessing alternative solutions. **Economic Computation and Economic Cybernetics Studies and Research**, n. 47, p. 5–20, 2013.

ZHANG, H.; CHU, X.; WEN, X. **4G Femtocells**. Springer New York, 2013. DOI: 10.1007/978-1-4614-9080-7.