



UNIVERSIDADE FEDERAL DE GOIÁS (UFG)
INSTITUTO DE INFORMÁTICA
DOUTORADO EM CIÊNCIA DA COMPUTAÇÃO

VICTOR HUGO LÁZARO LOPES

Soluções Baseadas em Aprendizado de Máquina para Alocação de Recursos em Redes Sem Fio de Próxima Geração

GOIÂNIA
2024

VICTOR HUGO LÁZARO LOPES

Soluções Baseadas em Aprendizado de Máquina para Alocação de Recursos em Redes Sem Fio de Próxima Geração

Tese apresentada ao Programa de Pós-Graduação do Instituto de Informática da Universidade Federal de Goiás (UFG), como requisito parcial para obtenção do título de Doutor em Ciência da Computação.

Área de concentração: Ciência da Computação.

Linha de pesquisa: Sistemas de Computação.

Orientador: Prof. Dr. Kleber Vieira Cardoso

Co-Orientador: Prof. Dr. Aldebaro Barreto da Rocha
Klautau Júnior

GOIÂNIA

2024

VICTOR HUGO LÁZARO LOPES

Soluções Baseadas em Aprendizado de Máquina para Alocação de Recursos em Redes Sem Fio de Próxima Geração

Tese defendida no Programa de Pós-Graduação do Instituto de Informática da Universidade Federal de Goiás (UFG) como requisito parcial para obtenção do título de Doutor em Ciência da Computação, aprovada em 06 de Maio de 2024, pela Banca Examinadora constituída pelos professores:

Prof. Dr. Kleber Vieira Cardoso
Instituto de Informática – UFG
Presidente da Banca

Prof. Dr. Aldebaro Barreto da Rocha Klautau Júnior
Faculdade de Engenharia da Computação e Telecomunicações – UFPA

Prof. Dr. Antonio Carlos de Oliveira Júnior
Instituto de Informática – UFG

Prof. Dr. Flávio Geraldo Coelho Rocha
Escola de Engenharia Elétrica, Mecânica e de Computação – UFG

Prof. Dr. Yuri Carvalho Barbosa Silva
Departamento de Engenharia de Teleinformática – UFC

Prof. Dr. José Ferreira de Rezende
COPPE – UFRJ

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador(a).

Victor Hugo Lázaro Lopes

Possui graduação em Sistemas de Informação pelo Instituto Federal de Educação, Ciência e Tecnologia de Goiás (2006) e mestrado em Engenharia Elétrica pela Universidade de Brasília (2015). Atualmente é professor efetivo do Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Campus Inhumas, Goiás. Tem experiência na área de Ciência da Computação e Engenharia de Software, com ênfase em Sistemas de Informação, Análise de Sistemas, Bancos de Dados, Algoritmos, Redes de Computadores e Inteligência Artificial. Atualmente possui como temas de interesse os problemas de alocação de recursos em redes sem fio de próxima geração e inteligência artificial/aprendizado de máquina aplicado às redes de acesso por rádio.

Dedico esse trabalho à minha esposa, Heliatrice Costa Menezes Lopes, e minha filha, Gabriela Costa Menezes Lopes, que muitas vezes se privaram de minha presença nesse período.

Agradecimentos

A Deus, por me conceder as condições necessárias para vencer essa etapa. À minha esposa e filha, por me darem apoio e carinho, em todos os momentos. Ao meu orientador, prof. Kleber Vieira Cardoso, e ao meu coorientador, prof. Aldebaro Barreto da Rocha Klautau Júnior, por acreditarem em meu potencial, pelos incentivos, paciência, oportunidades e pelas valorosas contribuições e ensinamentos, fundamentais para o desenvolvimento da pesquisa. À banca avaliadora, pelo compromisso com o avanço científico e acadêmico. Aos demais professores do curso, pelas valorosas disciplinas e ensinamentos. Ao IFG, pela oportunidade concedida através do afastamento. Por fim, aos amigos e colegas, importantes parceiros de pesquisa.

Resumo

Lopes, Victor Hugo Lázaro. **Soluções Baseadas em Aprendizado de Máquina para Alocação de Recursos em Redes Sem Fio de Próxima Geração**. GOIÂNIA, 2024. 169p. Tese de Doutorado. Doutorado em Ciência da Computação, Instituto de Informática, Universidade Federal de Goiás (UFG).

As Redes 5G e posteriores têm sido projetadas para dar suporte a serviços desafiadores. Mesmo diante de importantes avanços já introduzidos, os métodos de alocação e gerenciamento de recursos se mostram tarefas críticas nesse contexto. Embora os métodos de alocação de recursos baseados em otimização exata tenham uma longa história nas redes sem fio, diversos aspectos envolvidos nessa evolução exigem abordagens que consigam superar as limitações existentes. Trabalhos recentes têm mostrado o potencial de métodos de alocação de recursos baseados em AI/ML. Nessa abordagem, estratégias de alocação de recursos podem ser construídas com base no aprendizado, em que as relações complexas intrínsecas a tais problemas podem ser aprendidas por meio da experiência de agentes interagindo com o ambiente. Neste contexto, a presente tese buscou investigar as abordagens baseadas em AI/ML para o desenvolvimento de métodos dinâmicos de alocação e gerenciamento de recursos. Dois problemas relevantes foram considerados, sendo o primeiro relativo ao escalonamento de usuários e a alocação de recursos de rádio em redes MIMO multibandas, e o segundo voltado aos desafios de alocação de recursos de rádio, computacionais e de infraestrutura envolvidos no posicionamento de VNFs em vRAN desagregada. Para o primeiro problema, um agente baseado em DRL foi proposto. Para o segundo problema, duas abordagens foram propostas, sendo a primeira baseada em um método de otimização exata para a definição da solução de posicionamento de VNFs, e a segunda baseada em um agente DRL para o mesmo propósito. Componentes aderentes à arquitetura O-RAN foram propostos, criando o controle necessário para o monitoramento e a definição de novas soluções de posicionamento de forma dinâmica, considerando aspectos de cobertura e demanda das células. Simulações demonstraram a viabilidade das propostas, com importantes melhorias observadas em diferentes métricas.

Palavras-chave

Alocação de recursos, escalonamento baseado em DRL, aprendizado de máquina, vRAN desagregada, MIMO multibanda.

Abstract

Lopes, Victor Hugo Lázaro. **Machine Learning-Based Solutions for Resource Allocation in Next-Generation Wireless Networks**. GOIÂNIA, 2024. 169p. PhD. Thesis. Doutorado em Ciência da Computação, Instituto de Informática, Universidade Federal de Goiás (UFG).

5G and beyond networks have been designed to support challenging services. Despite important advances already introduced, resource allocation and management methods remain critical tasks in this context. Although resource allocation methods based on exact optimization have a long history in wireless networks, several aspects involved in this evolution require approaches that can overcome the existing limitations. Recent research has shown the potential of AI/ML-based resource allocation methods. In this approach, resource allocation strategies can be built based on learning, in which the complex relationships of these problems can be learned through the experience of agents interacting with the environment. In this context, this thesis aimed to investigate AI/ML-based approaches for the development of dynamic resource allocation and management methods. Two relevant problems were considered, the first one related to user scheduling and the allocation of radio resources in multiband MIMO networks, and the second one focused on the challenges of allocating radio, computational, and infrastructure resources involved in the VNF placement problem in disaggregated vRAN. For the first problem, an agent based on DRL was proposed. For the second problem, two approaches were proposed, the first one being based on an exact optimization method for defining the VNF placement solution, and the second one based on an DRL agent for the same purpose. Moreover, components adhering to the O-RAN architecture were proposed, creating the necessary control for monitoring and defining new placement solutions dynamically, considering aspects of cell coverage and demand. Simulations demonstrated the feasibility of the proposals, with important improvements observed in different metrics.

Keywords

Resource allocation, DRL-based scheduling, machine learning, disaggregated vRAN, multiband MIMO.

Sumário

Lista de Acrônimos	10
Lista de Figuras	18
Lista de Tabelas	20
1 Introdução	21
2 Referencial teórico	25
2.1 Alocação de recursos em redes móveis sem fio	25
2.1.1 Objetivos da alocação de recursos	26
2.1.2 Alocação e gerenciamento de recursos de rádio	28
2.1.3 Alocação e gerenciamento de recursos de infraestrutura	31
2.1.4 Visão geral sobre os métodos clássicos de alocação de recursos	32
2.1.5 Visão geral sobre os métodos de alocação de recursos baseados em aprendizado de máquina	36
2.2 Rede de acesso por rádio	39
2.2.1 Arquiteturas e suas evoluções	41
Arquitetura descentralizada	42
Arquitetura centralizada	43
Arquitetura virtualizada	45
Arquitetura aberta e desagregada	46
2.3 Soluções de prateleira para ML e RL	50
2.4 Simuladores	53
2.4.1 Simuladores em nível de link	54
Vienna 5G Link-Level	56
2.4.2 Simuladores em nível de sistema	58
Vienna 5G System-Level	59
Simu5G	61
2.5 Conclusão	63
3 Escalonamento de usuários baseado em DRL para MIMO massivo multibandas	64
3.1 Introdução	64
3.2 Definição do problema	68
3.3 O agente DRL proposto	73
3.4 Resultados e discussões	76
3.5 Conclusão	80

4	Alocação de recursos em vRAN desagregada levando em consideração aspectos da cobertura	85
4.1	Introdução	85
4.2	Abordagem baseada em otimização exata	87
4.2.1	O modelo do sistema	88
	Modelo para computar o consumo de recursos computacionais na vRAN	90
4.2.2	Formulação do problema	90
4.2.3	A solução proposta	92
4.2.4	Resultados e discussões	96
4.3	Abordagem baseada em otimização não exata com AI/ML	100
4.3.1	O modelo do sistema	101
4.3.2	A solução proposta	102
	MDP	103
	A solução DRL proposta	104
	Os elementos O-RAN propostos	105
4.3.3	Resultados e discussões	106
	Desempenho durante o treinamento	109
4.3.4	Desempenho durante a operação da vRAN	109
4.4	Conclusão	112
5	Conclusão	113
5.1	Resultados obtidos	115
5.2	Trabalhos futuros	116
	Referências Bibliográficas	118
6	Apêndice I	141
6.1	Métodos de otimização baseados em aprendizado de máquina	141
6.2	Otimização via aprendizado supervisionado versus aprendizado por reforço	143
6.2.1	Aprendizado supervisionado	144
6.2.2	Aprendizado por reforço	145
6.2.3	Aprendizado por reforço profundo	147
6.2.4	Abordagem <i>learning-to-optimize</i>	149
7	Apêndice II	153
7.1	Métodos clássicos de otimização	153
7.1.1	Otimização combinatória e multiobjetivo	154
7.1.2	Heurísticas e meta-heurísticas	158
7.1.3	Otimização estocástica	162
7.1.4	Otimização <i>online</i>	165

Lista de Acrônimos

2G segunda geração de rede móvel celular

3GPP *Third Generation Partnership Project*

4G quarta geração de rede móvel celular

5G quinta geração de rede móvel celular

6G sexta geração de rede móvel celular

A2C *Advantage Actor Critic*

A3C *Asynchronous Advantage Actor Critic*

ACER *Sample Efficient Actor-Critic with Experience Replay*

ACKTR *Actor Critic using Kronecker-Factored Trust Region*

ABC *Artificial Bee Colony*

AI *Artificial Intelligence*

AMC *Automatic Modulation Classification*

AP *Access Point*

AWGN *Additive White Gaussian Noise*

B5G *Beyond quinta geração de rede móvel celular (5G)*

BER *Bit Error Rate*

BBU *Base Band Unit*

BLER *Block Error Rate*

BW *bandwidth*

CAPEX *Capital Expenditures*

CBR *Constant Bitrate*

CN *Computing Node*

CO *Constrained Optimization*

CoMP *Coordinated Multi-Point*

COP *Combinatorial Optimization Problem*

CQI *Channel Quality Indicator*

CR *Cognitive Radio*

C-RAN *Centralized Radio Access Network (RAN)*

CRN *Cognitive Radio Network*

CS *Coordinated Scheduling*

CSI *Channel State Information*

CU *Central Unit*

D2D *Device-to-Device*

DBA *Deep Belief Architecture*

DDPG *Deep Deterministic Policy Gradient*

DL *Deep Learning*

DNN *Deep Neural Network*

DQN *Deep Q-Network*

DRC *Disaggregated RAN Combination*

DRL *Deep Reinforcement Learning*

DRL-SRA *DRL-based Scheduling and Resource Allocation*

DU *Distributed Unit*

D-RAN *Distributed RAN*

eICIC *enhanced Inter-Cell Interference Coordination*

eMBB *enhanced Mobile Broadband*

eNB *evolved Node B*

ERB *Estação Rádio Base*

FEC *Forward Error Correction*

FDD *Frequency-Division Duplexing*

GA *Genetic Algorithm*

GD *Gradient Descent*

gNB *next-generation Node B*

GOPS *giga operations per second*

HARQ *Hybrid Automatic Repeat reQuest*

HER *Hindsight Experience Replay*

HetNet *rede heterogênea*

HLS *High Layer Split*

ICI *Inter-Cell Interference*

ICIC *Inter-Cell Interference Coordination*

ILS *Iterated Local Search*

IoT *Internet das Coisas*

K-NN *K-Nearest Neighbors*

KPI *Key Performance Indicator*

L2O *Learning-to-Optimize*

LLS *Link-level Simulator*

LOS *Line Of Sight*

LTE *Long Term Evolution*

LTE-A *LTE Advanced*

MAC *Camada de Acesso ao Meio*

MBS *Macro Base Station*

MCS *Modulation Coding Scheme*

MDP *Markov Decision Process*

MEC *Mobile Edge Computing*

MF *macroscopic fading*

MH *Model Host*

MIP *Mixed Integer Programming*

MIMO *Multiple-Input Multiple-Output*

mmWave *Millimeter-Wave*

MNO *Mobile Network Operator*

ML *Machine Learning*

MLP *Multilayer Perceptron*

MOCOP *Multi-Objective Combinatorial Optimization Problem*

MOOP *Multi-Objective Optimization Problem*

MT *Maximum Throughput*

MU-MIMO *Multi-User Multiple-Input Multiple-Output (MIMO)*

near-RT RIC *near-real-time RAN Intelligent Controller (RIC)*

NIC *Network Interface Card*

non-RT RIC *non-real-time RIC*

NFV *Network Function Virtualization*

NR *New Radio*

NLOS *Non Line Of Sight*

OAI *OpenAirInterface*

O-CU *Open RAN (O-RAN) Central Unit (CU)*

O-Cloud *O-RAN cloud computing platform*

O-DU *O-RAN Distributed Unit (DU)*

O-eNB *O-RAN evolved Node B (eNB)*

OFDMA *multiplexação por divisão de frequências ortogonais*

OFDM *multiplexação por divisão de frequências ortogonais*

OPEX *Operational Expenditures*

O-RAN *Open RAN*

O-RU *O-RAN Radio Unit (RU)*

OTA *over-the-air*

PAH *Plan-At-Home*

PDCP *Packet Data Convergence Protocol*

PF *Proportional Fairness*

PHY *Camada Física*

PMI *Precoding Matrix Indicator*

PPO *Proximal Policy Optimization*

PRB *Physical Resource Block*

PSO *Particle Swarm Optimization*

QoE *Quality of Experience*

QoS *Quality of Service*

QSI *Queue State Information*

RAN *Radio Access Network*

RB *Resource Block*

RF *Radio Frequency*

RI *Rank Indicator*

RIC *RAN Intelligent Controller*

RIS *Reconfigurable Intelligent Surface*

RL *Reinforcement Learning*

RLC *Radio Link Control*

RNC *Radio Network Controller*

RO *Robust Optimization*

ROI *region of interest*

RR *Round Robin*

RRAM *Radio Resource Allocation and Management*

RRH *Remote Radio Head*

RRU *Remote Radio Unit*

RSRP *Reference Signal Received Power*

RU *Radio Unit*

SA *Stand-Alone*

SAC *Soft Actor Critic*

SB3 *Stable Baselines 3*

SBA *Service-Based Architecture*

SBS *Small Base Station*

SDN *Software-Defined Network*

SDR *Software-Defined Radio*

SE *Spectrum Efficiency*

SGD *Stochastic Gradient Descent (GD)*

SINR *Signal to Interference plus Noise Ratio*

SL *Supervised Learning*

SLS *System-level Simulator*

SMO *Service Management and Orchestration system*

SMOOP *Stochastic Multi-Objective Optimization Problem (MOOP)*

SNR *Signal to Noise Ratio*

SON *Self-Organizing network*

SOP *Stochastic Optimization Problems*

SRA *Scheduling and Resource Allocation*

SVM *Support Vector Machine*

TB *Transport Block*

TD3 *Twin Delayed DDPG*

TDD *Time-Division Duplexing*

TH *Training Host*

TRPO *Trust Region Policy Optimization*

TS *Time Slot*

TSP *Traveling Salesman Problem*

TTI *Transmission Time Interval*

UDN *Ultra Dense Network*

UE *User Equipment*

UL *uplink*

ULA *Uniform Linear Array*

UMi *Urban Micro*

UMTS *Universal Mobile Telecommunications System*

URLLC *Ultra Reliable Low Latency Communication*

USRP *Universal Software Radio Peripheral*

VANET *Vehicular Ad-hoc Network*

V2X *Vehicle-to-Everything Communication*

VCCS *Vienna Cellular Communications Simulators*

VNF *Virtualized Network Function*

VNF-C *VNF Controller*

VNF-P *VNF Placement*

vRAN *Virtualized Radio Access Network*

WMMSE *Weighted Minimum Mean Squared Error*

Lista de Figuras

2.1	Visão geral da arquitetura de uma rede celular 3G UMTS.	40
2.2	Visão geral da arquitetura de uma rede celular 5G NR <i>Stand-Alone</i> .	41
2.3	Visão geral da arquitetura de uma D-RAN.	43
2.4	Visão geral da arquitetura de uma rede C-RAN.	44
2.5	Visão geral da arquitetura de uma v-RAN.	45
2.6	Visão geral da arquitetura de uma v-RAN desagregada.	47
2.7	Visão lógica da arquitetura O-RAN, incluindo os seus componentes, interfaces e laços de controle.	48
2.8	Fluxo de trabalho do Vienna 5G LLS.	57
2.9	Fluxo de trabalho do Vienna 5G SLS.	60
2.10	A estrutura temporal do Vienna 5G SLS, em que o tempo de simulação é dividido em blocos, segmentos e TSs.	60
2.11	Visão geral sobre as capacidades do Simu5G.	62
2.12	Estrutura básica de uma NR NIC no Simu5G.	63
3.1	Escala temporal e organização do quadro: T_s , T_c e T_e são as durações de um <i>slot</i> , bloco e episódio, respectivamente.	69
3.2	SRA multibandas para MU-MIMO utilizando $F = 2$ frequências: f_1 e f_2 .	71
3.3	Exemplo ilustrando os usuários alocados de acordo com o modelo de alocação de recursos multiusuário e multibanda proposto, com $F = 2$, $K = 6$ e $BW_2 > BW_1$.	71
3.4	Interações entre a Estação Rádio Base (ERB) e o <i>User Equipment</i> (UE), com as setas azuis indicando o uso do canal de controle, e as setas vermelhas indicando a transmissão usando a portadora em f_c definida pelo escalonador, para o bloco m .	72
3.5	O agente DRL-SRA: as entradas são as informações de <i>Channel State Information</i> (CSI) e <i>Queue State Information</i> (QSI), e a saída são os usuários que vão ser servidos, por banda de frequências.	75
3.6	Convergência dos modelos em função do tempo de treinamento.	79
3.7	Comparando a vazão média por UE em diferentes cargas de rede: (a) ≈ 110 Mbps (b) ≈ 150 Mbps.	82
3.8	Comparando atraso médio de pacotes por UE nos <i>buffers</i> da ERB, em diferentes cargas de rede: (a) ≈ 110 Mbps (b) ≈ 150 Mbps.	83
3.9	Comparando a perda média de pacotes por UE em diferentes cargas de rede: (a) ≈ 110 Mbps (b) ≈ 150 Mbps.	84
4.1	Visão geral da topologia da vRAN desagregada considerada, demonstrando o domínio da topologia, o domínio das células e clusters criados.	89
4.2	DRCs que habilitam o CoMP.	90

4.3	A visão geral da solução proposta, com a inclusão dos componentes na arquitetura O-RAN: o rApp e o xApp.	94
4.4	Experimento 1: a eficiência espectral em função do total de clusters de ERBs, considerando os usuários de borda (em azul) e de centro (em verde).	98
4.5	Experimento 2: Vazão média por usuário ao longo do tempo enquanto os clusters de ERBs são reconfigurados dinamicamente.	99
4.6	A arquitetura O-RAN e os componentes propostos.	102
4.7	Visualização de exemplo da configuração da topologia implementada no Simu5G para os experimentos.	108
4.8	A variação de carga das ERBs. As linhas tracejadas representam as localizadas na zona suburbana e as linhas sólidas correspondem à zona urbana.	108
4.9	Evolução da recompensa durante a fase de treinamento.	109
4.10	Eficiência espectral em razão do número de UEs no decorrer de um dia.	110
4.11	Visão mais detalhada da eficiência espectral.	111
4.12	Demanda computacional nos CNs.	112
6.1	Exemplo de um sistema de <i>Reinforcement Learning</i> (RL), baseado em um MDP	146
6.2	Visão geral de uma DQN, adaptado de [Bengio, Lodi e Prouvost 2021].	148
7.1	Exemplo de um problema mono-objetivo.	155
7.2	Exemplo de um problema multiobjetivo.	156
7.3	Exemplo de um problema com 2 objetivos, e as suas possíveis soluções.	156
7.4	Exemplo de uma <i>Iterated Local Search</i> (ILS), iniciando com uma solução inicial s , que resulta na solução \hat{s} após uma busca local, em que se aplica uma perturbação que leva à solução s' , resultando na solução \hat{s}' após uma nova busca local.	161

Lista de Tabelas

2.1	Trabalhos relevantes que abordam a alocação de recursos com o emprego de métodos tradicionais.	35
2.2	Pesquisas relevantes que abordam a alocação de recursos com o emprego de métodos baseados em <i>Machine Learning</i> (ML).	39
2.3	Características das ferramentas de ML e RL.	51
2.4	Algoritmos disponíveis no <i>Stable Baselines 3</i> (SB3) e suas características de espaços de ações e de observação aceitos.	53
2.5	Simuladores em nível de link	55
3.1	Definições do estado do agente DRL, em que as últimas duas colunas indicam as dimensões e a frequência de atualização.	74
3.2	Parâmetros e configurações empregados nas simulações do DRL-SRA.	78
4.1	Elementos do modelo.	93
4.2	Parâmetros considerados nas simulações.	97
4.3	Visão geral do trabalho relacionado.	101
4.4	Legenda para índices na Tabela 4.3.	101
4.5	Parâmetros e configurações utilizados nas simulações.	107
6.1	<i>Surveys</i> recentes que abordam métodos de otimização com suporte de ferramentas de aprendizado de máquina.	143

Introdução

A sempre crescente busca por melhores e mais abrangentes serviços de redes móveis celulares tem guiado a evolução dessas redes. Em uma perspectiva histórica, coube aos avanços estabelecidos até a quarta geração de rede móvel celular (4G) garantir a massificação do acesso ao grande público, principalmente devido à constante melhoria dos serviços existentes, bem como a inclusão de novos serviços que marcaram cada uma das gerações. Mesmo diante de significativos avanços observados, as expectativas para as redes móveis de próxima geração, que incluem as redes 5G e *Beyond 5G* (B5G), apresentam grandes desafios à indústria e à academia. A expansão dos serviços, o suporte à conexão de dispositivos em massa, as comunicações ultraconfiáveis e de baixíssima latência, bem como as demandas por taxas de dados cada vez maiores são alguns dos requisitos desafiadores a serem atendidos.

No âmbito desses avanços, os métodos eficientes para o gerenciamento e alocação de recursos emergem como elementos críticos. A complexidade inerente à gestão dos recursos de rádio, largura de banda e controle de taxa de dados, aliada à crescente adoção do paradigma da virtualização e dependência de componentes baseados em software dessas arquiteturas, exige constantes inovações. Adicionalmente, a crescente densificação nessas redes, bem como a aguardada massificação de dispositivos conectados, trazem demanda para métodos de alocação e gerenciamento de recursos que priorizem o controle de interferências.

Tradicionalmente, técnicas de otimização formam a base para as soluções de alocação de recursos em redes de comunicação [Liang et al. 2019, Nguyen et al. 2021]. Abordagens exatas clássicas de otimização mono-objetivo e multiobjetivo ainda são amplamente utilizadas [Vijayalakshmi e Anandan 2019, Hao et al. 2020], enquanto abordagens estocásticas e *online* também começam a se consolidar [Bianchi et al. 2009, Lu et al. 2013, Duchi 2018, Mertikopoulos 2019, Alnabelsi, Salameh e Albataineh 2020, Wang, Magnússon e Johansson 2021]. Entretanto, as abordagens exatas de otimização são muitas vezes inadequadas devido à complexidade dos problemas [Liang et al. 2019, Wang et al. 2020], ou dependem de conhecimento especializado para a sua modelagem, implementação e adaptação. Ainda, a incapacidade de encontrar soluções para instâncias

grandes do problema, ou o alto custo computacional requerido para isso, são exemplos de limitações enfrentadas por esse tipo de abordagem.

Nesse contexto, abordagens baseadas em heurísticas e meta-heurísticas surgem como alternativas flexíveis e capazes de contornar as limitações inerentes aos métodos exatos de otimização. Elas demonstram habilidade em lidar com a complexidade dos problemas, proporcionando soluções viáveis mesmo quando a otimalidade não pode ser assegurada. No entanto, é necessário reconhecer que, apesar de sua versatilidade, essas abordagens apresentam desafios próprios que as limitam. Em muitos casos, as soluções obtidas podem não atender aos requisitos de eficiência exigidos, revelando limitações na capacidade de garantir resultados adequados. Em instâncias de grandes proporções do problema, o custo computacional associado à busca por soluções de alta qualidade não podem ser desprezados [Li e Pan 2015, Karimi-Mamaghan et al. 2022]. Além disso, a reusabilidade dessas estratégias em contextos semelhantes pode ser problemática, demandando grande esforço para customizações e adaptações.

Recentemente, técnicas de aprendizado de máquina (*Machine Learning* - ML) e aprendizado por reforço (*Reinforcement Learning* - RL) começaram a ser exploradas como abordagens não-exatas para problemas de otimização [Tian et al. 2020, Li, Zhang e Wang 2020, Mazyavkina et al. 2021, Nie, Li e Liu 2021, Bartsiokas et al. 2022]. Elas são similares às heurísticas e meta-heurísticas no que diz respeito à ausência de garantia de otimalidade. Por outro lado, esse tipo de abordagem é capaz de aprender sobre a estrutura do problema, incluindo aspectos que podem passar despercebidos por humanos, até mesmo especialistas. Isso pode trazer benefícios, tais como reutilização de abordagens, simplificação do processo de customização, manutenção da qualidade das soluções após o treinamento, dentre outras. Naturalmente, abordagens baseadas em ML/RL também trazem questões específicas, por exemplo, necessidade de grande volume de dados para aprendizado e ausência de arcabouços bem definidos para a sua adoção. No entanto, o potencial já demonstrado pelas técnicas de otimização baseadas em aprendizado de máquina e os notáveis avanços que a área de inteligência artificial (*Artificial Intelligence* - AI)/ML/RL tem apresentado desde o início do século motivam investigar sua capacidade de efetivamente fazer parte do conjunto de ferramentas adotadas para os problemas de alocação e gerenciamento de recursos.

Nesse sentido, o presente trabalho apresenta duas questões de pesquisa de interesse, que norteiam as investigações desempenhadas, sendo ambas relacionadas às redes 5G e posteriores:

1. Há benefícios em se investigar problemas de alocação e gerenciamento de recursos baseados em abordagens clássicas de otimização e usando estratégias baseadas em AI/ML?

2. É possível usar ferramentas de AI/ML de prateleira para a solução de problemas de alocação e gerenciamento de recursos?

Nesse contexto, o presente trabalho visa contribuir para a investigação sobre a efetividade de abordagens baseadas em AI/ML para resolução de problemas de otimização relacionados a alocação e gerenciamento de recursos em redes 5G e posteriores. Assim, as investigações foram desenvolvidas em duas fases. Na primeira fase, foi abordado o problema de escalonamento e alocação de recursos de rádio em redes MIMO com múltiplas bandas, considerando usuários em mobilidade. Na segunda fase, foram investigados os desafios de alocação de recursos de rádio, computação e de infraestrutura envolvidos no problema de posicionamento de funções de redes virtualizadas em redes de acesso desagregadas.

Para a primeira fase, um método baseado em *Deep Reinforcement Learning* (DRL) foi proposto. Para os testes e avaliações da proposta, um sistema foi criado, considerando tanto o sistema de comunicação, como o sistema RL. O sistema de comunicação foi criado considerando dados de canais gerados com o QuaDRiGA [Jaeckel et al. 2014, Burkhardt et al. 2014], um gerador de canais que considera um modelo de canal com geometria 3D em sistemas com múltiplas células. Assim, ao considerar canais com consistência nos domínios do tempo, do espaço e da frequência, o cenário se torna robusto e aderente aos ambientes 5G [Pang et al. 2022]. Outro fator relevante é o emprego de múltiplas bandas, considerando tanto a banda sub 6 GHz quanto a banda de ondas milimétricas. O sistema RL foi construído utilizando soluções de prateleira. Para o treinamento do agente proposto, um ambiente customizado foi desenvolvido, de forma a representar o espaço de observação, as ações a serem tomadas e a função de recompensa. Simulações realizadas permitiram avaliar a viabilidade da proposta. O desempenho do agente proposto superou o desempenho em diferentes métricas, considerando métodos clássicos do estado da arte, empregados como referencial comparativo.

Já na segunda fase, duas abordagens foram investigadas. Em ambos os casos, um importante objetivo almejado é a obtenção de soluções de posicionamento que agreguem determinadas divisões funcionais das pilhas de protocolos da *Virtualized Radio Access Network* (vRAN) em nós estratégicos da topologia, respeitando as limitações impostas (atraso máximo tolerado e as capacidades dos recursos computacionais e de infraestrutura envolvidos). Tal agregação cria *clusters* de ERBs em que um serviço de mitigação de interferências entre as células pode ser habilitado. Outro objetivo é a obtenção e implantação dinâmica de novas soluções de posicionamento, permitindo que a vRAN se adapte às condições de demanda e cobertura. Assim, busca-se otimizar o desempenho da vRAN, e em especial para os usuários de borda, normalmente experimentando piores condições de canais devido as interferências.

Nesse sentido, duas abordagens foram propostas, implementadas e avaliadas. Na primeira abordagem, o modelo do sistema é apresentado, seguido da proposição de um método baseado em otimização exata para a definição de novas soluções de posicionamento. A abordagem considera a proposição de componentes e laços de controle aderentes à arquitetura O-RAN, além de estabelecer dois importantes processos de monitoramento da vRAN para atingir os objetivos almejados. Na segunda abordagem, um método baseado em um agente DRL foi proposto, seguido da proposição de novos componentes a serem incluídos na arquitetura O-RAN. Em ambos os casos, simuladores de redes 5G foram empregados, permitindo criar ambientes integrados para o treinamento e a avaliação das propostas. Novamente, simulações demonstraram tanto a viabilidade das propostas, quanto os ganhos em termos de melhorias na eficiência espectral.

O restante do documento está organizado conforme descrito a seguir.

- No Capítulo 2, o referencial teórico é apresentado. Ele envolve a definição do problema de alocação e gerenciamento de recursos em redes móveis sem fio, estabelecendo os seus objetivos, bem como descrevendo os tipos de recursos considerados neste trabalho. Também apresenta os aspectos fundamentais das redes de acesso por rádio, as soluções de prateleira para ML e RL, concluindo com a apresentação dos simuladores empregados durante as investigações aqui descritas.
- O Capítulo 3 apresenta o primeiro problema considerado (primeira fase), que envolve o escalonamento de usuários e a alocação de recursos de rádio em redes MIMO multibandas. O problema é descrito e resolvido através de um agente DRL proposto, sendo avaliado e discutido.
- O Capítulo 4 apresenta o segundo problema considerado (segunda fase), envolvendo o posicionamento de *Virtualized Network Functions* (VNFs) em uma vRAN desagregada, considerando aspectos da cobertura das células. As duas abordagens empregadas são descritas e formuladas, sendo ambas avaliadas e discutidas.
- Por fim, o Capítulo 5 apresenta a conclusão sobre as pesquisas e investigações aqui descritas, e enumera as contribuições geradas como publicações. Além disso, alguns trabalhos futuros são sugeridos.

Referencial teórico

Esse capítulo apresenta a fundamentação teórica dentro do escopo da pesquisa. Iniciando pelos conceitos fundamentais envolvendo os problemas de alocação e gerenciamento de recursos em redes móveis sem fio e seguindo para as redes de acesso por rádio. Dessa forma, é possível obter uma visão sobre os elementos considerados nos modelos de sistema investigados durante as pesquisas. Tais conceitos também são úteis nas próprias formulações das soluções propostas em cada um dos problemas considerados. Além disso, também são apresentados detalhes importantes sobre as ferramentas de *Artificial Intelligence* (AI)/ML consideradas durante a construção e avaliação das soluções propostas, bem como o conjunto de ferramentas de simulação, fundamentais tanto para a obtenção de um melhor entendimento dos problemas de alocação aqui considerados, bem como servindo de base para a implementação e avaliação das soluções propostas.

2.1 Alocação de recursos em redes móveis sem fio

Os sistemas de comunicação sem fio precisam ser projetados para operar de forma a otimizar o uso dos recursos a eles destinados. Diversas técnicas e métodos de alocação e gerenciamento de recursos têm sido aplicados, propostos e continuamente melhorados neste sentido, como na adoção de diversos métodos de multiplexação nos domínios do tempo e da frequência, no emprego de arranjos celulares geograficamente separados para o reuso espectral, nas técnicas ortogonais e não ortogonais de acesso ao espectro, e no emprego de múltiplas antenas, por exemplo. Entretanto o desenvolvimento de métodos dinâmicos de alocação e gerenciamento dos recursos de rádio disponíveis se mostra essencial.

O surgimento de tecnologias como os rádios definidos por software (ou *Software-Defined Radios* - SDRs), as redes definidas por software (ou *Software-Defined Networks* - SDNs), fatiamento de recursos de rede e a virtualização das funções de rede (ou *Network Function Virtualization* - NFV), entre outros, permitiram a completa separação dos planos de dados e de controle. Embora essas tecnologias tragam importantes melhorias em termos de flexibilidade, escalabilidade e racionalização de custos, entre outros aspectos,

acabam por incluir os recursos computacionais e de infraestrutura dentro do escopo da alocação de recursos.

A transição das arquiteturas de redes monolíticas, construídas com componentes embarcados baseados em hardware não flexíveis e proprietários, para as arquiteturas flexíveis e ágeis baseadas em componentes de software e virtualização, já é uma realidade. Em arquiteturas como a *Service-Based Architecture* (SBA) proposta para o 5G *New Radio* (NR) [Mademann 2018], em que os componentes são facilmente reconfiguráveis, atingir os objetivos almejados para tais redes pode ser facilitado, dada a capacidade de implementar alterações, criando estruturas habilitadoras de diversos serviços e funcionalidades, que podem estar contidos em funções virtualizadas, por exemplo.

Entretanto, tais fatores tendem a gerar ainda mais complexidade nas tarefas envolvidas no projeto, implementação e na própria operação destas novas redes. Assim, se torna evidente que métodos dinâmicos de alocação e gerenciamento de recursos se fazem necessários [Alwarafy et al. 2021]. Desta forma, esta seção visa descrever tais recursos, apresentando uma visão ampla sobre os próprios desafios envolvidos nas tarefas de gerenciamento destes, a fim de se atingir a otimização dos sistemas de comunicação aqui considerados.

2.1.1 Objetivos da alocação de recursos

A alocação de recursos pode ser empregada para diferentes propósitos, que envolvem a melhoria de certos parâmetros da rede como um todo, de um grupo de serviços ou usuários específicos, ou até para usuários de forma individual. Tais melhorias podem ser vistos como objetivos da alocação de recursos, ou aspectos de projeto, tais como [Demirci e Sagiroglu 2019, Lee e Qin 2019, Wang et al. 2019]:

Maximização da eficiência espectral: Trata-se de um dos principais objetivos buscados por diversas tecnologias de comunicação sem fio, dada a natureza limitada dos recursos espectrais. Está diretamente relacionada com a otimização das taxas de dados, e se relaciona tanto com a eficiência energética quanto com o controle de interferências. Pode ser observada como o máximo de recursos espectrais que podem ser utilizados em uma determinada banda de frequências, nos domínios do tempo e do espaço, o que se traduz na taxa de dados que pode ser transmitida sobre uma determinada largura de banda. É a medida de quão eficientemente um recurso espectral limitado é utilizado pelas camadas mais baixas da pilha de protocolos.

Mitigação de interferências: Representa algo extremamente desafiador, principalmente ao se vislumbrar os cenários sobrepostos e densificados, como em *Ultra Dense Networks*

(UDNs), redes heterogêneas (HetNets) e em redes com acesso múltiplo não ortogonal [Xu et al. 2020]. O controle de interferências é o processo pelo qual um sistema visa operar sem degradar as comunicações que compartilham de um mesmo recurso espectral, podendo ser tando de uma mesma tecnologia ou operador, ou distintas. Também possui relação com a eficiência espectral e energética, de maneira geral. Entretanto, embora a eficiência energética costume se preocupar com transmissões em *downlink*, o controle de interferências precisa observar as duas vias. O controle de interferências em sistemas cientes de *Quality of Service* (QoS), como em [Liang et al. 2012] é um exemplo da complexidade envolvida, em que os *Physical Resource Blocks* (PRBs) precisam ser alocados, de forma a não quebrar os requisitos de QoS de seus usuários, enquanto busca minimizar as interferências co-canal e co-camadas. Também deve considerar as interferências entre símbolos, resultante das características de propagação dos sinais, e as interferências entre usuários [Eldeeb et al. 2020].

Maximização de vazão: Pode ser obtida pela melhoria da eficiência espectral e mitigação de interferências, que geram impactos diretos na otimização das capacidades dos canais. Entretanto, pode ter que depender de diversos outros objetivos, principalmente em cenários mais complexos. Por exemplo, pode ser formulado como um problema de controle de admissão e associação às células em HetNets [Khan et al. 2020], ou pela alocação conjunta de recursos computacionais, de comunicação e de energia [Deng et al. 2020], entre outros. Também sofre impacto das medidas de justiça nas alocações, de forma que existe um compromisso entre a maximização da vazão agregada da rede, que tende a ser injusta com usuários experimentando piores taxas de dados, e a maximização da vazão de forma ponderada entre os usuários da rede [Xu et al. 2020].

Otimização de conectividade e justiça: A ampliação da capacidade de conectividade é algo esperado, sob pena de impossibilitar a implementação massiva de diversas tecnologias previstas. Embora possa ser, até certo ponto, obtida pelo incremento de pontos de acesso, por exemplo, está cada vez mais evidente a dependência de métodos robustos de alocação de recursos, dado que esbarra em diversos outros elementos limitados, tais como as capacidades das redes de transporte, limitações energéticas e espectrais, entre outros [Hassan et al. 2020]. O conceito de pode assumir uma natureza relativa (justa para quem?). Entretanto, no contexto dos sistemas de comunicação sem fio, visa ampliar a oportunidade de acesso aos recursos de comunicação por seus usuários, mesmo que estejam em diferentes condições num determinado momento. Quando considerada na alocação de recursos de potência, por exemplo [Xu et al. 2020], a alocação da mesma potência a todos os usuários conectados não se mostra muito justa, dado que pode impor consequências na taxa efetiva de dados de usuários experimentando uma baixa *Signal to*

Noise Ratio (SNR). Neste sentido, uma alocação justa visa observar algum parâmetro que permita alocar a quantidade necessária de recursos.

Minimização de demanda computacional: A racionalização dos recursos computacionais envolvidos nas arquiteturas precisa ser devidamente considerada. A alocação de recursos deve considerar aspectos relativos à capacidade computacional (processamento e memória, por exemplo) de todas as entidades envolvidas, tais como servidores de borda e de núcleo, roteadores e comutadores, entre outros.

Adicionalmente, outros objetivos podem ser almejados. A redução de sobrecarga em ERBs visa minimizar a quantidade de usuários conectados em uma mesma rede de acesso. A minimização de cargas em enlaces pode permitir a escalabilidade dos serviços e a instanciação de VNFs de redes desagregadas, mesmo em topologias com poucos enlaces para a rede de transporte. Tais objetivos podem ser atingidos por métodos de escalonamento e associação de usuários, bem como pelo próprio projeto das topologias, de maneira otimizada [Lee e Qin 2019, Demirci e Sagioglu 2019, Moraes et al. 2021]. A maximização da confiabilidade é esperada para diversos serviços desafiadores, tanto *enhanced Mobile Broadband* (eMBB) quanto *Ultra Reliable Low Latency Communication* (URLLC), sem deixar de considerar as aplicações 6G, e podem sofrer impactos de diversos fatores, tal como pelos erros no CSI [Khan et al. 2020], ou por métodos de escalonamento ineficientes [Divya e Jayanthi 2020], por exemplo. Por fim, há também a minimização de atrasos, bem como a maximização de receitas e lucros, ou a redução de custos, almejados por operadoras e provedores [Zamzam, Elshabrawy e Ashour 2019, Demirci e Sagioglu 2019], entre outros.

2.1.2 Alocação e gerenciamento de recursos de rádio

De maneira geral, o conjunto de atividades envolvidas em *Radio Resource Allocation and Management* (RRAM) visa não somente garantir a operação dos serviços entregues pelas redes sem fio da maneira otimizada e mais eficiente, no que diz respeito ao uso dos recursos de rádio disponíveis, mas também podem estar relacionados à própria capacidade de oferta destes serviços. Na medida em que as redes sem fio vão se tornando cada vez mais heterogêneas e complexas, RRAM se mostra crucial, tanto em tarefas relativas ao planejamento e implementação de infraestruturas, quanto da própria operação destes [Alwarafy et al. 2021].

Soluções de RRAM se fazem indispensáveis para a solução de parte significativa dos requisitos rigorosos planejados, cumprindo papel fundamental na garantia do incremento da capacidade de suporte à conectividade massiva que estas redes demandam, além de buscar importantes avanços quanto a melhoria na eficiência espectral, energética e da própria operação destes sistemas. Adicionalmente, sistemas otimizados e eficientes pelo

emprego de métodos de alocação de recursos adequadamente projetados favorecem a melhoria de parâmetros de qualidade, tais como QoS e *Quality of Experience* (QoE) dos seus usuários.

Os recursos de rádio são aqueles envolvidos no processo de comunicação propriamente dito, e que estão diretamente relacionados com as responsabilidades das camadas mais baixas da pilha de protocolos, em especial a Camada Física (PHY) e a Camada de Acesso ao Meio (MAC). Também se incluem alguns processos inerentes à essas camadas, como a associação de usuários e controle de taxa de dados, por exemplo. De maneira geral, para atender aos objetivos almejados pelo emprego de métodos de alocação e gerenciamento de recursos, tais como os descritos na Seção 2.1.1, alguns recursos, ou *Elementos de Alocação de Recursos* [Ejaz et al. 2020], precisam ser administrados, sendo os principais recursos de rádio descritos a seguir.

Potência de transmissão

Os impactos das decisões que envolvem a alocação de potência de transmissão são cruciais nos sistemas de comunicação sem fio, especialmente em suas versões mais modernas, dado o compromisso que deve ser observado entre capacidade de transmissão e cobertura, e o controle de interferências. Embora a capacidade de transmissão de um dado *link* seja limitada pela potência alocada, o incremento indiscriminado desta, mesmo que possível, não necessariamente implica em melhorias crescentes de taxa de dados.

A potência máxima dos *Access Points* (APs) e ERBs pode ser tanto limitada pelas suas próprias capacidades energéticas, quanto por regulações das agências governamentais, de forma que o controle de potência se mostra essencial em questões envolvendo a eficiência energética, principalmente quando se considera o consumo energético de dispositivos alimentados por baterias, nos quais o incremento da potência de transmissão está diretamente relacionado com um maior consumo energético.

Desta forma, é comum que estes sistemas implementem restrições sobre os recursos de potência. Por outro lado, o controle de potência também se mostra indispensável na garantia da melhoria de parâmetros como QoS da rede e QoE dos seus usuários. À medida em que os sistemas de comunicação sem fio vão se tornando cada vez mais heterogêneos e densificados, os métodos de alocação da potência de transmissão são cada vez mais demandados para soluções adaptativas e dinâmicas, com capacidade de respostas às rápidas variações nos canais de propagação, nas condições da rede, bem como nas demandas flutuantes em tempo de execução.

Recursos espectrais e controle de acesso

Embora o espectro de radiofrequências seja fisicamente ilimitado, diversos desafios, como aqueles relacionados à propagação dos sinais eletromagnéticos, limitam tais recursos à poucas bandas úteis às aplicações de interesse comercial, industrial, científico

e governamental. Assim, os recursos espectrais disponíveis, também conhecidos como recursos de largura de banda, assumem uma natureza limitada. Os dispositivos de uma determinada rede sem fio precisam ter uma mínima garantia de acesso à estes recursos para iniciar a transmissão/recepção de dados. Desta forma, métodos de alocação de recursos precisam desempenhar papéis fundamentais nas decisões que envolvem o controle de acesso ao meio, bem como na atribuição dos recursos espectrais aos usuários envolvidos.

Associação de usuários e gerenciamento de *hand-off*

De maneira geral, a associação de usuários (ou seleção de célula) é o processo de definição de qual ERB ou AP um determinado UE deve estar associado a cada instante de tempo, de maneira a garantir as demandas de QoS e QoE na entrega dos serviços prestados. Portanto, observa-se que tal processo afeta significativamente a alocação de recursos, principalmente em cenários muito dinâmicos e heterogêneos. Levando-se em consideração a diversidade de dispositivos conectados, como nos casos de uso projetados para Internet das Coisas (IoT) massiva, por exemplo, e em redes com topologias baseadas em múltiplas RANs provendo o suporte à conectividade em múltiplas bandas, é fácil observar se tratar de uma tarefa de alocação com complexidade crescente. Adicionalmente, os métodos de alocação de recursos precisam lidar com a possibilidade de *hand-off* entre as múltiplas células [Hussain et al. 2020].

A obtenção das informações necessárias à tomada de decisão, em tempo real, tais como estatísticas dos canais, demanda e tráfego, ocupação das RANs e demandas de QoS, por exemplo, invariavelmente tendem a gerar significativos atrasos e sobrecargas na rede. Além disso, geraram dificuldades nos próprios processos de gerenciamento dos recursos de rádio, degradando a performance destes sistemas [Alwarafy et al. 2021].

Controle de taxa de dados

Tradicionalmente, a adaptação da taxa de dados é desempenhada pela camada PHY [Caponi, Chiti e Fantacci 2004, Mondal, Saha e Jang 2013]. Está diretamente relacionada ao cumprimento dos requisitos de vazão, tanto da rede quanto dos seus usuários, e geralmente deve ser balanceada entre a maximização da vazão agregada da rede e a maximização balanceada e justa entre todos os usuários conectados. Embora possa ser simplesmente relacionado aos processos envolvidos na alocação de subportadoras, como em sistemas baseados em multiplexação por divisão de frequências ortogonais (OFDMA) [Niyato e Hossain 2006], por exemplo, também se refere à definição de qual *Modulation Coding Scheme* (MCS) deve ser empregado, dependendo dos parâmetros de largura de banda.

Em redes sem fio baseadas em múltiplas RANs, como vistos em cenários de HetNets, por exemplo, métodos eficientes de RRAM que habilitem a agregação da taxa de dados, em cenários de sobreposição de RANs, se mostram importantes no

cumprimento dos requisitos de desempenho almejados. Tal desafio se torna ainda mais complexo quando considerada a dificuldade em obtenção de estatísticas dos canais de forma completa e atualizada.

2.1.3 Alocação e gerenciamento de recursos de infraestrutura

Sob o ponto de vista arquitetural de uma rede sem fio, os serviços por ela disponibilizados são dependentes de uma série de funções de rede. São os blocos funcionais que possuem comportamentos e propósitos bem definidos, estando tradicionalmente acoplados a dispositivos dedicados de hardware proprietários (*middleboxes*) dispersos na infraestrutura das redes [Yi et al. 2018, Mostafavi, Hakami e Sanaei 2021], tais como roteadores, comutadores e servidores. Em tais dispositivos, a implementação de políticas de alto nível exigem constantes reconfigurações de baixo nível, muitas vezes de forma manual [Demirci e Sagioglu 2019], o que acaba por ampliar a complexidade de operação e adaptação destes sistemas, impedindo a dinamicidade requerida pelos serviços a serem implementados nas redes sem fio de próxima geração. Adicionalmente, tais características também impactam negativamente na eficiência no uso dos recursos computacionais, bem como em termos de custos associados à aquisição e operação.

O emprego de arquiteturas de redes que ampliem tal flexibilidade é alvo de estudos há anos, tanto pela indústria quanto pela academia. Envolve a busca por redes adaptáveis, programáveis, de maior capacidade gerencial, eficientes em termos de custos e favoráveis à inovação. Assim, as arquiteturas com suporte à *Network Function Virtualization* (NFV) e *Software-Defined Network* (SDN), entre outras, surgem como promissoras para atingir tais objetivos.

Uma filosofia importante em SDN é a separação funcional entre os planos de controle e de dados, permitindo a abstração da lógica de controle do comportamento da rede pelos dispositivos que encaminham pacotes. Ao delegar tais tarefas à controladores baseados em software, habilitam as redes programáveis. Tais componentes de software possuem independência de plataforma de hardware, podendo ser instanciados em dispositivos de propósito geral, sendo um importante fator na facilitação de inovação [Demirci e Sagioglu 2019]. Adicionalmente, SDN favorece o gerenciamento da rede, principalmente devido a sua natureza centralizadora.

Apesar de não haver dependência com SDN, NFV compartilha alguns de seus objetivos, alcançados pela possibilidade da virtualização das funções de rede, que podem residir em diferentes dispositivos dispersos na topologia da rede. Podem ser, inclusive, instanciadas em dispositivos virtualizados, hospedados em servidores de propósito geral de alta capacidade. Em ambientes baseados em virtualização, as entidades de software e hardware são desacoplados, o que significa que as funções virtualizadas são independen-

tes dos elementos físicos. Os operadores podem trocar dinamicamente estes componentes de software para criar diferentes capacidades, inclusive com a possibilidade de instanciação de diferentes VNFs em um mesmo recurso físico [Xie et al. 2016]. A complexidade operacional, o consumo energético e o uso dos recursos computacionais, bem como os custos de *Capital Expenditures* (CAPEX), e *Operational Expenditures* (OPEX), podem ser minimizados, além das melhorias em termos de tempo e complexidade envolvidos na instanciação de atualizações.

Embora estas arquiteturas favoreçam a flexibilidade requerida pelos serviços com requisitos de desempenho cada vez mais desafiadores, cria-se um novo tipo de complexidade envolvida nos processos de tomada de decisão nestas redes. Embora não haja um consenso em termos destas tarefas serem classificadas dentro do escopo da alocação e gerenciamento de recursos, assume-se como equivalentes. Os desafios envolvem o aprimoramento dos sistemas de comunicação sem fio pela otimização do uso dos recursos envolvidos. Desta forma, destacam-se neste trabalho os desafios de posicionamento de VNFs em vRAN aberta e desagregada, e os respectivos recursos computacionais e de rede empregados.

De maneira geral, o problema de posicionamento de VNFs visa definir o posicionamento de determinadas funções virtualizadas, de forma a atender certos requisitos dos serviços de rede, de seus usuários e/ou da própria rede. As limitações impostas devem ser respeitadas, tais como aquelas que envolvem os recursos empregados e aquelas relativas à sensibilidade aos atrasos, por exemplo. Métodos de posicionamento de VNFs podem alocar funções de forma a se aproveitar dos caminhos existentes em cada fluxo [Li e Qian 2015], ou realizar as alocações definindo quais caminhos devem ser utilizados, dentre todos os caminhos disponíveis [Morais et al. 2021].

Destacam-se os seguintes recursos envolvidos [Demirci e Sagiroglu 2019]: i) servidores; ii) comutadores; e iii) enlaces de comunicação. Adicionalmente, abordagens recentes em problemas de posicionamento de VNFs em redes de acesso desagregadas consideram o fluxo de dados que fluem entre o núcleo da rede e as unidades de rádio, e vice-versa, como um recurso a ser manipulado pelo método de alocação, como em [Almeida et al. 2021], em que a quantidade de fluxo que passa por um determinado enlace pode ser particionado e encaminhado por outros que compõem a topologia.

2.1.4 Visão geral sobre os métodos clássicos de alocação de recursos

Diversas são as formas de classificação na literatura para os métodos de alocação e gerenciamento de recursos. Sob aspectos de como tais soluções se adaptam ao ambiente em que o sistema de comunicação sem fio está inserido, eles podem ser estáticos ou dinâmicos. Outros critérios podem ser empregados, tais como nas arquiteturas de decisão

em que são baseadas, podendo ser centralizadas ou distribuídas, de forma cooperativa ou sem cooperação, se tais decisões ocorrem de maneira instantânea ou ergótica, se há busca pela otimalidade ou se as soluções são subótimas, entre outras [Alwarafy et al. 2021].

Os métodos estáticos são baseados em conhecimento estatístico obtido *a priori*, não sendo possível, portanto, obter uma reação e adaptação mediante as alterações dinâmicas que a rede esteja experimentando, tais como o tráfego da rede e demandas flutuantes dos seus usuários, alterações das qualidades dos canais e demandas de QoS, por exemplo. Podem se mostrar adequados à otimização de soluções de alocação de recursos em que não ocorram alterações estruturais significativas. Assim, uma maior complexidade na busca pela solução adequada pode ser aceitável, como no projeto e implementação de infraestrutura física de redes celulares, no posicionamento de VNFs em vRANs, ou na definição de políticas fixas para a alocação de potência de transmissão e atribuição de canais, por exemplo. Entretanto, tendem a contribuir para a subutilização dos recursos, e na degradação de determinados parâmetros de qualidade da rede.

Os métodos dinâmicos, ou abordagens adaptativas, favorecem o ajuste dinâmico das políticas de alocação e gerenciamento de recursos, na medida em que conseguem obter acesso aos parâmetros da rede, rastreando as suas alterações no decorrer do tempo. São bastante empregados no projeto e operação de diversos sistemas sem fio modernos, mostrando bons resultados na redução da complexidade e do custo envolvidos no planejamento manual das redes, enquanto oferecem melhorias em termos de otimização do uso dos recursos empregados [Alwarafy et al. 2021]. Podem ser centralizados, no qual as entidades da rede, tais como APs, ERBs, *gateways* e servidores de borda, por exemplo, sendo controlados por alguma entidade central, enquanto outros podem assumir arquiteturas distribuídas.

Dada a capacidade de reação às mudanças que ocorrem nestas redes, os métodos dinâmicos se mostram bastante úteis em tarefas de alocação e gerenciamento de recursos. Podem ser empregados em tarefas de controle de potência, mitigação de interferências, alocação de canais e recursos de largura de banda, codificação e decodificação, associação e escalonamento de usuários, entre outros. Também podem ser empregados nas tarefas de alocação de recursos computacionais em redes baseadas em virtualização e fatiamento de recursos, por exemplo. De maneira geral, tais métodos são formulados como problemas de otimização complexos, que buscam maximizar ou minimizar alguma função objetivo, enquanto respeita certas restrições. Neste tipo de solução, as abordagens mais empregadas envolvem, habitualmente, otimização matemática, heurísticas e meta-heurísticas, teoria dos jogos e otimização estocástica. Estes métodos são apresentados e discutidos em mais detalhes no Seção 7.1.

Embora observa-se uma redução no volume de publicações nos últimos anos, há ainda um grande número de trabalhos relevantes sendo publicados, sendo que alguns dos

mais importantes trabalhos são sumarizados na Tabela 2.1, e discutidos a seguir.

Uma heurística para a alocação com foco em justiça em sistemas baseados em OFDMA sobrecarregados é proposta em [Letchford, Ni e Zhong 2020], sendo um cenário extremamente dinâmico, em que a chegada e saída de usuários é estocástica. Dado que os autores assumem que os métodos de otimização exata, por se tratar de um problema *NP*-difícil, bem como diversas heurísticas tendem a falhar ao apresentar suas soluções, a proposta envolve uma heurística de busca local dinâmica, que considera critérios de justiça em sua política. Embora as simulações apontam uma viabilidade do emprego em soluções de tempo real, o trabalho depende de algumas suposições que podem não se aplicar em situações reais, tais como o completo conhecimento estatístico do comportamento de chegada dos usuários, de suas solicitações e demandas, e também assumem cenários com potência do ruído invariante no tempo. Adicionalmente, como a métrica de justiça é baseada em uma média da chegada de usuários em um determinado momento, a proposta não permite observar o seu desempenho em situações em que tal média seja desviada, principalmente em situações de grande crescimento da chegada de usuários.

Os autores em [Khan et al. 2020] investigam a associação às células em HetNets, tanto para *uplink* quanto para *downlink*, formulado como um problema de maximização da vazão em termos de controle de admissão e associação dos usuários às diferentes células, considerando tanto a *Macro Base Station* (MBS) quanto as *Small Base Stations* (SBSs) envolvidas, bem como pela definição da potência a ser alocada. Entretanto, o modelo formulado a partir de um método de otimização baseado em busca exaustiva se mostra *NP*-difícil, de forma que a complexidade computacional cresce, exponencialmente, com relação ao incremento do número de usuários.

Em [Lu, Huang e Hua 2020] o consumo energético total de um sistema sensível ao atraso é minimizado, empregando métodos conjuntos de alocação de recursos e escalonamento de usuários móveis, em um sistema de múltiplas células. A proposta empregando métodos de otimização tradicionais, considerando tanto a otimização exata, quanto uma baseada em heurística. Assim, busca-se minimizar o consumo energético pela redução de retransmissões, bem como pela alocação de potência nestas múltiplas células, de forma que o método emprega a estimativa de qual célula o usuário estará no próximo instante de tempo. Entretanto, dado também se tratar de uma problema *NP*-difícil, os autores não demonstram como o método exato se comporta em uma instância do problema de maior complexidade. Adicionalmente, a heurística adotada busca reduzir o espaço de soluções viáveis assumindo que certos usuários podem não ter mais demandas de transmissão, deixando se ser considerados na política de escalonamento.

Em [Zhang, Liu e Zeng 2019] o posicionamento de VNFs em redes 5G de serviços customizados baseados em fatiamento de recursos é considerado, em que um método

heurístico é empregado. A proposta leva em consideração o conhecimento das interferências geradas por cada VNF rodando em paralelo em um mesmo recurso computacional, e busca maximizar a vazão total das requisições aceitas. Como consideram serviços de veículos autônomos e transmissão de vídeo em altíssima definição, trata-se, portanto, de aplicações sensíveis ao atraso, e de alta confiabilidade. Embora o método apresente um custo computacional aceitável, o tempo de otimização degrada com o incremento da topologia considerada. Desta forma, apesar do método permitir lidar com a variação dinâmica do fluxo de dados requerido pela rede, o tempo de otimização em cenários mais complexos pode impactar em sua adoção em sistemas reais.

Tabela 2.1: Trabalhos relevantes que abordam a alocação de recursos com o emprego de métodos tradicionais.

Ano	Referência	Recursos abordados	Cenário
2019	[Zhang, Liu e Zeng 2019]	Computacional	5G
2020	[Khan et al. 2020]	Rádio	HetNet
2020	[Letchford, Ni e Zhong 2020]	Rádio	OFDMA
2020	[Lu, Huang e Hua 2020]	Rádio	Celular

Embora estes métodos apresentem soluções viáveis à algumas aplicações nos sistemas atuais, observa-se que os serviços a serem disponibilizados pelas redes sem fio de próxima geração devem impor grandes dificuldades e limitações, tais como [Hussain et al. 2020, Alwarafy et al. 2021]:

- Exigência do conhecimento completo e atualizado de informações como CSI e QoS, por exemplo. Tais dados podem ser inviáveis de serem obtidos de forma adequada, sendo que os erros de estimação destes tendem a incrementar a margem de erro das próprias soluções obtidas. Em cenários de rápida variação das condições dos canais, a obtenção de CSI se torna ainda mais desafiador.
- Embora os métodos baseados em modelos, que visam capturar o conhecimento envolvido na representação do problema, e buscam as soluções adequadas com emprego de otimização matemática, sejam analiticamente tratáveis, nem sempre apresentam resultados precisos, devido às características complexas do ambiente de propagação e imperfeições de *hardware*, por exemplo.
- Crescimento do espaço de otimização com o incremento da complexidade dos modelos a serem otimizados, especialmente em problemas de natureza combinatória e multiobjetivo.
- Baixa capacidade de adaptabilidade em problemas que exigem decisões em tempo real, principalmente em sistemas que sofram alterações no decorrer do seu uso, como em casos da entrada e/ou saída de usuários conectados em uma determinada célula. Adicionalmente, as soluções obtidas por métodos baseados em heurísticas

são sensíveis à própria escolha dos parâmetros do sistema, no qual qualquer alteração possa exigir um novo ciclo do processo de otimização.

- Soluções tendem a ser computacionalmente caras, principalmente em termos de tempo de execução, incluindo significativos atrasos e sobrecargas ao sistema, impedindo a sua adoção por soluções que não sejam tolerantes aos atrasos. Além disso, a complexidade computacional é proporcional ao tamanho e arquitetura da rede, bem como ao número de usuários atendidos, aos parâmetros empregados e à quantidade de elementos que compõem a função objetivo e suas restrições.
- Os problemas suficientemente complexos são, normalmente, não convexos, levando a soluções baseadas em múltiplos ótimos locais, no qual a obtenção da solução ótima global não é trivial, ou não é viável em aplicações que exigem soluções em tempo real.
- Métodos baseados em teoria dos jogos podem não ser adequados à redes caracterizadas pela heterogeneidade massiva. Adicionalmente, tais soluções geram grande complexidade e demanda por informações, de forma proporcional ao número total de *players*, o que tende a ser proibitivo nestas futuras redes. Tais problemas também podem representar dificuldades em termos de escalabilidade e convergência.

Diante disso, soluções mais eficientes, e que possam ser, de fato, implementadas na prática são requeridas [Hussain et al. 2020], em que o emprego de métodos de otimização baseados em ML, e que possam ser dirigidos por dados, tem sido abordado em diversas soluções envolvendo os desafios de alocação e gerenciamento de recursos, nos quais os resultados se mostram promissores [Su et al. 2019, Ye, Li e Juang 2019, Noor-A-Rahim et al. 2020, Alwarafy et al. 2021, Xu et al. 2021]. Tais propostas são abordadas na Subseção 2.1.5.

2.1.5 Visão geral sobre os métodos de alocação de recursos baseados em aprendizado de máquina

O emprego de métodos baseados em ML visam, entre outros aspectos, combater parte das limitações enfrentadas pelos métodos tradicionais de alocação de recursos, em que pode-se destacar [Wang et al. 2016, Wang et al. 2019]: i) monitoramento eficaz do ambiente dinâmico e em tempo real de uma maneira eficiente; ii) aprendizagem contínua de novos conhecimentos advindos da dinamicidade do ambiente externo; iii) facilidades nas modelagens e simulação de ambientes complexos, com ou sem a dependência de modelos formais; iv) acesso facilitado a dados e informações importantes para a tomada de decisão; e v) promoção de avanços em redes heterogêneas pela facilitação da troca de informações entre as entidades envolvidas.

Diversos são os trabalhos recentes que formam o atual estado da arte envolvendo os métodos de alocação de recursos com suporte de ferramentas de aprendizado de máquina. Embora apresentem diferentes aspectos, e se baseiem em diferentes demandas de alocação de recursos, tais trabalhos corroboram a visão atual da presente pesquisa, e apresentam importantes instrumentos a serem futuramente considerados. A Tabela 2.2 apresenta uma visão geral sobre os trabalhos mais relevantes nesse contexto, sendo os mais importantes discutidos a seguir.

A alocação de recursos computacionais em redes 5G baseadas em fatiamento de recursos é abordada em [Han e Schotten 2020]. De forma um pouco mais aprofundada, os autores em [Su et al. 2019] apresentam uma visão dos desafios fim a fim, considerando os desafios desde a RAN até o núcleo da rede, também abordando os recursos de rádio bem como os de computação, com o foco no emprego racional dos recursos, levando a economias financeiras (CAPEX e OPEX). Além de apresentarem uma importante discussão sobre os desafios da alocação de recursos, quando considerados métodos matemáticos e baseados em teoria dos jogos, os autores apontam importantes direções futuras a serem desbravadas, tais como: i) melhoramentos na periodicidade das alocações de recursos, que tendem a ser postergadas quando baseadas em métodos de otimização tradicionais, ou são executadas em períodos muito curtos, representando aumento de consumo de recursos de processamento. Assim, métodos inteligentes podem produzir políticas de longa duração; ii) a multiplicidade de tecnologias de acesso por rádio, que devem coexistir, trará importantes complexidades; e iii) a alocação de recursos fim a fim precisa conseguir lidar, conjuntamente, com os recursos abstratos executados em contêiner virtualizado e os PRBs, por exemplo.

A alocação dinâmica de recursos com o suporte de métodos de RL em *Cognitive Radio Networks* (CRNs) é investigada em [Wang et al. 2019]. Tais redes visam o uso oportunístico de recursos espectrais por usuários secundários, sem gerar interferências significativas aos usuários primários. Assim, o emprego de métodos dinâmicos de alocação de recursos, que garantam a correta operação das funções destas redes, se mostra essencial. Os autores demonstram que as soluções RL baseadas em *Q-learning* têm sido uma das mais empregadas para a alocação de recursos nestas redes, principalmente em demandas de acesso dinâmico ao espectro. Observa-se que tais métodos também podem abranger questões de segurança inerentes às CRNs, em que usuários maliciosos se passam por usuários primários, por exemplo. Adicionalmente, o trabalho mapeia as soluções de alocação de recursos que consideram *deep Q-networks*, que empregam redes neurais profundas como função de aproximação.

Os autores em [Lin e Zhao 2020] abordam o gerenciamento de recursos em redes sem fio de próxima geração, tais como 5G, B5G e sexta geração de rede móvel celular (6G), abordando os recursos de comunicação (RAN, infraestrutura e espectro,

por exemplo), recursos de computação e de *cache*, de forma a dar suporte às aplicações previstas para estas redes.

O gerenciamento de recursos em redes 5G com suporte de NFVs e SDNs é abordado em [Shafik, Matinkhah e Sanda 2020], no qual são sumarizadas as propostas existentes que consideram o uso de métodos de ML em computação em névoa, de borda e em nuvem, entre outras. De forma parecida, os autores em [Zamzam, Elshabrawy e Ashour 2019] consideram o gerenciamento de recursos em *Mobile Edge Computing* (MEC) utilizando ML.

O provisionamento de QoS por intermédio de técnicas de gerenciamento de recursos na camada MAC, em que os processos decisórios são suportados por DRL, é abordado por [Abbasi et al. 2021], o que incluem os desafios de acesso ao meio, controle de taxa de dados, o escalonamento e compartilhamento de recursos, gerenciamento de *slices* e controle de *handover*.

Os autores em [Hussain et al. 2020] apresentam um completo e atualizado *survey* sobre o emprego de ML e *Deep Learning* (DL) nos desafios de alocação de recursos em redes celulares como suporte aos dispositivos de IoT, mapeando os potenciais, as soluções vigentes e os desafios pendentes. As limitações das técnicas de alocação de recursos nas diferentes arquiteturas IoT são consideradas, de forma a justificar o emprego de métodos baseados em dados.

A complexidade envolvendo o gerenciamento de recursos em redes 5G HetNets é abordada em [Lee e Qin 2019], com foco no emprego de DRL. De forma parecida, os autores em [Xu et al. 2021] abordam o uso de métodos dinâmicos de alocação de recursos em que o compartilhamento espectral pode ser otimizado, enquanto a interferência intercamadas é reduzida. Os autores apontam na direção da necessidade urgente de estruturação de métodos de alocação de recursos baseados em aprendizado, principalmente pelo rápido incremento da complexidade aguardada. Assim, sugerem o emprego de ML na obtenção de CSI, bem como na definição da potência de transmissão ótima.

Adicionalmente, outras pesquisas abrangentes são de grande importância para a compreensão do estado da arte, tal como em [Ejaz et al. 2020], que aborda a alocação de recursos em *Centralized* RANs (C-RANs) para redes 5G, em [Noor-A-Rahim et al. 2020], que aborda os desafios envolvendo as redes veiculares, especialmente *Vehicle-to-Everything Communication* (V2X), bem como em [Jiang et al. 2021], no qual o desafio de prover transmissões de vídeo sobre *Vehicular Ad-hoc Networks* (VANETs) é investigado. Em [Marzouk, Barraca e Radwan 2020] a alocação de recursos com o objetivo de maximização da eficiência energética em redes 5G é sumarizado, assim como é realizado em [Sharma e Kumar 2021], mas com o foco em UDNs em redes 5G. O problema de escalonamento de usuários em redes 5G é tratado em [Mamode e Fowdur 2020]. Por fim, embora o foco empregado pelos autores

em [Demirci e Sagioglu 2019] seja os métodos tradicionais de otimização empregados no problema de posicionamento de NFVs em redes virtualizadas, algoritmos de ML também são considerados.

Tabela 2.2: Pesquisas relevantes que abordam a alocação de recursos com o emprego de métodos baseados em ML.

Ano	Referência	Recursos abordados	Cenário
2019	[Demirci e Sagioglu 2019]	Computacional	NFV
2019	[Lee e Qin 2019]	Rádio	HetNet
2019	[Su et al. 2019]	Radio e Computacional	5G
2019	[Wang et al. 2019]	Rádio	CR
2019	[Zamzam, Elshabrawy e Ashour 2019]	Computacional	CR
2020	[Ejaz et al. 2020]	Rádio	5G
2020	[Han e Schotten 2020]	Computacional	5G
2020	[Lin e Zhao 2020]	Rádio e Computacional	5G-6G
2020	[Shafik, Matinkhah e Sanda 2020]	Computacional	5G
2020	[Hussain et al. 2020]	Rádio	IoT
2020	[Noor-A-Rahim et al. 2020]	Rádio	V2X
2020	[Mamode e Fowdur 2020]	Rádio	5G
2020	[Marzouk, Barraca e Radwan 2020]	Rádio	5G
2021	[Abbasi et al. 2021]	Rádio	Celular
2021	[Sharma e Kumar 2021]	Rádio	5G-6G
2021	[Jiang et al. 2021]	Rádio	VANET
2021	[Xu et al. 2021]	Rádio	5G
2022	[Chaieb, Abdelkefi e Ajib 2022]	Rádio	5G-6G
2022	[Pang et al. 2022]	Rádio	IoT
2022	[Wang et al. 2022]	Rádio	V2X
2023	[Amiri et al. 2023]	Computacional	5G-6G
2023	[Ali e Jammal 2023]	Computacional	5G-6G

2.2 Rede de acesso por rádio

Sob o ponto de vista arquitetural, uma rede de comunicação sem fio é formada por componentes estrategicamente posicionados e interconectados através de elementos especializados, objetivando formar um sistema capaz de fornecer os serviços planejados. Por exemplo, a arquitetura de redes celulares *Universal Mobile Telecommunications System* (UMTS), apresentada na Figura 2.1, é composta de um núcleo (subdividido nos núcleos comutados por circuitos e por pacotes) [Johnson 2011], e a RAN. Essa, por sua vez, é composta de *Radio Network Controllers* (RNCs), das *Base Band Units* (BBUs) e estações rádio base (ERBs), compostas de *Remote Radio Units* (RRUs)/*Remote Radio Heads* (RRHs) e das antenas, responsáveis por emitir/receber os sinais em forma

de ondas eletromagnéticas que formam os sinais no ambiente de propagação. Assim, todos os componentes (incluindo o núcleo da rede), são interconectados através de interfaces especializadas (cabeadas ou sem fio), que formam a rede de transporte. De forma semelhante, a Figura 2.2 apresenta uma visão geral da arquitetura das redes 5G NR em sua versão *Stand-Alone* (SA), conforme descrito em [Cardoso et al. 2020]. Essa arquitetura representa, de fato, uma grande alteração estrutural, principalmente devido ao impacto da virtualização e implementação de componentes baseados em software, tanto no núcleo quanto na RAN, conforme será apresentado em maiores detalhes na Seção 2.2.1.

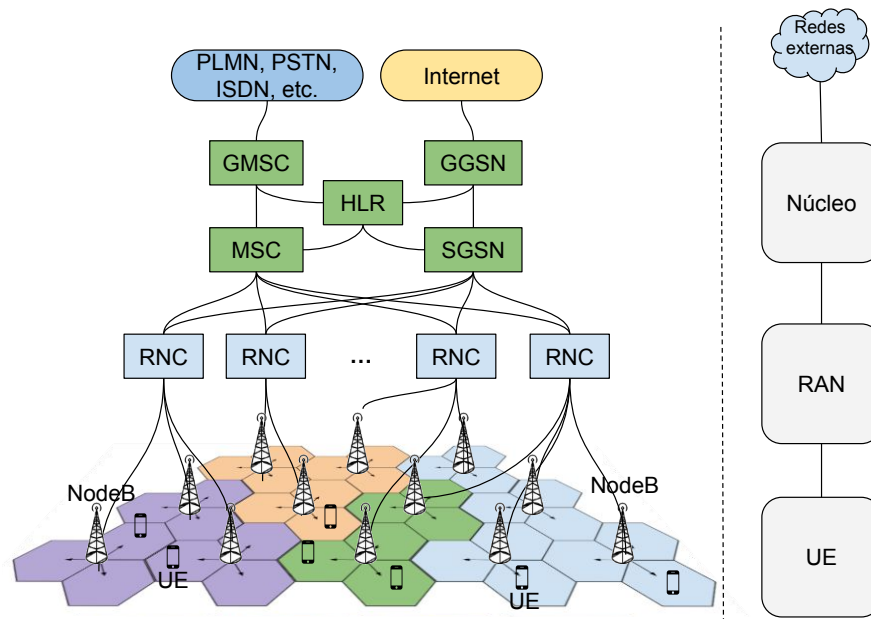


Figura 2.1: Visão geral da arquitetura de uma rede celular 3G UMTS.

A RAN é o componente de uma rede normalmente posicionado com maior proximidade dos usuários finais. Essencialmente, a RAN pode ser vista como um conjunto de componentes e protocolos que fornecem as tecnologias que habilitam a conectividade aos UEs, ao manter as interfaces aéreas da rede. Em outras palavras, cabe à RAN suportar as conexões de seus usuários para a entrega dos serviços da rede, gerenciando eficientemente os limitados recursos de rádio (e computacionais, nas versões mais modernas), realizando complexas tarefas e tomadas de decisão, tais como as envolvendo a alocação de recursos, escalonamento de usuários, balanceamento de carga, gerenciamento de mobilidade, entre outros [Habibi et al. 2019, Moraes et al. 2020].

Uma rede celular pode ser compreendida por um conjunto de áreas de cobertura de sinal sem fio de alcance e capacidade limitada denominadas células (Parte inferior da Figura 2.1). Em sua forma típica, uma célula é servida por pelo menos uma unidade de rádio, ou por um arranjo delas, formando setores (normalmente três setores por estação

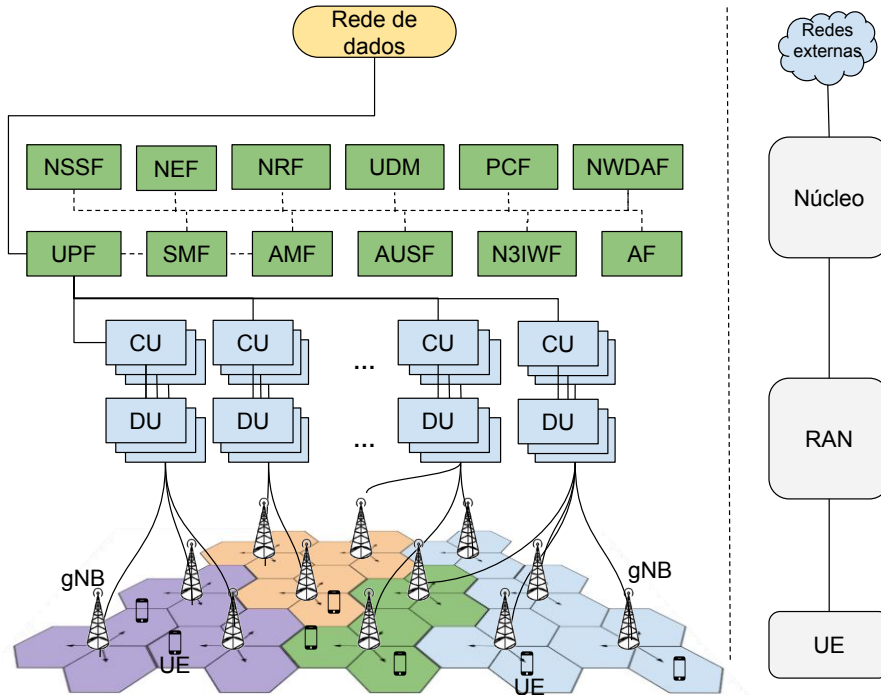


Figura 2.2: Visão geral da arquitetura de uma rede celular 5G NR *Stand-Alone*.

rádio base). Assim, uma RAN forma uma topologia, em que as células são organizadas de maneira a levar a cobertura de sinal a uma extensa área geográfica, fornecendo serviços de rede aos seus usuários finais. Observa-se, portanto, que a RAN cumpre uma série de papéis fundamentais para a própria operação da rede. Adicionalmente, o projeto, a implementação e a própria manutenção e operação das redes de acesso tornam-se tarefas complexas, refletindo em altos custos de implantação e operação (CAPEX e OPEX) [Costa-Pérez et al. 2013, Almeida et al. 2023, Zhao et al. 2023]. Nesse sentido, tanto a indústria quanto a academia têm direcionado esforços contínuos no aprimoramento da RAN desde a sua primeira implementação, esforços esses fundamentais para a própria evolução observada nas redes de comunicação sem fio, como será apresentado na Sec. 2.2.1.

2.2.1 Arquiteturas e suas evoluções

Sob um ponto de vista histórico, a RAN teve grandes avanços arquiteturais entre a primeira geração da telefonia celular até a proposição do 5G. Traçando um paralelo com a própria evolução dos serviços ofertados por essas redes, observa-se que tais avanços buscaram atender a tais evoluções, como na melhoria dos serviços de voz e inclusão de serviços de mensagens curtas de texto (SMS) na segunda geração (2G), a inclusão de serviços de dados e melhorias de cobertura de sinal na terceira geração (3G), a inclusão de banda larga móvel de alta velocidade e disponibilidade na quarta geração (4G) e a oferta dos serviços de banda larga móvel melhorada (eMBB), de comunicação ultra confiável

e de baixa latência (URLLC) e com suporte à conectividade massiva do tipo máquina (mMTC) na quinta geração (5G), entre outros aspectos. De fato, o núcleo também evoluiu nesse mesmo propósito.

Tais evoluções na RAN também podem ser observadas nas suas alterações arquiteturais ao longo do tempo. Várias abordagens arquiteturais foram desenvolvidas para atender às demandas crescentes por conectividade, eficiência, desempenho e suporte a novos serviços, em que a RAN migrou de uma arquitetura monolítica e estática, normalmente embarcada em uma solução *tudo em um* para arquitetura desagregada, fortemente guiada pelo paradigma da virtualização e nos conceitos de interfaces abertas, e independente de fornecedores. De maneira geral, pode-se diferenciar uma arquitetura RAN ao se distinguir em quais componentes as funções de banda base e de rádio são instanciadas, onde esses componentes estão localizados na topologia, e em como o fluxo de dados entre o núcleo e a RAN é estabelecido, como será descrito a seguir. Dentre as arquiteturas, destacam-se:

Arquitetura descentralizada

A RAN distribuída, ou *Distributed RAN* (D-RAN) é uma das primeiras arquiteturas propostas, empregada inicialmente na segunda geração de rede móvel celular (2G) [Brik, Boutiba e Ksentini 2022], em que tanto as funções de banda base quanto as funções de rádio encontram-se na própria ERB, próximo dos usuários. Assim, a ERB é responsável pelo processamento do sinal de banda base, pelo monitoramento e gerenciamento da rede de acesso, e por todas as funções de rádio, incluindo as funções de modulação, demodulação, filtros de sinais de rádio frequência, entre outros [Brik, Boutiba e Ksentini 2022]. Nas redes 3G e 4G as funções de banda base e de rádio foram separadas, sendo que as funções de rádio passam a ser implementadas na unidade de rádio (RRU ou RRH), que ficam posicionadas de maneira próxima (ou colocalizada) da ERB. Também é criada a unidade de banda base, ou BBU, responsável pelas funções de banda base. Tal arranjo se dá para cada ERB que compõe a RAN, de maneira que cada BBU provê os recursos necessários para a RRH associada, e utiliza um segmento da rede de transporte chamado *fronthaul* para interconectar os dois componentes, conforme demonstrado na Figura 2.3. Assim, em cada ERB, a unidade de rádio é atendida exclusivamente por uma BBU, sendo interconectadas pelo *fronthaul* [Brik, Boutiba e Ksentini 2022, Hossain et al. 2019]. Vale ressaltar que, embora a Figura 2.3 apresente duas células, cada uma servida por uma RRH e uma BBU, uma RAN pode ser composta de inúmeras células, onde múltiplos conjuntos de RRHs e BBUs podem ser instalados na ERB que atende à célula, a depender de aspectos como o número de setores e bandas de radiofrequência da célula, entre outros [Larsen et al. 2023].

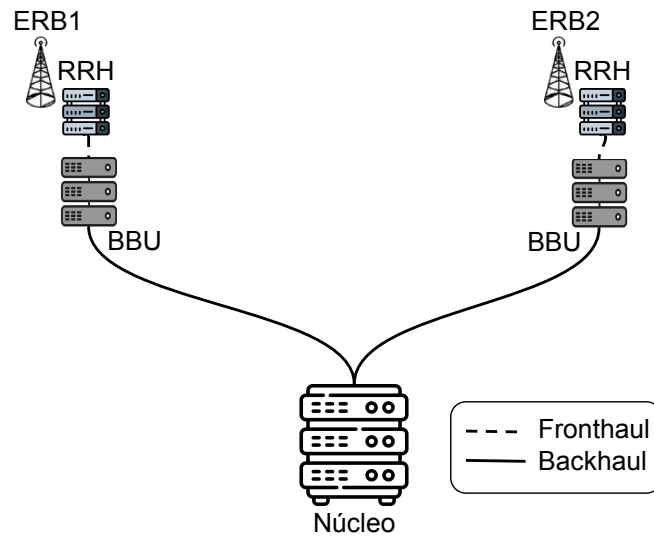


Figura 2.3: Visão geral da arquitetura de uma D-RAN.

Arquitetura centralizada

A RAN centralizada, também conhecida como *Centralized RAN (C-RAN)* [Brik, Boutiba e Ksentini 2022] ou *Cloud RAN* [Hossain et al. 2019]), é uma arquitetura que migra as funções de banda base para um conjunto centralizado e compartilhado de BBUs (*pool* de BBUs), enquanto mantém as unidades de rádio nas ERB. Nesse tipo de arquitetura, apresentado na Figura 2.4, o conjunto de BBUs pode ser posicionado distante das unidades de rádio, e integrados através de enlaces de alta capacidade (*fronthaul*). Assim, ocorre a centralização das BBUs de diferentes unidades de rádio, criando condições para uma melhor utilização dos recursos envolvidos, gerando melhores condições de gerenciamento, além de possibilitar reduções dos custos de implantação e operação da rede de acesso.

Tal arquitetura trouxe importantes avanços para as redes de acesso, principalmente devido ao emprego dos paradigmas da centralização e da virtualização. Em uma D-RAN, cada ERB é equipada com soluções embarcadas, normalmente monolíticas, o que representa um conjunto fixo de recursos computacionais e de rádio, que não podem ser compartilhados com outras ERBs da RAN. Assim, a RAN pode incorrer em gargalos pela falta de recursos suficientes em situações de alta demanda, ou em desperdício de recursos em baixa demanda [Tran e Pompili 2017]. Os autores em [Hossain et al. 2019] resumiram os benefícios associados ao uso de C-RAN, em que se destacam:

- Redução de custos, principalmente por fatores que levam a menores custos (e tempo) para implantação, instalação, adaptação e operação. Isso é possível pois os recursos computacionais são agregados em servidores de propósito geral, normalmente em nuvem.

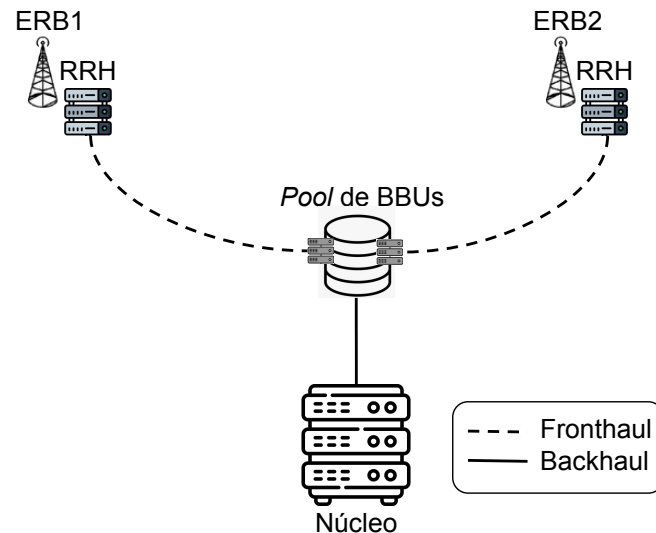


Figura 2.4: Visão geral da arquitetura de uma rede C-RAN.

- Melhorias de eficiência, tanto no que se refere à relação entre o consumo energético requerido para a operação da rede de acesso, quanto com relação ao próprio desempenho dela. Como as BBUs encontram-se instanciadas de forma agregada e compartilhada, é comum que a quantidade de BBUs requerida seja menor do que em uma D-RAN, levando a uma redução direta no consumo energético. Há, também, diversos custos energéticos indiretamente reduzidos, como no uso de refrigeração mais eficiente utilizado em servidores de grande porte, não disponíveis em uma D-RAN. Uma C-RAN também pode possibilitar melhorias na eficiência espectral, ao habilitar os métodos de controle e mitigação de interferências, como o *enhanced Inter-Cell Interference Coordination* (eICIC) e *Coordinated Multi-Point* (CoMP) na transmissão e/ou recepção, entre outros. Adicionalmente, a rede de acesso também pode favorecer aspectos como a redução de latência e aumento de vazão. Tarefas de *handover*, controle de falhas, retransmissões de pacotes e tarefas que requeiram de sincronização e cooperação entre ERBs diferentes, por exemplo, tornam-se facilitadas.
- A gestão e expansão são facilitadas, visto que a centralização em recursos compartilhados permite uma melhor escalabilidade, também favorecendo a capacidade de manutenção e implantação de melhorias, inclusive de hardware. Adicionalmente, a arquitetura favorece a inclusão de novos serviços e padrões.
- Suporte a adaptabilidade e flexibilidade, em que a RAN pode sofrer alterações para responder à demanda de um dado momento.

Entretanto, há algumas desvantagens relacionadas à arquitetura C-RAN. Ainda segundo o mapeamento realizado em [Hossain et al. 2019], eles envolvem questões de segurança e confiança, falhas e quebras de serviço comuns em ambientes em nuvem, e os

problemas relacionados a sobrecarga e demanda na rede de transporte, podendo incorrer em atrasos que inviabilizem certos serviços. Tais problemas serão discutidos em mais detalhes nas próximas seções.

Arquitetura virtualizada

A RAN virtualizada, ou vRAN, demonstrada na Figura 2.5, pode ser vista como um caso especial da C-RAN em nuvem, em que há o emprego mais aprofundado do paradigma da virtualização, principalmente no que diz respeito à virtualização das próprias funções de rede e das camadas da pilha de protocolos da RAN, isto é, as VNFs. Em outras palavras, graças ao emprego de conceitos como as redes definidas por software, ou SDNs, e o emprego de NFV em arquiteturas implantadas em servidores de propósito geral, compartilhados e em nuvem, as redes de acesso passam a ser formadas por recursos virtualizados [Azariah et al. 2022].

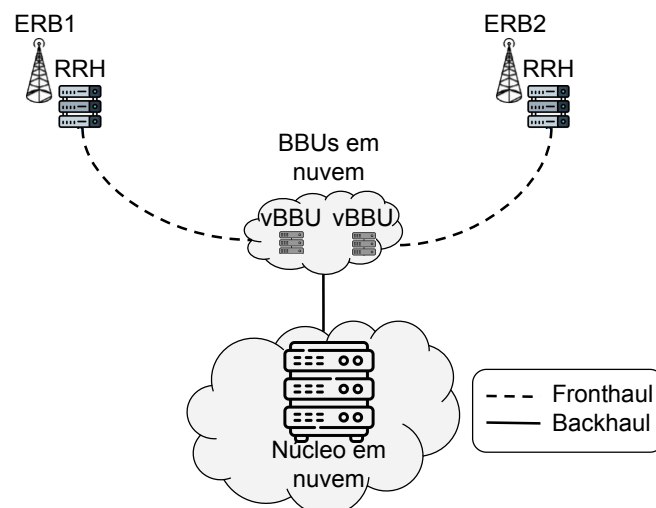


Figura 2.5: Visão geral da arquitetura de uma v-RAN.

Em uma RAN virtualizada, há o desacoplamento entre hardware e software [Larsen et al. 2023]. O conjunto de BBUs pode ser formado por vBBUs (*virtualized* BBUs), e as funções de rede são implementadas como VNFs. Os recursos virtualizados são implantados em servidores em nuvem como máquinas virtuais ou contêineres [Azariah et al. 2022]. Assim, os processos de gerenciamento e orquestração se tornam facilitados, dado que novos recursos virtuais podem ser disponibilizados, escalados ou desligados de forma dinâmica, levando à otimizações energéticas, de recursos e desempenho além de minimizar custos para o operador da rede. Por exemplo, um operador da rede pode escolher um conjunto adequado de hardware e/ou software que reduza o consumo energético da RAN [Larsen et al. 2023].

Entretanto, há algumas desvantagens a serem observadas. Como os recursos virtualizados e agregados em recursos computacionais compartilhados estão associadas a diferentes unidades de rádio, experimentando diferentes condições de demanda e carga, eles precisam ser gerenciados de maneira justa e eficiente, sem comprometer os requisitos dos serviços da rede, como o QoS [Azariah et al. 2022]. Adicionalmente, a vRAN depende de interfaces proprietárias, o que dificulta questões como a interoperabilidade e o uso de ambientes de hardware e software de múltiplos fornecedores.

Arquitetura aberta e desagregada

Como observado a respeito das desvantagens ainda existentes tanto na D-RAN quanto na v-RAN, uma nova arquitetura tem sido proposta: a RAN aberta. Trata-se de uma transição atualmente em curso nas arquiteturas de redes de acesso que busca estabelecer a capacidade de utilização de componentes não proprietários, vindos de diferentes fornecedores, ao adotar interfaces abertas que permitem suporte a serviços inteligentes e flexíveis, enquanto objetiva melhorias de desempenho e redução de custos de implantação e operação, entre outros aspectos [Zhang, Zhou e Erol-Kantarci 2022]. Adicionalmente, uma RAN aberta unifica as vantagens vindas da virtualização, e cria recursos que tanto suprem as dificuldades vistas nas outras arquiteturas, quanto abrem as portas para importantes avanços tecnológicos. Assim, a RAN aberta pode ser parametrizada e programada de forma flexível, empregando funções virtualizadas, interoperáveis e inteligentes.

Um outro conceito que surge nessa arquitetura é o da desagregação dos seus componentes. Assim, as funções da RAN, antes alocadas entre a RRH e a BBU, passam a ser instanciadas como VNFs em até três nós: a unidade central (ou *central unit* - CU), a unidade distribuída (ou *distributed unit* - DU) e a unidade de rádio (ou *radio unit* - RU). Uma visão genérica de uma RAN aberta e desagregada, ou NG-RAN (do inglês *Next Generation RAN*) [Morais et al. 2021, Kak et al. 2022] é apresentada na Figura 2.6. O caminho entre o núcleo da rede e as unidades de rádio é realizado através da rede de transporte, composta de três segmentos: o *backhaul*, o *midhaul* e o *fronthaul*. Tal arquitetura é fruto dos esforços de diversos grupos de pesquisa e padronização (3GPP, ITU e ETSI), bem como de alianças corporativas (O-RAN Alliance, TIP e GSMA) [Almeida et al. 2022].

Em uma vRAN desagregada a pilha de protocolos é composta por divisões funcionais, instanciadas nos diferentes nós da vRAN. Isso permite que os *Mobile Network Operators* (MNOs) escolham, para cada unidade de rádio, qual será a divisão funcional, em quais nós cada parte da divisão funcional selecionada será instanciada, e quais serão os caminhos utilizados para criar o fluxo de dados necessário, dentre os caminhos disponíveis. Tais escolhas precisam ser realizadas respeitando-se uma série de restrições, tais como o atraso máximo tolerado entre as partes da divisão funcional, largura de

banda disponível nos segmentos da rede de transporte e a capacidade computacional dos servidores em que os nós estejam rodando, entre outros.

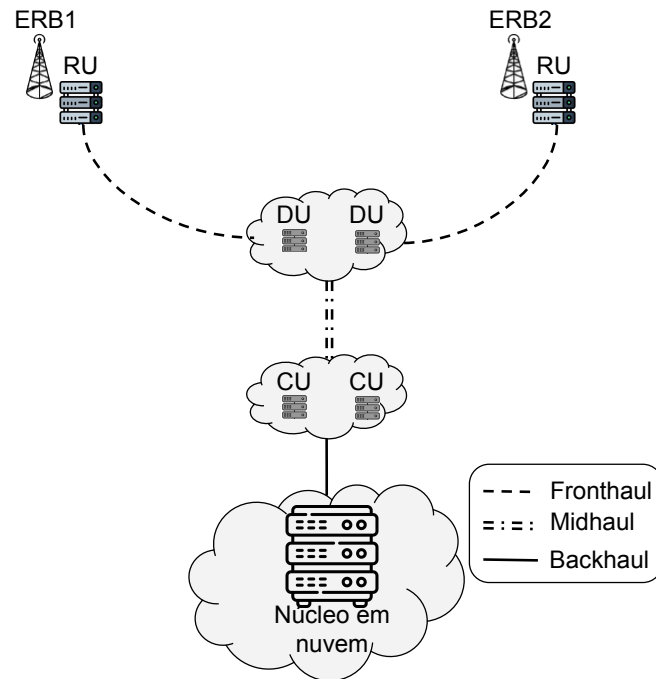


Figura 2.6: Visão geral da arquitetura de uma v-RAN desagregada.

Uma das mais proeminentes propostas para arquiteturas abertas e desagregadas é a O-RAN. A sua proponente, a *O-RAN Alliance*, é uma organização industrial fundada em 2018 com o objetivo de propor uma arquitetura a partir do LTE e do 5G NR [Polese et al. 2023], que deveria considerar todos os propósitos de interfaces abertas, virtualização, desagregação e controles inteligentes dirigidos por dados. Uma visão da arquitetura O-RAN é apresentada na Figura 2.7 (adaptado de [O-RAN Alliance 2023]), que apresenta os principais componentes, interfaces e a organização lógica.

Um dos principais componentes da arquitetura O-RAN é o *Service Management and Orchestration system* (SMO), que é baseado nos princípios da SBA. A arquitetura SBA apresenta as funções/papéis tanto do fornecedor quanto do consumidor de serviços, juntamente com interfaces padronizadas orientadas a serviços. Assim, as interfaces habilitam a interoperabilidade, de forma que os componentes da arquitetura sejam independentes da implementação. O SMO é responsável por tudo relativo ao domínio do gerenciamento da RAN, e a descrição da arquitetura é focada nos serviços que o SMO deve prover para suportar a RAN. Os objetivos primários do SMO envolvem [O-RAN Alliance 2023]: a) Prover interfaces para as funções de rede da O-RAN; b) Implementar e dar suporte às funções de otimização da RAN; e c) Prover gerenciamento da O-Cloud, orquestração e gerenciamento do fluxo de operação.

Uma importante inovação apresentada na arquitetura O-RAN é o uso de controladores inteligentes da RAN (RICs), que introduzem componentes programáveis e inteligentes para a orquestração da RAN, operando em loops fechados de controle com diferentes escalas de tempo [O-RAN Alliance 2023]. Para tal, os RICs permitem a inclusão de softwares com diferentes tipos de acesso a dados da RAN, agregando diversos KPMs (*Key Performance Measurements*), dados de carga e demanda, estado da infraestrutura, bem como dados adicionais de contexto de fontes externas à RAN [Polese et al. 2023]. Assim, aplicações dirigidas por dados (incluindo inteligência artificial e aprendizado de máquina) para o gerenciamento e otimização da RAN se tornam viáveis.

Mais precisamente, a arquitetura contém dois RICs, sendo o RIC de tempo não real (*Non-Real Time RIC*, ou *Non-RT*), e o RIC de tempo quase real (*Near-Real Time RIC*, ou *Near-RT*), sendo que o *Non-RT RIC* é membro do *framework SMO*, e recebe suporte e interfaces fornecidas pelo *Non-RT RIC framework*, como ilustrado na Figura 2.7.

O *Non-RT RIC* permite a inclusão de aplicações especializadas (*rApps*), estendendo as funcionalidades da RAN, em problemas de tempo não real (1 segundo ou mais). Através de uma interface aberta especializada (A1), ele também pode fornecer dados, orientações e gerenciamento de modelos de ML para o *Near-RT RIC*. Assim, entre outros aspectos, o *Non-RT RIC* pode gerar recursos que guiem as próprias operações e decisões do SMO, podendo até gerar impactos em diversos componentes da RAN, impactando até os dispositivos conectados [Polese et al. 2023].

Por sua vez, o *Near-RT RIC* opera em tarefas com menor escala de tempo (entre 10 ms e 1 segundo), sendo responsável por tarefas de controle e otimização dos componentes da RAN. Através da interface E2, possui acesso a dados em baixa granularidade vindos tanto da O-CU quanto da O-DU [O-RAN Alliance 2023]. Esse RIC permite a inclusão de aplicações de terceiros (*xApps*), enquanto garante-lhes acesso a dados em escala de tempo quase tempo real vindos da interface E2, podendo ser dados da célula, quanto em nível de UEs individuais. Assim, um *xApp* pode prover serviços de alto valor agregado. O *Near-RT RIC* pode fornecer informações importantes sobre a RAN para componentes externos, expondo-os em uma interface aberta (Y1). É importante ressaltar que a O-RAN também prevê loops de controle para tarefas em escala de tempo, como nas tarefas associadas às camadas mais baixas da pilha de protocolos, além das tarefas exclusivas da unidade de rádio. Entretanto, não há um RIC para tal escala de tempo.

A O-Cloud serve para suportar toda a infraestrutura virtualizada da O-RAN. Isto é, trata-se da plataforma física de computação em nuvem, que envolve todo o conjunto de nós que compõem a RAN (O-CU, O-DU e O-RU), as funções de rede que implementam o SMO e os RICs, todo o suporte necessário para os componentes de software envolvidos, tais como sistemas operacionais, máquinas virtuais, bancos de dados, entre outros, e as funções para gerenciamento e orquestração da RAN.

Por fim, a arquitetura proposta considera a O-eNB, que é a formalização da ERB compatível com as interfaces estabelecidas, bem como as interfaces padronizadas (3GPP). Assim, há o suporte tanto para a eNB (LTE) quanto para a ng-eNB (5G NR). Em ambos os casos, a O-eNB suporta a integração com as funções de rede instanciadas nos nós da RAN, e suporta a conexão entre O-DU e O-RU através da interface *Open Fronthaul* [O-RAN Alliance 2020, O-RAN Alliance 2023].

2.3 Soluções de prateleira para ML e RL

Em projetos de pesquisa científica em que há o emprego de recursos de inteligência artificial (IA), tais como aprendizado de máquina (ML) ou aprendizado por reforço (RL), uma decisão relevante que deve ser tomada envolve optar pela construção dos algoritmos necessários ou pela adoção de algoritmos, ferramentas e soluções oferecidos por projetos e/ou empresas, as quais chamamos de *soluções de prateleira* neste trabalho. De um lado, o emprego de algoritmos personalizados permite obter um maior controle sobre o processo de desenvolvimento e da capacidade de criar modelos mais aderentes ao problema. Entretanto, a construção de algoritmos próprios exige uma compreensão mais profunda dos fundamentos teóricos, gerando uma camada extra de complexidade quando se está buscando resolver problemas igualmente complexos. Do outro lado, o emprego de soluções prontas para o uso permite manter o foco no problema que está sendo resolvido, permite uma rápida implementação, prototipação, testes e aproveitamento de recursos já otimizados para problemas parecidos, ou de uma mesma classe de problemas. Além disso, apresenta uma boa robustez, dado que é desenvolvido e mantido por uma comunidade especializada, sendo também testado e avaliado por diversas pessoas.

Para os problemas tratados no decorrer da tese assumiu-se o uso dos algoritmos de prateleira, diante dos benefícios expostos. Parte fundamental do que levou à escolha das bibliotecas está localizada na própria natureza dos problemas aqui considerados, que podem ser modelados como processos de decisão de Markov (MDP). De maneira geral, um MDP fornece a estruturação necessária para a representação de problemas dinâmicos nos quais as ações tomadas por um agente, a partir de um determinado estado observado do ambiente, afetam as transições para os estados futuros. Assim, a escolha é direcionada às bibliotecas otimizadas para algoritmos de aprendizado por reforço, incluindo suporte a DRL.

Existem diversas bibliotecas disponíveis para ML, RL e DRL, tais como *Scikit-learn* [Pedregosa et al. 2011], *OpenAI Baselines* [Brockman et al. 2016], *TF-Agents* [Guadarrama et al. 2018] e *Stable-Baselines* [Hill et al. 2018, Raffin et al. 2021], em que as principais características são sumarizadas na Tabela 2.3, e discutidas a seguir.

Tabela 2.3: Características das ferramentas de ML e RL.

	Scikit-Learn	OpenAI Baselines	TF-Agents	Stable-Baselines
Integração com TensorFlow	N	S	S	S
Algoritmos Clássicos	S	S	S	S
Facilidade de Uso	S	S	N	S
Suporte	S	-	S	S
Documentação	S	-	S	S
Suporte à RL/DRL	N	S	S	S

Iniciado em 2007 como um projeto de um programa *online* para o desenvolvimento de software livre, o Scikit-learn tornou-se um módulo Python que contém uma ampla gama de algoritmos clássicos de aprendizado de máquina para problemas de aprendizado supervisionado e não supervisionado, tais como *Support Vector Machine* (SVM), *Random Forest*, *K-Nearest Neighbors* (K-NN), entre outros. Por representar uma biblioteca independente, ele não depende do TensorFlow. Por ser popular e relevante, trata-se de um projeto ativamente mantido, o que resulta em um bom suporte da comunidade, e de uma boa documentação. Entretanto, por decisões de projeto, não há suporte para RL/DRL. Assim, o seu uso pode ficar restrito a problemas de classificação, regressão e clusterização.

De iniciativa da empresa OpenAI¹, o OpenAI Baselines surge para fornecer implementações de algoritmos modernos de RL e DRL compatíveis com os ambientes OpenAI Gym² [Brockman et al. 2016] e Gymnasium³ [Towers et al. 2023]. Construído como um pacote Python, permite o uso do TensorFlow como *backend* para os algoritmos. Oferece facilidade para a implementação algoritmos novos (também desenvolvidos como pacotes Python), e já fornece diversos algoritmos clássicos, como *Advantage Actor Critic* (A2C), *Sample Efficient Actor-Critic with Experience Replay* (ACER), *Actor Critic using Kronecker-Factored Trust Region* (ACKTR), *Deep Q-Network* (DQN), *Proximal Policy Optimization* (PPO) e *Trust Region Policy Optimization* (TRPO). Embora seja um projeto legado, sem relevantes atualizações nos últimos anos, e sem suporte ativo, trouxe

¹<https://openai.com/>

²<https://github.com/openai/gym>

³<https://github.com/Farama-Foundation/Gymnasium>

facilidades para a comunidade de pesquisa nas tarefas de implementar, replicar, testar, refinar e comparar desempenhos perante diferentes algoritmos.

Desde que a empresa Google apresentou a sua iniciativa que resultou na biblioteca TensorFlow⁴, viu-se uma importante revolução na forma como soluções para problemas de aprendizado de máquina e aprendizado profundo são implementados e testados. Projetado para oferecer flexibilidade e desempenho, o TensorFlow se tornou uma plataforma completa com ênfase em redes neurais, comportando uma arquitetura versátil para a construção de modelos complexos, o que inclui suporte a problemas de classificação, regressão, visão computacional e aprendizado por reforço. Um outro benefício é a sua integração perfeita com hardware especializado, como GPUs e TPUs, otimizando o processo de treinamento e execução dos modelos. Originário desse mesmo projeto, os agentes TensorFlow (ou TF-Agents) foram propostos. Trata-se de uma biblioteca de código aberto desenvolvida como suporte a pesquisas avançadas envolvendo aprendizado por reforço e problemas clássicos de *Bandido Multiarmado* [Mahajan e Tenenetzis 2008]. Projetado como uma extensão nativa para o TensorFlow, o TF-Agents se destaca pela flexibilidade e escalabilidade, trazendo facilidades para a implementação e testes de novos algoritmos com a sua estrutura modular. A biblioteca já oferece alguns algoritmos clássicos, como DDPG, SAC, PPO, entre outros. Também possui comunidade forte, com boa documentação e suporte, o que inclui tutoriais e exemplos diversos.

Outra importante opção é o Stable-Baselines, fruto de esforços da comunidade de aprendizado por reforço, iniciado a partir do OpenAI Baselines. Isto é, também possuem suporte aos ambientes OpenAI Gym e Gymnasium, e uso de TensorFlow como *backend*. Segundo consta no próprio site da versão 2 do Stable-Baselines⁵, ocorreram alterações estruturais e de código com relação ao OpenAI Baselines, em que se destacam: a) estrutura unificada para todos os algoritmos; b) funções e classes documentadas; c) estilo de código unificado (PEP8); d) maior estrutura para testes e cobertura de código; e d) novos algoritmos. Possui uma documentação abrangente, e conta com exemplos prontos para uso que facilitam a implementação e experimentação. Com essa ferramenta, o pesquisador precisa apenas construir o ambiente RL para o treinamento (ambiente customizado), e escolher o algoritmo adequado ao problema. Adicionalmente, o Stable-Baselines também permite construir o próprio algoritmo RL, pois fornece uma classe base para tal. Os algoritmos também podem contar com suporte a multiprocessamento, em que múltiplos agentes podem ser paralelizados para otimizar o tempo de treinamento. A sua terceira versão, o SB3 passa a ter suporte ao PyTorch e Tensorboard, e uma maior integração com o *framework* RL Baselines3 Zoo, que provê uma coleção de agentes pré-

⁴<https://www.tensorflow.org/>

⁵<https://stable-baselines.readthedocs.io/en/master/>

Tabela 2.4: Algoritmos disponíveis no SB3 e suas características de espaços de ações e de observação aceitos.

	Espaço de ações		Espaço de observação		Multiprocessamento
	Discreto	Contínuo	Discreto	Contínuo	
A2C	Sim	Sim	Sim	Sim	Sim
DDPG	Não	Sim	Sim	Sim	Sim
DQN	Sim	Não	Sim	Sim	Sim
HER	Sim	Sim	Sim	Sim	Sim
PPO	Sim	Sim	Sim	Sim	Sim
SAC	Não	Sim	Sim	Sim	Sim
TD3	Não	Sim	Sim	Sim	Sim

treinados, além de ferramentas robustas para o treinamento e avaliação de modelos, bem como para a otimização de hiperparâmetros utilizados pelos algoritmos. A Tabela 2.4 apresenta os algoritmos disponíveis no SB3, bem como a relação dos tipos de espaço de ações e de observação.

2.4 Simuladores

Na área de redes e telecomunicações o desenvolvimento científico e tecnológico requer, em muitas das vezes, a experimentação e validação em ambientes minimamente aderentes aos cenários reais em que os artefatos oriundos desses esforços de pesquisa e desenvolvimento irão ser implementados. Entretanto, isso nem sempre é viável, por variados fatores. Isso se torna ainda mais significativo quando se trata do desenvolvimento de novas tecnologias para as redes 5G e B5G, em que a realização de experimentos reais pode exigir investimentos substanciais [Le et al. 2021], sem que haja garantia de avaliações em todos os possíveis cenários.

Como forma de suprir tais dificuldades, uma opção é o emprego de plataformas de testes (*testbeds*). Tratam-se de ambientes simulados que oferecem uma plataforma controlada e configurável para a execução de testes e validações. Buscam replicar as condições operacionais de redes específicas, permitindo a observação de cenários próximos à realidade, quando possível. Segundo [Esmaily e Kralevska 2021], as *testbeds* para redes móveis de próxima geração surgem como protótipos oriundos de pesquisas, principalmente na forma de implementações de componentes individuais da arquitetura. Entretanto, as *testbeds* têm evoluído no sentido de suportar conjuntos maiores de elementos de arquiteturas de redes móveis, incluindo elementos da RAN, do núcleo e da rede de transporte.

São diversos exemplos de *testbeds* atualmente descritas na literatura, tais como *5G4IoT* [Dzogovic et al. 2018], *5GTN* [Piri et al. 2016], *5G Tactile Inter-*

net Platform [Mekikis et al. 2019], *Mosaic5G* [Nikaein, Chang e Alexandris 2018], *Orion* [Foukas, Marina e Kontovasilis 2017] e *POSENS* [Garcia-Aviles et al. 2018], entre outros [Esmaily e Kravetska 2021]. Essas plataformas possuem em comum o emprego dos paradigmas NFV e SDN, além de considerarem o uso de ferramentas não proprietárias, de código aberto, e oriundas de projetos robustos e difundidos na comunidade de pesquisa, tais como *Openstack*⁶, *OpenAirInterface* (OAI)⁷ e *Open5GS*⁸. Entretanto, há diversos casos em que a escalabilidade e o custo tornam-se problemas significativos na adoção de tais plataformas. Por exemplo, em projetos que dependam da simulação de uma RAN com uma topologia de maior escala, o uso extensivo de ERBs emuladas em rádios definidos por software do tipo *Universal Software Radio Peripherals* (USRPs) (comum em muitas dessas plataformas) pode não ser viável.

Nesse sentido, uma outra possibilidade viável e eficaz é o emprego de simuladores de redes. Eles podem proporcionar ambientes virtuais que superam as limitações das *testbeds*, oferecem a flexibilidade necessária para modelar redes de grande porte, replicando cenários diversificados, que modelam com fidelidade aspectos de propagação dos sinais sem fio, e permitindo avaliações em maior escala. Além disso, simuladores proporcionam ambientes mais controlados para experimentações, eliminando as restrições físicas e financeiras que muitas vezes afetam a implementação em *testbeds* reais. Os simuladores se mostram essenciais na obtenção de métricas de performance, tais como a vazão potencial em *uplink* e *downlink*, as médias de *Bit Error Rate* (BER), *Signal to Interference plus Noise Ratio* (SINR) e *Channel Quality Indicator* (CQI), dados de usuários ativos, além de estatísticas importantes, como as probabilidades de *outage* e *handover*, taxas de transmissão e retransmissões, perdas, atrasos e latência [Gkonis, Trakadas e Kaklamani 2020]. Os simuladores são descritos em maiores detalhes a seguir.

2.4.1 Simuladores em nível de link

Uma das classes de simuladores de redes são os simuladores em nível de link, ou *Link-level Simulators* (LLSs). Eles se concentram em oferecer mecanismos detalhados para as camadas física e de enlace, sendo projetados para analisar e otimizar a comunicação ponto a ponto. Assim, levam em consideração fatores como qualidade de sinal, interferências e as características específicas do canal de comunicação e da propagação dos sinais. Segundo [Pratschner et al. 2018], as medições, análises e simulações em nível

⁶<https://www.openstack.org/>

⁷<https://openairinterface.org/>

⁸<https://open5gs.org/>

Tabela 2.5: Simuladores em nível de link

Simulador	MATLAB comms toolbox	SiMoNe	Aff3ct	Vienna 5G LLS
Linguagem	Matlab	C#	C++	Matlab
Licença	Comercial	Projeto particular	Uso acadêmico	Uso acadêmico
Código fonte	Fechado	Fechado	Aberto	Aberto

de link são fundamentais para o desenvolvimento de novos mecanismos para os sistemas de comunicação sem fio.

O uso de simuladores em nível de link se torna fundamental, pois permite obter subsídios para a construção das arquiteturas dos transceptores de rádio e painéis de antenas, para o próprio desenvolvimento dos padrões para sistemas sem fio [Pratschner et al. 2018] e para o desenvolvimento de componentes de outras camadas, por exemplo. Nesse sentido, tais simuladores devem possuir uma acurada representação do canal, estando apto a representar fielmente todos os componentes e mecanismos de propagação [Gkonis, Trakadas e Kaklamani 2020]. Um exemplo relevante envolve o uso do simulador *NYUSIM*⁹ em pesquisas com foco nas comunicações na banda de ondas milimétricas (*mmWave*) [Rappaport, Sun e Shafi 2017, Sun et al. 2018, Budalal e Islam 2023], e na investigação de cenários de aplicação 5G [Varshney et al. 2023].

Além do simulador *NYUSIM*, existem outros simuladores em nível de link descritos na literatura. A Tab. 2.5 apresenta as características dos principais simuladores, segundo mapeamento apresentado em [Eckhardt et al. 2022]. O pacote de ferramentas para comunicações do MATLAB (*MATLAB communications toolbox*)¹⁰ é um coleção de recursos para o design e verificação de sistemas de comunicações através de simulações fim a fim. Ele inclui mecanismos tanto para o uso de modulações e formas de onda já definidas em padrões conhecidos, quanto customizadas. Trata-se de uma poderosa ferramenta, e permite tanto a interação com outros módulos MATLAB, como o módulo de antenas, radiofrequência e de processamento paralelo (otimização de simulações). Também facilita a interação com *Software-Defined Radios* (SDRs) para testes *over-the-air* (OTA). O Aff3ct [Cassagne et al. 2019] pode ser visto como uma biblioteca de baixo nível, que oferece um conjunto de ferramentas dedicadas à codificação de canal, ou *Forward Error Correction* (FEC), com uma grande quantidade de códigos FEC e diversos tipos de decodificadores. Entretanto, o seu uso completo como LLS depende de implementações adicionais [Eckhardt et al. 2022]. O SiMoNE (*Simulator for Mobile Networks*) [Eckhardt et al. 2022] trata-se de um simulador LLS com foco na banda THz. Entretanto, trata-se de um projeto fechado, estando fora do escopo da Tese. Por fim, temos o simulador Vienna 5G Link-Level, descrito em mais detalhes a seguir.

⁹<https://wireless.engineering.nyu.edu/nyusim/>

¹⁰<https://www.mathworks.com/products/communications.html>

Vienna 5G Link-Level

O simulador Vienna 5G Link-Level [Pratschner et al. 2018] surge como uma evolução da sua versão LLS para *Long Term Evolution* (LTE). Ele incorpora-se ao conjunto de simuladores *Vienna 5G*¹¹, que por sua vez trata-se de um membro da suíte de simuladores *Vienna Cellular Communications Simulators* (VCCS), desenvolvida pelo grupo de pesquisa em comunicações móveis do instituto de tecnologia da Universidade de Viena, Áustria. Esse grupo de pesquisa tem uma longa história na produção desse tipo de simulador, com o início do desenvolvimento do simulador para LTE ainda em 2009. Assim, o novo simulador surge para estender as capacidades do simulador voltado para 4G de forma a comportar as especificidades da camada PHY do 5G.

O escopo do Vienna 5G LLS é a simulação da comunicação ponto a ponto de toda a cadeia envolvida entre o transmissor e o receptor: codificação de canal, processamento MIMO, modulação com multiportadoras, estimação de canal e equalização [Pratschner et al. 2018]. Dessa forma, não implementa modelo genérico para simular a transmissão de bits, e sim modelos de transmissão realistas e aderentes aos padrões tanto do LTE quanto do 5G NR, com possibilidade de uso de multiplexação por divisão de frequências ortogonais (OFDM), como WiMAX ou similares, por exemplo. Ao permitir a simulação tanto dos transmissores quanto dos receptores, encontra limites de escala apenas no que se refere aos recursos computacionais disponíveis.

Trata-se de um programa implementado em Matlab, com código fonte disponível sob licença de uso acadêmico, o que facilita a sua difusão e uso pela comunidade científica. A Figura 2.8 apresenta uma visão sobre o fluxo de trabalho de uma simulação, adaptado de [Pratschner et al. 2018]. O primeiro passo é carregar o arquivo que contém todas as informações necessárias para a simulação. Assim, tanto a topologia quanto o próprio experimento pode ser estabelecido. O simulador oferece uma série de cenários prontos, que podem ser facilmente adaptados para diferentes propósitos. Ao definir os nós da topologia, os usuários e os enlaces, todos os parâmetros da camada PHY são estabelecidos, incluindo a codificação do canal, a modulação, os parâmetros MIMO, a geração e estimação de canal, o processo de reporte e cálculo de CSI, o modelo de canal a ser considerado, a forma de onda das multiportadoras, e os equalizadores e decodificadores do receptor. Uma importante capacidade do simulador é permitir a coexistência de múltiplas tecnologias, de forma que os nós e enlaces possam ser definidos com parâmetros diferentes, apresentando uma capacidade acurada de modelagem de interferências [Pratschner et al. 2018].

Ainda segundo o processo demonstrado na Figura 2.8, assim que os parâmetros são definidos, e a topologia é criada, a simulação é realizada de maneira quadro por

¹¹<https://www.tuwien.at/etit/tc/en/vienna-simulators/vienna-5g-simulators/>

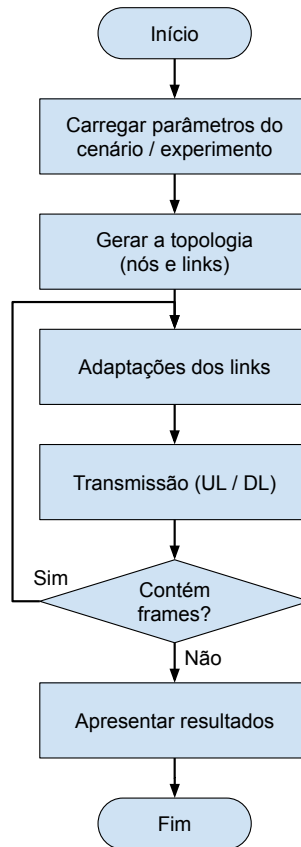


Figura 2.8: Fluxo de trabalho do Vienna 5G LLS.

quadro, considerando tanto aspectos como perda de percurso (*path loss*), potência de transmissão e velocidade (*Doppler shift* do canal do usuário). Durante a simulação, todas as operações de *downlink* e *uplink* são realizadas, para todos os enlaces, com suporte a medições e *feedbacks* de CSI para adaptações da transmissão MIMO e *Automatic Modulation Classification* (AMC). Durante os processos envolvidos nas transmissões, estatísticas são coletadas. A simulação é terminada quando não há mais quadros a serem transmitidos. Ao fim da simulação, o pesquisador terá acesso tanto às estatísticas coletadas, quanto aos resultados individuais dos quadros transmitidos.

O simulador oferece suporte tanto a *Frequency-Division Duplexing* (FDD) quanto a estrutura de quadros *Time-Division Duplexing* (TDD), assumindo a flexibilidade de espaçamento entre as subportadoras, conforme a numerologia 5G. Foi também incluído o suporte de nós intermediários na definição da topologia, habilitando o emprego de *Reconfigurable Intelligent Surfaces* (RISs). Além das características já descritas, o simulador possui as seguintes propriedades técnicas [Pratschner et al. 2021]:

- Múltiplos esquemas de codificação de canal.
- *Feedback* quantizado para adaptar o parâmetros de transmissão para o condições do canal: CQI, *Precoding Matrix Indicator* (PMI) e *Rank Indicator* (RI). Também permite parametrizar o atraso no feedback.

- Suporte a *codebook*, considerando até 4 antenas MIMO no LTE *Advanced* (LTE-A) e até 32 antenas para o 5G.
- Modelos de canal
 - Modelo de canal com desvanecimento duplo, parametrizável entre 500 MHz - 100 GHz, correlação espacial e seletividade temporal via modelos de Kronecker e Jakes, respectivamente.
 - Modelo espacial de canal baseado na TR38.901, com correlação espacial, temporal e de frequência. Define artificialmente a geometria através dos ângulos entre os usuários.
 - Perda de percurso baseada em distância.
 - *Line Of Sight* (LOS) baseado em geometria.
- Estimativas de canal baseadas em pilotos com diversos modos: estimador LS, conhecimento perfeito do canal e padrões consistentes com LTE e 5G.
- Diferentes esquemas de modulação, incluindo OFDM, f-OFDM, WOLA, FBMC, UFMC e portadora única.
- Diferentes modos de transmissão (TM): diversidade de transmissão e/ou recepção, multiplexação espacial de *looping* aberto e fechado, e mu-MIMO em *downlink* e *uplink*.
- Múltiplos equalizadores com esquemas de detecção MIMO: ZF, MMSE, *Sphere Decoder* e *Maximum Likelihood*.

2.4.2 Simuladores em nível de sistema

Uma outra categoria de simuladores de redes é o simulador em nível de sistema, ou *System-level Simulator* (SLS). Entre os seus objetivos encontram-se possibilitar a investigação de aspectos relacionados à rede como um todo, como na avaliação de performance de parâmetros como capacidade, latência e atrasos, cobertura, mobilidade, eficiência, entre outros, permitir a rápida prototipação e testes de métodos e capacidades inovadoras, e permitir o planejamento de novas redes [Müller et al. 2018].

Os simuladores em nível de sistema criam mecanismos para a análise em uma perspectiva mais abrangente, considerando o comportamento global da rede. Para tal, eles integram múltiplos componentes de rede, proporcionando uma visão geral do desempenho desses componentes em cenários complexos. Nas redes sem fio de próxima geração, tal recurso se torna essencial para examinar o desempenho da rede perante aspectos como mobilidade, escalonamento de usuários e o gerenciamento e alocação de recursos, entre outros. Dentre os principais simuladores em nível de sistema, dois se destacam, sendo ambos descritos a seguir.

Vienna 5G System-Level

O simulador em nível de sistema que também compõe a suíte de simuladores VCCS é o Vienna 5G SLS [Müller et al. 2018]. Assim como ocorreu com o desenvolvimento do Vienna 5G LLS, a versão em nível de sistema para 5G surge como uma evolução do simulador voltado ao LTE (Vienna LTE *simulators*). O simulador atua em um escopo voltado à geração flexível de topologias com geometria arbitrária, permitindo o uso de um grande número de ERBs e usuários, com a possibilidade de simulação de redes multicamadas. Isso é possível pois o simulador permite a modelagem da geometria da rede com diversos modelos e efeitos de propagação, em que a performance da rede simulada é obtida através na força do sinal recebido, combinado com as métricas de SINR.

De fato, a simulação sistêmica de uma rede requer uma série de abstrações. Como também é implementado em MATLAB, a camada PHY é abstraída ao receber como entrada informações vindas do próprio simulador em nível de link (Vienna 5G LLS). Já a camada MAC, por outro lado, não exige a implementação integral da pilha de protocolos, sendo focada nas tarefas de escalonamento e *feedback*. Assim, os autores do simulador assumem uma arquitetura modular, com componentes desenvolvidos no paradigma da orientação a objetos, o que favorece a escalabilidade das simulações (novamente, a limitação encontra-se na capacidade computacional empregada na simulação).

A Figura 2.9 apresenta a estrutura de blocos e o fluxo de trabalho de uma simulação, adaptado de [Müller et al. 2018]. Assim é possível observar as principais tarefas executadas em cada um dos blocos individuais. Assim como no Vienna 5G LLS, inicia-se definindo o cenário e todos os parâmetros que vão guiar o experimento. Dentre esses parâmetros, a definição do tempo de duração da simulação é bastante importante, que vai guiar a criação dos *Time Slots* (TSs), segmentos e blocos. Um TS é uma unidade de tempo que vai consumir uma iteração no laço de repetição principal da simulação, representando a estrutura de quadros (LTE-A, por exemplo). Um segmento é definido como o tempo em que o *macroscopic fading* (MF) é constante. A consistência do canal e a correlação espacial é levada em consideração aqui, dado que a duração do segmento é determinada tanto pela velocidade e trajetória do usuário, quanto pela descorrelação da distância especificada. Por sua vez, um bloco é uma sequência de segmentos consecutivos, conforme demonstrado na Figura 2.10. Assim, no início de cada segmento, os valores de MF são atualizados, permanecendo constante durante todo o segmento.

Assim que a simulação é inicializada, todos os elementos que possuem posicionamento são criados, considerando uma geometria 3-D. O simulador também permite estabelecer no cenário paredes e edificações. Assim que as ERBs e usuários são criados e posicionados, a região de interferências pode ser calculada na chamada *region of interest* (ROI).

Tendo todo o processo de pré-geração terminado, o simulador entra no fluxo de

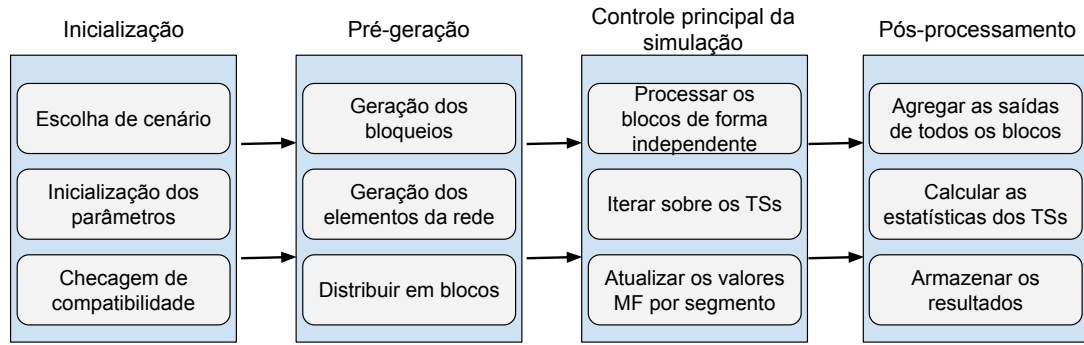


Figura 2.9: Fluxo de trabalho do Vienna 5G SLS.

controle principal da simulação. Cada bloco criado será processado individualmente, podendo ser tanto de forma consecutiva ou em paralelo (dependendo das configurações, dependências MATLAB e recursos computacionais). Cada bloco exige um laço de controle para processar todos os TSs de um segmento. Ao fim do segmento, as posições dos usuários em regime de mobilidade são atualizados, momento em que o MF correspondente, refletindo os impactos de propagação e interferências ocasionadas pelas novas posições. Nesse momento, todos os processos relativos ao *feedback* da qualidade do canal e escalonamento são desempenhados. Por fim, a cadeia de transmissão-recepção (*demapping* e *decoding*) é abstraída para emular a transmissão de bits, de forma que SINR calculada é mapeada em *Block Error Rate* (BLER) e vazão, considerando o MCS definido no escalonamento para a transmissão.

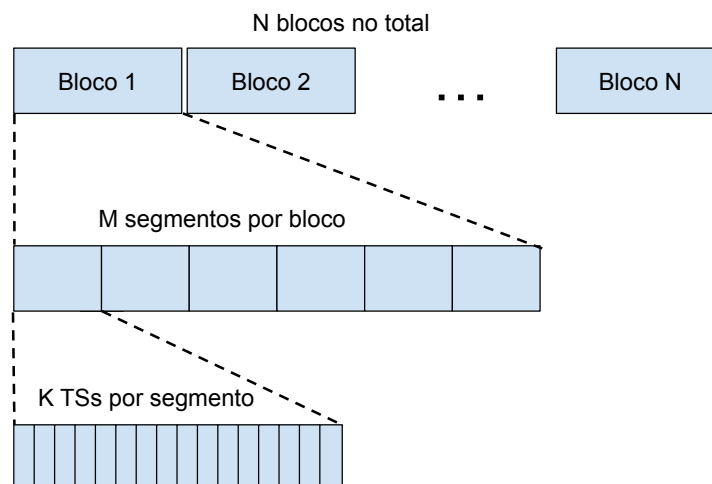


Figura 2.10: A estrutura temporal do Vienna 5G SLS, em que o tempo de simulação é dividido em blocos, segmentos e TSs.

Terminando-se o processamento de todos os blocos, a simulação é terminada. Os resultados individuais de cada TS é então processado de forma combinada entre todos os blocos. O simulador permite determinar quais dados e estatísticas serão coletados

e armazenados, de forma que apenas os resultados de interesse são processados, o que otimiza a simulação. Assim, além de armazenar em arquivo, os resultados são apresentados em uma interface gráfica.

Além dos aspectos já considerados, o simulador Vienna 5G SLS possui uma grande variedade de modelos de canal, que podem ser classificados em três tipos [Müller et al. 2018]: A) Modelos 2D estocásticos, como *Rayleigh* e *Rician*, *Additive White Gaussian Noise* (AWGN), entre outros; B) Modelos 3D baseados em geometria, como os modelos 3GPP para a banda sub-6 GHz e para a banda de até 100 GHz; e C) Interface para modelos externos, como o QuaDRiGA [Jaekel et al. 2014].

Simu5G

Outro relevante representante dos simuladores em nível de sistema é o Simu5G [Nardini et al. 2020, Nardini et al. 2020]. Trata-se de um simulador que considera o plano de usuário do 5G NR, desenvolvido no OMNeT++¹² integrado à biblioteca INET¹³, evoluído a partir de outro simulador do mesmo grupo, o SimuLTE¹⁴, voltado a redes 4G. É um simulador de código fonte aberto, de licença LGPL para uso tanto acadêmico quanto comercial, o que favorece a sua adoção pela comunidade científica. Desenvolvido em C++ e completamente customizável, oferece interfaces que permitem o reuso de recursos existentes e a inclusão de novos módulos. Por manter a compatibilidade com a versão anterior, permite explorar cenários de coexistência 4G/5G. Com simulações fim a fim, possui recursos para explorar redes TCP/IP, dispositivos padronizados de redes [Virdis et al. 2020], chegando até ao plano de dados da RAN e o núcleo 5G. Uma importante diferenciação com relação ao simulador Vienna 5G SLS é a implementação de funções de outras camadas acima da camada MAC, permitindo explorar cenários mais amplos e complexos.

A Figura 2.11 apresenta uma visão geral sobre as funcionalidades disponíveis no Simu5G. Os terminais de usuários podem ser tanto estáticos quanto móveis, sendo suportados diferentes modelos de mobilidade, inclusive emulando a mobilidade veicular. As ERBs podem ser estabelecidas como macro, micro e pico-células, operando em TDD/FDD com diferentes numerologias, agregação de portadoras, suporte a *handover*, com diversos algoritmos de escalonamento disponíveis, e contando com recursos nativos para a coordenação de interferências. O simulador apresenta diversos tipos de tráfego parametrizáveis, permitindo emular diferentes tipos de aplicações. Adicionalmente, há

¹²<https://omnetpp.org>

¹³<https://inet.omnetpp.org>

¹⁴<https://simulte.com/>

suporte tanto para comunicação *Device-to-Device* (D2D), quanto para a prototipação de aplicações MEC [Virdis et al. 2020].

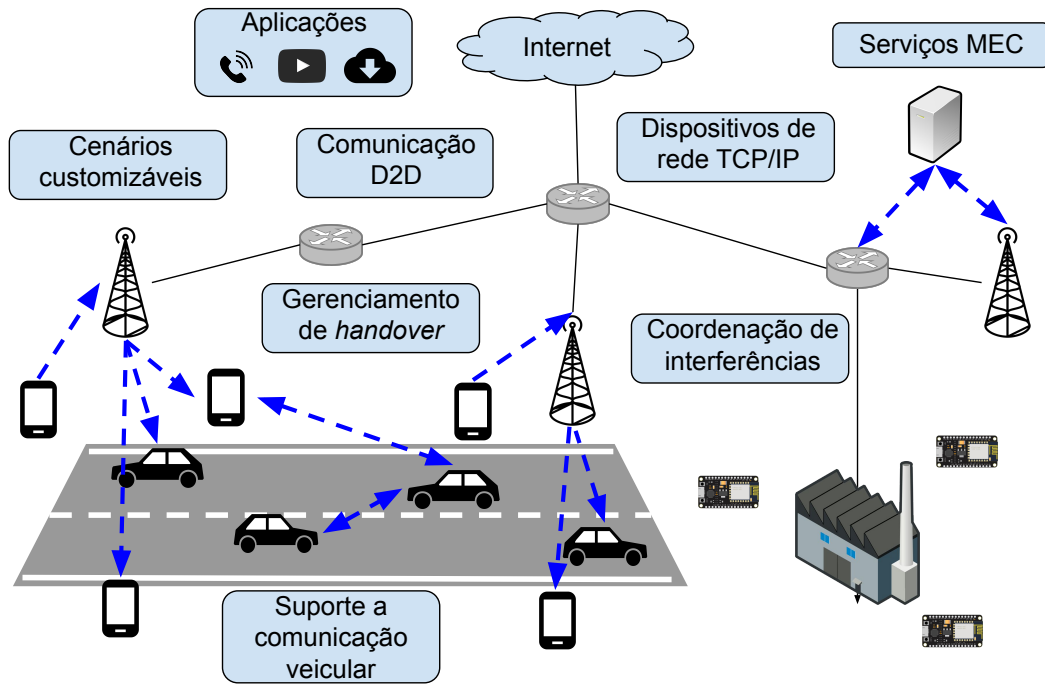


Figura 2.11: Visão geral sobre as capacidades do Simu5G.

Os principais componentes do simulador são o UE e a ERB, podendo atuar tanto como uma eNB ou uma gNB. Ambos elementos são munidos de *Network Interface Cards* (NICs), chamada de *NrNic*, demonstradas na Figura 2.12, adaptada de [Nardini et al. 2020]. A NIC do UE é munida de duas pilhas de protocolos (LTE e NR), habilitando a conectividade dupla. Já a NIC da gNB possui uma interface X2, permitindo emular aplicações voltadas à coordenação entre gNBs. Adicionalmente, a gNB possui uma segunda NIC, voltada à integração com demais dispositivos de rede via protocolo ponto a ponto TCP/IP. Cada protocolo da pilha de protocolos está implementada em seu correspondente módulo. A camada *Packet Data Convergence Protocol* (PDCP) está contida no módulo *NrPdcP*, a camada *Radio Link Control* (RLC) na *NrRlc*, e assim por diante.

Um ponto relevante no Simu5G é a sua implementação da camada MAC, principalmente pelo fato de suportar tanto FDD quanto TDD, agregação de portadoras (múltiplos *componentes de portadora*), e diferentes numerologias dentre aquelas definidas no padrão NR. Assim, o simulador emprega uma boa representação de quadros, subquadros e slots de tempo, ou *Transmission Time Intervals* (TTIs). No Simu5G diferentes índices de numerologia (μ) podem ser atribuídos a cada componente de portadora. Em FDD, cada componente de portadora terá recursos espectrais separados para *uplink* e para *downlink*, ao passo que, em TDD, o mesmo espectro será utilizado nas duas direções, com separação no tempo. Adicionalmente, de forma aderente ao padrão NR, o simulador permite o uso

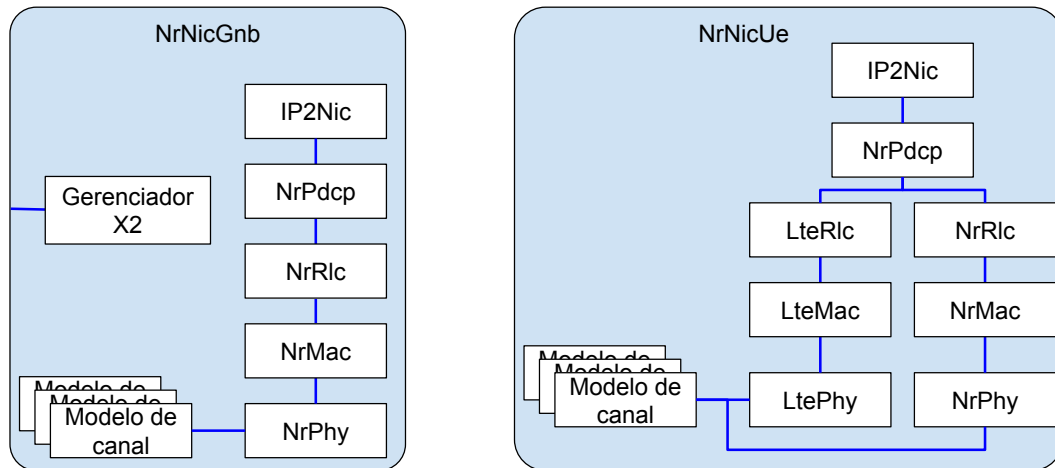


Figura 2.12: Estrutura básica de uma NR NIC no Simu5G.

flexível de formatos para os slots no TDD, fazendo com que a quantidade de bytes que podem ser transmitidos varie de acordo com o formato do slot TDD definido, bem como do MCS selecionado. Toda essa estruturação garante com que a gNB possa implementar um escalonador totalmente independente para cada componente de portadora, executando periodicamente a cada TTI. Além disso, são também tarefas consideradas para o módulo da camada MAC do simulador: escalonamento de usuários, reporte e cálculos de CQI, construção do MAC *Transport Block* (TB), alocação de *Resource Blocks* (RBs), definir os procedimentos do AMC e *Hybrid Automatic Repeat reQuest* (HARQ) [Nardini et al. 2020].

2.5 Conclusão

Após a apresentação do referencial teórico nesse capítulo, é importante discutir a relação entre os elementos aqui descritos e o desenvolvimento da pesquisa. Os detalhes sobre a alocação de recursos bem como sobre as redes de acesso por rádio, apresentados na Seção 2.1 e na Seção 2.2, respectivamente, formam a base dos modelos de sistemas considerados nas investigações descritas nos Capítulos 3 e 4. Em ambos capítulos são descritas soluções propostas para os problemas de alocação de recursos de rádio, de computação e de infraestrutura, em que os desenvolvimentos e as avaliações das propostas consideraram tanto as soluções de prateleira para AI/ML, descritos na Seção 2.3, quanto os simuladores descritos na Seção 2.4. Além disso, o conhecimento obtido a respeito das redes de acesso virtualizadas e desagregadas foi fundamental para a proposição de soluções aderentes à arquitetura O-RAN, sendo essa a arquitetura que se mostrou com grande relevância durante as pesquisas.

Escalonamento de usuários baseado em DRL para MIMO massivo multibandas

Neste capítulo, o problema de escalonamento e alocação de recursos para redes MIMO massivo multibandas, considerando múltiplos usuários, é apresentado. Neste problema, diante da impossibilidade de atender a todos os usuários conectados num mesmo instante, deve-se definir um subconjunto desses usuários que devem ser escalonados no o próximo tempo de transmissão e em quais faixas de frequência cada usuário deve ser alocado, considerando tanto os aspectos da qualidade do canal quanto da demanda requerida por cada usuário. Esse é um problema de otimização combinatória multiobjetivo, o qual foi resolvido por um agente baseado em DRL neste trabalho.

3.1 Introdução

As classes de serviços projetadas para as redes sem fio de próxima geração vêm impondo severas dificuldades na obtenção dos recursos necessários ao cumprimento de seus rigorosos requisitos. Embora MIMO tem sido empregado, há anos, nos mais variados sistemas de comunicação sem fio, cumprindo papel fundamental na otimização da eficiência espectral [Björnson e Jorswieck 2013], o emprego de sua versão massiva (MIMO massivo), operando em ondas milimétricas (mmWave), deve ser cada vez mais aguardado como solução viável ao desdobramento de tais requisitos, provendo acesso sem fio tanto aos usuários móveis ou fixos.

Porém, o emprego de ondas milimétricas impõe diversas limitações, principalmente devido suas características de propagação. Usuários que estejam experimentando boas condições de canais, usualmente aqueles que estejam mais próximos à estação rádio base, podem se beneficiar dos canais em mmWave, de forma que os demais usuários devam utilizar a banda sub-6 GHz. Assim, devido à menor largura de banda disponível nesta faixa espectral mais baixa, o emprego de múltiplas bandas pode garantir melhorias na vazão dos usuários finais, bem como garantir melhorias em termos de eficiência espectral para o sistema como um todo.

Em sistemas reais, em que o número de cadeias de *Radio Frequency* (RF) são menores que o número total de usuários a serem atendidos, num determinado instante e região geográfica, o emprego de métodos de escalonamento de usuários torna-se uma importante abordagem, principalmente em termos da obtenção dos benefícios de ganhos de diversidade e multiplexação proporcionados pelo MIMO massivo. Adicionalmente, os diversos recursos envolvidos nos sistemas de comunicação sem fio precisam ser alocados, empregando alguma política que vise otimizar os parâmetros de performance, o que geralmente se torna um problema de otimização multiobjetivo complexo.

Métodos de escalonamento e alocação de recursos (SRA) tradicionais, como o *Round Robin*, selecionam os usuários sem levar em consideração informações da camada física, tal como as informações sobre os seus canais [Ajib e Haccoun 2005, Asadi e Mancuso 2013]. Embora tais abordagens possam favorecer critérios de escalonamento e alocação baseado em justiça entre os usuários, e tendem a representar menor complexidade de implementação e operação, podem não favorecer melhorias de eficiência espectral e vazão do sistema [Bendaoud, Abdennebi e Didi 2014]. Diversos métodos têm sido propostos levando em consideração tais informações, também conhecidos como os escalonadores oportunistas [Asadi e Mancuso 2013], que seguem estratégias baseadas no estado atualizado do canal, bem como do histórico de vazão de cada usuário, e também podem incorporar comportamentos de alocação justa, como nas abordagens de *proportional fairness*.

Um passo ainda mais recente envolve métodos que lidam com dados advindos de múltiplas camadas, com foco em melhorias de performance [Khalek, Caramanis e Jr 2015, Femenias et al. 2017, Maaz et al. 2017, Vora e Kang 2018]. Apesar do ganho em termos de flexibilidade, novos níveis de complexidade são criados com tais abordagens, principalmente devido ao grande número de parâmetros operacionais que precisam ser observados. Assim, o custo computacional envolvido na otimização dos sistemas de comunicação, através do emprego destes métodos, tem motivado o desenvolvimento de abordagens baseadas em métodos de aprendizagem de máquina, que podem se beneficiar da grande quantidade de dados que podem ser obtidos dos sistemas modernos [Lin e Zhao 2020, Tang, Zhou e Kato 2020].

Neste sentido, o problema *Scheduling and Resource Allocation* (SRA) em redes sem fio baseadas em MIMO massivo multibandas, tratado como um problema de otimização combinatório e multiobjetivo, é investigado, em que um método baseado em DRL é proposto. Tal problema é aderente às soluções de RL, em que o escalonador precisa tomar decisões continuamente, enquanto busca se adaptar às dinamicidades do ambiente, potencialmente não adequado aos métodos de otimização tradicionais. Embora o emprego de *Markov Decision Process* (MDP) e de métodos de aprendizado baseados em tabelas também sejam considerados na literatura

tura no tratamento de problemas de otimização em sistemas de comunicação, como em redes SON [Pandana e Liu 2005], SRA [Djonin e Krishnamurthy 2007] e alocação de recursos para URLLC [Bhardwaj, Ginanjar e Kim 2020], por exemplo, a escalabilidade se torna um problema desafiador, principalmente devido ao crescimento da complexidade quando a quantidade de parâmetros a serem tratados se torna maior [Koutsopoulos e Tassiulas 2008, Choromanski et al. 2019].

Entretanto, alguns problemas de interesse em SRA possuem um grande número de parâmetros, estados e ações, não podendo ser modelados como MDP finito, tampouco como métodos de aprendizado tabular [Sutton e Barto 2018]. Adicionalmente, tais problemas impõem aos métodos baseados em RL um longo tempo de convergência, também tornando-os inapropriados. Porém, a combinação de RL e as redes neurais profundas (DNNs), dão origem ao DRL, conforme descrito na Seção (6.2.3). Assim, DRL tem sido extensivamente investigado para a solução de problemas de SRA nos últimos anos [Comsa et al. 2014, Ghadimi et al. 2017, Comsa, De-Domenico e Ktenas 2017, Zhu et al. 2017, Lee 2018, Yang et al. 2018, Chen et al. 2020, Shekhawat et al. 2020, Al-Tam, Correia e Rodriguez 2020, Tang, Zhou e Kato 2020], discutidos a seguir.

Trabalhos relacionados

Muitos dos trabalhos envolvendo RL aplicado à comunicação assumem um espaço de ações discreto. Por exemplo, uma arquitetura *Deep Belief Architecture* (DBA) é utilizada em [Tang, Zhou e Kato 2020] para a alocação dinâmica de recursos em HetNets 5G, com UEs se movendo em diferentes velocidades, de forma que a tarefa do agente envolve a divisão de subquadros TDD para *uplink* e *downlink*, num total de seis ações discretas. Os autores em [Ghadimi et al. 2017] assumem que o agente RL auxilia na estratégia de alocação de potência utilizando cinco ações discretas, no contexto de um método distribuído para adaptação de taxa de dados e controle de potência intercelular, empregando uma rede neural artificial, treinada para estimar os valores Q [Sutton e Barto 2018]. Assim, cada célula é controlada por um agente DRL, que utiliza medições locais em um esquema de observabilidade parcial, dado que observações do estado intercelular não estão disponíveis. Um cenário de simulação simplificado é empregado, considerando apenas uma ERB, com UEs fixamente posicionados em locais randômicos.

Em [Comsa, De-Domenico e Ktenas 2017], o objetivo do agente RL proposto é obter uma política de alocação de recursos que satisfaça diferentes objetivos de QoS, incluindo a minimização da taxa de perda de pacotes, garantia de taxa de *bits* e redução no atraso de pacotes. Uma arquitetura baseada em *Actor-Critic* é empregada para treinar um modelo que alcance as métricas definidas, selecionando entre três ações discretas, que indicam qual papel o escalonador deve assumir. Os resultados foram avaliados utilizando

apenas dados simulados a partir de células de antena única, sendo que cenários MIMO não são explorados, e com canais sem consistência temporal.

Um algoritmo de escalonamento de usuários baseado em DRL para sistemas *Multi-User* MIMO (MU-MIMO), com foco na otimização de cobertura e capacidade em cenários MIMO massivo é proposto em [Yang et al. 2018]. Um parâmetro que mede o agrupamento dos usuários pela força do sinal é empregado em conjunto com um limiar do QoS, de forma que uma configuração dinâmica possa ser desempenhada, em cada TTI, mediante uma rede neural profunda pré-treinada. Considerando um espaço de ações discreto, composto de quinze níveis discretos de SINR e níveis discretos da força do sinal de vinte UEs, e com espaço de observação composto dos *Key Performance Indicators* (KPIs) de eficiência espectral média da célula e eficiência espectral da borda da célula. Entretanto, os autores deixam de investigar como as decisões de alocação de potência podem afetar a performance dos usuários.

Um escalonador de recursos de rádio para as múltiplas configurações de nume-
rologias do 5G NR é proposto em [Al-Tam, Correia e Rodriguez 2020], em que um UE é selecionado para as atribuições de PRBs. Assumem um espaço de ações discreto, mas com dimensões definidas dinamicamente por características e parâmetros decorrentes do problema modelado, como o número de UEs, ERBs, *buffers*, fluxos de dados e funções de rede, impactando assim a escalabilidade e generalização [Attiah, Sohrabi e Yu 2020] dos modelos treinados, requerendo métodos de treinamento *online*. O espaço de observação é composto dos indicadores dos UEs para pacotes disponíveis no *buffer*, taxa de dados e justiça. O espaço de ações contém vetores codificados (*one-hot*) correspondentes ao PRB alocado a um UE específico. A latência do *buffer* dos UEs e a taxa de perda de pacotes não são avaliadas neste trabalho, que mantém seu foco nas métricas de vazão, sendo implementado considerando uma simulação de rede móvel semelhante ao empregado em [Comsa, De-Domenico e Ktenas 2017], não considerando MIMO.

Entretanto, nenhum dos trabalhos aqui referenciados consideram a alocação em múltiplas bandas na formulação do problema. Observa-se que tal alocação representa um grande potencial para contornar a pequena largura de banda disponível no espectro sub-6 GHz, bem como para explorar as características das diferentes faixas de frequência. O problema de decisão sobre qual banda disponível pode atender a um determinado usuário, em cada instante, da forma mais adequada, introduz uma dimensão extra ao problema de SRA, e pode se tornar um importante fator para dar suporte aos diferentes requisitos de usuários nas redes móveis de próxima geração. Desta forma, busca-se explorar tal potencial formulando o presente problema como um problema de otimização combinatória, com a proposição de um método baseado em DRL.

3.2 Definição do problema

Modelo do sistema de comunicação considerado

O sistema de comunicação adotado incorpora importantes fatores que influenciam a performance dos sistemas de comunicação sem fio, tal como a consistência espacial. A simulação destes sistemas dentro do escopo de treinamento de agentes com RL é, normalmente, computacionalmente proibitivo, dado que requerem uma grande quantidade de iterações até obter a convergência. Neste sentido, o trabalho emprega uma estratégia em termos de ambiente de treinamento e simulação, que utiliza os limites de capacidade para estimar a eficiência espectral (SE) [Björnson, Hoydis e Sanguinetti 2017], ao invés de se basear na simulação da transmissão de símbolos em método de Monte Carlo.

O sistema considera a operação em *downlink* MU-MIMO, assumindo um cenário de múltiplas células, em que cada estação rádio base (ERB) possui três setores, num total de L células, servindo K usuários conectados (ou ativos) por célula. Não há cooperação entre as ERBs, em que a interferência intercelular é tratada como ruído. O número total de antenas na ERB e em cada UE é definido por N_t e N_u , respectivamente. Por simplicidade, o MIMO massivo canônico [Björnson, Hoydis e Sanguinetti 2017] é assumido, no qual as ERBs são equipadas com painéis de antenas do tipo *Uniform Linear Array* (ULA) com $N_t = 64$, e cada UE com $N_u = 1$ antena omnidirecional. O sistema pode operar em múltiplas bandas de frequências, sendo que o agente está operando na célula $c = 1$, podendo sofrer interferências das outras $L - 1$ células.

Com foco em um cenário mais realista, assume-se transmissão em modo TDD, sendo FDD também plausível, caso o CSI seja adequadamente obtido. Considera-se as transmissões ocorrendo em blocos (ou quadros) de duração de T_s segundos, conforme demonstrado na Figura 3.1. Um episódio de T_e segundos de duração é composto de N_b blocos, sendo a duração da transmissão em *downlink* em um bloco ser de $\tau_d T_s$ segundos, em que τ_d é o número de *slots* de tempo (ou amostras) em *downlink*, e T_s o intervalo de amostragem. De forma similar, τ_p e τ_u são o total de *slots* por bloco dedicados à sequência de pilotos e transmissão de dados em *uplink*, respectivamente. Considera-se passos temporais discretos, em que $t \in \mathbb{Z}^+$, redefinindo para $t = 1$ a cada início de um episódio. Desta forma, o tempo relativo correspondente ao t -ésimo *slot* de um episódio é de $(t - 1)T_s$ segundos. Dado que *uplink* não é considerado, τ_u não é utilizado. A duração de T_c é escolhido de acordo com a definição do tempo de coerência do canal, que varia no decorrer de um bloco de acordo com o modelo de propagação adotado.

Modelagem dos canais

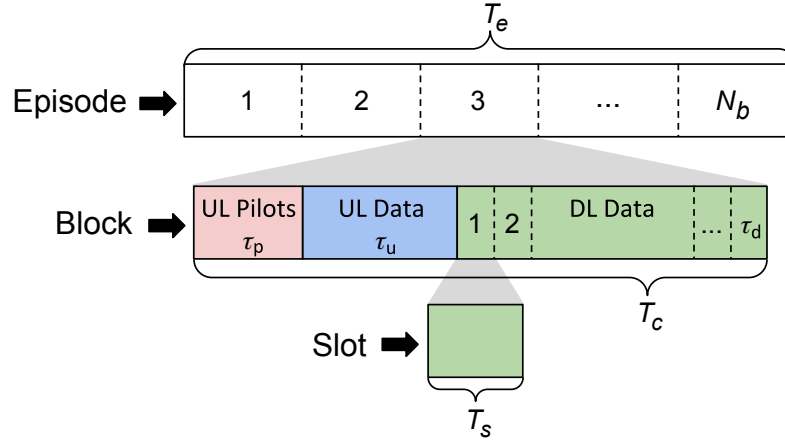


Figura 3.1: Escala temporal e organização do quadro: T_s , T_c e T_e são as durações de um *slot*, bloco e episódio, respectivamente.

Os cenários de simulação consideram canais gerados com a suíte de ferramentas QuaDRiGa [Jaeckel et al. 2014, Burkhardt et al. 2014]. Inicialmente, UEs são randomicamente posicionados em uma região a até 150 metros de distância da ERB, movendo-se a velocidades de acordo com a distribuição normal $|\mathcal{N}(10,3)|$. A cada segundo, o UE pode mudar a direção do seu movimento com uma probabilidade $P_{\text{turn}} = 0.2$. Uma vez que todos os caminhos são definidos, coeficientes dos canais são gerados, de acordo com um cenário 3GPP *Urban Micro* (UMi) [Mondal et al. 2015, Zhu et al. 2019], e correlacionados através dos UEs no espaço, tempo e banda de frequências. O intervalo de amostragem é $T_s = 1$ ms, sendo um episódio com $T_e = 2$ s de duração. A cada amostra, o canal é caracterizado como sendo LOS ou *Non Line Of Sight* (NLOS), dependendo da distância e dos *clusters* de espalhamento. As distribuições gerais dos coeficientes são distintas entre canais LOS e NLOS, sendo que caso estejam em LOS, o modelo da perda de percurso de larga escala segue um modelo *dual-slope* [Zhu et al. 2019], definido por

$$PL_{\text{UMi-LOS}} = \begin{cases} PL_1 & d_{2D} \leq d_{BP} \\ PL_2 & d_{2D} > d_{BP} \end{cases}, \quad (3-1)$$

sendo d_{2D} a distância bidimensional entre a ERB e o UE definida pelo 3GPP, e d_{BP} a distância de *breakpoint*, definida por

$$d_{BP} = BP_{SF}(h_{BS} - h_{Env})(h_{UT} - h_{Env})f_c, \quad (3-2)$$

sendo BP_{SF} o fator de escala de *breakpoint*, f_c a frequência central em GHz, h_{BS} , h_{Env} e h_{UT} as alturas da ERB, do ambiente e do UE, respectivamente. PL_1 e PL_2 são definidos p

$$PL_1 = 21 \log_{10}(d_{3D}) + 32.4 + 20 \log_{10}(f_c) + d_{3D} d_{BU}, PL_2 = PL_1(d_{BP}) + 40 \log_{10}\left(\frac{d_{3D}}{d_{BP}}\right), \quad (3-3)$$

onde d_{BU} é a distância entre a ERB e o UE.

Para canais em NLOS, a perda de percurso é definida por [Zhu et al. 2019] da seguinte forma:

$$PL_{UMi-NLOS} = \max(PL_{UMi-LOS}, PL'_{UMi-NLOS}), \quad (3-4)$$

sendo

$$PL'_{UMi-NLOS} = 35.3 \log_{10}(d_{3D}) + 22.4 + 21.3 \log_{10}(f_c) - 0.3(h_{UT} - 1.5). \quad (3-5)$$

O *Reference Signal Received Power* (RSRP) é calculado para cada possível conexão entre a ERB e os UEs [Mondal et al. 2015, 3GPP 2017]. Ao calcular as interferências, as 6 principais medições de RSRP dos sinais de interferência mais fortes são consideradas, semelhante a um relatório de medição de RSRP visto por uma ERB local.

Eficiência espectral

Conforme já considerado, pelo fato de não se levar em consideração a simulação da transmissão de símbolos, eficiência espectral (SE) é empregada no cálculo da taxa de bits. Existem fórmulas de capacidade bem estabelecidas para MIMO massivo, que habilitam o cálculo da SE em bits/s/Hz [Marzetta, Larsson e Yang 2016, Björnson, Hoydis e Sanguinetti 2017, Jr e Lozano 2018]. Assumindo-se uma capacidade limitada, quando considerando uma largura de banda BW, a taxa de bits para o usuário u , na célula b , é definida por $R_{bu} = SE_{bu}BW$, sendo a *Spectrum Efficiency* (SE) obtida com

$$SE_{bu}^{DL} = \frac{\tau_d}{\tau_c} \log_2 \left(1 + \frac{\rho_{bu} RSRP_{bu}}{I_{bu}^{inter} - (\rho_{bu} RSRP_{bu}) + \sigma_k^2} \right), \quad (3-6)$$

em que a interferência correspondente pode ser obtida com

$$I_{bu}^{inter} = \sum_{\substack{c=1 \\ c \neq b}}^L \sum_{u'=1}^K \rho_{cu'} RSRP_{cu'}. \quad (3-7)$$

Escalonamento de usuários e alocação de frequências

Assume-se que o UE limita-se a processar apenas uma banda de frequências em cada bloco, mesmo tendo a capacidade de trocar de frequência. A ERB pode transmitir

simultaneamente em todas as F bandas de frequências, cada qual denotada pela sua frequência de portadora f_c , como em $f_c \in \{2, 28\}$ GHz, por exemplo, e com BW_c .

O agente escalonador da célula alvo $c = 1$ é ilustrado na Figura 3.2, sendo que a notação $h_{bcu}[f_c, t]$ representa a ERB recebendo informações sobre o canal do usuário u nas bandas de frequências $f_c \in \{f_1, \dots, f_F\}$. Em cada *slot* $t \in \tau_s$ dentro de um bloco $m \in N_b$ (Figura 3.1), o agente escalonador precisa decidir sobre quais UEs precisam ser alocados, para cada banda de frequência $f \in F$ disponível, criando um subconjunto disjunto $\mathcal{K}_{m,f,t} \subset K$ de usuários a serem servidos. Assim como considerado em [Choi et al. 2018], emprega-se apenas um bloco de recurso físico (PRB) por *slot* de tempo e frequência, que pode ser espacialmente multiplexado através de diferentes usuários, como apresentado na Figura 3.3.

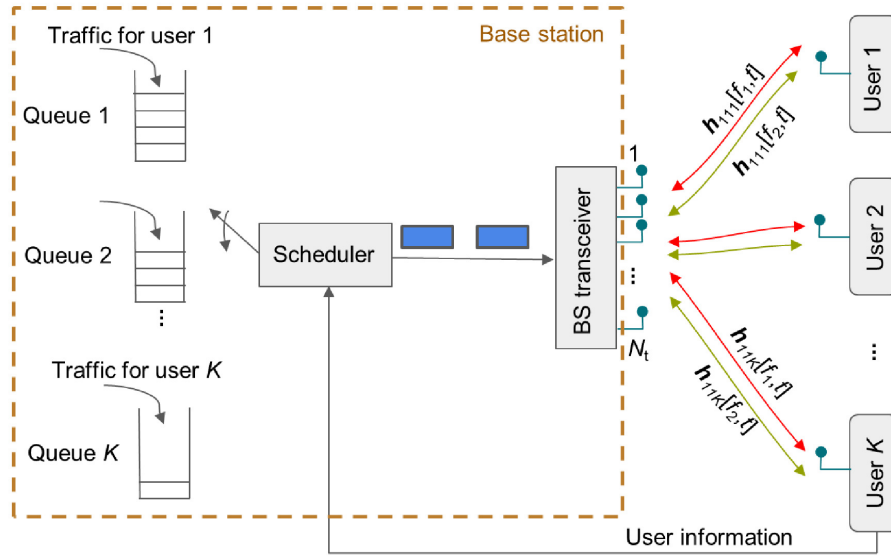


Figura 3.2: SRA multibandas para MU-MIMO utilizando $F = 2$ frequências: f_1 e f_2 .

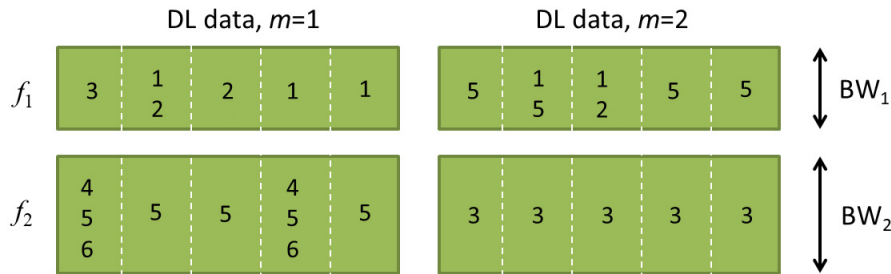


Figura 3.3: Exemplo ilustrando os usuários alocados de acordo com o modelo de alocação de recursos multiusuário e multibanda proposto, com $F = 2$, $K = 6$ e $BW_2 > BW_1$.

Assume-se que K_{\max} é o número máximo de usuários que podem ser servidos em um *slot* t na banda f , isto é, a cardinalidade máxima de $|\mathcal{K}_{m,f,t}|$ é K_{\max} . Um exemplo

do escalonamento é apresentado na Figura 3.3, para os dois blocos $m = 1$ e 2 , sendo $\mathcal{K}_{1,f_1} = \{1, 2, 3\}$ e $\mathcal{K}_{2,f_1} = \{1, 2, 5\}$ os usuários escalonados para o primeiro e segundo blocos, respectivamente. O usuário 6 é servido em 2 *slots* do primeiro bloco, como definido em $\mathcal{K}_{1,f_2,1} = \mathcal{K}_{1,f_2,4} = \{4, 5, 6\}$, não sendo escalonado no segundo bloco. Sem perda de generalidade, assume-se que, em contraste à Figura 3.3, K_{\max} sempre são escalonados em cada banda de frequências, de forma que $|\mathcal{K}_{m,f,t}| = K_{\max}, \forall t$. A ERB emprega a mesma potência de transmissão em *downlink* P_{DL} para banda de frequências, de forma que a potência total é FP_{DL} .

É assumido que todos os usuários ativos podem enviar informação de *feedback* através de um canal de controle, indicando sua qualidade de canal por meio do CSI para cada banda f_c . Tal canal de controle também é utilizado pela ERB para informar, antes da transmissão dos dados em *downlink* para o bloco m , os conjuntos $\mathcal{K}_{m,f}$ and $\mathcal{K}_{m,f,t}$ for t dentro do bloco m , conforme apresentado na Figura 3.4.

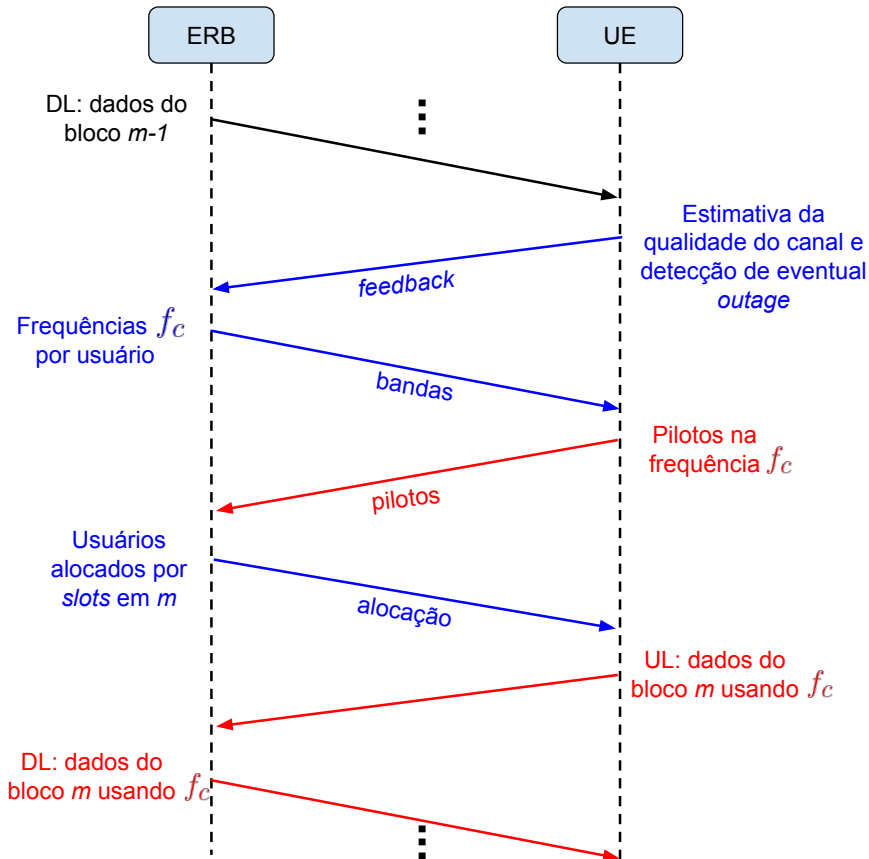


Figura 3.4: Interações entre a ERB e o UE, com as setas azuis indicando o uso do canal de controle, e as setas vermelhas indicando a transmissão usando a portadora em f_c definida pelo escalonador, para o bloco m .

Por simplicidade, assume-se uma ideal e completa observabilidade do cená-

rio [Avranas, Kountouris e Ciblat 2020], na qual o escalonador na ERB obtém o CSI completo para todos os K usuários ativos, para todas as F bandas de frequências para cada bloco m , sendo que todo o processo do escalonador realizado durante o tempo de coerência do canal. É importante ressaltar que o uso do estado do canal em uma observabilidade parcial do cenário, em que a informação de CSI atualizada somente é obtida para os usuários em $\mathcal{K}_{m-1,f}$ (isto é, aqueles alocados no bloco antecessor), é de importante para aplicações práticas, mas não é tratado na proposta.

Tráfego de dados e interferência

O tráfego de dados dos usuários é modelado como um processo de Poisson, sendo $\lambda_u[t]$ a média variável no tempo, para o usuário u . O tráfego de entrada é armazenado no *buffer* relativo ao usuário na ERB, ocorrendo o descarte de pacotes quando a sua capacidade é excedida. O *buffer* do usuário u tem o tamanho equivalente a S_u pacotes. Por simplificação da implementação, e dado que o ambiente não considera a transmissão de símbolos, como já considerado, simplesmente contabiliza-se a ocupação do *buffer* de cada usuário u , definida como $q_u \in \mathbb{N}, \forall u \in K$. A ERB conhece de cada K *buffer*: a sua ocupação, a idade do pacote mais velho¹, bem como o número de pacotes descartados.

Portanto, as simulações e, consequentemente, o agente escalonador, não se restringem a cenários como *buffer* cheio ou com *backlogs* infinitos, muitas vezes assumidos para permitir resultados analíticos. As simulações também podem incorporar tráfego em rajadas e tamanho de *buffer* limitado [Castaneda et al. 2016].

3.3 O agente DRL proposto

Sob uma perspectiva do aprendizado de máquina, o agente escalonador baseado em DRL proposto, nomeado de *DRL-based Scheduling and Resource Allocation* (DRL-SRA), é ilustrado na Figura 3.5. O estado de observação consiste de informações tanto dos canais quanto das filas de pacotes nos *buffers*, genericamente descritos como CSI e QSI, respectivamente, similar como considerado em [Shekhawat et al. 2020]. Desta forma, considera-se um agente intercamadas, dado que leva em consideração informações além da camada PHY, tais como do estado dos *buffers* dos usuários ativos, além de demais informações sobre os pacotes em fila.

¹Para saber a idade do pacote mais velho, a ERB armazena o tempo de chegada de todos os pacotes.

Assim, a entrada que o agente recebe adota os parâmetros descritos na Tabela 3.1, de forma que o estado s_t , observado no instante t , agrupa as informações de todos os K usuários conectados, em uma matriz de dimensão $K \times (F + 2)$, sendo dado por:

$$s_t = (\mathbf{o}_t, \mathbf{a}_t, \mathbf{se}_t), \quad (3-8)$$

em que os vetores \mathbf{o}_t , \mathbf{a}_t e \mathbf{se}_t possuem como elementos os valores para o *slot* t dos usuários $u = 1, \dots, K$ para a ocupação do *buffer* $o_t^u \in [0, 1]$, a idade (relativa) do pacote mais velho a_t^u , bem como a eficiência espectral $se_t^{f,u}$ para a banda de frequências f , respectivamente. O valor de a_t^u é obtido por

$$a_t^u = \frac{Oldest_t^u}{\xi}, \quad (3-9)$$

sendo $Oldest_t^u$ representando a idade do pacote mais velho no *buffer* do usuário u , e ξ , sendo a máxima idade em que um pacote pode permanecer no *buffer*, antes de ser descartado. Desta forma, dado que alguns dos elementos do estado são números reais, o agente DRL emprega uma rede neural para lidar com o espaço de observações contínuo.

Tabela 3.1: Definições do estado do agente DRL, em que as últimas duas colunas indicam as dimensões e a frequência de atualização.

Descrição	Categoria	Dim.	Atualização
Ocupação do <i>buffer</i> o_t^u	QSI	$1 \times K$	<i>slot</i>
Idade do pacote mais velho a_t^u	QSI	$1 \times K$	<i>slot</i>
Eficiência Espectral $se_t^{f,u}$	CSI	$F \times K$	bloco

Assume-se que o número total de fluxos de dados em *downlink* é limitado pela dimensão dos pilotos de *uplink* (*TDD*) ao invés do número N_t de antenas na ERB, sendo $|\mathcal{K}_{m,f,t}| < N_t$. Entretanto, o número total de $K > |\mathcal{K}_{m,f,t}|$ usuários conectados é maior que o total de usuários servidos em um mesmo tempo t e banda de frequências $f_c \in F$, de forma que o agente precisa operar num regime de escalonamento [Björnson, Hoydis e Sanguinetti 2017], em que o uso de informações intercambiadas se mostra ainda mais importante. A ERB organiza a informação de CSI na matriz \mathbf{C} , composta das medições de CQI reportadas pelos K UEs conectados, com a dimensão $K \times F$, sendo atualizada a cada bloco, como demonstrado na Figura 3.4.

Dado tratar-se de um problema de otimização combinatória, uma ação do agente DRL-SRA é representada por uma matriz de inteiros, de dimensões $F \times K_{\max}$, tomada para

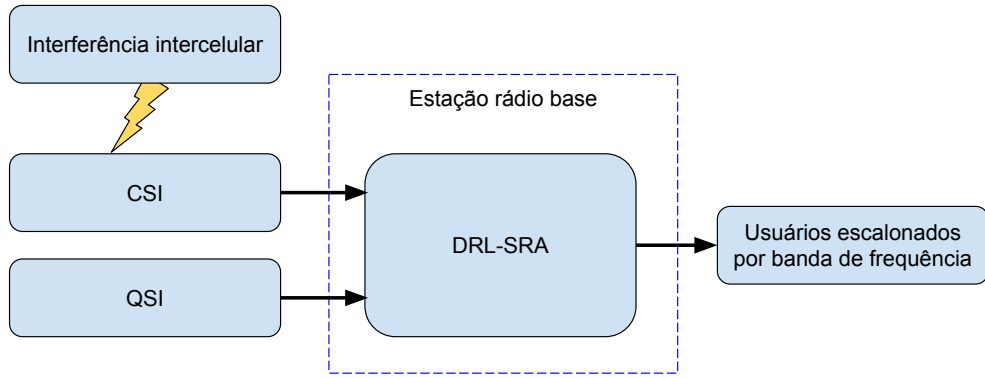


Figura 3.5: O agente DRL-SRA: as entradas são as informações de CSI e QSI, e a saída são os usuários que vão ser servidos, por banda de frequências.

cada observação do estado. Assim, o espaço de ações é discreto, e possui as dimensões

$$N_a = \prod_{i=0}^{F-1} \binom{K - iK_{\max}}{K_{\max}}, \quad (3-10)$$

o que indica que o espaço de ações pode ser extremamente grande. Por exemplo, assumindo $K = 10$, $K_{\max} = 3$ e $F = 2$, tem-se 4200 possíveis ações.

Quando o agente está em treinamento, uma recompensa é calculada para cada ação tomada, permitindo que se crie a política que irá guiar as ações futuras. Como observado em [Liang et al. 2019, Liang, Ye e Li 2019, Ye, Li e Juang 2019, Lu e Gursoy 2019], a solução de problemas de SRA com emprego de DRL é dependente do correto projeto de funções de recompensas, correlacionadas aos objetivos do agente. Na proposta do DRL-SRA, a função de recompensa é dada por

$$r_t = \frac{T_t}{B_{t-1}} - \frac{D_t}{T_t}, \quad (3-11)$$

em que $B_{t-1} = S \sum_{u=1}^K o_{t-1}^u$ é o somatório da ocupação do *buffer* de cada um dos K usuários, sendo S o tamanho do *buffer* em *bits*. De maneira similar, T_t e D_t são, respectivamente, a soma de todos os pacotes transmitidos e descartados no instante de tempo t .

Dado que a modelagem do ambiente RL aqui considerada traz certa flexibilidade aos processos de treinamento e avaliação dos agentes, diferentes algoritmos de RL podem ser considerados. Neste sentido, adotou-se um agente que é capaz de estimar a função de ação-valor $Q(s_t, a_t)$ utilizando atualizações iterativas [Sutton e Barto 2018]. Mais especificamente, um algoritmo DQN foi adotado [Mnih et al. 2013, Mnih et al. 2016], que emprega uma função de aproximação $Q(s_t, a_t; \theta_i)$ com os pesos (ou parâmetros) θ_i como uma *Q-network* da i -ésima iteração [Mnih et al. 2015].

DQN utiliza uma técnica chamada *experience replay*, baseada em um conjunto $\mathcal{D} = \{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ conhecido como *replay memory*, composto da experiência do agente em cada passo de treinamento (isto é, a cada *slot*) de múltiplos episódios $\mathbf{e}_t = (s_t, a_t, r_t, s_{t+1})$, em que s_t é o estado de observação, a_t a ação tomada, e r_t a recompensa. Após o processo de *experience replay* ser concluído, o agente aplica uma política ϵ -greedy para selecionar as ações a serem tomadas, sendo que cada passo de experiência pode ser empregado em várias atualizações dos pesos da rede neural [Mnih et al. 2013], de forma a otimizar o aprendizado do agente. Além disso, pela randomização das amostras, qualquer possível forte correlação existente nas amostras dos dados pode ser evitada, ou minimizada, reduzindo a variância durante o treinamento.

Devido ao custo computacional envolvido na obtenção e processamento de dados dos sistemas de comunicação sem fio, para a realização dos procedimentos de treinamento e validação dos modelos, os canais foram pré-computados, assim como considerado em [Dreifuerst et al. 2020], gerando um *dataset* composto dos canais MIMO gerados pelo QuaDRiGa, uma suíte de ferramentas de código aberto para simulação em *Matlab*. Adicionalmente, o *dataset* incorpora as informações sobre tráfego de dados dos usuários que chega aos *buffers*, habilitando a avaliação de diferentes cargas de tráfego. Os canais neste *dataset* são utilizados para obter os parâmetros necessários, tais como a eficiência espectral, em procedimentos de pós-processamento, em que se emprega uma plataforma de simulação desenvolvida em Python². A plataforma de simulação foi implementada com a biblioteca para DRL *Stable Baselines*³ [Hill et al. 2018], um *fork* do projeto *OpenAI Baselines* [Dhariwal et al. 2017], que emprega *TensorFlow* para as redes neurais e demais ferramentas.

Tal plataforma de simulação foi empregada tanto para o treinamento quanto para o teste da performance do agente DRL-SRA, considerando os aspectos dos subsistemas de comunicação e DRL já descritos. Quando o agente está sendo treinado, o Algoritmo (1) é utilizado, em que a recompensa é computada de acordo com a Equação (3-11), construindo a figura de mérito em cada episódio de treinamento e a *replay memory* \mathcal{D} utilizada pelo algoritmo DQN. Para o teste dos modelos treinados, o agente DRL-SRA selecionará uma ação a_t , dado o estado s_t observado, de acordo com a política ótima obtida durante o treinamento, sendo o subsistema de comunicação similar ao demonstrado no Algoritmo (1).

3.4 Resultados e discussões

²<https://github.com/LABORA-INF-UFG/DRL-SRA-Gym-SB>

³<https://github.com/DLR-RM/stable-baselines3>

Algoritmo 1: Treinamento do agente DRL-SRA.

Input: For the target cell, for frequency band $f = 1, \dots, F$, for all episodes $t = 1, \dots, N_e$: actual $se_t^{f,u}$.

Output: Rewards and figures of merit over time.

```

1 para each episode  $e = 1, \dots, N_e$  faça
2   para each TTI  $t$  faça
3     Scheduler takes action  $a = (\mathcal{U}_t, \mathbf{f}_t)$ ;
4     para each frequency band  $f$  in action  $a$  faça
5       Retrieve spectral efficiency  $se_t^{f,u}$ ;
6       Calculate effective rates  $R[k, t]$ ;
7       Get input traffic;
8       Calculate reward  $r_t$ ;
9       Update queues;
10      Update scheduler agent information;
11 Output results;
```

De forma a observar a capacidade do agente DRL-SRA de aprender como obter uma política de alocação, de forma a atingir os objetivos almejados, que envolvem a maximização da vazão agregada da rede (*sum-rate*), enquanto busca controlar parâmetros de perda de pacotes e atraso. Adicionalmente, busca-se observar o custo computacional envolvido no processo de treinamento, em termos de convergência. Para prover meios de comparação, três escalonadores tradicionais foram considerados, de forma a observar a flexibilidade do agente DRL-SRA em operar de forma similar, assumindo um comportamento mais complexo que tais escalonadores.

Todos os K usuários possuem $N_u = 1$ antena omnidirecional única, com a capacidade total de potência da ERB sendo igualmente distribuída entre todos os usuários, e em cada banda de frequências. Conforme já considerado, os modelos foram previamente treinados e salvos, sendo cada modelo treinado com um total de N_{ep} episódios sendo carregado para cada rodada de validação, de forma que o agente escolhe as ações de acordo com a observação do estado, até que todos os modelos sejam avaliados. Assim, obtêm-se os parâmetros de performance, tais como a vazão média por usuário, a perda média de pacotes por usuário e o atraso médio dos pacotes nos *buffers* por usuário, para cada modelo treinado.

Os três agentes escalonadores adotados para comparação, tradicionalmente observados na literatura [Schwarz, Mehlführer e Rupp 2010, Kawser et al. 2012], possuem diferentes objetivos e performances, quase sempre conflitantes, principalmente em termos da maximização da vazão e alocação baseada em justiça, sendo adequados às comparações aqui necessárias. São eles: *Round Robim* (RR), *Proportional Fairness* (PF) e *Maximum Throughput* (MT), sendo descritos a seguir, segundo [Kawser et al. 2012]. O

escalador RR toma suas decisões de escalonamento de forma cíclica sem levar em consideração informações sobre os canais dos usuários. Como vantagens esse escalador apresenta uma implementação baseada em procedimentos simples, em que a justiça no escalonamento é atingida. Entretanto, ele representa menor desempenho em termos de vazão agregada. O PF leva em consideração a alocação de mais recursos para aqueles usuários que estejam experimentando melhores condições de canais, levando em consideração uma média observada (por exemplo, a vazão média obtida ou o CQI médio reportado). Então, ele fornece um balanceamento entre a maximização da vazão agregada e a justiça. Por fim, o MT, também conhecido como *Maximum C/I*, as decisões de escalonamento consideram apenas aqueles usuários que possuem os melhores canais, em detrimento de todos os outros. Assim, torna-se o agente prioriza a maximização da vazão, comprometendo a justiça.

Os parâmetros empregados tanto nos treinamentos, quanto nas simulações para avaliação são descritos na Tabela 3.2. O algoritmo DQN foi instanciado com uma rede *Multilayer Perceptron* (MLP) com duas camadas de 256 *perceptrons* cada, empregando normalização entre as camadas. De forma a permitir a avaliação do impacto de diferentes cargas de tráfego da rede, as simulações foram realizadas considerando dois diferentes *datasets* de tráfego, cada um representando um padrão de carga por usuário, em que a quantidade média de pacotes que chegam ao *buffer* de cada usuário, em cada *slot* t é alterada. Assim, considera-se o primeiro *dataset* com uma média de chegada de pacotes por usuário de 110 Mbps, e o segundo com 150 Mbps, aqui considerados, genericamente, tráfegos *baixo* e *alto*, respectivamente.

Tabela 3.2: Parâmetros e configurações empregados nas simulações do DRL-SRA.

Descrição	Valor
F	{2, 28} GHz
BW	{20, 100} MHz
Número de usuários por f_c /bloco	2
K	10
Tamanho do <i>buffer</i> (por UE)	240 Kb
Duração do <i>slot</i> (T_S)	1 ms
Blocos por episódio	1000
Taxa de aprendizado	7e-3
Fator de desconto (γ)	0.90
Taxa média de chegada de pacotes	{110, 150} Mbps
DQN <i>policy</i>	MLP 2 x 256
Função de ativação	ReLU

A Figura 3.6 apresenta a evolução da recompensa enquanto os modelos são

treinados, considerando as duas cargas de rede já mencionadas. Em cada passo de tempo (em ms), a alocação dos UEs ocorre, conforme descrito no Algoritmo (1). Ambos modelos apresentam uma tendência de rápida convergência, ocorrendo em torno de 100 segundos. Como esperado, a estratégia *ϵ -greedy* empregada no agente DRL faz com que a recompensa oscile em torno da média, enquanto busca seguir no caminho em que a recompensa seja maximizada no longo prazo. Como forma de garantir que o agente cobrisse todas as possíveis ações, durante o treinamento, uma estratégia de decaimento do ϵ [Afifi e Karl 2020] foi adotada, de forma que nos primeiros 10% dos passos de tempo de treinamento, inicia-se com $\epsilon = 1.0$, que segue sendo reduzido até atingir $\epsilon = 0.02$, se mantendo fixo durante o restante do treinamento.

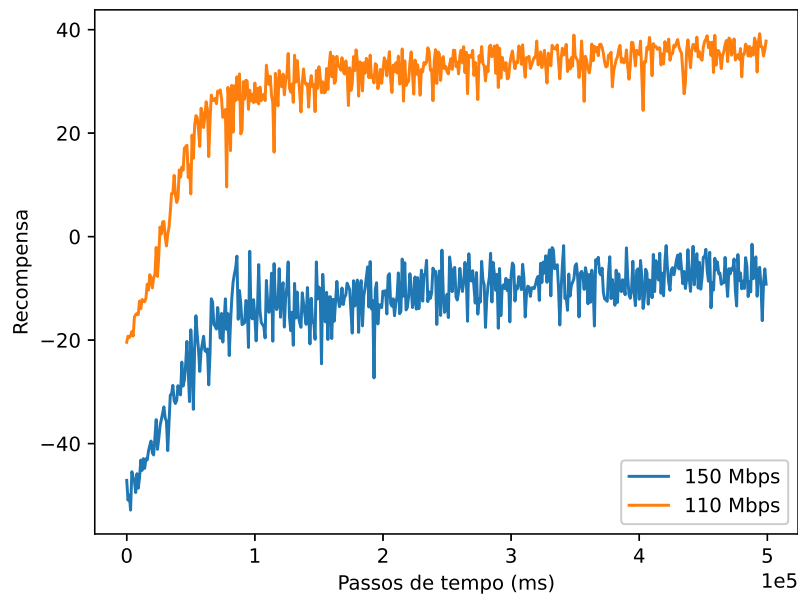


Figura 3.6: Convergência dos modelos em função do tempo de treinamento.

As Figuras 3.7, 3.8 e 3.9 apresentam as performances dos agentes escalonadores sob três métricas, em valores médios por UE: vazão, atraso nos *buffers* e perda de pacotes. Considera-se todos os usuários associados à ERB, empregando dois cenários de carga média de rede: baixo (≈ 110 Mbps) e alto (≈ 150 Mbps).

Em termos de vazão da rede, o método DRL-SRA proposto se mostra capaz de operar de forma semelhante ao agente MT, mesmo em cenário de maior carga na rede. Em cenário de menor tráfego, conforme observado na Figura 3.7(a), a convergência do modelo ocorre por volta de 1×10^5 passos de treinamento, indicando que apenas 1 segundo de 1 episódio pode ser suficiente, em que a convergência ocorre, na prática, ainda mais cedo do que fora observado na Figura 3.6. Em cenário de maior carga, visto na Figura 3.7(b), devido ao maior impacto da perda de pacotes (a capacidade do canal é

limitada) capturada pela função de recompensa (Equação 3-11), observa-se que um maior tempo de treinamento é necessário até a convergência. Adicionalmente, observa-se que, após a convergência ocorrer, o agente DRL-SRA consistentemente supera os agentes RR e PF nesta métrica, em ambos os cenários.

Com relação ao parâmetro de atraso médio no *buffer*, o método DRL-SRA proposto apresenta desempenho semelhante aos agentes RR e PF, que são os agentes com desempenho superior ao MT, conforme demonstrado nas Figuras 3.8(a) e 3.8(b). Desta forma, o agente DRL demonstra ter a habilidade de incorporar comportamentos mais elaborados do que de um agente convencional, mantendo a sua implementação simples. O agente MT exibe o maior atraso no *buffer*, notadamente mais elevado do que o observado no cenário de maior tráfego. Isto ocorre devido ao comportamento de alocação injusta imposto pelo agente MT, no qual os UEs com menor demanda, seja pela baixa eficiência espectral a qual estejam experimentando, ou pela baixa taxa de ocupação dos *buffers*, tendam a ter a alocação postergada, até que possam estar em condições que permitam oferecer uma maior taxa de dados. Embora o agente DRL-SRA apresente um desempenho levemente inferior aos agentes de melhor desempenho, no cenário de maior carga de tráfego, observa-se que se trata de uma diferença insignificante, e ilustram a adaptabilidade do agente proposto. Por fim, comportamentos semelhantes em termos de tempo de convergência dos modelos, nesta métrica, podem ser observados.

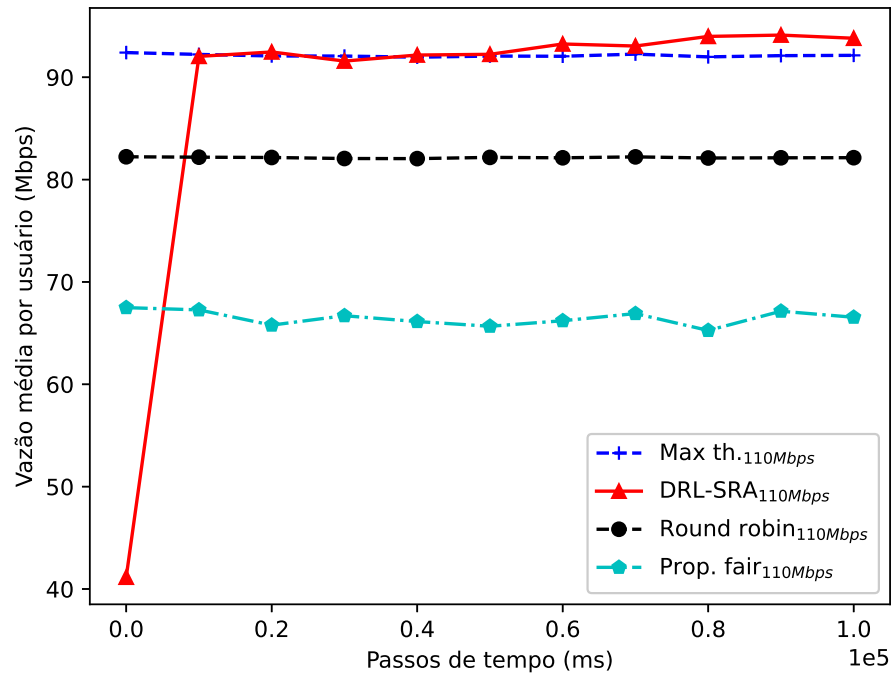
Por fim, em termos da métrica de perda de pacotes, apresentados nas Figuras 3.9(a) e 3.9(b), observa-se que o agente proposto é, novamente, capaz de apresentar resultado semelhante ao agente MT, logo que os modelos convergem. De fato, tal comportamento é esperado, dado que a perda de pacotes é fortemente correlacionada com a vazão avaliada nos cenários aqui considerados, o que é corroborado pelos desempenhos vistos nos agentes RR e PF.

3.5 Conclusão

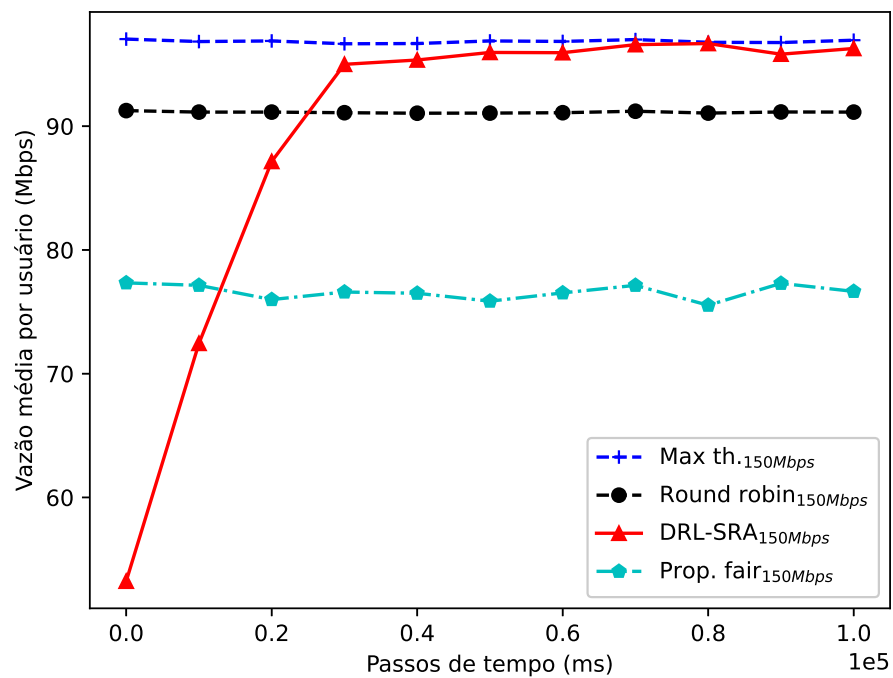
Esse capítulo apresentou uma solução baseada em DRL para o problema de alocação de recursos e escalonamento de usuários em redes MIMO multibandas. Trata-se de um agente inteligente que leva em consideração os dados dos canais e de demanda (isto é, CSI e CQI) para a tomada de decisões de escalonamento. Um ambiente para o treinamento e avaliação da proposta foi implementado, considerando tanto o subsistema de comunicação quanto o ambiente DRL para o treinamento do agente. O sistema de comunicação leva em consideração canais realistas e com consistência espacial e temporal, considerando usuários móveis. Por sua vez, o ambiente de treinamento do agente levou em consideração as soluções de prateleira descritas na Seção 2.3.

Experimentos mostram que o tempo requerido para treinar os modelos viabiliza a adoção do agente proposto em soluções que requeiram aprendizado de modo *online*, como na arquitetura proposta em [Han et al. 2020], que emprega um controlador próprio para ferramentas baseadas em ML na RAN. O agente proposto se mostrou adequado tanto para serviços de rede tolerantes ou sensíveis ao atraso [Dong et al. 2020], sem comprometer a vazão e a perda de pacotes dos seus usuários, conforme requerido para as redes 5G e de próxima geração, que requerem métodos de alocação e gerenciamento de recursos flexíveis e adaptáveis.

Embora a solução proposta não tenha sido formalizada e resolvida de maneira ótima através de um método de otimização, o próximo capítulo apresentará um problema que considera a formalização antes da proposição de uma abordagem baseada em AI/ML. Adicionalmente, os aprendizados e a experiência acumulada durante a proposição e validação da solução aqui descrita servem de base nas abordagens descritas a seguir.

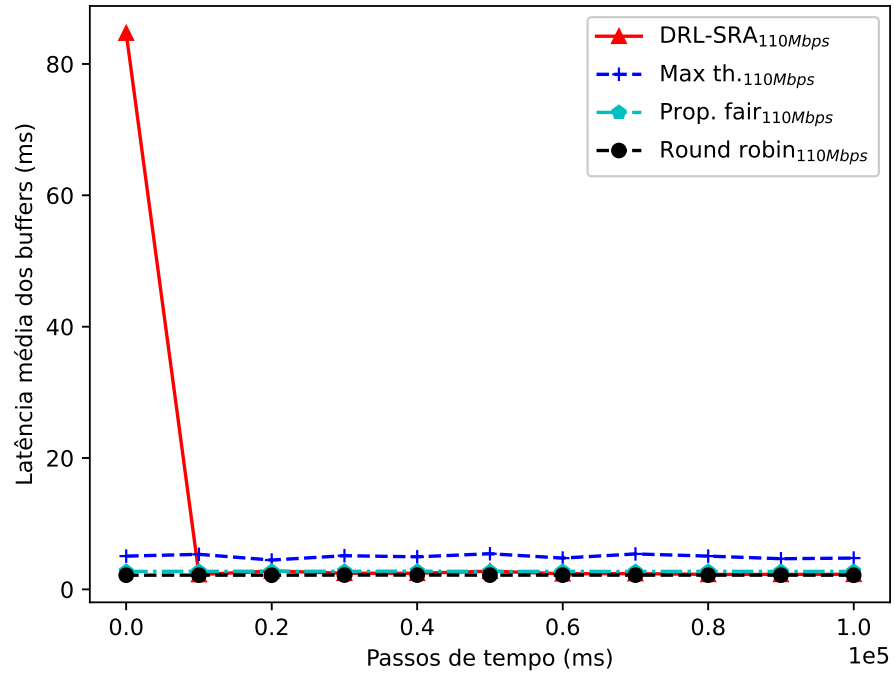


(a)

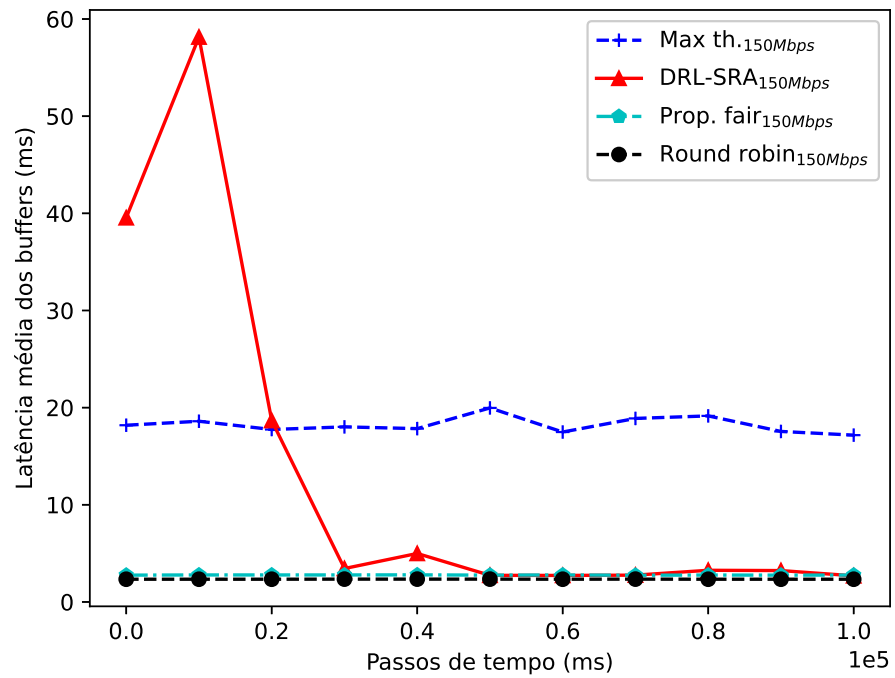


(b)

Figura 3.7: Comparando a vazão média por UE em diferentes cargas de rede: (a) ≈ 110 Mbps (b) ≈ 150 Mbps.

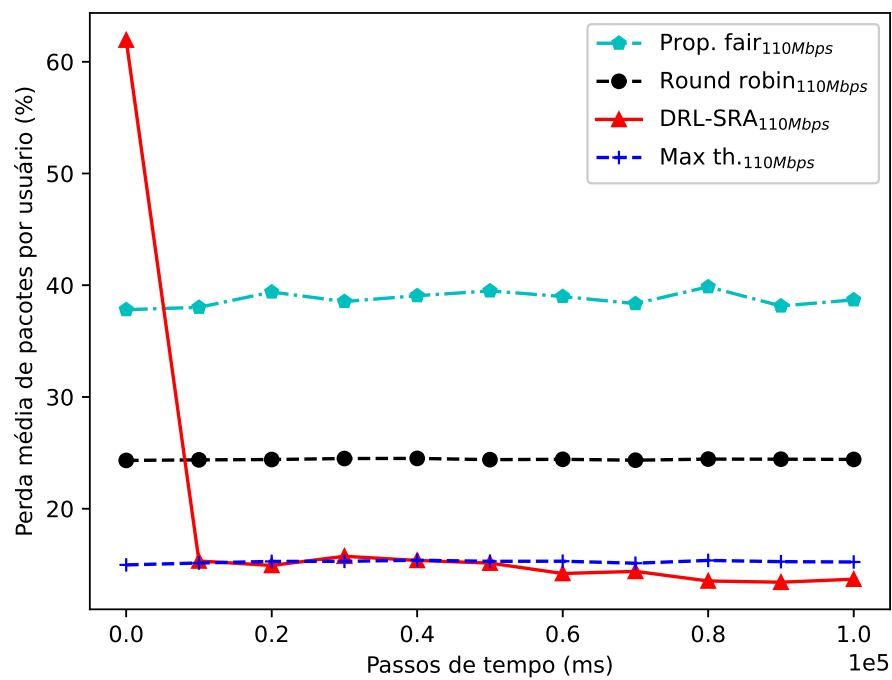


(a)

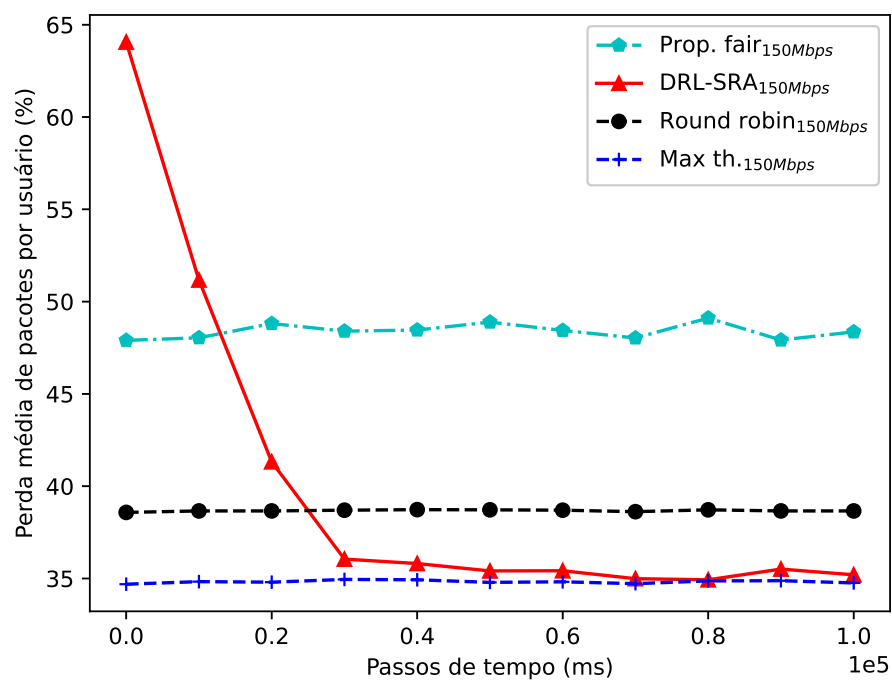


(b)

Figura 3.8: Comparando atraso médio de pacotes por UE nos *buffers* da ERB, em diferentes cargas de rede: (a) ≈ 110 Mbps (b) ≈ 150 Mbps.



(a)



(b)

Figura 3.9: Comparando a perda média de pacotes por UE em diferentes cargas de rede: (a) ≈ 110 Mbps (b) ≈ 150 Mbps.

Alocação de recursos em vRAN desagregada levando em consideração aspectos da cobertura

Neste capítulo, os problemas envolvendo os processos decisórios para a alocação de recursos e o posicionamento de VNFs em redes de acesso desagregadas, com ciência da qualidade do canal dos usuários servidos são considerados. Em tais problemas, diversos aspectos precisam ser conjuntamente observados nas soluções de posicionamento de VNFs de forma a proporcionar a agregação de funções virtuais de diferentes unidades de rádio (*Radio Units* - RUs), formando conjuntos que possam ser orquestrados com o intuito de mitigar interferências entre as células.

O capítulo apresenta duas abordagens empregadas para a solução desses problemas, resultando em propostas aderentes à arquitetura O-RAN. Na primeira, um método baseado em otimização exata é proposto, em conjunto com componentes e fluxos de controle a serem incluídos na arquitetura O-RAN. Assim, a RAN passa a ser monitorada de forma a garantir a implantação de soluções de posicionamento de VNFs de forma dinâmica, reagindo à condição dos canais e de demanda dos usuários associados. Na segunda abordagem, o problema é modelado como um MDP, seguindo da proposição de um agente baseado em DRL e de novos componentes aderentes à arquitetura O-RAN, de maneira a superar as limitações enfrentadas pela primeira abordagem.

Simulações foram realizadas, demonstrando que ambas as propostas trazem importantes ganhos em termos de redução de interferências e melhorias na eficiência espectral, conforme serão apresentados e discutidos a seguir.

4.1 Introdução

Nos últimos anos, diversos esforços de pesquisa, da indústria e das organizações de padronizações têm sido empregados visando superar as limitações impostas pelas implementações monolíticas e dependentes de hardware especializado das redes de acesso por rádio (RAN). Em paralelo, tais redes de acesso passam também a ser impactadas pelo paradigma da virtualização (NFV), de forma a habilitar o emprego de funções

virtualizadas de redes (VNFs), entre outros recursos virtualizados. Essas evoluções foram descritas em maiores detalhes na Seção 2.2. Assim, a centralização das funções de banda base em servidores virtualizados, e em nuvem, torna-se de grande interesse para os MNOs.

Ao assumir uma arquitetura virtualizada e desagregada, a vRAN passa a permitir a implementação de topologias que favorecem o aprimoramento no gerenciamento dos recursos associados, além de ampliar as capacidades para a oferta de serviços para a mitigação de interferências, entre outros. Nesse sentido, surgem iniciativas para o uso de interfaces abertas, como visto na arquitetura O-RAN [Polese et al. 2023, O-RAN Alliance 2020], descrita em maiores detalhes na Seção 2.2.1. Nessa arquitetura, os desafios de selecionar as divisões funcionais da pilha de protocolos da vRAN, para cada uma das unidades de rádio da topologia, e a escolha de onde (isto é, em qual nó da rede de acesso) posicionar cada parte dessas divisões funcionais, se tornam relevantes (conforme descrito na Seção 2.1.3).

Entretanto, as decisões que precisam ser tomadas para obter o posicionamento adequado das VNFs não são triviais. Tal problema precisa ser adequadamente solucionado de forma a atender às expectativas tanto de racionalização de custos por parte dos MNOs, quanto da qualidade dos serviços oferecidos (QoS e QoE, por exemplo). É importante salientar que a solução definida para o posicionamento de VNFs precisa respeitar as restrições impostas pelas aplicações e protocolos, além daqueles impostos pela própria topologia, como visto nas limitadas capacidades dos enlaces e dos servidores que hospedam os nós da topologia da vRAN.

Adicionalmente, duas características ampliam a complexidade da implementação e operação de uma vRAN desagregada e não podem ser negligenciadas: a flutuação na demanda das ERBs no decorrer do tempo e os aspectos relativos à qualidade dos canais dos usuários associados. De um lado, a demanda de cada ERB flutua durante o tempo, de forma que aquelas localizadas em regiões urbanas ou suburbanas podem sofrer grandes variações de demanda no decorrer de um mesmo dia, bem como no decorrer dos dias de uma semana, por exemplo. Do outro lado, os usuários também podem estar experimentando diferentes condições dos canais dentro da área de cobertura, o que pode representar uma diversidade de custos computacionais nas divisões funcionais estabelecidas. Tais características também precisam ser levadas em consideração na hora de estabelecer soluções de posicionamento de VNFs que favoreçam a mitigação de interferências, que impactam os usuários localizados mais próximos às bordas das células.

Esses aspectos precisam ser conjuntamente observados para a definição de soluções de posicionamento de VNFs. O uso de soluções estáticas, que não podem ser adaptadas no decorrer da operação da vRAN, ou aquelas que dependam da intervenção humana para qualquer alteração, podem representar tanto o desperdício de recursos, quanto

a degradação dos serviços oferecidos. Soluções estáticas sobredimensionadas representam desperdício de recursos quando a demanda é reduzida, comprometendo aspectos de eficiência energética e de custos (CAPEX e OPEX), por exemplo. Do outro lado, soluções estáticas podem se tornar subdimensionadas em caso de crescimento da demanda acima de um patamar previsto, podendo ocasionar a queda de desempenho da vRAN [Tamim, Shami e Ong 2023], principalmente devido a quebra de restrições de atraso impostas pelas divisões funcionais e até a interrupção dos serviços de processamento das funções de banda base [Bega et al. 2018].

Embora essa temática tenha sido foco de investigação recentemente, como em [Harutyunyan e Riggio 2018, Murti, Ali e Latva-aho 2021, Sarikaya e Onur 2021, Almeida et al. 2022, Chang et al. 2022, Awad et al. 2022, Amiri et al. 2023, Murti et al. 2023], nenhuma dessas propostas considera todos os aspectos aqui descritos de forma conjunta. Nesse sentido, duas abordagens foram propostas para a obtenção de soluções dinâmicas para o posicionamento de VNFs com o foco na criação de agrupamentos de RUs de maneira a habilitar a coordenação na transmissão para a mitigação de interferências entre as células, isto é, CoMP na transmissão. Ambas as abordagens são aderentes à arquitetura O-RAN, e levam em consideração aspectos da qualidade do canal dos usuários atendidos. A primeira é baseada em um método de otimização exata para a obtenção de uma nova solução de posicionamento, e a segunda abordagem baseada em um agente DRL, sendo ambas descritas a seguir.

4.2 Abordagem baseada em otimização exata

Na primeira abordagem, foi investigado como a informação da qualidade dos canais dos usuários atendidos pelas ERBs pode ser utilizado para produzir soluções que habilitem o CoMP. Isso é obtido quando as divisões funcionais da pilha de protocolos permite agregar as VNFs de diferentes RUs em um mesmo *Computing Node* (CN), de forma que as informações necessárias para o funcionamento do CoMP sejam acessadas localmente. Quando isso ocorre, um cluster de RUs é formado, de forma que o CoMP no modo *Coordinated Scheduling* (CS) [Bassoy et al. 2017, Marotta et al. 2019] na transmissão possa ser executado. Dessa forma, espera-se que a eficiência espectral, especialmente dos usuários de borda, seja melhorada, dado que as decisões de escalonamento (usuários e recursos espectrais) para a transmissão em DL de todas as RUs que compõem o cluster passam tomadas de forma coordenada. Assim, uma boa solução de posicionamento tende a ser aquela que forma o menor número de agrupamentos de RUs, sendo tais agrupamentos, consequentemente, assumindo o maior tamanho possível.

É importante ressaltar que o uso de CoMP não é restrito à vRAN desagregada implementada exclusivamente com soluções de posicionamento de VNFs que agregam

as funções estratégicas para o seu funcionamento. Entretanto, o seu emprego nesses cenários surge como uma relevante alternativa para superar um importante limitador na composição de agrupamentos de maiores proporções. Em seu formato original, a operação do CoMP CS requer que as ERBs do cluster encaminhem dados de CSI para a ERB âncora (aquela que irá coordenar as decisões de escalonamento), o que impõe limitações no tamanho dos *clusters* devido a sobrecarga gerada no *backhaul* pela troca de informações requerida [Bassoy et al. 2017, Marotta et al. 2019, Abdelhakam et al. 2023].

4.2.1 O modelo do sistema

O modelo do sistema adotado considera tanto as padronizações *Third Generation Partnership Project* (3GPP) para uma vRAN 5G [3GPP-TR21.916 2020] quanto as especificações da arquitetura O-RAN e os seus cenários de implantação [O-RAN Alliance 2020, O-RAN Alliance 2021, O-RAN Alliance 2022]. Uma visão geral pode ser observada na Figura 4.1, que apresenta as células formadas pelas ERBs, cada uma possuindo a RU (representada pelo triângulo), os nós que compõem a vRAN desagregada, implementados considerando os nós da O-RAN, e os enlaces que compõem a rede de transporte, formando uma topologia hierárquica. Sob a perspectiva das células e da área de cobertura, trata-se de um conjunto de *sites* de células de três setores (células hexagonais), operando com fator de reuso de frequência 1 [Kimura e Seki 2012, Yassin et al. 2017]. Dessa forma, há degradação na qualidade do sinal devido às interferências entre as células (ICI), impactando em maior nível os usuários localizados mais próximos às bordas. É importante ressaltar que outras técnicas de *Inter-Cell Interference Coordination* (ICIC) estão fora do escopo do trabalho.

No domínio da vRAN (parte superior da Figura 4.1), a topologia é composta de: i) um conjunto $\mathcal{B} = \{b_1, b_2, \dots, b_{|\mathcal{B}|}\}$ de RUs, cada uma contendo as funcionalidades de RF requeridas para atender aos serviços das demais camadas da pilha de protocolos; ii) um conjunto $\mathcal{C} = \{c_1, c_2, \dots, c_{|\mathcal{C}|}\}$ de CRs em que as VNFs das divisões funcionais selecionadas podem ser implantadas, considerando c_m^{Proc} a capacidade de processamento disponível; e iii) um conjunto $\mathcal{T} = \{t_1, t_2, \dots, t_{|\mathcal{T}|}\}$ de nós de transporte, que podem se conectar ao núcleo da rede, a quaisquer RUs e/ou CRs, ou entre si. É importante ressaltar que embora cada CR também possa ser representado através de suas capacidades de memória RAM e armazenamento, a capacidade de processamento tende a se esgotar antes dos demais, o que pode gerar impactos de atraso da rede e até interrupção dos serviços de processamento das funções de banda base.

Através do uso da rede de transporte da topologia, o fluxo de dados tem como fonte o núcleo da rede, sendo o destino as RUs, quando operando em DL, ou o inverso em *uplink* (UL). Embora o trabalho considere a operação em DL, ele também pode

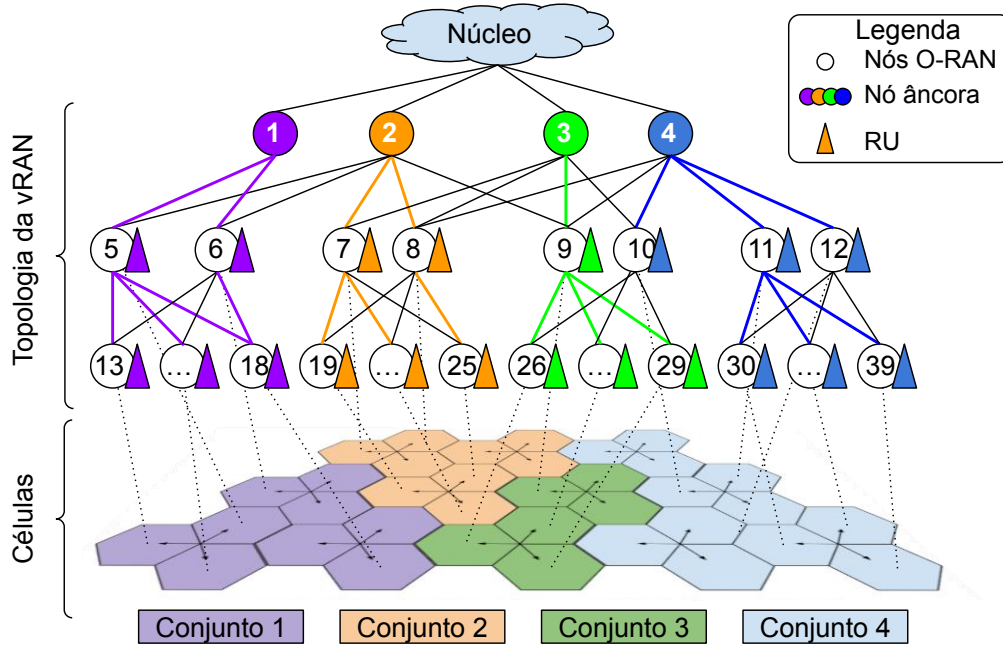


Figura 4.1: Visão geral da topologia da vRAN desagregada considerada, demonstrando o domínio da topologia, o domínio das células e clusters criados.

ser adaptado para considerar UL, realizando as devidas alterações nos componentes do sistema. O fluxo da rede precisa seguir um caminho $p \in \mathcal{P}_l$ dos k caminhos mais curtos entre a fonte e a RU $b_l \in \mathcal{B}$ de destino. Cada caminho p é formado por até três subcaminhos (ou segmentos da rede de transporte), o *backhaul* (p_{BH}), o *midhaul* (p_{MH}) e o *fronthaul* (p_{FH}), de forma que pelo menos um desses subcaminhos seja não nulo.

A pilha de protocolos da vRAN é considerada como implementadas em VNFs (exceto as funções de RF, conforme descrito anteriormente), enumeradas em ordem crescente, iniciando-se pela PHY Low como f_1 e terminando na RRC, como f_8 . Assim, fica definido $\mathcal{F} = \{f_1, \dots, f_8\}$ como o conjunto de VNFs da vRAN desagregada que precisa ser adequadamente alocadas. Como forma de garantir isso, emprega-se o conjunto $\mathcal{D} = \{D_1, D_2, \dots, D_{|\mathcal{D}|}\}$ de *Disaggregated RAN Combinations* (DRCs) adotadas pela indústria, em linha com o trabalho em [Morais et al. 2021]. O conjunto \mathcal{D} é demonstrado na Figura 4.2. Ela apresenta tanto as VNFs contidas em \mathcal{F} (representadas nas caixas em branco), bem como as divisões funcionais (representadas nas caixas coloridas). Assim, ao se definir o uso de um determinado DRC, ficam automaticamente definidas as divisões funcionais a serem consideradas, e em quais tipos de nós da O-RAN cada parte da divisão funcional poderá ser implantada. Vale ressaltar que são considerados no presente trabalho apenas os DRCs que habilitam o uso de CoMP.

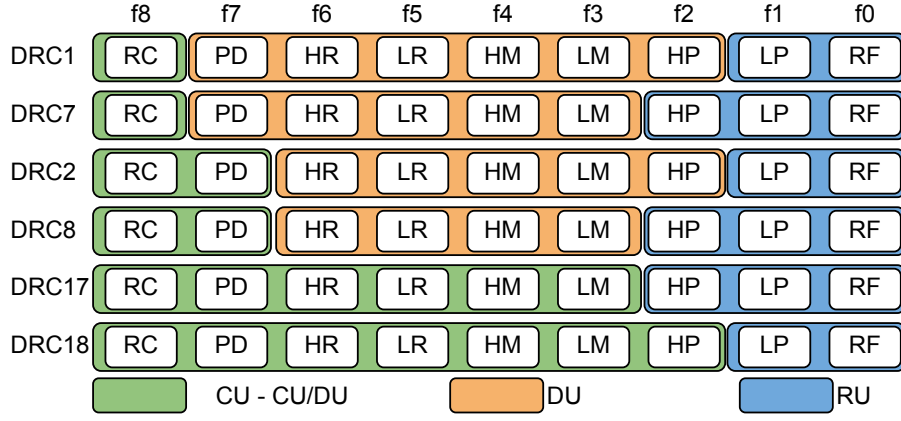


Figura 4.2: DRCs que habilitam o CoMP.

Modelo para computar o consumo de recursos computacionais na vRAN

Dada uma corrente solução de posicionamento de VNFs implantada na vRAN, podemos calcular o custo computacional instantâneo, isto é, o consumo atual dos recursos computacionais dos CNs envolvidos, em termos de *giga operations per second* (GOPS), para todas as VNFs em cada divisão funcional $s \in |S|$ alocado no CN $c_m \in |C|$, durante o TTI t , usando a seguinte equação (adaptada de [Sarıkaya e Onur 2021]):

$$\mathcal{Q}_{s,t}^{c_m} = \sum_{f=1}^F \sum_{u=1}^U \alpha_{s,f} \left(3a + a^2 + \frac{o_u r_u l}{3} \right) \frac{\Gamma_u}{5}, \quad (4-1)$$

onde a define o total de antenas de transmissão na RU, o_u e r_u são a máxima ordem de modulação e a taxa de codificação do usuário u , respectivamente, l é o número de camadas MIMO, $\Gamma_u \in \mathbb{Z}^+$ é o total de RBs alocados para o usuário u , e $\alpha_{s,f} \in \mathbb{R}^+$ é um fator de escala que normaliza o custo computacional da VNF f pertencente à divisão funcional s implantada e sendo executada em cada nó da vRAN. De acordo com a Equação 4-1, o MCS empregado para cada usuário (a partir das medições de CQI reportadas e do mapeamento SNR/CQI-MCS [Bakri 2021, Thyagarajan et al. 2021]), bem como o modo de transmissão em uso, são diretamente proporcionais ao consumo dos recursos computacionais pela divisão funcional alocada. Portanto, detalhes da cobertura das células não podem ser negligenciados durante o planejamento e administração da vRAN desagregada.

4.2.2 Formulação do problema

Para formular o problema, é definida a variável de decisão $x_l^{p,r} \in \{0,1\}$, que representa a configuração da alocação para servir a RU $b_l \in \mathcal{B}$ com o DRC $D_r \in \mathcal{D}$, considerando o caminho $p \in \mathcal{P}_l$. Assim, o objetivo é escolher as divisões funcionais, em quais CRs eles serão alocados, e quais caminhos serão utilizados, para cada ERB da

topologia, enquanto busca minimizar o número de clusters de ERBs (consequentemente, obtendo *clusters* maiores). O conjunto dessas decisões pode ser chamado de solução de posicionamento, em que tal solução precisa levar em consideração os requisitos de latência e vazão das quebras funcionais, além das capacidades de processamento dos CRs e dos enlaces envolvidos. A Tabela 4.1 sumariza os parâmetros e variáveis utilizadas no modelo de sistema, na formulação do problema e na solução proposta. Portanto, para representar o objetivo, definimos a seguinte equação:

$$\Phi = \sum_{c_m \in \mathcal{C}} \left[\frac{\sum_{b_l \in \mathcal{B}} \sum_{p \in \mathcal{P}_l} \sum_{D_r \in \mathcal{D}} (x_l^{p,r} \zeta(b_l, c_m))}{|\mathcal{B}|} \right]. \quad (4-2)$$

A Equação 4-2 permite calcular o número de clusters de RUs criados pela solução, sendo que $\zeta = \{0, 1\}$ indica que a RU b_l é parte do cluster ancorado em c_m . Observe que a parte do numerador representa o tamanho do cluster em c_m . No entanto, a informação que procuramos é se existe ou não um cluster em c_m . Neste caso, usamos a função teto para obter essas informações usando $|\mathcal{B}|$ como denominador, isto é, o tamanho máximo possível para o tamanho de um cluster, que é o maior número possível no numerador.

Como já descrito anteriormente, o trabalho leva em consideração a área de cobertura da RU, buscando garantir melhorias de desempenho na medida em que emprega um controle de interferências mediante o uso de CoMP. Para tal, definimos o subconjunto de RUs $\mathcal{B}_l \subseteq \mathcal{B}$, que mapeia todas as vizinhanças entre as células. A relação de vizinhança é realizada pelo grafo $G^* = \{V^*, E^*\}$, que é uma das entradas do modelo. O grafo de vizinhança possui como vértices as RUs, sendo que as arestas representam existência de interferências entre as elas. É importante ressaltar que não são considerados os níveis de interferência nessa abordagem. O mapeamento da vizinhança serve para definir clusters de forma coesa, considerando a área de cobertura das RUs. Assim, as RUs de um cluster devem gerar um subgrafo conectado do grafo G^* , ou seja, a área de cobertura de um cluster deve ser contínua. Isso é representado pelas seguintes equações:

$$x_l^{p,r} \zeta(b_l, c_m) \leq \sum_{b_j \in \mathcal{B}_l} x_j^{p,r} \zeta(b_j, c_m), \quad \forall c_m \in \mathcal{C}, b_l \in \mathcal{B}, p \in \mathcal{P}_l, r \in \mathcal{D} \quad (4-3)$$

$$\zeta(b_i, c_m) \zeta(b_l, c_m) \leq P(b_i, b_l, c_m), \quad \forall b_i, b_l \in \mathcal{B}, c_m \in \mathcal{C}. \quad (4-4)$$

A Equação 4-3 define que todas as RUs pertencentes a um *cluster* tenham, ao menos, uma RU vizinha (exceto em casos de D-RAN). Além disso, a Equação 4-4 estabelece que cada par de RUs do *cluster* precisa formar um subgrafo em G^* .

Para definir as restrições do modelo com relação aos recursos computacionais e de enlace, bem como sobre os requisitos exigidos pelas divisões funcionais, são consideradas as formulações apresentadas em [Morais et al. 2022]. Assim, as equações aqui consideradas são apresentadas abaixo e discutidas na sequência.

$$\sum_{p \in \mathcal{P}_l} \sum_{D_r \in \mathcal{D}} x_l^{p,r} = 1, \quad \forall b_l \in \mathcal{B} \quad (4-5)$$

$$\sum_{D_r \in \mathcal{D}} \sum_{b_l \in \mathcal{B}} \sum_{p \in \mathcal{P}_l} [x_l^{p,r} (y_{e_{ij}}^{p_{Bh}} \alpha_{Bh}^r + y_{e_{ij}}^{p_{Mh}} \alpha_{Mh}^r + y_{e_{ij}}^{p_{Fh}} \alpha_{Fh}^r)] \leq e_{ij}^{Cap}, \quad \forall e_{ij} \in \mathcal{E} \quad (4-6)$$

$$\sum_{e_{ij} \in \mathcal{E}} x_l^{p,r} y_{e_{ij}}^{p_{Bh}} e_{ij}^{Lat} \leq \beta_{Bh}^r, \quad \forall b_l \in \mathcal{B}, p \in \mathcal{P}_l, D_r \in \mathcal{D} \quad (4-7)$$

$$\sum_{e_{ij} \in \mathcal{E}} x_l^{p,r} y_{e_{ij}}^{p_{Mh}} e_{ij}^{Lat} \leq \beta_{Mh}^r, \quad \forall b_l \in \mathcal{B}, p \in \mathcal{P}_l, D_r \in \mathcal{D} \quad (4-8)$$

$$\sum_{e_{ij} \in \mathcal{E}} x_l^{p,r} y_{e_{ij}}^{p_{Fh}} e_{ij}^{Lat} \leq \beta_{Fh}^r, \quad \forall b_l \in \mathcal{B}, p \in \mathcal{P}_l, D_r \in \mathcal{D} \quad (4-9)$$

$$\sum_{f_s \in \mathcal{F}} \sum_{b_l \in \mathcal{B}} \sum_{D_r \in \mathcal{D}} \sum_{p \in \mathcal{P}_l} x_l^{p,r} u_m^p M(c_m, f_s, b_l) \gamma_m^s \leq c_m^{Proc}, \quad \forall c_m \in \mathcal{C}. \quad (4-10)$$

A Equação (4-5) estabelece a restrição de que apenas um caminho e configuração DRC pode ser definida para uma determinada RU. A Equação (4-6) determina que a capacidade de transmissão de cada enlace não pode ser excedida. As Equações (4-7) a (4-9) representam as restrições para os requisitos das divisões funcionais com relação às latências para o *backhaul*, o *midhaul* e o *fronthaul*, respectivamente. Por fim, a Equação (4-10) estabelece que a capacidade de processamento dos CRs não pode ser excedida. Assim, o modelo de otimização pode ser representado por:

minimizar Φ

sujeito a:

Equações (4 – 3) a (4 – 10)

4.2.3 A solução proposta

Em uma visão geral, para que o modelo de otimização descrito na Seção 4.2.2 possa estabelecer qual deverá ser a solução mais adequada para o posicionamento de VNFs, ele precisará ter acesso às informações necessárias como entrada do problema. Elas devem ser obtidas da topologia, o que inclui as informações sobre as capacidades dos enlaces e dos recursos computacionais envolvidos. Entretanto, é importante ressaltar que o trabalho objetiva a obtenção de soluções de posicionamento de forma dinâmica, em que

Tabela 4.1: Elementos do modelo.

Parâmetro	Definição
\mathcal{B}	Conjunto de ERBs (RUs)
\mathcal{B}_l	Subconjunto de RUs vizinhas à RU b_l
\mathcal{C}	Conjunto de CRs
\mathcal{T}	Conjuto dos nós de transporte
\mathcal{E}	Conjunto de enlaces
e_{ij}^{Cap}	capacidade de transmissão do enlace e_{ij}
e_{ij}^{Lat}	Latência estimada do enlace e_{ij}
\mathcal{P}_l	Conjunto dos k caminhos mais curtos
\mathcal{F}	Conjunto de VNFs
\mathcal{D}	Conjunto de DRCs
u_p^m	Indica se o CR $c_m \in \mathcal{C}$ faz parte do caminho $p \in \mathcal{P}_l$
$M(c_m, f_s, b_l)$	Indica se o CR $c_m \in \mathcal{C}$ está executando $f_s \in \mathcal{F}$ de $b_l \in \mathcal{B}$
$y_{e_{ij}}^{PBH}$	Indica se e_{ij} é parte do <i>backhaul</i>
$y_{e_{ij}}^{PMH}$	Indica se e_{ij} é parte do <i>midhaul</i>
$y_{e_{ij}}^{PFH}$	Indica se e_{ij} é parte do <i>fronthaul</i>
α_{BH}^r	A demanda associada (<i>bitrate</i>) no <i>backhaul</i> para o $D_r \in \mathcal{D}$
α_{MH}^r	A demanda associada (<i>bitrate</i>) no <i>midhaul</i> para o $D_r \in \mathcal{D}$
α_{FH}^r	A demanda associada (<i>bitrate</i>) no <i>fronthaul</i> para o $D_r \in \mathcal{D}$
β_{BH}^r	Latência máxima tolerada no <i>backhaul</i> para o $D_r \in \mathcal{D}$
β_{MH}^r	Latência máxima tolerada no <i>midhaul</i> para o $D_r \in \mathcal{D}$
β_{FH}^r	Latência máxima tolerada no <i>fronthaul</i> para o $D_r \in \mathcal{D}$
γ_m^s	A demanda computacional de $f_s \in \mathcal{F}$
c_m^{Proc}	Capacidade de processamento de $c_m \in \mathcal{C}$
$\zeta(b_l, c_m)$	Indica se o RU b_l está em um cluster ancorado em c_m
$P(b_i, b_j, c_m)$	Indica se existe um caminho de b_i até b_j passando pelos vértices do cluster em c_m .
Variável	Definição
$x_l^{p,r} = \{0, 1\}$	Representa qual par de $p \in \mathcal{P}_l$ e $D_r \in \mathcal{D}$ é selecionado para servir $b_l \in \mathcal{B}$

a vRAN poderá se adaptar em tempo de execução, conforme já descrito anteriormente. Nesse sentido, a solução proposta apresenta os recursos necessários a serem incluídos na arquitetura O-RAN, de forma a prover o ambiente que suporte tal objetivo.

A Figura 4.3 apresenta uma visão geral da arquitetura O-RAN modificada com a inclusão de novos componentes. Conforme já descrito na Seção 2.2.1, a arquitetura O-RAN possui o *non-real-time* RIC (non-RT RIC) e o *near-real-time* RIC (near-RT RIC). O primeiro compreende o arcabouço non-RT RIC, que expõe os serviços dos aplicativos rApps, enquanto que o segundo suporta vários aplicativos de terceiros (xApps). Os rApps e os xApps são aplicações modulares que podem fornecer serviços especializados, suportando tanto as tarefas de gerenciamento, quanto as adaptações necessárias para melhorias de vRAN. Neste sentido, a solução proposta envolve a inclusão de dois novos

componentes à arquitetura da rede: o rApp *VNF Placement* (VNF-P) e o xApp *VNF Controller* (VNF-C).

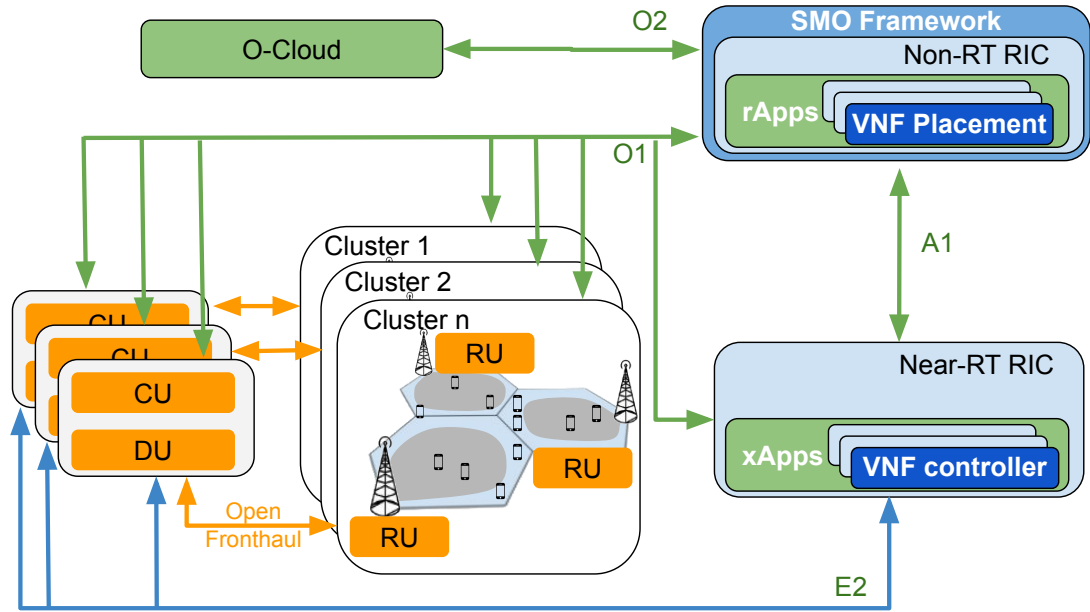


Figura 4.3: A visão geral da solução proposta, com a inclusão dos componentes na arquitetura O-RAN: o rApp e o xApp.

Cabe ao VNF-P todos os processos necessários para estabelecer uma nova solução de posicionamento de VNFs, isto é, ele é o componente responsável por executar o modelo de otimização estabelecido. O VNF-P é executado a partir de uma chamada requisitada pelo VNF-C, através da interface A1. Assim que a chamada é recebida, o VNF-P poderá obter informações sobre o estado da topologia disponíveis no arcabouço non-RT RIC, via interface O1, que integra-o diretamente aos nós da O-RAN.

Por sua vez, o VNF-C é responsável por monitorar a topologia em termos da utilização da capacidade computacional dos CRs, bem como do CQI médio reportado por cada usuário associado, considerando todas as ERBs. O monitoramento ocorre considerando dois processos independentes, sendo o primeiro deles o monitoramento aperiódico, e o segundo o monitoramento periódico.

De fato, o monitoramento aperiódico pode ser considerado um gatilho, acionado sempre que qualquer um dos CRs da topologia tiver esgotado a sua capacidade de processamento. Nesse momento, uma chamada ao VNF-P é realizada, dado que uma nova solução de posicionamento precisa ser estabelecida. Entretanto, como o esgotamento da capacidade de processamento pode levar aos efeitos danosos ao desempenho da vRAN [Bega et al. 2018], um procedimento de contenção é proposto, sendo também de responsabilidade do VNF-C. Nessa situação, o sistema emprega o MCS *downgrade* [Seytnazarov e Kim 2013, Bega et al. 2018] nas ERBs que estejam contribuindo

em maior grau para o consumo de recursos computacionais pelo cluster. Assim, a máxima ordem de MCS que pode ser alocada aos usuários da ERB passa a ser reduzida, alterando-se o parâmetro MCS_{max} . O procedimento continua a ser executado nas ERBs subsequentes do cluster (em ordem decrescente no consumo de processamento por ERB), até que a demanda de processamento caiba na capacidade dos CRs que hospedam as VNFs desse *cluster*.

O motivo para isso é que, segundo a Equação 4-1 adotada para calcular o custo computacional, o MCS alocado para os usuários contribui significativamente no custo computacional. Entretanto, isso também afetará o desempenho da ERB, principalmente em termos de vazão agregada. Porém, é importante ressaltar que isso tende a afetar, em menor grau, os usuários de borda, dado que eles já encontram-se em situações de pior qualidade de canal, o que por já limita o máximo MCS alocado. Em um outro ponto relevante, tal situação é transitória, e perdura somente até a implantação da nova solução de posicionamento. Ainda, caso ocorra nova alteração que reduza o consumo de processamento requerido pelo cluster, durante a espera pela nova solução de posicionamento, o VNF-C executa o procedimento inverso, isto é, o MCS *upgrade*.

No monitoramento periódico, o VNF-C monitora o estado da topologia a cada instante de tempo transcorrido na operação da vRAN, definido por um Δt relativo à quantidade de TTIs. O monitoramento consiste na obtenção dos parâmetros necessários para a entrada do modelo de otimização. Assim, uma chamada ao VNF-P é realizada, e uma nova solução de posicionamento será obtida. Entretanto, o objetivo agora é encontrar uma nova solução que seja mais adequada à situação da rede, considerando a última janela de observação. Por exemplo, transcorrido Δt TTIs, os CQIs médios reportados pelos usuários nesse período, para todas as ERBs, podem ser considerados para identificar possibilidades de aumentar o tamanho de um determinado cluster, caso o custo de processamento tenha reduzido. O mesmo pode ocorrer no caso contrário, ensejando a criação de novos clusters de ERBs.

É importante ressaltar que os dois processos são executados paralelamente e de forma independente. Nos dois casos, quando ocorrem chamadas ao VNF-P, elas são feitas de forma assíncrona. O motivo para tal decisão encontra-se na diferença de escalas de tempo entre o non-RT RIC e o near-RT RIC. Assim, o VNF-C pode realizar uma chamada assíncrona ao VNF-P, e continuar a desempenhar os processos de monitoramento, bem como iniciar o procedimento de contenção, caso seja necessário. Assim, assegura-se tanto a operação contínua da vRAN, quanto a sua adaptação dinâmica às condições da cobertura das células. Além disso, trata-se de uma arquitetura flexível, em que alterações futuras tanto no VNF-P quanto no VNF-C podem ser incluídas, visto que os dois componentes possuem um bom nível de desacoplamento.

Assim que o VNF-P é chamado, ele obtém as entradas para o problema, con-

forme descrito no modelo de otimização formulado na Seção 4.2.2, e inicia a busca pela solução ótima de posicionamento das VNFs. Assim que ela é definida, ela é imediatamente implantada na vRAN se, e somente se, for adequada à situação atual. Por exemplo, caso a nova solução de posicionamento estabeleça uma menor quantidade de clusters de ERBs (isto é, com clusters maiores), ela não pode ser implantada se ela demandar custos impeditivos de processamento nos CRs, considerando a atual demanda. Isso se justifica pois, durante a busca pela solução ótima, a situação da vRAN poderá ter sofrido alterações significativas, de forma que a nova solução não permaneça adequada.

4.2.4 Resultados e discussões

Para comparar o desempenho da proposta, foram realizadas simulações considerando um ambiente bastante detalhado, composto por um sistema de comunicação, responsável pela simulação da rede, incluindo todo o ambiente de propagação sem fio, e o sistema responsável pelo método de otimização, o que inclui o aparato necessário para a obtenção das entradas do problema, e o próprio *solver*. O sistema de comunicação foi implementado usando o simulador de nível de sistema Vienna 5G Link-Level [Pratschner et al. 2018, Müller et al. 2018], em sua versão para *downlink*, descrito em maiores detalhes na Seção 2.4.1. O simulador recebeu customizações para permitir tanto o acesso aos dados necessários para a implementação dos demais componentes, quanto a inclusão dos recursos para a contabilização de TTIs (isto é, estimar o Δt transcorrido) e processamento da Equação (4-1) para o custo computacional.

O xApp VNF-C foi implementado em Matlab[®], devido a facilidade para integração com o simulador de Vienna. Assim, o componente proposto pode ler as tabelas de relatório das medições de CQI do simulador do sistema de comunicação (e tudo relacionado ao mapeamento CQI-MCS), e qualquer outro parâmetro necessário para o modelo de otimização. Assim, é possível alterar, dinamicamente, a política de alocação de MCS do simulador de Vienna, de forma a permitir a definição do parâmetro MCS_{max} empregado no MCS *downgrade* e/ou *upgrade*, além de obter o acesso às contagens de TTIs e os resultados de custo de processamento.

Para emular a implantação de uma solução de posicionamento de VNFs, os clusters de ERBs são definidos diretamente na configuração da simulação. Assim, se o DRC escolhido permitir habilitar o CoMP, a simulação pode ser configurada para representar o cluster conforme definido pela solução de posicionamento, e a aplicação de CoMP no simulador é habilitada para as RUs que compõem o cluster. Além disso, a configuração estabelecida para a solução de posicionamento também é considerada para efeitos de contabilização do custo de processamento dos CRs envolvidos.

Por fim, o rApp VNF-P foi implementado em Python, permitindo a integração

o *solver* CPLEX[®]. Assim, o método de otimização descrito na Seção 4.2.2 pode ser representado e executado. A abordagem adotada para emular as chamadas recebidas de qualquer dos dois processos de monitoramento do VNF-C envolve a escrita das informações requeridas pelo problema em arquivos, que possam ser lidos pelo VNF-P.

Para verificar o desempenho da rede em relação a soluções de posicionamento VNFs com diferentes números de clusters de ERBs, um primeiro experimento foi desenvolvido. Nele, a métrica observada é a eficiência espectral por uso do canal (*bits/cu*), observada tanto para os usuários de borda (*Edge UEs*) e para os usuários fora da borda (*Non-edge UEs*). O primeiro caso do experimento envolve apenas ERBs definidas como D-RANs, em que CoMP não é habilitado. Os demais casos do experimento envolvem diferentes tamanhos de clusters, definido pelo valor de c . Assim, ao usar um número maior de clusters (por exemplo, $c=12$), menor será o tamanho de cada um, ou o contrário no caso em que c for menor (por exemplo, $c=2$). Quando existe apenas 1 cluster ($c=1$), o CoMP é global, no qual todas as ERBs da vRAN cooperam.

A topologia considerada tem 21 RUs (representando as ERBs), e 25 CRs, em uma organização hierárquica. No primeiro experimento, as restrições foram desconsideradas, com foco na observação do desempenho perante qualquer solução de posicionamento definida. Os demais parâmetros e configurações do simulador são listados na Tabela 4.2. É importante ressaltar que tratam-se de parâmetros e configurações do simulador, que não sofreram alterações para os experimentos. Foram realizadas diversas simulações, nas quais cada uma considera usuários posicionados aleatoriamente (para cada célula), segundo uma distribuição uniforme, com aproximadamente 20% dos usuários na borda das células.

Tabela 4.2: Parâmetros considerados nas simulações.

Parâmetro/configuração	Valor
Antenas (Tx/Rx)	4 / 2
Modelo de canal	ITU-R Pedestrian-A
Distância entre as ERBs	600 m
Número de células	21
Usuários por célula	4
Velocidade dos usuários	8 Km/h
Frequência	2.14 GHz
Bandwidth	20 MHz
MIMO tx mode	Closed Loop Spatial Multiplexing
Modelo de tráfego	Full buffer
MIMO layers	2
Escalonador	Round Robin

A Figura 4.4 apresenta a os resultados para esse experimento. Como pode ser observado, qualquer uma das soluções que realizam o agrupamento de ERBs, mediante

um adequado posicionamento das VNFs de forma a possibilitar a utilização de CoMP, tendem a melhorar a eficiência espectral de toda a rede, de forma diretamente proporcional ao tamanho do cluster. Embora a implementação de soluções que gerem clusters muito grandes (por exemplo, $c=1$) possa ser inviável em topologias reais, pois devem exigir enlaces e CRs com capacidades elevadas, as soluções viáveis provam ser capazes de melhorar o desempenho da rede como um todo. Comparada com a solução No-CoMP, a solução com 12 clusters ($c=12$) apresenta um aumento médio de 36,3% na eficiência espectral dos usuários de borda e de 17,4% para os usuários não-borda. Quando $c = 2$, os usuários de borda podem ter um desempenho 110% melhor e 41,5% melhor desempenho para os demais usuários.

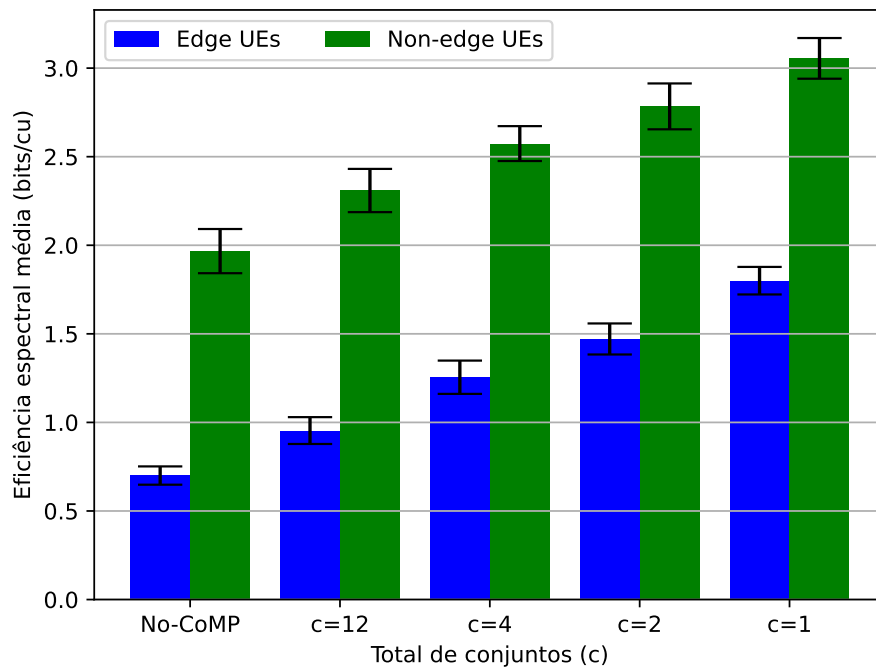


Figura 4.4: Experimento 1: a eficiência espectral em função do total de clusters de ERBs, considerando os usuários de borda (em azul) e de centro (em verde).

Em um segundo experimento, buscou-se avaliar a performance da proposta, considerando a operação dos componentes propostos. Assim, o monitoramento da vRAN é desempenhado, em que se observa a demanda média que cada ERB está recebendo em um determinado momento, conforme já descrito. Assim, a simulação passa a considerar as restrições ignoradas no experimento anterior. A mesma topologia foi considerada, bem como os parâmetros e configurações descritos. Para simular a operação do sistema, uma simulação é sempre iniciada considerando apenas D-RANs, de forma que CoMP não esteja habilitado. Devido ao monitoramento periódico ($\Delta t = 100$ TTIs), ou a qualquer momento quando acionado por monitoramento aperiódico, o VNF-C calcula as métricas

da rede e solicita uma nova solução sempre que necessário, conforme descrito na Seção 4.2.3. Para simplificar, assumiu-se que a solução é entregue pelo VNF-P antes do próximo monitoramento periódico. Na verdade, embora o tempo necessário para obter a solução ótima não possa ser sincronizado, esta abordagem procura apenas simplificar a experimento, sem perda de generalidade.

A Figura 4.5 apresenta os valores médios obtidos em diversas rodadas do experimento, considerando a vazão média por usuário, durante a execução do sistema. Em cada simulação, 4 usuários são colocados aleatoriamente em cada ERB, de acordo com uma distribuição uniforme. Os usuários se movem a uma velocidade constante, em linha reta, assumindo uma direção aleatória. Se um usuário entrar na área de cobertura de outra ERB, o *handover* é feito imediatamente. Novamente, *No-CoMP* indica a execução em que CoMP não está habilitado, e *Dinâmico* indica o uso da abordagem proposta, de forma que o tamanho dos clusters está continuamente sendo adaptado durante a execução da vRAN. As linhas sólidas representam os usuários de borda, e as tracejadas representam os demais usuários.

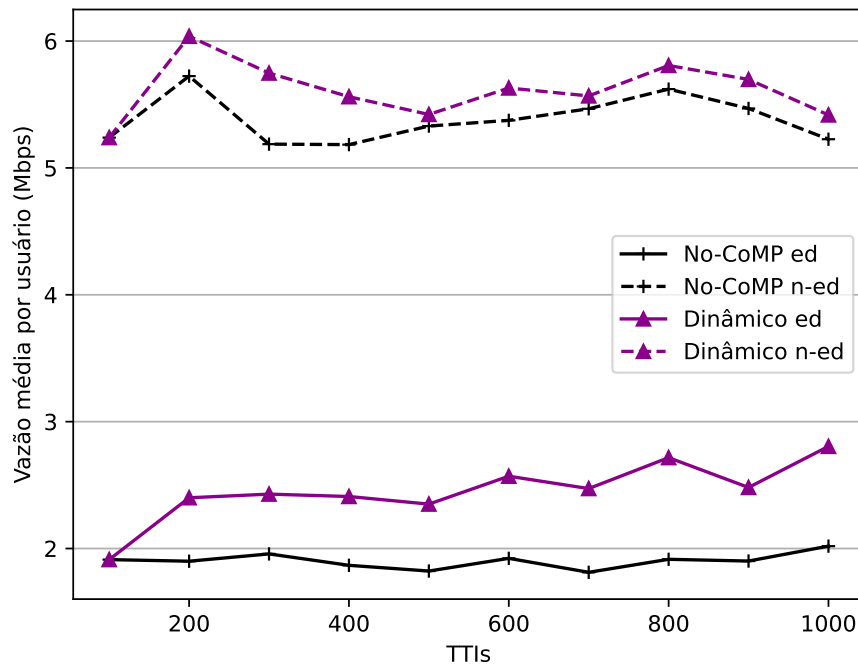


Figura 4.5: Experimento 2: Vazão média por usuário ao longo do tempo enquanto os clusters de ERBs são reconfigurados dinamicamente.

Pode ser observado que a proposta é capaz de melhorar o desempenho da rede como um todo. Conforme verificado no experimento anterior, o CoMP otimiza a eficiência espectral da rede, sem comprometer significativamente os usuários fora da borda. Ao tomar decisões sobre o tamanho dos clusters de ERB em paralelo ao sistema

em execução, e respeitando tanto a qualidade dos canais dos UEs ao longo do tempo, quanto as restrições do problema, o sistema pode garantir o tamanho adequado do cluster, reduzindo o impacto da interferência. Como pode ser observado, a abordagem proposta permite que os usuários de borda apresentem um desempenho médio 73% maior, e 26% maior para os demais usuários.

4.3 Abordagem baseada em otimização não exata com AI/ML

Em uma segunda abordagem, buscou-se estender a investigação realizada na primeira abordagem. Assim, a definição dinâmica de soluções de posicionamento de VNFs, considerando diversos aspectos conjuntamente, continua entre os objetivos, buscando as melhorias já descritas anteriormente. Entretanto, embora a primeira abordagem tenha estabelecido os fundamentos e componentes iniciais para atingir tal objetivo, ela pode incorrer nas limitações envolvendo o uso de métodos de otimização exata.

Assim, para topologias de grandes dimensões, a definição de novas soluções pode ser proibitiva, na escala de tempo exigida pelo problema. Assim, o emprego de métodos de otimização não exata surge como uma possibilidade. Adicionalmente, um entendimento mais aprofundado da arquitetura O-RAN levou à necessidade da proposição de novos componentes para lidar com esse problema, e exigindo uma abordagem adequada à essa arquitetura alterada.

Este tema tem chamado a atenção nos últimos anos [Sarikaya e Onur 2021, Murti, Ali e Latva-aho 2021, Almeida et al. 2022, Lopes et al. 2022, Awad et al. 2022, Amiri et al. 2023, Murti et al. 2023], porém, os autores não consideraram aspectos importantes que precisam ser abordados em conjunto. Os trabalhos relevantes considerados aqui estão resumidos na Tabela 4.3 e discutidos a seguir.

Alguns desses trabalhos visam resolver o problema de posicionamento de VNF formulando soluções que agregam divisões funcionais com diferentes objetivos [Almeida et al. 2022, Murti, Ali e Latva-aho 2021, Amiri et al. 2023, Murti et al. 2023]. Embora tais propostas visem maximizar a agregação de VNFs em O-CUs, habilitar o CoMP não faz parte dos seus objetivos.

A utilização de métodos baseados em otimização pode impor limites à sua adoção para lidar com soluções de posicionamento dinâmicas e adaptativas. Além disso, eles contam com modelagem ou ajustes que dependem do conhecimento especializado do problema. Nesse sentido, tem sido adotado o uso de métodos baseados em AI/ML/RL [Almeida et al. 2022, Murti, Ali e Latva-aho 2021, Chang et al. 2022,

Amiri et al. 2023, Murti et al. 2023]. Contudo, nenhuma das propostas anteriores considera todos os aspectos aqui considerados, de forma conjunta.

Tabela 4.3: Visão geral do trabalho relacionado.

Aspectos	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
CoMP	x	x					x	x
Clusterização de ERBs	x	x					x	x
Ciente da cobertura	x	x					x	x
O-RAN			x		x	x	x	x
rApp					x	x	x	x
xApp							x	x
SMO								x
DRL/ML	x		x	x	x	x		x
Recursos de rádio		x					x	x

Tabela 4.4: Legenda para índices na Tabela 4.3.

Índice	Referência
[1]	[Chang et al. 2022]
[2]	[Awad et al. 2022]
[3]	[Almeida et al. 2022]
[4]	[Murti, Ali e Latva-aho 2021]
[5]	[Amiri et al. 2023]
[6]	[Murti et al. 2023]
[7]	[Lopes et al. 2022]
[8]	Essa abordagem

4.3.1 O modelo do sistema

Assim como na primeira abordagem, a proposta considera uma vRAN desagregada baseada na arquitetura O-RAN [O-RAN Alliance 2023], conforme apresentado na Figura 4.1. Entretanto, uma versão mais detalhada da arquitetura é considerada em relação à primeira abordagem, apresentada na Figura 4.6.

A topologia é composta do conjunto \mathcal{B} de O-RUs, representando as O-RAN eNBs (O-eNBs) (isto é, as ERBs), munidas de antenas omnidirecionais e hospedando a subcamada Low PHY e todas as funcionalidades de RF necessárias para suportar o LLS. Cada ERB tem um subconjunto de células vizinhas, permitindo criar a matriz binária $\mathcal{K}^{|\mathcal{B}| \times |\mathcal{C}|}$, que mapeia cada b ao CN c definido como âncora do cluster em que b participa. Assim, para $k_{b,c} = 1$ indica que b participa do cluster ancorado em c , e $k_{b,c} = 0$ indica o caso contrário.

Os nós da O-RAN são hospedados nos CNs, representando a O-Cloud. Assim, há um conjunto \mathcal{C} de CNs, em que cada $c \in \mathcal{C}$ pode suportar as VNFs, de forma a não superar a sua capacidade de processamento disponível, definida por c^{Proc} . De maneira

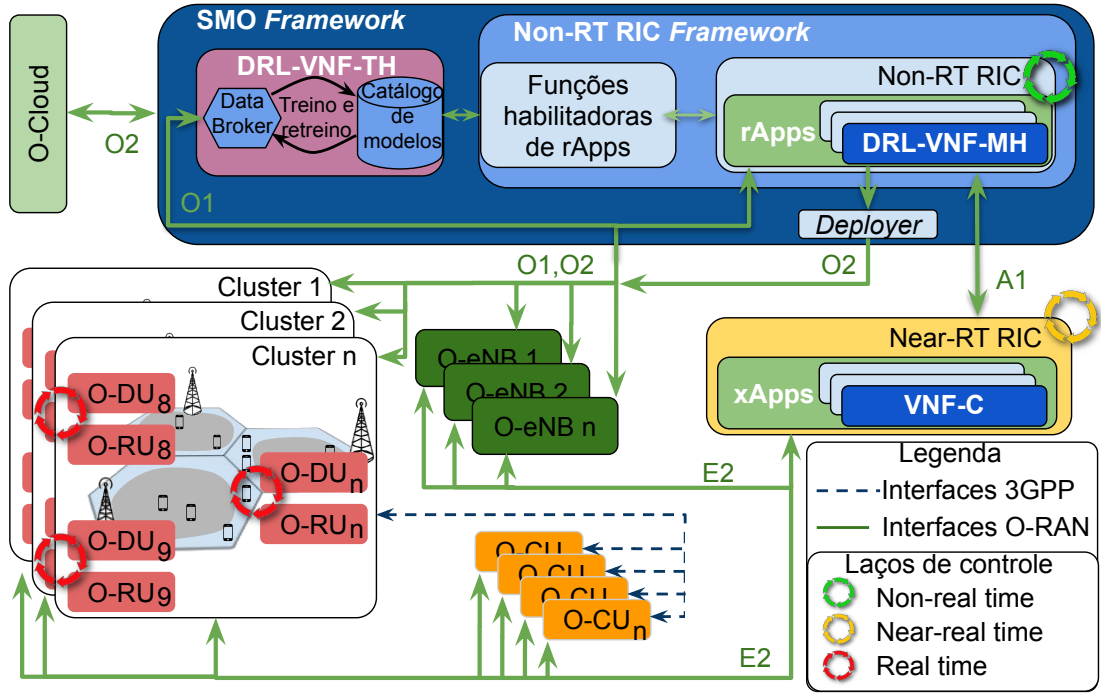


Figura 4.6: A arquitetura O-RAN e os componentes propostos.

mais específica, $\mathcal{C} = \mathcal{C}^{cu} \cup \mathcal{C}^{du} \cup \mathcal{C}^{ru}$, representando os subconjuntos de CNs hospedando as O-CUs, O-DUs e O-RUs, respectivamente. Cada nó está conectado através de um enlace $l \in \mathcal{L}$, com capacidade l^{Cap} e latência estimada l^{Lat} . O fluxo da rede precisa considerar um caminho p , tendo como fonte o núcleo da rede e como destino a ERB $b \in \mathcal{B}$ (isto é, operando em *downlink*). Cada $p \in \mathcal{P}$ é composto do *backhaul* (p_{BH}), *midhaul* (p_{MH}), e do *fronthaul* (p_{FH}).

Assim como considerado na primeira abordagem, a pilha de protocolos da vRAN é um conjunto de VNFs (exceto o protocolo RF), rotulados em ordem crescente, vindo de Low PHY (f_1) até a RRC (f_8), onde \mathcal{F} é o conjunto de VNFs da vRAN desagregada que devem ser alocados adequadamente. \mathcal{D} é o conjunto de DRCs, como em [Almeida et al. 2022]. Assim, uma solução de posicionamento de VNFs estabelece, para cada O-RU $b \in \mathcal{B}$, qual DRC $d_b \in \mathcal{D}$ é selecionado, em quais nós do conjunto \mathcal{C} cada parte da divisão funcional será implantada e quais links em \mathcal{L} serão usados para criar o fluxo necessário entre o núcleo da rede e o O-RU destino. Esta solução só pode ser estabelecida se respeitar as restrições de latência máxima toleradas por cada divisão funcional, e se enquadrar nas capacidades dos enlaces e CNs selecionados.

4.3.2 A solução proposta

Inicialmente, o problema foi formulado como um MDP [Sutton e Barto 2018]. Posteriormente, foi criado um ambiente DRL para treinar um modelo para resolver o

MDP. Para a operacionalização da proposta, quatro componentes O-RAN foram incluídos na arquitetura, conforme apresentado na Figura 4.6, e descrito a seguir.

MDP

O modelo MDP é representado pela tupla $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma\}$, onde \mathcal{S} é o espaço de estados, \mathcal{A} é o espaço de ação composto de ações discretas, \mathcal{P} é a matriz de distribuição de probabilidade de transição, de dimensões $\mathcal{S} \times \mathcal{A} \times \mathcal{S}$, \mathcal{R} é a função de recompensa, de dimensões $\mathcal{S} \times \mathcal{A}$, e $\gamma \in [0, 1]$ é o fator de desconto, que determina a importância das recompensas futuras em relação às recompensas imediatas. No modelo MDP proposto, \mathcal{S} e \mathcal{A} representam todos os estados e ações possíveis, respectivamente, conforme descrito a seguir.

- **Estados:** cada estado $s_t \in \mathcal{S}$ é definido como:

$$s_t = (\mathcal{T}, \mathcal{L}', \mathcal{C}', b^*), \quad (4-11)$$

onde o primeiro termo estabelece detalhes da solução de posicionamento implantada, o segundo e o terceiro termos descrevem detalhes sobre as condições dos conjuntos de enlaces e CNs da topologia, respectivamente, enquanto o último termo funciona como um ponteiro para a O-RU atual, ou seja, as ações são realizadas de forma individual e sequencial, conforme será descrito.

Mais especificamente, \mathcal{T} é definido como uma lista de tuplas da forma $(d_b, c_b, q_b^{\text{CU}}, q_b^{\text{DU}}) \forall b \in \mathcal{B}$, onde $d_b \in \mathcal{D}$ é a divisão funcional atual, $c_b \in \mathcal{C}$ é o O-CU onde o *High Layer Split* (HLS) está implantado, q_b^{CU} e q_b^{DU} são as últimas demandas de processamento conhecidas demandadas no O-CU e O-DU associados à O-RU b , respectivamente, obtidas da Equação 4-1. Em relação às condições dos enlaces, \mathcal{L}' é um vetor que define a capacidade ociosa de $l \in \mathcal{L}$. Dado que as ações são tomadas sequencialmente, para cada O-RU da topologia, \mathcal{C}' demonstra o custo computacional estimado para cada CN levando em consideração as ações já realizadas, e $b^* \in [0, 1]$ estabelece a proporção de O-RUs já alocados.

- **Ações:** Para cada O-RU $b \in \mathcal{B}$, é necessário estabelecer o DRC e em qual O-CU o HLS deverá ser implantado. Uma ação individual é uma tupla $a_t^b = (d_b, c_b^{\text{CU}}) \forall b \in \mathcal{B}$, onde o primeiro termo é o índice do DRC selecionado, sendo o segundo termo o índice do CN que hospeda o nó O-CU de interesse (do subconjunto $\mathcal{C}^{\text{CU}} \subset \mathcal{C}$). O espaço de ação $\mathcal{A} = \mathcal{D} \times \mathcal{C}^{\text{CU}}$, é representado pelo produto cartesiano do conjunto de DRCs e O-CUs. Novamente, como as ações são tomadas de forma sequencial, até que ocorra a alocação para cada uma das O-RUs da vRAN, o conjunto de todas as ações tomadas, isto é, $\Theta_t = \{a_t^b \mid \forall b \in \mathcal{B}\}$, é considerada a solução de posicionamento de VNFs estabelecida.

- **Recompensas:** As recompensas são baseadas nos seguintes sinais:

$$R(s_t, a_t^b) = \begin{cases} 0 & \text{if } |\Theta_t| < |\mathcal{B}| \\ -1 & \text{if } (\Lambda_C^{Proc}(\Theta_t) + \Lambda_C^{Lat}(\Theta_t) + \Lambda_C^{Cap}(\Theta_t)) > 0 \\ R'(\Theta_t) & \text{caso contrário.} \end{cases} \quad (4-12)$$

O primeiro sinal é considerado enquanto as ações individuais ainda não atingiram todas as O-RUs. O segundo sinal considera se alguma restrição na capacidade de processamento do CN, na latência ou na capacidade do link, respectivamente, foi quebrada. Finalmente, a recompensa da solução de posicionamento é calculada da seguinte forma:

$$R'(\Theta_t) = \sum_{b' \in \mathcal{B}'} |b'| \cdot \sum_{i \in b'} \sum_{u \in \mathcal{U}^i} thp(u), \quad (4-13)$$

onde $b' \subset \mathcal{B}'$ representa o subconjunto de ERBs nos clusters formados, e $thp(.)$ é a função para calcular a vazão recebida pelo usuário u associado à ERBs i no cluster b' .

É importante ressaltar que o problema foi relaxado ao não exigir a escolha dos caminhos, como realizado na primeira abordagem, sem perda de generalidade. Foi assumido que, uma vez escolhido o O-CU para um determinado O-RU, o caminho mais curto é utilizado tanto para o *backhaul* quanto para o *midhaul*. Além disso, ao assumir que todas as O-DUs estão co-localizadas a pelo menos um O-RU, e que qualquer O-RU está conectada a apenas uma O-DU, as seleções do caminho para o *fronthaul* e da O-DU não são necessárias.

A solução DRL proposta

Ao utilizar uma solução DRL, um agente pode aprender a lidar com o problema através da experiência obtida ao interagir com o ambiente. O problema pode ser resolvido sem modelá-lo matematicamente e não requer um grande conjunto de dados rotulados, conforme exigido por soluções de aprendizado supervisionada. Nesse sentido, um ambiente DRL foi criado, considerando todos os elementos da vRAN, conforme descritos na Seção 4.3.1, e também considerando MDP definido na Seção 4.3.2. Para treinar o agente, o algoritmo DRL selecionado foi o A2C [Mnih et al. 2016], um algoritmo baseado na arquitetura ator-crítico, implementado como uma variante síncrona e determinística do algoritmo *Asynchronous Advantage Actor Critic* (A3C), dos mesmos autores.

O algoritmo A2C é baseado em duas redes neurais, uma para o ator e outra para o crítico. O primeiro é responsável por calcular a distribuição de probabilidade sobre as possíveis ações relacionadas ao estado atual observado. O segundo deve calcular a recompensa esperada a partir daquele estado. Além disso, o algoritmo emprega a função de vantagem $A(s, a)$ para estimar a qualidade de uma ação realizada a em comparação com a ação média.

O processo de treinamento utiliza os dados de experiência coletados ao interagir com o ambiente. A rede ator é treinada usando uma função de perda de gradiente de política, e a rede crítica é treinada usando uma função de rede de valor. Nesse sentido, o agente é incentivado a tomar ações que levem a maiores recompensas no longo prazo e minimizem as diferenças entre os valores previstos pelo crítico e os retornos reais. Além disso, a vantagem é estimada utilizando a diferença entre as recompensas observadas e os valores previstos pelo crítico.

Os elementos O-RAN propostos

Agora serão descritos os quatro novos componentes propostos, operando de maneira aderente à arquitetura O-RAN, conforme mostrado na Figura 4.6. Três desses componentes estão posicionados no SMO e fazem uso das interfaces e recursos fornecidos pelo SMO *Framework*. Eles são responsáveis tanto pelo ciclo de vida do DRL, quanto pelas tarefas de obtenção e implantação da solução de posicionamento, estando alinhadas às definições estabelecidas em [O-RAN Alliance 2019, O-RAN Alliance 2020, O-RAN Alliance 2023]. O quarto componente reside no near-RT RIC e será descrito a seguir.

O DRL-VNF-TH é o componente operando como um *Training Host* (TH) [O-RAN Alliance 2019], que é responsável pelo treinamento e atualização dos modelos, empregando o DRL proposto. Posicionado no SMO, faz uso da interface *O1* para obter todo e qualquer dado necessário diretamente do *framework* non-RT RIC, das O-DUs e O-RUs, além dos dados das O-CUs. Dessa forma, tais dados estarão disponíveis no *data broker* implementado nesse componente. Assim, o DRL-VNF-TH terá o suporte necessário para realizar treinamentos online ou offline, bem como manter um catálogo de modelos já treinados.

Implantado dentro do non-RT RIC, o DRL-VNF-MH é o componente operando como um *Model Host* (MH), e implementado como uma aplicação do tipo rAPP. Uma primeira responsabilidade desse componente é atuar como um controlador do ciclo de vida dos modelos DRL, inclusive requisitando o treinamento de novos modelos e/ou retreinamento deles. Além disso, é responsável por realizar as inferências (isto é, o host de inferência) e definindo as soluções de posicionamento de VNFs guiado por um modelo DRL treinado (vindo do catálogo de modelos do DRL-VNF-TH). Por estar dentro

do SMO, ele possui integração direta a todos os seus recursos, o que permite a sua comunicação com o DRL-VNF-TH. Além disso, através da interface *A1*, ele pode operar como um serviço, recebendo chamadas e dados de componentes externos.

Ainda dentro do SMO, é proposto o terceiro componente, o *Deployer*, responsável por dar o suporte necessário à O-RAN *cloud computing platform* (O-Cloud) na implantação da solução de posicionamento de VNFs estabelecida, por intermédio da interface *O2*.

Por fim, o uso do componente VNF-C é proposto, sendo um near-RT RIC implementado como um xApp, responsável por monitorar a vRAN, através da interface *E2*. Trata-se de uma versão melhorada daquele proposto na primeira abordagem. Conforme já descrito anteriormente, ele emprega o monitoramento periódico, no qual é observado o estado da rede, gerando relatórios e chamadas ao DRL-VNF-MH através da interface *A1*. Assim, o rApp poderá avaliar se uma nova solução de posicionamento de VNFs precisa ser estabelecida e implantada (através do *Deployer*). Uma importante alteração no componente está no procedimento de contenção, executado assim que a capacidade de processamento de qualquer CN se esgota. Nessas situações, caberá ao VNF-C impor limites às ERBs com relação ao uso de qualquer dos recursos de rádio que gerem impactos na demanda de processamento, de acordo com a Equação 4-1 (custo computacional). Assim, ele poderá limitar a largura de banda máxima permitida, levando à reduções no uso de blocos de recursos de rádio, limitar a máxima ordem de MCS, ou até alterando o modo de transmissão da O-RU. Novamente, o foco é obter uma minimização do custo de processamento, até que uma nova solução de posicionamento de VNFs seja estabelecida.

4.3.3 Resultados e discussões

Para avaliar esta proposta foram realizadas simulações, conforme será descrito a seguir. Para emular o sistema de comunicação, o cenário da vRAN descrito na Seção 4.3.1 foi implementado no Simu5G [Nardini et al. 2020]. Trata-se de um simulador de nível de sistema ponta a ponta para redes 5G baseado nos arcabouços Omnet++ e Inet, bastante robusto e completo, descrito em maiores detalhes na Seção 2.4.2. O Simu5G foi customizado para fornecer acesso às informações necessárias ao ambiente DRL proposto, criando também os recursos para permitir os controles necessários para os componentes propostos e os procedimentos de contenção.

O ambiente DRL foi implementado em Python usando o SB3, diante de seus pontos positivos descritos na Seção 2.3. De forma mais detalhada, como o Simu5G permite a execução de experimentos a partir de outros programas, executando por linha de comando, todo o controle do sistema de comunicação passa a ser realizado pelo ambiente DRL desenvolvido. De fato, o Simu5G não permite alterar as configurações

de um experimento durante a sua execução. Entretanto, a integração construída permite emular tal necessidade. Assim, a alteração dinâmica da solução de posicionamento de VNFs pode ser obtida, sem prejuízos aos objetivos da simulação.

Adicionalmente, embora o Simu5G suporte apenas CoMP CS através da interface X2, foram considerados enlaces para essas interfaces nas quais não ocorram impactos de atrasos e sobrecarga para a propagação dos relatórios de medições exigidos pela coordenador do cluster para a tomada centralizada de decisões de escalonamento e alocação de recursos de rádio. Assim, o emprego de CoMP habilitado pela agregação de VNFs em um mesmo CN é emulado. Por esse motivo, o foco foi nas comparações em cenários em que o CoMP pode ser habilitado ou não. Em outras palavras, o CoMP poderá ser habilitado apenas quando todas as O-RUs de um determinado cluster estiverem com as divisões funcionais que agreguem as VNFs corretas em um mesmo CN. Além disso, comparações entre CoMP localizado ou distribuído estão fora do escopo deste trabalho.

A topologia considerada é a hierárquica, estando de acordo o que foi descrito na Seção 4.3.1, considerando os parâmetros e configurações utilizados nos treinamentos e validações conforme apresentados na Tabela 4.5. Um exemplo da topologia implementada no Simu5G pode ser vista na Figura 4.7. A cada experimento, os usuários são posicionados aleatoriamente dentro de cada célula, seguindo uma distribuição uniforme. Eles se movem, circularmente, ao redor da ERB, a uma velocidade entre 3 a 6 Km/h.

Tabela 4.5: Parâmetros e configurações utilizados nas simulações.

Parâmetros	Valores
$ \mathcal{B} $	9
$ \mathcal{C}^{cu} $, capacidade de processamento	3, 1200 GOPS
\mathcal{D}	{DRC-13, DRC-17, DRC-18}
Distância média entre as células	500 m
Modelo de mobilidade	Circular
Velocidade dos UEs (V - Km/h)	$V \sim \mathcal{N}(\mu, \sigma^2)$, onde $3 \leq V \leq 6$
Posicionamento dos UEs	Distribuição uniforme
Modelo de tráfego	<i>Constant Bitrate</i> (CBR) - 80 Mbps
Frequência	2 GHz
Potência de transmissão	36 dBm
Bandwidth	20 MHz (100 RBs)
Learning rate & gamma	0.0007 - 0.99
A2C Policy	MlpPolicy [64, 64]
Fading + shadowing	Habilitados
Ganho das antenas ERBs / UEs	8 dBi, 0 dBi
Noise figure ERBs / UEs	5 dB, 7 dB
Modelo de <i>path loss</i>	3GPP - TR 36.873

Em relação às demandas das ERBs, é necessário considerar tanto a demanda em termos da taxa de chegada de pacotes para cada usuário, quanto do número de usuários conectados. Para o primeiro caso, é considerada uma demanda uniforme, assumindo um tráfego CBR, com uma taxa de dados de 80 Mbps para cada usuário. Para o número de usuários conectados, foi empregado um modelo sintético derivado do conjunto de

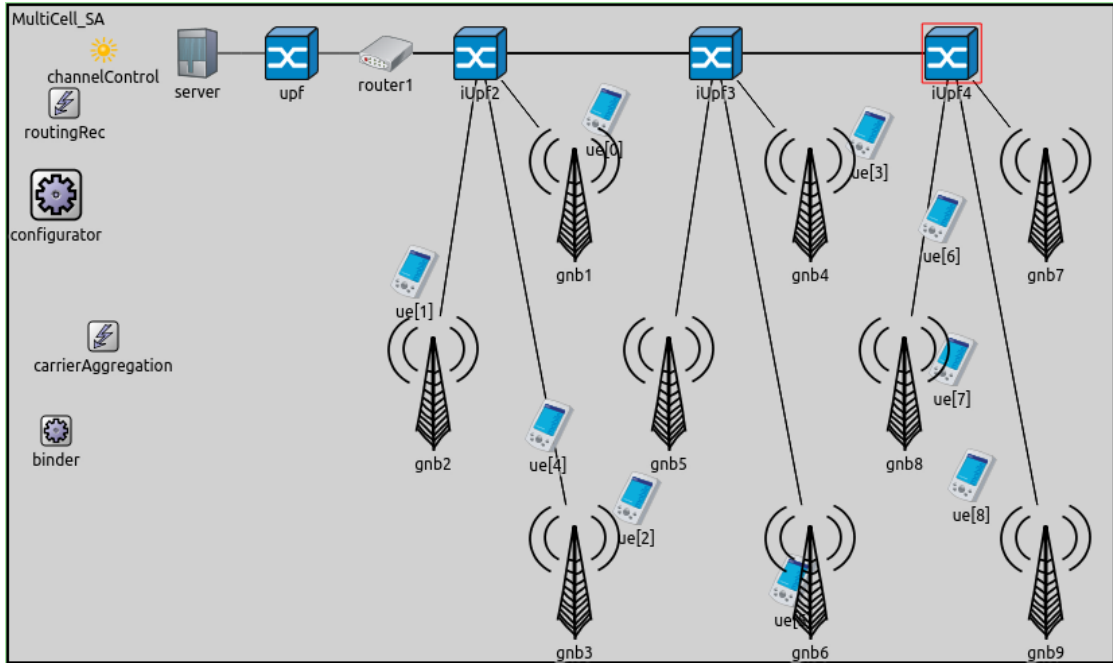


Figura 4.7: Visualização de exemplo da configuração da topologia implementada no Simu5G para os experimentos.

dados de [Chen et al. 2015], que considera duas regiões distintas, urbana e suburbana. Também, esse modelo considera a variação na quantidade de usuários conectados ao longo dos dias da semana, bem como durante as horas do dia. Assim, busca-se ter uma boa representação da dependência espaço-temporal. Uma visão da variação no número de usuários durante as horas de um único dia, conforme considerado nas simulações, é apresentada na Figura 4.8.

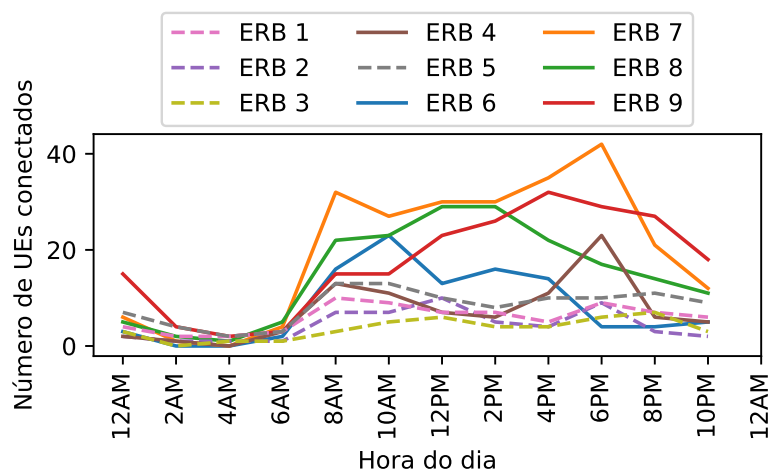


Figura 4.8: A variação de carga das ERBs. As linhas tracejadas representam as localizadas na zona suburbana e as linhas sólidas correspondem à zona urbana.

Desempenho durante o treinamento

Para a fase de treinamento foi utilizada a técnica de multiprocessamento (ambientes vetorizados) disponível no SB3, na qual são criados múltiplos agentes, cada um executando processos paralelos, mas treinando um único modelo global. Tal técnica otimiza o processo de treinamento pois divide a quantidade total de passos de treinamento entre os agentes. A Figura 4.9 mostra a evolução das recompensas médias por episódio de treinamento de um desses agentes. Como pode ser observado, a convergência ocorre em torno de 2 mil episódios de treinamento, sem melhorias nas recompensas a partir de então. Isso representa aproximadamente 200 mil etapas de treinamento (considerando 10 agentes).

Assim que a topologia for estabelecida, o operador pode utilizar um conjunto de dados que reflita as características da vRAN para treinar o modelo de forma *offline*, até atingir este nível de convergência. Com um modelo treinado, o sistema pode começar a operar de forma *online*. Entretanto, tanto o conjunto de dados pode ser aumentado como o modelo pode continuar a ser atualizado, adaptando-se às mudanças nas exigências da rede a curto e longo prazo.

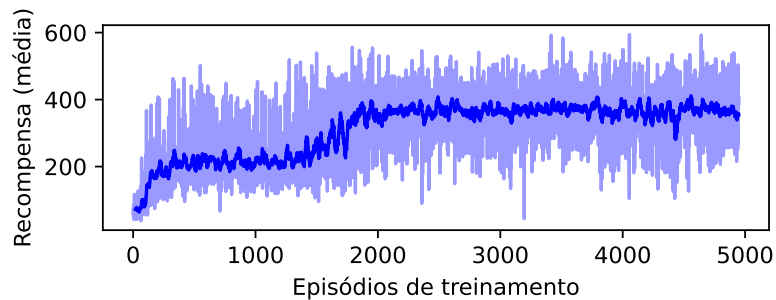


Figura 4.9: Evolução da recompensa durante a fase de treinamento.

4.3.4 Desempenho durante a operação da vRAN

Para a fase em que a vRAN será executada considerando a adaptação contínua e flexível da solução de posicionamento de VNFs proposta, foi empregado um modelo previamente treinado. Assim, ao ser chamado, o agente DRL-VNF-MH (isto é, o rApp proposto) obtém a solução de posicionamento através de previsões feitas a partir dos estados observados, exigindo a seleção de um total de $|\mathcal{B}|$ ações, o que envolve um *forward pass* pela política e as redes de funções de valor por ação necessária. Assim que a solução de posicionamento é obtida, ela é entregue ao componente *Deployer*, responsável pela implantação da solução de acordo com as decisões e os procedimentos de monitoramento da vRAN, conforme descrito na Seção 4.3.2. Assim, o modelo treinado permanece sem ser atualizado durante a execução da simulação.

As Figuras 4.10, 4.11 e 4.12 apresentam diferentes aspectos da performance da abordagem proposta. São avaliadas métricas de eficiência espectral e demanda/custo de processamento computacional, sendo que os dados apresentados são relacionados ao mesmo experimento. Nesse experimento, a operação da vRAN é simulada no decorrer de um dia inteiro, considerando a variação na quantidade de usuários conectados conforme apresentado na Figura 4.8, e considerando os demais parâmetros apresentados na Tabela 4.5. Diversas simulações foram empregadas, de forma que são apresentados os valores médios obtidos em cada simulação. Para simplificar o experimento, as simulações não consideram, de fato, a operação durante 24 horas. Isso foi simplificado considerando apenas o período em que ocorre a mudança entre as horas do dia. Para efeitos de comparação, três soluções estáticas de posicionamento de VNFs foram consideradas. A primeira, denominada *NoCoMP*, considera todas as ERBs operando como D-RANs, em que o CoMP não está habilitado. As demais são soluções de posicionamento que habilitam o CoMP, com diferentes tamanhos de clusters. A solução *Fixo 1* forma um único grande cluster, composto por todos os ERBs (isto é, CoMP global), no qual todos os HLSs são agregados na O-CU 1. A solução *Fixo 2* forma dois clusters, agregados em O-CU e O-CU 2. Finalmente, *VNF-DRL-P* é a solução proposta.

Uma visão geral do comportamento de todas as soluções ao longo do dia é apresentada na Figura 4.10. Pode ser observado que, quando a demanda é baixa (antes das 8h da manhã), a eficiência espectral é maior, o que pode ser explicado pelo fato de que é necessário uma menor quantidade de RBs para atender a demanda de dados. Além disso, há uma menor quantidade de usuários conectados nestes momentos, em que ocorre um menor impacto de interferências. Observa-se que o desempenho sem CoMP fica comprometido, principalmente em situações de maiores demandas, em que qualquer das soluções que empregam CoMP apresentam melhorias nessa métrica.

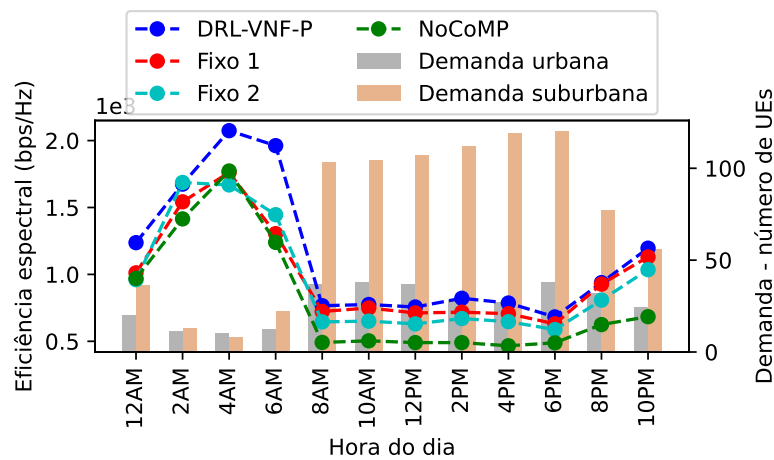


Figura 4.10: Eficiência espectral em razão do número de UEs no decorrer de um dia.

Para permitir uma melhor visualização das melhorias na eficiência espectral,

a Figura 4.11 apresenta maiores detalhes entre as 8h da manhã e as 2 horas da tarde. Novamente, qualquer uma das soluções que habilitam o CoMP apresentam importantes melhorias na eficiência espectral. Dentre as soluções fixas, o cluster global criado pela solução *Fixo 1* se mostra superior. De maneira mais específica, às 2h da tarde, a solução proposta oferece uma eficiência espectral $\approx 68\%$, $\approx 23\%$ e $\approx 15\%$ maior do que *NoCoMP*, *Fixo 2*, e *Fixo 1*, respectivamente. A explicação para esse desempenho está no fato de que pode existir uma combinação de clusters, diante de uma determinada situação de cobertura e demanda, que seja mais adequada do que aquela que cria clusters fixos, mesmo que envolvam todas as ERBs. Assim, o agente DRL proposto pode encontrar tais relações durante o processo de treinamento. Isso evidencia o fato de que podem existir soluções melhores do que as soluções estáticas, mesmo que criadas por especialistas.

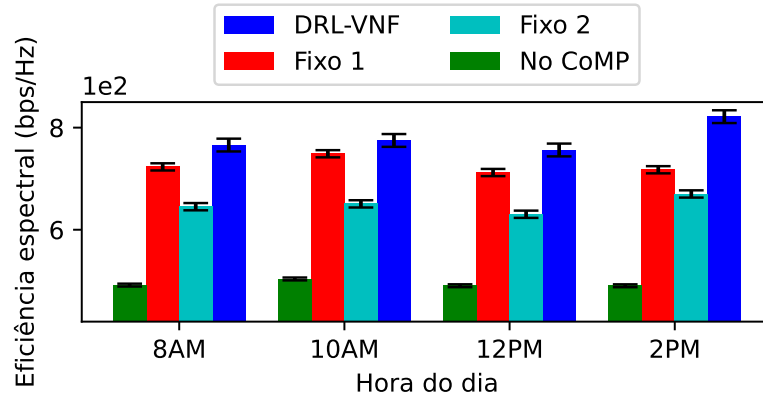


Figura 4.11: Visão mais detalhada da eficiência espectral.

Entretanto, há um ponto fundamental que precisa ser considerado, e envolve o custo computacional dessas soluções, que é apresentado na Figura 4.12. É importante observar que embora as soluções fixas também ofereçam melhorias na eficiência espectral, elas podem representar soluções inviáveis na prática. De fato, durante as simulações, apenas a solução proposta passou pela imposição das restrições envolvendo o esgotamento dos recursos de processamento dos CNs, das capacidades dos enlaces e atraso máximo máximo tolerado pelas divisões funcionais. Ainda, apenas a solução proposta considera o procedimento de contenção descrito na Seção 4.3.2. O foco, inicialmente, era observar o desempenho de diferentes tamanhos de clusters.

A Figura 4.12 apresenta a demanda de processamento médio nos CNs que hospedam os três O-CUs da topologia. A linha tracejada vermelha representa a capacidade de processamento dos CNs (1200 GOPS). As barras possuem três cores, cada uma representando um dos O-CUs, e três padrões, cada um representando, respectivamente, a solução proposta (listras transversais), a solução *Fixo 1* (liso) e o solução *Fixo 2* (bolinhas). A solução *Fixo 1* agrega todas as HLSs do cluster na O-CU 1. A solução *Fixo 2* divide entre o O-CU 1 e O-CU 2.

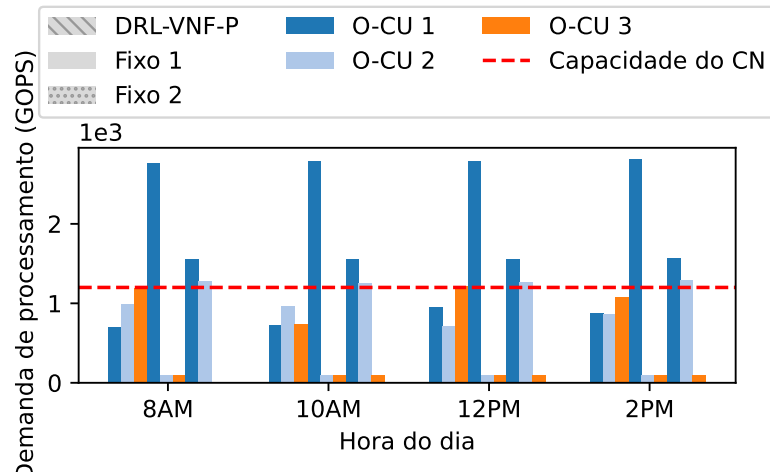


Figura 4.12: Demanda computacional nos CNs.

Assim, é possível observar que, de fato, as soluções fixas não são viáveis. Pode-se observar que apenas a solução dinâmica proposta não quebra esta restrição. Isto é possível devido à capacidade do agente em encontrar as melhores combinações de ERBs para formar os *clusters*, mesmo que resultem em tamanhos menores.

4.4 Conclusão

Nesse capítulo foi apresentado o problema de posicionamento de VNFs em vRANs desagregadas, levando-se em consideração a qualidade do canal dos usuários associados às células. Conforme discutido, há o interesse em métodos dinâmicos, em que a rede de acesso possa se adaptar durante a operação, reagindo às flutuações de demanda e qualidade média dos canais. Nesse sentido, as duas abordagens propostas atendem a esse objetivo, conforme demonstrado através dos experimentos relatados. É importante observar que não se trata de abordagens concorrentes, e sim complementares, obtidas no curso das pesquisas. Assim, a última abordagem (Seção 4.3) pode ser considerada uma versão aprimorada, sendo mais adequada para topologias de maior escala.

Conclusão

O presente documento apresenta os resultados obtidos durante as investigações a respeito dos métodos de alocação e gerenciamento de recursos em redes 5G e B5G com suporte de métodos baseados em aprendizado de máquina. Dentre os recursos de interesse, inicialmente, foram explorados os recursos de rádio. Em uma segunda etapa, a pesquisa evoluiu para a inclusão dos recursos de computação, comunicação e infraestrutura, como aqueles envolvidos em arquiteturas de redes sem fio baseadas em componentes de software.

Os estudos realizados permitiram obter uma clara visão da relação que muitos destes métodos possuem de soluções baseadas em abordagens tradicionais de otimização, incluindo os métodos de otimização exata, ou determinísticos, e aqueles baseados em métodos não determinísticos. Embora as abordagens determinísticas tenham uma longa história nos problemas tradicionais de alocação e gerenciamento de recursos, tais problemas tendem a apresentar um custo computacional proibitivo em instâncias maiores desses problemas. Nesse sentido, observou-se que as soluções apoiadas em métodos não determinísticos têm ganhado a atenção nos anos recentes.

Entretanto, diversos ainda são os problemas e limitações no emprego de métodos não determinísticos e baseados em aprendizado de máquina. Tais limitações se tornam mais complexas ao passo em que novos serviços vão sendo incluídos nas novas gerações das redes sem fio. Adicionalmente, os requisitos projetados para tais redes se mostram cada vez mais desafiadores, exigindo abordagens igualmente avançadas.

Neste sentido, a pesquisa se voltou à busca por abordagens e soluções que consigam lidar com tais problemas e suas limitações. Assim, seguindo a tendência vista nas mais diversas áreas, os métodos baseados em aprendizado de máquina se mostraram com grande potencial na solução dos problemas aqui investigados. Em especial, abordou-se os métodos de aprendizado por reforço (RL) e/ou por reforço profundo (DRL).

No desenvolvimento da pesquisa, um primeiro problema investigado envolveu a alocação de recursos em redes MIMO multibandas. No decorrer do trabalho, um método baseado em DRL foi proposto para realizar o escalonamento de usuários e a sua alocação em diferentes bandas de frequências, nomeado *DRL-based Scheduling and Resource*

Allocation (DRL-SRA). Para avaliar o desempenho e validar a proposta, simulações foram realizadas utilizando canais gerados no QuaDRiGA, considerando tanto a banda sub-6 GHz quanto em ondas milimétricas. Os resultados demonstram a viabilidade da proposta.

Durante a evolução da pesquisa, a alocação de recursos em redes de acesso por rádio desagregadas e virtualizadas se tornou um tópico de interesse. Conforme descrito no decorrer da presente tese, tal problema envolve as decisões necessárias que precisam ser tomadas para a alocação das divisões funcionais das VNFs da vRAN, e o seu posicionamento dentre os nós dispersos pela topologia. O problema também envolve a seleção dos caminhos que precisam ser estabelecidos, criando o fluxo de comunicação na rede de transporte. A pesquisa buscou avançar no estado da arte ao estabelecer como objetivo a criação dinâmica de soluções de posicionamento de VNFs levando em consideração aspectos da cobertura e qualidade dos canais dos usuários conectados. Assim, a vRAN poderá se adaptar às condições atuais da cobertura, no sentido de criar as condições necessárias para habilitar o controle de interferências (ICIC) através do CoMP. Outro objetivo almejado envolveu o estudo da arquitetura O-RAN, de forma a criar soluções aderentes a ela.

Essa etapa da pesquisa foi desenvolvida em duas fases, sendo a primeira voltada aos aspectos iniciais do problema, e a segunda voltada a uma solução mais robusta e escalável. Dessas etapas, duas abordagens foram criadas, descritas no Capítulo 4. Na primeira abordagem, uma solução baseada em otimização exata foi criada. Nela, dois componentes foram propostos a serem incluídos na arquitetura O-RAN, estabelecendo as bases para a introdução dos fluxos e processos para o monitoramento contínuo da vRAN. Assim, uma nova solução de posicionamento de VNFs é produzida sempre que as alterações nas condições da rede são capturadas pelo monitoramento periódico estabelecido, ou quando a capacidade dos recursos computacionais e de infraestrutura se esgotam. Nessa situação, a abordagem proposta também estabelece o processo para reduzir o consumo de processamento, até que uma nova solução seja definida e implantada. Simulações foram realizadas, considerando um importante simulador (em nível de enlace), comprovando tanto a viabilidade quanto a capacidade da proposta no cenário considerado.

Na segunda abordagem, buscou-se estender a proposta estabelecida na primeira abordagem. Parte da motivação se deu pela necessidade de se obter uma solução mais escalável e melhorada para o referido problema. Adicionalmente, estudos mais aprofundados da arquitetura O-RAN foram realizados, ampliando o entendimento dos componentes existentes e suas responsabilidades e interfaces. Dessa fase surge a segunda abordagem, sendo um método baseado em DRL para estabelecer as soluções de posicionamento de VNFs, sempre que for requisitado. O problema foi, inicialmente, formulado como um MDP, seguido da implementação de um ambiente para o treinamento do agente, considerando a sua integração com o sistema de comunicação. Assim, quatro novos componentes

foram propostos, sendo todos aderentes à arquitetura O-RAN, criando uma condição melhorada tanto para o monitoramento da vRAN, quanto para o gerenciamento do ciclo de vida dos modelos treinados e empregados na definição das novas soluções de posicionamento. Nessa abordagem, simulações foram realizadas em um moderno simulador em nível de sistema. Resultados também foram satisfatórios, considerando diferentes situações de carga nas ERBs, e a sua variação no decorrer das horas de um dia.

Dessa forma, a presente tese é fruto de todo o processo científico e tecnológico desempenhado no decorrer da pesquisa aqui relatada. Como resultados obtidos, a pesquisa gerou publicações, que são descritas a seguir.

5.1 Resultados obtidos

As investigações realizadas durante o desenvolvimento da tese geraram, como produtos finais, três publicações científicas diretamente relacionadas ao tema investigado, além dos recursos desenvolvidos através de programas, códigos, conjuntos de dados e alterações nos simuladores. Além disso, outros trabalhos científicos foram publicados tendo o presente estudante como co-autor, sendo tais trabalhos, em certo nível, também dentro do escopo de trabalho comum ao tema da presente tese. Dentre os trabalhos já publicados e/ou apresentados como primeiro autor, têm-se as seguintes publicações:

- LOPES, Victor Hugo L. et al. Deep Reinforcement Learning-Based Scheduling for Multiband Massive MIMO. *IEEE Access*, v. 10, p. 125509-125525, 2022.
- LOPES, Victor Hugo L. et al. A Coverage-Aware VNF Placement and Resource Allocation Approach for Disaggregated vRANs. In: *GLOBECOM 2022-2022 IEEE Global Communications Conference*. IEEE, 2022. p. 185-190.
- LOPES, Victor Hugo L. et al. O-RAN-Oriented Approach for Dynamic VNF Placement Focused on Interference Mitigation. In: *2024 IEEE International Conference on Communications (ICC): Mobile and Wireless Networks Symposium (IEEE ICC'24 - MWN Symposium)*

Além dessas publicações, o presente pesquisador vem contribuindo em outros trabalhos durante o desenvolvimento das pesquisas da presente tese. Tratam-se de colaborações relevantes ao tema da pesquisa, sendo tais trabalhos abaixo relacionados.

- ALMEIDA, G. M., **LOPES, V. H.**, KLAUTAU, A., and CARDOSO, K. V., Deep Reinforcement Learning for Joint Functional Split and Network Function Placement in vRAN. In *GLOBECOM 2022-2022 IEEE Global Communications Conference* (pp. 1229-1234). IEEE.

- NAHUM, C. V., **LOPES, V. H.**, DREIFUERST, R. M., BATISTA, P., CORREA, I., CARDOSO, K. V., KLAUTAU, A., and HEATH, R. W. (2023). Intent-Aware Radio Resource Scheduling in a RAN Slicing Scenario Using Reinforcement Learning. *IEEE Transactions on Wireless Communications*.
- ALMEIDA, G. M., **LOPES, V. H.**, BOTH, C., CORRÊA, S., KLAUTAU, A., AND CARDOSO, K. (2021). MPP-RAN: Posicionamento de funções de Rede Virtualizadas em Redes de Acesso de Nova Geração com Divisão de Fluxos. In: *Anais do I Workshop de Redes 6G*, (pp. 7-12). Porto Alegre: SBC. doi:10.5753/w6g.2021.17228

5.2 Trabalhos futuros

Por fim, ficam sugeridos alguns pontos que o presente pesquisador julga como relevantes para trabalhos futuros no escopo da tese. Primeiramente, podem ser abordados tópicos relativos ao uso de MDP e RL/DRL. Dentre eles, destacam-se os desafios envolvendo a escalabilidade, a generalização e a observabilidade parcial.

Em termos de escalabilidade, um conhecido desafio dos problemas modelados como MDP é a "maldição da dimensionalidade", relatada em diversos trabalhos na literatura [Phyu, Naboulsi e Stanica 2023, Laat, Kusters e Preuss 2023], em que o espaço de observação, bem como o espaço de ações, tendem a assumir dimensões que trazem impactos tanto em termos de tempo de convergência, quanto na dificuldade de representação de problemas de grande escala de parâmetros. Adicionalmente, há um desafio entre balancear a complexidade do ambiente DRL e a quantidade de parâmetros empregados na representação do espaço de observação. Embora os algoritmos DRL empregados nas soluções aqui propostas sejam otimizados para lidar com espaços de grandes dimensões [Elgendy et al. 2021], a aplicação de técnicas para a redução desses espaços se mostra relevante. Tal aspecto impacta tanto o agente DRL proposto para a alocação de recursos de rádio e o escalonamento de usuários para MIMO massivo multibandas, descrito no Capítulo 3 (*problema 1*), quanto o agente DRL proposto para o problema de posicionamento de VNFs descrito na Seção 4.3 (*problema 2*).

Também é importante obter soluções que consigam ser generalizadas para outras instâncias diferentes dos problemas. Tal necessidade também pode impactar na escalabilidade. Por exemplo, os modelos treinados para uma determinada topologia, para o *problema 2*, precisam ser retreinados em caso de alteração na topologia. Considerando que uma determinada topologia pode sofrer alterações durante a sua operação, como no caso da indisponibilidade de CNs e enlaces, por exemplo, o emprego de métodos que consigam generalizar para tais situações deve ser investigado.

Conforme observado durante o desenvolvimento da pesquisa, há um impacto do tipo de observabilidade em que o modelo terá acesso, que precisa ser considerado. Isso ficou evidente durante a construção do agente DRL para o *problema 1*. Da mesma forma, é importante investigar o impacto da observabilidade parcial no *problema 1*, em que as informações requeridas podem ser desatualizadas, ou não estarem completamente disponíveis no momento necessário para a tomada de decisão.

Adicionalmente, outros pontos relevantes podem ser investigados. Pode-se examinar como a predição do tráfego e mobilidade dos usuários podem impactar na tomada de decisões de escalonamento e alocação de recursos em ambos os problemas. O uso de um sistema MIMO multibandas também pode ser explorado no *problema 2*, permitindo obter um melhor entendimento sobre todo o impacto desse sistema na vRAN. Ainda, pode-se tornar viável estudos envolvendo um escalonador otimizado, nos moldes do agente proposto no *problema 1*. Isso se justifica pois, embora as abordagens propostas para o *problema 2* tenham gerado fluxos de controle que impactam a operação do escalonador utilizado, novos escalonadores não foram propostos, de fato. Por fim, é importante avaliar o impacto de topologias considerando recursos heterogêneos, como em casos de CNs com diferentes configurações e capacidades, por exemplo.

Referências Bibliográficas

- [3GPP 2017]3GPP. *Study on 3D channel model for LTE*. [S.l.], 06 2017. Version 12.5.0. Disponível em: <<https://portal.3gpp.org>>.
- [3GPP-TR21.916 2020]3GPP-TR21.916. *System Architecture for the 5G (Release 16)*. [S.l.], 2020.
- [Abbasi et al. 2021]ABBASI, M. et al. Deep Reinforcement Learning for QoS Provisioning at the MAC Layer: A Survey. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 102, p. 104234, 2021.
- [Abdelhakam et al. 2023]ABDELHAKAM, M. M. et al. Collaborative CoMP and Trajectory Optimization for Energy Minimization in Multi-UAV-assisted IoT Networks with QoS Guarantee. *Computer Networks*, Elsevier, v. 237, p. 110074, 2023.
- [Afifi e Karl 2020]AFIFI, H.; KARL, H. Reinforcement Learning for Virtual Network Embedding in Wireless Sensor Networks. In: IEEE. *2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)(50308)*. [S.l.], 2020. p. 123–128.
- [Ajib e Haccoun 2005]AJIB, W.; HACCOUN, D. An Overview of Scheduling Algorithms in MIMO-based Fourth-Generation Wireless Systems. *IEEE Network*, IEEE, v. 19, n. 5, p. 43–48, 2005.
- [Al-Ghamdi e Al-Masalmeh 2020]AL-GHAMDI, J. A.; AL-MASALMEH, E. R. Heuristics and Meta-Heuristics Optimization Methods in Solving Traveling Salesman Problem TSP. 2020.
- [Al-Tam, Correia e Rodriguez 2020]AL-TAM, F.; CORREIA, N.; RODRIGUEZ, J. Learn to Schedule (LEASCH): A Deep Reinforcement Learning Approach for Radio Resource Scheduling in the 5G MAC Layer. *IEEE Access*, IEEE, v. 8, p. 108088–108101, 2020.
- [Alexandropoulos et al. 2019]ALEXANDROPOULOS, S.-A. N. et al. Multi-Objective Evolutionary Optimization Algorithms for Machine Learning: A Recent Survey. In: *Approximation and Optimization*. [S.l.]: Springer, 2019. p. 35–55.

- [Ali e Jammal 2023]ALI, K.; JAMMAL, M. Proactive VNF Scaling and Placement in 5G O-RAN Using ML. *IEEE Transactions on Network and Service Management*, IEEE, 2023.
- [Almeida et al. 2021]ALMEIDA, G. F. de et al. MPP-RAN: Posicionamento de funções de rede virtualizadas em redes de acesso de nova geração com divisão de fluxos. In: *SBRC 2021 - W6G*. [S.l.: s.n.], 2021.
- [Almeida et al. 2023]ALMEIDA, G. M. et al. A Genetic Algorithm for Efficiently Solving the Virtualized Radio Access Network Placement Problem. In: IEEE. *ICC 2023-IEEE International Conference on Communications*. [S.l.], 2023. p. 1874–1879.
- [Almeida et al. 2022]ALMEIDA, G. M. et al. Deep Reinforcement Learning for Joint Functional Split and Network Function Placement in vRAN. In: IEEE. *GLOBECOM 2022-2022 IEEE Global Communications Conference*. [S.l.], 2022. p. 1229–1234.
- [Alnabelsi, Salameh e Albataineh 2020]ALNABELSI, S. H.; SALAMEH, H. A. B.; ALBATAINEH, Z. M. Dynamic Resource Allocation for Opportunistic Software-Defined IoT Networks: Stochastic Optimization Framework. *International Journal of Electrical and Computer Engineering*, IAES Institute of Advanced Engineering and Science, v. 10, n. 4, p. 3854, 2020.
- [Alwarafy et al. 2021]ALWARAFY, A. et al. Deep Reinforcement Learning for Radio Resource Allocation and Management in Next Generation Heterogeneous Wireless Networks: A Survey. *arXiv preprint arXiv:2106.00574*, 2021.
- [Amiri et al. 2023]AMIRI, E. et al. Deep Reinforcement Learning for Robust VNF Recon-figurations in O-RAN. *IEEE Transactions on Network and Service Management*, IEEE, 2023.
- [Asadi e Mancuso 2013]ASADI, A.; MANCUSO, V. A Survey on Opportunistic Scheduling in Wireless Communications. *IEEE Communications Surveys & Tutorials*, IEEE, v. 15, n. 4, p. 1671–1688, 2013.
- [Attiah, Sohrabi e Yu 2020]ATTIAH, K. M.; SOHRABI, F.; YU, W. Deep Learning Approach to Channel Sensing and Hybrid Precoding for TDD Massive MIMO Systems. *arXiv preprint arXiv:2011.10709*, 2020.
- [Ausiello et al. 2001]AUSIELLO, G. et al. Algorithms for the On-Line Traveling Salesman 1. *Algorithmica*, Springer, v. 29, n. 4, p. 560–581, 2001.
- [Avranas, Kountouris e Ciblat 2020]AVRANAS, A.; KOUNTOURIS, M.; CIBLAT, P. Deep Reinforcement Learning for Wireless Scheduling with Multiclass Services. *arXiv preprint arXiv:2011.13634*, 2020.

- [Awad et al. 2022]AWAD, A. M. et al. CoMP-Aware BBU Placements for 5G Radio Access Networks over Optical Aggregation Networks. *Applied Sciences*, MDPI, v. 12, n. 17, p. 8586, 2022.
- [Azariah et al. 2022]AZARIAH, W. et al. A Survey on Open Radio Access Networks: Challenges, Research Directions, and Open Source Approaches. *arXiv preprint arXiv:2208.09125*, 2022.
- [Bakri 2021]BAKRI, S. *Towards Enforcing Network Slicing in 5G Networks*. Tese (Doutorado) — Sorbonne université, 2021.
- [Bartsiokas et al. 2022]BARTSIOKAS, I. A. et al. ML-Based Radio Resource Management in 5G and Beyond Networks: A Survey. *IEEE Access*, IEEE, v. 10, p. 83507–83528, 2022.
- [Bassoy et al. 2017]BASSOY, S. et al. Coordinated Multi-Point Clustering Schemes: A Survey. *IEEE Communications Surveys & Tutorials*, IEEE, v. 19, n. 2, p. 743–764, 2017.
- [Bega et al. 2018]BEGA, D. et al. CARES: Computation-Aware Scheduling in Virtualized Radio Access Networks. *IEEE Transactions on Wireless Communications*, IEEE, v. 17, n. 12, p. 7993–8006, 2018.
- [Bendaoud, Abdennebi e Didi 2014]BENDAOU, F.; ABDENNEBI, M.; DIDI, F. Survey on Scheduling and Radio Resources Allocation in LTE. *arXiv preprint arXiv:1404.2759*, 2014.
- [Bengio, Lodi e Prouvost 2021]BENGIO, Y.; LODI, A.; PROUVOST, A. Machine Learning for Combinatorial Optimization: A Methodological Tour d’Horizon. *European Journal of Operational Research*, Elsevier, v. 290, n. 2, p. 405–421, 2021.
- [Bhardwaj, Ginanjar e Kim 2020]BHARDWAJ, S.; GINANJAR, R. R.; KIM, D.-S. Deep Q-Learning Based Resource Allocation in Industrial Wireless Networks for URLLC. *IET Commun.*, v. 14, n. 6, p. 1022–1027, 2020.
- [Bianchi et al. 2009]BIANCHI, L. et al. A Survey on Metaheuristics for Stochastic Combinatorial Optimization. *Natural Computing*, Springer, v. 8, n. 2, p. 239–287, 2009.
- [Björnson, Hoydis e Sanguinetti 2017]BJÖRNSON, E.; HOYDIS, J.; SANGUINETTI, L. Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency. *Foundations and Trends in Signal Processing*, Now Publishers Inc. Hanover, MA, USA, v. 11, n. 3-4, p. 154–655, 2017.
- [Björnson e Jorswieck 2013]BJÖRNSON, E.; JORSWIECK, E. *Optimal Resource Allocation in Coordinated Multi-Cell Systems*. [S.I.]: Now Publishers Inc, 2013.

- [Björnson, Hoydis e Sanguinetti 2017]BJÖRNSON, E.; HOYDIS, J.; SANGUINETTI, L. Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency. *Foundations and Trends® in Signal Processing*, v. 11, n. 3-4, p. 154–655, 2017. ISSN 1932-8346. Disponível em: <<http://dx.doi.org/10.1561/20000000093>>.
- [Blum et al. 2011]BLUM, C. et al. Hybrid Metaheuristics in Combinatorial Optimization: A Survey. *Applied Soft Computing*, Elsevier, v. 11, n. 6, p. 4135–4151, 2011.
- [Brik, Boutiba e Ksentini 2022]BRIK, B.; BOUTIBA, K.; KSENTINI, A. Deep Learning for B5G Open Radio Access Network: Evolution, Survey, Case Studies, and Challenges. *IEEE Open Journal of the Communications Society*, IEEE, v. 3, p. 228–250, 2022.
- [Brockman et al. 2016]BROCKMAN, G. et al. *OpenAI Gym*. 2016.
- [Brockman et al. 2016]BROCKMAN, G. et al. OpenAI Gym (2016). *arXiv preprint arXiv:1606.01540*, v. 476, 2016.
- [Budalal e Islam 2023]BUDALAL, A. A.; ISLAM, M. R. Path Loss Models for Outdoor Environment—With a Focus on Rain Attenuation Impact on Short-Range Millimeter-Wave Links. *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, Elsevier, v. 3, p. 100106, 2023.
- [Burkhardt et al. 2014]BURKHARDT, F. et al. QuaDRiGa: A MIMO Channel Model for Land Mobile Satellite. In: IEEE. *The 8th European Conference on Antennas and Propagation (EuCAP 2014)*. [S.l.], 2014. p. 1274–1278.
- [Cai et al. 2017]CAI, X. et al. An Adaptive Memetic Framework for Multi-objective Combinatorial Optimization Problems: Studies on Software Next Release and Travelling Salesman Problems. *Soft Computing*, Springer, v. 21, n. 9, p. 2215–2236, 2017.
- [Caponi, Chiti e Fantacci 2004]CAPONI, L.; CHITI, F.; FANTACCI, R. A Dynamic Rate Allocation Technique for Wireless Communication Systems. In: IEEE. *2004 IEEE International Conference on Communications (IEEE Cat. No. 04CH37577)*. [S.l.], 2004. v. 7, p. 4263–4267.
- [Cardoso et al. 2020]CARDOSO, K. V. et al. A Softwarized Perspective of the 5G Networks. *arXiv preprint arXiv:2006.10409*, 2020.
- [Cassagne et al. 2019]CASSAGNE, A. et al. Aff3ct: A Fast Forward Error Correction Toolbox! *SoftwareX*, Elsevier, v. 10, p. 100345, 2019.
- [Castaneda et al. 2016]CASTANEDA, E. et al. An Overview on Resource Allocation Techniques for Multi-user MIMO Systems. *IEEE Communications Surveys & Tutorials*, IEEE, v. 19, n. 1, p. 239–284, 2016.

- [Chaieb, Abdelkefi e Ajib 2022]CHAIEB, C.; ABDELKEFI, F.; AJIB, W. Deep Reinforcement Learning for Resource Allocation in Multi-band and Hybrid OMA-NOMA Wireless Networks. *IEEE Transactions on Communications*, IEEE, v. 71, n. 1, p. 187–198, 2022.
- [Chang et al. 2022]CHANG, C.-Y. et al. FlexDRAN: Flexible Centralization in Disaggregated Radio Access Networks. *IEEE Access*, IEEE, v. 10, p. 11789–11808, 2022.
- [Chen et al. 2020]CHEN, P.-C. et al. DDPG-Based Radio Resource Management for User Interactive Mobile Edge Networks. In: IEEE. *2020 2nd 6G Wireless Summit (6G SUMMIT)*. [S.l.], 2020. p. 1–5.
- [Chen et al. 2021]CHEN, T. et al. Learning to Optimize: A Primer and a Benchmark. *arXiv preprint arXiv:2103.12828*, 2021.
- [Chen et al. 2015]CHEN, X. et al. Analyzing and Modeling Spatio-temporal Dependence of Cellular Traffic at City Scale. In: IEEE. *2015 IEEE International Conference on Communications (ICC)*. [S.l.], 2015. p. 3585–3591.
- [Choi et al. 2018]CHOI, J. et al. Joint User Scheduling, Power Allocation, and Precoding Design for Massive MIMO Systems: A Principal Component Analysis Approach. In: IEEE. *2018 IEEE International Symposium on Information Theory (ISIT)*. [S.l.], 2018. p. 396–400.
- [Choromanski et al. 2019]CHOROMANSKI, K. et al. From Complexity to Simplicity: Adaptive ES-Active Subspaces for Blackbox Optimization. *arXiv preprint arXiv:1903.04268*, 2019.
- [Coello, Dhaenens e Jourdan 2010]COELLO, C. A. C.; DHAENENS, C.; JOURDAN, L. Multi-objective Combinatorial Optimization: Problematic and Context. In: *Advances in Multi-objective Nature Inspired Computing*. [S.l.]: Springer, 2010. p. 1–21.
- [Comsa, De-Domenico e Ktenas 2017]COMSA, I. S.; De-Domenico, A.; KTENAS, D. QoS-Driven Scheduling in 5G Radio Access Networks - A Reinforcement Learning Approach. In: . [S.l.: s.n.], 2017. p. 1–7.
- [Comşa et al. 2014]COMŞA, I. S. et al. Adaptive Proportional Fair Parameterization Based LTE Scheduling Using Continuous Actor-Critic Reinforcement Learning. In: . [S.l.: s.n.], 2014. p. 4387–4393.
- [Costa-Pérez et al. 2013]COSTA-PÉREZ, X. et al. Radio Access Network Virtualization for Future Mobile Carrier Networks. *IEEE Communications Magazine*, IEEE, v. 51, n. 7, p. 27–35, 2013.

- [Demirci e Sagiroglu 2019]DEMIRCI, S.; SAGIROGLU, S. Optimal Placement of Virtual Network Functions in Software Defined Networks: A Survey. *Journal of Network and Computer Applications*, v. 147, p. 102424, 2019.
- [Deng et al. 2020]DENG, X. et al. Wireless Powered Mobile Edge Computing: Dynamic Resource Allocation and Throughput Maximization. *IEEE Transactions on Mobile Computing*, IEEE, 2020.
- [Devanur et al. 2019]DEVANUR, N. R. et al. Near Optimal Online Algorithms and Fast Approximation Algorithms for Resource Allocation Problems. *Journal of the ACM (JACM)*, ACM New York, NY, USA, v. 66, n. 1, p. 1–41, 2019.
- [Dhariwal et al. 2017]DHARIWAL, P. et al. *OpenAI Baselines*. 2017. Disponível em: <<https://github.com/openai/baselines>>.
- [Divya e Jayanthi 2020]DIVYA, R.; JAYANTHI, V. Efficient Optimal Resource Allocation for Profit Maximization in Software Defined Network Approach to Improve Quality of Service in Cloud Environments. *Journal of Ambient Intelligence and Humanized Computing*, Springer, p. 1–10, 2020.
- [Djonin e Krishnamurthy 2007]DJONIN, D. V.; KRISHNAMURTHY, V. MIMO Transmission Control in Fading Channels—A Constrained Markov Decision Process Formulation with Monotone Randomized Policies. *IEEE Transactions on Signal Processing*, IEEE, v. 55, n. 10, p. 5069–5083, 2007.
- [Doering et al. 2019]DOERING, J. et al. Metaheuristics for Rich Portfolio Optimisation and Risk Management: Current State and Future Trends. *Operations Research Perspectives*, Elsevier, v. 6, p. 100121, 2019.
- [Dong et al. 2020]DONG, R. et al. Deep Learning for Radio Resource Allocation with Diverse Quality-of-Service Requirements in 5G. *IEEE Transactions on Wireless Communications*, IEEE, 2020.
- [Dreifuerst et al. 2020]DREIFUERST, R. M. et al. Optimizing Coverage and Capacity in Cellular Networks using Machine Learning. *arXiv preprint arXiv:2010.13710*, 2020.
- [Duchi 2018]DUCHI, J. C. Introductory Lectures on Stochastic Optimization. *The Mathematics of Data*, v. 25, p. 99–185, 2018.
- [Dzogovic et al. 2018]DZOGOVIĆ, B. et al. Building Virtualized 5G Networks using Open Source Software. In: IEEE. *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*. [S.l.], 2018. p. 360–366.

- [Eckhardt et al. 2022]ECKHARDT, J. M. et al. Modular Link Level Simulator for the Physical Layer of Beyond 5G Wireless Communication Systems. *Radio Science*, AGU, v. 57, n. 2, p. 1–15, 2022.
- [Ejaz et al. 2020]EJAZ, W. et al. A Comprehensive Survey on Resource Allocation for CRAN in 5G and Beyond Networks. *Journal of Network and Computer Applications*, Elsevier, v. 160, p. 102638, 2020.
- [Eldeeb et al. 2020]ELDEEB, H. B. et al. Optimal Resource Allocation and Interference Management for Multi-User Uplink Light Communication Systems with Angular Diversity Technology. *IEEE Access*, IEEE, v. 8, p. 203224–203236, 2020.
- [Elgendy et al. 2021]ELGENDY, I. A. et al. Advanced Deep Learning for Resource Allocation and Security Aware Data Offloading in Industrial Mobile Edge Computing. *Big Data*, Mary Ann Liebert, Inc., publishers 140 Huguenot Street, 3rd Floor New . . . , v. 9, n. 4, p. 265–278, 2021.
- [Enayaty-Ahangar, Albert e DuBois 2020]ENAYATY-AHANGAR, F.; ALBERT, L. A.; DU-BOIS, E. A Survey of Optimization Models and Methods for Cyberinfrastructure Security. *IJSE Transactions*, Taylor & Francis, v. 53, n. 2, p. 182–198, 2020.
- [Esmaeily e Kralevska 2021]ESMAEILY, A.; KRALEVSKA, K. Small-Scale 5G Testbeds for Network Slicing Deployment: A Systematic Review. *Wireless Communications and Mobile Computing*, Hindawi Limited, v. 2021, p. 1–26, 2021.
- [Femenias et al. 2017]FEMENIAS, G. et al. Downlink Scheduling and Resource Allocation for 5G MIMO-Multicarrier: OFDM vs FBMC/OQAM. *IEEE Access*, v. 5, p. 13770–13786, 2017.
- [Fontaine, Mannor e Perchet 2020]FONTAINE, X.; MANNOR, S.; PERCHET, V. An Adaptive Stochastic Optimization Algorithm for Resource Allocation. In: PMLR. *Algorithmic Learning Theory*. [S.l.], 2020. p. 319–363.
- [Foukas, Marina e Kontovasilis 2017]FOUKAS, X.; MARINA, M. K.; KONTOVASILIS, K. Orion: RAN Slicing for a Flexible and Cost-Effective Multi-Service Mobile Network Architecture. In: *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. [S.l.: s.n.], 2017. p. 127–140.
- [Frangioni 2005]FRANGIONI, A. About Lagrangian Methods in Integer Optimization. *Annals of Operations Research*, Springer, v. 139, n. 1, p. 163–193, 2005.
- [Gambella, Ghaddar e Naoum-Sawaya 2020]GAMBELLA, C.; GHADDAR, B.; NAOUM-SAWAYA, J. Optimization Problems for Machine Learning: A Survey. *European Journal of Operational Research*, Elsevier, 2020.

- [Garcia-Aviles et al. 2018]GARCIA-AVILES, G. et al. POSENS: A Practical Open Source Solution for End-to-End Network Slicing. *IEEE Wireless Communications*, IEEE, v. 25, n. 5, p. 30–37, 2018.
- [Ghadimi et al. 2017]GHADIMI, E. et al. A Reinforcement Learning Approach to Power Control and Rate Adaptation in Cellular Networks. In: *2017 IEEE International Conference on Communications (ICC)*. [S.l.: s.n.], 2017. p. 1–7.
- [Gigerenzer e Gaissmaier 2011]GIGERENZER, G.; GAISSMAIER, W. Heuristic Decision Making. *Annual Review of Psychology*, Annual Reviews, v. 62, p. 451–482, 2011.
- [Gkonis, Trakadas e Kaklamani 2020]GKONIS, P. K.; TRAKADAS, P. T.; KAKLAMANI, D. I. A Comprehensive Study on Simulation Techniques for 5G Networks: State of the Art Results, Analysis, and Future Challenges. *Electronics*, MDPI, v. 9, n. 3, p. 468, 2020.
- [Guadarrama et al. 2018]GUADARRAMA, S. et al. *TF-Agents: A library for Reinforcement Learning in TensorFlow*. 2018. <https://github.com/tensorflow/agents>. [Online; accessed 10-January-2024]. Disponível em: <<https://github.com/tensorflow/agents>>.
- [Gutjahr e Pichler 2016]GUTJAHR, W. J.; PICHLER, A. Stochastic Multi-Objective Optimization: A Survey on Non-Scalarizing Methods. *Annals of Operations Research*, Springer, v. 236, n. 2, p. 475–499, 2016.
- [Habibi et al. 2019]HABIBI, M. A. et al. A Comprehensive Survey of RAN Architectures Toward 5G Mobile Communication System. *IEEE Access*, IEEE, v. 7, p. 70371–70421, 2019.
- [Han e Schotten 2020]HAN, B.; SCHOTTEN, H. D. Machine Learning for Network Slicing Resource Management: A Comprehensive Survey. *arXiv preprint arXiv:2001.07974*, 2020.
- [Han et al. 2020]HAN, S. et al. Artificial-Intelligence-Enabled Air Interface for 6G: Solutions, Challenges, and Standardization Impacts. *IEEE Communications Magazine*, IEEE, v. 58, n. 10, p. 73–79, 2020.
- [Hanafi e Todosijević 2017]HANAFI, S.; TODOSIJEVIĆ, R. Mathematical Programming Based Heuristics for the 0–1 MIP: A Survey. *Journal of Heuristics*, Springer, v. 23, n. 4, p. 165–206, 2017.
- [Hanagandi e Nikolaou 1998]HANAGANDI, V.; NIKOLAOU, M. A Hybrid Approach to Global Optimization Using a Clustering Algorithm in a Genetic Search Framework. *Computers & Chemical Engineering*, Elsevier, v. 22, n. 12, p. 1913–1925, 1998.

- [Hao et al. 2020]HAO, X. et al. Joint Resource Allocation Algorithm Based on Multi-Objective Optimization for Wireless Sensor Networks. *Applied Soft Computing*, Elsevier, v. 94, p. 106470, 2020.
- [Harutyunyan e Riggio 2018]HARUTYUNYAN, D.; RIGGIO, R. Flex5G: Flexible Functional Split in 5G Networks. *IEEE Transactions on Network and Service Management*, IEEE, v. 15, n. 3, p. 961–975, 2018.
- [Hashimoto 2004]HASHIMOTO, K. *Técnicas de Otimização Combinatória Multiobjetivo Aplicadas na Estimação do Desempenho Elétrico de Redes de Distribuição*. Tese (Doutorado) — Universidade de São Paulo, 2004.
- [Hassan et al. 2020]HASSAN, M. Z. et al. Statistical-QoS Guarantee for IoT Network Driven by Laser-Powered UAV Relay and RF Backscatter Communications. *IEEE Transactions on Green Communications and Networking*, IEEE, v. 5, n. 1, p. 406–425, 2020.
- [Hendricks, Jannerup e Sørensen 2008]HENDRICKS, E.; JANNERUP, O.; SØRENSEN, P. H. *Linear Systems Control: Deterministic and Stochastic Methods*. [S.I.]: Springer Science & Business Media, 2008.
- [Hill et al. 2018]HILL, A. et al. *Stable Baselines*. [S.I.]: GitHub, 2018. <https://github.com/hill-a/stable-baselines>.
- [Hill et al. 2018]HILL, A. et al. *Stable Baselines*. aug 2018. Disponível em: <<https://stable-baselines.readthedocs.io/en/master/index.html>>.
- [Hossain et al. 2019]HOSSAIN, M. F. et al. Recent Research in Cloud Radio Access Network (C-RAN) for 5G Cellular Systems-A Survey. *Journal of Network and Computer Applications*, Elsevier, v. 139, p. 31–48, 2019.
- [Hussain et al. 2020]HUSSAIN, F. et al. Machine Learning for Resource Management in Cellular and IoT Networks: Potentials, Current Solutions, and Open Challenges. *IEEE Communications Surveys & Tutorials*, IEEE, v. 22, n. 2, p. 1251–1275, 2020.
- [Hwang e Masud 2012]HWANG, C.-L.; MASUD, A. S. M. *Multiple Objective Decision Making—Methods and Applications: A State-of-the-Art Survey*. [S.I.]: Springer Science & Business Media, 2012.
- [Jaeckel et al. 2014]JAECKEL, S. et al. QuaDRiGa: A 3-D Multi-Cell Channel Model with Time Evolution for Enabling Virtual Field Trials. *IEEE Transactions on Antennas and Propagation*, IEEE, v. 62, n. 6, p. 3242–3256, 2014.

- [Jaillet e Wagner 2008]JAILLET, P.; WAGNER, M. R. Online Vehicle Routing Problems: A Survey. In: *The Vehicle Routing Problem: Latest Advances and New Challenges*. [S.l.]: Springer, 2008. p. 221–237.
- [Jiang et al. 2021]JIANG, X. et al. Resource Allocation of Video Streaming Over Vehicular Networks: A Survey, Some Research Issues and Challenges. *IEEE Transactions on Intelligent Transportation Systems*, IEEE, 2021.
- [Johnson 2011]JOHNSON, C. *Radio Access Networks for UMTS: Principles and Practice*. [S.l.]: John Wiley & Sons, 2011.
- [Jr e Lozano 2018]JR, R. W. H.; LOZANO, A. *Foundations of MIMO Communication*. [S.l.]: Cambridge University Press, 2018.
- [Kak et al. 2022]KAK, A. et al. A Disaggregated O-RAN Platform for Network Slice Deployment and Assurance. In: IEEE. *IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. [S.l.], 2022. p. 1–2.
- [Karaboga, Akay e Karaboga 2020]KARABOGA, D.; AKAY, B.; KARABOGA, N. A Survey on the Studies Employing Machine Learning (ML) for Enhancing Artificial Bee Colony (ABC) Optimization Algorithm. *Cogent Engineering*, Taylor & Francis, v. 7, n. 1, p. 1855741, 2020.
- [Karimi-Mamaghan et al. 2022]KARIMI-MAMAGHAN, M. et al. Machine Learning at the Service of Meta-Heuristics for Solving Combinatorial Optimization Problems: A State-of-the-Art. *European Journal of Operational Research*, Elsevier, v. 296, n. 2, p. 393–422, 2022.
- [Kasgari e Saad 2018]KASGARI, A. T. Z.; SAAD, W. Stochastic Optimization and Control Framework for 5G Network Slicing with Effective Isolation. In: IEEE. *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*. [S.l.], 2018. p. 1–6.
- [Kastratia e Bibaa 2021]KASTRATIA, M.; BIBAA, M. A State-of-the-Art Survey of Advanced Optimization Methods in Machine Learning. In: *Proccedings of RTA-CSIT 2021*. [S.l.: s.n.], 2021.
- [Kawser et al. 2012]KAWSER, M. T. et al. Performance Comparison Between Round Robin and Proportional Fair Scheduling Methods for LTE. *International Journal of Information and Electronics Engineering*, v. 2, n. 5, p. 678–681, 2012.
- [Khalek, Caramanis e Jr 2015]KHALEK, A. A.; CARAMANIS, C.; JR, R. W. H. Delay-Constrained Video Transmission: Quality-Driven Resource Allocation and Scheduling.

- IEEE Journal of Selected Topics in Signal Processing*, v. 9, n. 1, p. 60–75, Feb 2015. ISSN 1932-4553, 1941-0484.
- [Khan et al. 2020]KHAN, H. et al. Deep Learning Assisted CSI Estimation for Joint URLLC and eMBB Resource Allocation. In: IEEE. *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. [S.I.], 2020. p. 1–6.
- [Khan et al. 2020]KHAN, H. Z. et al. Resource Allocation and Throughput Maximization in Decoupled 5G. In: IEEE. *2020 IEEE Wireless Communications and Networking Conference (WCNC)*. [S.I.], 2020. p. 1–6.
- [Kimura e Seki 2012]KIMURA, D.; SEKI, H. Inter-Cell Interference Coordination (ICIC) Technology. *Fujitsu Scientific & Technical Journal*, v. 48, n. 1, p. 89–94, 2012.
- [Kotary et al. 2021]KOTARY, J. et al. End-to-End Constrained Optimization Learning: A Survey. *arXiv preprint arXiv:2103.16378*, 2021.
- [Koutsopoulos e Tassiulas 2008]KOUTSOPOULOS, I.; TASSIULAS, L. The Impact of Space Division Multiplexing on Resource Allocation: A Unified Treatment of TDMA, OFDMA and CDMA. *IEEE Transactions on Communications*, IEEE, v. 56, n. 2, p. 260–269, 2008.
- [Krumke 2002]KRUMKE, S. O. Online Optimization: Competitive Analysis and Beyond. 2002.
- [Larsen et al. 2023]LARSEN, L. M. et al. Toward Greener 5G and Beyond Radio Access Networks—A Survey. *IEEE Open Journal of the Communications Society*, IEEE, v. 4, p. 768–797, 2023.
- [Laszczyk e Myszkowski 2019]LASZCZYK, M.; MYSZKOWSKI, P. B. Survey of quality measures for multi-objective optimization: Construction of complementary set of multi-objective quality measures. *Swarm and Evolutionary Computation*, Elsevier, v. 48, p. 109–133, 2019.
- [Le et al. 2021]LE, T. V. et al. Edge computing simulation platforms: A technology survey. In: SPRINGER. *Advances in Service-Oriented and Cloud Computing: International Workshops of ESOC 2020, Heraklion, Crete, Greece, September 28–30, 2020, Revised Selected Papers 8*. [S.I.], 2021. p. 18–28.
- [Lee e Lee 2017]LEE, J.-H.; LEE, J.-Y. Optimal beamforming-selection spatial precoding using population-based stochastic optimization for massive wireless MIMO communication systems. *Journal of the Franklin Institute*, Elsevier, v. 354, n. 10, p. 4247–4272, 2017.

- [Lee 2018]LEE, W. Resource allocation for multi-channel underlay cognitive radio network based on deep neural network. *IEEE Communications Letters*, IEEE, v. 22, n. 9, p. 1942–1945, 2018.
- [Lee e Qin 2019]LEE, Y. L.; QIN, D. A survey on applications of deep reinforcement learning in resource management for 5G heterogeneous networks. In: IEEE. *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. [S.l.], 2019. p. 1856–1862.
- [Lei 2021]LEI, C. Deep reinforcement learning. In: *Deep Learning and Practice with MindSpore*. [S.l.]: Springer, 2021. p. 217–243.
- [Letchford, Ni e Zhong 2020]LEITCHFORD, A. N.; NI, Q.; ZHONG, Z. A heuristic for fair dynamic resource allocation in overloaded OFDMA systems. *Journal of Heuristics*, Springer, v. 26, n. 1, p. 21–32, 2020.
- [Li e Pan 2015]LI, J.-q.; PAN, Q.-k. Solving the large-scale hybrid flow shop scheduling problem with limited buffers by a hybrid artificial bee colony algorithm. *Information Sciences*, Elsevier, v. 316, p. 487–502, 2015.
- [Li e Malik 2016]LI, K.; MALIK, J. Learning to optimize. *arXiv preprint arXiv:1606.01885*, 2016.
- [Li, Zhang e Wang 2020]LI, K.; ZHANG, T.; WANG, R. Deep reinforcement learning for multiobjective optimization. *IEEE transactions on cybernetics*, IEEE, v. 51, n. 6, p. 3103–3114, 2020.
- [Li e Qian 2015]LI, X.; QIAN, C. The virtual network function placement problem. In: IEEE. *2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. [S.l.], 2015. p. 69–70.
- [Liang, Ye e Li 2019]LIANG, L.; YE, H.; LI, G. Y. Spectrum sharing in vehicular networks based on multi-agent reinforcement learning. *IEEE Journal on Selected Areas in Communications*, IEEE, v. 37, n. 10, p. 2282–2292, 2019.
- [Liang et al. 2019]LIANG, L. et al. Deep-learning-based wireless resource allocation with application to vehicular networks. *Proceedings of the IEEE*, IEEE, v. 108, n. 2, p. 341–356, 2019.
- [Liang et al. 2012]LIANG, Y.-S. et al. Resource allocation with interference avoidance in OFDMA femtocell networks. *IEEE Transactions on Vehicular Technology*, IEEE, v. 61, n. 5, p. 2243–2255, 2012.

- [Lin e Zhao 2020]LIN, M.; ZHAO, Y. Artificial intelligence-empowered resource management for future wireless communications: A survey. *China Communications*, IEEE, v. 17, n. 3, p. 58–77, 2020.
- [Lin, Yu e Chen 2021]LIN, S.; YU, H.; CHEN, H. On-Line Optimization of Microgrid Operating Cost Based on Deep Reinforcement Learning. In: IOP PUBLISHING. *IOP Conference Series: Earth and Environmental Science*. [S.l.], 2021. v. 701, n. 1, p. 012084.
- [Lombardi e Milano 2018]LOMBARDI, M.; MILANO, M. Boosting combinatorial problem modeling with machine learning. *arXiv preprint arXiv:1807.05517*, 2018.
- [Lopes et al. 2022]LOPES, V. H. L. et al. A Coverage-Aware VNF Placement and Resource Allocation Approach for Disaggregated vRANs. In: IEEE. *GLOBECOM 2022-2022 IEEE Global Communications Conference*. [S.l.], 2022. p. 185–190.
- [Loureço, Martin e Stützle 2019]LOURENÇO, H. R.; MARTIN, O. C.; STÜTZLE, T. Iterated local search: Framework and applications. In: *Handbook of metaheuristics*. [S.l.]: Springer, 2019. p. 129–168.
- [Lu, Huang e Hua 2020]LU, L.; HUANG, L.; HUA, W. Energy efficient resource allocation and scheduling for delay-constrained multi-cell broadcast networks. *IEEE Access*, IEEE, v. 8, p. 164844–164857, 2020.
- [Lu et al. 2013]LU, X. et al. *Online optimization problems*. Tese (Doutorado) — Massachusetts Institute of Technology, 2013.
- [Lu e Gursay 2019]LU, Z.; GURSOY, M. C. Dynamic channel access and power control via deep reinforcement learning. In: IEEE. *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*. [S.l.], 2019. p. 1–5.
- [Lust e Teghem 2010]LUST, T.; TEGHEM, J. The multiobjective traveling salesman problem: a survey and a new approach. In: *Advances in Multi-Objective Nature Inspired Computing*. [S.l.]: Springer, 2010. p. 119–141.
- [Maaz et al. 2017]MAAZ, B. et al. Joint User Association, Power Control and Scheduling in Multi-Cell 5G Networks. In: *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. [S.l.: s.n.], 2017. p. 1–6.
- [Mademann 2018]MADEMANN, F. The 5G System Architecture. *Journal of ICT Standardization*, p. 77–86, 2018.
- [Mahajan e Teneketzis 2008]MAHAJAN, A.; TENEKETZIS, D. Multi-armed Bandit Problems. In: *Foundations and Applications of Sensor Management*. [S.l.]: Springer, 2008. p. 121–151.

- [Mamode e Fowdur 2020]MAMODE, M. I. S.; FOWDUR, T. P. Survey of Scheduling Schemes in 5G Mobile Communication Systems. *Journal of Electrical Engineering, Electronics, Control and Computer Science*, v. 6, n. 2, p. 21–30, 2020.
- [Marbach e Ijspeert 2005]MARBACH, D.; IJSPEERT, A. J. Online Optimization of Modular Robot Locomotion. In: IEEE. *IEEE International Conference Mechatronics and Automation, 2005*. [S.l.], 2005. v. 1, p. 248–253.
- [Marotta et al. 2019]MAROTTA, A. et al. Network Solutions for CoMP Coordinated Scheduling. *IEEE Access*, IEEE, v. 7, p. 176624–176633, 2019.
- [Marzetta, Larsson e Yang 2016]MARZETTA, T.; LARSSON, E.; YANG, H. *Fundamentals of Massive MIMO*. [S.l.]: Cambridge Univ Press, 2016.
- [Marzouk, Barraca e Radwan 2020]MARZOUK, F.; BARRACA, J. P.; RADWAN, A. On Energy Efficient Resource Allocation in Shared RANs: Survey and Qualitative Analysis. *IEEE Communications Surveys & Tutorials*, IEEE, v. 22, n. 3, p. 1515–1538, 2020.
- [Mazyavkina et al. 2021]MAZYAVKINA, N. et al. Reinforcement Learning for Combinatorial Optimization: A Survey. *Computers & Operations Research*, Elsevier, p. 105400, 2021.
- [Mekikis et al. 2019]MEKIKIS, P.-V. et al. NFV-Enabled Experimental Platform for 5G Tactile Internet Support in Industrial Environments. *IEEE Transactions on Industrial Informatics*, IEEE, v. 16, n. 3, p. 1895–1903, 2019.
- [Mertikopoulos 2019]MERTIKOPOULOS, P. *Online Optimization and Learning in Games: Theory and Applications*. Tese (Doutorado) — Grenoble 1 UGA-Université Grenoble Alpes, 2019.
- [Mnih et al. 2016]MNIH, V. et al. Asynchronous Methods for Deep Reinforcement Learning. In: BALCAN, M. F.; WEINBERGER, K. Q. (Ed.). *Proceedings of The 33rd International Conference on Machine Learning*. New York, New York, USA: PMLR, 2016. (Proceedings of Machine Learning Research, v. 48), p. 1928–1937. Disponível em: <<http://proceedings.mlr.press/v48/mniha16.html>>.
- [Mnih et al. 2016]MNIH, V. et al. Asynchronous Methods for Deep Reinforcement Learning. In: PMLR. *International Conference on Machine Learning*. [S.l.], 2016. p. 1928–1937.
- [Mnih et al. 2013]MNIH, V. et al. *Playing Atari with Deep Reinforcement Learning*. 2013.
- [Mnih et al. 2015]MNIH, V. et al. Human-Level Control Through Deep Reinforcement Learning. *Nature*, Nature Publishing Group, v. 518, n. 7540, p. 529–533, 2015.

- [Mondal et al. 2015]MONDAL, B. et al. 3D Channel Model in 3GPP. *IEEE Communications Magazine*, IEEE, v. 53, n. 3, p. 16–23, 2015.
- [Mondal, Saha e Jang 2013]MONDAL, R. K.; SAHA, N.; JANG, Y. M. Joint Scheduling and Rate Allocation for IEEE 802.15. 7 WPAN System. In: IEEE. *2013 Fifth International Conference on Ubiquitous and Future Networks (ICUFN)*. [S.l.], 2013. p. 691–695.
- [Morais et al. 2021]MORAIS, F. Z. et al. PlaceRAN: Optimal Placement of Virtualized Network Functions in the Next-generation Radio Access Networks. *arXiv preprint arXiv:2102.13192*, 2021.
- [Morais et al. 2022]MORAIS, F. Z. et al. PlaceRAN: Optimal Placement of Virtualized Network Functions in Beyond 5G Radio Access Networks. *IEEE Transactions on Mobile Computing*, IEEE, 2022.
- [Morais et al. 2020]MORAIS, F. Z. et al. When SDN Meets C-RAN: A Survey Exploring Multi-Point Coordination, Interference, and Performance. *Journal of Network and Computer Applications*, Elsevier, v. 162, p. 102655, 2020.
- [Mostafavi, Hakami e Sanaei 2021]MOSTAFAVI, S.; HAKAMI, V.; SANAEI, M. On Energy Efficient Resource Allocation in Shared RANs: Survey and Qualitative Analysis. *Computing*, v. 103, p. 917–991, 2021.
- [Murti et al. 2023]MURTI, F. W. et al. Deep Reinforcement Learning for Orchestrating Cost-Aware Reconfigurations of vRANs. *IEEE Transactions on Network and Service Management*, IEEE, 2023.
- [Murti, Ali e Latva-aho 2021]MURTI, F. W.; ALI, S.; LATVA-AHO, M. Deep Reinforcement Based Optimization of Function Splitting in Virtualized Radio Access Networks. *arXiv preprint arXiv:2105.14731*, 2021.
- [Müller et al. 2018]MÜLLER, M. K. et al. Flexible multi-node simulation of cellular mobile communications: the Vienna 5G System Level Simulator. *EURASIP Journal on Wireless Communications and Networking*, v. 2018, n. 1, p. 17, set. 2018.
- [Nardini et al. 2020]NARDINI, G. et al. Simu5G—An OMNeT++ library for end-to-end performance evaluation of 5G networks. *IEEE Access*, IEEE, v. 8, p. 181176–181191, 2020.
- [Nardini et al. 2020]NARDINI, G. et al. Simu5G: a system-level simulator for 5G networks. In: SCITEPRESS. *Proceedings of the 10th International Conference on Simulation and Modeling Methodologies, Technologies and Applications*. [S.l.], 2020. p. 68–80.

- [Nguyen et al. 2021]NGUYEN, H. T. et al. DRL-Based Intelligent Resource Allocation for Diverse QoS in 5G and toward 6G Vehicular Networks: A Comprehensive Survey. *Wireless Communications and Mobile Computing*, Hindawi, v. 2021, 2021.
- [Nie, Li e Liu 2021]NIE, H.; LI, S.; LIU, Y. Multi-Agent Deep Reinforcement Learning for Resource Allocation in the Multi-Objective HetNet. In: IEEE. *2021 International Wireless Communications and Mobile Computing (IWCMC)*. [S.l.], 2021. p. 116–121.
- [Nikaein, Chang e Alexandris 2018]NIKAEIN, N.; CHANG, C.-Y.; ALEXANDRIS, K. Mosaic5G: Agile and flexible service platforms for 5G research. *ACM SIGCOMM Computer Communication Review*, ACM New York, NY, USA, v. 48, n. 3, p. 29–34, 2018.
- [Niyato e Hossain 2006]NIYATO, D.; HOSSAIN, E. Adaptive fair subcarrier/rate allocation in multirate OFDMA networks: Radio link level queuing performance analysis. *IEEE Transactions on Vehicular Technology*, IEEE, v. 55, n. 6, p. 1897–1907, 2006.
- [Noor-A-Rahim et al. 2020]NOOR-A-RAHIM, M. et al. A survey on resource allocation in vehicular networks. *IEEE Transactions on Intelligent Transportation Systems*, IEEE, 2020.
- [O-RAN Alliance 2019]O-RAN Alliance. *O-RAN Working Group 2: AI/ML Workflow Description and Requirements*. [S.l.], 2019. v. 1.
- [O-RAN Alliance 2020]O-RAN Alliance. *Cloud Architecture and Deployment Scenarios for O-RAN Virtualized RAN*. [S.l.], Apr 2020.
- [O-RAN Alliance 2021]O-RAN Alliance. *O-RAN Working Group 1: O-RAN Use Cases Detailed Specification 6.0*. [S.l.], Oct 2021. O-RAN.WG1.Use-Cases-Detailed-Specification-v06.00.
- [O-RAN Alliance 2022]O-RAN Alliance. *O-RAN Working Group 1: O-RAN Architecture Description 7.0*. [S.l.], Oct 2022. O-RAN.WG1.O-RAN-Architecture-Description-v07.00.
- [O-RAN Alliance 2023]O-RAN Alliance. *O-RAN Working Group 1: O-RAN Architecture Description*. [S.l.], Mar 2023.
- [Pandana e Liu 2005]PANDANA, C.; LIU, K. R. Near-optimal reinforcement learning framework for energy-aware sensor communications. *IEEE Journal on Selected Areas in Communications*, IEEE, v. 23, n. 4, p. 788–797, 2005.
- [Pang et al. 2022]PANG, G. et al. Deep reinforcement learning for radio resource allocation in NOMA-based remote state estimation. In: IEEE. *GLOBECOM 2022-2022 IEEE Global Communications Conference*. [S.l.], 2022. p. 3059–3064.

- [Pang et al. 2022]PANG, L. et al. Investigation and comparison of 5G channel models: From QuaDRiGa, NYUSIM, and MG5G perspectives. *Chinese Journal of Electronics*, Wiley Online Library, v. 31, n. 1, p. 1–17, 2022.
- [Pedregosa et al. 2011]PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.
- [Phyu, Naboulsi e Stanica 2023]PHYU, H. P.; NABOULSI, D.; STANICA, R. Machine learning in network slicing-a survey. *IEEE Access*, IEEE, 2023.
- [Piri et al. 2016]PIRI, E. et al. 5GTN: A test network for 5G application development and testing. In: IEEE. *2016 European Conference on Networks and Communications (EuCNC)*. [S.l.], 2016. p. 313–318.
- [Plaat, Kusters e Preuss 2023]PLAAT, A.; KOSTERS, W.; PREUSS, M. High-accuracy model-based reinforcement learning, a survey. *Artificial Intelligence Review*, Springer, p. 1–33, 2023.
- [Polese et al. 2023]POLESE, M. et al. Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges. *IEEE Communications Surveys & Tutorials*, IEEE, 2023.
- [Pratschner et al. 2018]PRATSCHNER, S. et al. Versatile mobile communications simulation: The Vienna 5G link level simulator. *EURASIP Journal on Wireless Communications and Networking*, Springer, v. 2018, p. 1–17, 2018.
- [Pratschner et al. 2021]PRATSCHNER, S. et al. The Vienna 5g Link Level Simulator V1. 2. *nt. tuwien. ac. at/wpcontent/uploads/2020/05/5gll_usermanual. pdf*, Seen In, v. 19, p. 25, 2021.
- [Raffin et al. 2021]RAFFIN, A. et al. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, v. 22, n. 268, p. 1–8, 2021. Disponível em: <<http://jmlr.org/papers/v22/20-1364.html>>.
- [Rappaport, Sun e Shafi 2017]RAPPAPOORT, T. S.; SUN, S.; SHAFI, M. Investigation and comparison of 3GPP and NYUSIM channel models for 5G wireless communications. In: IEEE. *2017 IEEE 86th vehicular technology conference (VTC-Fall)*. [S.l.], 2017. p. 1–5.
- [Ren, Sun e Peng 2020]REN, Y.; SUN, Y.; PENG, M. Deep Reinforcement Learning Based Computation Offloading in Fog Enabled Industrial Internet of Things. *IEEE Transactions on Industrial Informatics*, IEEE, v. 17, n. 7, p. 4978–4987, 2020.

- [Sarikaya e Onur 2021]SARIKAYA, E.; ONUR, E. Placement of 5G RAN slices in multi-tier O-RAN 5G networks with flexible functional splits. In: IEEE. *2021 17th International Conference on Network and Service Management (CNSM)*. [S.l.], 2021. p. 274–282.
- [Sarmah 2020]SARMAH, D. K. A survey on the latest development of machine learning in genetic algorithm and particle swarm optimization. In: *Optimization in Machine Learning and Applications*. [S.l.]: Springer, 2020. p. 91–112.
- [Schneider e Kirkpatrick 2007]SCHNEIDER, J.; KIRKPATRICK, S. *Stochastic optimization*. [S.l.]: Springer Science & Business Media, 2007.
- [Schwarz, Mehlführer e Rupp 2010]SCHWARZ, S.; MEHLFÜHRER, C.; RUPP, M. Low complexity approximate maximum throughput scheduling for LTE. In: IEEE. *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*. [S.l.], 2010. p. 1563–1569.
- [Sergienko, Huliannytskyi e Sirenko 2009]SERGIENKO, I.; HULIANYTSKYI, L.; SIRENKO, S. Classification of applied methods of combinatorial optimization. *Cybernetics and Systems Analysis*, Springer, v. 45, n. 5, p. 732–741, 2009.
- [Seytnazarov e Kim 2013]SEYTNAZAROV, S.; KIM, Y.-T. Cognitive rate adaptation for high throughput ieee 802.11 n wlans. In: IEEE. *2013 15th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. [S.l.], 2013. p. 1–6.
- [Shafik, Matinkhah e Sanda 2020]SHAFIK, W.; MATINKHAH, M.; SANDA, M. N. Network resource management drives machine learning: a survey and future research direction. *Journal of Communications Technology, Electronics and Computer Science*, v. 30, p. 1–15, 2020.
- [Sharma e Kumar 2021]SHARMA, N.; KUMAR, K. Resource allocation trends for ultra dense networks in 5G and beyond networks: A classification and comprehensive survey. *Physical Communication*, Elsevier, p. 101415, 2021.
- [Shekhawat et al. 2020]SHEKHAWAT, J. S. et al. A reinforcement learning framework for QoS-driven radio resource scheduler. In: IEEE. *GLOBECOM 2020-2020 IEEE Global Communications Conference*. [S.l.], 2020. p. 1–7.
- [Singla, Ghosh e Shukla 2020]SINGLA, M.; GHOSH, D.; SHUKLA, K. A survey of robust optimization based machine learning with special reference to support vector machines. *International Journal of Machine Learning and Cybernetics*, Springer, v. 11, n. 7, p. 1359–1385, 2020.

- [Su et al. 2019]SU, R. et al. Resource allocation for network slicing in 5G telecommunication networks: A survey of principles and models. *IEEE Network*, IEEE, v. 33, n. 6, p. 172–179, 2019.
- [Sun e Yang 2019]SUN, C.; YANG, C. Learning to optimize with unsupervised learning: Training deep neural networks for URLLC. In: IEEE. *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. [S.I.], 2019. p. 1–7.
- [Sun et al. 2017]SUN, H. et al. Learning to optimize: Training deep neural networks for wireless resource management. In: IEEE. *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. [S.I.], 2017. p. 1–6.
- [Sun et al. 2018]SUN, H. et al. Learning to optimize: Training deep neural networks for interference management. *IEEE Transactions on Signal Processing*, IEEE, v. 66, n. 20, p. 5438–5453, 2018.
- [Sun et al. 2019]SUN, S. et al. A survey of optimization methods from a machine learning perspective. *IEEE transactions on cybernetics*, IEEE, v. 50, n. 8, p. 3668–3681, 2019.
- [Sun et al. 2018]SUN, S. et al. Propagation models and performance evaluation for 5g millimeter-wave bands. *IEEE Transactions on Vehicular Technology*, IEEE, v. 67, n. 9, p. 8422–8439, 2018.
- [Sun et al. 2019]SUN, X. et al. Enhancing the user experience in vehicular edge computing networks: An adaptive resource allocation approach. *IEEE Access*, IEEE, v. 7, p. 161074–161087, 2019.
- [Sutton e Barto 2018]SUTTON, R. S.; BARTO, A. G. *Reinforcement Learning: An Introduction*. Second edition. [S.I.]: MIT Press, Cambridge, MA. A Bradford Book, 2018.
- [Taha 2020]TAHA, K. Methods that optimize multi-objective problems: A survey and experimental evaluation. *IEEE Access*, IEEE, v. 8, p. 80855–80878, 2020.
- [Takahashi e Tomah 2020]TAKAHASHI, K.; TOMAH, S. Online optimization of AGV transport systems using deep reinforcement learning. *Bulletin of Networking, Computing, Systems, and Software*, v. 9, n. 1, p. 53–57, 2020.
- [Tamim, Shami e Ong 2023]TAMIM, I.; SHAMI, A.; ONG, L. ALAP: Availability-and Latency-Aware Protection for O-RAN: A Deep Q-Learning Approach. *IEEE Transactions on Network and Service Management*, IEEE, 2023.

- [Tang, Zhou e Kato 2020]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*, IEEE, v. 38, n. 12, p. 2773–2782, 2020.
- [Thyagarajan et al. 2021]THYAGARAJAN, A. K. et al. Snr-cqi mapping for 5g downlink network. In: IEEE. *2021 IEEE Asia Pacific Conference on Wireless and Mobile (APWi-Mob)*. [S.l.], 2021. p. 173–177.
- [Tian et al. 2020]TIAN, Y. et al. Solving large-scale multiobjective optimization problems with sparse optimal solutions via unsupervised neural networks. *IEEE transactions on cybernetics*, IEEE, v. 51, n. 6, p. 3115–3128, 2020.
- [Towers et al. 2023]TOWERS, M. et al. *Gymnasium*. Zenodo, mar. 2023. Disponível em: <<https://zenodo.org/record/8127025>>.
- [Tran e Pompili 2017]TRAN, T. X.; POMPILI, D. Dynamic radio cooperation for user-centric cloud-ran with computing resource sharing. *IEEE Transactions on Wireless Communications*, IEEE, v. 16, n. 4, p. 2379–2393, 2017.
- [Varshney et al. 2023]VARSHNEY, R. et al. Deep learning based wireless channel prediction: 5g scenario. *Procedia Computer Science*, Elsevier, v. 218, p. 2626–2635, 2023.
- [Verma, Pant e Snasel 2021]VERMA, S.; PANT, M.; SNASEL, V. A Comprehensive Review on NSGA-II for Multi-Objective Combinatorial Optimization Problems. *IEEE Access*, IEEE, v. 9, p. 57757–57791, 2021.
- [Vesselinova et al. 2020]VESSELINOVA, N. et al. Learning combinatorial optimization on graphs: A survey with applications to networking. *IEEE Access*, IEEE, v. 8, p. 120388–120416, 2020.
- [Vijayalakshmi e Anandan 2019]VIJAYALAKSHMI, K.; ANANDAN, P. A multi objective tabu particle swarm optimization for effective cluster head selection in wsn. *Cluster computing*, Springer, v. 22, n. 5, p. 12275–12282, 2019.
- [Virdis et al. 2020]VIRDIS, A. et al. End-to-end performance evaluation of mec deployments in 5g scenarios. *Journal of Sensor and Actuator Networks*, MDPI, v. 9, n. 4, p. 57, 2020.
- [Voorneveld 2003]VOORNEVELD, M. Characterization of pareto dominance. *Operations Research Letters*, Elsevier, v. 31, n. 1, p. 7–11, 2003.
- [Vora e Kang 2018]VORA, A.; KANG, K. Downlink Scheduling and Resource Allocation for 5G MIMO Multicarrier Systems. In: *2018 IEEE 5G World Forum (5GWF)*. [S.l.: s.n.], 2018. p. 174–179.

- [Wang et al. 2020]WANG, C. et al. Mdp-based distribution network reconfiguration with renewable distributed generation: Approximate dynamic programming approach. *IEEE Transactions on Smart Grid*, IEEE, v. 11, n. 4, p. 3620–3631, 2020.
- [Wang e Chen 2013]WANG, F.-S.; CHEN, L.-H. Heuristic optimization. *Encyclopedia of Systems Biology*, Springer, New York, p. 885–885, 2013.
- [Wang et al. 2016]WANG, W. et al. A survey on applications of model-free strategy learning in cognitive wireless networks. *IEEE Communications Surveys & Tutorials*, IEEE, v. 18, n. 3, p. 1717–1757, 2016.
- [Wang, Magnússon e Johansson 2021]WANG, X.; MAGNÚSSON, S.; JOHANSSON, M. On the convergence of step decay step-size for stochastic optimization. *arXiv preprint arXiv:2102.09393*, 2021.
- [Wang et al. 2022]WANG, X. et al. Machine learning enables radio resource allocation in the downlink of ultra-low latency vehicular networks. *IEEE Access*, IEEE, v. 10, p. 44710–44723, 2022.
- [Wang et al. 2020]WANG, X. et al. Drl-based energy-efficient resource allocation frameworks for uplink noma systems. *IEEE Internet of Things Journal*, IEEE, v. 7, n. 8, p. 7279–7294, 2020.
- [Wang et al. 2019]WANG, Y. et al. A survey of dynamic spectrum allocation based on reinforcement learning algorithms in cognitive radio networks. *Artificial intelligence review*, Springer, v. 51, n. 3, p. 493–506, 2019.
- [Weise 2009]WEISE, T. Global optimization algorithms-theory and application. *Self-Published Thomas Weise*, 2009.
- [Wu et al. 2021]WU, Y. et al. Learning improvement heuristics for solving routing problems.. *IEEE Transactions on Neural Networks and Learning Systems*, IEEE, 2021.
- [Xie et al. 2016]XIE, Y. et al. Service function chaining resource allocation: A survey. *arXiv preprint arXiv:1608.00095*, 2016.
- [Xie et al. 2020]XIE, Y. et al. Virtualized network function provisioning in stochastic cloud environment. In: IEEE. *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. [S.l.], 2020. p. 1–6.
- [Xu et al. 2021]XU, C. et al. Optimizing the long-term average reward for continuing mdps: A technical report. *arXiv preprint arXiv:2104.06139*, 2021.

- [Xu et al. 2020]XU, L. et al. Fairness-aware throughput maximization over cognitive heterogeneous noma networks for industrial cognitive iot. *IEEE Transactions on Communications*, IEEE, v. 68, n. 8, p. 4723–4733, 2020.
- [Xu et al. 2021]XU, Y. et al. A survey on resource allocation for 5G heterogeneous networks: current research, future trends and challenges. *IEEE Communications Surveys & Tutorials*, IEEE, 2021.
- [Yang 2011]YANG, X.-S. Metaheuristic optimization: algorithm analysis and open problems. In: SPRINGER. *International Symposium on Experimental Algorithms*. [S.l.], 2011. p. 21–32.
- [Yang et al. 2018]YANG, Y. et al. DECCO: Deep-learning Enabled Coverage and Capacity Optimization for Massive MIMO Systems. *IEEE Access*, p. 1–1, 2018.
- [Yang e Whinston 2020]YANG, Y.; WHINSTON, A. A survey on reinforcement learning for combinatorial optimization. *arXiv preprint arXiv:2008.12248*, 2020.
- [Yassin et al. 2017]YASSIN, M. et al. Survey of icic techniques in lte networks under various mobile environment parameters. *Wireless Networks*, Springer, v. 23, p. 403–418, 2017.
- [Ye, Li e Juang 2019]YE, H.; LI, G. Y.; JUANG, B.-H. F. Deep reinforcement learning based resource allocation for v2v communications. *IEEE Transactions on Vehicular Technology*, IEEE, v. 68, n. 4, p. 3163–3173, 2019.
- [Yi et al. 2018]YI, B. et al. A comprehensive survey of network function virtualization. *Computer Networks*, Elsevier, v. 133, p. 212–262, 2018.
- [Zamzam, Elshabrawy e Ashour 2019]ZAMZAM, M.; ELSHABRAWY, T.; ASHOUR, M. Resource management using machine learning in mobile edge computing: a survey. In: IEEE. *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*. [S.l.], 2019. p. 112–117.
- [Zappone, Renzo e Debbah 2019]ZAPPONE, A.; RENZO, M. D.; DEBBAH, M. Wireless networks design in the era of deep learning: Model-based, ai-based, or both? *IEEE Transactions on Communications*, IEEE, v. 67, n. 10, p. 7331–7376, 2019.
- [Zhang, Zhou e Erol-Kantarci 2022]ZHANG, H.; ZHOU, H.; EROL-KANTARCI, M. Federated deep reinforcement learning for resource allocation in o-ran slicing. In: IEEE. *GLOBECOM 2022-2022 IEEE Global Communications Conference*. [S.l.], 2022. p. 958–963.

- [Zhang, Liu e Zeng 2019]ZHANG, Q.; LIU, F.; ZENG, C. Adaptive interference-aware VNF placement for service-customized 5G network slices. In: IEEE. *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. [S.l.], 2019. p. 2449–2457.
- [Zhao et al. 2023]ZHAO, J. et al. Cost-effective deployment for fully-decoupled radio access networks: A techno-economic approach. In: IEEE. *2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)*. [S.l.], 2023. p. 1–6.
- [Zhou et al. 2021]ZHOU, T. et al. Multi-agent reinforcement learning for online scheduling in smart factories. *Robotics and Computer-Integrated Manufacturing*, Elsevier, v. 72, p. 102202, 2021.
- [Zhou et al. 2021]ZHOU, X. et al. Joint it-facility optimization for green data centers via deep reinforcement learning. *IEEE Network*, IEEE, 2021.
- [Zhu et al. 2017]ZHU, J. et al. A New Deep-Q-Learning-Based Transmission Scheduling Mechanism for the Cognitive Internet of Things. *IEEE Internet of Things Journal*, p. 1–1, 2017.
- [Zhu et al. 2019]ZHU, Q. et al. 3GPP TR 38.901 Channel Model. *Wiley 5G Ref: The Essential 5G Reference Online*, Wiley Online Library, p. 1–35, 2019.

Apêndice I

6.1 Métodos de otimização baseados em aprendizado de máquina

Diante dos diversos desafios e potenciais limitações observadas com o uso dos métodos tradicionais de otimização, a adoção de métodos baseados em ML tem ganhado força. Em problemas de otimização que possuem instâncias que tendem a representar um alto custo computacional, ou que resultem em soluções de difícil comprovação de otimalidade, o emprego destes métodos se mostra atraente. Tal interesse também tem surgido em situações em que ocorre a busca por melhores soluções para problemas já resolvidos, incluindo os problemas de otimização combinatória [Vesselinova et al. 2020, Wu et al. 2021].

Dentre diversos fatores motivadores, observa-se dois importantes aspectos: (i) uma classe de problema compartilha uma estrutura similar entre suas instâncias, em que a diferença normalmente ocorre apenas nos seus dados, que seguem determinada distribuição de probabilidade; e (ii) por meio do aprendizado supervisionado ou por reforço, pode-se descobrir os padrões subjacentes de uma determinada classe de problema, que podem ser usados para gerar algoritmos alternativos melhores do que os projetados por humanos.

Determinados problemas de otimização, principalmente os *Multi-Objective Combinatorial Optimization Problems* (MOCOPs), não possuem algoritmos de otimização exata, minimamente genéricos, bem como formulações baseadas em aproximação, que possam ser facilmente adaptados a problemas de uma mesma classe. Quando considerados os algoritmos de aproximação, soluções quase ótimas para problemas práticos e gerais podem ser alcançadas, em que podem fornecer apenas garantias teóricas para a qualidade das soluções produzidas. No entanto, tais algoritmos são apenas de valor teórico quando sua complexidade de tempo representa um polinômio de ordem superior, e tendem a não serem realistas para instâncias de problemas do mundo real [Vesselinova et al. 2020]. Desta forma, o emprego de soluções não exatas, principal-

mente aquelas baseadas em heurísticas e meta-heurísticas, tem sido amplamente considerado, principalmente pelo balanceamento entre a qualidade das soluções e o tempo de processamento. Entretanto, o projeto de tais métodos costuma requerer o conhecimento de domínio especializado, e envolve um pesado esforço de tentativa e erro, bem como ajustes até se obter sua melhor forma. No caso de problemas multiobjetivo, é comum que cada objetivo exija um esforço específico de projeto e, muitas vezes, um algoritmo próprio. Adicionalmente, os esforços de atualização e/ou adaptação tendem a ser igualmente complexos, toda vez que a formulação do problema se altera.

Observa-se na literatura que o emprego de métodos baseados em aprendizado de máquina é cada vez mais intensificado [Gambella, Ghaddar e Naoum-Sawaya 2020], em tarefas relacionadas aos mais variados problemas de tomada de decisões. Há, portanto, uma clara relação entre tais métodos e a área de pesquisa operacional, principalmente no uso de aprendizado de máquina, e suas diferentes ferramentas. Podem ser empregados para a solução de problemas de otimização complexos e de larga escala, tanto servindo como ferramenta de suporte complementar em abordagens de otimização tradicionais, ou explorando problemas de forma isolada.

Há um crescente interesse recente na investigação de soluções para tais problemas, com suporte destes métodos, conforme pode ser observado em uma coletânea de *surveys* publicados nos últimos três anos, que são sumarizados na Tabela 6.1. Conforme pode ser observado, a grande maioria destes trabalhos aborda problemas de otimização específicos, tais como os *Combinatorial Optimization Problems* (COPs) e MOOPs, e concentram-se em mapear as abordagens empregadas, bem como os desafios e áreas de pesquisa em aberto para tais problemas.

Alguns trabalhos visam apresentar uma visão sobre certos desafios conceituais, como em [Sun et al. 2019] e [Kastratia e Bibaa 2021], que detalham os desafios envolvendo estatística de primeira ordem, como *Gradient Descent* (GD), *Stochastic GD* (SGD), *momentum* e estatística de ordem superior, bem como os chamados métodos livres de derivadas, que incluem algumas heurísticas. Os autores em [Kotary et al. 2021] e [Singla, Ghosh e Shukla 2020] apresentam de maneira genérica as *Constrained Optimizations* (COs) e as *Robust Optimizations* (ROs), respectivamente. Por último, os autores em [Karaboga, Akay e Karaboga 2020] e [Sarmah 2020] mapeiam os desafios envolvendo soluções específicas, como os algoritmos baseados em *Artificial Bee Colony* (ABC), *Genetic Algorithms* (GAs) e *Particle Swarm Optimizations* (PSOs).

Por décadas, os métodos de otimização tradicionais têm sido empregados, com sucesso, nas mais variadas tarefas na área dos sistemas de comunicação sem fio. Muitos dos algoritmos empregados pertencem à classe dos métodos baseados em busca iterativa. Utilizam como entrada uma série de parâmetros obtidos em tempo real, tais como as medidas de qualidade de canais, níveis de interferência e demandas de QoS, por

exemplo, em que uma solução otimizada deve ser providenciada ao sistema. Apesar do fato de que tais soluções apresentam boas performances, a implementação em sistemas reais continua apresentando uma série de desafios [Sun et al. 2018], o que se mostra ainda mais desafiador para as demandas previstas nos sistemas sem fio de próxima geração.

Tabela 6.1: *Surveys* recentes que abordam métodos de otimização com suporte de ferramentas de aprendizado de máquina.

Ano	Survey	Tópicos
2018	[Lombardi e Milano 2018]	COP
2019	[Alexandropoulos et al. 2019]	MOOP
2019	[Sun et al. 2019]	GD, SGD e heurísticas
2020	[Gambella, Ghaddar e Naoum-Sawaya 2020]	COP, MIP e MOOP
2020	[Karaboga, Akay e Karaboga 2020]	ABC
2020	[Sarmah 2020]	GA e PSO
2020	[Singla, Ghosh e Shukla 2020]	RO
2020	[Vesselinova et al. 2020]	COP
2020	[Yang e Whinston 2020]	COP
2021	[Bengio, Lodi e Prouvost 2021]	COP
2021	[Kastratia e Bibaa 2021]	GD, SGD e heurísticas
2021	[Kotary et al. 2021]	COP
2021	[Mazyavkina et al. 2021]	COP

Desta forma, diante dos diversos desafios observados, e seguindo nesta crescente tendência da abordagem das limitações encontradas nos métodos de otimização tradicionais, este capítulo visa apresentar importantes considerações de tais abordagens, de forma a guiar os passos da presente pesquisa. Adicionalmente, determinados aspectos e problemas encontrados no atual estágio da pesquisa são apresentados, de forma a guiar parte dos próximos esforços do trabalho.

6.2 Otimização via aprendizado supervisionado versus aprendizado por reforço

Esta seção visa descrever, dentro do conjunto de métodos e/ou ferramentas de aprendizado de máquina, aquelas com o maior potencial no sentido de oferecer o suporte necessário à solução de tantos desafios enfrentados pelos métodos de gerenciamento, escalonamento e alocação de recursos para as redes sem fio de próxima geração.

Conforme pôde ser observado na pesquisa bibliográfica desempenhada, o emprego nessas demandas de *Supervised Learning* (SL), aprendizado por reforço (RL), e aprendizado por reforço profundo (DRL), tem crescido nos últimos anos. Desde que DRL

começou a ser empregado, com seus primeiros trabalhos encontrados a partir de 2016, logo se tornou a abordagem mais presente nos trabalhos, desde então. Entretanto, não se trata, necessariamente, de uma abordagem que substitui as demais, sendo ela própria o exemplo de uma abordagem surgida da junção de outros métodos já difundidos.

O emprego de métodos de aprendizado não supervisionado não é abordado, de maneira direta, dado o seu baixo uso em tarefas de alocação de recursos, conforme observado em apenas vinte e quatro trabalhos encontrados na literatura, no mesmo período considerado na pesquisa bibliográfica (última década). Tais métodos trazem a vantagem de empregar, para o treinamento de seus modelos, conjuntos de dados de treinamento que não contém os rótulos que indicam a qual classe cada entrada pertence. Assim, torna a tarefa de obtenção de dados para treinamento mais fácil, principalmente em aplicações do mundo real. Assim, eles são muito empregados em tarefas de classificação, agrupamentos, compressão de dados e redução de dimensionalidade, principalmente. Entretanto, mesmo sendo possível encontrar o emprego de tais métodos como parte integrante do conjunto de ferramentas utilizadas, tal como visto em técnicas baseadas na metodologia *learning-to-optimize*, julgou-se não ser, por enquanto, alvo da presente pesquisa.

6.2.1 Aprendizado supervisionado

No contexto do aprendizado de máquina, ou *Machine Learning* (ML), o aprendizado supervisionado é a tarefa de encontrar uma função que ofereça o mapeamento ideal entre os pares de um conjunto de entrada (parâmetros e/ou características) e um rótulo alvo, sendo a tarefa de obtenção desta função chamada de aprendizado. Tal aprendizado é feito mediante a análise de um conjunto de dados de treinamento, devidamente associados aos seus rótulos, de forma que a função de mapeamento aprendida visa permitir que, a cada nova entrada de um conjunto ainda desconhecido, um rótulo possa ser estabelecido, no que é conhecido como generalização.

O processo de aprendizagem, de maneira geral, pode ser visto como um problema de otimização [Bengio, Lodi e Prouvost 2021]. Uma função de perda, capaz de medir a discrepância entre o rótulo predito e o rótulo real ao qual o conjunto de entrada pertença, é normalmente empregada neste tipo de otimização, e visa ser minimizada no decorrer do treinamento. Tal função de perda pode ser definida conforme o problema, tal como de regressão e classificação, por exemplo, bem como pelo próprio método de otimização definido.

Neste sentido, um problema de aprendizado supervisionado pode ser definido,

de maneira geral, conforme [Bengio, Lodi e Prouvost 2021]

$$\min_{\theta \in \mathbb{R}^p} \mathbb{E}_{X, Y \sim P} \ell(Y, f_{\theta}(X)), \quad (6-1)$$

sendo X e Y variáveis aleatórias que seguem uma distribuição conjunta de probabilidade P , e representam as entradas e os rótulos alvo, respectivamente, ℓ a função de perda a ser minimizada (sendo dependente da tarefa), obtida para cada amostra de entrada, e f_{θ} a função paramétrica de aprendizado de máquina empregado, podendo assumir um modelo linear com pesos $\theta \in \mathbb{R}^p$, a qual se pretende aprender.

A distribuição de probabilidade é, normalmente, desconhecida e inacessível, e deve ser aproximada por uma distribuição de probabilidade empírica, obtida de um conjunto finito de dados de treinamento $D_{treino} = \{(x_i, y_i)\}_i$, de forma que o problema de otimização pode ser reescrito como

$$\min_{\theta \in \mathbb{R}^p} \sum_{(x, y) \in D_{treino}} \frac{1}{|D_{treino}|} \ell(y, f_{\theta}(x)). \quad (6-2)$$

Desta forma, assim que uma redução na variação dos valores da função de perda é observada, obtêm-se o modelo treinado. Diz-se que um determinado modelo generaliza se os valores obtidos na Equação (6-2) se traduzem em baixos valores na Equação (6-1). Entretanto, como a Equação (6-1) se mantém inacessível, emprega-se uma avaliação do modelo treinado com a equação

$$\sum_{(x, y) \in D_{teste}} \frac{1}{|D_{teste}|} \ell(y, f_{\theta}(x)), \quad (6-3)$$

mediante um conjunto de dados de teste D_{teste} , que seja disjunto de D_{treino} .

6.2.2 Aprendizado por reforço

Quando se pensa sobre a natureza do ato humano de aprender, talvez a forma mais natural seja aquela baseada na observação resultante da interação com um determinado objeto ou ambiente. Embora uma parte significativa do aprendizado obtido por um indivíduo seja oriundo da transmissão de conhecimento, outra parte é obtida por uma série de interações com objetos e ambientes, em diversas situações, em uma clara relação entre causa e efeito de determinadas ações, e indicam, empiricamente, como o indivíduo deve reagir para atingir os objetivos almejados, e quais estratégias devem ser consideradas para tarefas de longo prazo.

Em ML tal forma de aprender é conhecida como aprendizado por reforço, ou *Reinforcement Learning* (RL). De maneira geral, em um de RL emprega-se um agente, capaz de tomar determinadas ações, enquanto interage com um ambiente que

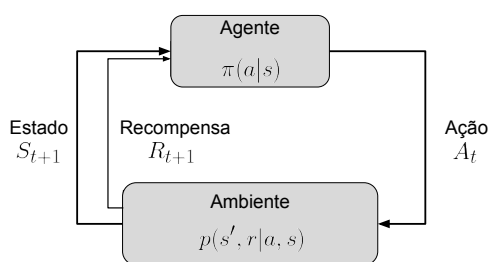


Figura 6.1: Exemplo de um sistema de RL, baseado em um MDP

replica o problema a ser aprendido, através de um processo de decisão de Markov (MDP) [Bengio, Lodi e Prouvost 2021]. Assumem um formato conforme apresentado na Figura 6.1, em que a evolução do ambiente é descrita pela dinamicidade de p , e não precisa ser previamente conhecida pelo agente, podendo ser aprendida diretamente ou indiretamente, pelo conhecido dilema de explorar *versus* explotar, em que o agente precisa escolher entre explorar novos estados do cenário, ou tirar proveito do melhor cenário observado.

Ao contrário do que se observa em métodos de aprendizado supervisionado, ou não supervisionado, que lidam com tarefas de encontrar estruturas escondidas em coleções de dados com exemplos das instâncias, com ou sem etiquetas que as classifiquem, um método de aprendizado por reforço busca maximizar uma determinada recompensa, ao contrário de tentar encontrar tais estruturas. Embora alguém possa ficar tentado a pensar na aprendizagem por reforço como um tipo de aprendizagem não supervisionada, porque não se baseia em exemplos de comportamento correto, o mais comum é classificá-lo como um terceiro paradigma de aprendizado de máquina [Sutton e Barto 2018].

Além do agente e do próprio ambiente, representado pelo espaço de observação, um sistema RL é composto de quatro elementos principais: uma política, um sinal de recompensa, uma função de valor e, opcionalmente, um modelo do ambiente. O estado precisa ser capaz de caracterizar completamente o cenário, em cada passo de tempo, no sentido de que estados futuros dependem apenas de estados passados por meio do estado atual (a propriedade de Markov), sendo que quando isso não for possível, diz-se que o cenário é parcialmente observável.

Uma política define a maneira de o agente se comportar em um determinado momento, sendo um tipo de mapeamento entre os parâmetros obtidos dos estados conhecidos do ambiente e as possíveis ações a serem tomadas, quando tais estados ocorrerem, podendo ser tanto uma função quanto uma tabela de pesquisa [Sutton e Barto 2018], ou demandar de mecanismos mais robustos, dependendo da implementação adotada. É um elemento fundamental da arquitetura do agente RL, de forma que uma política, por si só, deve ser suficiente para determinar o comportamento do agente, e costumam assumir

estruturas estocásticas, definindo probabilidades para cada ação.

Um sinal de recompensa é, geralmente, uma função capaz de representar os objetivos, bem como as restrições, do problema sendo aprendido. A cada passo de tempo, o agente recebe uma recompensa, sendo esta um número que quantifica quão próximo ao objetivo aquela ação tomada se encontra. Assim, o agente busca maximizar a recompensa, no longo prazo, devendo priorizar aquelas ações que geram a melhor recompensa, dado o estado em que o ambiente se encontra. O sinal de recompensa é, portanto, o principal mecanismo para gerar a alteração da política, de forma que as ações que resultam em menores recompensas tendem a ser preteridas pelo agente. Também, de maneira geral, o sinal de recompensa tende a ser funções estocásticas que representam os estados do ambiente e as ações a serem tomadas.

Uma função de valor, por sua vez, também se relaciona com as recompensas, e indicam quão boas as ações podem ser no longo prazo. De maneira geral, um valor de um estado representa o total de recompensas que o agente espera acumular no futuro, partindo de um determinado estado. Uma recompensa determina a imediata relação com os estados do ambiente, ao passo que a função de valor indica o desejo de longo prazo dos estados, após levarem em consideração os estados que tendem a ser encontrados, e as recompensas que devem vir deles. Tal medida permite, entre outros aspectos, que se siga eventualmente num caminho de menores recompensas imediatas, com vistas a atingir cenários que gerem melhores resultados num futuro. Adicionalmente, uma função de valor segue a equação de Bellman, representando uma solução por aproximação [Bengio, Lodi e Prouvost 2021].

Já os modelos do ambiente permitem realizar o planejamento, sendo um elemento que generaliza o comportamento do ambiente [Sutton e Barto 2018], permitindo realizar inferências sobre os próximos estados e as possíveis recompensas, sendo úteis na definição do curso das ações. Neste sentido, os métodos RL que empregam modelos do ambiente são chamados métodos baseados em modelos, ou livres de modelo, caso contrário, sendo estes métodos puramente baseados em tentativa e erro.

6.2.3 Aprendizado por reforço profundo

O aprendizado por reforço profundo, ou *Deep Reinforcement Learning* (DRL), é baseado na junção de *Deep Neural Networks* (DNNs) e RL, em que uma DNN pode ser empregada como o próprio agente do sistema RL [Bengio, Lodi e Prouvost 2021], ou para representar os estados, ou até como aproximador dos componentes do sistema RL [Lei 2021], tal como a função de valor $\hat{v}(s; \theta)$, a política $\pi(a|s; \theta)$ e o próprio modelo do sistema, caso exista, sendo θ os parâmetros/pesos da DNN.

Embora uma DNN demande de um conjunto de dados para treinamento, e bus-

que aprender os parâmetros que descrevem o relacionamento funcional entre os dados, quando considerado o seu uso em uma abordagem DRL, tal conjunto de dados não é requerido, dado o potencial de aprendizado pela interação com o ambiente provido pelo RL. Para problemas em que o número de possíveis estados, bem como as possíveis ações a serem tomadas pelo agente, possuam dimensões grandes, abordagens baseadas em MDP incorrem em problemas de dimensionalidade amplamente conhecidos [Wang et al. 2020, Xu et al. 2021], o que tende a gerar custos computacionais proibitivos, ou resultar em soluções ineficientes. Nesses casos, o emprego de métodos de aproximação costuma ser uma opção, o que torna o uso de DNN para tal tarefa algo encorajador. Assim, ao combinar DNN e RL, os agentes podem aprender por si próprios, criando políticas para maximizar as recompensas no longo prazo.

Segundo [Bengio, Lodi e Prouvost 2021], DRL pode ser solucionado por duas categorias: DQN e gradiente de política. DQN são métodos baseados em valor, normalmente baseados em tabelas, em que uma função de valor Q é obtida, permitindo calcular um valor Q para cada ação dentro do espaço de ações, e deve guiar a escolha das ações futuras (política). Já o gradiente de política é um método baseado em política, em que a política ótima pode ser aprendida diretamente da função de política, que é capaz de gerar a distribuição de probabilidade para as ações. Em uma DQN, um método Q -learning é combinado com uma DNN, que determina a função de valor Q para as ações, e pode ser observado na Figura 6.2.

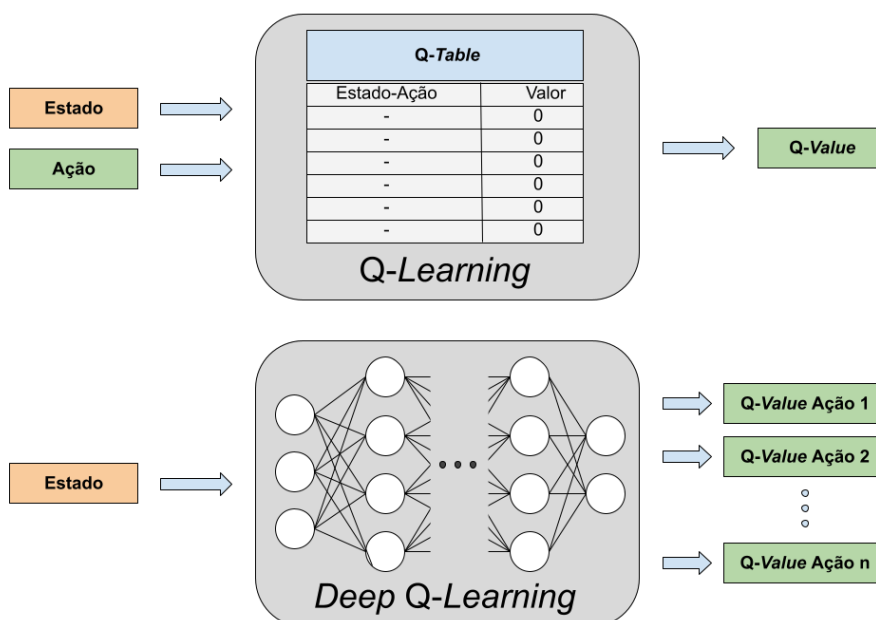


Figura 6.2: Visão geral de uma DQN, adaptado de [Bengio, Lodi e Prouvost 2021].

DRL se mostra uma importante abordagem para diversos problemas em sistemas de comunicação sem fio [Zappone, Renzo e Debbah 2019], se mostrando útil tanto

em tarefas com espaços de observação e de ações discretos quanto contínuos, bem como em problemas de otimização multiobjetivo e combinatória.

6.2.4 Abordagem *learning-to-optimize*

Em adição a todos os problemas e complexidades já considerados com relação ao emprego dos métodos de otimização tradicionais, há ainda um outro fator: a dependência de diversos processos manuais. A abordagem atual para projetar tais soluções se mostra um processo trabalhoso, em que se deve estudar o problema e desenvolver um algoritmo guiado por uma mistura de intuição, visão teórica e/ou empírica, muitas vezes de forma multidisciplinar, e considerar outros domínios, tais como quanto aos paradigmas gerais de projeto e desenvolvimento. Concluído o algoritmo para o problema em questão, é preciso analisar a sua performance, e projetar meios de avaliação de suas soluções, possivelmente comparando com outros algoritmos conhecidos, sendo a própria validação desta qualidade uma tarefa das mais complexas [Laszczyk e Myszkowski 2019]. Se o algoritmo falhar, deve-se descobrir a causa subjacente, e encontrar maneiras adequadas para superar as deficiências descobertas, e refatorar a implementação realizada. Assim, este processo segue, iterativamente, até chegar a um algoritmo superior aos algoritmos existentes, considerando um problema factível. Neste sentido, as autoras em [Li e Malik 2016] apresentam uma importante questão: podemos automatizar tal processo?

Posto desta forma, observa-se que tal processo compartilha características semelhantes às aquelas envolvidas em desafios de aprendizado de máquina por reforço, ou *Reinforcement Learning* (RL), isto é, é possível aprender qual algoritmo de otimização é melhor pela observação de sua execução, em que a convergência às soluções ideais já é um desafio conhecido. Assim, um algoritmo de otimização em particular pode ser compreendido como uma política, de forma que uma possível abordagem envolve aprender como um algoritmo de otimização se comporta e, então, encontrar uma política ótima, sendo esta uma tarefa que pode ser resolvida usando qualquer método de RL, por exemplo.

Tal abordagem (ou paradigma) emergente é conhecida como *Learning-to-Optimize* (L2O) [Li e Malik 2016, Sun et al. 2017, Chen et al. 2021], que se aproveita das ferramentas providas pela área do aprendizado de máquina (não se restringindo a apenas RL) para desenvolver métodos de otimização, visando reduzir as iterações trabalhosas da engenharia manual. Trata-se de um procedimento dirigido por dados, com a capacidade de obtenção de soluções que podem revolver, de forma eficiente, problemas similares aqueles considerados durante o treinamento [Chen et al. 2021].

Embora um método baseado em aprendizado de máquina possa carecer de uma

base teórica sólida, o que normalmente é bem definido em métodos de otimização tradicionais, a capacidade de apresentar melhora do seu desempenho, durante o processo de treinamento, é um fator definitivamente motivador. O seu emprego se mostra adequado tanto para a redução do tempo gasto na obtenção da solução ótima (ou quase), sem melhoria da sua qualidade, quanto para a obtenção de soluções de maior qualidade, sem se preocupar com redução do tempo gasto. Assim, a abordagem L2O se mostra uma possibilidade interessante, tanto em soluções com instâncias de problemas *NP*-difíceis, até para aqueles que demandam de soluções *online*.

De maneira geral, para os casos de otimização com variáveis contínuas, considera-se um problema de otimização genérico, normalmente com restrições, que pode ser modelado conforme

$$\begin{aligned} \min \quad & f(x) = (x; z), \\ \text{sujeito a} \quad & x \in \mathcal{X}, \end{aligned} \quad (6-4)$$

em que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é a função objetivo contínua (potencialmente não convexa), $x \in X$ os parâmetros do problema, em que $X \in \mathbb{R}^n$ é o espaço viável, e $z = \theta \in \mathbb{R}$ o vetor de parâmetros do problema [Sun et al. 2018, Sun e Yang 2019]. Desta forma, o objetivo envolve mapear as relações entre x e z .

Os autores em [Sun et al. 2018] apresentam um método baseado em uma DNN como aproximador no mapeamento de funções não lineares envolvidas no gerenciamento de interferências em redes sem fio, mostrando resultados próximos aos obtidos por um algoritmo *Weighted Minimum Mean Squared Error* (WMMSE), que normalmente representa um problema não convexo. Um importante benefício observado no trabalho é que a aproximação empregada representa, de fato, uma grande redução em complexidade computacional, dado que as camadas da DNN empregam poucas operações. Em uma primeira contribuição do trabalho, os autores consideram o problema modelado como

$$\begin{aligned} \min \quad & f(x) = (x^2 - z)^2, \\ \text{sujeito a} \quad & x \in \mathcal{X} \end{aligned} \quad (6-5)$$

sendo $x \in \mathbb{R}$ a variável de otimização e $z \in \mathbb{R}$ o parâmetro do problema, de forma que a equação

$$x^{t+1} = x^t - \alpha \nabla f(x^t) = x^t - 4\alpha x^t((x^t)^2 - z), t = 0, \dots, T, \quad (6-6)$$

demonstra o processo iterativo de um método baseado em gradiente descendente (GD), em que $t \in T$ é o passo de iteração, x^0 uma solução inicial qualquer, e $\alpha > 0$ o tamanho do passo. Dado um conjunto de treinamento com as amostras $\{z^i, (x^T)^i\} \forall i \in I$, o mapeamento $(x^0, z) \rightarrow x^T$, que representa o comportamento do GD, pode ser aproximado com acurácia, com uma rede neural com apenas três camadas. Os resultados demonstram que a qualidade das soluções obtidas pelo método proposto, medidas em

termos de *sum-rate* médio, podem ser até 15% inferiores ao WMMSE, mas representando um tempo computacional de apenas 3% do tempo requerido pelo WMMSE, em um cenário integralmente simulado.

Em [Sun e Yang 2019] uma abordagem semelhante é apresentada, também empregando uma DNN, considerando a alocação de recursos em redes sensíveis ao atraso, em especial voltado para aplicações URLLC. Em aplicações reais, com emprego de métodos de otimização tradicionais, o espaço de soluções precisa ser recriado toda vez que os parâmetros se alteram, muitas vezes de forma *online*, tais como os parâmetros dos canais, por exemplo, mostrando-se uma tarefa desafiadora. Dado que as soluções obtidas são, frequentemente, imprecisas, não podendo ser diretamente aplicadas para garantia de QoS, o arcabouço proposto busca empregar as soluções obtidas, de forma não supervisionada, como sinal de supervisão.

Em uma formulação baseada em RL, as autoras em [Li e Malik 2016] apresentam um arcabouço em que qualquer algoritmo de otimização contínua pode ser considerado uma política π , de forma que tal algoritmo possa ser aprendido, utilizando busca guiada por políticas. Como é comum em soluções de RL, apoiam-se na estrutura de um MDP de horizonte finito, com espaços de ações e de estado contínuos, definido pela tupla $(\mathcal{S}, \mathcal{A}, p_0, p, c, \gamma)$, em que \mathcal{S} representa o espaço de estados, \mathcal{A} o espaço de ações, p_0 a distribuição de probabilidades do estado inicial, p a densidade de probabilidades de transição, isto é, que descreve a transição de estados e ações, c é a função que mapeia os estados e os custos, e $\gamma \in [0, 1]$ o fator de desconto. Assim, o objetivo é aprender uma política estocástica $\pi^* : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}^+$, baseada em uma densidade de probabilidade condicional sobre as ações, dado o estado atual, de modo que o custo acumulado esperado seja minimizado por

$$\pi^* = \arg \min_{\pi} \mathbb{E}_{s_0, a_0, s_1, \dots, s_T} \left[\sum_{t=0}^T \gamma^t c(s_t) \right], \quad (6-7)$$

sendo a esperança obtida com relação à distribuição conjunta sobre a sequência de estados e ações (trajetória), sendo tal distribuição definida por

$$q(s_0, a_0, s_1, \dots, s_t) = p_0(s_0) \prod_{t=0}^{T-1} \pi(a_t | s_t) p(s_{t+1} | s_t, a_t). \quad (6-8)$$

O problema de encontrar a minimização do custo da política é definido como o problema de pesquisa de política [Li e Malik 2016]. Como resolver tal problema, com métodos exatos, tende a ser intratável em muitos dos casos de problemas complexos e de grande escala, ele é resolvido por aproximação. Entretanto, em muitos problemas do mundo real p , que caracteriza a dinamicidade, é desconhecido, e difícil de ser estimado. Por tal razão, as autoras se apoiam em uma versão guiada desta busca, que funciona de forma

alternada entre a tarefa de computar uma distribuição alvo, ao longo das trajetórias, em que é incentivada a minimização dos custos, e de realizar a aprendizagem da política, de uma forma supervisionada, de modo que as trajetórias de amostra de execução da política sejam próximas às trajetórias de amostra traçadas a partir da distribuição alvo.

Após demonstrar que os problemas de otimização se diferenciam, de maneira geral, na escolha da política π , assume-se que se for possível aprendê-la, um processo automatizado para o aprendizado de um algoritmo de otimização pode ser construído. Assim, o emprego de RL no aprendizado da política π é proposto, sendo necessário definir uma função de custo, que deve penalizar as políticas que apresentem comportamentos indesejáveis durante a execução. Adicionalmente, deve-se penalizar as políticas que apresentem lenta convergência, caso exista uma métrica de performance definida para o problema. Para tanto, assumindo a minimização da função objetivo, define-se o custo em um estado como o valor objetivo no local atual, o que encoraja a política a atingir o mínimo da função objetivo o mais rápido possível. Os resultados apresentados demonstram que a proposta apresenta resultados semelhantes aos obtidos por algoritmos baseados em GD e *momentum*, tanto para problemas convexos e não convexos.

Apêndice II

7.1 Métodos clássicos de otimização

A área que busca quantificar e qualificar as possíveis soluções de um dado problema é conhecida como pesquisa operacional. De maneira geral, a pesquisa operacional lida com métodos de otimização, oriundos da teoria da otimização [Doering et al. 2019], sendo estes baseados nos processos que buscam encontrar soluções adequadas para um determinado problema [Weise 2009]. Os métodos de otimização têm cada vez mais sido aplicados aos mais diferentes propósitos nas últimas décadas. De fato, diversas contribuições teóricas e práticas têm sido propostas para a solução de problemas que envolvam a tomada de decisão nas mais diversas áreas do conhecimento, como nas engenharias e telecomunicações, medicina, biologia, finanças, segurança e computação, entre outros. E a quantidade de ferramentas disponíveis para auxílio à solução destes problemas cresceu de forma proporcional à quantidade de aplicações possíveis.

Trata-se, sobretudo, de uma ferramenta empregada no processo de tomada de decisões, que visa considerar um possível curso de ações a serem tomadas, dentre todas aquelas disponíveis [Hwang e Masud 2012, Doering et al. 2019]. Em muitos dos problemas encontrados no mundo real, a multiplicidade de critérios a serem julgados na escolha das melhores ações torna a tomada de decisões algo complexo e desafiador. Embora a otimização matemática tenha sido, cada vez mais, solucionada de forma assistida por computadores, implementando algoritmos cada vez mais especializados, certas classes de problema continuam a conter instâncias em que tempo de processamento da otimização é, de fato, proibitivo, ou não controlável.

Os desenvolvimentos recentes nesta temática podem ser divididos, basicamente, entre os modelos de otimização e os métodos de otimização [Enayaty-Ahangar, Albert e DuBois 2020]. Dentre os modelos de otimização, destacam-se a programação linear, programação inteira, programação não linear e processos de decisão de Markov, entre outros. Dentre os métodos de otimização, duas abordagens são consideradas: os métodos determinísticos (também

conhecidos como exatos ou não estocásticos) e os métodos não determinísticos [Hanagandi e Nikolaou 1998, Hendricks, Jannerup e Sørensen 2008, Duchi 2018].

Ao contrário do que ocorre nos métodos determinísticos, que sempre levam à mesma solução, caso partam de um mesmo ponto inicial, dada a possibilidade de se obter a sequência de pontos que cobrem todas as possíveis soluções do problema, de forma analítica, a um custo computacional normalmente alto, os métodos não determinísticos, que incluem os métodos heurísticos, meta-heurísticos e estocásticos, além de diversos casos envolvendo modelos *online*, buscam encontrar soluções mais próximas o possível das soluções ótimas. Cabe ao tomador de decisões avaliar o compromisso entre o custo envolvido no processo de busca das soluções no espaço viável e a qualidade aceitável da solução requerida pelo problema.

Uma outra importante diferenciação é que as soluções encontradas pelos métodos exatos podem ter sua otimalidade matematicamente comprovada, ao passo que tal comprovação é, normalmente, inviável em métodos não exatos. Mesmo que os métodos exatos consigam atingir uma solução ótima, em um tempo aceitável, é comum que empreguem um significativo esforço em tarefas relacionadas à prova da otimalidade da solução encontrada [Hanafi e Todosijević 2017]. Tal fato é determinante no incremento da complexidade computacional de certas instâncias dos problemas sob otimização.

Nesta temática, existem diversas taxonomias e classificações encontradas na literatura [Sergienko, Huliannytskyi e Sirenko 2009]. Entretanto, de forma a prover uma visualização sobre os desafios e potenciais soluções que permeiam os desafios da alocação e gerenciamento de recursos em redes sem fio de próxima geração, esta seção concentra-se na descrição dos modelos de otimização aderentes à tais desafios. Adicionalmente, a visão sobre os problemas de otimização multiobjetivo, bem como de otimização combinatória, são apresentadas, dado se tratar dos tipos de problemas normalmente enfrentados na temática abordada na pesquisa.

7.1.1 Otimização combinatória e multiobjetivo

Os problemas de otimização multiobjetivo, ou MOOPs, também conhecidos como Programação Multiobjetivo [Hwang e Masud 2012, Taha 2020], Otimização Vetorial, Otimização Multicritério/Multiatributo ou Otimização de Pareto, fazem parte de uma classe de problemas que ocorrem de forma regular no mundo real, em que a função objetivo a ser maximizada ou minimizada envolve dois ou mais objetivos a serem otimizados simultaneamente [Li, Zhang e Wang 2020], muitas vezes de forma antagônica. Envolve, portanto, o desafio da identificação de uma solução adequada, diante de situações com diferentes tipos de desejos. Assim, um algoritmo de solução de problemas de objetivo único não é aplicável [Taha 2020]. De maneira simplificada, um MOOP pode ser definido

como:

$$\max/\min \quad z(x) = (f_1(x), f_2(x), \dots, f_k(x)), \quad \forall k \geq 2, \quad (7-1)$$

$$\text{sujeito a} \quad x \in X, \quad (7-2)$$

em que $f_k(x)$ representa a k -ésima função objetivo (ou função utilidade) do problema, X representa o espaço de soluções viáveis (ou espaço objetivo), e $Z = \{z(x) : x \in X\}$ sendo a imagem de X em \mathcal{R}^k , de forma que $z = (z_1, z_2, \dots, z_K) \in Z$ o vetor objetivo.

Quando $K = 1$, tem-se um problema de otimização mono-objetivo. A Figura 7.1 demonstra as soluções obtidas para um problema mono-objetivo qualquer, sem restrições, sendo a função objetivo $f_1(x) = (x - 5)^2$, na qual a otimização empregada busca a sua minimização, sendo a solução ótima obtida para $x = 5$. Na Figura 7.2 vê-se um exemplo de um problema multiobjetivo qualquer, de forma que se a otimização buscar a minimização/maximização conjunta das K funções, métodos simples que empregam condições de primeira e segunda ordem (ou operadores de Lagrange, para os casos não lineares [Frangioni 2005]) podem ser suficientes para a obtenção da(s) solução(ões) ótima(s). Entretanto, caso tais funções sejam conflitantes, métodos mais robustos precisam ser empregados, de forma a explorar o espaço de soluções viáveis.

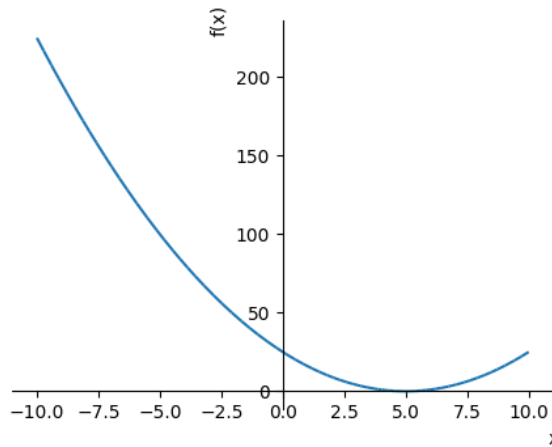


Figura 7.1: Exemplo de um problema mono-objetivo.

Nestas situações de conflito entre as K funções objetivo do referido problema, o método de otimização deve ponderar entre as possíveis soluções encontradas, de forma a encontrar qual solução é melhor, dentre elas. Isto é, busca encontrar a solução ótima dentre aquelas mapeadas no espaço de soluções viáveis. Em outras palavras, devido ao típico conflito de objetivos, a noção de solução ótima costuma não existir em MOOPs [Lust e Teghem 2010]. Desta forma, surge o conceito de dominância e otimalidade de Pareto [Voorneveld 2003].

A Figura 7.3 (adaptada de [Hashimoto 2004]) apresenta o espaço objetivo factível $Z = \{z(x) : x \in X\}$ de um problema de minimização qualquer, composto de dois objetivos

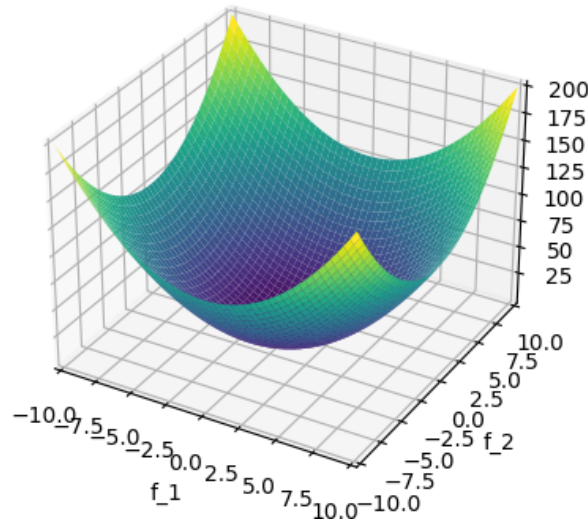


Figura 7.2: Exemplo de um problema multiobjetivo.

conflitantes, sendo eles $z_1(x) = f_1(x)$ e $z_2(x) = f_2(x)$. Conforme se observa, tanto A quanto B representam pontos (vetor objetivo) contidos na fronteira (ou frente) de Pareto, que comporta todas as soluções factíveis ótimas, isto é, que apresentam um mesmo valor para a função objetivo modelada para o problema. O ponto A demonstra um melhor valor para f_2 que o ponto B, ao passo que o ponto B demonstra uma melhor valor para f_1 que o ponto A. Assim, a redução do valor de uma função objetivo gera o incremento da outra.

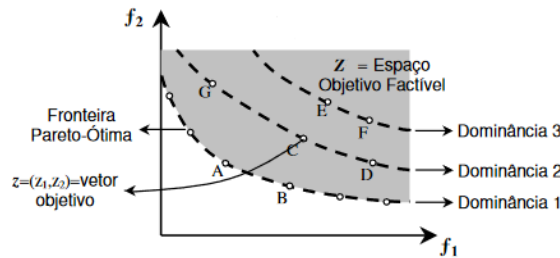


Figura 7.3: Exemplo de um problema com 2 objetivos, e as suas possíveis soluções.

A dominância de Pareto estabelece a relação entre possíveis soluções, servindo como um dos importantes meios de se medir a qualidade das soluções contidas no espaço de soluções factíveis [Laszczyk e Myszkowski 2019]. Considerando duas soluções quaisquer, $x, y \in Z$, para um problema de minimização, sendo I o conjunto de critérios, têm-se as seguintes relações:

- Se $f_i(x) \leq f_i(y) \quad \forall i \in I$, e se $f_i(x) < f_i(y) \quad \exists i \in I$ diz-se que a solução x domina a solução y . Por sua vez, y é dominado pela solução x ;
- Se $f_i(x) \not\leq f_i(y)$, e $f_i(y) \not\leq f_i(x) \quad \forall i \in I$, diz-se que as soluções x e y são indiferentes entre si.

Uma solução é dita Pareto ótimo se ela não é dominada por nenhuma outra solução do espaço de soluções, sendo que nenhuma melhoria do resultado da função objetivo pode ser obtida sem degradar a solução. Na Figura 7.3, as soluções A e B compõem a frente de Pareto ótima, ou simplesmente são soluções eficientes, dominando todas as demais soluções, e sendo indiferentes entre si. Já as soluções C, D e G, dominam as soluções E e F, e são dominadas por A e B. O ponto ideal é aquele que representa, para cada função objetivo avaliada isoladamente, o melhor valor possível, o que tende a ser um ponto fora do espaço de soluções viáveis, sendo um ponto não realista, o que costuma ser chamado de ponto utópico [Coello, Dhaenens e Jourdan 2010]. Por outro lado, o ponto que marca a pior solução possível, para ambas as funções objetivo consideradas, é chamado de ponto de Nadir. Diversos destes conceitos podem ser empregados na medição da qualidade das soluções obtidas [Laszczyk e Myszkowski 2019].

A solução de um MOOP, de maneira geral, consiste em determinar o conjunto eficiente de soluções, um subconjunto do conjunto eficiente, ou ainda, alguma solução tão próxima quanto possível de soluções que compõem a frente de Pareto ótima. Vem desta demanda parte significativa da complexidade computacional que certos problemas de otimização incorrem. Em problemas com instâncias que gerem um espaço de soluções viáveis muito grande, e que o método de otimização seja exato, a quantidade de buscas de soluções tende a levar o tempo de processamento do algoritmo a tempos exponenciais.

Algoritmos exatos garantem o retorno da solução ótima em um tempo finito ou concluem que ela não existe se o problema é insolúvel. Os algoritmos mais populares são a busca exaustiva, algoritmo *branch-and-bound*, o método *branch-and-cut*, análise sequencial e eliminação de alternativas, e a programação dinâmica [Sergienko, Hulianyskyi e Sirenko 2009].

Embora muitos dos MOOPs sejam baseados em variáveis de decisão contínuas, em diversas aplicações do mundo real as variáveis discretas de decisão precisam ser consideradas. Os problemas de otimização combinatória, ou COPs, são problemas caracterizados pelo emprego de variáveis de decisão discretas, mesmo que possam empregar funções objetivo e restrições de qualquer forma. Tais problemas têm atraído a atenção dos grupos da matemática e ciências da computação, entre outras, há anos, devido a sua ampla aplicabilidade em problemas reais, tais como no roteamento, programação, planejamento, processos de tomada de decisão, transporte e telecomunicações, entre outros [Coello, Dhaenens e Jourdan 2010, Cai et al. 2017, Vesselinova et al. 2020, Mazyavkina et al. 2021].

Os problemas de otimização combinatória multiobjetivo, ou MOCOPs, são parte integrante de uma classe particular dos MOOPs, no qual as soluções do problema são obtidas mediante o arranjo combinatório das variáveis de decisão, sendo formulados como

[Coello, Dhaenens e Jourdan 2010, Verma, Pant e Snasel 2021]:

$$\max/\min \quad y = f(x) = (f_1(x), f_2(x), \dots, f_k(x)) \quad \forall k \geq 2 \quad (7-3)$$

$$\text{sujeito a} \quad x \in X^* \quad (7-4)$$

$$g_i(x) \leq 0, \quad \forall i = 1, 2, \dots, m \quad (7-5)$$

$$h_j(x) = 0, \quad \forall j = 1, 2, \dots, p \quad (7-6)$$

onde $x = (x_1, x_2, \dots, x_n) \in X$ é um vetor de decisão n -dimensional no espaço de decisões X , Z é a imagem de X (ou espaço objetivo), em que $y = (y_1, y_2, \dots, y_k) \in Z$ é um vetor objetivo k -dimensional em \mathcal{R}^k , f é a função de mapeamento da k -ésima função objetivo a ser otimizada (maximização ou minimização), g_i é a i -ésima restrição de desigualdade e h_j a j -ésima restrição de igualdade, k , m e p representam o número de funções objetivo, restrições de desigualdade e restrições de igualdade, respectivamente. Por fim, X^* é o conjunto de todas as soluções factíveis que satisfazem (7-5) e (7-6) e que descrevem a estrutura combinatória do problema em questão.

Apesar de sua aplicação estar presente nas mais diversas áreas, tais problemas de otimização continuam a apresentar diversos desafios. O principal objetivo da otimização combinatória é conceber algoritmos eficientes para resolver tais problemas [Vesselinova et al. 2020]. Sob o ponto de vista computacional, uma solução baseada em algoritmo é eficiente se a quantidade de passos elementares empregados na solução do problema cresce de forma polinomial, de acordo com o tamanho n da entrada. Os problemas que podem ser resolvidos, de forma determinística, em um tempo polinomial, ou que possuam instâncias que possam ser verificadas em tempo polinomial, são chamados de problemas em \mathcal{P} . Entretanto, a maioria dos MOCOPs faz parte da classe \mathcal{NP} , que são aqueles intratáveis em tempo polinomial, ou até da classe \mathcal{NP} -difícil [Vesselinova et al. 2020]. Desta forma, métodos de otimização especializados precisam ser considerados.

7.1.2 Heurísticas e meta-heurísticas

Embora os métodos de otimização exata tenham a capacidade de resolver tanto os problemas mais simples, quanto os mais complexos, em que a convergência à solução ótima global se torna garantida, e pode ser matematicamente comprovada [Schneider e Kirkpatrick 2007], alguns dos problemas sob otimização podem conter certas instâncias em que o tempo de computação pode crescer exponencialmente [Al-Ghamdi e Al-Masalmeh 2020].

Em problemas do mundo real, certos tipos de restrições (ou variáveis do problema) podem aparecer durante ou após à sua modelagem, pelos mais diversos motivos,

seja pelo desconhecimento, pela falta de clareza nos passos iniciais da modelagem, devido a inclusão de novos requisitos, ou até pela alteração de leis que impactem a empresa, por exemplo, fazendo com que as soluções obtidas deixem de ser válidas. Nestas situações, é comum que a estrutura do algoritmo precise ser inteiramente remodelada, podendo fazer com que os esforços anteriores de pesquisa e desenvolvimento, bem como os recursos empregados na obtenção das soluções previamente estabelecidas, sejam inteiramente desperdiçados. Adicionalmente, devido a alta customização que tais algoritmos normalmente requerem, a impossibilidade de que um determinado algoritmo de otimização possa ser simplesmente empregado em outro problema idêntico, mas que utilize outros parâmetros, se torna uma realidade.

Importantes avanços nas áreas da pesquisa operacional e da ciência da computação acrescentaram novas abordagens de solução destes problemas à teoria de otimização, tais como as heurísticas e meta-heurísticas. Segundo descreve a psicologia, um processo decisório envolve a lógica, estatística ou heurísticas, sendo estas as ferramentas mentais (ou processos cognitivos) empregadas em decisões não racionais, em que se ignora parte da informação, com o objetivo de facilitar o processo de escolha [Gigerenzer e Gaissmaier 2011]. No contexto da otimização, uma heurística emprega um procedimento computacional que busca determinar uma solução ótima (ou quase) tentando, de forma iterativa, melhorar uma solução candidata com relação a uma determinada medida de qualidade [Wang e Chen 2013]. Já as meta-heurísticas são modelos gerais que podem ser facilmente adaptados para abordar uma ampla gama de problemas [Doering et al. 2019], sendo normalmente relacionados a algoritmos que atuam com o foco dos algoritmos heurísticos, mas que não são projetados especificamente para um problema particular [Blum et al. 2011].

Os métodos baseados em heurísticas fazem poucas (ou nenhuma) suposições sobre o problema que está sendo otimizado. Podem realizar buscas em grandes espaços de soluções candidatas, com vistas à encontrar soluções ótimas ou quase ótimas, a um custo computacional razoável. Porém, a garantia da convergência ao ótimo global não é obrigatória, embora desejável. Em muitos casos, concentra-se em obter soluções viáveis, normalmente sendo ótimos locais, tão próximas quanto o possível das soluções ótimas globais.

De forma parecida, os métodos baseados em meta-heurísticas têm se mostrado capazes de obter soluções quase ótimas, em um custo computacional relativamente menor, tornando ambas abordagens alternativas atraentes para a solução de problemas de otimização nas diversas áreas do conhecimento [Doering et al. 2019]. Um caso ideal é quando a meta-heurística pode ser usada sem qualquer conhecimento dependente do problema, em que o problema sob otimização é tratado no mais alto nível possível, com foco na simplicidade e generalidade [Lourenço, Martin e Stützle 2019].

Uma importante vantagem de tais abordagens é a facilidade normalmente encontrada nas tarefas envolvendo a implementação e adaptação dos algoritmos, inclusive para a sua adequação quanto às alterações ocorridas na modelagem do problema. Restrições adicionais podem ser implementadas com maior facilidade, e com a possibilidade do emprego do próprio algoritmo para outros problemas, com fácil ajuste dos parâmetros necessários.

Soluções de otimização baseadas em meta-heurísticas podem ser classificadas entre dois grupos distintos [Blum et al. 2011, Yang 2011, Doering et al. 2019]: (i) meta-heurísticas inspiradas na natureza, que se dividem entre os algoritmos evolutivos, tais como os algoritmos genéticos, programação genética e evolução diferencial, e entre os algoritmos de inteligência de enxame (ou baseados em população), como a otimização por enxame de partículas, otimização de colônia de formigas ou abelhas, busca *cuckoo*, busca harmônica, algoritmos inspirados em morcegos e algoritmos *firefly*; e (ii) meta-heurísticas não naturais, que incluem os métodos baseados em busca local, como a busca tabú, busca de vizinhança variável, busca local por iteração, iteração *greedy*, recozimento simulado, procedimento de busca adaptativa gananciosa, etc..

Um exemplo de meta-heurística é a busca local por iteração, ou ILS [Lourenço, Martin e Stützle 2019], isto é, um método baseado em população, com busca local. Tal meta-heurística é baseada em uma simples ideia, em que o algoritmo ILS produz uma solução inicial empregando alguma política previamente determinada, modelada de forma aderente ao problema, ao invés de utilizar uma solução inicial aleatoriamente gerada, como realizado em outras soluções deste tipo. Assim, um processo de iteração é realizado, em que uma perturbação é empregada, gerando uma possível nova solução. Tal perturbação visa obter soluções viáveis, advindas de uma espécie de campo de atração gerado por algum ótimo local.

O Algoritmo (2), adaptado de [Blum et al. 2011], apresenta um pseudocódigo genérico para uma ILS, corroborado pela Figura 7.4 (adaptada de [Lourenço, Martin e Stützle 2019]). Conforme se observa, uma solução inicial é estabelecida, e uma busca local é realizada. A construção da solução inicial deve ser computacionalmente barata e, sempre que possível, representar o melhor ponto de partida possível. Dentro do processo iterativo, uma perturbação é gerada, resultando na solução perturbada s' . O processo de perturbar uma solução é, normalmente, não determinístico, e possui como característica oferecer ao algoritmo a capacidade de escapar da atração gerada pelo mínimo local, mas sem ser forte o suficiente ao ponto de, na prática, representar um reposicionamento no espaço de soluções puramente aleatório (como em um método de reinício randômico de busca local). Assim, a solução \hat{s}' precisa ser avaliada, podendo ser, ou não, aceita como uma nova solução.

Embora o emprego de heurísticas e meta-heurísticas têm sido de grande apelo

Algoritmo 2: Busca local por iteração.

```

1  $s \leftarrow \text{GerarSolucaoInicial}();$ 
2  $\hat{s} \leftarrow \text{BuscaLocal}(s);$ 
3 while não convergir do
4    $s' \leftarrow \text{Perturbar}(\hat{s}, \text{historico});$ 
5    $\hat{s}' \leftarrow \text{BuscaLocal}(s');$ 
6    $\hat{s} \leftarrow \text{AplicaCritérioDeAceitacao}(\hat{s}', \hat{s}, \text{historico});$ 

```

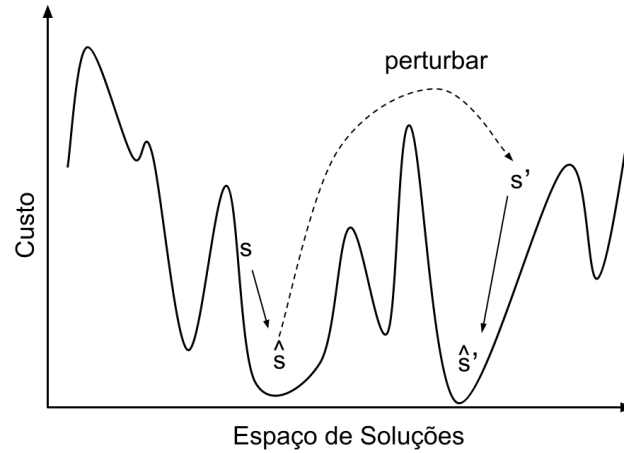


Figura 7.4: Exemplo de uma ILS, iniciando com uma solução inicial s , que resulta na solução \hat{s} após uma busca local, em que se aplica uma perturbação que leva à solução s' , resultando na solução \hat{s}' após uma nova busca local.

nos últimos anos, sendo abordagens que não podem ser desconsideradas quando se trata de problemas do mundo real. Visto que podem possuir instâncias em que a solução impõe um tempo de processamento proibitivo, elas ainda confiam principalmente no conhecimento e na experiência de especialistas humanos, que podem ser limitados e incapazes de fornecer soluções de alta qualidade [Wu et al. 2021].

Quando se torna claro que tais abordagens atingem o seu limite, tanto em termos de melhorias das soluções quase ótimas encontradas, ou da otimização do tempo desempenhado em encontrá-las, uma possível solução envolve o emprego de métodos híbridos (ou meta-heurísticas híbridas), misturando diferentes heurísticas e/ou meta-heurísticas, bem como empregando métodos exatos em conjuntos com os métodos não exatos [Blum et al. 2011]. Entretanto, o desenvolvimento de uma abordagem híbrida efetiva é, de maneira geral, uma tarefa difícil, que também requer expertise de diferentes áreas da otimização. Adicionalmente, observa-se que a generalização dos métodos de otimização que serão compostos nem sempre é trivial.

Apesar do rápido desenvolvimento das meta-heurísticas, sua análise matemática continua parcialmente não resolvida, em que diversos desafios continuam em aberto. A dificuldade costuma envolver o fato de que a interação dos vários componentes de seus

algoritmos costuma gerar problemas não lineares, complexos e estocásticos [Yang 2011].

À medida que as meta-heurísticas se tornaram cada vez mais sofisticadas, o caso ideal da sua aplicação, que considera soluções não dependentes de conhecimento do problema, tende a ser deixado de lado na busca por maior (ou melhor) desempenho. Como consequência, o conhecimento específico do problema (além daquele embutido na heurística inerente) deve agora ser incorporado às meta-heurísticas, a fim de atingir o nível de estado da arte. Assim, a fronteira entre as heurísticas e meta-heurísticas se torna, cada vez mais, tênue, incorrendo no risco da perda tanto da simplicidade quanto da generalidade encontradas nestas soluções.

7.1.3 Otimização estocástica

Diversos tipos de problemas do mundo real requerem, em sua modelagem, a inclusão de dados dinâmicos e, muitas vezes, advindos de eventos incertos. Neste sentido, há um crescente interesse observado na comunidade científica pelo emprego de métodos de otimização que considerem a incerteza e dinamicidade em suas formulações matemáticas [Bianchi et al. 2009]. Assim, a otimização estocástica surge, sendo métodos de otimização em que parte da informação sobre o problema é parcialmente conhecida, e costumam depender do conhecimento da distribuição de probabilidade em que o problema esteja inserido.

Trata-se de um arcabouço, ou abordagem geral, para a modelagem de problemas que envolvem um determinado grau de incerteza sobre variáveis que compõem o problema. Ao contrário dos métodos de otimização determinísticos, que lidam com parâmetros conhecidos, os métodos estocásticos somente podem garantir a otimalidade global da solução em termos de uma certa probabilidade, pois lida com parâmetros desconhecidos ou incertos, ou que podem assumir valores baseados em certa probabilidade.

Tais problemas podem possuir formas parecidas com aqueles observados em modelos de otimização determinísticos. Entretanto, com a otimização estocástica, busca-se obter vantagem do fato de que, embora tais dados não possam ser conhecidos de forma determinística, as distribuições de probabilidade que os governam podem ser conhecidas, *a priori*, ou podem ser estimadas.

Embora seja muito comum o seu emprego em demandas das ciências econômicas, encontra-se na literatura o seu emprego em diferentes áreas, como no problema de alocação de recursos [Sun et al. 2019, Alnabelsi, Salameh e Albataineh 2020, Fontaine, Mannor e Perchet 2020], fatiamento de recursos em redes 5G [Kasgari e Saad 2018], seleção de feixes e pré-codificação em redes MIMO massivo [Lee e Lee 2017], bem como em demandas associadas ao problema de posicionamento de VNFs [Xie et al. 2020], entre outros.

Os mais simples problemas de otimização estocástica, ou *Stochastic Optimization Problems* (SOPs), empregam um valor esperado conjuntamente na formulação da função objetivo, de forma que uma decisão só é possível antes da observação de algum evento aleatório influenciando o resultado. Na área financeira, por exemplo, em que o investidor precisa decidir sobre a alocação de ativos ou pela decomposição da sua carteira de investimentos, e ao fim de um determinado período, ele observará o resultado da carteira (lucros ou perdas), observa-se que a única certeza encontra-se no fato de que o resultado advém, de fato, de um conjunto de incertezas, desconhecidas no momento da tomada de decisões. Este resultado aleatório pode ser representado matematicamente por

$$f(x, Y), \quad (7-7)$$

em que x descreve a decisão a ser tomada, e Y a variável aleatória modelada do cenário incerto. Em termos matemáticos, a Equação (7-7) representa uma versão simplificada de uma variável aleatória de valor real, atribuída a cada cenário $\omega \in \Omega$, sendo Ω o espaço amostral em (Ω, \mathcal{A}, P) . Como cada cenário ω determina a realização da variável aleatória $Y(\omega)$, sendo Y completamente determinada por $Y(\omega)$, a formulação pode, então, ser simplificada como $f(x, \omega)$. Assim, um SOP, em sua versão mais simplificada, pode ser modelado pelo valor esperado obtido por [Gutjahr e Pichler 2016]

$$\max/\min \mathbb{E}[f(x, \omega)], \quad (7-8)$$

sendo sujeito a $x \in \mathcal{X}$, em que \mathcal{X} representa o espaço de soluções possíveis (ou espaço de decisão), que descreve as decisões a serem tomadas.

Em determinados contextos, como no caso do SOP da alocação da carteira de investimentos, a formulação em Equação (7-8) ainda não é capaz de apresentar soluções adequadas, dado que ainda há elementos a serem modelados e considerados, tal como o risco, que também é incerto. Neste tipo de problema, o risco ρ pode ser medido, sendo tal medida também adaptada para outros problemas. Desta forma, a função objetivo pode ser reescrita como

$$\begin{aligned} &\max/\min \mathbb{E}[f(x, \omega)], \\ &\text{sujeito a } \rho(f(x, \omega)) \leq r, \quad x \in \mathcal{X}, \end{aligned} \quad (7-9)$$

de forma que uma decisão x só será válida se o risco associado à $f(x, \omega)$ não exceda um determinado limiar definido por r . Desta forma, dependendo de r , o problema se torna possivelmente inviável. Adicionalmente, também é possível formular o problema de forma a buscar a minimização de ρ , enquanto estabelece um determinado limiar μ para a própria

função objetivo, tal como em

$$\begin{aligned} \min \quad & \rho(f(x, \omega)), \\ \text{sujeito a} \quad & \mathbb{E}[f(x, \omega)] \geq \mu, \quad x \in \mathcal{X}. \end{aligned} \quad (7-10)$$

Ainda no exemplo aqui considerado, a formulação na Equação (7-10), bem como na restrição estabelecida, buscar minimizar o risco das decisões em uma determinada carteira de investimentos torna-se possível, de forma a atingir um lucro mínimo definido por μ , sendo que, novamente, dependendo de μ o problema torna-se possivelmente inviável.

Outros tipos de SOPs podem exigir uma formulação escalar, ou que considere uma função biobjetivo. Para o primeiro caso, considerando um problema de maximização, que pode ser modelado conforme

$$\begin{aligned} \max \quad & (1 - \gamma) \cdot \mathbb{E}[f(x, \omega)] - \gamma \cdot \rho(f(x, \omega)), \\ \text{sujeito a} \quad & x \in \mathcal{X}, \end{aligned} \quad (7-11)$$

sendo γ assumindo o intervalo $0 \leq \gamma \leq 1$ o parâmetro escalar que define o grau de ajuste para o risco, e para o segundo caso, também em maximização, descrito em uma otimização vetorial

$$\begin{aligned} \max \quad & (\mathbb{E}[f(x, \omega)], -\rho(f(x, \omega))), \\ \text{sujeito a} \quad & x \in \mathcal{X}. \end{aligned} \quad (7-12)$$

Ao considerar os problemas com múltiplos objetivos, sendo conjuntamente também um SOP, têm-se a possibilidade de sua modelagem como um *Stochastic MOOP* (SMOOP), que pode assumir o formato

$$\begin{aligned} \max/\min \quad & (f_1(x, \omega), f_2(x, \omega), \dots, f_m(x, \omega)), \\ \text{sujeito a} \quad & x \in \mathcal{X}, \end{aligned} \quad (7-13)$$

sendo $f_j : X \rightarrow \mathbb{R} (j = 1, \dots, m)$ as $m \geq 2$ funções objetivo, em que se assume o conhecimento da distribuição de probabilidade conjunta das variáveis aleatórias.

O algoritmo genético (ou GA) é um dos mais populares algoritmos para a solução de problemas de otimização estocástica, normalmente sendo empregado na solução de problemas complexos e de larga escala, em vários campos. O algoritmo genético *multi-start* é uma versão aprimorada, que aumenta a capacidade de busca pela solução ótima global. Entretanto, tais algoritmos, bem como os algoritmos baseados em recozimento simulado, por exemplo, devido ao esforço empregado para evitar os mínimos locais, tendem a ser apropriados para problemas de decisão *offline*, dado que tendem a gerar convergências lentas [Marbach e Ijspeert 2005, Wang, Magnússon e Johansson 2021].

Desta forma, observa-se que a otimização estocástica surge como potencial solução para problemas do mundo real, em que as características aleatórias do problema não permitem a sua solução por métodos determinísticos. Trata-se, portanto, não de um tipo de abordagem que entra para o rol de ferramentas intercambiáveis, mas sim de uma classe de abordagem de problemas de decisão que devem ser, claramente, consideradas, quando assim o problema indicar. Entretanto, problemas SOPs, especialmente os SMOOPs, também apresentam limitações, incluindo aquelas já observadas em métodos exatos de otimização, que possuem instâncias em que custo computacional inviabiliza o seu emprego.

Segundo [Gutjahr e Pichler 2016], um SMOOP precisa, convenientemente, considerar uma distribuição de probabilidade finita, de forma que o valor esperado, bem como as medições de risco, reduzem o espaço de decisões a ser varrido, também tendendo a reduzir a complexidade do problema, ao passo que o transforma em um problema linear. Há, também, severos impactos quanto ao tamanho das N variáveis, ou objetivos, principalmente em problemas NP-difíceis, em que o tempo requerido cresce exponencialmente em razão de N . Entretanto, tal fato, bem como a simplificação de cenários normalmente encontrada em diversas modelagens [Gutjahr e Pichler 2016], tendem a transformar as soluções obtidas em inapropriadas (ou incompletas) para a tomada de decisões. O clássico arcabouço de otimização estocástica em dois estágios, quando considerado em SMOOPs, se mostra complexo, e de difícil modelagem, normalmente resultando em soluções que se reduzem à problemas de objetivo único. Adicionalmente, qualquer erro no conhecimento da distribuição poderia tornar o algoritmo subótimo e às vezes até inviável [Devanur et al. 2019].

7.1.4 Otimização *online*

De maneira geral, em métodos de otimização ditos tradicionais, assume-se o completo conhecimento de todos os dados que descrevem uma instância do problema. Assim, é possível iterar sobre eles, a fim de se computar as melhores soluções possíveis. Entretanto, há diversos problemas do mundo real em que isso não ocorre, sendo que as decisões precisam ser tomadas antes de que toda a informação necessária esteja disponível. Quando toda a informação necessária para o processo de tomada de decisão não está disponível, no exato momento em que ela é requerida, uma nova classe de problemas de otimização surge, podendo ser resolvidos pelos métodos de otimização *online*. Um algoritmo é dito *online* se ele toma decisões de otimização, isto é, determina soluções parciais [Krumke 2002], sempre que solicitado, não importando a quantidade de informação que esteja disponível, ao passo que os algoritmos *offline* obtêm acesso aos dados antes da tomada de decisão precisar ser desempenhada.

O problema do caixeiro viajante, ou *Traveling Salesman Problem* (TSP), é um importante problema nas áreas de pesquisa operacional e otimização matemática. Em um típico TSP, há um conjunto de cidades, devidamente instanciadas com suas métricas posicionais, em que, dada uma cidade de origem, a tarefa é encontrar um roteiro de duração total mínima, começando e terminando na origem, de forma que se visite cada cidade pelo menos uma vez. Nota-se, portanto, tratar-se de um típico problema *offline*, dado que toda informação necessária é conhecida, de forma totalmente determinística. Diversas restrições podem ser incluídas em um TSP, de forma que ele pode ser especializado para representar diferentes tipos de problemas de otimização [Jaillet e Wagner 2008]. Entretanto, quando tais especializações transformam a forma de obtenção dos dados necessários ao processo de otimização, assumir que todo esse conhecimento requerido estará disponível torna a maioria das soluções não realistas.

Embora existam diversas abordagens que visam lidar com problemas de otimização em cenários de incerteza ou com parcial observação dos dados, tais como os métodos de otimização estocástica, tais abordagens costumam requerer o conhecimento acurado das distribuições de probabilidade que descrevem tais dados. Em situações mais realistas, em muitos dos problemas do mundo real, pode não haver dados suficientes para realizar tais estimativas, como normalmente ocorre em problemas em novas indústrias ou mercados, ou até em cenários atípicos em problemas já conhecidos. Neste sentido, a otimização *online* surge como uma abordagem diferente para a otimização sob incerteza, se mostrando mais apropriada para problemas de tomada de decisão sequencial, onde as distribuições probabilísticas não estão disponíveis e/ou não são confiáveis no instante em que as decisões precisam ser tomadas.

É importante ressaltar que um método de otimização *online* não deve ser, necessariamente, observado como um método em que a tomada de decisões ocorre em tempo real. Para tal, um outro tipo de otimização deve ser considerado, a otimização em tempo real [Jaillet e Wagner 2008]. Essencialmente, a otimização em tempo real resolve problemas onde a instância do problema também é revelada de forma incremental, mas o foco está em empregar algoritmos práticos (rápidos), normalmente se apoiando em heurísticas ou meta-heurísticas, que resolvem rapidamente e podem entregar boas soluções, mesmo que sem garantias teóricas de otimalidade. Entretanto, é comum que nos problemas de tempo real, um algoritmo *online* seja empregado para entregar a próxima parte da solução dentro de um limite de tempo muito curto, ou seja, reagindo em tempo real [Krumke 2002]. Assim, o comportamento correto de um sistema de tempo real depende da qualidade da solução, bem como do tempo necessário para produzir a solução. Uma solução fornecida tarde demais pode ser inútil ou, em alguns casos, até mesmo perigosa porque não se ajusta aos parâmetros atuais do sistema que podem variar ao longo do tempo.

A forma mais amplamente aceita de medir o desempenho de algoritmos *online* é pela análise competitiva. A qualidade de uma determinada estratégia *online* é medida pela razão de pior caso entre o tempo necessário pelo algoritmo *online* para uma sequência de solicitações e o tempo ideal necessário para um algoritmo que conhece a sequência com antecedência. Essa proporção é chamada de *proporção competitiva* do algoritmo *online*. Assim, um algoritmo é dito ρ -competitivo (para $\rho \neq 0$) se para cada instância do problema, seu tempo de conclusão é, no máximo (para o caso de minimização), ou no mínimo (para o caso de maximização), ρ vezes o tempo de conclusão ideal, ou (quase) ótimo, para a mesma instância [Ausiello et al. 2001, Lu et al. 2013], tal que

$$Custo_{online}(i) \leq \rho \cdot Custo_{ideal}(i), \quad \forall i \in I, \quad (7-14)$$

$$Custo_{online}(i) \geq \rho \cdot Custo_{ideal}(i), \quad \forall i \in I, \quad (7-15)$$

representam a relação de minimização e maximização, respectivamente, sendo ρ estabelecido por

$$\rho = \max_{i \in I} / \min_{i \in I} \frac{Custo_{online}(i)}{Custo_{ideal}(i)}. \quad (7-16)$$

Os autores em [Ausiello et al. 2001] apresentam uma proposta de otimização para um TSP *online*, modelado como um problema de roteamento e escalonamento. Nesta proposta, enquanto o vendedor está viajando, novas cidades a visitar podem ser comunicadas a ele, de forma que o seu objetivo é visitar todas as cidades, minimizando o tempo total gasto. Tal formulação é bem aderente aos diversos problemas encontrados em redes de computadores, por exemplo. Na referido trabalho é proposto o algoritmo *Plan-At-Home* (PAH), que atinge uma razão competitiva de 2 para qualquer espaço métrico pertencente ao conjunto \mathcal{M} , que estabelece os possíveis subcaminhos, sendo o a origem, e $d(\cdot, \cdot)$ a métrica de distância entre os dois pontos. PAH segue, então, os seguintes passos:

1. Sempre que o caixeiro está na origem, ele passa a seguir um roteiro ótimo que atende a todas as solicitações a serem atendidas e retorna à origem.
2. Se em algum momento uma nova solicitação for apresentada no ponto l , o caixeiro realizará uma de duas ações, dependendo de sua posição atual p :
 - (a) Se $d(l, o) > d(p, o)$, então o caixeiro volta imediatamente para a origem, onde retorna à situação do Caso 1.
 - (b) Se $d(l, o) \leq d(p, o)$, então o caixeiro ignora a solicitação, seguindo o seu caminho até retornar à origem, onde novamente entra no Caso 1.

Assim, observa-se que o algoritmo proposto não pode ser computado em tempo

polinomial, dado que seus subcaminhos ainda precisam ser resolvidos como um TSP tradicional.

A análise competitiva de pior caso é um dos principais arcabouços adotados para a solução de problemas de otimização *online* para os desafios da alocação de recursos, se mostrando bastante robusto [Devanur et al. 2019]. Entretanto, isso leva a limites igualmente pessimistas, que tendem a impossibilitar a obtenção de soluções localizadas o mais aproximado possível dos ótimos globais.

Uma outra abordagem é baseada no conceito de computar o arrependimento do agente para estabelecer a análise da qualidade da solução. Como um problema de otimização *online* é baseado na cadeia de ações sequenciais, o objetivo segue no sentido de minimizar a perda agregada de um conjunto sequencial de funções de perda desconhecidas. Um protótipo genérico é demonstrado no Algoritmo (3) (adaptado de [Mertikopoulos 2019]), em que seus passos são descritos como:

1. para cada estágio $t = 1, 2, \dots, T$ o agente otimizador seleciona uma ação X_t do subconjunto convexo \mathcal{X} ;
2. o otimizador incorre na função de perda $\ell_t(X_t)$, desconhecida *a priori*;
3. baseado na perda, e/ou noutras informações recebidas, o otimizador atualiza sua ação;

Algoritmo 3: Protótipo de um problema de otimização *online* genérico.

```

1 requer conjunto de ações  $\mathcal{X}$ , sequência  $\ell_t : \mathcal{X} \rightarrow \mathbb{R}$ 
2 selecione  $X_t \in \mathcal{X}$ ;
3 para  $t = 1, 2, \dots, T$  faça
4   computar  $\ell_t(X_t)$ ;
5   atualizar  $X_t \leftarrow X_{t+1}$ ;
```

Assim, considerando um problema de otimização estático qualquer, que busca a minimização da função objetivo, tal que

$$\begin{aligned} & \text{minimize} && f(x), \\ & \text{sujeito a} && x \in \mathcal{X}, \end{aligned} \tag{7-17}$$

sendo $f: \mathcal{X} \rightarrow \mathbb{R}$ uma função objetivo estática. Vendo o presente problema como um problema de decisão sequencial, toda função de perda encontrada pelo otimizador será, portanto, igual à própria função objetivo [Mertikopoulos 2019], isto é, $\ell_t(x) = f(x), \forall t = 1, 2, \dots, T$. O *gap* de otimalidade em uma sequência de ações $X_t \in \mathcal{X}$, após T passos,

pode ser obtida por

$$Gap(T) = \sum_{t=1}^T f(X_t) - T \min_{x \in \mathcal{X}} f(x) = \sum_{t=1}^T \ell(X_t) - \min_{x \in \mathcal{X}} \sum_{t=1}^T f(X_t) = \max_{x \in \mathcal{X}} \sum_{t=1}^T [\ell_t(X_t) - \ell_t(x)], \quad (7-18)$$

também visto como o arrependimento do agente, sendo este a diferença entre a perda total obtida pelo agente após os T passos, e a perda observada na melhor ação observada.

Ao lidar com problemas de otimização estocástica, modelados, também, de forma *online*, em que um problema genérico é modelado conforme a Equação (7-9), também assumindo um problema de decisão sequencial, assumindo que um espaço randômico independente e identicamente distribuído $\omega_t \in \Omega$, em $F : \mathcal{X} \times \Omega \rightarrow \mathbb{R}$ como a função objetivo estocástica, e com uma função de perda definida como $\ell_t(x) = F(x; \omega_t), \forall t = 1, 2, \dots, T$, o *gap* médio de otimalidade é obtido por:

$$\overline{GAP}(T) = \mathbb{E} \left[\max_{x \in \mathcal{X}} \sum_{t=1}^T [\ell_t(X_t) - \ell_t(x)] \right] \quad (7-19)$$

Assim como ocorre nas demais áreas tratadas pelos métodos de otimização tradicionais, os problemas de larga escala também se mostram desafiadores quando considerados em problemas *online*, sendo ainda mais difíceis em situações em que os dados encontram-se dispersos, mas a arquitetura em que o problema está inserido exige a tomada de decisão de forma centralizada [Zhou et al. 2021]. Em problemas de gerenciamento de infraestrutura computacional, por exemplo, a abordagem de otimização *online* geralmente incorre em procedimentos de busca exaustiva, impondo problemas de escalabilidade com relação ao crescimento dos parâmetros e do horizonte de tempo considerado para a otimização [Zhou et al. 2021]. Neste sentido, esforços recentes têm apontado na direção do tratamento dos desafios envolvidos nos modelos de otimização *online* com suporte de ferramentas de aprendizado de máquina [Ren, Sun e Peng 2020, Takahashi e Tomah 2020, Lin, Yu e Chen 2021, Zhou et al. 2021, Zhou et al. 2021], principalmente em cenários em que há a possibilidade de se adotar agentes capazes de aprender as políticas de otimização diretamente da observação dos dados brutos.