Brasil 6G

Projeto Brasil 6G

Estado da Arte em Inteligência Artificial Aplicada a Redes 6G

Inatel
Instituto Nacional de Telecomunicações

RNP

cpqd

Universidade Federal do Ceará

INF
Instituto de Informática

UFG
Universidade Federal de Goiás

Universidade Federal do Rio de Janeiro

Universidade Federal do Pará

UFSC

UFU
Universidade Federal de Uberlândia

UNICAMP

03/02/2022

Histórico de Atualizações:

Versão	Data	Autor(es)	Notas
1	24/01/2022	Aldebaro Klautau Anderson Reis Rufino Marins Arismar Cerqueira Sodré Junior Cleverson Veloso Nahum Ciro José Almeida Macedo Cristiano Bonato Both Davi da Silva Brilhante Felipe Augusto P. de Figueiredo Joanna Carolina Manjarres José Ferreira de Rezende Kleber Vieira Cardoso Luan Gonçalves Lucas dos Santos Costa Luciano Leonel Mendes Luiz Augusto Melo Pereira Mariana Mello Rausley Adriano Amaral de Souza Roberto Michio Marques Kagami Rodrigo Moreira Sand Luz Correa Sheila Cássia da Silva Janota Victor Hugo L. Lopes	Elaboração de conteúdo
2	03/02/2022	Aldebaro Klautau Davi da Silva Brilhante Kleber Vieira Cardoso Luciano Leonel Mendes Victor Hugo L. Lopes	Revisão de texto

Lista de Tabelas

Lista de Figuras

Acrônimos

2G Segunda Geração de Rede Móvel Celular
3G Terceira Geração de Rede Móvel Celular
3GPP *3rd Generation Partnership Project*
4G Quarta Geração de Rede Móvel Celular
5G Quinta Geração de Rede Móvel Celular
6G Sexta Geração de Rede Móvel Celular
ADC *Analog-to-digital Converter*
AF *Application Function*
AM Aprendizado de Máquina
AMF *Access and Mobility Management Function*
AMP *Approximate Message Passing*
AN *Artificial Noise*
ANATEL Agência Nacional de Telecomunicações
ANN *Artificial Neural Network*
AoA *Angle of Arrival*
AoD *Angle of Departure*
API *Application Programming Interface*
ARIMA *AutoRegressive Integrated Moving Average*
AROW *Adaptive Regularization of Weight Vectors*
ASHT *Asymptotic Simple Hypothesis Test*
AUC *Area Under the Curve*
AWGN *Additive White Gaussian Noise*
B5G *Beyond* 5G
BBU *Baseband Unit*
BCJR *Bahl–Cocke–Jelinek–Raviv*
BER *Bit Error Rate*
BIC *Bayesian information criterion*
BP-SHMM *Beta Process Sticky Hidden Markov Model*
BiRNN *Bi-Directional Recurrent Neural Network*
BS *Base Station*
BWP Bandwidth Part
C2PO *Convex 1-bit Precoder*
CAPEX *Capital Expenditure*
CDMA *Code Division Multiple Access*

CFO *Carrier Frequency Offset*

CL *Compressed Learning*

CLDNN *Convolutional Long Short-Term Deep Neural Networks*

CU *Central Unit*

CN *Core Network*

CNN *Convolutional Neural Network*

CP *Cyclic Prefix*

CPU *Central Processing Unit*

CQI *Channel Quality Indicator*

CR *Cognitive Radio*

C-RAN *Cloud-RAN*

CS *Compressed Sensing*

CSI *Channel State Information*

CSMF *Communication Service Management Function*

D2D *Device-to-Device*

DAG-SVM *Directed Acyclic Graph - Support Vector Machine*

DARPA *Defense Advanced Research Projects Agency*

DBN *Deep Belief Network*

DBSCAN *Density-based Spatial Clustering of Applications with Noise*

DCS *Deep Cooperative Sensing*

DDPG *Deep Deterministic Policy Gradient*

DFT *Discrete Fourier Transform*

DL *Deep Learning*

DMV Detector por Máxima Verossimilhança

DNN *Deep Neural Network*

DoA *Direction of Arrival*

DPMM *Dirichlet Process Mixture Model*

DQN *Deep Q Network*

DRL *Deep Reinforcement Learning*

DSRC *Dedicated Short-Range Communications*

DSDV *Destination-Sequenced Distance-Vector Routing*

DU *Distributed Unit*

E2EMD *End-to-End Service Management Domain*

ED *Energy Detection*

EI *Edge Intelligence*

ELPC *Extremely Low-Power Communications*

EM *Expectation Maximization*

eMBB *enhanced Mobile Broadband*

ENI *Experiential Networked Intelligence*

EPC *Evolved Packet Core*

ERB Estação Rádio Base

ERLLC *Extremely Reliable and Low Latency Communication*

ETSI *European Telecommunications Standards Institute*

EVD *Eigenvalue Detection*

EVM *Error Vector Magnitude*

FC *Fusion Center*

FCNN *Fully-Connected Neural Network*

feMBB *further-eMBB*

FL *Federated Learning*

FLD *Fisher Linear Discriminant*

FNN *Feedforward Neural Network*

FTN *Faster-than-Nyquist*

GAN *Generative Adversarial Network*

GMM *Gaussian Mixture Model*

GPS *Global Positioning System*

GRU *Gated Recurrent Unit*

HD *Hard Decision*

IA Inteligência Artificial

ICI *Inter-carrier interference*

IDMA *Interleave Division Multiple Access*

IEEE *Institute of Eletrical Electronic Engineers*

IIoT *Industrial Internet of Things*

IMT-2020 *International Mobile Telecommunications* 2020

INI *Inter-Numerology Interference*

IoT *Internet of Things*

IQ *In-phase and Quadrature*

IRS *Intelligent Reflecting Surface*

ISI *Inter-symbol Interference*

ISTN *Integrated Space and Terrestrial Network*

ITU *International Telecommunication Union*

KMC *K-Means Clustering*

KNN *K Nearest Neighbours*

KPI *Key Performance Indicator*

LBT *Listen-Before-Transmit*

LDHMC *Long-Distance and High-Mobility Communications*

LGPD Lei Geral de Proteção de Dados

LRT *Likelihood Ratio Test*

LS *Least Squares*

LSTM *Long Short-Term Memory*

LTE *Long Term Evolution*

MAB *Multi-Armed Bandit*

MAC *Medium Access Control*

MANO *MAnagement and Network Orchestration*

MBBLL *Mobile Broad Bandwidth and Low Latency*

MBRLLC *Mobile Broadband Reliable Low Latency Communication*

MD *Management Domain*

MD-IMA *Intelligent Multiple Access*

MDP *Markov Decision Process*

MEC *Multi-access Edge Computing*

MED *Maximum Eigenvalue Detection*

MIMO *Multiple-Input Multiple-Output*

MISO *Multiple-Input Single-Output*

ML *Machine Learning*

MLP *Multi-layer Perceptron*

MMSE-FDE *Minimum Mean Squared Error-Frequency Domain Equalization*

MMSE *Minimum Mean Squared Error*

mMTC *Massive Machine-Type Communications*

mmWave *millimeter wave*

MPA *Message Passing Algorithm*

MS *Mean-Shift*

MSE *Mean Squared Error*

MTC *Machine-Type Communications*

MUD *Multi-User Detection*

mULC *Massive Ultra-Reliable Low-Latency Communication*

muRLLC *massive-uRLLC*

MUSA *Multi-User Shared Access*

NB *Naive Bayes*

NB-IoT *Narrow-band IoT*

NF *Network Function*

NFV *Network Functions Virtualization*

NFVI *NFV Infrastructure*

NFVO *NFV Orchestrator*

NGMN *Next Generation Mobile Networks*

NG-RAN *Next Generation Radio Access Network*

NOMA *Non-Orthogonal Multiple Access*

NR *New Radio*

NSMF *Network Slice Management Function*

NSSMF *Network Slice Subnet Management Function*

NWDAF *Network Data Analytic Function*

O-RAN *Open Radio Access Network*

OAM *Operations, Administration and Maintenance*

OFDM *Orthogonal Frequency-Division Multiplexing*

OFDMA *Orthogonal Frequency Division Multiple Access*

OMA *Orthogonal Multiple Access*

ONU Organização das Nações Unidas

OPEX *Operational Expenditure*

OSI *Open Systems Interconnection*

OSPF *Open Shortest Path First*

OTN *Optical Transport Network*

PAPR *Peak-to-Average Power Ratio*

PCA *Principal Component Analysis*

PCF *Policy ControlFunction*

PD-NOMA *Power-Domain Non-Orthogonal Multiple Access*

PDMA *Pattern Division Multiple Acces*

PHY *Physical Layer*

PRBs *Physical Resource Blocks*

PU *Primary User*

QoE *Quality of Experience*

QoS *Quality of Service*

RAN *Radio Access Network*

RT *Ray-Tracing*

RB *Resource Block*

RBF-SVM *Radial Basis Function Neural Network Support Vector Machine*

RF *Radiofrequency*

RFC *Random Forest Classifier*

RFML *Radio Frequency Machine Learning*

RFMLS *Radio Frequency Machine Learning Systems*

RGB-D *Red, Green, Blue and Depth*

RIC *RAN Inteligence Controller*

RIP *Routing Information Protocol*

RL *Reinforcement Learning*

RNN *Recurrent Neural Network*

ROC *Receiver Operating Characteristic*

RoF *Radio over Fiber*

RRH *Remote Radio Head*

RRM *Radio Resource Management*

RTT *Round Trip Time*

RU *Radio Unit*

SBA *Service-Based Architecture*

SCMA *Sparse Code Multiple Access*

SD *Soft Decision*

SDN *Software-Defined Networking*

SEI *Specific Emitter Identification*

SFC *Service Function Chain*

SIC *Successive Interference Cancellation*

SINR *Signal to Interference plus Noise Ratio*

SIR *Signal to Interference Ratio*

SLA *Service Level Agreement*

SMC Computação Multipartidária Segura

SMF *Session Management Function*

SNR *Signal-to-Noise Ratio*

SRA *Scheduling and Resource Allocation*

SSL *Secure Socket Layer*

SU *Secondary User*

SVM *Support Vector Machine*

SWIPT *Simultaneous Wireless Information and Power Transfer*

TAS *Transmit Antenna Selection*

TCP *Transmission Control Protocol*

TDD *Time Division Duplex*

TDMA *Time Division Multiple Access*

THz *Terahertz*

t-SNE *t-Distributed Stochastic Neighbor Embedding*

TTI *Transmission Time Interval*

TVWS *TV White Space*

UAVs *Unmanned Aerial Vehicles*

UDR *Unified Data Repository*

UE *User Equipment*

uHDD *Ultra-High Data Density*

uHSLLC *Ultra-High-Speed with Low-Latency Communication*

ULBC *Ultra-reliable Low-latency Broadband Communication*

uMBB *Ubiquitous Mobile Broadband*

umMTC *ultra-mMTC*

uMUB *Ubiquitous Mobile Ultra-Broadband*

UPF *User Plane Function*

URLLC *Ultra Reliable Low Latency Communications*

V2X *Vehicle-to-Everything*

VANT *Veículo Aéreo Não Tripulado*

VLC *Visible Light Communications*

VNE *Virtual Network Embeding*

VNF *Virtualized Network Function*

VPN *Virtual Private Network*

XMD *Cross-modulation Distortion*

ZF *Zero Forcing*

ZSM *Zero-touch Network and Service Management*

WDM *Wavelength Division Multiplex*

Sumário

1 Introdução

Este relatório apresenta o estado-da-arte na adoção de soluções pautadas em IA para a 6G. A evolução das tecnologias de redes de comunicações móveis é capitaneada pelo 3GPP, em interação com a ITU. Como indicado na Fig. 1, uma geração corresponde a um intervalo de vários anos, durante o qual uma série de *releases* define um processo evolutivo contínuo.

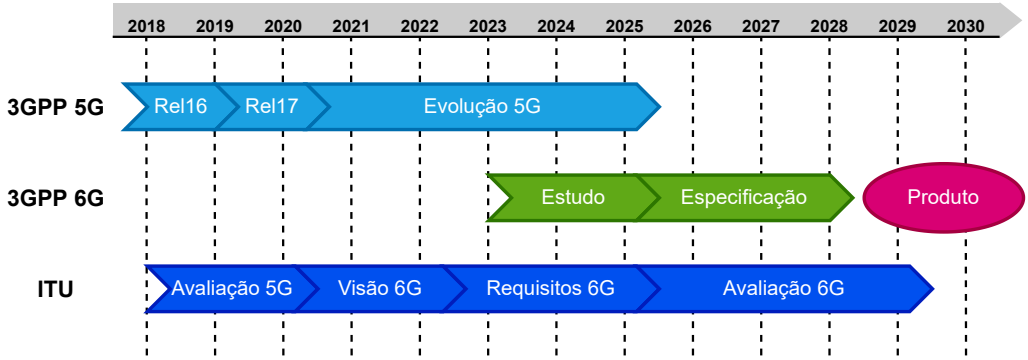


Figura 1: Lançamentos e ações do 3GPP e ITU em direção ao 6G.

Assim, IA será discutida no contexto de tecnologias inerentes às redes 6G como transmissão de dados usando Terahertz, superfícies inteligentes e disponibilidade de localização com grande acurácia, mas também com tecnologias associadas às predecessoras 4G e 5G. De fato, o escopo e as tecnologias de IA em 6G serão consequências diretas das atuais pesquisas e implantações de redes 5G.

Além do legado das gerações anteriores, este relatório também leva em conta a integração entre IA e tecnologias como *big data*. A Fig. 2 ilustra um ambiente de IA onde a tecnologia de comunicações não é especificada. Apesar de algumas das tecnologias abordadas neste relatório terem uso em redes distintas da 6G, o foco principal é o uso de IA para otimizar as redes 6G ou tecnologias 6G para habilitar aplicações de IA.

O uso de IA em 6G é pautado por tecnologias emergentes como hologramas e robótica. Novos requisitos e características de funcionamento da rede são exigidos para que as novas aplicações possam funcionar corretamente e alguns desses requisitos são discutidos nos próximos parágrafos.

O desenvolvimento da tecnologia de hologramas vêm sendo implementado em projetos como Microsoft HoloLens [15] e o uso amplo dessa tecnologia deve se tornar realidade ainda nessa década. A renderização de hologramas de alta definição através de redes móveis trará uma experiência verdadeiramente imersiva, tornando realidade a telepresença baseada em hologramas em reuniões e permitindo a interação educacional com objetos ultra-realísticos. Todavia, o uso de aplicações com hologramas requer uma taxa de bits na ordem de terabits por segundo, mesmo utilizando-se compressão de imagem e vídeo [5]. E para verdadeira imersão nas aplicações é necessário uma latência extremamente baixa.

As aplicações de realidade estendida que combinam realidade aumentada, virtual e mistas começaram a ser exploradas no 5G, mas com características similares aos serviços de vídeo de redes 5G. Para oferecer o suporte a dispositivos de realidade estendida com campo de visão de 360 graus serão necessárias taxas de transmissão muito maiores que as necessárias para a transmissão de vídeos 2D. Para uma experiência de imersão ideal, vídeos com maior qualidade, taxa de quadros e profundidade de cores são necessárias, levando a uma demanda de taxas de

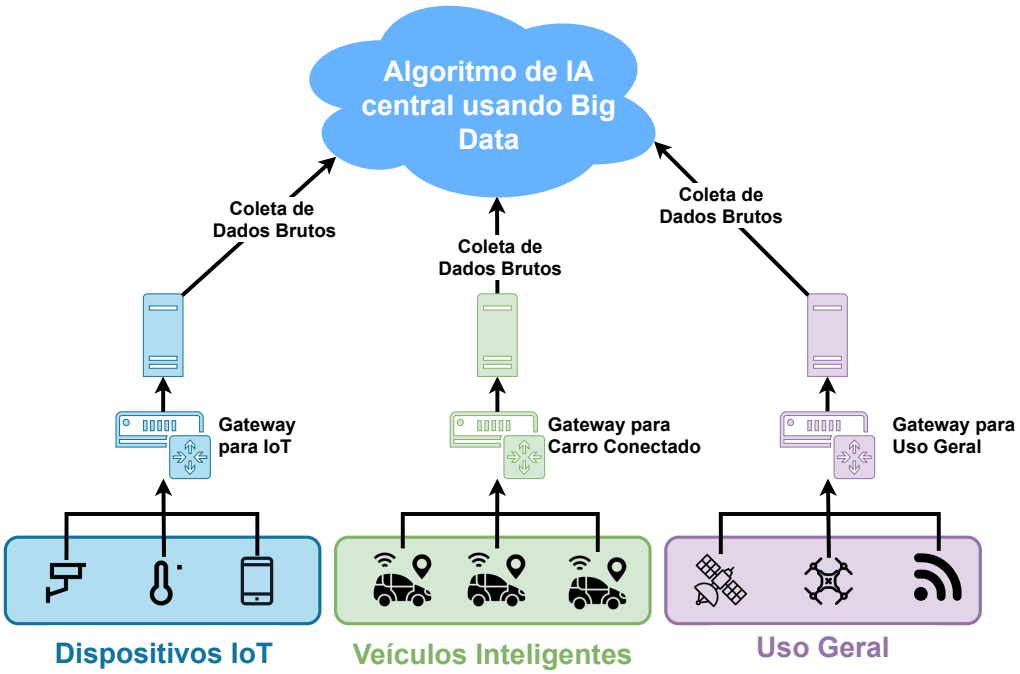


Figura 2: Sistemas inteligentes conectados.

1.6 Gbps por dispositivo [16]. Esses requisitos levariam a rede 5G rapidamente a saturação, limitando o uso dessas aplicações.

A Internet tátil deve prover latências extremamente baixas, atingindo um valor máximo de 1 ms, pautado pelo sentido humano [17]. Em combinação com a alta confiabilidade, disponibilidade, segurança e algumas vezes com altas taxas de transmissão, uma vasta gama de aplicações disruptivas funcionando em tempo real podem ser implementadas. A internet tátil tem especial interesse na área da indústria 4.0, onde por exemplo um operador poderia controlar uma máquina em um ambiente que requer alta precisão ou então poderia ser utilizada por médicos em cirurgias a distância através da utilização de robôs.

A possibilidade de aplicações envolvendo experiências multisensoriais são uma possibilidade do 6G, onde seriam explorados não só a experiência visual e auditiva dos usuários, mas também o envolvimento de outros sentidos como paladar, olfato e tato. Alguns exemplos poderiam ser aplicações que permitissem verificar a textura de um material ou mesmo o cheiro de um perfume [18]. Além desses exemplos, aplicações multisensoriais podem ser exploradas em cirurgias remotas e jogos imersivos.

Outra área de aplicações a ter um grande crescimento nas redes 6G, é a área das aplicações de computação pervasiva, onde aplicações passam a confiar principalmente na tomada de decisões baseadas em IA como em aplicações de visão computacional, reconhecimento de face e voz, processamento de linguagem natural, e mapeamento e localização simultâneos. Para oferecer suporte a essas aplicações as redes 6G deverão oferecer computação pervasiva baseada em IA como um serviço [19] utilizando computação distribuída ao longo de recursos na nuvem, borda e dispositivos-fim.

Transportes inteligentes devem ser amplamente explorados dentro do contexto de 6G, dado o crescimento na utilização de transportes autônomos e drones para prover segurança, eficiência e contribuir com a diminuição da poluição. Dispositivos autônomos conectados a rede 6G

necessitarão de alta-confiabilidade da rede e uma baixa latência para que as decisões de direção possam ser tomadas em tempo real sem gerar problemas para os passageiros. A maior difusão de drones e enxames de drones para oferta de serviços como entregas e vigilância também devem alavancar a utilização de aplicações baseadas em veículos autônomos.

Uma conectividade ubíqua global também deve ser explorada dentro do contexto do 6G, já que as gerações que o antecederam possuíam um foco maior em áreas metropolitanas, especialmente cenários internos. De qualquer forma, uma relevante parte da população vive em áreas remotas e rurais que não possuem acesso básico a serviços de informação e comunicação. Então um dos desafios que as aplicações do 6G devem suportar é oferecer cobertura às áreas remotas do planeta, oferecendo uma qualidade de serviço aceitável e com custo que possa ser implementado na realidade. Devido aos custos de implementação, é extremamente problemático realizar a cobertura dessas áreas utilizando redes terrestres, e por isso cresce o interesse em satélites geoestacionários com órbitas mais próximas a Terra para habilitar comunicação de baixo custo e com alta taxa de transmissão [20].

Tendo em vista os exemplos de áreas de aplicações fornecidos, fica claro que as redes 5G não serão capazes de fornecer o suporte necessário para essa nova gama de aplicações. Com isso, espera-se que as redes 6G possam ir além da comunicação personalizada em direção à plena realização do paradigma da Internet das Coisas (*Internet of Things*, IoT), conectando não apenas pessoas, mas também recursos de computação, veículos, dispositivos, sensores e até agentes robóticos.

Além das distintas aplicações de 6G e sua influência no uso de IA, leva-se em conta neste relatório as distintas escalas de tempo envolvidas. Como indica a Fig. 3, dependendo do elemento da rede 6G onde IA é aplicada, tem-se algoritmos operando em escalas de tempo variando de milisegundos a meses.

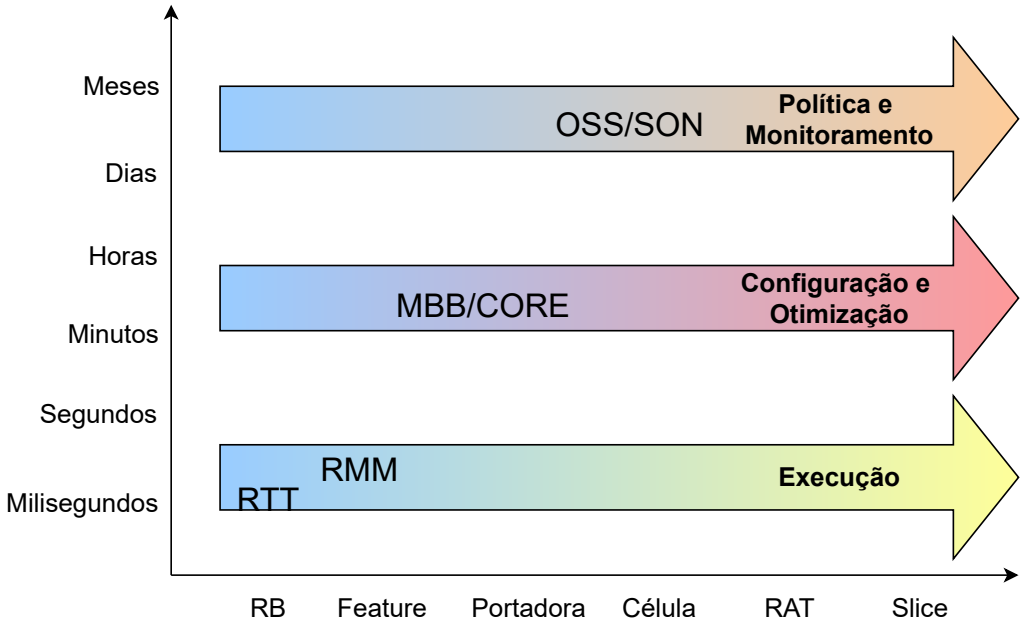


Figura 3: Elementos da rede 6G onde IA será aplicada e correspondente escala de tempo.

Perante a variedade de usos de IA em 6G, as seções deste relatório buscam identificar prioritariamente:

- Limitações e desafios de 5G e gerações anteriores;

- Aplicações e problemas onde IA se mostra útil ou tem potencial;
- Novas aplicações e casos de uso de IA em 6G;
- Tecnologias que complementam IA em 6G.

O documento encontra-se organizado da seguinte forma. Na Seção 2 são apresentadas as integrações e casos de uso de IA focados na redes de acesso, apresentando o uso de IA para funções como gerenciamento de feixes em *Multiple-Input Multiple-Output* (MIMO), alocação de recursos e estimação de canais. Na Seção 3 são apontadas as principais pesquisas na área de redes de transporte para o 6G, abordando a utilização de técnicas de IA para aumentar a eficiência, automatização, gerenciamento e redução de custos. A Seção 4 discute a utilização de IA no núcleo das redes 6G, abordando tópicos como a elasticidade, balanceamento de carga e orquestração das funções do núcleo da rede. Na Seção 5 são apresentadas as especificações e padronizações para o desenvolvimento de implantação da IA em sistemas 5G e 6G, abordando tanto os padrões para coleta e disseminação de dados, como a geração de conhecimento e arquiteturas voltadas à integração de IA à gerência de redes. A Seção 6 apresenta a estrutura da rede 6G a ser utilizada no suporte à aplicações de IA, abordando métodos como o particionamento de redes neurais, aprendizado federado e inteligência na borda. Por fim, a Seção 7 conclui o relatório sumarizando as tecnologias apresentadas e as expectativas com relação ao papel da IA dentro do contexto de redes 6G.

2 IA em Redes de Acesso

Neste capítulo discute-se a utilização de técnicas de IA na rede de acesso de redes 6G. A rede de acesso têm o objetivo de realizar a conexão de dispositivos a rede móvel através de sinais de rádio, apresentando uma grande abrangência de funcionalidades e cenários possíveis. Ao longo deste capítulo serão enfatizadas algumas das aplicações de IA voltadas para o aperfeiçoamento das funções de rede de acesso como o gerenciamentos de feixes em MIMO, adaptação e alocação de recursos multidimensionais nas camadas PHY/MAC, Sensoriamento espectral, esquemas de *Non-Orthogonal Multiple Access* (NOMA), gerenciamento de mobilidade em redes de ondas milimétricas, estimação e equalização de canais, segurança da camada física e o uso de gêmeos digitais e extração de características eletromagnéticas.

2.1 Gerenciamento de feixes em MIMO

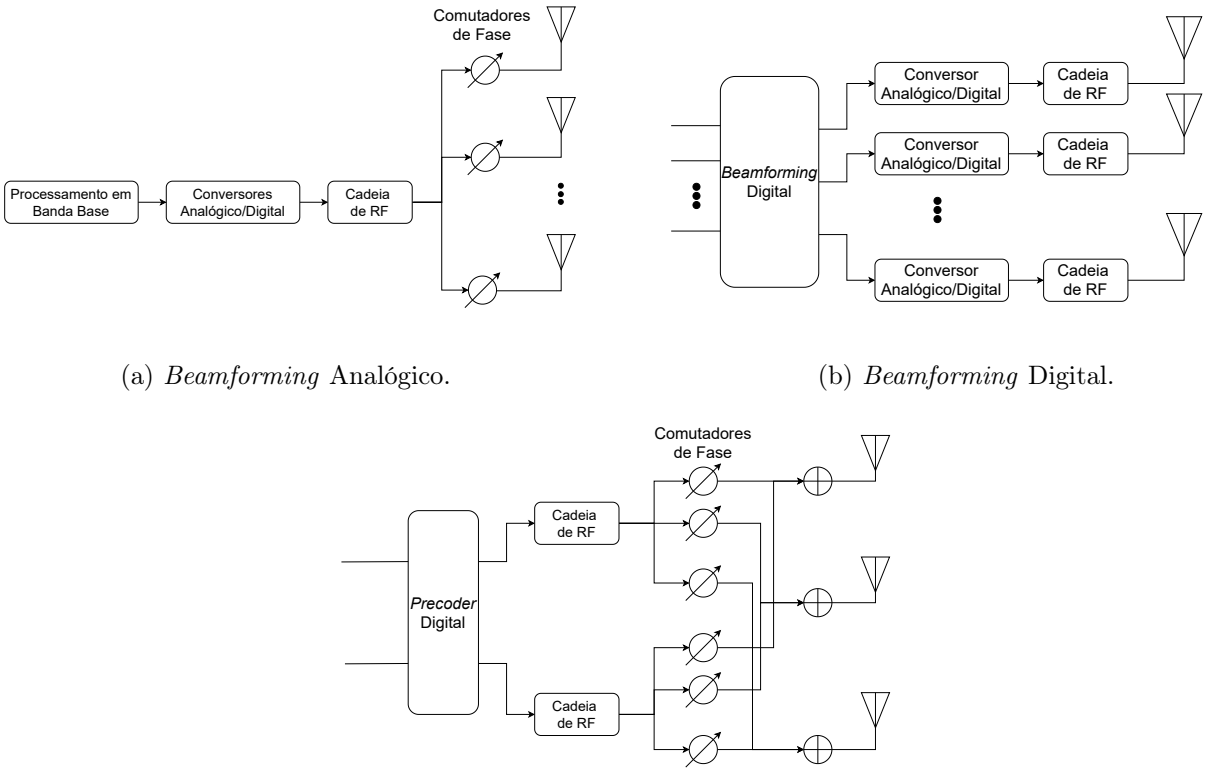
Davi da Silva Brilhante, *Joanna Carolina Manjarres*, *José Ferreira de Rezende*
dbrilhante@land.ufrj.br, joanna@land.ufrj.br, rezende@land.ufrj.br

A IA mostra-se útil quando a configuração de um enlace de comunicação torna-se complexa, tal como quando o número de antenas aumenta consideravelmente. O uso de múltiplas antenas compondo sistemas MIMO em redes sem fio vem se tornando cada vez mais comum à medida que o número de usuários e a largura de banda de frequência aumentam significativamente a cada ano [21]. Quando empregadas, técnicas de MIMO podem proporcionar reuso espacial, aumentar o ganho do sinal recebido e diminuir a interferência co-canal. Tais fatores aumentam a eficiência espectral agregada da rede [22].

Um desafio em arranjos de antenas MIMO é o *beamforming* direcional. O *beamforming* é realizado por meio da interação dos sinais irradiados por cada antena do arranjo de antenas para, através de interferências construtivas e destrutivas, modificar o padrão de irradiação para um determinado propósito. Ao alterar o ganho e a fase dos sinais transmitidos em cada elemento do arranjo de antenas é possível alterar a direção e o formato do padrão de irradiação. Por exemplo, um transmissor pode incrementar por um fator constante a fase do sinal transmitido em cada elemento do seu arranjo de antenas e assim direcionar o feixe principal da antena na direção de um único dispositivo receptor, aumentando a diretividade e reduzindo o efeito multi-percurso [23].

O *beamforming* pode assumir três tipos de arquitetura: analógica, digital e híbrida, conforme mostrado na Figura 4. No *beamforming* analógico, ajustes de fase no sinal são feitos na cadeia de *Radiofrequency* (RF) para compensar a forma como os raios irão irradiar (incidir) a partir (sobre) as antenas transmissoras (receptoras), mas só há um sentido para o fluxo de dados. Na arquitetura digital, o ajuste de fase é feito ainda com o sinal em banda base, o que confere maior precisão ao processo e também a possibilidade de múltiplos fluxos de dados simultâneos. O processo em banda base, por outro lado, exige processamento digital de sinais intenso e um conversos digital/analógico para cada cadeia de RF, o que torna a arquitetura digital dispendiosa em termos de custo e energia. O *beamforming* híbrido, combina o *beamforming* digital e o *beamforming* analógico. A motivação para o *beamforming* híbrido é a possibilidade de reduzir o custo, a complexidade e o consumo de energia, mas ainda ter mais de um fluxo de dados simultâneo [24].

Contudo, encontrar a direção ótima para realizar uma transmissão em um sistema MIMO é um problema complexo. Para ter o máximo de desempenho de um sistema MIMO é necessário obter amostras de canal para cada par de antenas entre receptor e transmissor com o intuito



(a) *Beamforming* Analógico.

(b) *Beamforming* Digital.

(c) *Beamforming* Híbrido.

Figura 4: Arquiteturas de *Beamforming*.

de elevar o ganho do sistema e contornar os efeitos adversos do canal. O processo de estimação de canal torna-se mais custoso, e pode vir a ser inviável, a medida que o número de antenas aumenta. Somado a isto, direcionar um sistema MIMO também depende das limitações de *hardware* do equipamento transceptor e do cenário e aplicação a que estes dispositivos se destinam [25]. Por isso, é comum usar os mecanismos de *codebook* que pré-definem quais padrões de irradiação podem ser utilizados por um arranjo de antenas [26]. Os *codebooks* são matrizes e cada coluna dessa matriz, chamadas também de *codewords*, apresenta um padrão de irradiação diferente.

Ainda que o espaço de possibilidades seja reduzido ao adotar um *codebook*, ainda é considerado custoso o processo de seleção de *codewords* ou seleção de feixes (do inglês, *beam selection*), como é comumente adotado na literatura. Tomemos como exemplo um método ingênuo de seleção de feixes, também chamado de exaustivo. O método exaustivo procura em cada feixe, um a um, a combinação entre transmissor e receptor que resultará no máximo valor de uma métrica utilizada como critério, como por exemplo o ganho do canal transmissor/receptor. Considerando que transmissor e receptor tenham o mesmo número N de antenas, a complexidade da seleção de feixes do método exaustivo é da ordem de N^2. Apesar de o método exaustivo garantir sempre o resultado ótimo, ele torna-se impraticável devido ao tempo exponencialmente crescente a medida que o número de feixes ou padrões de irradiação aumentam [27] e os requisitos de latência ultra-baixa previstos para o 6G, em torno de $1 - 10\mu s$ [28].

Os problemas em MIMO relatados acima tornam-se ainda mais notáveis nas comunicações nas banda de *millimeter wave* (mmWave) e tera hertz. Essas duas bandas estão localizadas no espectro de frequências nas faixas de 30 a 300 GHz e de 0.1 a 100 THz, respectivamente, e são

consideradas tecnologias promissoras devido à quantidade expressiva de espectro de frequência pouco utilizado nessas bandas [29]. No entanto, o benefício de ocupar uma parte ainda pouco explorada do espectro vem com a alta atenuação em espaço livre como contrapartida. Essa alta atenuação é contornada utilizando antenas MIMO altamente direcionais, cujo ganho compensa a perda de percurso, mas demanda métodos de seleção de feixe precisos e eficientes, para garantir os requisitos de taxa de dados e atraso demandados pelas aplicações [30]. Outro desafio nas bandas de mmWave e tera hertz é a baixa capacidade de difração e o bloqueio severo na maioria dos materiais. Medidas em [31] mostraram que a atenuação em vidro tingido pode chegar a 40,1 dB e em tijolos 28,3 dB. Ademais, o bloqueio causados por corpos humanos pode causar entre 30 e 40 dB de atenuação [32] e reduzir em até 32% a taxa de dados em redes móveis em ambientes externos [33].

Hoje em dia, os algoritmos de *Machine Learning* (ML) permitem que as redes sem fio aprendam a extrair informações ao interagir com grandes quantidades de dados, tornando-se uma potencial ferramenta em casos em que não há solução por meio de abordagem tradicional analítica ou que a solução exige a configuração manual de muitos parâmetros, permitindo a alguma das técnicas de ML contribuir na estimação destes parâmetros. Assim estes algoritmos são vistos pela academia e indústria como essenciais para as redes de comunicação, podendo ser aplicados em detecção de anomalias e falhas na rede, assim como na predição de cenários futuros. Além disso, esses algoritmos permitem à rede: se adaptar a ambientes que variam frequentemente, obter *insights* de problemas complexos com grandes quantidades de dados e, em geral, descobrir padrões [34]. Técnicas de ML são frequentemente estudadas em aplicações MIMO [35, 36] que, como já dito, são de importância fundamental para as comunicações sem fio modernas e demandam muitos recursos da rede (tempo e banda) que devem ser usados de modo eficiente.

Logo, com o auxílio de técnicas de IA, o gerenciamento de feixes atua com acesso a informações de contexto, obtidas de forma alternativa ao convencional uso de sinais de piloto, por exemplo. Imagens, coordenadas de geoposicionamento e dados dos demais usuários são exemplos de informações que podem ser usadas para realizar o gerenciamento de feixes. De forma simples, podemos dizer que para um conjunto de dados de entrada, os algoritmos de IA fazem o mapeamento dessas informações para o domínio dos feixes. Pode-se questionar a disponibilidade das informações a serem utilizadas, contudo, a própria rede já dispõe de vários indicadores (*Key Performance Indicator* (KPI)) que podem ser analisados de forma conjunta ao invés de usar apenas os dados do enlace com apenas um usuário. Outros formatos de informação como a localização do usuário e imagens estão se tornando cada vez mais plausíveis, apesar das questões de privacidade do usuário. Como resultado dessa junção entre IA e o gerenciamento de feixes, tem-se a potencial redução do tempo para realizar as operações relacionadas à seleção de feixes e a otimização dos mecanismos de *beamforming* de acordo com o cenário.

O 6G é um cenário favorável para ambas tecnologias, IA e *beamforming*. Devido à dinâmica alta e flexibilidade previstas para o 6G, as técnicas existentes de *beamforming* e seleção de feixes ainda satisfazem os requisitos de resposta ágil, adaptabilidade e modelagem do ambiente. Como veremos a seguir, com o auxílio de técnicas de ML o gerenciamento de feixes adquiriu características mais dinâmicas, como a adaptação *on-line* de *codebooks*, e eficazes, como a seleção de feixes realizada em uma fração de tempo da seleção por meio de busca exaustiva e com desempenho comparável à tal técnica.

2.1.1 *Beam selection* em sistemas MIMO com arquitetura analógica

O problema de seleção de feixes consiste em encontrar o melhor par de feixes para que transmissor e receptor possam se comunicar explorando a melhor configuração possível da antena, considerando o cenário em que estão inseridos. Para tal, conta-se com *codebooks* previamente especificados no transmissor e no receptor. A partir desses *codebooks*, os feixes que levam ao maior ganho para o canal existente entre transmissor e receptor devem ser selecionados. Como já dito, esse problema torna-se inviável de ser abordado de modo exaustivo, ou seja, testando par a par os feixes entre transmissor e receptor, necessitando de um longo tempo de treinamento de feixes e retardando a comunicação de dados úteis. Outras abordagens além da exaustiva foram levantadas na literatura, como a hierárquica, assim como diversas heurísticas e aquelas utilizando IA.

As abordagens em IA para o problema de seleção de feixes o categorizam como um problema de classificação. Para esse tipo de problema, predominam as abordagens de aprendizado supervisionado, ou seja, várias instâncias de um vetor de dados x são associadas a uma saída conhecida y (também chamada de rótulo ou *label*) e o modelo de ML aplicado é treinado a fim de determinar uma regra geral que mapeie as entradas e saídas de um conjunto de dados de treinamento. Posteriormente, na fase de testes, o modelo de ML deve estimar as saídas para novas entradas, estimando a probabilidade $p(y|x)$ ou propriedades particulares da distribuição de probabilidades existente entre esses dois vetores [37].

No problema de seleção de feixes o vetor de entradas x é geralmente composto de dados como a posição do usuário, configuração do entorno, situação da rede, etc. A partir do conjunto de dados de entrada os algoritmos de ML estimam as configurações de feixe para transmissor e/ou receptor de modo a otimizar algum parâmetro. As abordagens tradicionais para o problema são limitadas à potência recebida ou SINR, como o acesso inicial do 5G *New Radio* (NR) ou o *beamforming* hierárquico do *Institute of Eletrical Electronic Engineers* (IEEE) 802.11ad. No 5G NR, há um período para transmissão de mensagens controle no *downlink*, quando há o envio de sequências de treinamento em cada feixe da antena da ERB e a estação móvel toma a decisão de qual feixe deverá ser utilizado para a comunicação entre eles baseado na potência recebida [38]. Com uma abordagem de ML, uma vez que o modelo foi treinado na ERB, é possível escolher o feixe de transmissão ótimo com tempo menor do que exaustivamente e visando otimizar diferente parâmetros, como veremos adiante.

No 6G, com o aumento expressivo no número de dispositivos conectados e a demanda ainda maior por capacidade e baixa latência, os sistemas MIMO e o problema de seleção de feixes devem apresentar soluções eficientes para dar vazão a essa nova demanda. Da Quarta Geração de Rede Móvel Celular (4G) para o 5G, houve um aumento de no máximo 4 antenas para um máximo de 64 antenas, o que habilitará o aumento de até 1000 vezes na capacidade de transmissão de dados [39]. Haja vista o maior dinamismo e exigências mais estritas em em termos de desempenho, o 6G dependerá ainda mais da união entre MIMO e ML.

Diferentes tipos de dados de entrada podem ser empregados no treinamento dos modelos de ML. Em [40], os autores abordam a seleção de feixes em redes veiculares explorando variações no conjunto de dados de entrada do tipo informação de contexto. Informações de contexto podem ser coordenadas de posição ou geolocalização da estação móvel, seu deslocamento, informações sobre o ambiente em que essa estação está localizada, entre outros. Neste trabalho, o conjunto de informações de contexto teve variados tipos de coordenadas e inserção ruídos na localização dos veículos, além de testar diferentes tamanhos do vetor de antenas e número de pares de feixes recomendados.

Nesse caso, o método *Random Forest Classifier* (RFC) pode atingir até 99% da vazão máxima, mesmo com arranjos de antenas 16 × 16 se comparado com outros métodos de ML, como *Gradient Boosting*, *Deep Learning* (DL) e *Radial Basis Function Neural Network Support Vector Machine* (RBF-SVM) e obteve obteve acurácia de 95% na recomendação dos 3 melhores pares de feixe transmissor/receptor para todos os arranjos de antenas testados. A seleção de feixes a partir de informações de contexto que é um problema de classificação altamente não-linear, por isso as redes neurais profundas podem lidar com este problema de forma adequada, pois suas múltiplas camadas. *Rezaie et al* usam esta técnica em [41] e a seleção de feixes é tratada como um problema de classificação *multi-label*. Os autores treinaram uma rede neural profunda usando a posição e a orientação do receptor para a seleção de feixes. Outros tipos de informação de contexto que podem ser exploradas por métodos de ML para o problema de seleção de são a potência recebida, o *Angle of Arrival* (AoA) [42], *Direction of Arrival* (DoA) [43], os ganhos dos múltiplos percursos que chegam até a estação móvel [44].

Outro dado que pode ser usado de modo estratégica são as imagens, que podem ser facilmente obtidas com o recente barateamento desse tipo de sensores. Além disso, métodos de ML são bastante eficientes e amplamente empregados em processamento e extração de informações a partir de imagens. Em [45] são usadas informações de contexto, por exemplo, a forma, a posição e até mesmo os materiais dos edifícios, carros e árvores circundantes. Estes dados são obtidos por múltiplas imagens obtidas por câmeras *off-line* a fim de construir uma imagem 3D. Esta imagem é a entrada da rede neural profunda, a qual tem como objetivo se adaptar a diferentes ambientes. A rede tem como saída os vetores com os índices ótimos de formação de feixe do transmissor/receptor. Outro tipo de abordagem, como em [46] e [47], imagens são formadas a partir da potência recebida pelos diferentes feixes e tratadas como um problema de busca de pico de calor em uma imagem. A imagem é criada a partir das matrizes de potência de recepção, sendo elas transformadas num mapa de calor de potência, portanto, cada matriz associada a diferentes feixes recebidos possui um único mapa de potência.

Outra estratégia que pode ser utilizada para gerar dados de treinamento é o uso de bandas em frequências abaixo de 6 Ghz ou sub-6GHz. Devido ao efeito de multi-percurso, as bandas sub-6GHz não são frequentemente exploradas em sondagens de canal e em sistemas de MIMO massivo, mas pode-se estabelecer conhecimento sobre a rede mesmo nessas bandas. Em [48], é considerada uma rede de comunicação heterogênea, onde coexistem pequenas *Base Station* (BS) mmWave com macro BS de sub-6 GHz. Através de sinais básicos extraídos do canal sub-6GHz o projeto de uma rede neural profunda é aplicada, com o intuito de dividir o problema em dois sub-problemas, um para a seleção da BS e outro para a seleção do feixe. A seleção da BS foi tratada como um problema de classificação, enquanto a seleção do feixe foi mapeado como um problema de regressão. Em [49] uma rede neural profunda foi usada para estimar a ocorrência de bloqueios na banda mmWave e determinar quais pares de feixes otimizariam a comunicação entre os dispositivos. De maneira semelhante, mas empregando também imagens de câmeras próximas às estações bases, uma rede neural foi aplicada em [50] com o mesmo objetivo de detectar bloqueios e estimar os melhores pares de feixe para a transmissão entre estações base e usuários espalhados em um cenário urbano.

Além do aprendizado supervisionado, a seleção de feixes também é frequentemente modelada através de um algoritmo de aprendizagem por reforço. A aprendizagem por reforço compreende um agente interagindo com um ambiente e obtendo respostas do sistema de aprendizagem e suas experiências, em termos de recompensas e penalidades. Esses algoritmos são compostos por duas fases. Durante a primeira fase, o agente explora o ambiente e as recompensas obtidas nessas interações. Na segunda fase, o agente traça uma estratégia baseado nas recompensas

coletadas na fase anterior, de modo a maximar as recompensas coletadas nas novas interações que será submetido. Os autores em [51] usam informações de contexto para auxiliar a tomada de decisão de um algoritmo *Multi-Armed Bandit* (MAB) de aprendizado *online*, onde as estações base são capazes de aprender a taxa de dados de cada feixe, obtendo desempenho superior aos demais algoritmos testados. Em [52], essa mesma abordagem foi aplicada na seleção de feixes 3D para VANTs usando dados de tráfego do *Google Maps*. Na abordagem apresentada em [53], os autores propõem usar a posição do usuário como informação de contexto para treinar as direções de feixes mais promissores. É usada a técnica de MAB com consciência de risco para reduzir a probabilidade de desalinhamento grave do feixe durante o aprendizado. Assim, são usados dois algoritmos em uma solução de aprendizado online de duas camadas, uma que seleciona o par de feixes e outro que as refina. Os resultados demonstram que a solução integrada possui um aprendizado rápido.

O problema de seleção de feixes revela-se relevante para a evolução das redes sem fio, principalmente no que tange à mobilidade, como em redes veiculares e redes para VANT que serão ainda mais comuns no 6G. Para essas redes é necessário que mecanismo de seleção de feixes se adapte aos bloqueios dinâmicos e padrões de tráfego, como em [43]. Apesar do número expressivo de trabalhos dedicados a esse tema, a seleção de feixes ainda é vista como uma problema isolado, centrado na otimização de métricas como potência recebida, capacidade e taxa de dados. A literatura ainda é carente de abordagens que, por exemplo, minimizem a interferência [54] ou permitam transmissões concorrentes [55]. Além disso, o uso de tecnologias emergentes, como LIDAR [56] e *Intelligent Reflecting Surface* (IRS) [57], podem dar mais subsídios para abordar o problema de seleção de feixes. Por fim, a criação de conjuntos de dados com canais MIMO pode facilitar a aplicação de ML em MIMO, fornecendo dados para serem utilizados durante a fase de treinamento [58].

2.1.2 Projeto do dicionário *codebook design*

Sistemas MIMO dependem de esquemas de *beamforming* direcional, que codificam ou decodificam os sinais para serem transmitidos através de múltiplas antenas e aproveitar-se desse recurso para aumentar o desempenho da rede. Para gerar um padrão de irradiação apropriado, o *beamformer* precisa obter informações sobre o estado do canal (com ou sem realimentação). O processo pelo qual o *beamforming* direciona o padrão de irradiação do sistema MIMO usando amostras do canal é também chamado de treinamento de feixes (do inglês, *beam training*).

O custo e o alto consumo de energia dos circuitos em alta-frequência inviabilizam a arquitetura de *beamforming* digital para arranjos de antenas com um grande número de elementos. Sendo assim, a maioria dos sistemas MIMO tende a seguir as arquiteturas de *beamforming* analógica ou híbrida. Essas arquiteturas de *beamforming*, devido as suas restrições de *hardware*, são utilizadas com o auxílio de *codebooks* de feixes previamente definidos, geralmente com um feixe por palavra código. Todavia, esses *codebooks* podem não ser eficientes em todos os cenários em que um transceptor MIMO seja aplicado. Para aumentar o desempenho da rede, é desejável que um *codebook* se adapte às condições em que o transceptor estará exposto [59].

Um exemplo de *codebook* genérico é o *codebook Discrete Fourier Transform* (DFT), que possui um feixe apontando em cada direção do espaço 3D. Esse *codebook*, apesar de simples e robusto, apresenta algumas limitações: ainda que cubra todas as direções, muitas delas podem não ter uso direto e aumentar o tempo da fase de treinamento de feixes [60]; por serem genéricos, esses *codebooks* podem ter o desempenho comprometido por imperfeições do *hardware* do transceptor [59]. Esses fatores levaram então a academia e indústria a pesquisar *codebooks*

adaptativos, gerados com auxílio de IA.

A maneira mais direta de se adaptar ao é utilizar indicadores já existentes ou amostras do próprio canal. Em [61], uma rede neural extrai características de propagação das amostras de canal e essas características são usadas para classificar as amostras através do algoritmo *K-means*. Assim, o *codebook* pode ser formado usando as características e seus valores que são válidos simultaneamente. Também por essa perspectiva de adaptação, não só ao cenário, mas também às limitações de hardware, uma rede neural artificial foi proposta em [62] para gerar *codebooks* cujos ajustes de fase refletissem os pesos da rede neural. A rede neural proposta obteve desempenho superior ao *codebook* DFT, principalmente em situações com mais de 16 feixes e quando múltiplos feixes foram ativados simultaneamente.

Por motivos de limitações no armazenamento e aquisição das informações que alimentam os métodos citados acima, os autores em [63] propusera um algoritmo de aprendizado *offline* que treina a partir de amostras geradas artificialmente. O resultado cada amostra permite atualizar a geração das amostras seguintes, até encontrar um ótimo para os problemas de otimização formulados para criar matrizes de *precoder* e *combiner* analógicos. Em [59], uma rede neural para deriva um *codebook* ótimo usando valores complexos, em conjunto com uma rede neural auto-supervisionada que não exige informações de canal pré-existentes, viabilizando o processo de aprendizagem online. Baseado nos pilotos recebidos em uma transmissão no *uplink*, com a arquitetura proposta, as *codewords* que gerem o maior ganho para o piloto recebido são escolhidas e ajustadas conforme o método de retro-propagação. Em [64], um algoritmo de aprendizado profundo utiliza apenas a potência recebida e mais nenhum dado a respeito do canal. Na primeira fase, este método define uma ação ótima, sem levar em conta as restrições. Em uma segunda fase, usando o algoritmo *K Nearest Neighbours* (KNN), a ação ótima é aproximada das ações mais viáveis, que serão avaliadas na fase seguinte e então a *codeword* é definida e a estratégia de aprendizado é atualizada.

Outro modo de empregar IA em projeto de *codebooks* é otimizar uma métrica de desempenho. Em [65], os *codebooks* foram projetados para aumentar a taxa de dados através da minimização da soma das distâncias da informação de canal real para as informações estatísticas do canal. O processo de *clustering* é baseado no bem difundido algoritmo *K-means*. Depois, diferentes métodos podem ser utilizados para montar os *codebooks* a partir dos centroides obtidos. Por outro lado, em [66], os autores visam, através de métodos de aprendizado por reforço profundo, definir um*precoder* pertencente a um conjunto pré-definido de modo a minimizar a BER. , dando ao método maior adaptabilidade.

A adaptabilidade conferida pelas diversas técnicas de ML aos *codebooks* vai de encontro aos objetivos estipulados para o 6G e suas características já propícias à IA. Contudo, ainda são poucos os trabalhos que assumam premissas mais estritas, como as que serão encontradas nos dispositivos comerciais. Por exemplo, ter acesso somente a informações de alto nível disponível nas camadas superiores. Ainda assim, ML pode ainda ser integrada a algoritmos de formação de *codebook* existentes de modo a otimizar os parâmetros desses algoritmos quando aplicados a determinados ambientes. Essas abordagens tornam esses algoritmos mais eficientes, adaptáveis e simples de serem aplicados. Em [67], uma rede neural profunda foi usada para ajustar os parâmetros do algoritmo *Convex 1-bit Precoder* (C2PO). Abordagens semelhantes podem ser encontradas em [68] e [69].

2.1.3 Precoding e combining em MIMO com arquitetura híbrida ou digital

Precoding e *combining* são técnicas que exploram a diversidade espacial de transmissão quando são utilizadas múltiplas antenas. Isto é, receptores em posições distintas no espaço recebem sinais distintos ao mesmo tempo, em uma mesma transmissão. O *precoding* (*combining*) atua no lado do transmissor (receptor), codificando (decodificando) os sinais transmitidos (recebidos) com ajustes de fase que maximizem o ganho da informação transmitida (recebida). Daqui para frente nesta subseção quando nos referirmos ao *precoder* estaremos também nos referindo ao *combiner*, exceto quando mencionados individualmente.

Diferente das frequências mais baixas, nas bandas de mmWave e THz, devido ao maior número de elementos de antenas montadas no mesmo arranjo, torna-se complexo e custoso realizar as cadeias de RF para cada um dos elementos do arranjo. Por outro lado, o *precoding* realizado de forma inteiramente digital na etapa de banda base aumenta a complexidade do *beamformer* com o número de *Analog-to-digital Converter* (ADC) necessários. Alternativamente, propôs-se realizar parte do *precoding* digitalmente em banda base e parte analogicamente na cadeia de RF, dando origem a arquitetura de *beamforming* híbrida, mais viável que a arquitetura puramente digital, mas que soma as restrições analógicas, como a quantização de ângulos nos ajustes de fase. Essas novas restrições podem refletir no projeto do *precoding*, o que deixa em aberto um novo tópico de pesquisa.

Também por consequência do maior número de antenas exigidos pelas comunicações nas bandas de mmWave e THz, as técnicas conhecidas de estimação de canal que dependem da sondagem de cada combinação de elementos de antena entre transmissor-receptor, ficam inviáveis, dada a sobrecarga que o treinamento dos equipamentos para o *beamforming* traria. Por isso, é necessário investigar algoritmos de baixa complexidade para estabelecimento da matriz de *precoding*, especialmente algoritmos que sejam multi-usuário. Para isso, um método promissor é o emprego de IA, que pode a partir de informações diversas a respeito do canal, do usuário ou da BS, determinar a formação de uma matriz de *precoding* ótima segundo algum critério de interesse, como a eficiência espectral [70], por exemplo.

Cada vez mais populares, as redes neurais são com frequência empregadas nos projetos de *precoder*, uma vez que as redes neurais podem chegar a resultados altamente precisos mesmo em aplicações não-lineares e complexas. Para exemplificar, os autores em [71] utilizam uma rede neural de aprendizado profundo para gerar amostras de canais artificiais e também para treinar um *precoder* híbrido com essas amostras, comparando os resultados com um ambiente simulado. Em [72], o *precoder* também é gerado a partir de canais artificiais usando uma rede neural convolucional, obtendo melhores resultados do que as soluções heurísticas, de aprendizado profundo e *Multi-layer Perceptron* (MLP) que foram comparadas no artigo. Porém, amostras de indicadores reais de rede são abundantes na maioria dos casos, como o AoA e *Angle of Departure* (AoD) [73], os pilotos presentes em diferentes configurações de quadros [74] e amostras do canal [75], e podem também alimentar redes neurais e resultar em *precoders* precisos e adaptados às condições se fazem necessárias.

Diferentes estratégias podem ser empregadas na geração de *precoder*. Como o arranjo de antenas possui vários elementos irradiantes, em alguns casos é possível formar sub-arranjos. Em [76] os autores propõem um método em duas etapas para formação de um *codebook* para *beamforming* híbrido com sub-arranjos de arranjos dinâmicos. Na primeira etapa, um algoritmo de clusterização hierárquica aglomerativa para agrupar as antenas do arranjo de modo a explorar as variações características do canal seletivo em frequências. Na segunda etapa, um algoritmo baseado em *Principal Component Analysis* (PCA) gera um *codebook* ótimo de

baixa dimensionalidade e com resposta em frequência plana a partir de um *codebook* seletivo em frequência. Outra forma de projetar *codebooks* foi proposta em [77], dividindo um *codebook* multi-usuário em *precoder* interno e *precoder* externo. O *precoder* interno é voltado para a multiplexação espacial, enquanto o *precoder* externo é voltado para a divisão espacial, ou seja, o interno divide-se em setores de usuários e o externo divide os usuários dentro de cada grupamento. Uma rede neural profunda é empregada para resolver o problema do *precoder* externo. A abordagem do artigo mantém o número de grupos fixo e o desempenho é próximo do ótimo estabelecido, que faz uma busca exaustiva pelo melhor *codebook*.

Alguns autores criticam método tradicional de estimar o canal e especificar *codebooks* separadamente. Em [78], por exemplo, é proposto um método de *Deep Neural Network* (DNN) que utiliza diretamente os pilotos recebidos em banda base para um desenho fim-a-fim da matriz de *precoding*. Em [79], um *precoding* para *beamforming* com otimização conjunta é proposta. A matriz de *precoding* é criada através de um método de entropia cruzada e posteriormente, para reduzir as interferências entre usuários, podem ser usados os algoritmos *Zero-forcing* ou diagonalização de blocos, para o caso de usuários com uma e múltiplas antenas, respectivamente.

O projeto *precoding* e *combining* auxiliado por técnicas de ML mostra-se um caminho possível para prover adaptabilidade e desempenho necessários às comunicações em altas frequências. Além disso, é possível atender a sistemas multi-usuários, contribuindo para os avanços em direção ao 6G, cuja a capacidade planejada da rede está além da capacidade atingida hoje. Passos concretos estão sendo dados para que as técnicas de ML se confirmem como método para o desenho de *precoding*, como a integração com o 5G NR [80] e a interação com IRS [81]. Contudo, ainda faltam na literatura alternativas de aprendizado em tempo real e que sejam aplicadas em equipamentos reais, o que são desafios a serem explorados pela academia e indústria.

2.2 Adaptação das Camadas Física e de Enlace

Anderson Reis Rufino Marins, Roberto Michio Marques Kagami
anderson@inatel.br, robertomk@inatel.br

Da primeira geração de redes móveis, com o objeto singelo de desafixar e conectar pessoas, à quinta geração, com o intento maior de interconectar quaisquer elementos na rede, a evolução de serviços e aplicações tem estabelecido uma série de requisitos particulares e dependentes das respectivas características de uso. Por consequência natural, este ciclo evolutivo demandará uma adaptação cada vez mais especializada de parâmetros, além dos já existentes na camada física e de enlace, para o melhor atendimento das futuras redes 6G.

2.2.1 Cenário: alta flexibilidade vs. complexidade de parametrização

Tomando como base as características de uso nas redes 5G, há uma estratificação em três principais categorias, baseadas nas necessidades de banda, latência, confiabilidade e presença massiva de elementos. Desta forma, surgiram as conhecidas classes de uso denominadas: *enhanced Mobile Broadband* (eMBB), URLLC e *Massive Machine-Type Communications* (mMTC).

Ainda assim, com a velocidade da evolução de tecnologias, surgimento de inovações e diferentes necessidades, já estão sendo apresentados novos atributos, ainda mais específicos e de especiais demandas, que passarão a gerar novas classes a serem apontadas na esteira da sexta geração. Alguns dos exemplos que já estão sendo prognosticados são como os que se seguem: secure URLLC, *Mobile Broad Bandwidth and Low Latency* (MBBLL), massive URLLC, *Extremely*

Reliable and Low Latency Communication (ERLLC), reliable eMBB, *ultra-mMTC* (umMTC), *Long-Distance and High-Mobility Communications* (LDHMC) e *Extremely Low-Power Communications* (ELPC) [82].

Diante deste contexto, a infraestrutura das futuras gerações de redes móveis necessitará uma flexibilidade ainda maior. Por consequência, um número significativo de parâmetros estabelecerá um grau de complexidade tal que, como tem sido a atual tendência, demandará o emprego de algoritmos de inteligência artificial.

A estratégia não seria diferente na camada física, onde diversos parâmetros têm que tornar o atendimento das características de uso não somente realizável como também o mais otimizado possível.

A forma de onda a ser entregue ao enlace é determinante quanto à ocupação dos recursos físicos básicos, ou seja, banda, tempo, espaço, potência, etc. Portanto, sua adequada configuração constitui um fundamento essencial de suporte às diversas aplicações da rede.

Certamente, a direção destas configurações é dada não somente pelas características de uso. Também é dada pela necessária avaliação de todas as questões pertinentes ao estado do enlace, particularidades comportamentais dos elementos que a integram e do sistema como um todo.

2.2.2 Numerologias e especificações

A flexibilidade da 5G pode ser considerada como o fator de maior diferenciação com relação ao *Long Term Evolution* (LTE). Este grau de aperfeiçoamento aponta para a ideia de direção para qual poderá evoluir a parametrização das futuras redes 6G no que diz respeito às numerologias. Para o 5G, estas foram definidas no Release 15 do 3GPP [83], sendo base para o tratamento paramétrico das redes futuras a serem abordadas nas próximas seções. São tratadas na especificação os seguintes itens:

- Espaçamento de subportadoras;
- Duração de símbolo;
- Prefixo cíclico;
- Duração de *slot*;
- Duração de quadro;
- Duração de sub-quadro.

Cada um desses parâmetros serão brevemente descritos a seguir:

Espaçamento de portadoras: O espaçamento de subportadoras não é mais fixado em 15 kHz, como no LTE. Em lugar disso, são cobertas escalas de 2^{μ} 15 kHz, o que permite ajuste da duração do quadro e da largura de faixa das subportadoras para diferentes aplicações. Os espaçamentos para as frequências Sub-6 GHz (450 a 6000MHz) são obtidos com $\mu \in \{0,\ 1\,\text{e}\,2\}$. Já para frequências mm-Wave (24.2 a 52.6 GHz), emprega-se $\mu \in \{2,\ 3,\ \text{e}\ 4\}$. O espaçamento de 480 kHz, não foi homologado e é objeto de estudo para as futuras versões do padrão 5G. A Tabela 1 apresenta os espaçamentos previstos.

Tabela 1: Parâmetros da estrutura tempo-frequência.

SCS μ	ΔF(kHZ)	Símbolos/*slot*	*Slots*/subquadro	*Slots*/quadro	CP
0	15	14	1	10	Normal
1	30	14	2	20	Normal
2	60	14	4	40	Normal
2	60	12	4	40	Estendido
3	120	14	8	80	Normal
4	240	14	16	120	Normal

Número de *slots*: Um *slot* corresponde ao recurso temporal que pode ser alocado para um usuário no 5G. Cada *slot* pode possuir 12 ou 14 símbolos OFDM, tal como pode ser visto na Tabela 1. O número de *slots* no quadro (*frame*) aumenta de acordo com o parâmetro μ. Então, assim como no LTE, cada quadro tem 10 ms de duração e cada sub-quadro (*subframe*) possui 1 ms de duração.

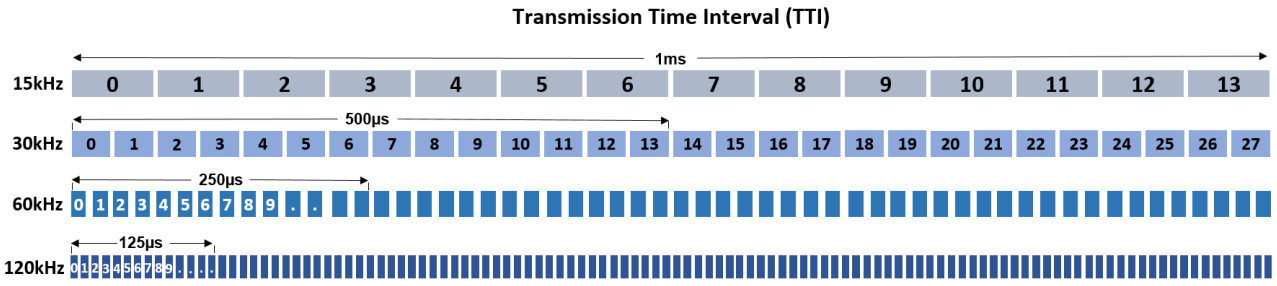


Figura 5: Duração do slot dentro de um subframe

Mini-*slots*: Um mini-*slot* pode conter 7, 4 ou 2 símbolos *Orthogonal Frequency-Division Multiplexing* (OFDM). A vantagem do mini-*slot* é que eles podem ser inseridos na estrutura do quadro sem a necessidade de aguardar a formação do quadro inteiro. Este procedimento é bastante interessante para aplicações que necessitem baixa latência.

***Slots* multifuncionais:** Outro grau de liberdade é dado pelo sentido dos *slots* que poderão ser direcionados para *downlink*, *uplink* ou podem ainda ser flexíveis. A cada símbolo OFDM pode ser atribuída a direção dos dados dependendo do balanço de tráfego necessário entre *downlink* e *uplink*. Este balanço pode ser otimizado conforme a necessidade de um determinado serviço.

Multiplexação de numerologias: A multiplexação de diferentes numerologias pode ser aplicada em um mesmo quadro, empregando a técnica denominada Bandwidth Part (BWP). A Figura 6 ilustra esse processo.

Uma dificuldade que este tipo de estrutura pode causar é uma interferência de subportadoras de uma numerologia em outra. Bandas de guarda podem ser necessárias para evitar a quebra de ortogonalidade no quadro.

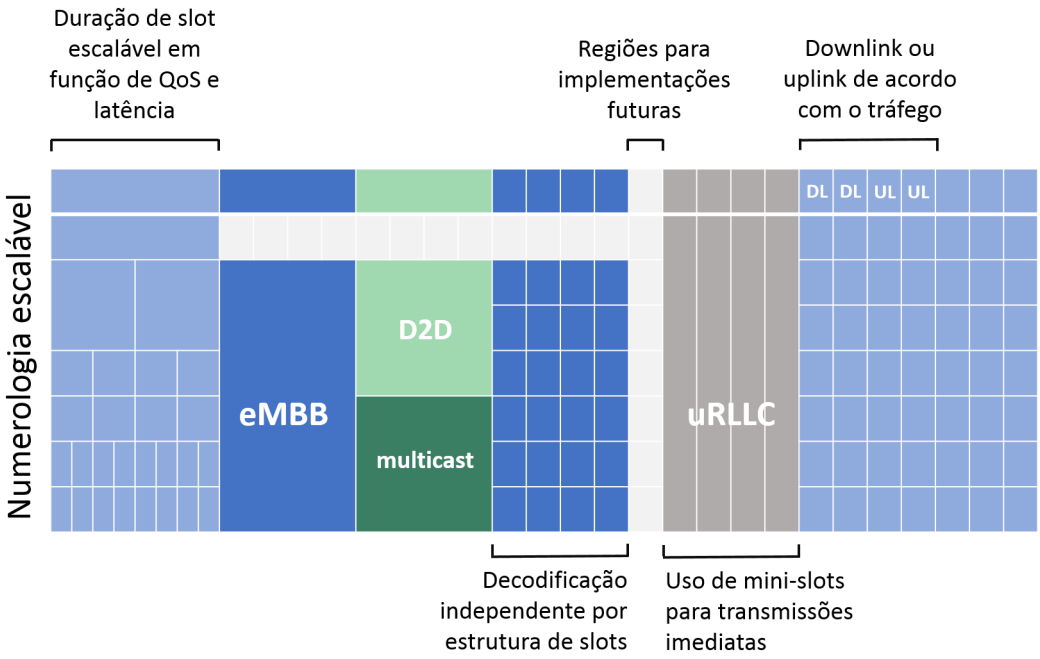


Figura 6: Multiplexação de diferentes numerologias [1]

Codificação de canal e multiplexação: A especificação TS 38.212 [84] trata da multiplexação e codificação de canal, descrevendo os canais de dados, transporte e controle da rede. A Tabela 2 apresenta os esquemas de modulação que podem ser empregados pelo 5G.

Tabela 2: Esquemas de modulação

Modulação	Bits/símbolo
$\pi/2$-BPSK	1
QPSK	2
16QAM	4
64QAM	6
256QAM	8

2.2.3 Análise de dados e geração de indicadores

A quantidade de recursos disponibilizada para cada usuário vai depender fundamentalmente das condições circunstanciais de cada enlace. Para a garantia do atendimento de requisitos e da qualidade dos serviços a estação base deve receber em tempo real, e com a máxima acurácia possível, todos os indicadores de estado dos canais.

Atualmente, os dados indicadores da qualidade dos enlaces são informados nos campos de *Channel Quality Indicator* (CQI) dos canais de controle. A evolução de novos indicadores de qualidade será de suma importância para a tomada de decisões paramétricas. Além disso, servirão de apoio para reportar características do enlace e fornecerão apoio para os algoritmos determinísticos de estimação e detecção, por exemplo.

Algumas iniciativas práticas já vem sendo desenvolvidas, como em [85] em que é apresentada uma implementação em lógica programável concebida para ser parte de uma arquitetura

proposta e denominada *context-aware cognitive transceiver* (CTR). Esta concepção habilita os transceptores a mudar suas configurações de forma automática, conforme as características de enlace detectadas. Como base de parâmetros de canal para indicar qual melhor configuração para determinado enlace e serviço, estão a estimação de *Signal-to-Noise Ratio* (SNR) e *Doppler Spread*. As opções para a adequação de parâmetros incluem três diferentes modos de utilização dos símbolos pilotos, duas identificações de canais diferentes e dois modos de detecção de dados. Quando o SNR é baixo e a mobilidade do usuário é alta, os resultados apontam um incremento de até 700% de vazão de pico e 40% de ganho de vazão a nível sistêmico quando comparados com arquiteturas convencionais de transceptores. Percebe-se, então, o quanto um procedimento para a escolha criteriosa de parâmetros, que se aproxime do estado da arte, pode proporcionar em termos de ganho para o sistema nas mais diferentes demandas de serviço.

Fatores como a medida de SINR [86], *Delay Spread* [87] e *Doppler Spread* [88], podem ser estimados com a assistência de algoritmos de aprendizado de máquina. Ao mesmo tempo, dados sobre esta variância podem ser investigados à luz de redes neurais, gerando indicadores de confiabilidade dos enlaces, o que pode ser elemento chave na geração da melhor alternativa quanto ao atendimento de determinado tipo de serviço.

Neste aspecto, é importante atentar a outros fatores referentes à comunicação destes indicadores para a estação rádio-base. Se as variações das características do enlace são rápidas, uma maior taxa de envio será necessária. Um canal ótimo e com baixa variação pode utilizar uma menor taxa, disponibilizando mais dados para as aplicações.

Considere-se o futuro cenário de configuração da camada física, em que a flexibilização de parâmetros atue basicamente em três frentes: forma de onda, numerologia e estrutura de quadros. Nesta última opção, este tipo de decisão a ser tomada em respeito ao janelamento de monitoramento e periodicidade de transmissão para os canais de controle, pode ser também tomada a partir de esquemas utilizando aprendizado de máquina.

O desempenho de detecção de sinal é muito prejudicado em termos de interferência interportadoras, ou *Inter-carrier interference* (ICI), e interferência intersimbólica, *Inter-symbol Interference* (ISI), quando a propagação com atraso por multicaminhos tem componentes significativos que excedem o comprimento do período cíclico, *Cyclic Prefix* (CP). Existem, atualmente, alguns algoritmos de aprendizado de máquina que são relativamente bem sucedidos na mitigação de efeitos ocasionados por um tamanho insuficiente deste, como em [89]. Embora os resultados sejam interessantes, estes ainda ficam significativamente aquém do desempenho obtido quando o tamanho de CP é suficientemente dimensionado.

Algumas numerologias da tecnologia 5G contam com tamanhos diferentes (normal e estendido), mas a gama de valores deste parâmetro tende a ser maior para que uma melhor eficiência seja atendida de acordo com as características do enlace. Um tamanho menor possível sempre é interessante para uma melhoria de eficiência espectral. Desta forma, é importante que sejam identificadas as componentes mais significativas em termos de atraso e de potência para posterior parametrização do tamanho de prefixo cíclico mais adequado. Em [87] é apresentada uma proposta de rede neural profunda para esta identificação.

Uma questão que é particularmente importante para a determinação de simulações, no que diz respeito à geração de provas de conceito quanto à parametrização, é a modelagem de canais. É tarefa fundamental para a pesquisa e implementação de sistemas de comunicação, principalmente quanto à camada física. A utilização de novas e mais altas frequências tem dificultado ainda mais o levantamento de modelos, sendo necessário um profundo conhecimento de propagação de sinais de radiofrequência e campos eletromagnéticos. É uma elaboração muito complexa e que compreende uma gama elevada de parâmetros.

O número crescente de novos cenários, novas bandas de frequência e número de células irão gerar conjuntos de dados massivos. Entre estes, medições de canal irão gerar grandes quantidades de informações sobre as características presentes nos enlaces. Algoritmos de aprendizado de máquina utilizando redes neurais podem ser utilizados para processar este grande conjunto de dados para aprender as estruturas de canais existentes. Há, inclusive, proposições para que sejam gerados modelos unificados e compartilhados para os sistemas de comunicação sem fio.

Neste campo, também existem técnicas de inteligência artificial para geração de dados de canais aleatórios para simulações, mas que tem características iguais à de canais reais. Para a implementação de algoritmos capazes de estabelecer tamanha conformidade com canais verdadeiros, há técnicas como a denominada *Generative Adversarial Network* (GAN), que é, por exemplo, proposta e executada em [90].

Esta técnica se baseia em duas redes neurais que competem entre si. No caso desta implementação, há uma rede geradora e outra denominada discriminadora. Um banco de dados com canais reais é comparado pela rede neural discriminadora com os diversos modelos calculados pela geradora. É então determinado, para cada par, qual é classificado como real. O aprendizado é realizado até que ambas as redes atinjam um estágio que permita à geradora um grau elevado de semelhança para validação e, à discriminadora, um grau de exigência de conformidade apurado.

2.2.4 Predição e controle

Métodos de predição são matéria de vários estudos nas mais variadas áreas envolvendo controle de processos. Ao atingir um grau de confiabilidade de alta acurácia, a fluência de funcionamento, grau de acerto em tomadas de decisão e prevenção de falhas oferecem o melhor aproveitamento possível dos recursos disponíveis. Diante do complexo cenário de comunicações móveis, isto não seria diferente. O emprego de técnicas de aprendizado é uma poderosa ferramenta neste sentido.

Indicador de qualidade de canal: Com o maior emprego de ondas milimétricas, o comportamento de medida para determinação do CQI tem uma característica mais instável, gerando uma ineficiência maior, diferentemente do que ocorre em enlaces utilizando frequências menores. Propostas neste sentido são apresentadas como em [2]. Entre outras topologias, o estudo aponta o uso do método baseado em *Long Short-Term Memory* (LSTM). É uma arquitetura de rede neural artificial recorrente (RNN) que possui conexões de feedback. Nesta topologia é possível processar não somente dados únicos como também sequências de dados. É um dos métodos mais utilizados em modelos de predição.

O princípio de funcionamento pode ser descrito como duas porções de RNN, onde uma é responsável por criar uma representação vetorial fixa a partir de uma sequência e a outra se utiliza deste dicionário para fazer a predição de sequências futuras. A Figura 7 apresenta a arquitetura do modelo LSTM.

Foram implementados dois cenários de simulação para avaliar o desempenho dos algoritmos de predição. Um deles utilizando somente um usuário que se movimenta com velocidades constantes de 10ms a 70ms e outro, com os mesmos parâmetros de velocidade, com quatro usuários independentes.

O atraso atribuído ao enlace para envio de informação de CQI foi de 0,5ms, por se tratar do valor de 1 slot para a numerologia utilizada. A frequência é pertencente à faixa de mmWave em uma banda de 100MHz.

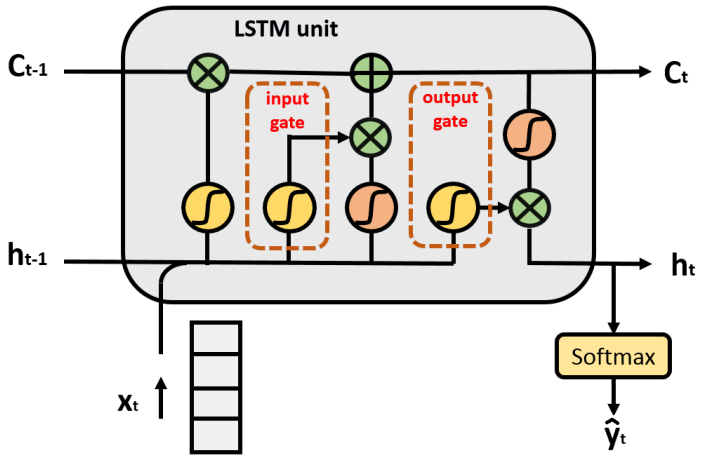


Figura 7: LSTM - Arquitetura

Outro modelo testado para a predição foi o *Feedforward Neural Network* (FNN), que é uma rede neural tradicional com um hidden layer. Os resultados comparativos são ilustrados na Figura 8. É possível perceber uma melhora significativa na vazão quando utilizada a topologia de rede neural LSTM.

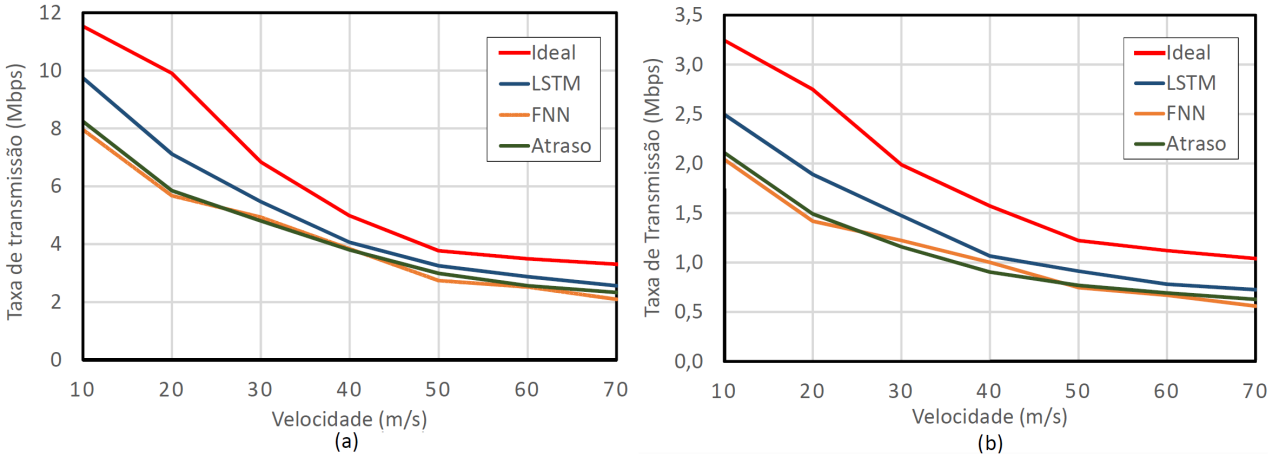


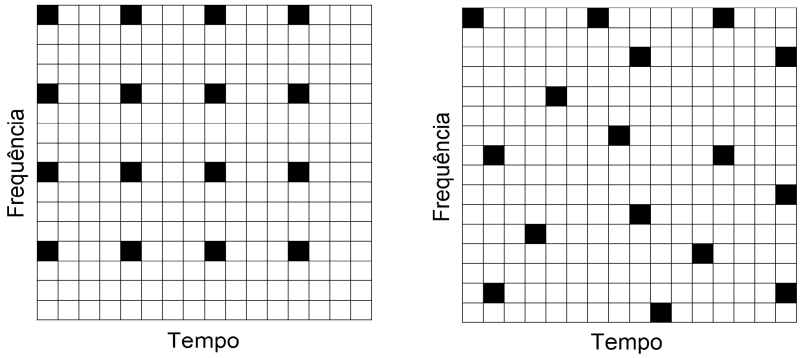
Figura 8: Performance de vazão: (a) Um usuário (b) Multi-usuário [2]

Esta mesma questão também é tratada em [91]. Há um estudo de impacto no atraso de *Channel State Information* (CSI) e de sistemas atualmente existentes que empregam algoritmos de aprendizado. Além disso, é feita uma análise comparativa a respeito tanto da performance como das topologias, quantidade de camadas ocultas (*hidden layers*) e complexidade computacional em cada caso.

Mapeamento de canais piloto: A utilização de canais piloto em uma estrutura de quadros OFDM para realização de estimação de canal é uma prática amplamente empregada. Contudo, algumas das desvantagens desta aplicação reside no compromisso entre a precisão mais acurada no cálculo do canal, nos *Resource Block* (RB)s onde estão os dados, e a eficiência na utilização do espectro. Se o espaçamento entre pilotos é menor, melhor é a estimação de canal dos

símbolos com dados, visto que esta é calculada utilizando técnicas de interpolação. Porém, com o aumento no número de pilotos, a eficiência espectral diminui. Características específicas de cada enlace determinam um maior ou menor espaçamento. Por exemplo: uma maior velocidade de um usuário acarreta maiores variações de canal e desvanecimento seletivo, necessitando um distanciamento temporal menor entre pilotos.

Para uma escolha adequada deste mapeamento, também podem ser utilizados algoritmos de aprendizado de máquina. Um tratamento de dados, como a redução de dimensionalidade de coeficientes do canal estimado pode gerar informações para a determinação das variações e velocidades envolvidas no enlace. Em alguns estudos, como [92], há a indicação de que a solução ótima destes espaçamentos tem uma distribuição não uniforme. A Figura 9 mostra dois exemplos de mapeamento de símbolos pilotos.



(a) Uniforme (b) Não uniforme

Figura 9: Distribuição de símbolos pilotos.

Neste trabalho, o grid formado pela distribuição de símbolos no espaço-tempo é tratado como se fosse uma imagem onde os símbolos pilotos representariam, em baixa resolução, uma imagem de alta resolução que comporia a resposta do canal incluindo os RBs. Sendo assim, a escolha é baseada nas posições que apresentam a maior significância para obtenção do menor erro de reconstrução possível quando efetuada a estimação deste canal específico. Ocorre que o número de combinações existentes torna esta escolha inviável em termos de processamento. Assim, um algoritmo de seleção de características baseado em aprendizado de máquina, ou *Deep Learning-based feature selection*, é utilizado para realizar a escolha da melhor distribuição de pilotos dentro de um conjunto pré-determinado de opções que representam um escopo de canais mais frequentes.

Intervalo de tempo de transmissão: A flexibilidade de *Transmission Time Interval* (TTI) já presente na quinta geração, é de fundamental importância para cumprimento das demandas futuras em termos de aplicação, principalmente naquelas que serão baseadas nas características como as definidas como URLLC e mMTC. A melhor escolha de opções quanto ao TTI tem sido avaliadas em trabalhos, sendo alguns fazendo uso do aprendizado de máquina.

Nos estudos encontrados em [3], por exemplo, são analisados vários tipos de cenários com diferentes parâmetros de intervalo para determinação de quais escolhas seriam mais adequadas quanto ao aspecto de atraso. O algoritmo utilizado é baseado na categoria de florestas aleatórias (*Random Forest based Ensemble TTI Decision Algorithm*) e o processo é denominado pelos autores como *Self-adaptive Flexible TTI scheduling* (SAF-TS).

Na Figura 10 há o comparativo entre a estratégia de agendamento utilizando a rede neural proposta e a escolha de valores fixos de TTI. É possível perceber a melhora em termos de performance quanto ao aspecto de latência nos três cenários propostos: de baixo, médio e alto volume de tráfego.

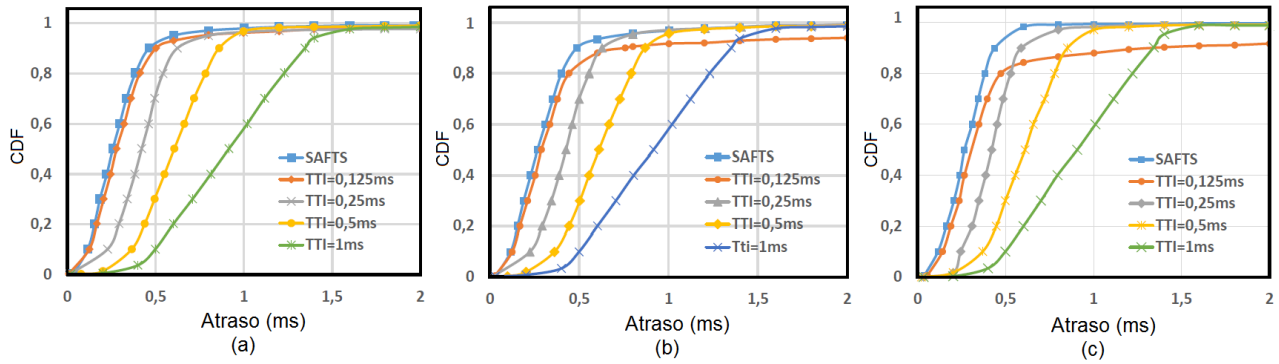


Figura 10: Latência de serviços URLLC em função do tráfego (a) baixo (b) médio (c) alto [3].

Controle de potência: A eficiência espectral e a eficiência de consumo de energia são questões a serem resolvidas para atender as novas exigências do sistema de comunicação. Vários estudos vêm sendo feitos para melhoria da eficiência energética no quesito de otimização de potência de transmissão e taxa de dados de usuário através do controle de potência. Porém para tal, é necessário que se tenha um conhecimento preciso e instantâneo do modelo do sistema [93]. Para melhoria da eficiência energética, uma nova arquitetura para as estações de base tradicionais são estabelecidas, separando-as em *Cloud-RAN* (C-RAN). Dentro das C-RAN há duas partes: a *Baseband Unit* (BBU) a qual faz todo o processamento e fica centralizada na célula, e a *Remote Radio Head* (RRH) responsável pelos transceptores de sinal que ficam distribuídas por toda a célula.

Com o aumento dos dados trafegados na rede, a instalação de RRH se torna mais densa e a comunicação de *Device-to-Device* (D2D) é utilizada. Com o uso da comunicação D2D há uma redução da energia de consumo do aparelho pois a potência de transmissão é reduzida [94] e o reuso dos recursos do celular gera um aumento na eficiência espectral do sistema.

Para garantir um melhor desempenho no controle de potência é utilizado de IA para gerenciamento, conforme em [94] é utilizado o algoritmo de Reinforcement Learning. Porém, com o alto número de RRHs em funcionamento, há um aumento na interferência entre células o que degrada drasticamente a comunicação.

Para mitigar esse problema, é utilizado um algoritmo de Q-learning para que a BBU gerencie qual e quantas RRHs estarão em funcionamento, conseguindo-se também reduzir consideravelmente a energia gasta por todo o sistema de comunicações móveis [94]. Abaixo, são apresentados os resultados comparando três soluções, onde "SINR" é o alvo ajustado proposto em [95], "Controle de potência baseado em Q-learning I" [96], "Chaveamento de RRH" [97] e "Algoritmo proposto em Q-learning" [94]. Foi considerado um cenário multicelular em ambiente C-RAN com os parâmetros apresentados na Tabela 3.

Tabela 3: Parâmetros da simulação

Parâmetros	Valor
Largura de Banda Total	100MHz
Densidade Espectral de Potência de Ruído	-174dBm/Hz
Restrição de SINR	0,5dBm
Potência de transmissão máxima	23dBm
Restrição de probabilidade máxima de interrupção	0,05
Limite mínimo de probabilidade de interrupção	0,01
Declive de RRH	4,0

Esses valores estão baseados nas especificações do 3GPP. São consideradas 9 RRHs e dispositivos que foram distribuídos aleatoriamente em 3 células de 500m cada. Foi usado o simulador SUMO para gerar o conjunto de dados de dispositivos de mobilidade de pedestres.

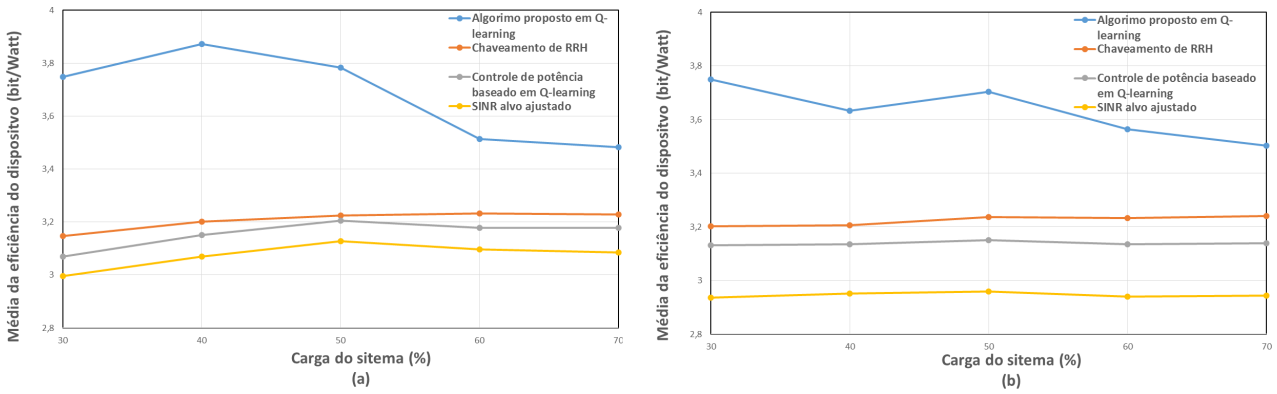


Figura 11: Comparação de eficiência de energia de dispositivo: (a) 150m e (b) 350m

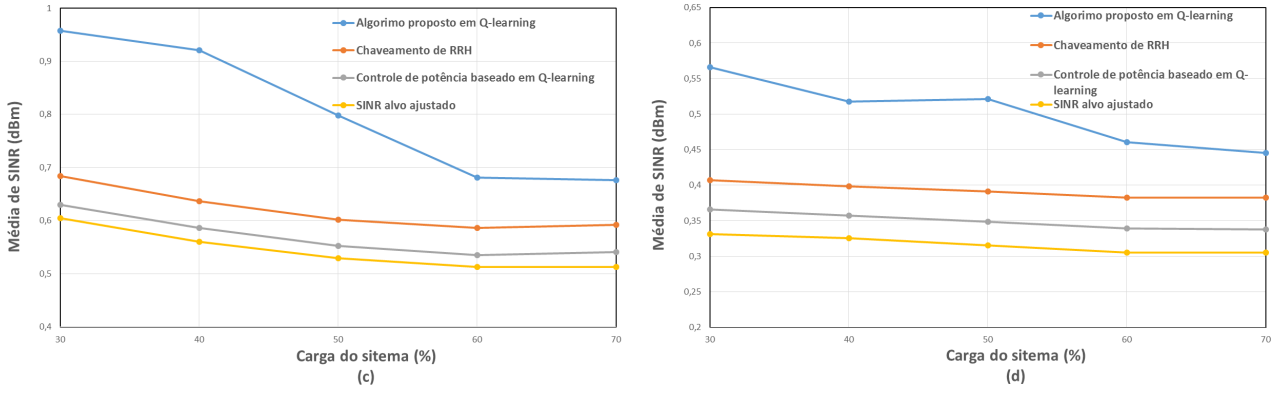


Figura 12: Comparação das médias de SINRs: (a) 150m e (b) 350m

A Figura 11 apresenta as respostas de variação de carga do sistema pela média de eficiência energética do dispositivo e a Figura 12, a variação de carga pela média do SINR onde se destaca o Algoritmo proposto em Q-learning, pois atua tanto na BBU para controle do número de RRHs ativas, quanto na utilização de comunicação D2D com controle de potência de transmissão.

Controle de interferência inter-numerológica: A flexibilidade de numerologia é de fundamental importância para o melhor atendimento de todos os alvos de qualidade esperados para os serviços nas redes móveis futuras. Em contrapartida, alguns novos efeitos surgem se houver um emprego não devidamente controlado dos parâmetros. Uma delas é a interferência inter-numerológica, ou *Inter-Numerology Interference* (INI).

Embora complexa, uma esquematização do uso de diferentes numerologias, no que se refere à mitigação destes efeitos, pode ser executada com algoritmos utilizando redes neurais. Em [98] é proposta uma estrutura com decisões baseadas em aprendizado supervisionado de máquina. Os dados que são levados em conta para o treinamento da rede neural são: tipo de serviço, distância do usuário, espaçamento de subportadora, presença de banda de guarda, nível de potência, banda e tipo de canal.

A distância do usuário é importante, por exemplo, quanto à questão de confiabilidade de um enlace. Usuários localizados nas bordas de uma célula possuem maior suscetibilidade à interferência inter-células. Nos gráficos da Figura 13 é possível perceber a influência dos fatores espaçamento de portadora e presença de banda de guarda na interferência inter-numerológica. Estes são aspectos significativos para as escolhas paramétricas e de alocação dos usuários.

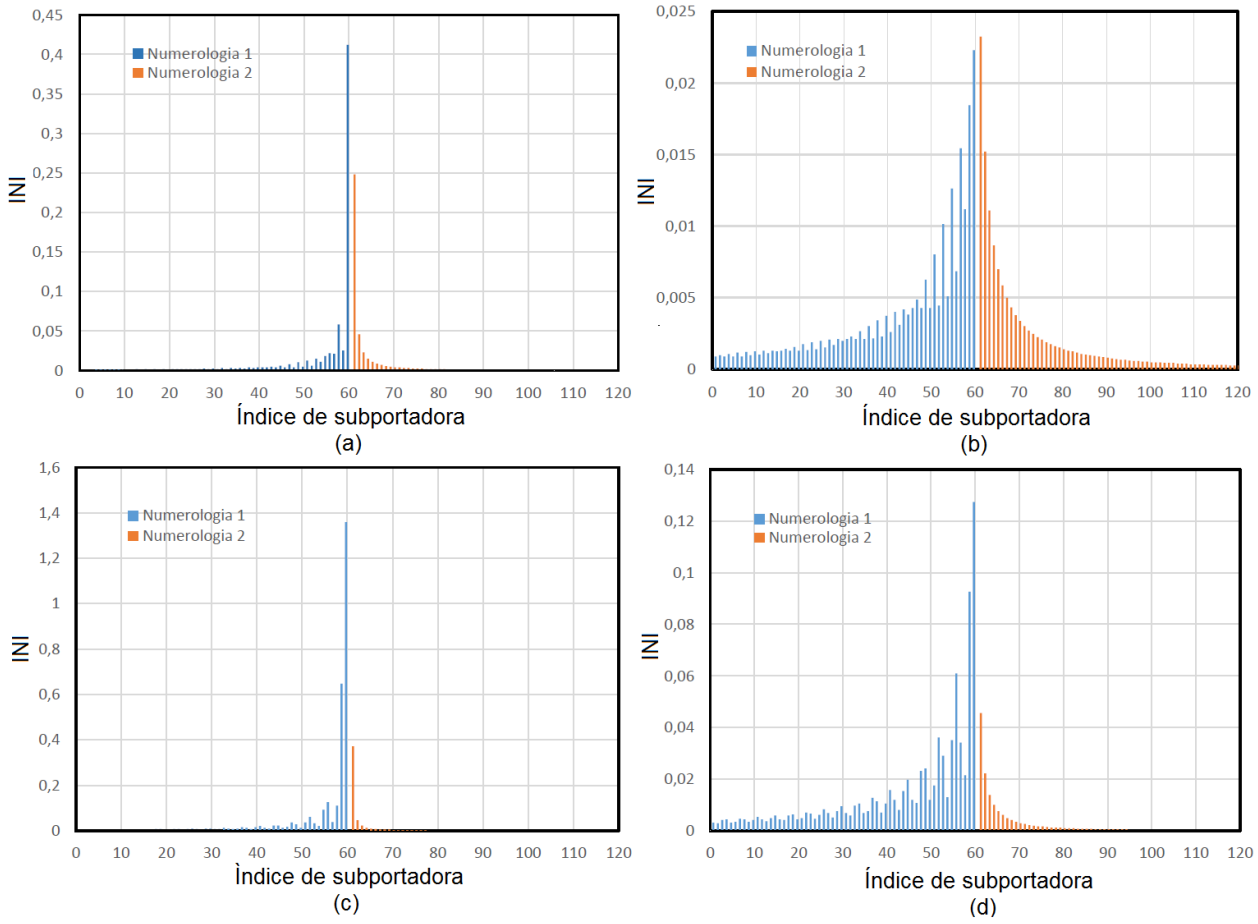


Figura 13: Interferência inter-numerológica: (a) sem BG, Num.1 = 15kHz, Num.2 = 30kHz. (b) com BG, Num.1 = 15kHz, Num.2 = 30kHz. (c) sem BG, Num.1 = 15kHz, Num.2 = 60kHz. (d) com BG, Num.1 = 15kHz, Num.2 =60kHz.

A diferença de potência entre usuários com numerologia diferente aumenta a relação sinal-interferência *Signal to Interference Ratio* (SIR). Levando em conta os fatores supra-citados, é

executada uma simulação onde são utilizados vários usuários com perfil de uso URLLC e vários outros não pertencentes a este perfil. A seguir, três categorias de algoritmos foram executadas, onde uma alocação é realizada de forma aleatória, outra priorizando usuários URLLC que estejam nas bordas da célula (algoritmo 1) e outra priorizando a minimização de diferença de potência entre os usuários (algoritmo 2).

Os resultados estatísticos para os usuários não pertencentes à borda da célula são bastante semelhantes. Contudo, para os usuários localizados na borda, há uma nítida melhora quanto ao aspecto de diminuição de SIR. A Figura 14 apresenta o levantamento estatístico deste cenário.

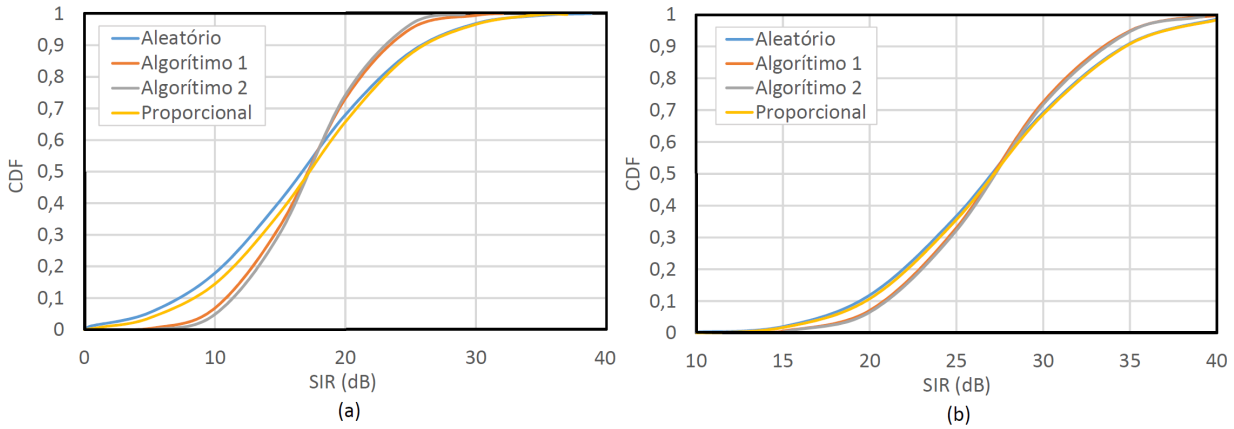


Figura 14: Relação sinal-interferência (SIR) em função dos algoritmos: (a) Sem banda de guarda. (b) Com banda de guarda.

Foram citados aqui alguns exemplos do que vem sendo proposto e investigado em termos de utilização de IA para o estado da arte no tratamento e geração de dados e indicadores, assim como predição e controle dos processos envolvidos na especialização de serviços no contexto das camadas física e de enlace. Invariavelmente, os resultados obtidos são melhores que os métodos tradicionais, o que aponta para direção a ser percorrida no desenvolvimento das redes 6G. O grande desafio para estas camadas, porém, é o processamento em tempo real necessário para implementação das novas técnicas. A melhor abordagem de soluções certamente passará pelo balanço na execução de aprendizado com a utilização de modelos já otimizados para inferência, como também pelo seu uso como apoio às técnicas convencionais existentes.

2.3 Alocação de recursos *cross-layer* ou multidimensionais

Victor Hugo L. Lopes, Cleverson Veloso Nahum, kleber Vieira Cardoso
victor.lopes@ifg.edu.br, cleversonahum@ufpa.br, kleber@inf.ufg.br

De maneira geral, parte dos problemas de otimização dos sistemas de comunicação sem fio, quando observados nas camadas mais baixas da pilha de protocolos, envolve os desafios de escalonamento de usuários e alocação de recursos. Estes problemas de otimização envolvem, invariavelmente, a busca pela melhoria de determinados atributos dos sistemas, tais como capacidade, potência de transmissão, acesso à canais, controle de interferências, entre outros, respeitando-se certas restrições inerentes ao sistema a ser otimizado.

Os desafios envolvendo a administração dos recursos de rádio tendem a se tornar cada vez mais complexos ao passo que novos casos de uso vão sendo projetados, gerando requisitos de

desempenho rigorosos, e que tendem a gerar uma extrema ampliação do número de usuários e/ou dispositivos disputando os sempre escassos recursos, de forma que a proposição de métodos de alocação e escalonamento robustos para tais demandas precisa avançar.

Mesmo se tratando de tarefas nem sempre triviais, em especial nos casos de grande dimensionalidade de atributos e/ou graus de liberdade a serem abordados pela otimização da política de alocação de recursos em questão, e que observa-se na literatura diversas soluções propostas baseadas em heurísticas, tais problemas costumam possuir características similares aos problemas de aprendizado automatizado, de forma que o emprego de ferramentas baseadas em aprendizado de máquina mostra-se viável [99, 100].

Neste sentido, esta subseção visa descrever os aspectos inerentes aos desafios da administração de recursos de rádio com suporte de IA, que garantam a evolução requerida para acomodar os importantes requisitos projetados para as redes sem fio de próxima geração, tais como *Beyond* 5G (B5G) e 6G, provendo uma visão sobre o atual estado da arte, e as possibilidades futuras nesta temática.

2.3.1 Escalonamento de usuários e alocação de recursos PHY/MAC até o 5G

Apesar de o espectro de radiofrequências ser fisicamente inesgotável, características de propagação reduzem significativamente sua porção útil aos sistemas de comunicação sem fio, cada vez mais obrigados a operar em faixas de frequências licenciadas de alto custo e/ou extremamente povoadas. Assim sendo, os sistemas de comunicação sem fio são projetados para operar de forma a maximizar a eficiência do uso destes limitados recursos [101], o que se torna ainda mais importante quando observadas as características da evolução das redes celulares, culminando nos serviços previstos para as redes 5G, que propõe alta capacidade de dispositivos conectados, baixa latência e alta taxa de dados, em que parte significativa da melhoria da eficiência/satisfação de usuários e sistemas passa, obrigatoriamente, pela adoção de métodos dinâmicos de alocação e gerenciamento dos recursos disponíveis.

Tais métodos de alocação e gerenciamento de recursos envolvem a tomada de decisão baseando-se em certos tipos de estratégias (ou políticas), de forma a buscar a maximização de determinado atributo de desempenho do sistema, ao passo que não deteriora significativamente outro atributo de desempenho intrinsecamente relacionado. Quando se tratam de recursos e funções das camadas *Physical Layer* (PHY) e *Medium Access Control* (MAC), em que destacam-se o escalonamento de recursos de rádio em tempo e frequência para transmissão em *uplink* e *downlink*, multiplexação de dados através de múltiplas portadoras, bem como a própria adaptação de links, codificação e decodificação, modulação e demodulação, processamento multiantena, controle de potência e mitigação de interferências [102], entre outras, é por intermédio destas camadas que tais políticas devem ser otimizadas.

Como exemplo, dado que uma das formas geralmente aplicadas para se obter uma melhoria da eficiência espectral de um dado *link* é pela permissão de que múltiplos dispositivos o utilizem de forma compartilhada, sabe-se que a eficiência espectral será limitada pela potência de transmissão disponível. Desta forma, uma política ótima buscará alocar as potências de cada usuário ativo, ou até mesmo escalonar quais usuários devam ser alocados, de forma a garantir o desempenho esperado, sendo que um controle inapropriado pode permitir níveis indesejados de interferência e a consequente degradação da comunicação.

De forma diferente das gerações anteriores, projeta-se para o 5G NR os objetivos de prover suporte otimizado para um grande número de diferentes tipos de serviços, com requisitos de usuário diversos, em que podem ser organizados em três diferentes classes de casos de uso [4,

102, 103]:

- eMBB: grande incremento de taxas de dados para usuários móveis, com eficiência espectral aprimorada e cobertura ampliada, dando suporte à conectividade centrada no usuário, com grande densidade de usuários conectados, permitindo acesso em alta velocidade a recursos de multimídia, serviços e dados, mesmo quando em mobilidade em alta velocidade.
- mMTC: também descrito como IoT (Celular) Massivo [103, 104], projetado para dar suporte ao massivo crescimento de dispositivos conectados, gerando tráfego *Machine-Type Communications* (MTC), superando o tráfego gerado por humanos, sendo o foco prover conectividade a esse grande número de dispositivos, que possuem como características as transmissões esporádicas, de baixo tráfego e não sensível ao atraso, com heterogêneas capacidades, custos, consumo energético e potências de transmissão.
- URLLC: trata-se de comunicações críticas [104], em que os rigorosos requisitos projetados tanto para latência e confiança também perpassam o MTC, dando suporte a aplicações como *Industrial Internet of Things* (IIoT) e os processos industriais sensíveis ao atraso, cirurgias robóticas e remotas, veículos autônomos, entre outros.

Estas classes de casos de uso, e os seus cenários de aplicação, desencadeiam uma série de requisitos desafiadores paras as redes 5G (Figura 15, adaptado de [4]), principalmente em termos de atendimento aos requisitos de *Quality of Service* (QoS) e *Quality of Experience* (QoE), em que a busca por mecanismos extremamente otimizados para a alocação dos recursos PHY/MAC se acentua.

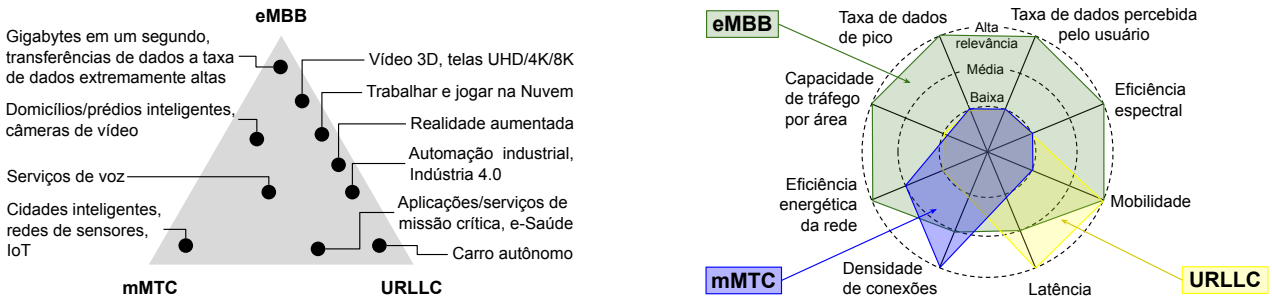


Figura 15: Cenários e requisitos definidos para as redes 5G [4].

O emprego de múltiplas antenas já é importante, de fato, desde o LTE, mas será a partir do 5G NR que ele passará a desempenhar um papel cada vez mais importante, sendo fundamental para o cumprimento dos requisitos de desempenho almejados, permitindo, entre outros aspectos, lidar com a interferência enquanto permite o uso simultâneo de um mesmo recurso de rádio (banda de frequência, por exemplo) por dispositivos espacialmente separados, ao permitir o direcionamento de feixes de sinais na direção dos usuários pretendidos, ao passo que limita a interferência para os demais usuários. Quando considerado o emprego de ondas milimétricas (*mmWave - millimeter wave*) em conjunto com as múltiplas antenas, o potencial de desdobramento dos referidos requisitos se torna ainda maior.

Para garantir o cumprimento das rigorosas restrições de latência e confiabilidade exigidas pelos serviços URLLC, a coexistência num mesmo recurso de rádio com serviços eMBB deve

ocorrer, de forma que a eficiência espectral pode ser seriamente comprometida pelo fato de que os serviços URLLC devem sempre ocupar recursos de rádio de forma exclusiva e prioritária em relação aos serviços eMBB [105], em que a alocação de recursos se torna ainda mais desafiador quando considerada a mobilidade de dispositivos e tal heterogeneidade. Neste sentido, os autores em [106] apresentam um método para alocação de RBs entre usuários móveis URLLC e eMBB coexistindo em um mesmo feixe de ondas milimétricas, com emprego de *Deep Reinforcement Learning* (DRL) baseado em memória longa de curto prazo (LSTM), em que embora os experimentos não permitam observar resultados em cenários densificados, mostrou grande potencial em termos de controle de latência e perda de pacotes, bem como de melhorias em vazão. Em [107] o problema de *Scheduling and Resource Allocation* (SRA), de forma a atender diferentes requisitos de QoS, é enfrentado com DRL, tanto em cenários com completo conhecimento do CSI, quanto quando não há tal conhecimento, em que os resultados são superiores aos métodos de otimização combinatórios convencionais.

A alocação de recursos e o *Radio Resource Management* (RRM) de forma a satisfazer os acordos de nível de serviço (SLAs - *Service Level Agreement*s) em redes com emprego de fatiamento de redes na *Core Network* (CN) e *Radio Access Network* (RAN) é apresentada em [108], em que a IA é utilizada para prever o comportamento da rede celular e tomar melhores decisões no ajuste dos mecanismos RRM, e incluindo um orquestrador para a RAN, ciente dos KPIs, com a responsabilidade de prover ajustes no controle de admissão e escalonamento de pacotes, garantindo que tais indicadores sejam satisfatórios. Tal abordagem visa facilitar o processo de acomodar serviços distintos como aqueles previstos para o 5G NR, em que as redes lógicas fim-a-fim podem também coexistir.

A alocação de recursos multi-dimensionais é abordada em [109], no qual um *Intelligent Multiple Access* (MD-IMA) em cenário de redes celulares *Time Division Duplex* (TDD) é empregada para cumprir os requisitos diversos de QoS em tempo real, utilizando os recursos de rádio em domínios heterogêneos. Para tal, uma rede neural LSTM é utilizada para a predição da dinâmica de longo prazo da rede e inferir as mudanças nos requisitos de QoS, enquanto um algoritmo *Deep Deterministic Policy Gradient* (DDPG) é adotado para otimização dos recursos de rádio em tempo real, de acordo com as flutuações da situação da rede, sendo útil para a redução da latência de processamento.

2.3.2 Escalonamento de Recursos PHY/MAC em rede móvel 6G

De forma a permitir uma projeção sobre os desafios do escalonamento de recursos das camadas PHY e MAC para as redes 6G, a Figura 16 apresenta as classes de casos de uso descritas em [5], em que as classes previstas para o 5G precisam ser otimizadas e conjugadas para dar suporte às novas aplicações, sendo elas:

- *Ubiquitous Mobile Broadband* (uMBB): originado pela junção de eMBB e mMTC, com vistas ao suporte das comunicações *on-board* de alta qualidade e conectividade ubíqua e global, com importante incremento de capacidade da rede e taxa de transmissão em pontos de acesso como suporte a serviços disruptivos, tendo o pico de taxa de dados, a taxa de dados percebida pelo usuário, a mobilidade, o pico na eficiência espectral, a capacidade de tráfego por área, a largura de banda, cobertura e CAPEX/OPEX como KPIs críticos.

- *Ultra-reliable Low-latency Broadband Communication* (ULBC): seguindo na mesma tendência da coexistência URLLC e eMBB vista nos desafios de alocação de recursos no 5G,

tais serviços devem exigir não apenas prioridade nas alocações que garanta baixa latência e alta confiabilidade, mas também uma extrema capacidade de vazão. Os KPIs críticos previstos envolvem a latência e confiabilidade, o pico de taxa de dados, a taxa de dados percebida pelo usuário, o pico na eficiência espectral, a capacidade de tráfego por área, a largura de banda, a cobertura, segurança e privacidade.

- *Massive Ultra-Reliable Low-Latency Communication* (mULC): combinando as características do mMTC e URLLC, espera-se facilidades na implementação massiva de sensores e atuadores, provendo melhorias na cobertura e eficiência espectral, sem abrir mão das vantagens vistas em URLLC.

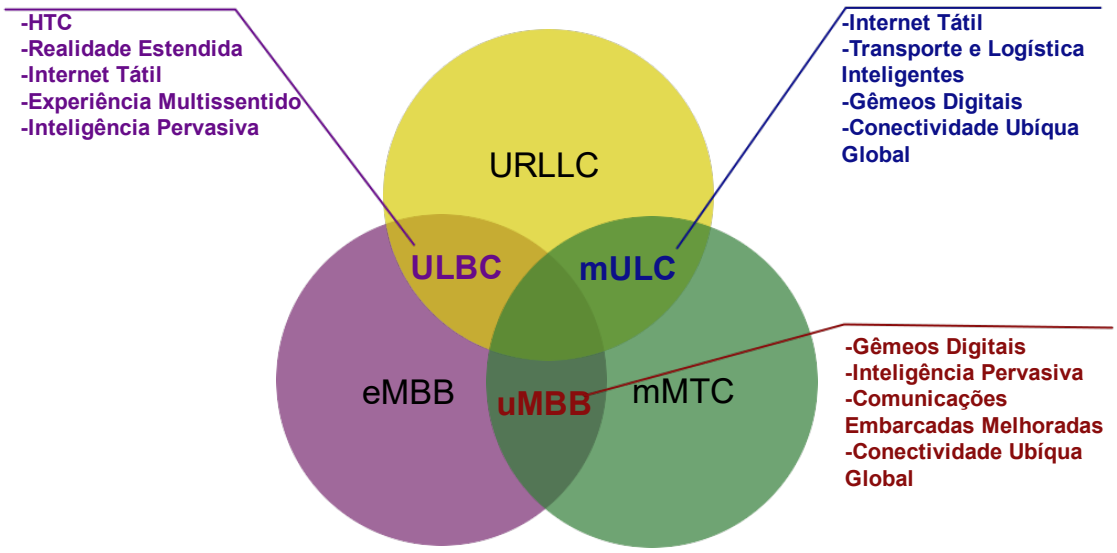


Figura 16: Cenários 5G (em preto) otimizados e conjugados para suporte aos casos de uso e aplicações 6G (adaptado de [5]).

Embora outras classes de casos de uso possam ser observados na literatura, tais como *Ubiquitous Mobile Ultra-Broadband* (uMUB), *Ultra-High-Speed with Low-Latency Communication* (uHSLLC) e *Ultra-High Data Density* (uHDD) [110], *further-eMBB* (feMBB), umMTC, *massive-uRLLC* (muRLLC), ERLLC e *Mobile Broadband Reliable Low Latency Communication* (MBRLLC) [111], entre outros apenas baseados em melhorias nos três cenários previstos para 5G, os requisitos fundamentais para habilitar os casos de uso dentro de tais classes permitem formar uma visão sobre os desafios impostos ao escalonamento de recursos nas camadas aqui destacadas.

A implementação desses novos serviços vai impor uma extrema dificuldade no gerenciamento dos recursos de rádio, em uma escala ainda mais complexa do que aquela encontrada para o 5G. Para se obter uma visão sobre tais dificuldades, podemos observá-las a partir dos requisitos (ou KPIs) fundamentais para o cumprimento dos seus objetivos, tais como:

Latência: Diversas são as dimensões que compõem a latência em sistemas celulares, sendo resultante do somatório dos atrasos gerados pela infraestrutura (*backhaul delay*), pela transmissão/retransmissão pela interface aérea e controles de erros (*transmission delay*, *over-the-air*

delay, *round-trip delay* ou *E2E delay*), pela fila (*queueing delay*), pelo processamento (*processing delay*) [105], que formam a latência do plano de usuário e a latência do plano de controle [5]. Embora as evoluções nas mais diversas áreas, como nas novas modulações, codificações/decodificações, frequências eletromagnéticas com maiores capacidades, incremento do número de antenas, a evolução na capacidade dos dispositivos e a inclusão de métodos dos mais diversos baseados em IA contribuam para a redução da latência em todas as suas dimensões, os requisitos de latência que apontam para tempos inferiores a 1 ms previsto para o 5G são especialmente desafiadores.

Mobilidade: Enquanto projeta-se para o 5G a mobilidade em termos da oferta de serviços de rede a usuários móveis com a mais alta velocidade ($\approx$ 500 Km/h), atendendo a um dado QoE mínimo, projeta-se para o 6G uma velocidade ainda maior ($\approx$ 1000 Km/h) quando considerando o uso por aviões comerciais. Entretanto, com a inclusão dos novos casos de uso, observa-se a necessidade de oferta destes serviços a uma maior variedade de dispositivos móveis, nos mais variados perfis de mobilidade, como nos casos que envolvem o transporte e logística inteligentes, em conjunto com *Vehicle-to-Everything* (V2X) e alguns serviços uMBB e mULC, bem como nos casos de uso de dispositivos para Internet tátil e realidade estendida com usuários móveis, por exemplo.

Pico da taxa de dados e taxa de dados percebida pelo usuário: Dirigidos tanto pelos avanços tecnológicos, tais como pela adoção de superfícies refletivas inteligentes (IRS) e comunicação em THz, por exemplo, quanto pelas demandas dos usuários/aplicações, com a expectativa de atingir 1 Tbps de pico da taxa de dados e 10 vezes mais velocidade percebida pelo usuário em relação ao 5G. Assim, observa-se que o emprego de métodos aprimorados de alocação de recursos PHY/MAC serão fundamentais para atingir tais KPIs, dado que dependem da ideal alocação de todos os recursos de rádio disponíveis [5]. A garantia da coexistência das aplicações 6G (Fig. 16) ocasionará em casos complexos de otimização multiobjetivos [112], dado que tendem a possuir objetivos conflitantes, em que o emprego de IA se torna fundamental.

Cobertura e capacidade de tráfego por área: Algumas das tecnologias habilitadoras do 6G devem alterar alguns conceitos já difundidos, como o conceito de cobertura, que deve ser drasticamente ampliado, saindo dos cenários terrestres 2D para uma completa conjunção 3D de sistemas terrestres, aéreos e de satélites, formatando uma cobertura ubíqua e global. O emprego conjunto de VANTs e IRSs para ampliar a cobertura de UEs móveis em redes *cell-free massive MIMO* [113] é um exemplo da complexidade à qual o gerenciamento de recursos de rádio deve ser confrontado. Adicionalmente, a ampliação da capacidade de tráfego das redes 6G, isto é, a medida de tráfego que a rede é capaz de acomodar por unidade de área, também é igualmente complexo nestes novos cenários de aplicação.

Eficiência espectral e energética: Grandes melhorias também são aguardadas em termos de eficiência espectral e energética, em que se projeta que as redes 6G devem ultrapassar em 3 vezes o que é projetado para as redes 5G em eficiência espectral, e de 10 a 100 vezes quando se trata da eficiência energética [5], sendo a alocação de recursos de forma otimizada por IA fundamental nestas tarefas de complexa otimização por métodos baseados em modelos. A cobertura ubíqua e global projetada para dar suporte aos casos de uso previstos pode representar um grande entrave na eficiência energética de todo o sistema, em que métodos dinâmicos de

alocação de recursos podem ser empregados na economia energética de pontos de acesso não utilizados [114, 115], por exemplo.

2.3.3 IA para escalonamento de recursos na camada PHY/MAC em redes 6G

Os métodos de IA apresentam vantagens importantes com relação aos métodos baseados em heurísticas para a alocação de recursos multidimensionais que podem ser sumarizadas em quatro principais características [116]: A primeira delas é a capacidade de fornecer soluções através do aprendizado baseado nos dados da rede, sem necessidade do suporte de modelos estatísticos. Dada a grande complexidade e variabilidade de cenários previstos na 6G, e os diferentes requisitos para cada um desses cenários, o aprendizado diretamente dos dados disponibilizados pela rede torna a aplicação de métodos de IA extremamente vantajosa frente a métodos baseados em modelos pré-determinados. A segunda característica é a capacidade de obter aproximações a métodos de otimização NP-difícil com uma complexidade aceitável e tomar decisões em menor espaço de tempo. A terceira característica é a menor sensibilidade aos parâmetros do sistema dado que todas as informações relacionadas a um determinado processo podem ser utilizadas para o treinamento do método de IA. Por fim a capacidade de otimização de problemas multiobjetivos sem a necessidade de dividir em subproblemas.

A alocação de recursos é usualmente baseada em modelos e sua efetividade depende diretamente da acurácia na representação do seu comportamento. Com a evolução das redes móveis, os modelos matemáticos de representação desses sistemas crescem em complexidade. Um exemplo do aumento de complexidade das redes móveis é o modelo de escalonamento dos recursos de rádio que evoluiu de um modelo básico de filas para modelos sofisticados baseados em processos de decisão de Markov, habilitando o suporte a múltiplos tipos de serviços de dados com diferentes requisitos de taxa e latência [116]. Apesar da evolução, existem problemas complexos que não podem ser descritos com precisão por modelos estatísticos, um exemplo é o caso da mitigação de interferência que é afetada por um grande número de detalhes técnicos subjacentes como a distribuição de tráfego e mobilidade de usuários. Além das limitações adquiridas da abstração da descrição de modelos, outro problema é a alta complexidade da resolução de problemas e o custo computacional para solucioná-los em tempo hábil.

Uma das alternativas para a solução de problemas de alocação de recursos utilizando métodos de otimização com complexidade NP-difícil é a utilização de métodos de IA como aproximações dos métodos de otimização, de forma a reduzir o custo computacional para a resolução do problema. Um exemplo de uso de IA para esse fim é apresentado em [108] onde uma rede neural é treinada para aproximar o método ótimo de alocação de recursos de rádio em uma rede móvel que utiliza fatiamento da rede para diferenciar os serviços fornecidos aos dispositivos conectados. O método utilizando a IA obteve resultados próximos aos obtidos com o método de otimização e com uma complexidade computacional menor, tornando o método mais apropriado para uma implementação real devido ao menor tempo necessário para a tomada de decisão do escalonador de recursos de rádio.

Um outro exemplo da utilização de métodos de IA para alocação de recursos de rádio e também de potência é apresentado em [117], onde uma estrutura de cascata de redes neurais profundas é utilizada para aproximar um método de otimização de alocação de recursos de rádio para diversos requisitos de qualidade de serviço. A cascata é composta por duas redes neurais distintas, onde a primeira rede tem a função de alocar os recursos de rádio dentre os usuários ativos, e a segunda têm a função de alocar a potência de transmissão necessária para que os dispositivos atendam aos seus requisitos. A rede neural é treinada em diferentes condições de

cenário possibilitando a sua atuação em diferentes ambientes, mas sem a eficácia mantida nos resultados treinados. Com o intuito de melhorar o desempenho em cenários desconhecidos, é explorado a técnica de transferência de aprendizado, onde as primeiras camadas das redes neurais têm os seus pesos mantidos e o treinamento é executado alterando o peso das camadas remanescentes. Dessa forma a rede têm um período de aprendizado sob as novas condições de ambiente lhe possibilitando um melhor desempenho.

Ambos os trabalhos [108, 117] utilizam redes neurais profundas com aprendizado supervisionado, no qual a rede neural é utilizada como aproximação de um método de otimização com o intuito de diminuir a complexidade computacional e o tempo necessário para o processo de alocação de recursos. A Figura 17 representa uma estrutura geral para rede neural completamente conectada utilizada para aproximação de métodos de otimização. A entrada da rede neural corresponde a informações obtidas da rede como o canal dos dispositivos, interferência entre células, nível de ocupação dos *buffers* e outras informações pertencentes à rede de acesso e dispositivos conectados a rede em geral. A rede neural pode possuir uma ou mais camadas ocultas, dependendo da complexidade do problema de otimização sendo abordado. As saídas da rede neural correspondem a aproximação ao método de otimização, podendo representar a escolha de alocação de dispositivos em recursos de rádio, potência, feixe de transmissão ou a adaptação de parâmetros da rede como o esquema de código e modulação. Para o treinamento da rede neural, normalmente são simulados cenários de rede móvel onde as informações da rede e as respostas do método de otimização são armazenados, e posteriormente a rede neural é treinada utilizando essas informações e a saída do método de otimização é utilizada como base para a saída correta da rede neural. Por fim a eficácia da utilização de métodos supervisionados para alocação de recursos na camada PHY/MAC está diretamente relacionada a eficiência na representação do cenário da rede móvel na simulação baseada em modelo.

Como contraponto aos métodos baseados em modelos, os métodos baseados em dados oferecem uma alternativa mais apropriada a alta diversidade de cenários previstas para o 6G e a alocação de recursos das camadas PHY/MAC de acordo com os requisitos de aplicações variadas, dada as limitações de elaboração de modelos que possam descrever esses cenários considerando toda a sua complexidade. Em [118] são exploradas a utilização de técnicas de aprendizado profundo através da utilização de modelos baseados em dados para a predição de mobilidade, escalonamento de recursos de rádio e associação de usuários considerando dispositivos móveis. Os resultados obtidos para as três funcionalidades utilizando o método baseado em dados foram comparados com métodos baseados em modelos, onde percebeu-se a vantagem do primeiro através da obtenção de melhores desempenhos além de maior capacidade de generalização a novos cenários.

O uso de técnicas de aprendizado supervisionado podem ser utilizados em conjunto com métodos baseados em dados, mas isso requer a rotulação da informação obtida através dos dados e a definição de resultados esperados para os métodos de otimização. Para fazer essa rotulação é então necessário utilizar algum método de otimização, o que acaba limitando a performance do modelo à otimização utilizada [119]. Dessa forma, apesar de aplicável, as técnicas de aprendizado supervisionado não são as mais recomendadas para as tarefas de alocação de recursos nas camadas PHY/MAC em modelos baseados em dados [120]. As técnicas de aprendizado baseado em reforço por outro lado possuem um sistema baseado em recompensas, no qual não é necessário rotular uma resposta correta, mas escolher ações que maximizem as recompensas obtidas [121].

A Figura 18 ilustra o funcionamento da técnica de aprendizado por reforço aplicada a alocação de recursos nas camadas PHY/MAC. O ambiente da rede móvel fornece a informação

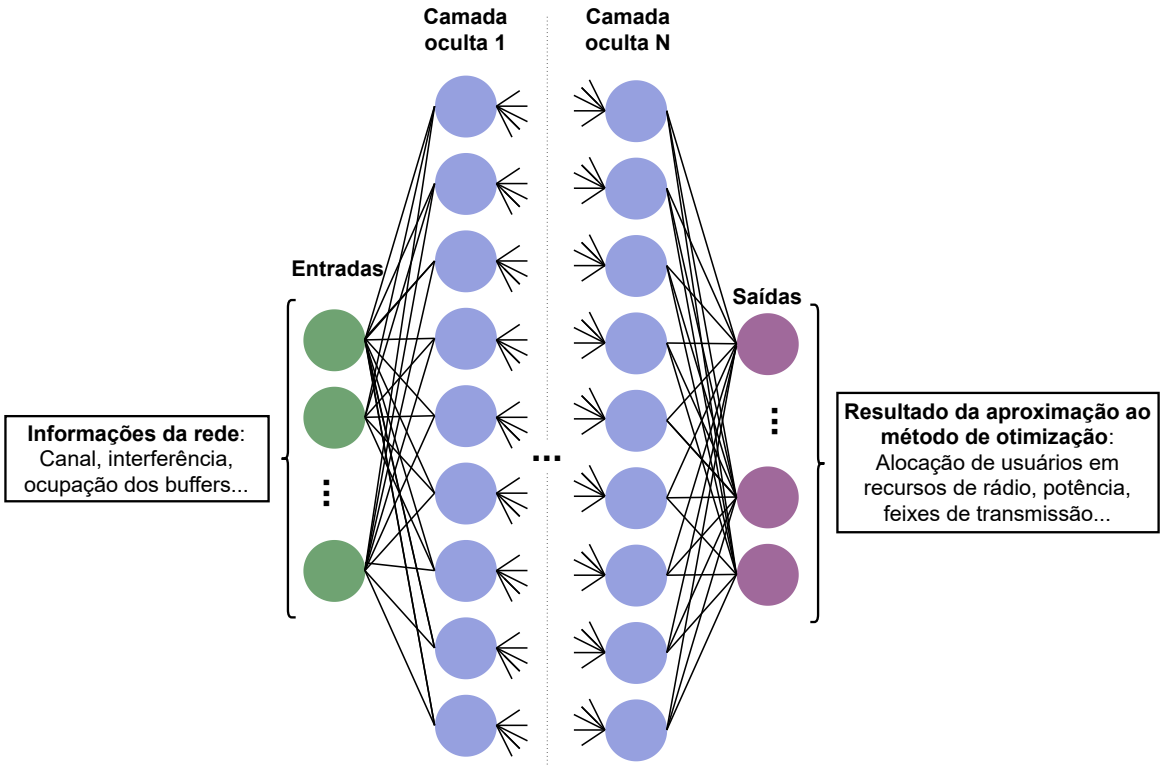


Figura 17: Estrutura de rede neural completamente conectada usualmente utilizada para aprendizado supervisionado de alocação de recursos da camada PHY/MAC.

do estado S da rede móvel para o agente, esse estado pode conter qualquer informação que caracterize um momento do sistema como a razão sinal-ruído dos dispositivos, taxa de *download*/*upload* e níveis de interferência. A ação A representa a escolha realizada pelo agente de aprendizado por reforço com relação a atividade sendo executada, por exemplo, a ação de um agente treinado para alocação de recursos de rádio pode ser a escolha de um dispositivo para ocupar uma determinada quantidade de recursos de rádio. A recompensa é um valor que indica quão bem sucedida foi a última ação A aplicada no ambiente da rede móvel, levando o ambiente para um novo estado S. Por exemplo, a recompensa de um agente que busca minimizar a interferência intra-celular através da alocação de potência para os dispositivos poderia ser considerada inversamente proporcional a interferência medida, de forma a incentivar o agente a sempre diminuir os níveis de interferência.

Alguns exemplos do uso de técnicas de aprendizado por reforço aplicadas a alocação de recursos da camada PHY/MAC são a mitigação de interferência [122], escalonamento de recursos de rádio [123] e a adaptação de *links* em comunicações de veículo para veículo [124].

Apesar das vantagens dos métodos baseados em dados, uma das grandes dificuldades encontradas é a disponibilidade de base de dados para treinamento de agentes para execução de funções da camada PHY/MAC e a validação de dados, dado que grande parte dos trabalhos consideram diferentes tipos de informações obtidas da rede e sob diferentes condições, além de que o treinamento de modelos de IA em redes móveis em ambientes de produção não são propícios devido a usual demora para que a agente possa convergir e se adaptar a novas condições do ambiente da rede móvel [125].

A utilização de gêmeos digitais é uma importante alternativa para criação de base de dados mais consistentes e prover maior segurança e flexibilidade para explorar novas políticas para

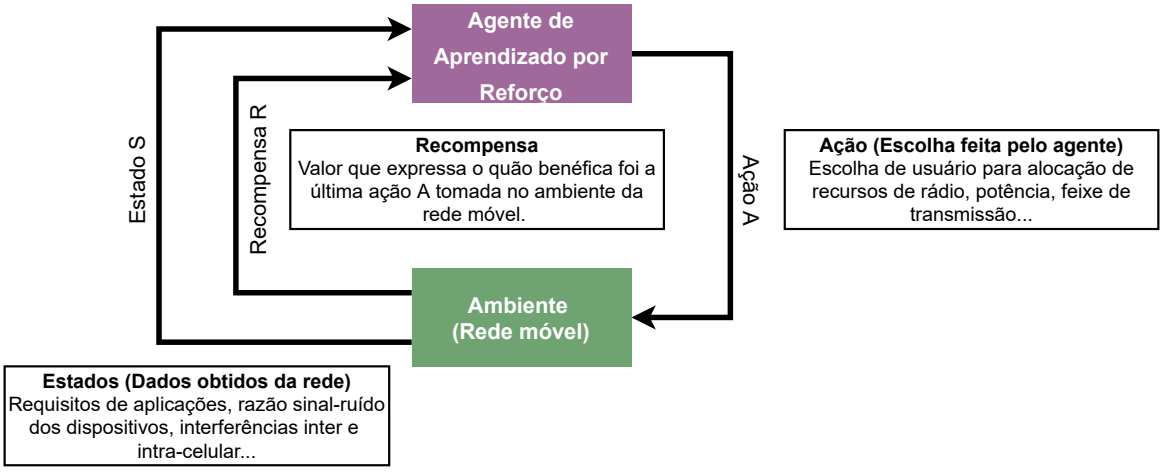


Figura 18: Estrutura da técnica de aprendizado por reforço utilizada para alocação de recursos da camada PHY/MAC.

as funções da PHY/MAC [126]. Os gemêos digitais consistem na criação de uma cópia digital do ambiente real da rede móvel de forma a refletir sua dinâmica, características, componentes críticos e o seu ciclo de funcionamento [127]. Com a utilização de gêmeos digitais para cenários de redes móveis focando nas funções das camadas PHY/MAC, podem ser obtidas informações mais realistas da rede móvel real do que a proposta por modelos estatísticos através da utilização de modelos 3D criados utilizando fotogrametria para representar o ambiente digitalmente [128]. Dessa forma cada ambiente pode ser devidamente explorado de acordo com as suas características e permitindo explorar a máxima performance dos métodos de IA.

A Figura 19 representa a utilização de gêmeos digitais para treinamento de um agente de IA para controle da rede de acesso. O cenário real é a rede móvel com todos os dispositivos conectados a ela, possuindo uma dinâmica de funcionamento específica para as condições do ambiente como o nível de interferência, razão sinal-ruído, e diferentes dispositivos e aplicações utilizando a rede. As informações e características obtidas da rede real são usadas como base para criação do gêmeo digital em uma plataforma de simulação ou emulação, dessa forma o gêmeo digital deve ser o mais fiel possível ao cenário real. Um agente de IA utilizando técnicas de aprendizado por reforço é responsável pela alocação de recursos da camada PHY/MAC da rede simulada/emulada do gêmeo digital, de forma a escolher as ações que melhor atendem os requisitos do cenário. O uso de gêmeos digitais garantem uma maior segurança para explorar novas políticas de funcionamento para o agente sem comprometer a rede real, quando o agente IA do gêmeo digital converge para uma solução estável e satisfatória para o cenário, é realizada uma atualização da política do agente IA do cenário real. Tanto a atualização das informações do cenário de rede real no gêmeo digital quanto a atualização de políticas do agente IA do gêmeo digital para o cenário real são processos periódicos e constantes, com o intuito de garantir o melhor funcionamento da rede móvel.

Sensores podem ser utilizados para obter informações importantes do cenário da rede móvel de forma a aperfeiçoar o funcionamento de funções de rede da camada PHY/MAC, como o uso de informações de sensores LIDAR para suporte no processo escolha de feixes de transmissão em redes móveis com ondas milimétricas como apresentado em [129]. Nesse artigo as informações obtidas do sensor LIDAR são utilizadas para o treinamento de uma rede neural profunda para predizer os feixes em visada direta ou indireta e também para classificar os melhores feixes disponíveis. Para a construção do cenário de simulação realístico, foram combinadas três

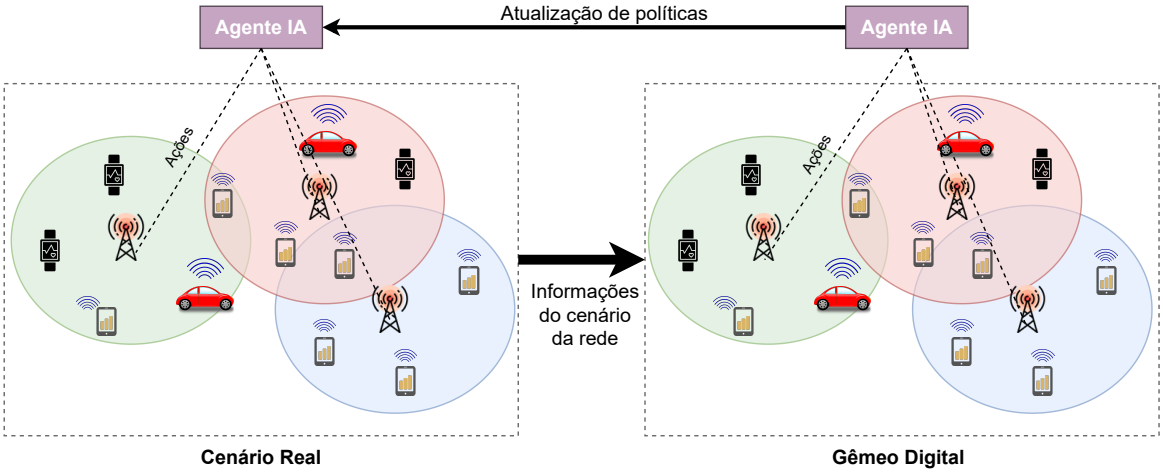


Figura 19: Estrutura de Gêmeo Digital aplicada a rede de acesso utilizando IA para alocação de recursos da camada PHY/MAC.

ferramentas para simulação de tráfego de carros usando a ferramenta SUMO [130], LIDAR usando a ferramenta BlenSor [131] e os traçados de raios utilizando a ferramenta Remcom Wireless InSite.

Informações das aplicações também podem ser utilizadas para otimização do processo de alocação de recursos das camadas PHY/MAC, em [132] propõe uma arquitetura para escalonar recursos de rádio em aplicações de realidade virtual interativas. As informações de localização e ângulo de visão do usuário são utilizadas para definir grupos de transmissão de informação e otimizar a alocação de recursos de rádio efetuada por um método de aprendizado por reforço.

2.4 Sensoriamento espectral via aprendizado de máquina

Lucas dos Santos Costa, Sheila Cássia da Silva Janota, Felipe Augusto P. de Figueiredo, Rausley Adriano Amaral de Souza, Luciano Mendes
lucass@inatel.br, sheila.cassia@mtel.inatel.br, felipe.figueiredo@inatel.br, rausley@inatel.br, luciano@inatel.br

Esta subseção tem por objetivo relacionar os esforços mais recentes que aplicam IA na tarefa do sensoriamento espectral em rádios cognitivos. Após uma breve revisão sobre o tema sensoriamento espectral, tem-se uma descrição do estado da arte em IA em suas aplicações nos problemas de sensoriamento espectral, especialmente aquelas com potencial de aplicação nos sistemas 6G.

2.4.1 Sensoriamento espectral

O espectro radioelétrico é a faixa do espectro eletromagnético destinada à radiocomunicação. É um bem público, limitado e dividido em faixas bem definidas a fim de prover diversos serviços e atividades de telecomunicações tendo em vista a interoperabilidade entre um serviço e outro. O recurso é gerenciado globalmente pela ITU [133], agência da Organização das Nações Unidas (ONU) [134], que padroniza e regula as telecomunicações internacionais. Entretanto, o uso do espectro radioelétrico é também administrado separadamente em cada país. No Brasil, por exemplo, a Agência Nacional de Telecomunicações (ANATEL) [135] é o órgão responsável pela administração das faixas de frequência, definidas em tratados e acordos internacionais e

aprovadas pela ITU, com base na emissão anual do plano de atribuição, destinação e distribuição de faixas de frequência.

A rápida evolução dos sistemas de comunicação viabilizou um aumento explosivo de demanda por novos serviços de telecomunicações nos últimos 50 anos. Esse aumento foi considerado o principal responsável pelo atual cenário de escassez espectral. O problema da falta de faixas espectrais para atender a esta crescente demanda tornou-se ainda mais evidente com os adventos das redes 5G e da IoT, pois para o bom funcionamento de ambas é imperativo que haja interoperabilidade entre uma quantidade enorme de dispositivos transceptores [136]. Adicionalmente, pode-se prever que a implementação das redes 6G demandará medidas eficazes que possam mitigar o problema da escassez de recursos espectrais tendo em vista os cenários previstos para as redes 6G em relação ao 5G [137].

Apesar da escassez espectral, algumas pesquisas mostram que o espectro radioelétrico se encontra contraditoriamente subutilizado, visto que faixas espectrais ociosas podem ser frequentemente encontradas em determinadas localizações, em intervalos de tempo específicos [138–140]. A busca de soluções para estes cenários revelou que uma mudança na atual política de alocação ao espectro poderia mitigar tanto o problema da escassez quanto o problema da subutilização espectral. A partir daí, uma nova política de alocação dinâmica de banda começou a ser considerada pelos órgãos reguladores. O direito de uso de determinada faixa espectral é adquirido por meio de concessão na política de acesso fixo em vigor, e a concessão é dada exclusivamente a usuários contratantes, conhecidos como *Primary User*s (PUs), durante todo o período contratado. Já a nova política de acesso dinâmico, no entanto, baseou-se no compartilhamento oportunista do espectro com usuários não contratantes, também chamados de *Secondary User*s (SUs). Este novo modelo de acesso admite que transmissões secundárias possam ser feitas utilizando canais PUs desde que essas transmissões sejam feitas de maneira inteligente (sem prejudicar as transmissões primárias). Por isso, leva-se em conta duas principais formas de atuação para a rede secundária, que são: i) transmissões secundárias podem ser feitas simultaneamente às primárias desde que as interferências causadas pela rede secundária não ultrapassem o patamar de interferência máximo preestabelecido pela rede primária; ou ii) transmissões secundárias podem ser feitas sem restrições de interferência com as primárias desde que canais primários possam ser identificados como ociosos.

Nesse contexto surgiu o conceito de sensoriamento espectral como uma possível solução para o problema da escassez e subutilização do espectro. O sensoriamento espectral pode ser entendido como uma inteligência incorporada a um dispositivo transceptor de tal forma que o capacite a observar o espectro e detectar oportunidades de se fazer transmissões secundárias via canais primários ociosos. Dado que esta é a função-chave de um *Cognitive Radio* (CR), um CR é, no contexto do acesso secundário a canais primários, um SU inteligente, capaz de entender o ambiente em que está inserido e prover acesso oportunista às faixas primárias do espectro temporariamente inativas. A função primária de um CR, ou SU, é sensoriar determinada faixa do espectro e decidir sobre seu estado de ocupação. Comumente, adotam-se apenas duas hipóteses para a decisão de ocupação: canal desocupado, em que a decisão é representada pela hipótese $\mathcal{H}_0$, e canal ocupado, em que a decisão é representada pela hipótese alternativa $\mathcal{H}_1$.

A decisão de ocupação pode ser tomada por meio de apenas um CR, no sensoriamento espectral não cooperativo, ou por meio de um conjunto de CRs no sensoriamento espectral cooperativo. Porém o desempenho do sistema no modelo não cooperativo pode ser severamente degradado caso o rádio em questão esteja operando, por exemplo, sob desvanecimento multipercurso, sombreamento, terminal primário escondido ou uma combinação destes fenômenos [141]. Todavia, o modelo cooperativo é capaz de melhorar o desempenho dessa decisão. A maior

acurácia da decisão neste caso se deve à diversidade espacial promovida por CRs em diferentes posições geográficas e, portanto, possivelmente em melhores condições de sensoriamento. A Figura 20 ilustra um cenário em que se tem a coexistência de uma rede primária e uma rede secundária de CRs, cada uma demarcada por sua respectiva área de cobertura. A rede primária contém um transmissor primário, denotado por Tx, e dois receptores primários: Rx_1 e Rx_2. Já a rede secundária está equipada com uma BS e quatro SUs: CR_1, CR_2, CR_3 e CR_4.

Apesar do CR_3 estar próximo ao transmissor primário, em um dado momento os sinais refletidos pelo obstáculo que impede a chegada do sinal primário ao CR_2 pode fazer com que a decisão de ocupação tomada pelo CR_3 seja em favor de $\mathcal{H}_1$ quando na verdade o transmissor está inativo, ou seja, sob $\mathcal{H}_0$, por exemplo. É claro que a decisão do CR_2 também será incerta, pois esse rádio se encontra em uma situação de sombreamento. Assim como o CR_3, o CR_1 também pode estar sob desvanecimento multipercurso em um dado momento devido às reflexões de sinal geradas pelo mesmo obstáculo que bloqueia a chegada do sinal primário ao CR_2. Já o CR_4 está muito distante do transmissor primário e, portanto, fora da área de cobertura da rede primária. Consequentemente, este rádio está operando em uma situação de terminal primário escondido e sua decisão também não será confiável. Se, por exemplo, o CR_4 decidir em favor de $\mathcal{H}_0$ quando o canal estiver sob a hipótese $\mathcal{H}_1$, suas transmissões poderão causar interferências prejudiciais ao receptor primário Rx_2. Com base no que foi descrito sobre a Figura 20, nota-se que nenhum CR é capaz de tomar decisões de ocupação sempre acuradas a todo momento de forma isolada. Por este motivo costuma-se adotar o sensoriamento espectral cooperativo, uma vez que é menos provável que mais de um CR esteja operando sob condições ruins de sensoriamento ao mesmo tempo [142].

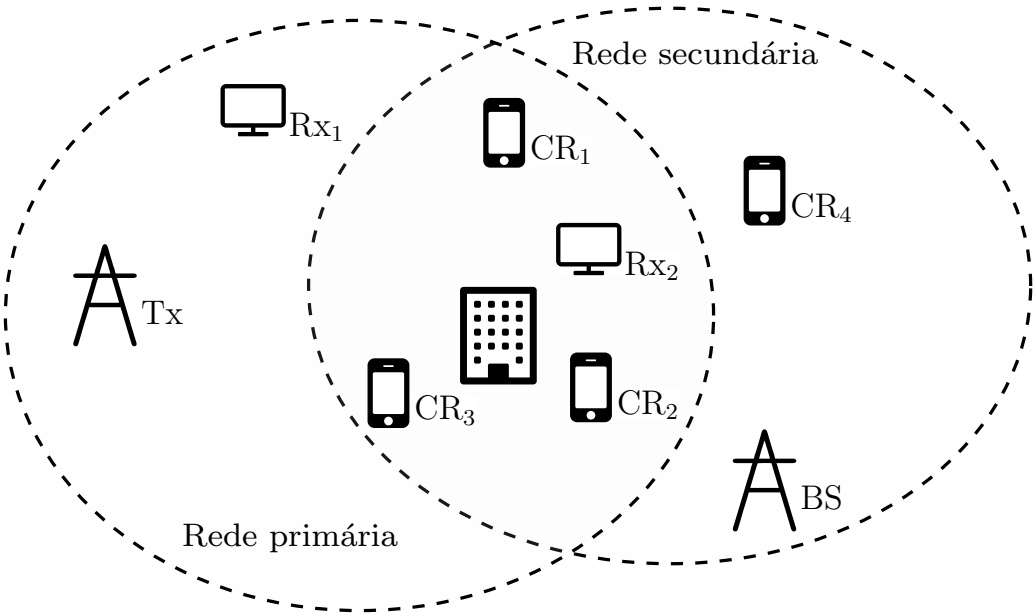


Figura 20: Ilustração de um possível cenário de sensoriamento espectral.

O sensoriamento espectral cooperativo pode ser centralizado, assistido por retransmissão, ou distribuído [141], conforme descrito na Figura 21. No modelo centralizado, Figura 21(a), os CRs realizam o sensoriamento em determinada faixa espectral, coletam amostras do sinal recebido, e compartilham informações de sensoriamento via canais de controle com um nó de processamento geral chamado *Fusion Center* (FC). O FC faz a combinação dos dados recebidos por meio de uma operação chamada fusão, toma a decisão de ocupação, e também usa os canais de controle

para informar aos CRs sobre esta decisão e assim gerenciar a atividade de rede secundária. O sensoriamento espectral cooperativo assistido por retransmissão, Figura 21(b), é semelhante ao modelo centralizado. A diferença é que neste caso os CRs podem compartilhar as informações de sensoriamento caso seus respectivos canais de controle não estejam em boas condições de transmissão ao FC. Ainda assim o FC recebe estas informações por meio dos canais de controle dos CRs em melhores condições de transmissão, processa todas as informações recebidas, toma a decisão de ocupação e controla a atividade da rede secundária. Já no modelo distribuído, Figura 21(c), não existe um FC.

Neste caso, as informações de sensoriamento são compartilhadas exclusivamente entre os CRs participantes da cooperação. Um dos CRs pode assumir o papel do FC e a decisão de ocupação é obtida por meio de interações entre os rádios até que uma decisão conjunta definitiva seja alcançada. Observe que na Figura 21(b) os canais de controle dos CRs 3 e 4 não se encontram em boas condições de transmitir dados ao FC, pois suas transmissões não podem alcançá-lo. Assim as informações de sensoriamento destes rádios podem ser enviadas ao FC tanto pelo CR_1 quanto pelo CR_2 cujos canais de controle se encontram em boas condições para o envio de dados.

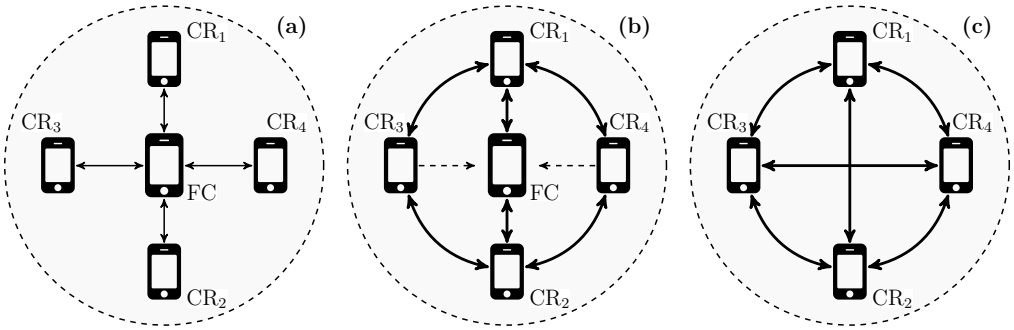


Figura 21: Ilustração do sensoriamento espectral cooperativo centralizado (a), assistido por retransmissão (b) e distribuído (c).

O tipo de informação compartilhada entre CRs e/ou FC define dois esquemas de fusão: a fusão de dados, também conhecida como *Soft Decision* (SD), e a Fusão de Decisões, também chamada *Hard Decision* (HD). Na fusão de dados tais informações podem ser as próprias amostras de sensoriamento ou quantidades derivadas dessas amostras (p. ex., medidas de energia). Por outro lado, na fusão de decisões, as decisões locais, tomadas individualmente por cada CR, são as informações a serem compartilhadas. Na fusão de decisões, a decisão final de ocupação, tomada pelo FC, é obtida por meio da combinação/fusão das decisões individuais recebidas. Sua forma geral de combinação pode ser descrita pela regra k-em-m, em que k representa o número de decisões recebidas em favor da hipótese $\mathcal{H}_1$; e m representa o número total de decisões recebidas, ou seja, o número de CRs em cooperação. Na regra k-em-m o FC toma a decisão final em favor da hipótese $\mathcal{H}_1$ se pelo menos k das m decisões individuais recebidas também forem em favor de $\mathcal{H}_1$. Com $k = 1$ tem-se a conhecida regra de decisão "OU", em que o FC decide em favor de $\mathcal{H}_1$ se ao menos uma das decisões individuais recebidas também for em favor de $\mathcal{H}_1$. Com $k = m$ tem-se a regra de decisão "E". Uma decisão final em favor de $\mathcal{H}_1$ é adotada pelo FC apenas se todas as decisões individuais recebidas também forem em favor de $\mathcal{H}_1$ neste caso. Por fim, adotando-se $k = \lfloor m/2 + 1 \rfloor$ tem-se a chamada regra de decisão majoritária (MAJ). Nessa regra, uma decisão final em favor de $\mathcal{H}_1$ é tomada pelo FC apenas se mais da metade das decisões individuais recebidas também forem em favor de

$\mathcal{H}_1$, em que o símbolo "*floor*", denotado por $\lfloor x \rfloor$, representa o inteiro menor ou igual a x. Em quaisquer dessas regras toma-se uma decisão final em favor de $\mathcal{H}_0$ caso a respectiva condição não seja atendida.

Independentemente do esquema de fusão adotado, qualquer decisão de ocupação é obtida por meio da aplicação de uma estatística de teste, T. A decisão é alcançada comparando-se o resultado da estatística de teste com um limiar de decisão predefinido, γ. Se $T > \gamma$, decidi-se em favor de $\mathcal{H}_1$: ou em favor de $\mathcal{H}_0$ caso contrário. Os desempenhos das decisões são comumente medidos em função da probabilidade de falso alarme, $P_{\text{fa}} = \Pr\{\text{decisão} = \mathcal{H}_1|\mathcal{H}_0\} = \Pr\{T > \gamma|\mathcal{H}_0\}$, e da probabilidade de detecção, $P_{\text{d}} = \Pr\{\text{decisão} = \mathcal{H}_1|\mathcal{H}_1\} = \Pr\{T > \gamma|\mathcal{H}_1\}$. A probabilidade de falso alarme é a probabilidade de haver uma decisão em favor de $\mathcal{H}_1$ (canal ocupado) dado que o canal sob sensoriamento está sob hipótese $\mathcal{H}_0$ (canal desocupado). Alternativamente, P_{fa} também pode ser entendida como sendo a probabilidade da estatística de teste, T, ser maior que o limiar de decisão, γ, dado que o canal sob sensoriamento está sob hipótese $\mathcal{H}_0$. A probabilidade de detecção, por sua vez, é a probabilidade de haver uma decisão em favor de $\mathcal{H}_1$ dado que o canal sob sensoriamento está de fato sob esta hipótese, o que é equivalente à probabilidade de se ter $T > \gamma$ dado $\mathcal{H}_1$.

As análises de desempenho são geralmente feitas de forma gráfica por meio da curva *Receiver Operating Characteristic* (ROC), ou pela área abaixo dessa curva, aqui denotada simplesmente como *Area Under the Curve* (AUC). As análises de desempenho via AUCs como alternativa às ROCs são particularmente úteis quando é necessário analisar diversas ROCs com desempenhos muito próximos e/ou que se sobrepõem em algum ponto do gráfico. Uma ROC é obtida traçando-se pontos de P_{fa} versus P_{d} com a variação do limiar de decisão, γ, em um intervalo de valores. A Figura 22 mostra exemplos possíveis de desempenhos em termos de ROCs e AUCs para três diferentes testes estatísticos: T_1, T_2, e T_3. Com base nos resultados pode-se concluir que o teste estatístico T_1 possui menor poder estatístico de detecção, pois para qualquer valor de P_{fa} o respectivo valor de P_{d} é menor ou igual ao valor obtido pelos testes estatísticos T_2 e T_3. Por exemplo, para $P_{\text{fa}} = 0{,}1$ o respectivo valor de P_{d} no teste estatístico T_1 é $P_{\text{d}} = 0{,}8$, enquanto que $P_{\text{d}} = 0{,}9$ nos testes estatísticos T_2 e T_3. Já os desempenhos dos testes estatísticos T_2 e T_3 se cruzam em $P_{\text{fa}} = 0{,}1$ e por isso ambos possuem o mesmo desempenho nesse ponto. Veja que para $0 \lessapprox P_{\text{fa}} < 0{,}1$ o teste estatístico T_2 possui melhor desempenho que o T_3. Porém, o contrário acontece para $0{,}1 < P_{\text{fa}} \lessapprox 0{,}8$, ou seja, o teste estatístico T_3 possui melhor desempenho que o T_2 neste caso. Nestes casos o uso das AUCs pode ser útil. Observando os resultados em termos de AUCs, pode-se dizer que o teste estatístico T_3 possui poder estatístico de detecção ligeiramente superior ao T_2, já que seu respectivo valor de AUC é ligeiramente maior.

As técnicas de sensoriamento podem ser classificadas como técnicas de banda larga ou técnicas de banda estreita conforme a largura de banda do canal a ser sensoriado [141]. As técnicas de banda estreita [143–145] são capazes de detectar a presença do sinal primário em apenas uma banda por vez, e a ocupação secundária também é feita por meio de apenas um SU por vez [146]. Já as técnicas de banda larga [144, 145, 147–149] podem detectar a presença de sinais primários em múltiplas bandas [146] de forma simultânea ou sequencial e a ocupação secundária pode ser feita por múltiplos SUs [146] conjuntamente.

A Tabela 4 mostra alguns exemplos de técnicas de sensoriamento de banda larga e de banda estreita bem discutidos na literatura. Ressalta-se que a técnica de detecção por energia, ou *Energy Detection* (ED), e a técnica de detecção por autovalores, ou *Eigenvalue Detection* (EVD), podem ser utilizadas tanto no sensoriamento de banda larga como no de banda estreita, sendo a EVD uma das mais recentes e promissoras técnicas de sensoriamento [144]. Algumas

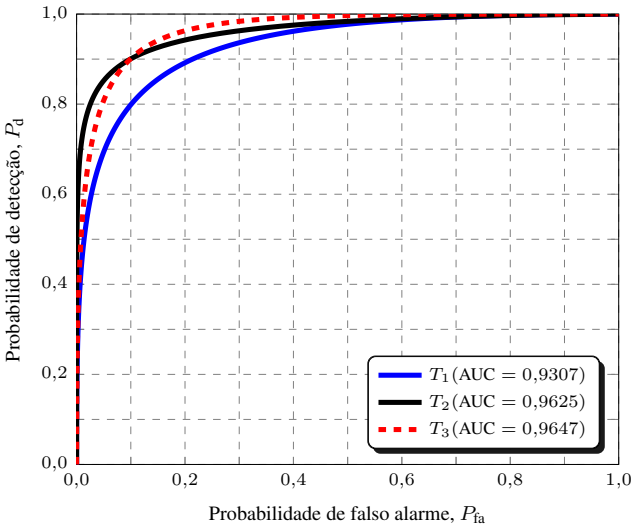


Figura 22: ROCs/AUCs hipotéticas para os testes estatísticos T_1, T_2 e T_3.

Tabela 4: Classificação de técnicas de sensoriamento conforme a largura de banda do canal.

Sensoriamento espectral	
Técnicas de banda Larga [144, 145, 147–149]	Técnicas de banda estreita [143–146]
Detecção por energia	Detecção por energia
Detecção por transformada de *wavelet*	Detecção por filtro casado
Detecção pela técnica *compressed sensing*	Detecção por ciclo-estacionariedade
Detecção por autovalores	Detecção por autovalores

dessas técnicas foram desenvolvidas pressupondo algum conhecimento *a priori* relacionado às características do sinal primário e/ou do ruído de recepção, o ruído aditivo Gaussiano branco, ou *Additive White Gaussian Noise* (AWGN) [150]. Por exemplo, a detecção por propriedades ciclo-estacionárias (*cyclostationary feacture detection*) depende do conhecimento das características cíclicas das frequências do sinal primário, e a detecção por filtro casado (*matched filter detection*) requer o conhecimento da forma de onda do sinal primário e da resposta ao impulso do canal de sensoriamento. Ambas são conhecidas como técnicas de sensoriamento não-cegas [151] por este motivo. As técnicas que operam sem conhecimentos *a priori* sobre o sinal primário, mas necessitam de informações sobre a potência do ruído, são conhecidas como semicegas, como a ED e a detecção pelo máximo autovalor, *Maximum Eigenvalue Detection* (MED) [144], da matriz de covariância do sinal recebido durante o sensoriamento, por exemplo. Já as técnicas que operam sem qualquer conhecimento *a priori* sobre o sinal primário ou ruído são chamadas de cegas. A detecção pela transformada de *wavelet* e alguns tipos de detecção por autovalores [144] se encaixam nesta categoria. As técnicas cegas têm grande importância para o sensoriamento espectral porque o conhecimento *a priori* sobre as características do sinal primário ou ruído podem ser difíceis ou até mesmo impossível de serem obtidas na prática. Além do mais, uma das premissas do sensoriamento espectral é que os SUs deveriam ser capazes de detectar oportunidades de ocupação secundária aos canais primários de forma autônoma, sem dispor de qualquer informação advinda da rede primária [152, 153].

O sensoriamento espectral baseado em CRs é uma possível solução para os problemas de subutilização e escassez espectral, e a escassez é uma consequência das evoluções (p. ex., do 2G

ao 4G) dos sistemas de comunicação sem fio. Essas evoluções não só implicam no aumento do número de dados transmitidos, mas também na transmissão de dados heterogêneos, como sinais de voz, vídeos, textos e imagens, por exemplo. Como se não bastassem, os adventos do 5G e da IoT, e as previsões para as futuras redes 6G, revelam a necessidade de um aumento expressivo para a demanda de volume de dados, bem como a necessidade de transmissão de dados ainda mais diversificados. Por isso, já existem estudos apontando que as técnicas convencionais de detecção talvez não sejam capazes de operarem satisfatoriamente nesses cenários devido ao natural aumento da complexidade e heterogeneidade dos sistemas [154]. Estes são dois motivos importantes que têm impulsionado as pesquisas sobre a utilização de técnicas baseadas em ML nos sistemas de sensoriamento espectral.

2.4.2 Algoritmos de aprendizado de máquina usados no sensoriamento espectral

Com a intensificação das pesquisas sobre IA nos últimos anos, as técnicas baseadas em ML têm sido cada vez mais estudadas no contexto de sensoriamento espectral [153] como uma alternativa às técnicas convencionais de detecção: como as citadas na Tabela 4, por exemplo. Uma importante vantagem da detecção baseada em ML é a capacidade da rede secundária de aprender sobre o ambiente de sensoriamento de forma implícita, com a experiência, identificando padrões nas informações de sensoriamento sob as hipóteses $\mathcal{H}_0$ e $\mathcal{H}_1$ sem de fato necessitar de conhecimentos *a priory* sobre o sinal primário ou ruído. Por consequência disso, o sensoriamento espectral via ML pode ser visto como sendo uma tarefa de classificação e decisão da rede secundária na qual a rede secundária deve identificar características particulares nas informações de sensoriamento e classificá-las como estando sob as hipóteses $\mathcal{H}_1$ ou $\mathcal{H}_0$ a fim de decidir sobre presença ou ausência do sinal primário no canal sensoriado [155], respectivamente. O processo de aprendizado assume o papel primário neste tipo de detecção, pois as ações da rede secundária são influenciadas por informações de sensoriamentos anteriores [153] utilizadas como informações de treinamento.

A literatura de sensoriamento espectral tem explorado predominantemente dois tipos particulares de algoritmos baseados em ML [153]: i) os algoritmos de aprendizado supervisionado (*supervised learning*) e ii) os algoritmos de aprendizado não supervisionado (*unsupervised learning*). Os algoritmos de aprendizado supervisionado necessitam ser alimentados com exemplos (informações de treinamento) e rótulos. Os rótulos são informações contendo respostas desejadas: por exemplo informações sobre as hipóteses $\mathcal{H}_0$ ou $\mathcal{H}_1$ a qual pertence cada exemplo de treinamento. O objetivo destes algoritmos é treinar um modelo de aprendizado alimentado com as respostas corretas e então utilizá-lo na identificação dos padrões das informações de sensoriamento em tempo real. Por outro lado, nos algoritmos de aprendizado não supervisionado o treinamento é feito sem os rótulos e utilizando apenas os exemplos de treinamento. Ressalta-se que o algoritmo de aprendizado supervisionado pode oferecer melhor desempenho que o não supervisionado devido às informações adicionais contidas nos rótulos. No entanto, considerando as premissas do sensoriamento espectral, as informações para os rótulos podem ser difíceis ou até mesmo impossíveis de serem obtidas na prática. Em decorrência disso, os algoritmos de aprendizado não supervisionado podem ser a melhor opção dependendo de seu objetivo específico em um dado cenário de sensoriamento [153]. A seguir, apresenta-se uma revisão bibliográfica contendo diversas aplicações específicas destes algoritmos em cenários de sensoriamento espectral.

Os autores em [153] discutem sobre a implementação de vários algoritmos baseados em ML na caracterização do problema de aprendizado em CRs destacando a importância da IA

na viabilização dos sistemas de comunicação inteligentes. A atuação dos CRs é definida pelas tarefas de classificação e de decisão, e as discussões dividem as técnicas em algoritmos de aprendizado supervisionado e não supervisionado. As tarefas de decisão são responsáveis pelo estabelecimento de regras de decisão relacionadas ao estado de ocupação da rede primária e/ou políticas de sensoriamento espectral, controle de potência, ou modulação adaptativa. Já as tarefas de classificação são responsáveis pela identificação e classificação de diferentes modelos de observação. Um estudo sobre as vantagens e desvantagens das técnicas mais comuns e mais utilizadas possibilita a identificação daquela que é a melhor opção em função do método de aprendizado em contextos específicos, bem como do aprendizado de uma tarefa específica, ou característica específica desejada. Os algoritmos de aprendizado não supervisionados *Dirichlet Process Mixture Model* (DPMM) e *Reinforcement Learning* (RL) são estudados no contexto de classificação e decisão, respectivamente, e os algoritmos de aprendizado supervisionados *Support Vector Machine* (SVM) e *Artificial Neural Network* (ANN) são estudados no contexto de classificação. O estudo ainda apresenta alguns desafios relacionados ao problema de aprendizado dos CRs juntamente com possíveis estratégias de solução.

Quatro técnicas baseadas em ML são propostas em [156] em um modelo de estimação da energia contida nas amostras do sinal recebido no sensoriamento espectral cooperativo. A função dos CRs é uma tarefa de classificação na qual a rede secundária deve distinguir padrões nos sinais recebidos e classificá-los como estando sob as hipóteses $\mathcal{H}_0$ ou $\mathcal{H}_1$ a fim de decidir sobre o estado de ocupação dos canais sensoriados. Cada CR calcula o nível de energia presente nas amostras de sensoriamento, tal como é feito na técnica ED, e envia esta informação a um FC para tomada de decisão cooperativa. Foram propostas a utilização dos algoritmos de aprendizado supervisionados SVM e KNN, e os algoritmos de aprendizado não supervisionados *Gaussian Mixture Model* (GMM) e *K-Means Clustering* (KMC) na detecção de sinais primários. Os autores propuseram ainda um esquema de pesos para o algoritmo KNN no qual os pesos são obtidos em função da AUC do vetor característico formado no FC a partir dos valores de energia enviados pelos CRs. Os desempenhos são avaliados em termos do número de CRs, do tempo de treinamento, do atraso de classificação e das ROCs. Uma comparação com os desempenhos obtidos com as regras de decisão "OU" e "E", e com o algoritmo de aprendizado supervisionado *Fisher Linear Discriminant* (FLD), mostra que os esquemas propostos são capazes de alcançar melhores resultados. Especificamente, um destaque é dado aos algoritmos SVM e KMC, que possuem melhores desempenhos que os demais em termos de ROC, sendo o desempenho do algoritmo KMC ligeiramente inferior ao algoritmo SVM. Já o algoritmo KNN ponderado mostrou-se um bom candidato a ser utilizado nos sistemas de sensoriamento que necessitam de treinamento em tempo real, já que o menor tempo de treinamento é a principal vantagem desta proposta.

O teste ED é uma das técnicas de detecção mais utilizadas no sensoriamento espectral cooperativo ou não cooperativo, em canais de banda larga ou banda estreita, principalmente devido à sua simplicidade de implementação e capacidade de distinguir qualquer tipo de sinal sob ruído AWGN. O teste é implementado simplesmente comparando-se o nível de energia presente nas N amostras coletadas do sinal sensoriado com o limiar de decisão predefinido, γ; e o resultado da comparação é usado para decidir se tais amostras estão sob as hipóteses $\mathcal{H}_0$ ou $\mathcal{H}_1$. A estatística de teste do teste ED pode ser obtida simplesmente calculando-se $T_{\text{ED}} = \sum_{n=1}^{N} |y(n)|^2$, em que $|\cdot|$ representa a operação de módulo e $y(n)$ a n-ésima amostra do sinal recebido. Uma decisão de ocupação em favor de $\mathcal{H}_1$ é adotada se $T_{\text{ED}} > \gamma$. Decide-se em favor de $\mathcal{H}_0$, caso contrário. A maior desvantagem do ED, no entanto, consiste na dificuldade do estabelecimento de um limiar de decisão adequado para satisfazer um requisito de desempenho

mínimo desejado, pois o cálculo do limiar requer o conhecimento *a priori* exato da potência do ruído AWGN. Na prática, entretanto, tem-se apenas uma estimativa do valor exato da potência de ruído, e o menor erro de estimação costuma ser bastante prejudicial aos desempenhos de detecção no teste ED. Além do mais, alguns cenários de sensoriamento exigem que a estimação da potência de ruído seja feita em tempo real devido às mudanças do ambiente de propagação e movimentação dos CRs dentro da área de cobertura da rede secundária, o que dificulta ainda mais o estabelecimento adequado do valor de γ. Sabendo disso, na pesquisa feita em [157], os autores propõem um algoritmo de aprendizado não supervisionado capaz de fazer o ajuste adequado do limiar de decisão em tempo real a fim de garantir que um desempenho alvo de detecção seja alcançado e mantido. A proposta foi desenvolvida para o sensoriamento de sinais de banda larga e tem como base outras duas técnicas: uma técnica de sensoriamento chamada *Radiobot*, e uma técnica de limiar de decisão adaptativo em tempo real baseada em aprendizado supervisionado. A topologia da rede secundária é distribuída e a detecção é feita a partir de uma combinação entre o teste ED e a detecção ciclo-estacionária. Contudo, a principal vantagem da técnica proposta é a capacidade de fazer a detecção sem quaisquer conhecimentos *a priori* sobre o ruído ou características ciclo-estacionárias do sinal primário, pois o método utiliza apenas as amostras do sinal sensoriado. Os resultados são avaliados em termos de ROCs e mostram que a técnica proposta é capaz de atingir o desempenho desejado.

O teste ED também é considerado em [158] em um esquema de sensoriamento espectral cooperativo centralizado com fusão de dados, SD, e fusão de decisões, HD. A fusão de decisões é empregada utilizando-se a regra geral de combinação k-em-m com o valor de k ajustado para que o melhor desempenho de detecção possível seja atingido. No esquema SD, cada CR calcula o nível de energia das amostras do sinal sensoriado e envia este valor ao FC.

Já no esquema HD, cada CR calcula uma estatística de teste local, toma uma decisão de ocupação individual e envia essa decisão ao FC. Na detecção convencional SD, o FC calcula a estatística de teste do teste ED, compara com o limiar de decisão predefinido, e toma a decisão de ocupação cooperativa. Na detecção convencional HD, o FC combina as decisões individuais recebidas e toma a decisão final de ocupação. Como alternativa às abordagens convencionais de combinação das informações recebidas no FC, o estudo propõe uma técnica de ML, baseada em uma rede neural convolucional profunda, *Convolutional Neural Network* (CNN), chamada *Deep Cooperative Sensing* (DCS). Ao contrário do que ocorre nas abordagens convencionais, neste caso a melhor estratégia de combinação do FC é aprendida autonomamente a partir das informações recebidas, que são os valores de energia no esquema SD e as decisões individuais no esquema HD. No cenário adotado, considerou-se o sensoriamento de múltiplas bandas de forma simultânea, canais de sensoriamento com sombreamento correlacionado e transmissor primário e CRs movimentando-se com velocidade definida em certa área preestabelecida. Na técnica DCS desenvolvida, a detecção é feita em duas etapas conforme a estrutura da rede neural proposta. Na primeira etapa uma parte convolucional é responsável por extrair características espaciais das informações recebidas. Já na segunda etapa é feita uma classificação dos dados de saída da parte convolucional por meio da qual as decisões de ocupação são obtidas. Adicionalmente, a técnica SVM também é implementada para fins de comparação entre os resultados das simulações. Os resultados das simulações são avaliados em termos do tempo de execução, das ROCs, e também das AUCs em cada técnica. As ROCs e AUCs revelam que a técnica proposta possui o melhor desempenho na fusão SD, seguida pela fusão HD, técnica SVM e fusão HD convencional. Com base nos resultados, a única desvantagem da técnica proposta foi o maior tempo de execução requerido, apesar desse tempo ainda poder ser considerado baixo e, portanto, favorável à aplicação da técnica em tempo real.

Os autores em [159] empregam um algoritmo desenvolvido para reconhecimento de modulações, chamado *Convolutional Long Short-Term Deep Neural Networks* (CLDNN), na proposta de uma técnica de detecção que utiliza DL no sensoriamento espectral não cooperativo e cooperativo centralizado. Eles propõem um detector baseado na estimação da energia do sinal recebido, nomeado *DetectNet*, com ação individual em cada CR, capaz de detectar a presença de sinais primários por meio da exploração de características fundamentais dos sinais modulados sem a necessidade de conhecimentos *a priori* sobre o estado de ocupação do canal sensoriado ou ruído AWGN. Os testes de desempenho incluem diferentes tipos de modulação e diferente níveis de relação sinal ruído, SNR, bem como o teste ED convencional com a regra de decisão "OU" e outras três técnicas utilizando redes neurais, nomeadas como: LSTM, CNN e DNN. Os resultados são analisados em termos da capacidade de generalização para distinguir diferentes sinais, P_{fa} e P_{d}, número de amostras, e também em termos da SNR *Wall*, que por definição corresponde ao menor valor de SNR do sinal recebido abaixo do qual torna-se impossível detectar a presença de sinais primários de forma confiável, não importando o número de amostras colhidas do sinal recebido durante os períodos de sensoriamento. Como esperado, uma vez que a técnica proposta utiliza o algoritmo CLDNN, desenvolvido especificamente para distinguir diferentes tipos de sinais modulados, nota-se que o detector proposto possui bom desempenho em termos de capacidade de generalização para diferentes ordens do mesmo tipo de modulação, mas é consideravelmente afetado quando testado sob diferentes tipos de modulações. Os resultados também mostram que o detector *DetectNet* possui desempenhos compatíveis com os obtidos pelo teste ED convencional na fusão de decisões pela regra "OU" em termos de P_{d}, no entanto com desempenhos significativamente melhores em termos de P_{fa}, o que confirma sua superioridade. Além disso, nota-se que o detector *DetectNet* é capaz de reduzir a SNR *Wall* em relação ao ED convencional.

Em [160], os autores apresentam uma solução para o sensoriamento espectral em banda larga. Trata-se de um método pouco flexível baseado na transformada de *wavelet* contínua (*Continuous Wavelet Transform*) e na análise de multiresolução (*Multiresolution Analysis*) combinada com a dimensão fractal de Higuchi (uma medida não linear) para detectar a transmissão de usuários primários. Para automatizar todo o processo de detecção e melhorar o sensoriamento espectral em banda larga do rádio cognitivo abordado em [160], técnicas de ML são utilizadas em [161] a fim de detetar a presença de usuários primários e determinar se uma porção do espectro está ocupada ou livre. Isso é possível por meio da análise do comportamento de símbolos, ruído e acurácia das bordas de frequência (mudanças de estado) detectadas. A técnica *K-means* de ML não-supervisionado atua como classificador, e é o que apresenta resultados mais estáveis na simulação durante a fase de detecção de bordas de frequência e percentagem de sucesso em um ambiente com níveis altos de ruído. Um ponto negativo sobre o *K-means* é que é necessário indicar o número K de grupos (*clusters*) a serem classificados, o que apresenta uma dificuldade sendo que os *frames* que são avaliados pelo algoritmo são aleatórios. Considerando as técnicas de ML apresentadas neste estudo deseja-se que sejam implementadas em uma plataforma de rádio definida por *software*, para se obter uma descrição de acurácia dos algoritmos propostos em que parâmetros como o tempo de evacuação (métrica de desempenho que analisa o tempo a partir do momento em que um usuário secundário detecta o primário até o instante em que todos os usuários secundários que causam interferência evacuam o canal) podem ser medidos.

Um classificador *Naive Bayes* é proposto como uma técnica de aprendizado supervisionado em [162] para tratar o problema de sensoriamento espectral em sistemas OFDM. O classificador é utilizado para reduzir o número de candidatos e determinar a ocupação do espectro, a fim de treinar o modelo de ML e reduzir o tempo de sensoriamento do espectro. A eficácia do

método proposto é demonstrada por simulação e verifica-se que para um valor de SNR igual a -18 dB o sistema apresenta um desempenho muito próximo ao *Likelihood Ratio Test* (LRT) ótimo, em termos de ROC. A abordagem deste estudo remove a dependência inerente a SNR, ou seja, para valores de SNRs baixos é possível se obter um ótimo desempenho na detecção do espectro. Comparando o método proposto em relação aos métodos convencionais utilizados para sensoriamento, como ED, detecção de correlação baseada em CP e detecção baseada em *Asymptotic Simple Hypothesis Test* (ASHT), os métodos convencionais são mais sensíveis à incerteza de ruído, por isso não apresentam um bom desempenho e não funcionam corretamente para valores de SNRs baixos.

A técnica NOMA introduz alta complexidade nos modelos de sensoriamento espectral cooperativo tradicionais e sua implementação na camada física é mais complicada pois aumenta a quantidade de estados dos canais nos sistemas de comunicação sem fio, o que torna a modelagem matemática mais difícil. Por isso soluções baseadas em ML são propostas em [163] para resolver o problema da complexidade do modelo de sensoriamento espectral cooperativo para cenários NOMA em redes de rádio cognitivo. Dentre as várias soluções propostas, a que apresentou melhor desempenho em termos de tempo médio de treinamento, melhor capacidade anti-interferência e melhor precisão de detecção foi o algoritmo supervisionado *Directed Acyclic Graph - Support Vector Machine* (DAG-SVM). Isso significa que as soluções de sensoriamento espectral cooperativo baseadas em ML têm complexidade computacional muito baixa e podem atender aos requisitos de tempo real do sistema de rádio cognitivo.

Para atacar o problema do sensoriamento espectral cooperativo, os autores em [164] propõem um modelo Bayesiano *Beta Process Sticky Hidden Markov Model* (BP-SHMM) para redes heterogênea de CR de grande escala. Este modelo Bayesiano permite capturar a correlação espaço-temporal nos dados coletados em diferentes momentos e locais por vários usuários secundários. Ao comparar o desempenho de detecção do modelo proposto BP-SHMM em relação a quatro mecanismos convencionais nomeadamente ED, GMM com *Expectation Maximization* (EM), *Bayesian information criterion* (BIC) e o algoritmo *Mean-Shift* (MS), considerando diferentes números de SUs e PUs ativos, bem como várias potências de transmissão. Os resultados da simulação mostram que o BP-SHMM supera significativamente os outros métodos existentes em termos de ROC e isso valida a sua eficácia.

Os autores de [165] propõem um *framework* de sensoriamento espectral de banda estreita baseado em DL. Ao contrário do sensoriamento espectral baseado em DL existente que utiliza informações especializadas [158,166–168], o método proposto no trabalho usa sinais brutos como entradas para uma CNN. Embora esse sensoriamento espectral baseado em aprendizado profundo seja eficaz ao operar no mesmo cenário em que os dados de treinamento foram coletados, os autores observaram que o desempenho do sensoriamento é degradado quando aplicado em um cenário com diferentes sinais e propagação, ou seja, diferentes condições do canal sem fio. Para melhorar a robustez, eles incorporam a aprendizagem por transferência, que usa pequenas quantidades de dados adicionais para adaptar os modelos já treinados a novas configurações do canal sem fio. Os resultados das curvas ROC apresentadas mostram que a transferência de aprendizado melhora significativamente a robustez do sensoriamento espectral quando pequenas quantidades de dados rotulados são utilizadas para adaptar os modelos treinados a novas condições, superando o desempenho da detecção de energia e se aproximando do sensoriamento ideal.

Em [167], assumindo dados de treinamento limitados e sem conhecimento das estatísticas do canal e dos transmissores, os autores usam uma GAN para gerar dados de treinamento sintéticos adicionais com o intuito de melhorar a precisão de um classificador e adaptar os

dados de treinamento à dinâmica do espectro (e.g., mudança do canal e/ou do transmissor). Neste problema, dados de treinamento rotulados e classificadores, i.e., modelos treinados para detectar a presença ou ausência de usuários, estão disponíveis para um cenário específico, mas conforme as condições do espectro mudam, nenhum dado de treinamento está disponível em um novo cenário. Para enfrentar esse desafio, as amostras sinteticamente geradas pela GAN podem ser usadas para treinar os classificadores RFC e SVM para o novo cenário por meio de adaptação de domínio. Os resultados apresentados mostram que o aumento dos dados de treinamento melhora significativamente a precisão do classificador. Inicialmente, a abordagem utilizando GAN é utilizada para sensoriamento espectral de banda estreita, mas os autores afirmam que ela pode ser diretamente portada para sensoriamento de banda larga.

Os autores de [169] utilizam uma CNN para desenvolver um *framework* para sensoriamento espectral de banda larga cooperativo. Em sua proposta, ao invés da modelagem matemática usada no sensoriamento espectral cooperativo tradicional, os autores propõem uma abordagem que combina os resultados de detecções individuais dos SUs de forma autônoma com uma CNN, independentemente de os resultados de detecção individuais serem quantizados ou não. A CNN usada neste trabalho é treinada com amostras coletadas por vários SUs durante o sensoriamento espectral de diferentes bandas. O *framework* proposto opera independentemente do tipo de decisão utilizada pelos SUs, ou seja, opera com SD ou HD, além de levar em consideração as correlações espaciais e espectrais dos canais sem a necessidade de derivações matemáticas. Os resultados apresentados por eles mostram que o *framework* proposto atinge maior precisão de detecção, que resulta em uma menor probabilidade de falso alarme, em comparação com abordagens convencionais, especialmente em condições de detecção adversas.

Os estudos citados nesta revisão bibliográfica mostram o potencial da combinação entre a IA e o sensoriamento espectral na composição de uma solução inovadora de interoperabilidade entre os diversos serviços de telecomunicações por meio da redução da escassez e subtilização das faixas de frequência destinadas à radiocomunicação. Tais estudos, juntamente com várias referências citadas em cada um deles, apontam diferentes vantagens das técnicas baseadas em ML em relação às técnicas convencionais de detecção. Algumas delas são i) o melhor desempenho de detecção, ii) a adaptabilidade ao dinamismo dos ambientes de propagação e às mudanças de alguns parâmetros sistêmicos da rede primária, iii) a maior confiabilidade de detecção em baixos níveis de SNR, e iv) a maior robustez frente a problemas de incerteza de ruído. Adicionalmente, outra importante vantagem do sensoriamento espectral via técnicas de ML é a capacidade proporcionada à rede secundária de reduzir problemas de alta complexidade matemática às tarefas de aprendizado empírico, podendo ainda tais tarefas serem satisfatoriamente aprendidas em tempo real devido ao baixo tempo de aprendizado requerido por certos algoritmos. As discussões sobre esses estudos levam a crer que a IA será de fato uma ferramenta essencial na viabilização do acesso secundário a canais ociosos do espectro radioelétrico via sensoriamento espectral.

Apesar de todos os trabalhos mencionados, no entanto, é possível perceber o baixo número de pesquisas sobre sensoriamento espectral via ML em comparação com o alto número de pesquisas envolvendo apenas as técnicas convencionais de sensoriamento disponíveis na literatura. Logo, conjectura-se que o aumento deste número possa revelar a necessidade do desenvolvimento de novas técnicas de ML a fim de superar as técnicas convencionais já exaustivamente estudadas em inúmeros contextos, cenários e circunstâncias diversas. Além do mais, o aumento deste número pode mostrar a necessidade de se fazer combinações entre as técnicas de ML e as técnicas convencionais de detecção a fim de alcançar algum desempenho alvo em cenários específicos ou estabelecer uma relação de compromisso entre desempenhos, complexidade, co-

nhecimentos *a priori* disponíveis e tempo de treinamento requerido, por exemplo. Os autores em [170] apontam para esta direção quando propõem uma técnica híbrida de sensoriamento baseada em uma ANN combinada com os testes convencionais ED e LRT. As análises de comparações entre as diversas técnicas, tipos de algoritmos, e tarefas de aprendizado, bem como entre as arquiteturas de redes neurais propostas para o sensoriamento espectral, também são outro ponto importante que parece ainda não ter sido suficientemente explorado na literatura. O trabalho desenvolvido em [171], por exemplo, talvez seja o único estudo sobre sensoriamento espectral via ML dedicado exclusivamente a comparações entre arquiteturas de redes neurais publicado até o momento. Os autores desse trabalho comparam quatro redes neurais em termos de desempenhos de detecção, complexidades computacional, requerimentos de dados de treinamento e requerimentos de memória. Os resultados mostram que as redes CNN, *Recurrent Neural Network* (RNN) e *Bi-Directional Recurrent Neural Network* (BiRNN) possuem desempenhos similares quando não há restrições de recursos de memória, número de dados de treinamento e complexidade computacional. Já os desempenhos da rede *Fully-Connected Neural Network* (FCNN) são inferiores aos demais, a não ser no caso em que há restrições quanto a complexidade computacional máxima permitida.

Por fim, com base nas investigações feitas para compor esta revisão, é importante destacar que as pesquisas direcionadas especificamente aos estudos de soluções para o 5G e o 6G também levam a deduzir que o sensoriamento espectral via ML será de fato a opção mais viável para o provimento de acesso secundário ao espectro radioelétrico, como pode ser verificado nos exemplos em [154, 172–175], em relação ao 5G, e em [176–180], em relação ao 6G.

2.4.3 Uso de Inteligência Artificial em Compressed Sensing

Atualmente, o mercado de telecomunicações está passando por uma mudança onde o uso de um grande número de dispositivos com baixa capacidade de processamento e limitação na fonte de energia será fundamental para a coleta de informações que irão viabilizar diversas aplicações. O *International Mobile Telecommunications* 2020 (IMT-2020) já previu essa demanda para as redes 5G, concentrando as aplicações baseadas na comunicação de dados entre máquinas no cenário denominado de mMTC. No entanto, diversos novos casos de uso voltados para redes 6G irão demandar uma quantidade massiva de dispositivos responsáveis pela coleta de dados em cenários como Gêmeos Digitais, Monitoramento Global, Zonas Seguras Invisíveis, entre outras [181].

A técnica conhecida como CS (ou *compressive sensing*) [182–184] é uma candidata em potencial para redução da complexidade na coleta de dados provenientes de um grande conjunto de dispositivos distribuídos em um dado ambiente. Nesta técnica, os dados provenientes do sensoriamento são sub-amostrados a uma taxa menor do que a taxa de Nyquist, de modo que a informação intrínseca sobre o sensoriamento seja preservada. Essa compressão do sinais de sensoriamento é realizada através de uma transformação linear, baseada em uma matriz de compressão que resulta em um vetor comprimido com dimensão menor do que o vetor de dados original. A informação original sobre o sensoriamento pode ser reconstruída a partir do vetor comprimido, porém o processo de expansão pode ser computacionalmente complexo e requer a resolução de um problema de otimização convexo, o que pode ser inviável dadas as restrições de capacidade de processamento disponível.

Uma possível solução para esse problema consiste em utilizar algoritmos de aprendizado de máquina e IA para recuperar a informação desejada diretamente do vetor comprimido, evitando assim a expansão do sinal. Embora detectores não lineares, como o Detector por Máxima

Verossimilhança (DMV), possam ser usados para realizar essa detecção, os algoritmos de IA podem ser utilizados para essa tarefa com uma complexidade consideravelmente menor e com desempenho em termos de taxa de acerto próximas do caso ótimo. Além disso, os algoritmos de IA podem absorver os erros de estimação de canal e até mesmo suprimir a necessidade do envio de informações piloto, sem perdas significativas de desempenho, melhorando a eficiência espectral e até mesmo superando o DMV nas situações reais de operação da rede. Esse ramo de estudo que emprega IA para detectar informações de sinais comprimidos no âmbito do CS recebeu o nome de *Compressed Learning* (CL). O CL considera, primordialmente, o uso de DNNs para realizar a classificação dos padrões presentes nos vetores comprimidos recebidos.

CS tem uso em diversas tecnologias 6G, tal como na estimação de canais de sistemas operando em ondas milimétricas ou terahertz. Por exemplo, em [185], os autores usam o método UCB para *armed bandits* medidas do canal MIMO de forma eficiente através do projeto de dicionários eficientes. A seguir, CS é discutida no âmbito de sensoriamento espectral.

O uso oportunista do espectro, de forma harmoniosa com outras tecnologias legadas, será uma das principais características das redes de acesso empregadas em 6G. Para viabilizar essa coexistência, a rede 6G deverá coletar dados sobre a ocupação espectral na sua região de cobertura e decidir, de forma inteligente, quais faixas de frequências podem ser utilizados para atender as demandas de seus usuários, sem causar impacto na comunicação das redes primárias. Isso irá demandar um elevado número de informações sendo enviadas por sensores distribuídos em uma dada região é o sensoriamento espectral para viabilizar a exploração das bandas de frequências ociosas, principalmente na faixa de TV, também chamadas de *TV White Space* (TVWS). As informações sobre a ocupação espectral obtidas por cada um dos sensores da rede precisam ser enviadas para um centro de fusão capaz de processar os dados de todos os sensores e que tenha acesso ao bando de dados georreferenciado para a tomada final de decisão. A transmissão destes dados para o centro de fusão é um grande desafio e resulta em uma sobrecarga considerável em termos de sinalização para a rede de comunicação.

No caso da aplicação de sensoriamento espectral em um cenário de ocupação dinâmica do canal, C ERBs primárias compartilham o espectro empregando o protocolo LBT. Sensores distribuídos na região espacial realizam medições do espectro empregando algoritmos de sensoriamento espectral, como aqueles descritos na subseção 2.4.1. Os resultados destes sensoriamentos são encaminhados para um *gateway*, que é responsável por enviar as informações para o centro de fusão. Assume-se que o *gateway* fica posicionado em uma localidade privilegiada em relação aos nós sensores e que o uso de um código corretor de erro potente reduz a probabilidade de erro a níveis negligenciáveis. O *gateway* emprega uma matriz de compressão e envia o vetor comprimido para o centro de fusão, que utiliza uma DNN para detectar o padrão que determina qual dentre as C ERBs primárias está transmitindo em uma determinada janela de tempo. A Figura 23 apresenta o modelo no qual o CS é empregado para o sensoriamento espectral e o CL é usado pelo centro de fusão para detectar a classe representada pelo vetor comprimido.

O projeto da DNN consiste em determinar os seguintes hiper-parâmetros:

- Número de neurônios da camada de entrada;
- Número de neurônios da camada de saída;
- Número de camadas escondidas;
- Número de neurônios por camada escondida;

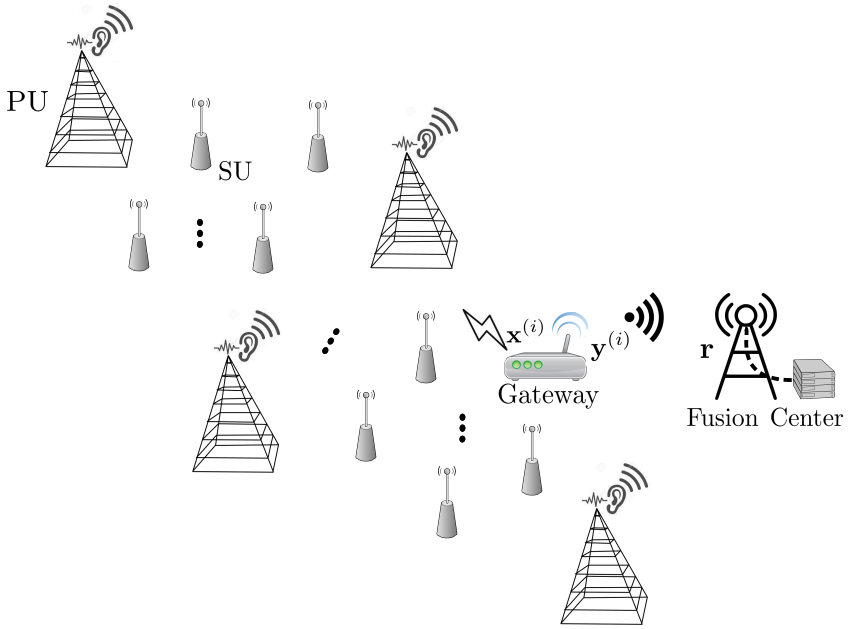


Figura 23: Modelo sistêmico para o uso de CS no sensoriamento espectral de ERBs primárias que empregam o protocolo LBT.

- Funções de ativação dos neurônios;
- Algoritmos de aprendizagem;
- Taxa de aprendizagem;
- Dimensão do conjunto de treinamento.

Para o caso do CL, o número de neurônios da camada de entrada é definido pela dimensão do vetor comprimido e o número de neurônios da camada de saída é definido pelo número de classes que podem ser detectadas, ou seja, do número de ERBs primárias em operação na região. Isso significa que é necessário conhecer C para o dimensionamento da DNN, o que pode ser considerado uma desvantagem do CL, uma vez que o número de classes pode variar ao longo do tempo em um cenário muito dinâmico. Já os demais hiper-parâmetros devem ser determinados por uma busca de tentativa por uma especificação que resulte em desempenho satisfatório, observando-se os cuidados para evitar a divergência da função de erro durante o processo de treinamento por *overfitting* ou *underfitting*.

Resultados de simulação apresentados na literatura [186] mostram que o uso de DNN para a detecção da classe observada pelo sensoriamento espectral supera o desempenho do DMV em termos de probabilidade de erro de detecção em cenários onde erros de estimação de canal estão presentes ou nos casos em que a sinalização piloto é reduzida para economia de recursos da rede. Também é importante destacar que esse melhor desempenho em condições reais de operação é obtido com uma complexidade de implementação menor. Esses resultados mostram que o uso de algoritmos de IA é essencial para a implementação do CS, uma vez que esses algoritmos garantem o desempenho satisfatório em condições reais de operação e com um menor custo computacional para o centro de fusão.

2.5 Esquemas de Múltiplo Acesso Não Ortogonal

Mariana Mello, Luciano Mendes
mariana.mello@dtel.inatel.br, luciano@inatel.br

A próxima geração de redes sem fio enfrentará desafios significativos para dar suporte ao tráfego de dados heterogêneos de grande escala. Em cenários de mMTC será necessário lidar com uma grande quantidade de dispositivos equipados com baterias. Como estes dispositivos possuem energia limitada, é imprescindível a utilização de receptores de baixa complexidade e sinais com baixa *Peak-to-Average Power Ratio* (PAPR) para aumentar a eficiência do amplificador de potência [187]. Além disso, para reduzir a latência e sobrecarga em comunicações de *uplink* cujo tráfego da rede consiste em rajadas curtas, o acesso ao meio sem concessão deverá ser empregado. Já em cenários de eMBB, a principal questão a ser tratada é a eficiência espectral de pico elevada, portanto a complexidade do receptor na ERB pode ser alta, enquanto a PAPR dos sinais transmitidos e a complexidade do receptor móvel devem ser baixos [187]. Em aplicações de URLLC, como sistemas de comunicação veicular, os requisitos exigentes de baixa latência e baixo índice de perda de pacotes deverão ser atendidos.

As técnicas convencionais de *Orthogonal Multiple Access* (OMA), como *Time Division Multiple Access* (TDMA) e *Orthogonal Frequency Division Multiple Access* (OFDMA), atendem a um único usuário em cada bloco de recurso ortogonal. Devido a ortogonalidade presente nestas técnicas, receptores de baixa complexidade podem ser empregados para separar os sinais dos diferentes usuários. No entanto, como a quantidade de recursos ortogonais é limitada, os sistemas OMA não são capazes de atender a um grande número de usuários, tornando um fator limitante em cenários onde conectividade massiva é necessária [188]. Além disso, a utilização do OMA significa que é inevitável que um dos recursos seja ocupado exclusivamente por um usuário, mesmo que este usuário tenha condições do canal muito ruins. Uma vez que o OMA prioriza a não interação entre os dados de diferentes usuários, isso tem um impacto negativo na capacidade, eficiência espectral e taxa de transferência do sistema como um todo.

Estudos recentes mostraram que o NOMA tem potencial para ser aplicado em vários cenários de comunicação de 5G e além, incluindo os cenários de mMTC e IoT. Deste modo, o NOMA é uma das técnicas de acesso de rádio mais promissoras nas comunicações sem fio 6G. Ao contrário da técnica OMA, o NOMA permite que sinais de vários usuários compartilhem o recurso de tempo-frequência por meio da superposição, ao passo que introduz intencionalmente interferência no sinal transmitido. Desta forma, o NOMA ajuda a aumentar a capacidade e a eficiência espectral do sistema ao oferecer um uso mais eficaz dos recursos disponíveis, ao custo de maior complexidade no receptor. Comparado ao OMA, o NOMA oferece um conjunto de potenciais benefícios, como [188, 189]

1. Aumento da eficiência espectral, devido ao uso simultâneo do mesmo recurso de tempo-frequência por vários usuários;

2. Suporte a conectividade massiva, ao acomodar mais usuários para lidar com a sobrecarga do sistema;

3. Menor latência devido à transmissão simultânea, uma vez que o usuário não precisa esperar por um intervalo de tempo programado dedicado para transmitir suas informações;

4. Melhor utilização da heterogeneidade das condições do canal. Devido a multiplexação intencional de usuários fortes com usuários fracos, o ganho de desempenho do NOMA sobre OMA é maior quando os canais possuem ganhos bastante distintos.

5. Além de conseguir manter a justiça do usuário e a QoS diversificada através da alocação de recursos mais flexível entre os usuários fortes e fracos.

Além disso, existem algumas evidências de melhoria de desempenho quando NOMA é integrado com outras técnicas eficazes de comunicação sem fio, como comunicações cooperativas [190], MIMO [191], *beamforming* [192], etc.

Várias técnicas NOMA têm sido propostas, as quais podem ser classificadas em duas categorias, NOMA de domínio do código e NOMA de domínio da potência. As técnicas NOMA de domínio de código foram desenvolvidas a partir do *Code Division Multiple Access* (CDMA) clássico e distinguem os múltiplos usuários com ajuda de sequências de espalhamento não ortogonais específicas do usuário. As principais técnicas NOMA de domínio do código são brevemente descritas a seguir:

- *Interleave Division Multiple Access* (IDMA) [193]: baseia-se em um intercalador de *chip* único e específico do usuário para distinguir os sinais de diferentes usuários. Portanto, o IDMA pode ser visto como CDMA intercalado por *chip*, que tem o benefício de um ganho de diversidade, uma vez que se um ou dois *chips* forem corrompidos, a sequência de propagação correspondente ainda pode ser recuperada com o auxílio de uma estratégia de detecção multiusuário iterativa *chip* por *chip* de baixa complexidade.

- *Sparse Code Multiple Access* (SCMA) [194]: é uma técnica de espalhamento não ortogonal baseada em livro código multidimensional de estrutura esparsa. No SCMA, o procedimento de mapeamento de para símbolo QAM e espalhamento são combinados, e os bits de entrada são mapeados diretamente para palavras código multidimensionais de conjuntos de livros código SCMA. No receptor, o SCMA elimina a interferência por um *Multi-User Detection* (MUD) baseado no algoritmo de *Message Passing Algorithm* (MPA).

- *Multi-User Shared Access* (MUSA) [195]: adota sequências de espalhamento não binárias de valor complexo para distinguir os dados de diferentes usuários. Devido à liberdade adicional fornecida pela parte imaginária, o comprimento da sequência de espalhamento complexa pode ser curto. No receptor, o *Successive Interference Cancellation* (SIC) é usado para cancelar a interferência entre os usuários, supondo uma sincronização perfeita.

- *Pattern Division Multiple Acces* (PDMA) [196]: os dados dos usuários são mapeados para um grupo de elementos de recursos de acordo com um padrão definido. A ideia principal por trás da técnica PDMA é obter as vantagens da diversidade sem perder em termos de eficiência espectral [197]. No entanto, a combinação linear dos dados dos usuários altera as características do sinal e aumenta a complexidade no lado do receptor. Fatores de escala de potência e deslocamentos de fase podem ser incluídos na matriz PDMA para proteger os dados dos usuários e tornar a detecção mais fácil. No lado receptor, o algoritmo SIC pode ser usado para separar os dados dos usuários multiplexados no mesmo elemento de recurso, por apresentar um bom equilíbrio entre complexidade e desempenho em termos de qualidade de detecção de multiusuários.

O conceito do *Power-Domain Non-Orthogonal Multiple Access* (PD-NOMA) [198] remete à concepção de PDMA proposta em [199], onde múltiplos usuários têm diferentes potências de recepção no *uplink* para que o SIC possa ser empregado. A ideia do esquema PD-NOMA é garantir que diferentes usuários possam ser atendidos pela estação rádio base ao mesmo código/tempo/frequência, mas com níveis de potência diferentes. No PD-NOMA, a alocação de

potência é realizada considerando duas prioridades de sistema; justiça dos usuários e requisitos de QoS dos usuários. Como podem existir variações significativas entre as condições do canal de diferentes usuários devido o efeito de próximo-distante, a detecção multiusuário no receptor é realizada pelo SIC. O uso do algoritmo SIC implica que quando o sinal de um usuário é decodificado, a mensagem de outros usuários é tratada como interferência, portanto, o sinal dos usuários restantes será decodificado com a vantagem de não haver interferência do usuário já decodificado.

Um dos principais desafios do uso das técnicas NOMA é a detecção, que deve ser capaz de lidar com a interferência gerada entre os usuários. O SIC pode ser considerado o principal método de detecção NOMA aplicado aos receptores em transmissões NOMA de *uplink* e *downlink*. Para executar o SIC no receptor é muito importante conhecer perfeitamente o CSI. No entanto, adquirir o CSI perfeito ou quase perfeito é uma tarefa desafiadora. Portanto, o SIC é limitado pela complexidade no receptor e pelos problemas de propagação de erro. Uma solução para contornar estes problemas é incorporar técnicas de IA ao NOMA.

Neste contexto, em [200], o aprendizado profundo foi usado para abordar as deficiências do método SIC em sistemas de comunicação MIMO-NOMA. Em contraste com o SIC tradicional, que divide o processo de detecção em blocos separados, incluindo estimativa de canal, detecção *Minimum Mean Squared Error* (MMSE), demodulação, decodificação de canal e decisão de sinal, o método de aprendizagem profunda pode realizar todos esses procedimentos como um único processo [200]. Portanto, após o treinamento da DNN, o sinal das antenas receptoras pode ser enviado diretamente para o detector MIMO-NOMA baseado em aprendizado profundo, sem necessidade de nenhum processamento de sinal extra. Os resultados da avaliação do desempenho em termos de taxa de erro de símbolo do sistema proposto foram comparados ao MIMO-NOMA tradicional baseado em SIC. Também foi analisado o impacto de parâmetros de sistema como modulação, alocação de energia e tamanho do mini-*batch*, no desempenho do detector MIMO-NOMA proposto. Em todas as situações avaliadas em [200], o detector proposto superou o método SIC tradicional.

Sob a consideração de que os decodificadores SIC são imperfeitos na prática, em [201] foi proposto um esquema para o sistema MIMO-NOMA de *downlink* baseado em aprendizado profundo. Foi construído um pré-codificador não linear e decodificadores SIC empregando FNNs. O esquema proposto permite que os sinais transmitidos para os usuários sejam devidamente pré-codificados na ERB na maneira de superposição e os sinais recebidos possam ser decodificados com precisão pelos usuários [201]. Além disso, o treinamento do pré-codificador e dos decodificadores SIC propostos foi realizado de forma a minimizar o *Mean Squared Error* (MSE) total entre os sinais desejados dos usuários e seus sinais decodificados. O desempenho do esquema proposto é avaliado em termos de MSE e *Bit Error Rate* (BER), e comparado com a abordagem linear da literatura [202]. Os resultados mostraram que o esquema proposto é capaz de abordar a decodificação imperfeita SIC de forma eficaz e garantir confiabilidade alta.

Ainda sobre a estimação de CSI, em [203] o aprendizado profundo foi proposto para melhorar a estimativa de canal e, assim, diminuir erro da estimativa de canal que se propaga pelo receptor SIC e compromete o desempenho do MUD em um sistema MUSA. A DNN é empregada como uma unidade de processamento não linear adicional sobre a estimativa *Least Squares* (LS) para corrigir o erro de estimativa geral, aprendendo a frequência do canal, resultando em canais estimados melhorados. Os resultados numéricos mostraram que a DNN melhora a estimativa do canal LS e, consequentemente, melhora o desempenho geral do sistema e permite que o MUSA consiga alcançar um fator de sobrecarga razoável.

Em [204], o aprendizado profundo foi empregado no sistema NOMA para obter o CSI. Uma

rede LSTM baseada em aprendizado profundo foi incorporada em um sistema NOMA típico, para que as características do canal fossem detectadas automaticamente. Os resultados das simulações mostraram que sistema NOMA auxiliado por LSTM proposto pode alcançar melhor desempenho em termos de taxa de erro de bloco e taxa de dados de soma.

O desempenho do sistema NOMA também depende das estratégias de alocação de recursos. Um algoritmo de alocação de recursos adequado pode maximizar a taxa de soma geral e otimizar o consumo de energia. Uma solução eficiente é utilizar o aprendizado de máquina e IA para superar o problema de alocação de recursos em sistema NOMA, principalmente em cenários onde a condição de canal muda rapidamente. Em [205], o aprendizado profundo foi utilizado para resolver um problema de minimização de energia no *downlink* em sistema NOMA multi-portadora compatível com PDMA e *Simultaneous Wireless Information and Power Transfer* (SWIPT) com receptores baseados em comutação de tempo. O objetivo foi desenvolver uma solução conjunta de alocação de recursos, que consiste em encontrar a solução ótima da atribuição de subportadora, alocação de potência e da relação de comutação de tempo, que minimize a potência total de transmissão e apresente uma complexidade computacional aceitável. Para isto, foi estabelecida uma *Deep Belief Network* (DBN) para cada parâmetro a ser otimizado. De modo que, para os ganhos de canal apresentados na entrada das DBNs, a aproximação da solução ótima do problema de otimização de energia é formada com base na saída de cada DBN. Os resultados numéricos mostraram que a abordagem proposta [205] pode produzir uma solução que é semelhante àquelas derivadas tanto pelo método de busca exaustiva quanto pelo algoritmo genético, enquanto reduz significativamente o tempo de computação necessário.

No entanto, o uso de aprendizado profundo supervisionado em algoritmos de alocação de recursos requer um conjunto de dados de treinamento correto, cujas as soluções ideais são difíceis de se obter através de simulações por se tratar de um problema não convexo e NP-*hard*. Por exemplo, em [205], o conjunto de treinamento foi gerado através do algoritmo genético, que apresenta um alto custo computacional. Além disso, o treinamento *offline* da rede neural profunda costuma a ser demorado, devido a quantidade de camadas e neurônios.

Uma saída viável para problemas de alocação de recursos, cuja a tomada de decisão deve ser dinâmica e em tempo real, é o uso da aprendizagem por reforço. Neste tipo de aprendizagem não há a necessidade de conhecer *a priori* os pares de dados rotulados de entrada/saída e as informações do modelo. Além disso, o aprendizado por reforço pode resultar em uma política de decisão quase ótima que maximiza o desempenho a longo prazo do sistema por meio de interações constantes. Em [206], os autores propõem um algoritmo baseado em DRL, composto de duas etapas, para a atribuição conjunta de subportadora e a alocação de energia em um sistema NOMA de *uplink* visando maximizar a eficiência energética e garantir a qualidade de serviço para todos os usuários. Na primeira etapa do algoritmo é projetada uma *Deep Q Network* (DQN) para emitir a política de atribuição de subportadora ideal, usando as condições do canal atual como entrada. Na segunda etapa, uma rede DDPG, capaz de fornecer saídas contínuas, é empregada para selecionar dinamicamente a potência de transmissão de todos os usuários. Por fim, toda a política de alocação de recursos é ajustada através da atualização dos pesos das redes neurais de acordo com o *feedback* do sistema. O desempenho da abordagem proposta é comparado com a abordagem de alocação de energia fixa e com a abordagem de alocação de energia com valores quantizados. Os resultados simulados mostram que o DRL pode fornecer melhor eficiência energética sob várias limitações de potência de transmissão se comparado as essas outras abordagens.

Além disso, o NOMA pode ser combinado com a sinalização *Faster-than-Nyquist* (FTN) a fim de explorar os benefícios conjuntos destas técnicas. A sinalização FTN é uma técnica

promissora que pode melhorar a eficiência espectral de futuras redes móveis por meio da quebra da ortogonalidade entre os símbolos, causando ISI intencional. O FTN é capaz de aumentar a taxa de transmissão sem consumir mais largura de banda ou aumentar o número de antenas dos transceptores, ao custo de maior complexidade no receptor. Assim, no FTN-NOMA os sinais de transmissão baseados em FTN de vários usuários podem ser sobrepostos em um mesmo recurso. Métodos convencionais de detecção não são capazes de lidar adequadamente com as interferências intra (ISI) e inter-usuários introduzidas pelo esquema de transmissão FTN-NOMA. Por isso, em [207] foi proposto um MUD auxiliado por aprendizado profundo para o sistema FTN-NOMA com base na detecção de janela deslizante. A DNN proposta estima os bits transmitidos diretamente dos sinais recebidos, com complexidade computacional muito menor do que os detectores FTN convencionais, como o *Bahl–Cocke–Jelinek–Raviv* (BCJR) e o receptor *Viterbi*. Os resultados numéricos revelaram que o desempenho de BER do esquema proposto supera o algoritmo *Minimum Mean Squared Error-Frequency Domain Equalization* (MMSE-FDE) e pode se aproximar do método de máxima verossimilhança que tem o desempenho ideal nos casos OMA e NOMA. No entanto, os resultados em [207] foram limitados ao canal AWGN e a modulação binária. Futuras contribuições deverão considerar métodos de detecção auxiliados por aprendizado profundo de FTN-NOMA em situações de canais multi-percursos e esquemas multi-portadoras, além de considerar ordens de modulações mais altas.

Em esquemas NOMA de domínio do código, o projeto do livro-código também tem forte influência sobre o desempenho do sistema. Projetar livros-código manualmente é problemático, uma vez que as palavras-códigos contidas neles não são ortogonais entre si e são constituídas por valores complexos multidimensionais [208]. O uso de técnicas de IA pode permitir a derivação autônoma de um livro-código eficiente para um sistema NOMA de domínio do código. Neste sentido, em [208], foi proposto um SCMA auxiliado por aprendizado profundo, no qual o livro-código que minimiza a BER é adaptativamente construído e uma estratégia de decodificação é aprendida usando a estrutura de um *autoencoder* baseado em DNN. Os resultados numéricos em [208] mostraram que o esquema proposto supera o esquema convencional SCMA e a decodificação MPA em termos de BER e complexidade computacional.

O uso de algoritmos de IA em sistemas IDMA ainda é pouco explorado na literatura. A estimativa de canal baseada na rede neural treinada pelo algoritmo de Levenberg-Marquardt foi proposta para estimar os coeficientes de canal em sistemas OFDM-IDMA em [209]. O desempenho do estimador proposto foi comparado ao desempenho dos algoritmos LS e MMSE. Embora, o MMSE tenha apresentado melhor desempenho, o estimador baseado em redes neurais proposto tem como principal vantagem não precisar conhecer a priori as estatísticas do canal e as informações de ruído. A Tabela 5 sumariza as aplicações de IA em NOMA apresentados anteriormente e destaca os pontos fortes e fracos de cada abordagem.

Apesar dos algoritmos de IA serem empregados em sistemas NOMA para melhorar o desempenho, ainda existem alguns desafios a serem resolvidos para as redes 6G. Um deles é a escolha eficiente dos hiper-parâmetros da arquitetura de aprendizado profundo, que inclui número de camadas, número de neurônios por camada, função de ativação e taxa de aprendizado, para reduzir o custo computacional. Grande parte dos algoritmos de aprendizado profundo empregados na literatura em NOMA resultam em alta complexidade computacional. Além disso, a qualidade e quantidade dos dados disponibilizados para o treinamento também são importantes para o bom desempenho do algoritmo de aprendizagem e do sistema NOMA como um todo. O projeto e treinamento adequado da rede neural pode impactar positivamente na complexidade e latência durante a etapa de teste, resultando em uma convergência rápida se comparado com as soluções iterativas e de força bruta empregadas em sistemas NOMA convencionais.

Tabela 5: Aplicações de algoritmos de IA em esquemas NOMA.

Referência	Algoritmo de IA	Aplicação	Pontos Fortes	Pontos Fracos
[203]	Aprendizado profundo - DNN	Melhorar a estimativa de canal.	Melhora a estimativa LS do canal e, consequentemente, melhora o desempenho geral do sistema MUSA.	Limitação do SIC, usuários com SNRs específicas. Matriz livro-código precisa conter sequências com baixa correlação entre si.
[207]	Aprendizado profundo - DNN	Detectar o sinal FTN-NOMA e separar fontes.	Desempenho melhor do que o MMSE-FDE (detector convencional) e complexidade menor que o detector baseado em Viterbi. Combinação do NOMA com modulação não ortogonal.	Cenário não realista. Ordem de modulação baixa, canal AWGN e poucos usuários.
[200]	Aprendizado profundo - DNN	Detectar sinal em sistemas MIMO-NOMA de *downlink* (Estimação de CSI e detecção de sinal simultânea).	O sistema proposto pode processar o sinal MIMO-NOMA tradicional diretamente em vez de implementar um receptor SIC. O processo de estimativa de canal e detecção de sinal é realizado simultaneamente. Além do ganho de desempenho, o uso de IA ajuda na redução da sobrecarga do sinal de referência para aumentar a taxa de transferência no sistema de *downlink*.	A análise foi realizada apenas para o esquema PD-NOMA. Além disso, não foi analisado o desempenho do detector proposto na situação de múltiplos *clusters*.
[209]	Algoritmo Levenberg-Marquardt - MLP	Estimar CSI.	Estimação cega do canal. Melhor desempenho que o estimador LS.	Considera um sistema ortogonal. O desempenho foi analisado apenas para canal seletivo invariante. Além disso, o desempenho da abordagem proposta é pior que o MMSE e piora conforme o número de usuários no sistema aumenta.
[205]	Aprendizado profundo - DBN	Alocação de recursos sob condições de canal dinâmicas para minimização da potência.	Requer menos tempo para se aproximar da solução ideal do que os algoritmos iterativos convencionais e, portanto, facilita o atendimento ao requisito de latência ultrabaixa.	Conjunto de treinamento é obtido utilizando o algoritmo genético, que necessita de um tempo computacional muito grande para convergir.
[208]	Aprendizado profundo - DNN	Derivação autônoma do livro-código e decodificação SCMA. (*autoencoder*)	A derivação autônoma do livro-código SCMA é aplicável a qualquer número de usuários e recursos. Além disso, o livro-código SCMA proposto possui melhor desempenho do que o convencional. O detector baseado em DNN possui desempenho comparável ao MPA, porém com custo computacional menor.	Não apresenta o desempenho para canais com desvanecimento. Cenário não realista.
[206]	Aprendizado por reforço - DQN e DDPG	Alocação de recursos sob condições de canal dinâmicas para maximização a longo prazo da eficiência energética.	Não precisa conhecer *a priori* os pares de dados rotulados de entrada/saída e informações do modelo. Também não há treinamento *offline*, que pode ser custoso computacionalmente. Melhor desempenho que as abordagens de alocação de energia fixa e quantizada.	A abordagem não foi comparada com outros algoritmos de aprendizagem de máquina, como o aprendizado profundo. Além disso, a dimensão da DQN aumenta exponencialmente com a quantidade de ações.

Outro dilema em relação ao uso de algoritmos de aprendizado profundo em esquemas NOMA é o consumo de energia durante a etapa de treinamento, que pode ser também um fator limitante. Treinar redes neurais de maneira eficiente e rápida, consumindo o mínimo de energia possível, ainda é uma questão bastante desafiadora.

Por fim, as redes de próxima geração devem incorporar o NOMA à outras técnicas de comunicação sem fio, assim como considerar cenários mais realistas, heterogêneos e dinâmicos. Nestes cenários, o uso da aprendizagem por reforço pode oferecer benefícios significativos em termos de taxa de transferência, latência e alocação de recurso, uma vez que é capaz de aprender com o *feedback* em tempo real, bem como com suas experiências históricas, o padrão variável do ambiente ao qual está inserido. Portanto, apesar de não ser extensivamente empregado na literatura NOMA quanto o aprendizado profundo, o aprendizado por reforço é uma abordagem bastante promissora em esquemas NOMA 6G.

2.6 Gerenciamento de mobilidade em redes mmWave

Davi da Silva Brilhante
dbrilhante@land.ufrj.br

O 6G trará novos desafios quanto aos dispositivos terminais que serão os usuários da rede. Além das aplicações para *smartphones*, novos casos de uso estão previstos para o 6G, como automação industrial com robôs que se movem nos chãos de fábrica, sistemas de transporte inteligentes, veículos autônomos, com direção inteligente e VANTs [28]. Esses novos casos de uso demandam novas garantias de desempenho da rede, como latência ultra-baixa, cobertura total e ultra-banda larga. Portanto, a mobilidade é vista como um desafio ainda não alcançado em sua totalidade para o 5G, mas que será um marco tecnológico para o 6G.

O gerenciamento de mobilidade consiste em dar suporte ao usuário para que este não perca conexão com a rede a qual está associado. As redes sem fio, seja qual for a tecnologia, têm seu alcance limitado pela potência de transmissão máxima de cada célula ou ponto de acesso, que em geral limita o alcance das células a alguns quilômetros. Portanto, um usuário que se desloca grandes distâncias ou está transitando na borda da célula, sofre mudança de célula ou *handover*. O *handover* demanda que a rede realize operações de transferência dos dados que estavam sendo trocados entre o usuário e a rede. Idealmente, o *handover* é transparente ao usuário e este não percebe a interrupção do serviço causada pela troca de célula.

As redes heterogêneas fazem uso de diferentes tipos de célula, com alcances distintos, para reduzir a interferência entre células, aumentando a eficiência espectral da rede [210]. Células com menor potência, mas em bandas do espectro mais favoráveis e sofrendo menor interferência, como pico-células, potencialmente provêm maior taxa de dados a um usuário do que uma macro-célula, com maior alcance e sofrendo maior interferência. Células menores também possuem menor custo para os provedores de Internet móvel e consomem menos energia para operar [211]. Por isso, as redes móveis se tornam cada vez mais densas, ou seja, contam com mais células transmitindo em menor potência. Em contrapartida, células com menor alcance levam a mais *handovers* e assim a interrupções no serviço mais frequentes [212].

Para habilitar a mobilidade em redes nas bandas de mmWave e tera hertz é preciso realizar de forma eficiente o rastreamento do feixe (do inglês, *Beam Tracking*), de modo a reagir às mudanças frequentes de ambientes dinâmicos. Aliado a isto, as redes em mmWave e na banda de tera hertz, como já dito, dependem de antenas direcionais, possuem baixa capacidade de difração para contornar objetos e alta perda por absorção nos mesmos. Esses aspectos inerentes a essas tecnologias de comunicação requerem da rede novas capacidades relacionadas ao bloqueio do sinal. As estratégias mais comuns para lidar com o bloqueio são: detecção de bloqueio, predição do bloqueio e mitigação dos efeitos do bloqueio.

Com o bloqueio como novo fator para gerar perdas no sinal, além da perda de propagação naturalmente elevada que reduz ainda mais o alcance das células, os algoritmos tradicionais de *handover* baseados em diferença de potência recebida não apresentam desempenho satisfatório quando nos cenários de mmWave e comunicações em tera hertz [213]. Usualmente, esses algoritmos levam a *handovers* desnecessários ou antecipados, aumentando a probabilidade de usuário ter o acesso à rede interrompido. Portanto, técnicas de IA podem usar dados externos à rede, como velocidade e sentido do movimento do usuário, imagens de câmeras, posicionamento via *Global Positioning System* (GPS) e etc. [214], para auxiliar no processo de tomada de decisão que é a realização *handovers*, tornando-o mais eficientes e oferecendo maior suporte aos usuários que estão em mobilidade.

Uma estimativa da qualidade do enlace de um esquema pró-ativo de *handover* para pedestres

foi proposta em [215] usando imagens de câmeras *Red, Green, Blue and Depth* (RGB-D) e o método de aprendizado *online Adaptive Regularization of Weight Vectors* (AROW) para predizer a aproximação de outros pedestres que gerariam bloqueio, mitigando assim seu efeito através do gerenciamento do *handover*. Abordagens semelhantes são encontradas também em [216–220]. Em [216], *Q-learning* é aplicado em uma rede para pedestres auxiliada por câmeras para estabelecer a política ótima de *handover*. Um arcabouço para *handover* baseado em aprendizado por reforço aliado a redes neurais foi proposto em [217] e [218], também usando imagens, e superou o método tradicional de *handover* baseado em potência em termos de taxa de dados. Em [219] e [220], os autores adicionaram múltiplas câmeras ao arcabouço anteriormente citado, obtendo desempenho melhor do que com uma única câmera.

Voltado para redes veiculares, os autores de [221] implementaram uma rede neural profunda baseada em *Gated Recurrent Unit* (GRU) que usa o histórico de feixes de um usuário para predizer bloqueios e pró-ativamente realizar o *handover*, que obteve acurácia na predição dos *handovers* em mais de 90% dos casos. Em [222], os autores consideram um sistema *dual-band* em que cada estação base conta com transceptores tanto na banda sub-6GHz quanto em mmWave, explorando a omnidirecionalidade do canal na banda de sub-6GHz para aumentar a taxa de sucesso dos *handovers* usando o classificador *Extreme Gradient Boosting*. Baseado nos feixes previamente utilizados na rede, uma rede neural recorrente é proposta em [223] para realizar rastreamento de feixes, obtendo acurácia na predição de 85% para um feixe, 68% para 3 feixes e 60% para 5 feixes. A abordagem *dual-band* também foi aplicada em [224] que, alimentando um método de ML baseado em função Kernel, realiza predições da posição do veículo e para acelerar o *handover*, uma série histórica de dados e um algoritmo KNN decidem como o *handover* procederá.

2.7 Estimação de canal, equalização e detecção de sinais

Felipe Augusto Pereira de Figueiredo
felipe.figueiredo@inatel.br

Estimação de canal, equalização e detecção de sinais são três tarefas cruciais interrelacionadas para atingir a capacidade de canal em sistemas de comunicação sem fio. Tais tarefas são, convencionalmente, implementadas e otimizadas individualmente. Abordagens baseadas em ML permitem a otimização individual dessas tarefas, bem como sua otimização conjunta, que é uma tarefa substancialmente complicada em sistemas convencionais, mas que pode ser bastante simplificada através do uso de algoritmos de ML, como, por exemplo, redes neurais.

Modelos de aprendizado de máquina podem ser usados para estimar ou prever parâmetros de rádio associados a usuários específicos. Por exemplo, em sistemas MIMO massivo, i.e., sistemas com arranjos de antenas com um grande número de elementos, tanto a detecção quanto a estimativa de canal levam a problemas de busca de alta dimensão, que podem ser resolvidos por modelos de aprendizagem [35].

A família de técnicas de aprendizagem supervisionada (e.g., regressão, KNN, SVM, redes neurais, etc.) depende de modelos e rótulos conhecidos que podem apoiar a estimativa de parâmetros desconhecidos. Estas técnicas podem ser aplicadas aos problemas de estimação de canal, equalização e detecção de sinais.

2.7.1 Estimação de Canal

Técnicas de ML usadas no processamento de imagens, visão computacional e processamento de linguagem natural são adaptadas em vários trabalhos para a tarefa de estimação de canal, onde as correlações entre tempo, frequência e espaço dos canais são exploradas durante aprendizagem. Em [225], os autores utilizam abordagens como super-resolução e restauração de imagens para interpolação de canal e supressão de ruído, tratando a resposta em frequência temporal de um canal com desvanecimento como uma imagem 2D de baixa resolução.

A tecnologia conhecida como MIMO massivo, envolve o uso arranjos de antenas com um grande número de elementos e transceptores totalmente digitais nas ERBs. Ela é uma tecnologia prática cujos conceitos principais são adotados em redes 5G [226]. A estimativa de canal usando sequências piloto transmitidas no sentido de *uplink* de sistemas MIMO convencionais e massivos é um problema bem estudado no caso de hardware ideal tanto na ERB quanto nos equipamentos de usuário, *User Equipment*s (UEs) [227]. No entanto, na prática, deficiências do transceptor, como não-linearidades em amplificadores, desequilíbrio I/Q e erros de quantização são inevitáveis [228].

É possível derivar um estimador MMSE Bayesiano ciente de distorção que utiliza as estatísticas de distorção de primeira e segunda ordens para estimar os canais, mas ao fazer isso, a distorção é tratada como ruído colorido independente, embora dependa do canal. Além disso, deve-se observar que derivar o estimador MMSE é geralmente muito difícil no caso de deficiências de hardware não-lineares.

Existem vários trabalhos que modelam e analisam o impacto de não-linearidades de hardware em sistemas MIMO massivo usando estimadores MMSE Bayesianos. Entretanto, esses estimadores tratam a distorção como ruído colorido independente e utilizem apenas suas estatísticas de primeira e segunda ordens [229]. Esta abordagem resulta em estimadores sub-ótimos para o problema da estimação de canais que envolvam não-linearidades de hardware.

Em [229], os autores apresentam uma abordagem baseada em aprendizagem profunda, i.e., *deep learning*, que melhora a qualidade das estimativas de canal, levando em consideração as características de distorções não-lineares presentes na ERB e UE, ao invés de considerar apenas o canais sem fio. O trabalho mostra como uma abordagem de aprendizagem baseada em dados pode ser combinada com o conhecimento especializado do campo das comunicações sem fio para explorar a estrutura do hardware do transceptor e, assim, superar os abordagens baseadas em modelos sub-ótimos.

No entanto, as técnicas existentes de estimação de canal baseadas em aprendizado profundo têm uma deficiência comum. Como o DNN deve ser treinado *offline* devido aos requisitos de longos períodos de treinamento e grandes bases de treinamento, incompatibilidades entre canais reais e canais na fase de treinamento podem causar uma degradação do desempenho. Em pesquisas futuras, o treinamento *online* e a criação ou obtenção de dados de treinamento que correspondam às condições do canal do mundo real podem ser uma abordagem promissora para superar esse problema [230].

2.7.2 Estimação de Canal e Detecção de Sinal Conjunta

Tradicionalmente, a estimação do canal e a detecção dos sinais transmitidos são dois procedimentos separados no receptor. A CSI é inicialmente estimada por meio de sinais piloto antes da detecção dos símbolos transmitidos. Então, com a CSI estimada, os símbolos podem ser recuperados no receptor.

Em [231], os autores propõem uma estratégia baseada em aprendizagem profunda para estimação e detecção conjunta em sistemas OFDM. Os resultados apresentados por eles mostram que a abordagem proposta tem um melhor desempenho de estimação de canal com uma sobrecarga de sinalização reduzida (i.e., menor número de pilotos e nenhum prefixo cíclico) e é capaz de lidar com ruído de corte (*clipping*) não linear.

Outra abordagem para estimação e detecção conjunta de sinais foi proposta em [232]. Especificamente, uma DNN de cinco camadas totalmente conectadas é incorporada em um receptor OFDM para estimativa e detecção de canal conjunta, tratando o canal como uma "caixa preta". A DNN é treinada para reconstruir os símbolos transmitidos utilizando os sinais recebidos correspondentes aos símbolos e os pilotos como entradas. Portanto, as informações do canal podem ser inferidas implicitamente pela DNN e usadas para prever os símbolos transmitidos diretamente sem precisar estimar explicitamente a CSI. Os resultados demonstram que a abordagem para estimativa e detecção conjunta baseada em DNN supera a abordagem baseada MMSE quando o número de pilotos é insuficiente, sem prefixo cíclico e com distorções não-lineares. A vantagem é que quando esses efeitos adversários ocorrem, desta abordagem baseada em dados pode aprender a lidar com esses efeitos de forma supervisionada, ou seja, atualizando os parâmetros para minimizar a função de custo, o que melhora a robustez do sistema com relação a circunstâncias indesejadas.

2.7.3 Detecção em sistemas MIMO

Na detecção de sinais vindos de sistemas MIMO, os métodos iterativos, que são baseados em detectores Bayesianos ótimos, mostraram desempenho superior com complexidade de computação moderada. No entanto, esses detectores geralmente impõem suposições sobre a distribuição do canal, o que limita o desempenho em muitos ambientes complexos. Ao incorporar abordagens baseadas em aprendizagem de máquina, a adaptabilidade dos detectores pode ser melhorada, uma vez que os parâmetros do modelo podem ser refinados de acordo com os dados específicos. Em [233], os autores propõem um *framework* que combina alguns procedimentos iterativos com um detector *Approximate Message Passing* (AMP) ortogonal para detecção de sinais. Os resultados mostram que o *framework* proposto pode ser facilmente treinado em um período mais curto e com menos dados de treinamento em comparação a uma DNN, melhorando o desempenho do detector AMP ortogonal em canais Rayleigh e MIMO correlacionados. Portanto, essa abordagem pode ser escalonada para sistemas MIMO massivo com grande potencial para serem aplicadas a canais variantes no tempo.

2.7.4 Equalização e Detecção de Atividade Conjunta

Com o rápido desenvolvimento da tecnologia IoT, o número de dispositivos tem crescido vertiginosamente. No entanto, a tecnologia IoT para redes 5G, ou seja, o *Narrow-band IoT* (NB-IoT), não suporta acesso massivo de dispositivos. Neste contexto, é desejado projetar a tecnologia IoT para redes 6G de tal forma que o acesso massivo seja suportado. Para realizar acesso massivo com baixa latência, tal tecnologia deve adotar um protocolo de acesso aleatório sem-concessão, i.e., *grant-free* [234]. Especificamente, os dispositivos podem acessar as redes 6G diretamente após enviarem suas sequências piloto exclusivas para a ERB.

Intuitivamente, a chave do acesso aleatório sem-concessão é a detecção dos dispositivos ativos na ERB. Como as sequências piloto não são ortogonais no contexto do acesso massivo, a detecção de dispositivos ativos não é um problema trivial. Considerando que apenas dispositivos ativos enviam sequências piloto, a detecção de dispositivos ativos pode ser visto como um

problema típico de recuperação de sinais esparsos. Portanto, muitas abordagens baseadas em *compressed sensing* são empregadas para detectar os dispositivos ativos [235]. Em [236], os autores propõem um algoritmo de aprendizado profundo baseado em modelo para detecção de atividade e estimativa de canal conjunta baseado no princípio da passagem aproximada de mensagens, AMP. A vantagem desse algoritmo é que ele não requer informações a priori sobre as probabilidades de dispositivos ativos e a variância do canal e pode melhorar significativamente o seu desempenho com um número finito de dados de treinamento.

2.8 Camada física com AI aplicada fim-a-fim

Luan Gonçalves
luan.goncalves@itec.ufpa.br

A camada física (*Physical Layer*, PHY) refere-se ao conjuntos de tecnologias de transmissão de uma rede, podendo ser implementada através de várias tecnologias com diferentes características (transmissão via satélite, cabo coaxial, radiotransmissão, par metálico, fibra ótica, etc.). Essa camada é cercada de conhecimentos especializados que, por um longo período, foram responsáveis pela sua estagnação em termos de melhorias de performance, dado que tarefas como modulação, demodulação e estimativa de canal possuem aproximações matemáticas capazes de representar satisfatoriamente os eventos físicos relacionados.

Com o surgimento de diferentes perfis de comunicação, que vão de necessidades pessoais a industriais, houve o impulsionamento da evolução dos meios de comunicação móvel. Consequentemente, a PHY passou por um intenso processo de flexibilização através da inserção de parâmetros que podem ser configurados de forma a atender diferentes requisitos de qualidade de serviço (*Quality of Service*, QoS). O efeito colateral dessa flexibilização é o aumento da complexidade de gerenciamento da PHY. De acordo com a afirmação feita em [237], pontos de transmissão 2G, 3G, 4G e 5G possuem, aproximadamente, 500, 1000, 1500 e 2000 parâmetros, respectivamente, a serem otimizados. Logo, a otimização manual desses parâmetros torna-se impraticável.

Recentemente, em [238] um framework de DL foi utilizado para a otimização da formação de beams de canais *Multiple-Input Single-Output* (MISO) de downlink. Neste trabalho, o framework de DL beneficia-se de conhecimentos especializados como a dualidade uplink-downlink e estruturas conhecidas de soluções ótimas, a qual apresenta um bom equilíbrio entre complexidade e custo computacional. No entanto, este trabalho não aborda os efeitos de imperfeições no CSI e cenários de células múltiplas.

Considerando a complexidade inerente a estimativa de CSI, os frameworks de DL se apresentam como soluções promissoras. Em [239], o canal é considerado como uma caixa preta e uma rede neural profunda é utilizada para estimação e detecção conjunta de canais. Os resultados apresentados sugerem que as redes neurais profundas são mais robustas que os métodos convencionais. No momento da escrita do presente texto, a tendência para a otimização da PHY basea-se no paradigma de redes autoencoder, introduzido em [240] e aprofundado em [6]. Em [240], os autores utilizam redes autoencoder para modelar sistemas de comunicação, conforme apresentado na Figura 24. Desta forma, a referida abordagem pode até ser aplicada a modelos de canal e funções de perda para os quais as soluções ótimas são desconhecidas. Motivados pelo fato de que tal abordagem assume que o modelo de canal é diferenciável, Fayçal Ait Aoudia e Jakob Hoydis [6] propuseram estratégias de treinamento diferentes para o transmissor e o receptor sem que seja necessário nenhum tipo de conhecimento a respeito do canal. Na abordagem proposta, o transmissor e o receptor têm acesso às amostras de treinamento (m_τ) e são

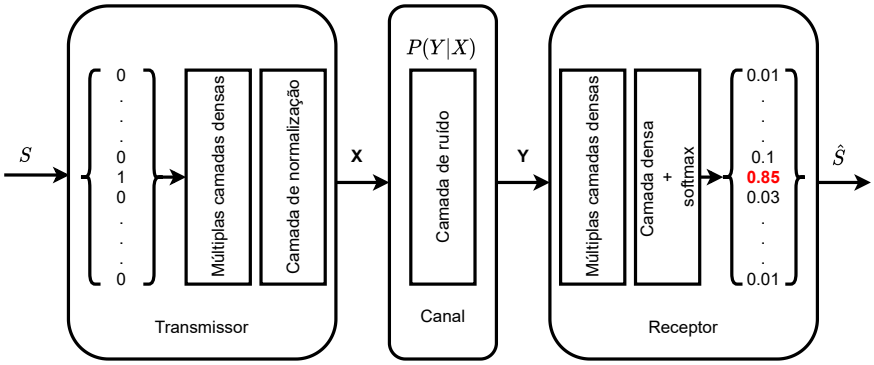


Figura 24: Representação de um sistema de comunicação com canal AWGN através de uma rede *autoencoder*. A entrada S é codificada usando a codificação *one-hot coding* e a saída $\hat{S}$ é a mensagem mais provável da distribuição de probabilidade de todas as mensagens possíveis (adaptada de O'Shea *et al.* [6]).

treinados através de estratégias de aprendizado supervisionado e por reforço, respectivamente, como observado nas Figuras 25 e 26.

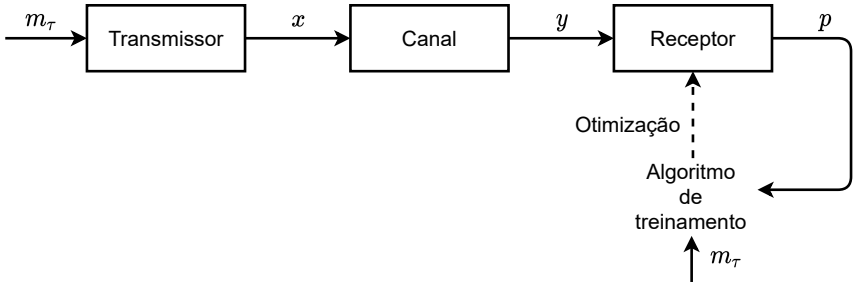


Figura 25: Receptor treinado de forma supervisionada (adaptada de Aoudia *et al.* [6]).

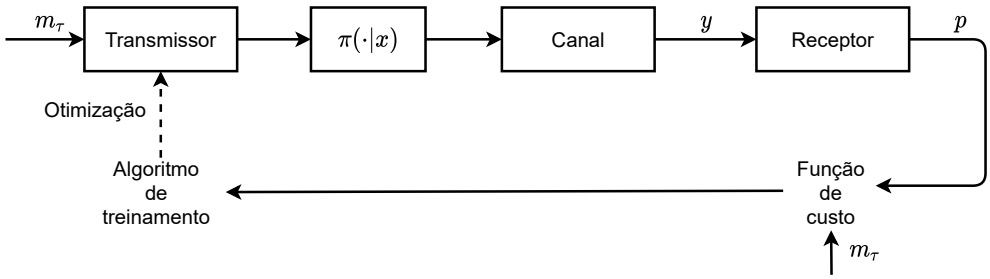


Figura 26: Transmissor treinado como uma tarefa de aprendizado por reforço (adaptada de Aoudia *et al.* [6]).

2.9 Uso de IA para segurança em camada física

Roberto Michio Marques Kagami
robertomk@inatel.br

O aprendizado de máquina tem assumido um papel tão significativo nas questões envolvendo o ambiente de radiofrequência, que as necessidades e as diversas pesquisas de tecnologia relacionadas a esta área tem sido agrupadas e organizadas no que alguns autores se referem como

Radio Frequency Machine Learning (RFML), ou ainda *Radio Frequency Machine Learning Systems* (RFMLS). Em propostas de estruturação [241], os tópicos podem ser reunidos em frentes referentes tanto às aplicações em si, como em questões envolvendo infraestrutura e desenvolvimento de soluções, tratamento e criação dos dados para IA, confiabilidade e segurança. Neste último tema, por exatamente se tratar do meio de radiofrequência, são abordadas técnicas de aprendizado de máquina para a segurança à nível de camada física. Este tópico tem recebido também a denominação de *PHY Security* e tem começado a receber especial atenção, visto que surgem cada vez mais atividades de *hacking*, *cracking* e escuta nos sistemas de telecomunicação como um todo.

Mesmo se constituindo em uma linha de frente em termos de defesa, são em outras camadas que estão presentes os procedimentos correntes de segurança. Há criptografia fim-a-fim para a camada de aplicações, *Secure Socket Layer* (SSL) para a camada de transporte e *Virtual Private Network* (VPN) para a camada de rede, somente para citar alguns exemplos.

Pelas características de heterogeneidade e alta escalabilidade, principalmente em sistemas de IoT, segurança é um dos itens mais críticos para a implementação na camada física das redes 6G. Ao mesmo tempo, é um dos itens menos investigados, talvez por sua complexidade de implementação. Contudo, vem recebendo um número maior de pesquisas mais recentemente em razão de um maior domínio de métodos e algoritmos de inteligência artificial.

Algumas proteções quanto à segurança na camada física podem ser adotadas de forma intrínseca, como utilizando IRS e *Visible Light Communications* (VLC). Basicamente, soluções como esta ajudam a restringir o campo de alcance de prováveis intrusos. Contudo, é possível fazer uso de técnicas de IA para ampliar e fortalecer ainda mais as necessárias proteções. Nas subseções abaixo são tratados alguns tópicos a respeito:

2.9.1 Assinatura de sinal de radiofrequência

Mesmo que mensagens exatamente iguais sejam emitidas por diferentes transceptores via radiofrequência, estes sinais não são idênticos. Eles possuem pequenas e quase imperceptíveis características que os diferem. Porém, estas pequenas variações são detectáveis com técnicas utilizando processamento digital de sinais e aprendizado de máquina. Muito dificilmente esta assinatura, cuja análise é conhecida como *RF Fingerprinting*, poderá ser mimetizada ou clonada por um outro transceptor intruso. Elas provêm de diferenças no balanceamento de componentes *In-phase and Quadrature* (IQ), imperfeições de amplificadores quanto a linearidade de fase e magnitude, diferenças de portadora e sincronismo e outras particularidades provenientes de processos de fabricação. Para uma maior robustez, pode-se ainda somar a isto as características de canal e de ângulo de chegada do sinal transmitido, por exemplo. De posse destes dados é então realizada a identificação do emissor, ou processo também conhecido como *Specific Emitter Identification* (SEI).

Técnicas de estimação de desbalanceamento de componentes IQ tem sido bastante estudadas, como em [242] e, usando redes neurais CNN, como em [243]. São apontadas adversidades nesta estimação com níveis mais altos de ruído. Mas em estudos realizados em [7], por exemplo, há uma proposta de modelo de reconhecimento de emissores com base nesta característica que apresenta boa resposta também com baixos valores de SNR.

Neste trabalho, a coleta de dados de vários usuários considerados autorizados e de um usuário considerado intruso foi realizada durante um período de 5 minutos. Os sinais obtidos dos usuários autorizados foram então filtrados e utilizados para o treinamento de uma rede neural do tipo CNN para a extração de características. Posteriormente, é realizado um processo de

redução de dimensionalidade em que é empregada uma técnica do tipo *t-Distributed Stochastic Neighbor Embedding* (t-SNE). Basicamente, é um método estatístico utilizado para visualização de dados com muitas dimensões, criando um mapeamento destes em âmbito bidimensional ou tridimensional. Após este tratamento dos sinais, há finalmente um processo de clusterização, utilizando o método comumente empregado para este fim denominado *Density-based Spatial Clustering of Applications with Noise* (DBSCAN). Este bloco gera então as referências necessárias para a identificação dos usuários intrusos. A Figura 27 é um diagrama em blocos desta arquitetura proposta.

Figura 27: Ex. de arquitetura de detecção de intrusos por assinatura de sinal [7].

Os resultados deste estudo demonstram uma boa eficência do método proposto. Principalmente pelo fato de a detecção não depender do conhecimento prévio de dados de intrusos ou da inserção destes no treinamento da rede neural, conseguindo uma distinção mesmo com classes ou *clusters* nunca observados.

Várias outras técnicas vem sendo desenvolvidas para identificação de emissores utilizando outras características, como a diferença de portadora, ou *Carrier Frequency Offset* (CFO). Nesta alternativa, uma boa eficiência também foi obtida em trabalhos como em [244]. Da mesma forma, um bom desempenho já foi observado em identificação baseada em não linearidades de amplificação utilizando uma rede neural do tipo CNN, como na proposta de tratamento deste tipo de dados realizada em [245].

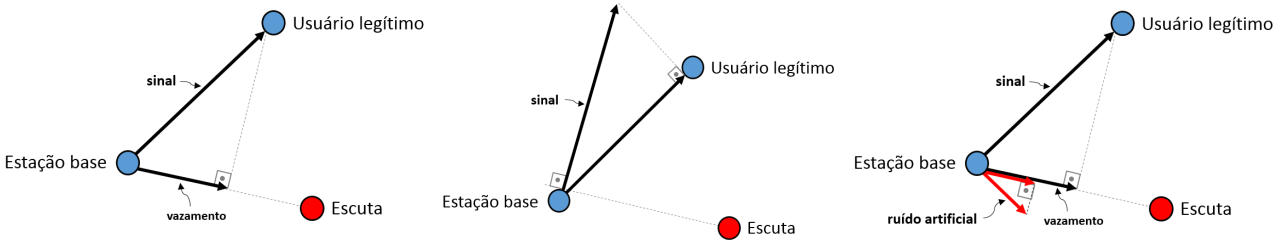
Se forem tomadas as características já descritas, por exemplo, agregando-as para o processamento e análise de um sistema semelhante ao descrito pela Figura 27, é possível prever que resultados ainda mais significativos serão alcançados em termos de eficácia e sucesso na identificação de emissores intrusos e no bloqueio de sinal para receptores em escuta.

2.9.2 Pré-codificadores e ruído artificial

Um dos princípios para aumentar a privacidade de uma mensagem é transmití-la de forma que esta se apresente mais ruidosa para um possível canal de escuta, ou *wiretap channel*, do que para o destino pretendido. Para isso, considerando que os caminhos são diferentes, o emissor pode enviar um sinal de forma que sua característica seja favorável para as propriedades do canal principal e contrária às propriedades de um canal de escuta.

De posse da informação de CSI, no canal principal, é possível encontrar um pré-codificador MIMO que estabeleça uma otimização de taxa de dados neste sentido, ou seja, que propicie o aumento de capacidade do canal. Este aumento gera, em contraposição, uma diminuição da capacidade do canal de escuta. A diferença destas representa a chamada capacidade de sigilo, ou *Secrecy Capacity*. Uma das vantagens deste método, em relação à criptografia, é que este não requer distribuição e gerenciamento de chaves de segurança. Também é interessante por desonerar os equipamentos de usuário quanto ao processamento, aspecto bastante favorável em um cenário de IoT massivo, por exemplo.

Se a informação de CSI do canal de escuta se encontra disponível, é possível direcionar o feixe de tal forma a minimizar o envio de sinal no sentido deste usuário intruso, ou seja, utilizando-se um pré-codificador do tipo *Zero Forcing* (ZF). Este método é ainda melhor que uma conformação de feixe convencional, porém causa uma diminuição da eficiência energética, dado que a potência não é totalmente direcionada para o usuário legítimo. A Figura 28 ilustra os métodos mencionados, incluindo também outra técnica conhecida como ruído artificial, ou *Artificial Noise* (AN).



(a) Conformação de feixe (b) Pré-codificador ZF (c) Ruído Artificial

Figura 28: Técnicas de aumento de capacidade de sigilo.

A dificuldade que se apresenta, então, é a de determinação dos pré-codificadores para que o incremento da capacidade de sigilo seja o maior possível. Dada uma alta complexidade de implementação, métodos utilizando aprendizado de máquina tem surgido, como o apresentado em [246]. Neste trabalho é utilizada uma rede neural profunda do tipo CNN para otimização de um pré-codificador desenvolvido para agregar ruído ao canal de uma possível escuta. É necessário que este ruído seja adicionado na região relativa ao espaço nulo (*null space*) gerado pelo pré-codificador. Assim, o receptor legítimo não percebe esta porção ruidosa do sinal, enquanto que um receptor intruso terá seu sinal degradado.

Este tipo de abordagem sempre requer que os processos de aprendizado tenham uma rápida convergência, como também apresentem boa estabilidade. A inicialização de pesos e *bias* é um importante aspecto neste sentido. Para que a implementação se torne factível na prática em termos de processamento, uma estratégia de estado inicial com aprendizado supervisionado é proposta, seguida de uma segunda fase em que o aprendizado passa a ser não-supervisionado. O desempenho deste esquema de AN foi comparado com o de esquemas convencionais, apresentando resultados sempre superiores, tanto com algoritmos idealmente treinados, como também em cenários com algoritmos onde foi utilizado o aprendizado prático proposto.

Estas técnicas ainda podem ser acompanhadas de outros métodos, como o de seleção de antenas transmissoras, ou *Transmit Antenna Selection* (TAS). A diferença para o método convencional, que busca eficiência do sistema via a redução de transmissores, é que o alvo, para os termos de segurança, passa a ser o incremento da capacidade de sigilo. Neste tipo de esquema, há também algoritmos de aprendizado, como o apresentado em [247]. A solução para um esquema TAS consiste, basicamente, em resolver um problema de classificação multi-classe. Nesta proposta, são descritos dois esquemas de aprendizado de máquina: um baseado em Bayes ingênuo, *Naive Bayes* (NB), e outro baseado em máquina de vetor de suporte (SVM). Os desempenhos dos algoritmos, comparados ao método convencional, são semelhantes, porém a vantagem apontada pelo estudo é quanto à questão da realimentação do sistema, onde a sobrecarga de dados diminui significativamente.

2.10 Gêmeos Digitais para Camada Física de Redes Móveis

Aldebaro Klautau, Cleverson Nahum
aldebaro@ufpa.br,cleversonahum@ufpa.br

Estendendo o conteúdo apresentado anteriormente na Seção 2.3.3, o uso de gêmeos digitais nas redes 6G não se limitará apenas a alocação de recursos das camadas PHY/MAC, mas poderá ser explorado para diversos fins como a previsão de comportamentos de rede móvel, movimentação de usuários e até mesmo o teste de novas políticas da rede para explorar possíveis falhas e evitar a quebra do funcionamento da rede móvel, garantindo uma maior segurança e disponibilidade da rede. Um gêmeo digital é caracterizado como uma cópia digital de alta fidelidade de um ambiente real onde essa cópia digital sincroniza suas informações constantemente com o ambiente formando um ciclo fechado de interação [248].

Em sistemas tão complexos como uma rede móvel 5G e 6G, é importante saber os possíveis resultados de cada politica de rede, algoritmos e parâmetros configurados antes de implementá-los em produção para evitar problemas que possam vir a prejudicar o funcionamento da rede, principalmente com relação a dispositivos que lidam com aplicações sensíveis a perda de pacotes e falhas na rede, como aplicações de hospitais e veículos autônomos. Dessa forma, em um sistema ideal todas as entradas e saídas podem ser preditas com determinado nível de segurança. Se essas saídas e seus impactos pudessem ser preditos com precisão, essa informação seria de grande valia para ajudar na tomada de decisões para encontrar estratégias apropriadas para diferentes países, evitando custos e erros de investimento irreversíveis [249]. Dessa forma os gêmeos digitais surgem como ferramenta estratégica para conseguir realizar a implementação e adaptação de redes móveis a diferentes cenários as tornando mais eficientes.

A figura 29 ilustra o funcionamento de um gêmeo digital aplicado ao cenário de uma cidade física inteligente onde há vários dispositivos conectados e um ambiente bem específico correspondente a uma determinada zona da cidade. Nesse caso todos os dispositivos, estações de rádio-base e antenas são digitalmente representados dentro de um cenário virtual que representa a cidade e veículos, de forma que o gêmeo digital possa apresentar uma dinâmica de funcionamento muito próxima à cidade física. Uma vez estabelecido o sincronismo entre o real e a cópia virtual, o gêmeo digital pode ser utilizado para diversos fins, como para testes preliminares com novos algoritmos e políticas de rede, de forma a avaliar o desempenho e decidir se é vantajoso ou não aplicá-lo no ambiente real. O uso de gêmeos digitais para redes móveis têm ganhado interesse significativo por pare de empresas de telecomunicações como Ericsson [250] e Huawei [251].

O uso de gêmeos digitais permite a geração de dados na cópia digital que são similares ao ambiente real, habilitando o aprendizado de agentes de IA orientado à dados ao invés de baseados em sistemas supervisionados, permitindo a criação de agentes de IA para diferentes cenários e a otimização dos mesmos para localidades específicas como sugerido na sub-subseção 2.3.3.

Uma das principais dificuldades da implementação de gêmeos digitais é a modelagem de cenários reais dentro de cópias virtuais para possibilitar o uso dessa tecnologia. Iniciativas como a plataforma NVIDIA Omniverse [252] permitem a modelagem, simulação e visualização de produtos complexos, aprimorando o funcionamento da rede de acesso de redes móveis, reduzindo o tempo de colocação no mercado e facilitam a introdução de novas funcionalidades. O sistema de modelagem e computação é baseado em recursos de jogos 3D e imagens geradas por computador, permitindo uma modelagem do gêmeo digital fiel ao cenário real sendo virtualizado incluindo detalhes como: alta resolução e geometria urbana ou interna complexa; materiais de superfícies detalhados que influenciam a propagação de RF, como revestimentos metálicos e

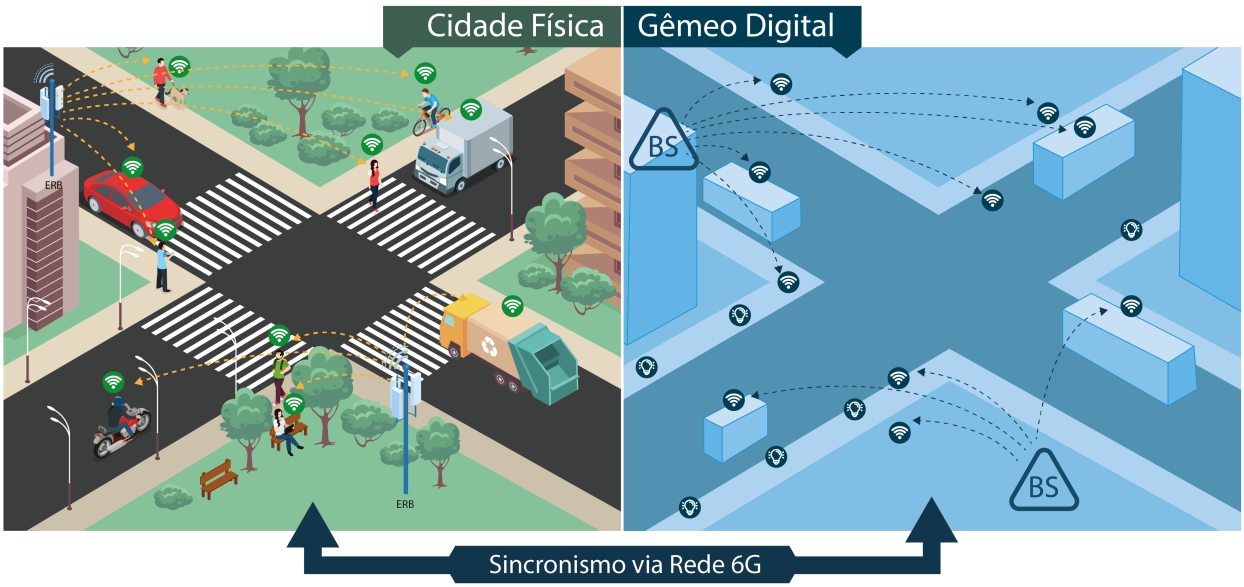


Figura 29: Cidade física e seu gêmeo digital correspondente.

recursos de metal em locais com fábricas; mobilidade de usuários e recursos dinâmicos de cena como tráfego automotivo.

A cidade de Bristol na Inglaterra têm avançado seu progresso para se tornar uma cidade inteligente [253], e com esse fim um gêmeo digital foi criado para visualizar e predizer como os sinais de rádio 5G serão propagados ao longo da cidade, conectando veículos e dispositivos. O modelo do gêmeo digital utiliza um modelo 3D de alta acurácia contendo o terreno, prédios e árvores da cidade, juntamente com modelos avançados de tracejado de raios de propagação de RF de forma a predizer a área de cobertura da rede para cada estação de rádio-base [249]. Uma vez que a rede 5G está implementada, o gêmeo digital pode monitorar a performance da rede continuamente e testar novas funcionalidades na cópia virtual aplicando otimizações em tempo real com o auxílio de métodos baseados em heurística e também agentes de IA.

Outro exemplo de utilização de gêmeos digitais, é o caso do gêmeo digital Spirent 5G [254] que almeja emular uma rede 5G para testar o comportamento e a performance de veículos conectados a rede com um controlador de ambiente realístico usando um modelo de direção 3D. A emulação permite que as fabricantes entendam como os veículos se comportam em diferentes cenários de rodovias e também sob diversas condições de conectividade da rede 5G [249]. Toda a informação gerada pelo gêmeo digital pode ser utilizada por agentes de IA com o intuito de otimizar tanto as funções da rede móvel como também de funcionamento dos veículos.

2.11 Extração de Características Eletromagnéticas

Aldebaro Klautau
aldebaro@ufpa.br

A indústria de jogos e outras estão impulsionando o desenvolvimento de ferramentas sofisticadas para criar mundos virtuais, compostos de Modelos 3-D, motor (*engine*) de física e outros componentes. O cenário 3-D do mundo virtual pode ser criado do zero por modeladores de design 3-D ou a partir de dados importados do mundo real. Por exemplo, o novo Plug-in *Cesium*

para o *Unreal Engine* da *Epic Games*[1] integra informações fotogramétricas obtidas de drones em modelos 3-D disponíveis via *Cadmapper* [2] e outros sites. Isso complementa ferramentas como *Twinmotion*,[3] o qual facilita a construção de mundos virtuais 3-D.

Em [255] as imagens de câmeras 360º são exploradas para predizer informações de propagação do canal. Dentro destes ambientes 3-D, as simulações com traçado de raios ou *Ray-Tracing* (RT) permitem simular canais com boa acurácia.

Essas simulações de RT exigem a identificação do material das superfícies, a fim de simular adequadamente a interação eletromagnética das ondas com os objetos. A disposição e diversidade desses materiais impactam diretamente na qualidade dos canais [256]. Como essa atribuição é feita manualmente, poucos materiais são normalmente adotados. Os próximos parágrafos descrevem pesquisa em andamento para atribuir automaticamente tais materiais a objetos por meio de segmentação semântica com redes neurais profundas.

A segmentação semântica é uma abordagem moderna que realiza a classificação em nível de pixel, e permite determinar tanto a classe de um objeto quanto os limites de cada objeto [257]. As abordagens atuais deste método utilizam o deep learning para superar a segmentação tradicional de objetos, permitindo classificar os pixels não apenas por suas cores, mas também considerando o contexto da região [258]. Em [259], é adotada a segmentação semântica em imagens obtidas através do plug-in Cesium para Unreal a fim de identificar os diferentes tipos de superfícies que compõem o cenário, e então associar automaticamente um material a cada superfície.

A Fig. 30a e a Fig. 30b mostram uma imagem tirada do Césio e sua segmentação, respectivamente. Esta segmentação usou uma implementação no PyTorch de DNN para segmentação semântica, treinada no conjunto de dados de análise de cena MIT ADE20K [260]. Neste exemplo, é possível verificar que o algoritmo foi capaz de determinar o contorno do asfalto. Por outro lado, as regiões referentes a edifícios, automóveis e alguma vegetação, foram associadas à mesma classe. Isso se deve à relativa má qualidade das imagens do Cesium, e precisará ser contornado pelo treinamento de novas redes neurais.

(a) Imagem original de uma rua obtida através do Cesium.

(b) Versão segmentada por DNN da imagem para fins de atribuição automática de propriedades eletromagnéticas aos objetos.

[1] Anúncio de lançamento: https://cesium.com/blog/2021/03/30/cesium-for-unreal-now-available/.
[2] https://cadmapper.com.
[3] https://www.unrealengine.com/en-US/twinmotion.

3 IA em Redes de Transporte

Neste capítulo apresenta-se as principais pesquisas na área de redes de transporte para 6G, utilizando técnicas de IA para aumentar a eficiência, automatização, gerenciamento, redução de custos, entre outras funcionalidades. Inicialmente, é discutida a desagregação da *Next Generation Radio Access Network* (NG-RAN) nas redes Crosshaul, para posteriormente, discutir como deverá ocorrer o posicionamento das *Virtualized Network Functions* (VNFs) no Crosshaul para 6G. Além disso, é abordado como IA é vislumbrado para sistemas com fibra óptica, i.e., meio de comunicação fundamental para as redes Crosshaul. Além disso, com essa mesmas perspectiva de utilização de IA para o futuro das redes 6G, são investigadas as pesquisas sobre as redes terrestres, veiculares, subaquáticas, aéreas e, finalmente, espaciais.

Observa-se que em [261], os autores sugerem três perspectivas de inteligentização que as redes 6G deverão habilitar, são: comunicação, *networking* e computação. Além disso, os autores pontuam requisitos sob a perspectiva da pilha de protocolos do *framework* conceitual *Open Systems Interconnection* (OSI). Considerando a pilha de protocolos, eles posicionam os requisitos das futuras redes móveis e relacionam as camadas e seus serviços mutuamente. Em todos os níveis da pilha de protocolos espera-se atuação de mecanismos de inteligentização da rede. Como na camada física, mecanismos de IA poderão ser amplamentes utilizados para aprimoramento da transmissão de sinais como mecanismos inteligentes de codificação e decodificação de sinais, estimadores inteligentes de capacidade e congestionamento de canais. Os autores vislumbram oportunidade de implementar LSTM para aprimorar a performance de desempenho dos componentes de *encoding* e *decoding* em sistemas NOMA.

Dentre as penitencialidades de técnicas de IA para endereçar problemas específicos nas camadas da pilha, chama-se atenção as penitencialidades na camada de transporte. Nessa camada os requisitos de inteligentização são pontuados para lidar com agrupamento de tráfego, predição de tráfego, roteamento inteligente, fatiamento de rede inteligente e controle inteligente de tráfego. Além disso, construir soluções de inteligência *cross-layer* é um desafio que deverá ser endereçado na realização dos novos *frameworks* arquiteturais das redes móveis [261].

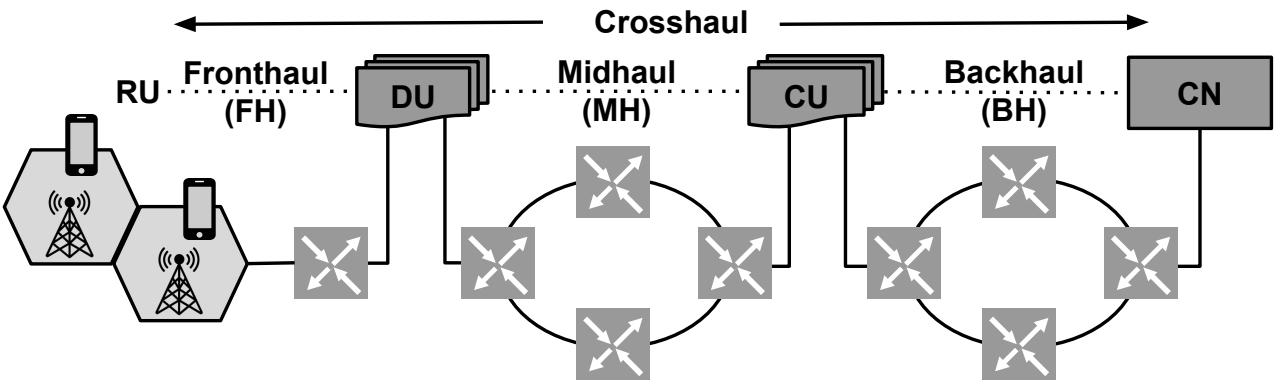
3.1 Redes Crosshaul

Cristiano Bonato Both
cbboth@unisinos.br

A desagregação da RAN entre *Radio Unit* (RU), *Distributed Unit* (DU) e *Central Unit* (CU) passa por forte sinergia com a rede Crosshaul, formando as sub-redes Fronthaul, Midhaul e Backhaul [262] e devem ser uma realizada nas redes 6G. A Figura 30 apresenta a sinergia entre a desagregação da RAN e a rede Crosshaul, detalhando o posicionamento conceitual. Apesar do modelo conceitual, a alta flexibilidade da arquitetura NG-RAN demanda do Crosshaul possibilidades de suporte de desagregação, desde que as camadas atendam aos requisitos necessários. A rede Backhaul é a precursora das redes de transporte para atendimento às redes móveis e foi introduzido com a tecnologia LTE para prover comunicação entre as estações base e o núcleo da rede [263]. Sendo posicionado entre a CU e o núcleo da arquitetura NG-RAN. No que se refere à desagregação proposta na NG-RAN, o posicionamento da rede Backhaul é mantida sem alterações. Entretanto, o posicionamento dinâmico da CU pode acarretar em um tráfego, com caraterísticas da sub-rede Backhaul mais próximos da borda ou do núcleo da rede [264]. Neste ponto, a utilização de IA para auxiliar a tomada de decisão sobre o posicionamento da CU é indicado na literatura [265]. A sub-rede Fronthaul foi introduzida com a arquitetura C-RAN para

o desacoplamento dos nós RRH e BBU. Desta forma, a sub-rede Fronthaul conecta as RRHs a um pool centralizado de BBUs [263]. Entretanto, para a arquitetura NG-RAN, Fronthaul é posicionada nos protocolos mais baixos da pilha de protocolo, entre as opções de desagregação que compõem a RU e os protocolos PHY e MAC. Nesse contexto, a sub-rede Midhaul foi desenvolvida para ser aplicado na nova.

Figura 30: Arquitetura conceitual da rede *Crosshaul*



A evolução das redes de transporte para 6G, incluindo a rede Crosshaul, consolida as arquiteturas multi-camadas, orientadas a dispositivos de domínio óptico e elétrico. Dispositivos estes, concentrados nas tecnologias de Multiplexação por divisão de comprimento de onda (*Wavelength Division Multiplex* (WDM)) para redes de domínio ópticos e switches/roteadores para redes de domínio elétrico, baseados majoritariamente em pacotes. Neste contexto, a arquitetura multi-camada proporciona a utilização de topologias hierárquicas (ou topologias em árvore), com a realização de sobreposição do domínio elétrico sob o óptico. Sendo assim, reduzindo o número de salto entre dispositivos, o compartilhamento de recursos de enlace e, principalmente, o impacto em latência, uma vez que, o domínio óptico opera com latência na ordem de microsegundos [264, 266, 267].

A arquitetura NG-RAN possui dependência direta do Crosshaul e as garantias que o mesmo deve prover para os diferentes requisitos, incluindo latência. Neste sentido, devido a heterogeneidade das tecnologias e possibilidades de topologias de rede, a rede Crosshaul introduz alta complexidade para atingir soluções ótimas. Assim como, estratégias distintas para resolução de problemas similares. Por exemplo, a literatura tem como enfoque amplamente a integração entre as diferentes camadas da rede Crosshaul e os requisitos da sub-rede Fronthaul com a desagregação da RAN [268]. Em um primeiro momento, a motivação foi baseada na arquitetura C-RAN e a centralização máxima da BBU e CU, como investigado em Garcia-Saavedra et al. [269]. Devido a dificuldade de presivibilidade em uma topologia altamente distribuída, assimétrica e com largura de banda impulsionada pela divisão da RAN (camadas de rede Backhaul e Fronthaul), a restrição de latência não foi largamente explorada e deve ser fortemente investigada em redes 6G. Por exemplo, Molner et al. [270], exploraram, além da integração do Backhaul e Fronthaul, a otimização de caminho e o posicionamento da RAN. Além disso, Murti et al. [271] apresentam uma abordagem de DRL para otimizar a divisão de funções em uma RAN. Neste sentido, a latência tem papel central, como restrição, para o posicionamento da CU e, consequente, distribuição da DU. Além disso, um direcionamento para a utilização de recursos ópticos foi endereçado por Musumeci et al. [272]. Neste trabalho, os autores propuseram uma agregação do tráfego do Fronthaul em elemento de domínio óptico para flexibilizar o

posicionamento da CU (neste estudo, CU e DU integradas), orientado pelas tecnologias *Optical Transport Network* (OTN) e WDM. Tendo em vista que o domínio óptico opera com latência na ordem de microsegundos. A topologia definida para o Crosshaul na arquitetura é fundamental para o desempenho do sistema 6G. Por exemplo, a topologia deve levar em consideração os conceitos multi-camada baseados em redes ópticas desenvolvidas para as redes de transporte, orientada pelas três sub-redes que compõe o Crosshaul.

3.2 Posicionamento e implantação de VNF

Cristiano Bonato Both
cbboth@unisinos.br

Apesar da evolução vislumbrada pela virtualização da NG-RAN e sua distribuição, o melhor posicionamento das VNFs ou VNF-P da NG-RAN é considerado um dos principais problemas e desafios para as redes 6G [273]. O problema de posicionamento de VNFs implica em escolher qual o melhor local para implantar as instâncias VNFs nos recursos de uma rede física [274]. O posicionamento das VNFs é um problema combinatório desafiador, pois envolve um grande número de decisões de posicionamento discretas. Esse tipo de problema de decisão é conhecido como NP-*Hard*, devido a dificuldade de encontrar um cenário ideal ou próximo ao ideal [273]. Por exemplo, se n for o número total de VNFs e m for o número total de recursos computacionais, o número de mapeamento possível poderá ser m^n [275].

O tópico de posicionamento é estudado amplamente em ambientes de computação em nuvem, uma vez que é considerada uma operação crítica, devido a tomada de decisão para determinar o melhor recurso físico apropriado para hospedar as máquinas virtuais (incluindo VNFs) [276,277]. Para a computação em nuvem, os requisitos considerados mais relevantes são a eficiência energética, os recursos computacionais (rede, armazenamento e computação) e o suporte à qualidade de serviço. O problema do posicionamento pode ser dividido em duas partes: a primeira parte é a admissão de novas solicitações para o provisionamento e a alocação nos recursos computacionais, enquanto a outra parte é a otimização (por exemplo, monitoração de desempenho e processos de migração). A complexidade do posicionamento tem relação direta com os requisitos de entrada de uma instanciação, sendo estes, muitas vezes, imprevisíveis [275].

Concentrando no problema VNF-P, a entrada consiste em uma cadeia de funções de serviços (*Service Function Chain* (SFC)), que é composta por um conjunto de VNFs. Enquanto a rede de substrato (recursos de rede e computacionais) fornece as restrições físicas em termos de largura de banda e capacidade. Nesse contexto, o termo capacidade não está relacionado apenas a recursos computacionais, como número de núcleos de *Central Processing Unit* (CPU) e memória. Pelo contrário, refere-se aos recursos de encaminhamento e processamento de pacotes. Cada cadeia de serviço é um fluxo de pacotes de rede que flui através de uma sequência de VNFs a uma determinada taxa. O tráfego de rede para uma determinada cadeia de serviços deve ser orientado a uma pré-definida sequência de VNFs [277].

Cada cadeia de serviço em um problema de posicionamento de VNFs pode ser vista como uma rede virtual a ser mapeada em uma rede física. No entanto, o posicionamento dos VNFs difere ao passo que cada nó (VNF) em uma rede virtual pode ser mapeado para várias instâncias que são colocadas em nós diferentes na rede física. O posicionamento do VNF combina os problemas de mapeamento de cadeias de serviço em redes virtuais e mapeia as redes virtuais resultantes na rede física. Do ponto de vista arquitetural do *Network Functions Virtualization* (NFV) *MAnagement and Network Orchestration* (MANO), a solução do problema de posicionamento

dos VNFs deve ser desenvolvida no *NFV Orchestrator* (NFVO), sendo este, o componente responsável por manusear a alocação de recursos na *NFV Infrastructure* (NFVI) [273].

Em redes 6G, o tempo gasto em uma solução de otimização tende a ser proibitivo e o emprego de métodos de otimização não exatos se tornam desejável. Por exemplo, observa-se que a estrutura do problema de posicionamento de VNFs apresenta características que habilitam o emprego de métodos baseados em IA, que podem lidar com problemas complexos com ou sem a necessidade de se apoiar em modelos formalmente definidos [278]. Tendo em vista que o problema de posicionamento de VNFs pode ser modelado como uma otimização combinatória, como em Murti et al. [271], uma solução baseada em DRL se mostra uma possibilidade atraente para redes 6G. Por exemplo, a função objetivo e as restrições do problema podem ser aprendidas por um agente, enquanto interage com um ambiente que replica o problema, mapeando as recompensas recebidas pelas ações tomadas diante de observações do estado, buscando construir uma política de decisão que visa maximizar tal recompensa no longo prazo. Desta forma, o desafio envolve a definição do espaço de observações, o espaço de ações para o agente, e a função de recompensa que agrega os objetivos e restrições, eliminando a necessidade da própria formulação matemática do problema.

3.3 IA para sistemas com fibra óptica

Luiz Augusto Melo Pereira, Luciano Leonel Mendes, Arismar Cerqueira Sodré Junior
luiz.augusto@dtel.inatel.br,lucianol@inatel.br, arismar@inatel.br

A convergência de sistemas ópticos e sem fio representa uma solução chave tanto para 5G quanto para 6G, enquanto o AM, as comunicações com luz visível e em Terahertz ganham espaço [279–281]. A tecnologia *Radio over Fiber* (RoF) vem sendo considerada como uma solução potencial para suportar comunicações de alta vazão e para a implantação de sistemas que demandem múltiplas células com processamento centralizado dos sinais de RF. A centralização de funções da rede permite o compartilhamento de recursos, favorece uma alocação dinâmica e permite simplificar as operações e manutenções da rede, resultando em redução nos custos operacionais do sistema.

A transmissão de sinais utilizando RoF permite combinar a capacidade dos sistemas ópticos com a mobilidade e a flexibilidade das comunicações sem fio existentes. Dentre as vantagens da transmissão por fibras ópticas, destacam-se altíssima largura de banda, baixo consumo de potência, baixa atenuação, imunidade a interferências eletromagnéticas e capacidade para operação de múltiplos serviços. Os sistemas RoF permitem empregar enlaces ópticos para distribuir sinais de RF para diversas ERBs, necessitando apenas dos estágios de conversão óptico-elétrica, amplificação e transmissão.

A Figura 31 descreve os principais mecanismos de degradação em sistemas óptico-sem fio que empregam modulação direta e externa da portadora óptica, incluindo as não-linearidades da fibra, a dispersão cromática e a resposta não-linear de moduladores eletro-ópticos. Os sistemas RoF são muitas vezes utilizados em redes de acesso, por isso não demandam o uso de potências ópticas elevadas. Isto resulta em uma redução das degradações causadas pelos efeitos não-lineares da transmissão por fibras ópticas e a dispersão cromática [282]. Portanto, a resposta não-linear do modulador eletro-óptico é a principal fonte de degradação dos sistemas RoF [283–286]. Particularmente, em sistemas multi-banda a resposta não-linear do modulador eletro-óptico causa intermodulação entre as subportadoras e modulação cruzada entre os componentes I e Q do sinal transmitido [282, 287].

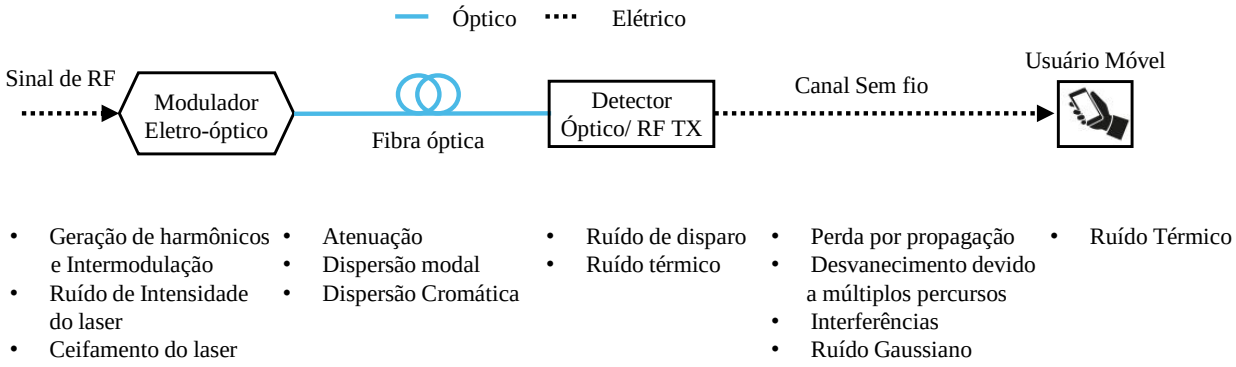


Figura 31: Principais fontes de degradação em sistemas óptico-sem fio (adaptado de [8]).

Diversas soluções na literatura tratam da linearização de sistemas RoF, tanto no domínio óptico, quanto no elétrico [288]. Como exemplo de esquemas de linearização no domínio óptico convém citar as seguintes técnicas: *feedforward* [289]; *dual parallel modulation* [290]; *mixed polarization* [291]; *dual electro-absorption* [292]; *gain modulation scheme* [293]. Embora as técnicas citadas demonstrem ser capazes de reduzir a emissão de espúrios e aumentar a região linear de operação do sistema, a complexidade de implementação limita o uso destes tipos de linearizadores. De maneira geral, as técnicas de linearização no domínio óptico requerem a adição de diversos componentes, aumentando o custo de implementação, além de necessitarem de um controle preciso, adaptativo e de resposta rápida em diversos pontos de operação dos linearizadores. Tais restrições impulsionaram as pesquisas de técnicas de linearização no domínio elétrico, como por exemplo o uso de processamento digital de sinais, visando reduzir a complexidade do processo de linearização de sistemas RoF.

Dentre as técnicas que empregam processamento digital de sinais, a pré-distorção digital tem sido reconhecida como uma solução eficiente para combater as distorções dos sistemas RoF [294–297]. Em [298], os autores propuseram a linearização de um sistema RoF aplicado como *fronthaul* de uma arquitetura de rede centralizada 5G. Em tal rede, as não linearidades são provenientes da resposta não linear do modulador eletro-óptico e dos amplificadores elétricos das ERBs, que por sua vez, distorcem consideravelmente o sinal, especialmente para as formas de onda de sinais 5G, que apresentam elevada PAPR. O uso de pré-distorção digital possibilitou mitigar os efeitos de não linearidades dos amplificadores, os quais demostraram elevado nível de degradação, bem como as não linearidades inerentes do processo de modulação da portadora óptica. Recentemente, os pesquisadores do Instituto Nacional de Telecomunicações (Inatel), propuseram a utilização de processamento digital de sinais para geração e pré-distorção de sinais em um sistema 5G distribuído com o auxilio de uma rede óptica de um provedor local de Internet [299]. Os experimentos práticos demonstraram a viabilidade da solução proposta para a linearização, distribuição e transmissão de sinais 5G.

Neste contexto, os avanços na capacidade computacional, o aumento na quantidade de conjuntos de dados e o crescimento sem precedentes na complexidade dos sistemas de comunicações, têm impulsionado as pesquisas que envolvem o uso de aprendizado de máquina em diversos níveis de sistemas com fibras ópticas. As técnicas de processamento de sinais que empregam algoritmos de aprendizado de máquina, ao contrário das técnicas convencionais, são baseadas em dados. Ou seja, durante o treinamento, as características e comportamentos das distorções do sistema são aprendidos e, posteriormente, utilizados para mitigar tais degradações,

visando melhorar o desempenho do sistema. Uma vantagem importante de métodos baseados em aprendizado de máquina, em relação aos convencionais, consiste na excelente aderência dos algoritmos para operar com sistemas não-lineares e a capacidade de compensar os efeitos conjuntos de diversos tipos de distorção [300].

Diversos algoritmos de aprendizado de máquina têm sido propostos na literatura usando diferentes técnicas e com diferentes objetivos. Os autores de [301] apresentaram a classificação de técnicas de aprendizado de máquina que são aplicadas em diversos casos de uso. Dentre estes casos cita-se: mitigação de não-linearidades; reconhecimento da ordem de modulações; monitoramento da qualidade de transmissão; gerenciamento da rede; detecção e identificação de falhas. De maneira abrangente, Khan e colaboradores relataram em [302] uma perspectiva de uso de aprendizado de máquina em redes ópticas. O objetivo do trabalho foi identificar os potenciais gargalos das redes ópticas, os quais possibilitariam a penetração de algoritmos de aprendizado de máquina para o aumento de desempenho e capacidade da rede.

Mais especificamente, em sistemas com fibras ópticas que empregam a tecnologia RoF, os trabalhos presentes na literaturam são mais escassos. Uma grande parcela destes trabalhos tratam do uso de aprendizado de máquina para mitigação de não linearidades do sistema RoF. Em [300] os autores discutiram sobre a aplicação de técnicas de aprendizado de máquina para compensar dispersões do sistema e não-linearidades do RoF. As não-linearidades são provenientes do processo de modulação da portadora óptica e receberam atenção especial. Adicionalmente, os autores agruparam diversos algorítimos de aprendizado de máquina, utilizando redes neurais ou não, incluindo o *K-means*, *K-nearest neighbours*, *Suport vector machine*. Tais algorítimos e técnicas podem ser aplicados como equalizador, decodificador ou o demultiplexador, os quais resultaram em melhorias no desempenho do sistema RoF quando comparado à técnicas convencionais.

Trabalhos recentes demonstraram o poder das redes neurais para a compensação de distorções presentes no sistema RoF. Em [303], Najarro e colaboradores descreveram um modelo baseado em redes neurais para a compensação de distorções por meio da obtenção da resposta inversa do sistema RoF. Os autores comprovam a eficácia do método, por meio de simulações e da redução da magnitude do vetor de erro do sistema. Já em [304], os autores relataram uma solução de pré-distorção digital baseada em redes neurais com retro-propagação para sistemas RoF. A eficácia do modelo proposto é comprovada por meio da comparação com técnicas que empregam séries de Volterra. Os resultados demonstraram que o modelo baseado em redes neurais é capaz de reduzir o *Error Vector Magnitude* (EVM) e a emissão fora da faixa de interesse. Os autores de [305] reportaram a utilização de computação evolutiva para configurar de maneira adaptativa os parâmetros de transmissão de um sistema RoF, resultando em melhoria na qualidade do sinal recebido.

Além das distorções resultantes da geração de produtos de intermodulação, sistemas RoF multi-banda apresentam distorções adicionais entre os componente I e Q dos sinais transmitidos. Esta degradação é conhecida como *Cross-modulation Distortion* (XMD) e resulta em compressão na constelação do sinal recebido. Alguns algorítimos de aprendizado de máquina, não somente os que empregam redes neurais, vêm sendo propostos na literatura para combater este tipo de degradação [306]. Liu e colaboradores em [307] observaram que a XMD torna-se mais significante com o aumento da ordem de modulação de sinais. Logo, autores propuseram uma rede neural com função de ativação com multi-níveis, a qual é mais propícia para sistemas que empregam sinais com elevada ordem de modulação. Além disso, a rede neural proposta foi capaz de mitigar os efeitos da XMD e suprimir efeitos de rotação de fase na constelação.

Ainda considerando sistemas multibandas, em [308] foi apresentada uma rede neural de pós

distorção, aplicada após a fotodetecção do sinal de RF, para reduzir os efeitos de intermodulação e da XMD. Estes dois tipos de distorções limitam a região de operação livre de espúrios. A rede neural proposta foi capaz de simultaneamente reduzir os produtos de intermodulação e XMD, resultando em um aumento de 34 dB na região linear de operação do sistema RoF em questão. Adicionalmente, os autores declaram que a rede neural proposta possui elevada capacidade de generalização e pode ser potencialmente empregada para reduzir outros tipos de degradações lineares e não lineares.

Em um cenário de operação com múltiplos usuários, os efeitos degenerativos causados por produtos de intermodulação e XMD tornam-se ainda mais severos e aumentam consideravelmente as interferências entre os usuários. Contudo, os autores de [309] demostram que é possível reduzir tais interferências utilizando uma rede neural equalizadora. Resultados numéricos e experimentais comprovam a eficácia da rede neural em mitigar as interferências entre múltiplos usuários bem como a redução da taxa de erro de bits dos sistema.

3.4 Redes Terrestres

Rodrigo Moreira
rodrigo.moreira@ufu.br

Em [9] é proposto um *framework* arquitetural dividido em quatro blocos estruturais: Plataforma, Funcional, Especialização e Orquestração. A divisão em quatro bolocos está ilustrada na Figura 32. Em cada bloco projeta-se utilização de mecanismos de inteligentização. O Bloco Plataforma compreende componentes de nuvens heterogêneas para prover um ambiente aberto para inovação com escalabilidade. Além disso, o bloco Plataforma prevê que uma arquitetura para redes *6G* deve ser centrada em dados e possuir mecanismos de aceleração em *hardware*. O Bloco Funcional considera a convergência RAN-CORE, células livres e uma arquitetura centrada na informação e em inteligência artificial. O Bloco Especialização contêm mecanismos para oferecer fatiamento de rede extremo, isto é, oferecendo às aplicações parcela da rede. O bloco de Orquestração possui um ecossistema para lidar com o gerenciamento e monetização de recursos.

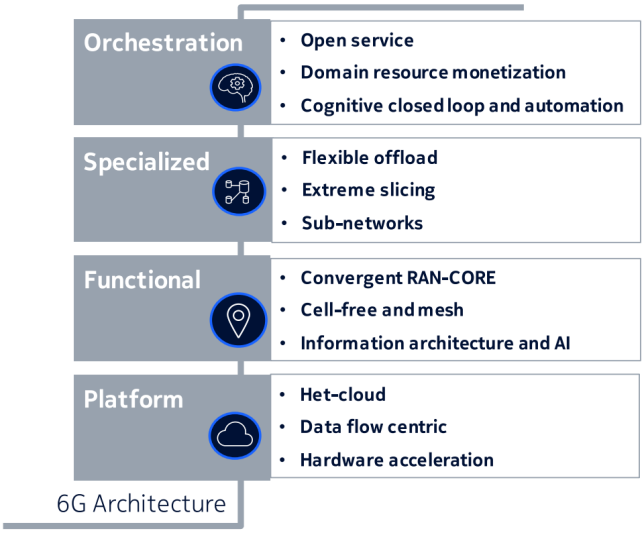


Figura 32: *Framework* Arquitetural *6G*: Blocos Estruturais [9].

A Inteligência Artificial toca todos esses quatro blocos arquiteturais propostos e ilustrados na Figura 32. Em especial, a camada de rede e core se beneficiam dos métodos de inteligência artificial porque, conforme se vislumbra para a rede *6G*, haverá um aumento de dados devido a quantidade de dispositivos e o suporte da arquitetura para subredes. Logo, mecanismos cognitivos de roteamento entre pontos finais deverão ser baseados em fluxo de dados. Espera-se haver uma separação de contextos de transação e sessões do contexto da aplicação, como: localização, aplicação, dispositivo utilizado e funções de rede utilizadas. Nesse sentido, torna-se mandatório uma rede com mecanismos cognitivos para roteamento de trafego e colocação ótima de funções de rede e serviços. A camada de orquestração também deverá conter componentes de inteligência artificial, para lidar com tarefas de automação e gerenciamento inteligente.

3.4.1 Redes Veiculares

A comunicação entre veículos, sobre redes *Vehicle to Vehicle (V2V)*, tem recebido significativos avanços da comunidade científica. Especialmente, no desenvolvimento e avaliação de técnicas de aprimoramento das comunicações nesse contexto. Infraestruturas de comunicação foram propostas, como o *Dedicated Short-Range Communications* (DSRC), LTE e a extensão LTE-V, que é uma rede LTE especializada para comunicação veicular padronizada pela entidade 3GPP. Uma particularidade dessas infraestruturas de rede especializadas em contextos específicos é a alta flexibilidade, capacidade de respostas rápidas e controle de colisão sensitivo.

Propostas fundamentadas em IA para endereçar os desafios impostos à infraestruturas de redes V2V baseiam-se predominantemente em Alocação de Recursos de Rede e Controle de Tráfego de Rede. Além das técnicas de IA serem capazes de lidar com Alocação de Recursos de Rede de forma inteligente, elas oferecem suporte tecnológico para endereçar problemas específicos como: Reposta Rápida, quando o ambiente muda abruptamente causando *handoffs*; Modelagem Acurada do Ambiente, que prevê utilização de modelos matemáticos para mensurar a utilização de recursos; Auto-adaptação da rede, provendo técnicas para exploração adaptativa e gradual de recursos subjacentes.

Para lidar com problemas de Controle de Tráfego de Rede, técnicas de IA são utilizadas, considerando que os requisitos de atrasos extremamente baixos e alta capacidade de rede não são plenamente satisfeitos pelas técnicas propostas e estabelecidas do estado da arte. Algoritmos convencionais de roteamento como *Open Shortest Path First* (OSPF), *Routing Information Protocol* (RIP) e *Destination-Sequenced Distance-Vector Routing* (DSDV) foram tentativas para endereçar os desafios de controle no nível de rede.

Outros desafios como previsão de estado futuro, que controla o tráfego de rede considerando um estado futuro; recuperação de correlação, refere-se a recuperação de um estado da rede; exploração proativa e auto-aprendizado, modificar a estratégia de controle de tráfego é demorado e propenso a erros. Por isso, vislumbra-se para o 6G mecanismos baseados em IA que explorarão a modificação do ambiente e exercerái influência sobre o plano de controle proativamente [36].

3.5 Redes Subaquáticas

Rodrigo Moreira
rodrigo.moreira@ufu.br

A integração tridimensional que se discute para o futuro das redes móveis, que incluem incluindo espaço, terra e subaquáticos, traz consigo desafios que deverão ser endereçados no âmbito da especificação e padronização da rede 6G. Em especial, as redes subaquáticas deverão

prover conectividade ubíqua nesses ambientes para prover minimização de fronteiras de conectividade habilitando acessos de qualquer lugar. Aplicações de contexto militar ou comercial poderão explorar essa conectividade com novos serviços, mas a gerência e operação de redes com essas características requer incluir novos blocos.

No nível de rede, uma diferença distinguível entre a rede móvel 5G e sua sucessora, 6G, é a inteligentização, que considera a incorporação de mecanismos inteligentes habilitados por inteligência artificial e aprendizado de máquina. A inteligentização e integração das dimensões arquiteturais de rede é fundamental para habilitar a conectividade para serviços de Internet em ambientes subaquáticos. Prover a comunicação nesses ambientes requer novos mecanismos que lidem adequadamente com as diferenças de propagação. Nesse ponto, uma rede de transporte para o ambiente subaquático valerá-se de técnicas de IA para endereçar os problemas físicos apontados.

A IA comporá substancialmente o *framework* das novas gerações de redes móveis em especial no transporte. Assim, espera-se que para as redes subaquáticas, as tecnologias de IA potencializem mecanismos como de alocação de recursos, previsão de tráfego, personalização de conectividade, melhoramentos de segurança e redundância de transmissão. No nível da rede de transporte, questões como gerenciamento de mobilidade e conectividade também cedem espaço para a inteligentização. Além disso, requisitos como orquestração e gerenciamento dinâmico de fatias de rede, realização de redes autônomas e alocação ótima de recursos multinível são apresentados à comunidade científicas como problemas ainda não endereçados.

Para lidar com eles, espera-se utilizar conceitos emergentes como *Software-Defined Networking* (SDN), NFV e computaçãod e borda combinados e eventualmente estendidos para atenderem o requisito de inteligentização. Aplicando técnicas de inteligência artificial como DNN, *Self-Organizing Maps*, *Enhanced Q-Learning*, *Deep Reinforcement Learning*, vislumbra-se a possibilidade de realizar uma integração tridimensional de redes , inclusive no nível de transporte, de forma a tornar o gerenciamento e orquestração inteligentes e sensíveis a requisitos refinados dos usuários [310].

3.6 Redes Aéreas

Rodrigo Moreira
rodrigo.moreira@ufu.br

Uma tecnologia promissora para novas arquiteturas 6G é a *Unmanned Aerial Vehicles* (UAVs). Ela permite confiabilidade e flexibilidade na comunicação, sobretudo em locais com catástrofes da natureza. Aplicações civis ou militares podem valer-se dos UAVs dado sua possibilidade de escalar a depender do contexto. Ocorre que a realização desses mecanismos exigem transmissão de alta velocidade, confiabilidade e auto cura.

Algumas abordagens como [311–314] mitigaram os desafios inerentes a esse tipo de rede. No que toca a inteligência artificial como componente fundamental nas na rede 6G, considera-se como desafio a aplicação de mecanismos para prever a mobilidade de usuários e dos UAVs. Além disso, a distribuição de carga, que poderá habilitar os UAVs de forma dinâmica e otimizar dinamicamente sua trajetória. Além disso, vislumbra-se a possibilidade de utilizar inteligência artificial para prover mecanismos de cache em UAVs e gerenciadores inteligentes de energia. A IA mostra-se como um componente fundamental para aplicar inteligência na integração nessa dimensão de rede que é considerada no 6G.

3.7 Redes Espaciais

Rodrigo Moreira
rodrigo.moreira@ufu.br

A integração de redes do que se prevê com a ISTN tem sido considerada como componente fundamental do *framework* da rede 6G. A Figura 33 ilustra uma ISTN e a integração que se espera. Especificar métodos e tecnologias para construir uma ISTN de alta capacidade e baixo custo é o desafio preponderante da comunidade e industria nesse contexto. *Backbones*, isto é, a rede de trasporte aérea possue limitações de velocidade dos *links* e frequência.

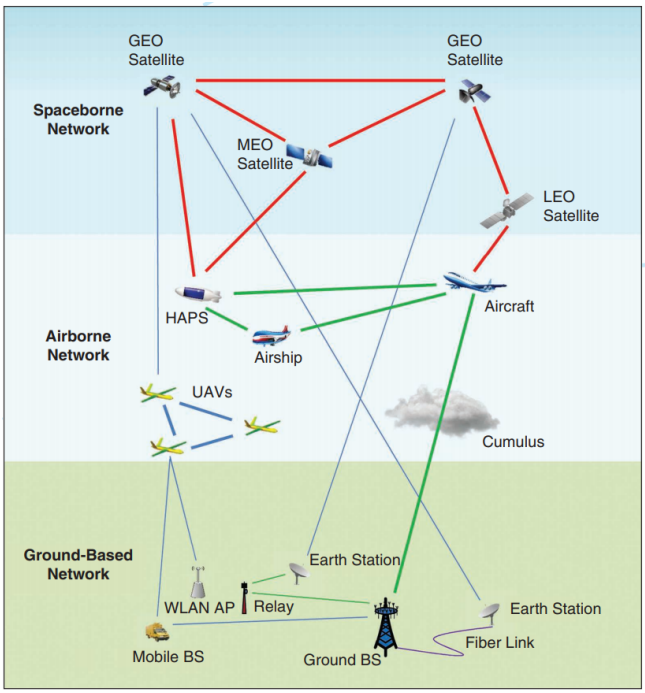


Figura 33: Uma arquitetura ISTN típica [10].

Esforços na realização de redes no espaço já ocorrem. A entidade *Defense Advanced Research Projects Agency* (DARPA), no contexto do projeto *Optical Experimental Network Experiment Program* alcançou taxas de transmissão entre satélites de 10 *Gbps* no espaço [10]. Nesse sentido, o aprimoramento dessas técnicas de transmissão e mecanismos de inteligentização na rede de transporte no espaço permitirão que arquitetura 6G integre no seu bloco estrutural as redes do espaço, para cumprir dentre tantos requisitos maior cobertura que a 5G.

Ocorre que no contexto de transporte de dados e de convergência dos meios de conectividade, desafios são reconhecidos como o gerenciamento de mobilidade. Existem diversos padrões de mobilidade que precisam ser lidados de maneira transparente pelos mecanismos de controle da rede. Além disso, protocolos de transporte são amplamente utilizados na comunicação de redes terrestres e espaciais. Mas, os protocolos das redes terrestres baseiam-se predominantemente na premissa de melhor esforço, assim como na rede de transporte da Internet. Redes de satélites possuem uma topologia determinística e dinâmica. Além disso, os tempos de viagem *Round Trip Time* (RTT) são maiores, o que pode impactar substancialmente o comportamento dos mecanismos já estabelecidos em redes terrestres, como o *Transmission Control Protocol* (TCP).

O roteamento também impõe-se como um desafio para as redes do espaço. Habilitar a conectividade em longas distâncias faz a tarefa de roteamento inter-satélite ser mandatório. Os nós desse tipo de rede possuem limitação de espaço e energia, o que implica na capacidade de processamento e armazenamento. Técnicas de IA perpassa esses desafios na perspectiva micro e macro. Na perspectiva micro vislumbra-se a utilização e inteligentização da operação e gerenciamento dos componentes que realizam as redes do espaço. Em uma perspectiva macro, técnicas de IA são promissoras porque se coloca como uma ferramenta para automatizar e otimizar tecnologias já existentes, como fatiamento de redes e monitoramento do serviço e adaptação da fatia de rede a depender da condição corrente.

Além disso, no contexto de redes do espaço, vislumbra-se a possibilidade de integrar mecanismos de IA para facilitar a análise e predição de mobilidade, além de habilitar técnicas de agrupamento de recursos. Além disso, soluções baseadas em IA são vistas como fundamentais para a realização da ISTN.

4 IA no Núcleo e na Orquestração de Recursos e Serviços

Neste capítulo, discute-se a utilização de IA no núcleo das redes 6G. Inicialmente, a elasticidade e balanceamento de carga são abordados considerando as novas funções do núcleo introduzidas no *Release* 15 do 3GPP [315], tais como *Access and Mobility Management Function* (AMF), *User Plane Function* (UPF), *Session Management Function* (SMF), entre outras. Posteriormente, o gerenciamento e orquestração dessas funções são discutidos em uma perspectiva de serviços fim-a-fim, utilizando o conceito de *network slicing*. Finalmente, destaca-se, neste contexto, a utilização de VANTs para 6G, considerando a computação de borda da rede e a utilização de IA como elemento habilitador para suportar as características dinâmicas desses dispositivos.

4.1 Elasticidade e Balanceamento de Carga no Núcleo

Cristiano Bonato Both
cbboth@unisinos.br

A sexta geração de telecomunicações móveis está começando a ser idealizada e diversos aspectos da sua implementação estão em aberto. Essa nova geração terá como principal objetivo atender a realidade que está por vir, *i.e.*, alavancada pelos dispositivos IoT, com previsão para 2025 de mais de 79,4 zettabytes trafegados por ano e cerca de 41,6 bilhões de dispositivos conectados [316]. Pensando nesses aspectos e visando estar preparado para essa nova realidade, o 3GPP lançou a *Release* 16. Um dos itens de adequação é a arquitetura baseada em serviços para o núcleo que, dentre outras características, desacopla os serviços de forma que cada um tenha uma responsabilidade específica, facilitando a multiplicação de serviços para atender os mais diversos e dinâmicos cenários. Além disso, essa arquitetura é subdividida em plano de controle e plano de dados, composta por serviços com responsabilidades bem definidas e todos definidos em nível lógico, ou seja, em software [315]. Uma vez estando em software, estes serviços podem ser implementados em estruturas convencionais, como *datacenters* com hardware genérico ou mesmo em nuvens.

Essa migração apresenta uma relação custo-benefício implícita, pois de um lado temos que, quanto mais perto do hardware, menor é a utilização de recursos computacionais, porém com alto custo. Já no outro lado, quanto mais abstrato em software, menor o custo, porém com maior consumo de recursos computacionais. Esse é um problema comumente resolvido através da utilização de elasticidade, pois a alocação dos recursos é dinâmica e diretamente proporcional às necessidades do sistema em um determinado instante ou período.

Existem trabalhos na literatura que exploram a utilização de replicação de serviços para permitir o paralelismo distribuído de ações do núcleo da telecomunicação móvel [316]. Pensando no contexto atual, é imprescindível que a comunicação do núcleo com a rede de acesso seja totalmente transparente. Esta transparência também é válida, quando se fala da NG-RAN, pois neste caso todos os softwares já desenvolvidos, bem como os que ainda serão, precisariam se adequar ao modelo, o que não é positivo, em termos de compatibilidade. Dessa forma, nas futuras redes 6G é fundamental projetar balanceadores de carga inteligentes que habilita a movimentação de serviços de um computador para outro, para deixar o serviço mais rápido ou minimizar o tempo de resposta [317, 318].

Outro ponto extremamente importante é elasticidade dos serviços do núcleo, de tal forma que a utilização de recursos computacionais possa ser reduzida. Dentre os trabalhos em que

houveram a aplicação de elasticidade, os trabalhos de [319–322] utilizaram o modelo de elasticidade reativa, que só realiza uma ação de elasticidade, quando algum indicador é atingido. Entretanto, este modelo pode gerar ineficiência na utilização dos recursos, devido a realização de uma alocação ou desalocação tardia. Desta forma, entende-se que o modelo que introduz maior benefício é o proativo utilizando técnicas de IA, que busca antever o comportamento da carga de trabalho e permite que a reconfiguração dos recursos possa ser realizada de forma antecipada e, assim, reduzindo a utilização de recursos.

O trabalho que utiliza proatividade foi proposto por [323]. Este modelo, além de utilizar o núcleo *Evolved Packet Core* (EPC), é baseado em épocas de tamanho fixo e a ação de elasticidade ocorre somente no início de cada época. Desta forma, se o comportamento se modificar no meio de uma época, o sistema apenas será redefinido na próxima época, podendo gerar ineficiência na utilização de recursos e na capacidade de atendimento. Uma forma de contingenciar este comportamento seria ter épocas de tamanho pequeno, porém, neste modelo o cálculo de predição só considera valores obtidos na última época, logo esta estratégia pode trazer perda de precisão na predição. Entende-se por oportunidade de pesquisa, um modelo de elasticidade horizontal proativo para o núcleo *Service-Based Architecture* (SBA) previsto no *Release* 15 e que deverá ser adotado em redes 6G. Esse modelo deve ser totalmente transparente para a rede de acesso e que, no intuito de melhorar a utilização de recursos, efetue as ações de elasticidade o mais próximo possível das necessidades do sistema. Além disso, essa elasticidade deve ser aplicada para diferentes serviços do núcleo, como AMF, UPF, SMF, entre outros serviços críticos.

Para prover elasticidade proativa, uma estratégia normalmente utilizada é baseada em tendência para determinar a carga futura. Dessa forma, pode-se considerar um histórico de medições anteriores para a definição da tendência de carga (aumento ou redução). Por exemplo, pode-se utilizar técnicas como *AutoRegressive Integrated Moving Average* (ARIMA), por ser um modelo que representa uma série temporal baseada em tendência que, diferentemente de outras séries, leva em consideração a variação do comportamento ao longo do tempo. Entretanto, deve-se investigar outras técnicas de IA para determinar a melhor elasticidade proativa das redes 6G. A solução de elasticidade inteligente deve estar alinhada com as atuais e futuras arquitetura padronizadas, descritas na Seção 5. Um exemplo é a arquitetura NFV-MANO, que possui a responsabilidade de gerenciar a infraestrutura que suporta as redes virtualizadas, bem como orquestrar a alocação de recursos necessários para as funções de rede, tanto virtualizadas como não virtualizadas [324].

4.2 Gerência e Orquestração de Serviços Fim-a-Fim

Sand Luz Correa
sandluz@ufg.br

5G introduz a noção de uma rede móvel flexível, programável e capaz de atender serviços com requisitos díspares e desafiadores, acelerando a transformação digital em diversos segmentos verticais. Um elemento crucial para a realização das redes 5G e de próxima geração é o conceito de *network slicing*. Proposto pela *Next Generation Mobile Networks* (NGMN) como um conceito fim-a-fim, o termo *network slicing* refere-se à criação e operação de múltiplas redes logicamente independentes sobre uma infraestrutura de telecomunicação física e compartilhada [325]. Cada rede lógica, denominada *network slice* ou simplesmente *slice*, consiste em um conjunto de VNFs, bem como recursos de computação, armazenamento e rede para

executá-las. Essas funções e recursos formam uma rede lógica instanciada para atender, de forma customizada, as características específicas (e.g., QoS, *Service Level Agreement* (SLA) e KPIs) do serviço representado pelo *slice*. Adicionalmente, os diferentes *slices* podem ser total ou parcialmente isolados uns dos outros em relação ao controle, tráfego e recursos utilizados [326].

Para atender todos esses requisitos, *network slicing* depende amplamente de tecnologias como NFV, SDN e IA [327]. NFV permite que o serviço fim-a-fim, representado pelo *slice*, seja criado como um conjunto de VNFs encadeadas. Essas VNFs são executadas em hardware de propósito geral, aliviando os custos de implantação. SDN explora a separação entre o plano de controle e plano de dados para facilitar o encaminhamento de dados entre VNFs encadeadas, facilitando a programabilidade da rede. O gerenciamento e a orquestração de *network slices*, por sua vez, abrange o posicionamento, configuração e provisionamento dinâmico de recursos para as redes virtuais (VNFs) com diferentes requisitos. A complexidade dessas operações, no entanto, exige mecanismos autônomos. IA, realizada através de técnicas de AM, é a escolha natural para apoiar a gerência e orquestração em *network slicing*.

Seguindo a nomenclatura proposta pelo 3GPP, o ciclo de vida de um *slice* é composto por quatro fases: preparação, instanciação, execução e descomissionamento [328]. Na fase de preparação, são descritos os requisitos do serviço fim-a-fim associado ao *slice*. Na fase de instanciação, o *slice* deve ser implantado sobre a infraestrutura, de forma que a implantação atenda os requisitos descritos na fase de preparação. Na fase de execução, o *slice* entra em operação e seu comportamento deve ser monitorado para garantir sua execução de forma otimizada. Finalmente, na fase de descomissionamento, o *slice* é encerrado e seus recursos liberados. Além disso, para gerenciar e orquestrar o ciclo de vida de *slices*, o 3GPP propõe três funções de rede inseridas na arquitetura de serviço 5G: *Communication Service Management Function* (CSMF), *Network Slice Management Function* (NSMF) e *Network Slice Subnet Management Function* (NSSMF). Estas funções estão ilustradas na Figura 34. A CSMF é responsável por converter os requisitos de um serviço fim-a-fim em requisitos de *slice*, gerando um descritor para o *slice* a ser criado. A NSMF recebe o descritor gerado pela CSMF e o decompõe em requisitos de sub-rede. Funções NSSMF presentes nas diferentes sub-redes (e.g., RAN e CN) usam os requisitos recebidos da NSMF e instanciam as VNFs relacionadas à sua sub-rede de acordos com os requisitos recebidos. De fato, para gerenciar e orquestrar essas VNFs, a NSSMF depende de uma arquitetura NFV-MANO [324] implantada na sua sub-rede.

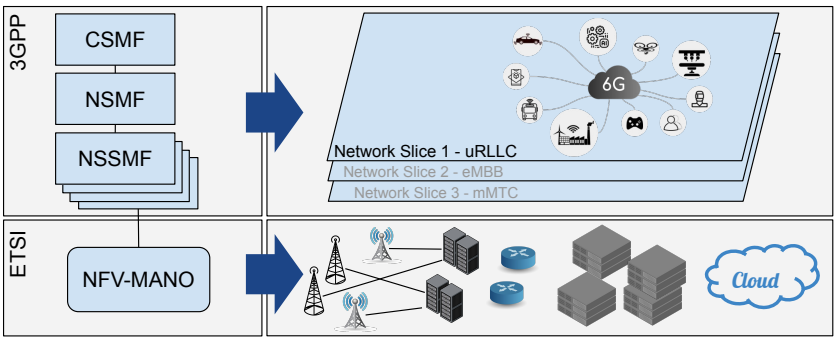


Figura 34: Arquitetura para gerenciamento e orquestração de *slices* segundo o 3GPP.

Como será apresentado na Seção 5, o 3GPP começou a delinear o caminho para integrar IA na arquitetura descrita acima. Iniciativas semelhantes tem ocorrido em outros órgão de padronização, como *European Telecommunications Standards Institute* (ETSI) e ITU. Paralelamente, diversos trabalhos na academia estão complementando os trabalhos dos órgãos de

padronização, propondo soluções algorítmicas baseadas em IA/AM para diversos problemas relacionados à gerência e orquestração de *slices*. Os autores em [329], categorizam o uso de IA no gerenciamento e orquestração de *slices* em duas dimensões: os dados usados no aprendizado e os algoritmos de aprendizado usados nas diferentes fases do ciclo de vida do *slice*. Em relação à primeira dimensão, três tipos de dados são considerados, conforme descrito a seguir.

- Demanda: os ganhos de multiplexação alcançados pela combinação eficiente de diferentes *slices* na mesma infraestrutura requerem necessariamente o aprendizado da demanda do usuário. Particularmente, a compreensão das demandas temporais e espaciais do serviço fim-a-fim representado pelo *slice* é essencial para as políticas de alocação de recursos, principalmente na fase de instanciação e execução. Os segmentos de rede geralmente considerados em *network slicing* são a RAN e CN [330]. Dados da demanda na CN podem ser coletados numa granulosidade maior, na ordem de minutos, uma vez que as tecnologias de virtualização usadas atualmente e a sobrecarga imposta por várias realocações desencorajam reconfigurações frequentes neste segmento. No entanto, a coleta de dados sobre a demanda na RAN pode ocorrer em um nível mais fino, uma vez que a antecipação mais rápida da carga de trabalho neste segmento permite melhores decisões sobre escalonamento futuro de curto prazo.

- Infraestrutura: na RAN, a infraestrutura compartilhada normalmente envolvem recursos de rede, computação e de armazenamento. Neste segmento, recursos de rede são geralmente representados por *Physical Resource Blocks* (PRBs), recursos de computação envolvem servidores usados para hospedar funções de radio e serviços *Multi-access Edge Computing* (MEC) instanciados como VNFs, enquanto recursos de armazenamento consistem em *caches* provisionadas em BSs. Na RAN, é importante compreender/antecipar a demanda dos recursos de rede, uma vez que eles são os recursos limitantes neste segmento. Por outro lado, na CN, a infraestrutura compartilhada normalmente inclui recursos de computação e armazenamento. Neste segmento, estes recursos são usados para hospedar serviços (VNFs) do núcleo da rede, os quais podem atender um ou múltiplos *slices*. Como a configuração dos serviços do núcleo é bastante flexível, é importante compreender como cada serviço se comporta em relação à carga de trabalho para determinar a quantidade de instâncias requeridas nos diferentes períodos de operação da rede.

- Requisitos: um dos desafios das redes de próxima geração é o tempo de implantação do serviço fim-a-fim. Algumas técnicas de IA podem ser aplicadas para facilitar a configuração do *slice* durante as fases de preparação e instanciação, reduzindo o tempo de implantação do serviço. Particularmente, IA pode ser empregada na função CSMF para traduzir requisitos de serviço em requisitos voltados para recursos. Essa tradução pode ocorrer inclusive no nível de detalhamento de configurações de máquinas virtuais ou contêineres.

Em relação à dimensão dos algoritmos de aprendizado, em geral, eles são aplicados para tratar três categorias principais de problemas em *network slicing* [327], como descrito a seguir.

- Posicionamento de VNFs: um *slice* envolve um conjunto de VNFs que se estende da RAN até a CN. Uma decisão a ser tomada na fase de instanciação do *slice* consiste no posicionamento dessas VNFs ao longo dos diferentes segmentos de rede. Esse problema é tratado pela NSMF e consiste em dois sub-problemas: o mapeamento de nós virtuais em

nós físicos e o mapeamento de *links* virtuais (conectando nós virtuais) em caminhos físicos [330]. Esse problema é normalmente denominado *Virtual Network Embeding* (VNE). Em VNE, dois grafos são tomados como entrada do problema: um representando a infraestrutura física e outro representando o conjunto de VNFs a ser instanciado. Nós da infraestrutura física possuem recursos (e.g., CPU e largura de banda) que serão usados para hospedar os elementos virtuais. O propósito do problema é posicionar as VNFs que compõem o *slice* em nós físicos da infraestrutura e mapear os *links* lógicos entre elas em caminhos físicos, de acordo com diferentes objetivos e respeitando os requisitos definidos para o *slice*. Grande parte dos trabalhos propostos têm como objetivo maximizar o lucro do provedor de rede. Neste caso, a receita e o custo incorridos com o mapeamento bem como a taxa de aceitação de requisições são os fatores que influenciam diretamente no lucro final do provedor. Este último é definido como a razão entre o número de requisições aceitas e o número total de requisições em um dado período de tempo. Em [331], os autores formulam o problema de VNE como um *Markov Decision Process* (MDP), onde um agente recebe as requisições para criação de *slice* e ganha uma recompensa quando a requisição é mapeada para a infraestrutura física. O objetivo do agente é maximizar as recompensas recebidas. Primeiramente os autores posicionam as VNFs na infraestrutura física. Para isso, eles utilizam o método de Monte Carlo para resolver o MDP. Em seguida, um algoritmo de caminho mais curto (*shortest path*) é usado para conetar as VNFs. Recentemente, soluções baseada em DRL [332] foram propostas para tratar o problema de VNE e melhorar o desempenho do agente.

- Controle de admissão: um mecanismo de controle de admissão de *slice* é um algoritmo que executa na NSMF e tem como objetivo aceitar ou rejeitar requisições de inquilinos para criação de *slices* na infraestrutura do provedor de rede [327]. Mecanismos dessa natureza são importantes para evitar que os recursos do provedor sejam sobrecarregados em um cenário de alta demanda por requisições de criação de *slices*. É função do controle de admissão de *slice* selecionar as requisições para criação de *slices* que serão aceitas em um determinado instante do tempo e provisionar recursos para os *slices* admitidos. Portanto, esses mecanismos devem ser executados antes da fase de instanciação dos *slices*. A estratégia empregada no mecanismo, por sua vez, terá um grande impacto no lucro do provedor de rede e na utilização geral dos recursos. Uma estratégia muito conservadora pode levar a uma baixa utilização da infraestrutura e, consequentemente, redução de receita. Por outro lado, uma estratégia excessivamente agressiva pode levar a violações de SLA num cenário de pico de demanda em múltiplos *slices* simultaneamente. De maneira geral, dois tipos de estratégias de decisão são geralmente empregadas para o problema de admissão de *slices*: decisões baseadas em políticas e decisões baseadas em leilão [333]. Nas decisões baseadas em políticas, o provedor de rede apresenta uma lista de custos para os diferentes tipos de *slices*. Os *slices* são tipificados por *templates* definidos especificamente para cada tipo de serviço (e.g., eMBB, mMTC e URLLC). Cada *template* define um conjunto de recursos (e.g. CPU e/ou largura de banda) necessários para instanciar o *slice*, o tempo de vida do *slice* e o QoS associado. Ao requisitar a criação de um *slice*, o inquilino escolhe o *template* desejado. A decisão sobre admitir ou não um *slice* é feita de acordo com a política do provedor de rede definida sobre o estado corrente do sistema. Esse é normalmente definido a partir de informações como quantidade de recursos ociosos, conjunto de *slices* ativos e a fila de requisições em espera. Decisões baseadas em políticas são tipicamente modeladas como MDP, onde uma política mapeia cada estado do sistema

em uma ação e uma recompensa correspondente. Particularmente, para o problema de admissão de *slices*, os autores em [334] mostram que a função de recompensa é geralmente não convexa sobre um grande espaço de solução. Nesse cenário, RL é uma técnica de AM conhecidamente eficiente. Portanto, alguns trabalhos propuseram o uso de RL para construir mecanismos de controle de admissão de *slices*. Em [335], os autores utilizam Q-*Learning*, enquanto os autores em [336] utilizam uma estratégia baseada em MAB. Técnicas como DRL [337] e Algoritmos Genéticos [338] também foram utilizadas para a solução desse problema. Por outro lado, nas decisões baseadas em leilão, o provedor de rede apresenta uma lista dos *slices* disponíveis e os inquilinos oferecem lances para os *slices* de interesse. Os lances são periodicamente coletados e avaliados pelo provedor de rede e os *slices* são concedidos aos vencedores. Em [339], os autores usam um algoritmo baseado em RL, distribuído nos diversos inquilinos, para implementar um mecanismo de controle de admissão de *slices* baseado em leilões. Comparado com decisões baseadas em políticas, decisões baseadas em leilões podem reduzir significativamente a complexidade computacional do problema de admissão de *slices* [333]. No entanto, a efetividade do método depende de um projeto cuidadoso do mecanismo de leilão e seu regulamento. Por fim, é importante ressaltar que o controle de admissão de *slices* não tem como foco a admissão de usuários finais. Esses, serão admitidos posteriormente por uma algoritmos tradicional de controle de admissão, executado pelo inquilino, caso o *slice* seja admitido e instanciado na infraestrutura do provedor de rede.

- Elasticidade: Diferentemente do controle de admissão, estratégias de elasticidade de *slices* focam no provimento de recursos para *slices* já admitidos e ativos na infraestrutura, sendo aplicadas na fase de execução dos *slices*. A propriedade de elasticidade é definida como a capacidade de um *slice* aumentar ou diminuir dinamicamente a quantidade de recursos alocada para um *slice*, de acordo com a carga de trabalho percebida. Tal propriedade é importante, uma vez que a carga de trabalho de um *slice* pode variar significativamente ao longo do tempo. A elasticidade de *slice* é geralmente classificada em dois tipos: vertical e horizontal. A maioria dos trabalhos definem elasticidade vertical como sendo a capacidade de aumentar ou diminuir dinamicamente os recursos atribuídos a uma máquina virtual ou contêiner, enquanto a elasticidade horizontal refere-se à habilidade de aumentar ou diminuir dinamicamente instâncias de máquinas virtuais ou contêineres. Operações de elasticidade podem ser implementadas como decisões baseadas em políticas. Nesse caso, o provedor de rede oferece um conjunto de recursos e a decisão de quais recursos alocar para cada *slice* em execução depende da quantidade de recursos disponível, da demanda instantânea de cada *slice* ativo e sua respectiva taxa de serviço. Na maioria dos trabalhos, a elasticidade é abordada com relação aos recursos de computação devido à cloudificação da CN (completamente) e da RAN (parcialmente). Além disso, a maioria dos trabalhos [340, 341] utilizam RL para resolver esse problema.

Em resumo, algoritmos de IA/AM são essenciais para prover autonomicidade às funções de gerenciamento e orquestração de *slices* e avançarmos em direção a uma rede gerenciada e operada com o mínimo de intervenção humana. No entanto, três desafios principais precisam ser superados para que esse objetivo seja alcançado em sua plenitude [342]. O primeiro desafio refere-se à falta de *datasets* de alta qualidade que possam ser utilizados no processo de aprendizado. A validação e acurácia dos modelos aprendidos dependem diretamente da disponibilidade de tais *datasets*. O segundo desafio refere-se à interpretabilidade dos modelos aprendidos. A adoção de modelos de IA para gerência e operação das redes de próxima geração depende,

em grande parte, da compreensão da relação causa-e-efeito entre decisões tomadas e dados de entrada que levaram a tais tomadas de decisões. Infelizmente, os algoritmos de IA mais promissores para diversos problemas relacionados à gerência e orquestração de *slices*, como algoritmos de DRL, geram modelos com alta complexidade e pouca interpretabilidade. Finalmente, para suportar latências muito baixas e alta confiabilidade, funções de gerenciamento e orquestração de *slices* devem tomar decisões acuradas em tempo real ou quase em tempo real. Embora técnicas emergentes de IA/ML como DRL apresentam alta acurácia, o tempo de treinamento requerido por essas técnicas ainda é alto.

De fato, embora introduzido nas redes 5G, espera-se que o gerenciamento e a orquestração inteligente de múltiplos *slices* alcance maturidade nas futuras redes 6G. Adicionalmente, características implícitas às redes 6G, como a integração de diferentes meios de comunicação (e.g., espaço, ar e terra), o suporte nativo a serviços de IA e o compromisso com a sustentabilidade, trarão novos desafios à realização do conceito de *network slicing*. No primeiro caso, o gerenciamento de *slices* sobre redes terrestres, aerotransportadas e de satélites não apenas aumentará a quantidade de recursos a serem gerenciados, como também exigirá uma coordenação criteriosa de segmentos de rede altamente heterogêneos. No segundo caso, a crescente demanda por IA em diferentes tipos de aplicações exigirá que a rede ofereça IA como um serviço. Dessa forma, serviços de IA também farão parte de um ou múltiplos *slices*, como os serviços de rede e serviços MEC. Serviços de IA, por sua vez, apresentam novos requisitos (e.g., acurácia do modelo e velocidade de aprendizado) que diferem dos requisitos tradicionais de redes. Finalmente, modelos acurados de IA exigem um grande consumo de processamento e memória e, portanto, um alto consumo de energia elétrica. Portanto além do tradicional compromisso entre multiplexação dos ganhos do provedor de rede e atendimento aos SLAs acordados, mecanismos de *network slicing* nas redes 6G deverão levar em consideração também uma dimensão relacionada à eficiência energética para prover serviços de IA de forma sustentável.

4.3 VANT como Parte da Infraestrutura de Recursos

Ciro José Almeida Macedo
ciro.macedo@ifg.edu.br

A utilização de VANT com diferentes propósitos, em diferentes aplicações, tem se tornado cada vez mais comum nos últimos tempos. Os recentes avanços tecnológicos associados a esses dispositivos, tais como, autonomia no tempo de voo, aumento na capacidade de transporte de carga útil, baixo custo de fabricação e trajetórias de voo com alta precisão, tem possibilitado a utilização de VANTs como uma ferramenta para uso em potencial em aplicações civis e militares.

No âmbito das telecomunicações, pesquisas recentes discutem diversos aspectos associados à utilização de VANTs como parte da infraestrutura de comunicação de redes móveis. Dentre os temas abordados, se destacam investigações voltadas para as temáticas de computação de borda, tecnologias de comunicação e, principalmente, à utilização de IA como elemento habilitador, dado as características dinâmicas inerentes à natureza destes dispositivos. Em [343], os autores exploram a utilização de VANTs como estações de base voadoras. O trabalho apresenta os detalhes da implementação de uma infraestrutura de comunicação LTE completamente funcional, através da utilização de dois VANTs. Se destacam aspectos associados à interoperabilidade da rede e em especial, à possibilidade de um processo de implantação ágil e flexível de novas infraestruturas de comunicação para redes móveis através da utilização de VANTs como estações de base voadoras.

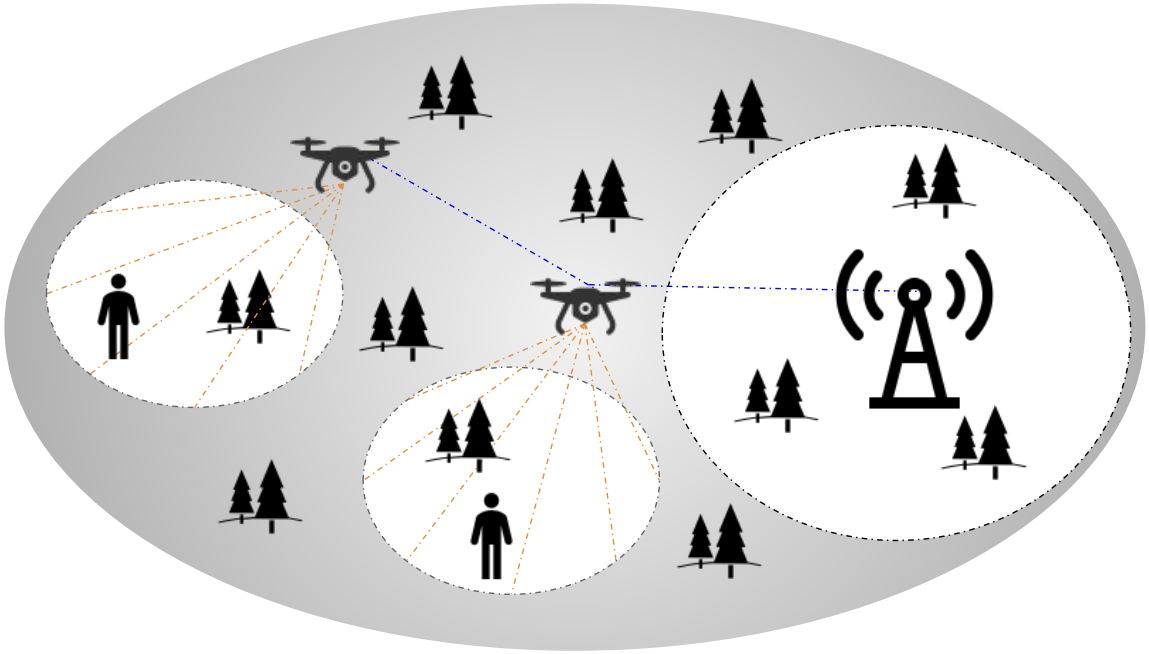


Figura 35: Operação de busca e salvamento assistida por VANTs.

Em um cenário onde um eventual desastre natural venha afetar uma determinada região, colocando a vida de pessoas em risco, o tempo de resposta associado ao inicio das operações de busca e salvamento é um fator de extrema importância. A Figura 35 ilustra uma situação, onde a infraestrutura de comunicação terrestre existente, provê conectividade apenas para a área delimitada pela circunferência tracejada que envolve a BS. Contudo o espaço de busca se estende para além da área que possui conectividade. Nestes cenários, a possibilidade da utilização de VANTs como elemento auxiliar em operações de busca e salvamento apresenta considerável relevância e vem sendo constantemente investigada.

Em [344], os autores consideram um cenário de catástrofe natural e investigam a temática do posicionamento de estações de base terrestres móveis, colaborando com estações de base posicionadas em VANTs. Os resultados apresentados indicam que a flexibilidade e agilidade dos VANTs possibilitam uma melhor distribuição das estações de base terrestres, além de permitir uma melhor utilização dos recursos de largura de banda, especialmente nas bordas da área de cobertura, onde se concentram as operações de busca e salvamento. Os autores investigam ainda, características associadas à quantidade de estações de base terrestres necessárias para cobrir uma determinada área, considerando aspectos associados à interferência. Neste sentido, os resultados confirmam que com a introdução de VANTs, menos estações de base terrestres são necessárias para prover conectividade à área desejada.

Considerando a interação de VANTs com uma infraestrutura de comunicação 5G, o trabalho [345] tem como temática central duas questões-chave. A primeira delas, os autores investigam de que maneira as redes de comunicação sem fio podem prover suporte à utilização de VANTs, considerando aplicações de uso pessoal ou profissional. No segundo caso, os autores investigam como os VANTs podem ser utilizados como elemento de suporte às redes de comunicação sem fio, provendo, por exemplo, um aumento sob demanda na capacidade da rede ou possibilitando um maior alcance de cobertura. Sob uma perspectiva de comunicação, o trabalho agrupa os VANTs em duas categorias: elementos habilitadores de mobilidade e nós aéreos de suporte a comunicação sem fio.

A interoperabilidade das VNFs que compõe o núcleo 5G, também vem sendo exploradas em pesquisas recentes envolvendo VANTs. Em [346], os autores exploram um cenário cujo o objetivo é estender o alcance de uma rede móvel através da utilização de VANTs. Para tanto, os autores movem a VNF responsável por controlar o plano de dados de usuário posicionando-a em um VANT, que atua como um elemento âncora para dispositivos móveis. Com essa abordagem, o trabalho demonstra a viabilidade de se estender as funcionalidades do núcleo da rede com a utilização de VANTs, provendo conectividade a UEs posicionados em uma determinada área.

De uma forma geral, os recentes avanços tecnológicos tem contribuído para o crescente interesse por parte da comunidade científica em fazer uso dos VANTs, como parte da infraestrutura de comunicação de redes móveis futuras. A flexibilidade destes dispositivos, possibilita explorar, por exemplo, um cenário de implantação de estações de base sob demanda, ou seja, quando e onde for necessário. Os benefícios de se prover infraestrutura de comunicação para suprir demandas momentâneas de tráfego de dados dificilmente previsíveis são consideráveis, em especial, no que diz respeito à redução de custos do tipo *Operational Expenditure* (OPEX) e *Capital Expenditure* (CAPEX). Contudo, alguns desafios necessitam ser superados. As principais preocupações que direcionam as pesquisas atuais, estão associadas à otimização da mobilidade e do posicionamento dos VANTs, à eficiência energética, à recarga destes dispositivos quando em operação e a aspectos relacionados a segurança. Considerando o aumento no número de VANTs em operação, a segurança se torna um fator crucial. Não apenas no que diz respeito a evitar quedas e possíveis ferimentos envolvendo pessoas, mas também à proteção dos dados capturados por estes dispositivos e que são posteriormente transferidos para uma rede terrestre. A utilização de hardware e software adequados pode mitigar de forma considerável os problemas associados à segurança dos dados, contudo, o custo adicional para se incorporar estas soluções nos VANTs pode se tornar um fator restritivo.

Tão importante quanto os desafios técnicos anteriormente citados, são os desafios relacionados a aspectos regulatórios dos VANTs. Temas, como privacidade, segurança publica, procedimentos administrativos e licenciamento, vêm sendo discutidos em diversos países ao redor do mundo, dando origem a um conjunto de normas regulatórias, que tem como principal objetivo controlar e, consequentemente, limitar a utilização de VANTs. Embora na maioria destes países as regulamentações propostas apresentem semelhanças pontuais, tais como, altura de voo, peso dos dispositivos e distanciamento mínimo dos seres humanos, alguns aspectos regionais inviabilizam a construção de uma visão unificada, através da qual se possa extrair um conjunto básico de regras comuns, para se fazer uso em todo o mundo.

Por fim, o atual estado da arte envolvendo a utilização de VANTs em redes móveis futuras, demonstra que a temática encontra-se em um estágio inicial de desenvolvimento, onde diversos obstáculos necessitam ser superados. O contínuo interesse e os avanços em pesquisas direcionadas à evolução das redes móveis de comunicação, aos avanços das técnicas de IA, bem como os avanços em temáticas associadas a aspectos regulatórias, pode fazer com que a utilização dos VANTs como parte de uma infraestrutura de comunicação se torne uma realidade.

5 Padronização para IA

Órgãos de padronização de telecomunicações, como ITU [347], ETSI [348], 3GPP [349], bem como alianças entre operadoras e fabricantes como O-RAN [350] *Alliance*, publicaram especificações sobre o projeto, desenvolvimento e implantação de IA/AM em sistemas 5G. Juntas, essas especificações abrangem um amplo escopo (incluindo a RAN, CN, bem como MEC) e certamente influenciarão a forma como IA será incorporada na arquitetura das redes 6G. A seguir, descrevemos as principais iniciativas que estão sendo desenvolvidas neste contexto. Essas iniciativas abrangem padrões para a coleta e disseminação de dados, geração de conhecimento e arquiteturas voltadas à integração de IA à gerência de redes.

5.1 Coleta de Dados e Disseminação de Informação

Sand Luz Correa
sandluz@ufg.br

O sistema 5G, padronizado em sua maioria pelo 3GPP, é composto pelo núcleo (CN), construído a partir do paradigma SBA, e a rede de acesso denominada NG-RAN [315]. Essa arquitetura permite o posicionamento flexível de ambientes virtuais nos segmentos de rede que compõem um sistema e define como as VNFs são criadas e implantadas. Além disso, a arquitetura usa amplamente o conceito de computação em nuvem para desenvolver, implantar e gerenciar serviços.

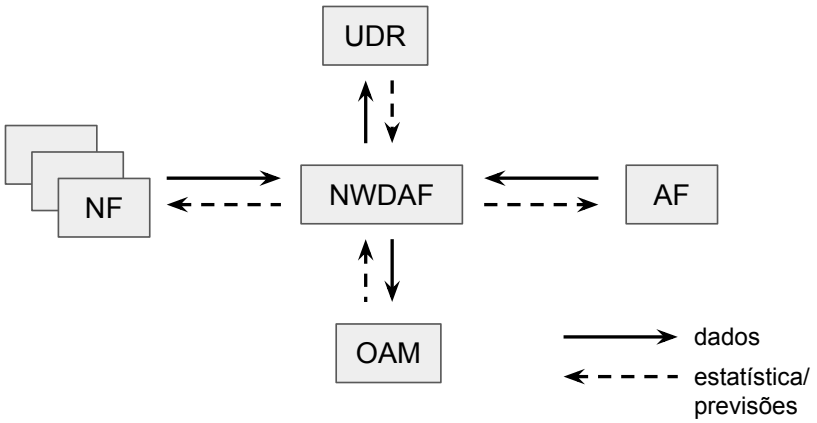


Figura 36: Arcabouço 3GPP para coleta e disseminação de dados.

Nessa arquitetura, o componente denominado *Network Data Analytic Function* (NWDAF) é a abordagem do 3GPP para atender a perspectiva de IA/AM nas redes 5G, sendo descrito nas versões 15 [315] e 16 [349] da especificação da arquitetura. Como ilustrado na Figura 36, este componente é responsável por coletar dados de vários módulos do sistema 5G, incluindo *Network Functions* (NFs), *Application Functions* (AFs), *Unified Data Repository* (UDR) e o módulo *Operations, Administration and Maintenance* (OAM). Tipicamente, dados coletados de NFs (e.g., AMF, SMF, *Policy ControlFunction* (PCF), e UDR) referem-se a equipamentos (UEs) e sessões de usuários, enquanto dados relacionados ao estado global da rede são coletados a partir do OAM. Os dados coletados são processados para fornecer diversas informações da rede e de seus usuários. De maneira geral, essas informações podem ser classificadas em dois tipos: estatísticas (e.g., sobre tráfego da rede, utilização de recursos, métricas de desempenho

da rede e mobilidade de usuários) e previsão sobre a distribuição de tráfego temporal e espacial e da localização do dispositivo em um instante futuro. A coleta e a disponibilização de dados na NWDAF pode ser realizada de forma síncrona (modelo requisição/resposta) ou assíncrona (o modelo publicação/subscrição). Adicionalmente, para ser escalável, a coleta de dados pode ocorrer em diversos níveis de granularidade.

A NWDAF é definida como uma função de plano de controle na arquitetura baseada em serviço definida pelo 3GPP. A função pode ser implantada como uma única ou múltiplas VNFs no mesmo domínio. No último caso, as múltiplas instâncias podem representar uma NF centralizada (i.e., todas as instâncias oferecem as mesmas informações analíticas) ou descentralizada (i.e., cada instância oferece uma informação analítica diferente).

5.2 Transformação de Dados e Geração de Conhecimento

Sand Luz Correa
sandluz@ufg.br

O componente NWDAF proposto pelo 3GPP especifica um padrão para os aspectos relacionados com a coleta e disseminação de dados em redes móveis de última geração, onde o núcleo da rede é composto por uma coleção de serviços que executam de forma virtualizada e seguindo preceitos do paradigma de computação em nuvem. Contudo, outro aspecto importante na incorporação de IA/AM nas redes móveis diz respeito à transformação dos dados coletados. Essa transformação gera conhecimento útil para guiar a tomada de decisão autônoma nos componentes destinados à gerência da rede. A arquitetura ENI [351], desenvolvida pelo ETSI, tem esse objetivo.

A arquitetura ENI, mostrada na Figura 37, foi projetada para atender diversos aspectos de gerenciamento das futuras redes móveis, incluindo gerenciamento de infraestrutura, operação da rede, gerenciamento e orquestração de serviços e garantias de serviço [352]. Para atender todos esses aspectos, a arquitetura ENI é formada por três módulos principais: *Processamento de Entrada*, *Análise* e *Geração de Saída*. O módulo *Processamento de Entrada* é composto pela função de *ingestão de dados* e a função de *normalização*. A primeira é responsável por coletar dados da infraestrutura e de diversos serviços e funções de rede, enquanto a segunda traduz os dados, recebidos da função de *ingestão*, para um formato ENI interno usado pelas demais funções da arquitetura.

O módulo *Análise* forma o núcleo da arquitetura ENI, sendo formado por duas categorias de funções. A primeira categoria, denominada *processamento e gestão de conhecimento*, inclui a função de *gerenciamento de conhecimento*, responsável pelo modelo de informação usado para representar os sistemas gerenciados; a função de *ciência do contexto*, que descreve o ambiente em que os sistemas gerenciados estão inseridos; e a função de *gerenciamento de cognição*, a qual é responsável pela interpretação do significado dos dados recebidos do módulo *Processamento de Entrada*, bem como da compreensão das condições do ambiente em que esses dados foram produzidos. A segunda categoria, denominada funções de *ciência e atuação*, envolve funções destinadas ao *reconhecimento da situação corrente dos sistemas gerenciados* e funções de *gerenciamento de políticas* que guiam a tomada de decisão.

Finalmente, o módulo *Geração de Saída* é responsável por transformar as políticas e recomendações, geradas pelo módulo *Análise* e representadas no formato ENI interno, em um formato que possa ser entregue aos sistemas gerenciados.

A arquitetura ENI é completada com um *broker* cuja função é abstrair as diferentes interfaces presentes nos sistemas gerenciados. Esse componente é essencial uma vez que os sistemas

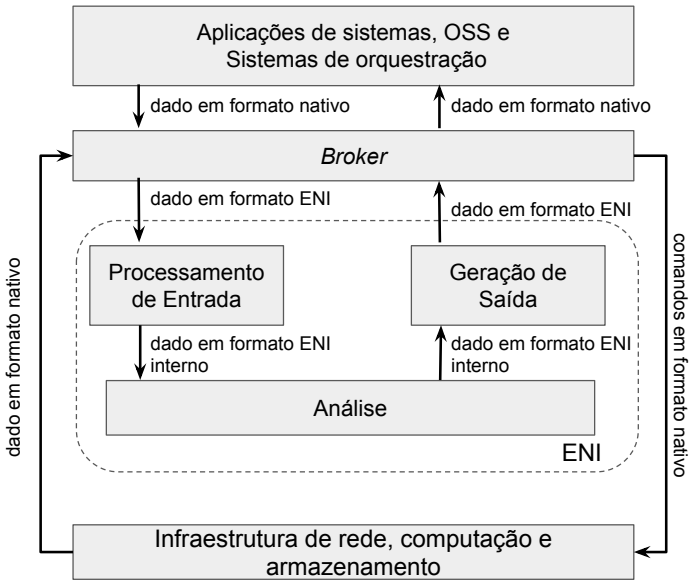


Figura 37: Arcabouço ENI para geração de conhecimento.

gerenciados representam um conjunto heterogêneo de hardware e software de domínio de rede, computação e armazenamento. É função do *broker* receber dados sobre os sistemas gerenciados em seu formato nativo e traduzir esses dados para o formato compreendido pelo módulo *Processamento de Entrada*. De maneira similar, é também função do *broker* receber dados do módulo *Geração de Saída* e convertê-los em comandos ou políticas compreendidos pelos sistemas gerenciados.

5.3 Arquiteturas para Integração de IA à Gerência de Redes

Sand Luz Correa, Cristiano Bonato Both, Ciro José Almeida Macedo
sandluz@ufg.br, cbboth@unisinos.br, ciro.macedo@ifg.edu.br

Redes móveis de próxima geração serão amplamente baseadas em paradigmas como NFV, SDN e *network slicing*, os quais permitem grande flexibilidade para provimento de serviços de rede customizados para seguimentos ou verticais específicos [5]. Esses serviços serão providos sobre uma infraestrutura de virtualização compreendendo diversos domínios tecnológicos (e.g., domínios de rede, computação e armazenamento), cada um dos quais envolvendo diversos aspectos de gerenciamento. Além de diversos domínios tecnológicos, serviços nas redes de próxima geração também poderão atravessar diferentes domínios administrativos (e.g., domínios administrados por provedores de redes móveis, provedores MEC, provedores de computação em nuvem). Portanto, o gerenciamento de serviços nas redes de próxima geração é particularmente desafiador e exige uma gerência autônoma e inteligente das redes. Para atender essa necessidade, diversos órgãos de padronização também estão especificando arquiteturas onde técnicas de IA/AM possam ser integradas, de forma holística e consistente, às funções de gerência e orquestração das redes de próxima geração [353]. A seguir, descrevemos algumas iniciativas neste sentido.

ITU desenvolveu uma arquitetura unificada para funções de AM e funções de rede. Essa arquitetura é mostrada na Figura 38 e consiste em três subsistemas principais: *pipeline de AM*, *subsistema de gerência* e *ambiente de teste* [347, 354]. O subsistema *pipeline de AM* é

um conjunto de nós lógicos, cada um com uma funcionalidade específica. Esses nós podem ser combinados para formar uma aplicação de AM. O primeiro nó do *pipeline*, denominado nó fonte (SRC), é a origem dos dados para o *pipeline*. Geralmente esse nó é representado pelo UE ou por funções de rede da arquitetura 5G (e.g., SMF, AMF). O nó coletor (C) coleta dados de um ou mais nós SRC. O nó de pre-processamento (PP) é responsável pelas funções de limpeza, agregação e preparação do dado para um formato adequado para o modelo de AM. Este último é representado pelo nó modelo (M). O nó política (P) permite a aplicação de políticas à saída gerada pelo nó modelo. O nó distribuição (D) é responsável por identificar os nós destinos e distribuir a saída gerada pelo modelo a esses nós. Finalmente, o nó destino (SINK) é o alvo da saída do modelo, ou seja, onde serão aplicadas as ações determinadas pelo modelo de AM. Portanto, o encadeamento de nós de um *pipeline* forma um fluxo de execução para aplicar uma técnica de AM sobre um destino (alvo) específico. O *subsistema de gerência* estende as funções de gerenciamento e orquestração de funções e serviços de rede para os nós do *pipeline*, unificando como serviços de rede e AM são instanciados e gerenciados na arquitetura. Para atingir esse objetivo, o *subsistema de gerência* inclui um módulo declarativo onde um *pipeline* pode ser definido, bem como um orquestrador de funções de AM (MLFO), responsável por gerenciar os *pipeline*. Finalmente, o *subsistema de ambiente de teste* é um domínio isolado para treinamento e teste de *pipeline* antes que eles sejam implantados na infraestrutura real.

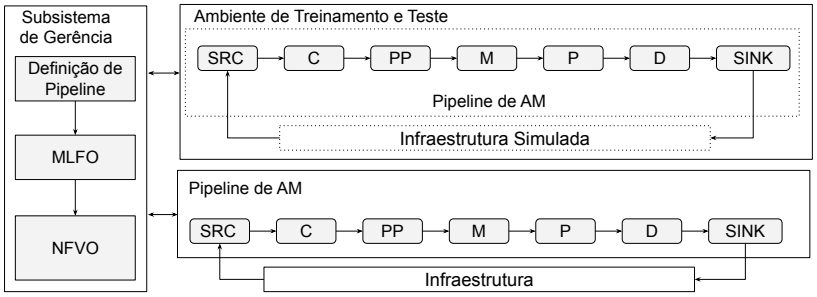


Figura 38: Arquitetura unificada para IA/AM proposta pelo ITU.

A arquitetura unificada proposta pelo ITU está alinhada com a arquitetura de serviços para a rede 5G proposta pelo 3GPP. Cada nó do *pipeline* na arquitetura unificada pode ser realizado como um serviço ou função de rede conforme definido na arquitetura 3GPP. Esse serviço, por sua vez, pode ser implantado de forma independente como uma VNF interagindo com outros nós por meio de interfaces bem definidas. Um *pipeline* pode ser realizado com base na função NWDAF. Por exemplo, um *pipeline* de AM pode ter AMF como o nó SRC; PCF como o nó SINK e os outros nós hospedados na NWDAF.

O-RAN *Alliance* introduziu um conjunto complementar de padrões NG-RAN, onde a estação base (gNB) é dividida em três partes: (i) Unidade de Rádio O-RAN (O-RU), (ii) Unidade Distribuída O-RAN (O-DU), e (iii) Unidade Central (O-CU-CP, O-CU-UP). Essas divisões para tecnologias de acesso por rádio podem ser projetadas, desenvolvidas e implantadas através de múltiplas VNFs que devem ser adequadamente posicionada e encadeada [103]. A Figura 39 ilustra a arquitetura O-RAN em alto nível. A arquitetura é composta por tês módulos principais: Arcabouço de Gerenciamento e Orquestração de Serviço (SMO), responsável pela gerência e orquestração das VNFs; Funções de Rede O-RAN, conjunto de VNFs intanciadas incluindo O-CU-CP, O-CU-UP, O-RU, O-DU; e O-Cloud, uma plataforma de computação em nuvem que provê a infraestrutura de virtualização na qual VNFs são instanciadas [355]. O SMO interage com os demais módulos da arquitetura através das interfaces O2, A1, O1 e Open Fronthaul

M-plane. A interface O1 permite a comunicação entre o SMO e uma VNF, sendo obrigatória em todas as VNFs. IA/AM é incorporada na arquitetura O-RAN através dos componentes *RAN Inteligence Controller* (RIC), habilitado para IA em tempo não real (*non-RT*) e quase em tempo real (*near-RT*) [356]. O componente *non-RT* RIC é uma função lógica dentro do SMO, sendo responsável pelo controle e otimização em tempo não real de elementos e recursos da RAN e pelo fluxo de execução de IA/AM (e.g., treinamento de modelo, inferência e atualização de modelo) na arquitetura. O componente *near-RT* é instanciado como uma VNF, sendo adequado para o gerenciamento de recursos de rádio e melhora das funções operacionais, como controle de transferência contínua, gerenciamento de QoS e gerenciamento de conectividade.

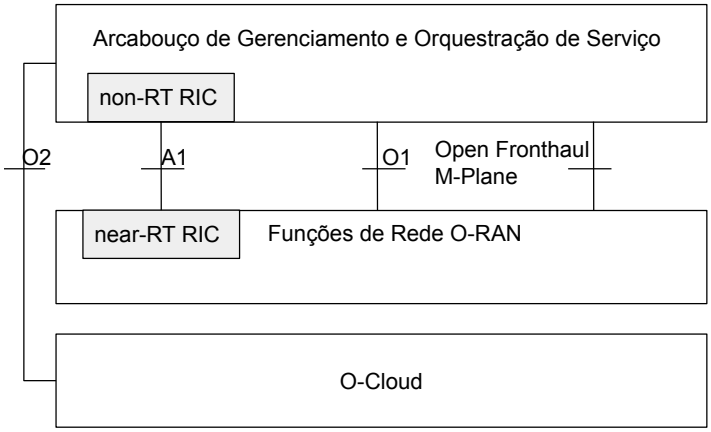


Figura 39: Arquitetura O-RAN em alto nível.

As arquiteturas para integrar IA/AM e funções de gerência de rede proposta pelo ITU e O-RAN *Alliance* são voltadas para cenários onde os serviços estão implantados em um único domínio administrativo. O gerenciamento e orquestração de serviços fim-a-fim de forma autônoma é o foco da arquitetura *Zero-touch Network and Service Management* (ZSM), desenvolvida pelo ETSI [357].

A arquitetura ZSM é projetada para dar suporte ao gerenciamento de serviços de rede em ambientes multi-domínios de forma automática e autônoma. Para atingir esse objetivo, a arquitetura usa amplamente os princípios de modularidade, extensibilidade, separação de interesse e composição de serviços. Como mostrado na Figura 40, a arquitetura ZSM é composta por múltiplos *Management Domain*s (MDs), cada um responsável pela gerência de um domínio administrativo, bem como um *End-to-End Service Management Domain* (E2EMD), responsável pelo gerenciamento fim-a-fim de serviços de clientes que atravessam múltiplos domínios.

Cada MD, incluindo o E2EMD, expõe um conjunto de serviços ou funções de gerenciamento, os quais podem ser acessados através de interfaces padronizadas. Funções de gerenciamento são agrupadas de acordo com suas funcionalidades (e.g., coleta de dados, análise de dados, controle e orquestração) e a comunicação entre elas é realizada através de entidades denominadas *estruturas de integração*. Uma *estrutura de integração* provê funções para registro, descoberta, seleção e invocação de funções de gerenciamento, oferecendo suporte para comunicação síncrona e assíncrona dentro de cada MD (*estrutura de integração intra-domínio*), como também entre domínios (*estrutura de integração inter-domínio*). Dados coletados pelas funções de gerenciamento ou gerados pelas funções de análise são mantidos por *serviços de dados*. A função principal dos *serviços de dados* é a persistência e o compartilhamento de dados dentro de cada MD. Além dos *serviços de dados intra-domínio*, existem também os *serviços de da-*

dos inter-domínios, os quais permitem o compartilhamento de dados com algumas funções de gerenciamento autorizadas.

Em resumo, na arquitetura ZSM, o E2EMD e o conjunto de MDs formam uma estrutura hierárquica de dois níveis: cada MD individual gerencia diretamente os recursos de infraestrutura dentro de um único domínio de rede, enquanto o E2EMD compõe os serviços de gerenciamento dos acpMD para oferecer suporte aos serviços de rede de foma fim-a-fim. Além disso, a arquitetura está em alinhamento com a especificação ETSI NFV e com o 3GPP, onde um MD pode ser representado por um domínio NFV MANO ou um domínio de gerenciamento 3GPP.

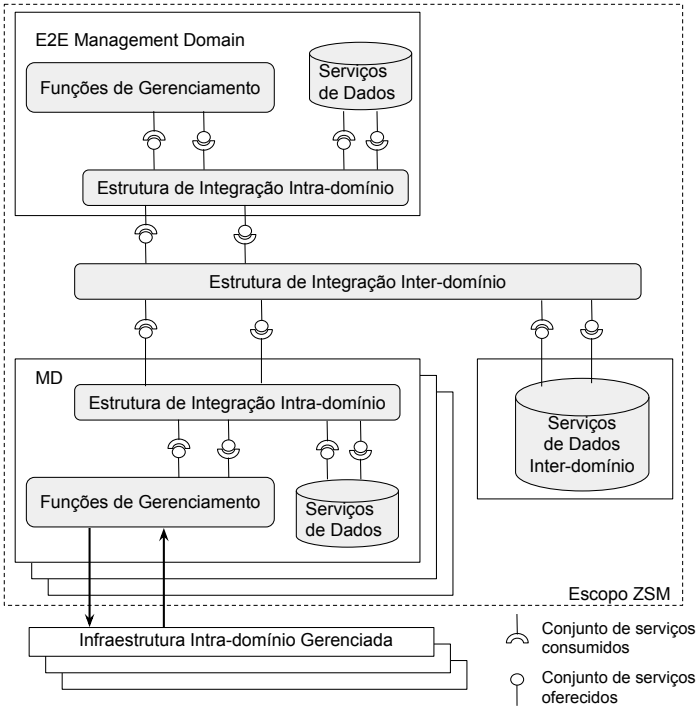


Figura 40: Arquitetura *Zero Touch Management* (ZSM).

É importante ressaltar, no entanto, que todas as iniciativas para integrar IA à gerência de redes ainda estão em um estágio inicial. Os arcabouços e arquiteturas apresentadas nesta seção são preliminares e visam, principalmente, a introdução de vários blocos de construção em um nível muito alto de abstração. Portanto, tais arquiteturas ainda estão longe de uma solução completa e pronta para implantação. Padrões mais maduros e detalhados são esperados para as redes 6G. Além de arquiteturas mais concretas, faltam também métodos padronizados para mensurar o desempenho da aplicação de técnicas de IA na gerência da rede [358]. Os métodos aplicados atualmente são centrados em métricas de comunicação, ou seja, o quanto a aplicação de uma técnica de IA melhora o desempenho da rede do ponto de vista dos recursos de comunicação. No entanto é importante mensurar também o custo computacional e a quantidade de armazenamento requeridos pela técnica. Como técnicas de IA para gerência de rede são normalmente aplicadas para diminuir a intervenção humana nestas atividades, é importante também levar essa métrica em consideração ao definir padrões para mensurar o desempenho da aplicação de técnicas de IA. Finalmente, a privacidade dos dados será uma questão muito importante para as redes 6G. Atualmente, os arcabouços e arquiteturas disponíveis não propõem blocos de construção para lidar com a privacidade dos dados.

6 Redes 6G como Suporte a Aplicações de IA

Neste capítulo serão apresentados possíveis funcionalidades das redes 6G que visam garantir o suporte ao funcionamento e aperfeiçoamento de diferentes técnicas de IA. Nos capítulos anteriores foram enfatizadas diversas técnicas de IA como base para desempenhar funções dentro das redes 6G, enquanto que ao longo deste capítulo serão enfatizados os recursos e funcionalidades das redes 6G que permitem a implementação de diferentes técnicas de IA utilizando os recursos da rede. Três funcionalidades serão apresentadas, sendo elas o particionamento de redes neurais, aprendizado federado e inteligência na borda, de forma a explorar as diferentes formas que a rede 6G pode potencializar o funcionamento das técnicas de IA.

6.1 Particionamento de redes neurais

Cleverson Nahum, Luan Gonçalves
cleversonahum@ufpa.br, luan.goncalves@itec.ufpa.br

Nos últimos anos, o particionamento de redes neurais profundas vem ganhando atenção de pesquisadores pois permite implementar tais redes com uma combinação de processamento local em dispositivos de baixo poder computacional e processamento em nuvem. Em [11] e [12], os autores analisam a viabilidade de estratégias de particionamento de redes neurais profunda entre dois dispositivos diferentes (Figura 41), levando em consideração diferentes técnicas para quantização dos valores dos sinais de ativação geradas pela camada correspondente ao ponto de particionamento da DNN. O desempenho dos métodos foram analisados em simulações que compararam a razão sinal-ruído das ativações comprimidas e a acurácia da rede para diferentes tamanhos de bloco nos métodos de compressão propostos. Os resultados obtidos sugerem que o particionamento é capaz de reduzir a taxa necessária para transmissão de informação no canal, e prover um grau de liberdade extra no posicionamento das funcionalidades de IA no sistema.

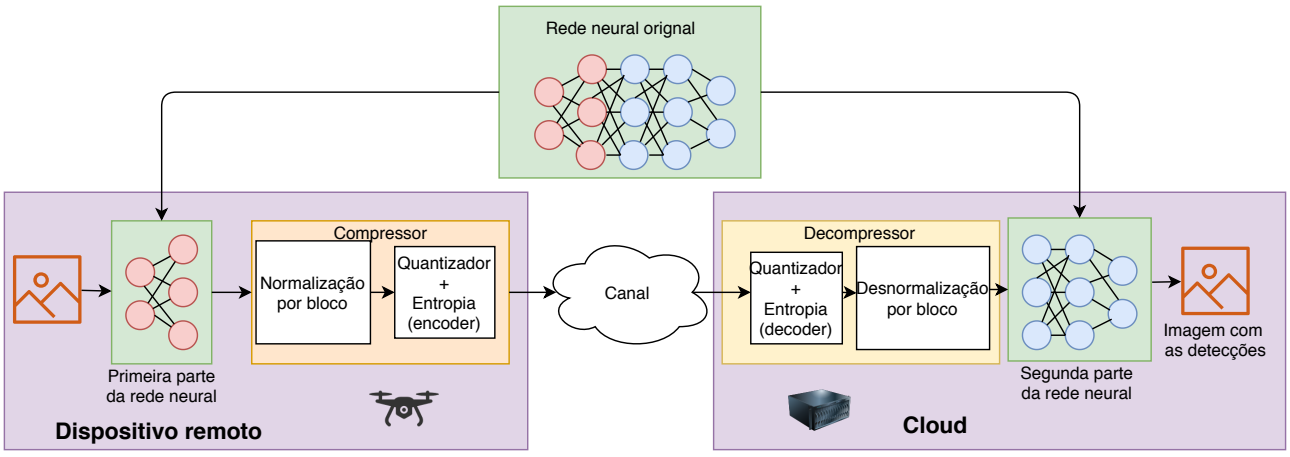


Figura 41: Representação geral do sistema analisado em [11] e [12]. O dispositivo remoto implementa as primeiras camadas da rede neural e um sistema computacional em nuvem implementa as últimas camadas da rede neural (adaptado de [12]).

A escolha do ponto de particionamento da rede neural define as características das informações (por exemplo, a dimensão e distribuição de probabilidade dos sinais de ativação) a serem transmitidas. Além disso, este ponto dita o custo computacional que estará sendo demandado de cada elemento responsável pelo processamento das partes da rede. Por exemplo, caso o

dispositivo remoto possua baixa capacidade de processamento, o particionamento da rede pode ser feito de forma a alocar poucas camadas da rede neural no dispositivo remoto e a maioria delas na nuvem, já que essa última possui um maior poder computacional. Contudo, esse cenário tende a exigir maior taxa de bits do canal de comunicação. Em cenário alternativo, o dispositivo remoto possui maior capacidade computacional. Assim, mais camadas poderiam ser associadas ao dispositivo remoto e a taxa de bits ser diminuída. O artigo [11] avalia os ganhos e perdas de cada variação no particionamento tanto ao nível de recursos computacionais, quanto dos recursos da rede de comunicações.

A Fig. 42 apresenta os resultados obtidos em [12] ao utilizar a normalização de blocos, quantização escalar e codificação por entropia das ativações da rede neural. É possível perceber que tanto a variação de SNR e número de bits usados para o tamanho do bloco a ser normalizado afetam a acurácia média do sistema, apesar de obter uma acurácia ligeiramente menor quando apenas 2 bits são utilizados no compressor, o ganho na taxa de transmissão da rede é mais significativo. De forma que dependendo da situação da rede e do nível de acurácia necessários na aplicação, as ativações podem sofrer maior ou menor compressão.

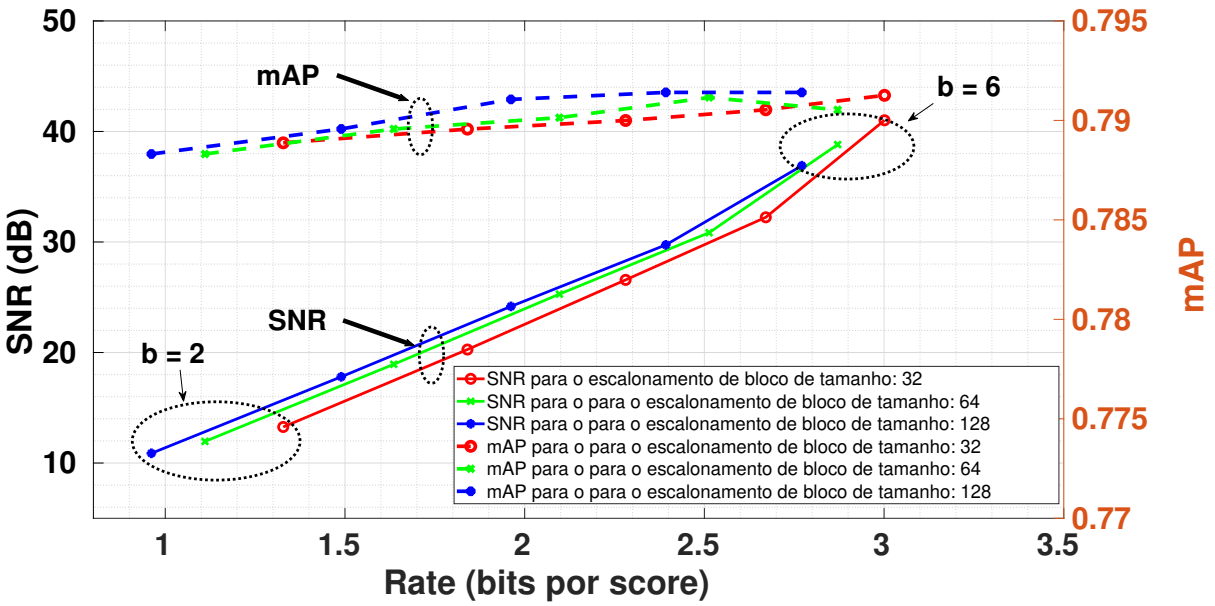


Figura 42: Resultado obtido com a utilização do sistema de compressão dos sinais de ativação da rede neural em [12].

6.2 Aprendizado Federado

Victor Hugo L. Lopes
victor.lopes@ifg.edu.br

Embora o emprego de IA em redes 5G e B5G/6G, com os mais diversos propósitos e em diferentes camadas da pilha de protocolos, se projete como uma realidade, as soluções baseadas no processamento dos dados necessários aos processos de treinamentos de modelos em entidades centralizadas nem sempre é viável em todas as aplicações nestes sistemas de comunicação sem fio, principalmente devido à questões de privacidade no acesso de certos dados e aos custos da transmissão de dados brutos [359]. Neste sentido, surge um apelo para arquiteturas que

permitam a otimização dos sistemas e serviços nestas novas redes nas quais o treinamento dos modelos de IA ocorram de forma descentralizada, em que o Aprendizado Federado (FL, do inglês *Federated Learning*) se mostra viável.

Inicialmente proposto como um método de otimização descentralizado para aprendizado de máquina, levando em consideração conceitos de otimização federada (*Federated Optimization* [360]), FL [361, 362] surge com a proposta de permitir o treinamento de modelos centralizados de alta qualidade, mas mantendo os dados de treinamento nos próprios dispositivos em que eles são gerados, e que estão dispersos na rede, com foco em uma comunicação eficiente, minimizando o tráfego necessário para o envio destes dados de treinamento para as unidades centralizadas, ao passo que resolve problemas de privacidade, segurança e propriedade destes dados [363]. Trata-se, portanto, de uma solução de aprendizado de máquina que fornece meios para lidar com dados de treinamento (*datasets*) distribuídos em dezenas de dispositivos na rede (ou bilhões [359]), onde apenas modelos localmente treinados são transferidos às unidades centralizadoras, cumprindo as principais características:

- Dispositivos com capacidades heterogêneas e naturalmente dispersos treinam modelos locais, utilizando seus próprios dados.
- Modelos globais são criados, treinados e/ou agregados [13] baseados nos modelos locais coletados, e não precisam considerar um *dataset* centralizado.
- Os dispositivos dispersos são alimentados com os modelos globais para uso local.
- Há diversas restrições de segurança e privacidade dos dados.
- Há limitações de recursos de comunicação e de tempo para a coleta e disseminação de dados brutos.

O funcionamento básico do FL pode ser observado na Figura 43, na qual os números ilustram a sequência de procedimentos requeridos. De maneira geral, uma arquitetura baseada em FL gera maior capacidade para o sistema de comunicação sem fio no tratamento de questões relacionadas aos aspectos de IA em termos de eficiência e efetividade, bem como sobre a segurança e privacidade [364]. Levando-se em consideração as classes de casos de uso e os requisitos previstos para as redes 6G (Fig. 16), em que se sobressai a conectividade massiva de dispositivos executando diversos tipos destes serviços, e a importante redução da latência para as aplicações de tempo real e sensíveis ao atraso, o emprego de FL como suporte na obtenção da eficiência e efetividade se torna evidente. A necessidade do encaminhamento de dados para o processamento e treinamento centralizado tende a incluir maior latência no sistema, também gerando impactos na eficiência energética e espectral e na própria operação destes novos serviços, principalmente quando se vislumbra os ambientes extremamente densificados. Tal impacto também se apresenta quando se considera nós centrais que precisam receber tais dados, em que gargalos de transmissão podem ser gerados. Adicionalmente, diversos tipos de aplicações também podem obter ganhos pelo uso local de modelos treinados/agregados de alta qualidade, que podem carregar características aprendidas em dados que a aplicação não teria acesso, ou que o dispositivo não teria capacidade computacional para o treinamento dos modelos.

Já quando se considera os aspectos de segurança e privacidade, diversas aplicações 6G (ULBC e uMBB, por exemplo) baseadas em dados sensíveis, principalmente aqueles que integram os serviço centrados no usuário, trazem diversas implicações, inclusive legais, como requerido pela Lei Geral de Proteção de Dados (LGPD), por exemplo. De forma semelhante,

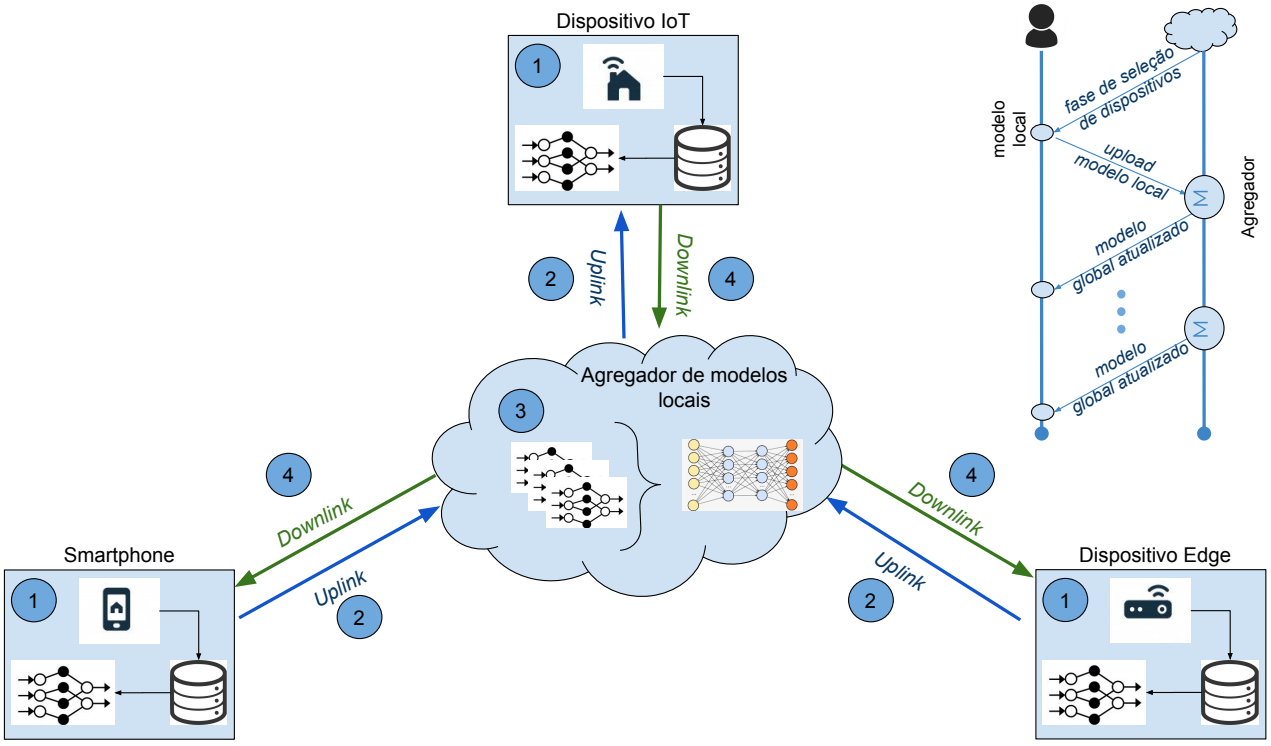


Figura 43: Procedimentos de uma arquitetura baseada em Aprendizado Federado (Adaptado de [13]).

as aplicações mULC e uMBB que integram todo o escopo de aplicações de indústria 4.0 também carregam informações sensíveis, com sérios impactos em questões como segredo industrial, por exemplo. Desta forma, em aplicações em que um grande volume de dados sensíveis precisa ser utilizado para prover modelos de IA robustos e úteis, o emprego de FL se mostra essencial, de forma que a proteção dos dados é fortalecida pela capacidade de compartilhamento das atualizações dos modelos, como parâmetros e gradientes de redes neurais, por exemplo, ao invés dos dados completos [365].

Apesar dos recentes avanços observados nesta temática, estudos apontam alguns grandes desafios a serem enfrentados de forma a permitir o pleno emprego de FL como parte integrante das arquiteturas de sistemas 6G com forte emprego de IA, tais como [13, 14, 359, 363–365]:

- Recursos de comunicação caros: embora o emprego de bandas do espectro de radiofrequências atualmente menos povoadas e com largura de banda maiores, como no emprego de mmWave e *Terahertz* (THz), por exemplo, tais recursos tendem a se tornar ainda mais caros, em termos de eficiência espectral e energética, quando a operação dos sistemas 6G elevar consideravelmente a densidade de ocupação de dispositivos conectados, de forma que se torna essencial o desenvolvimento de métodos com comunicação eficientes, em que ocorram trocas de informações entre os dispositivos conectados e as entidades da rede na menor quantidade possível nas etapas de treinamento e propagação dos modelos, reduzindo o número total de iterações e o tamanho das mensagens trocadas em cada iteração [365].

- Heterogeneidade de sistemas: a conectividade ubíqua e global aguardada para os sistemas 6G deve contribuir significativamente para a diversificação dos dispositivos, princi-

palmente em termos de suas capacidades de processamento, armazenamento, memória, conectividade e energéticas. Tal heterogeneidade gera impactos na operação de sistemas baseados em FL, tais como na tolerância a falhas, e na variação contínua da quantidade de usuários conectados, e na própria qualidade dos modelos globais, dado que dependem da qualidade dos modelos locais [13]. Desta forma, estes sistemas devem se antecipar às situações de baixo número de participantes, ampliar a tolerância a *hardwares* heterogêneos, serem robustos o suficiente para descartar certos dispositivos conectados. A heterogeneidade também pode ocorrer com relação aos dados obtidos por dispositivos não identicamente distribuídos, de forma que o número de pontos de coletas de dados pode variar drasticamente, requerendo recursos estatísticos suficientemente robustos na captura das relações através destes dados e as suas representações no mundo real, que não gerem maiores complexidades nas realizações de otimizações distribuídas, modelagem, análises teóricas e avaliações de soluções.

- Preocupações com a privacidade: Embora os métodos recentes tenham como objetivo aumentar a privacidade em FL usando ferramentas como a Computação Multipartidária Segura (SMC) ou privacidade diferencial, essas abordagens geralmente fornecem privacidade ao custo da redução do desempenho do modelo ou da eficiência do sistema. Compreender e equilibrar essas compensações, tanto teórica quanto empiricamente, é um desafio considerável na realização de sistemas privados de FL. Embora o fato de não trafegar as informações sensíveis de forma bruta resolva parte dos problemas de segurança e privacidade, o compartilhamento dos modelos treinados também pode representar ameaças, como na obtenção maliciosa de padrões em textos e informações financeiras, por exemplo [365].

A adoção de arquiteturas de sistemas de comunicação sem fio de próxima geração que favoreçam as aplicações baseadas em IA se torna essencial. Neste sentido, a arquitetura definida pelo 3GPP para o núcleo das redes 5G já abre diversas possibilidades de que futuras arquiteturas facilitadoras favoreçam a implementação e operação de métodos de IA, como se torna esperado para o 6G. Na arquitetura de núcleo 5G NR, a função de rede NWDAF (Fig. 36) deve fornecer mais capacidade de exposição de dados para funcionalidades baseadas em IA, o que favorece o uso de métodos inteligentes de gerenciamento do sistema, permitindo que os operadores automatizem diversas funções gerenciais e tarefas de configuração, por exemplo, reduzindo a necessidade de interação humana [356, 359]. Sendo o NWDAF capaz de acessar qualquer NF, o que deve permitir o seu futuro uso como facilitador em diferentes tarefas envolvendo FL.

Neste sentido, a arquitetura proposta em [14] (demonstrada na Figura 44) permite vislumbrar as redes 6G como parte fundamental no suporte às aplicações de IA nestes ambientes de conectividade ubíqua, no que os autores consideram ser a evolução da arquitetura SBA do 5G NR para uma arquitetura hiperflexível baseada em IA ubíqua. Nesta arquitetura, subconjuntos de UEs são selecionados e configurados para o treinamento de modelos segundo critérios e objetivos estabelecidos pela rede, como para o atendimento de demandas específicas de determinadas aplicações, por exemplo. Os modelos agregados podem ser propagados para os servidores de nuvem, ou entre os servidores de borda, como forma de transferência de conhecimento, estando disponíveis para outras aplicações e/ou funções de rede. Desta forma, as soluções de IA nestas novas redes tendem a ter um significativo incremento na eficiência das comunicações envolvidas nos processos de transmissão de dados, treinamento e propagação de modelos, além da ampliação do suporte à dados heterogêneos, resolvendo parte dos problemas da preservação da privacidade e segurança dos dados sensíveis. Além disso, uma arquitetura

6G com soluções de FL proporcionará o suporte a implementações de redes com escala massiva, também contribuindo para as arquiteturas flexíveis necessárias ao atendimento dos requisitos projetados.

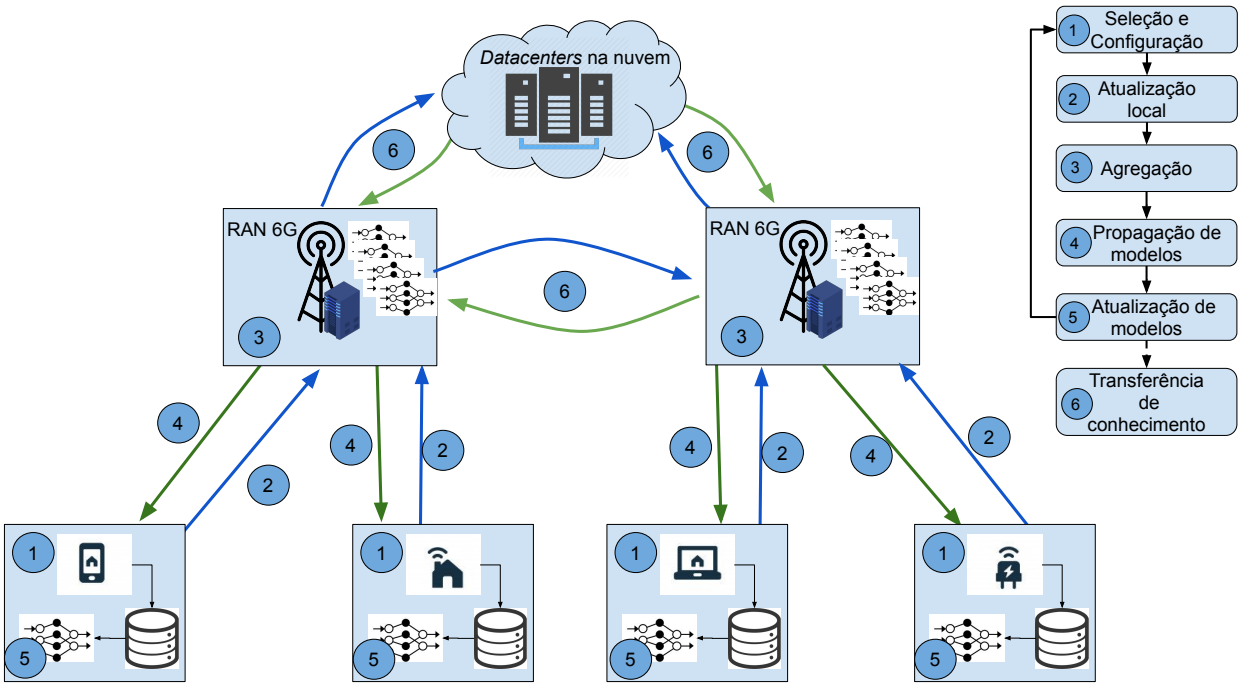


Figura 44: Arquitetura 6G baseada em FL (Adaptado de [14]).

6.3 Inteligência de Borda

Kleber Vieira Cardoso
kleber@ufg.br

Inteligência de Borda ou *Edge Intelligence* (EI) vem sendo tratada como um dos principais componentes a ser incluído em Redes 6G [366–368], consolidando a integração entre comunicação e computação iniciada nas Redes 5G e introduzindo o suporte nativo de IA/AM à infraestrutura. De fato, EI representa dois aspectos de IA/AM: 1) o suporte da infraestrutura às aplicações de IA/AM e 2) o uso de IA/AM na infraestrutura. O primeiro aspecto é o que tem sido mais amplamente discutido e diz respeito a uma mudança significativa na filosofia convencional de separar computação e comunicação em redes sem fio tradicionais [369]. Para Redes 6G, EI propõe abordar de maneira conjunta comunicação e sistemas IA/AM de borda. Conforme será discutido posteriormente, essa abordagem implica em alterações importantes no projeto da rede. O segundo aspecto envolve ampliar a adoção de soluções baseadas em IA/AM para abordar questões relacionadas à comunicação de dados, em especial as relacionadas a alocação de recursos. Diferente de sistemas de nuvem tradicionais, os sistemas de computação de borda possuem recursos limitados e comumente envolvem *hardware* heterogêneo e muitas vezes podem interagir diretamente com a comunicação sem fio. Esse cenário apresenta dificuldades e oportunidades adequadas para soluções baseadas em IA/AM, conforme será detalhado posteriormente.

Inicialmente, EI se propõe a resolver algumas questões relacionadas ao uso de IA/AM que estejam em sistemas de nuvem. Em geral, os dados usados por IA/AM em sistemas de nuvem são obtidos a partir de equipamentos que se conectam à infraestrutura de rede sem fio, tais como dispositivos de usuários (e.g., celulares, *tablets* e equipamentos vestíveis) e dispositivos IoT. Por um lado, IA/AM em sistemas de nuvem contam com diversas vantagens desse tipo de sistema, tais como alta escalabilidade, desempenho e robustez. Por outro lado, o envio de dados para a nuvem introduz o atraso necessário para atravessar os componentes de rede que separam os dispositivos que se conectam à borda e os sistemas de nuvem. Além disso, o envio regular de alto volume de dados para a IA/AM em sistemas de nuvem tende a criar gargalos ao longo do caminho, aumentando o atraso fim-a-fim e, eventualmente, provocando perdas devido ao congestionamento da rede. Trazer IA/AM para a borda resolve os dois problemas, permitindo que os dados sejam utilizados próximos de onde são produzidos e provendo o retorno necessário com menor atraso. De fato, para algumas aplicações que dependem de atrasos muito baixos, tais como carros autônomos e realidade aumentada/virtual/misturada/estendida, o benefício de adotar EI é notável. Várias aplicações utilizam o reconhecimento de pessoas ou objetos em escalas de tempo mais altas e poderiam, portanto, utilizar IA/AM em sistemas de nuvem. No entanto, esse tipo de aplicação também pode se beneficiar de IA/AM na borda por evitar o envio de um grande volume de informações para a nuvem, sobretudo à medida em que aumenta significativamente a quantidade de dados enviados por dispositivo e o número de dispositivos (e.g., câmeras e sistemas de varredura a laser) que precisam utilizar IA/AM.

Outra questão relevante em que EI pode oferecer contribuições diz respeito à privacidade dos dados. Cenários em que o sistema de borda pertence ao usuário ou é considerado confiável, há menos risco e, portanto, menos restrições ao envio dos dados que são fundamentais para as aplicações e serviços baseados em IA/AM. Em cenários em que o sistema de borda pertence a terceiros considerados não confiáveis, é necessário a adoção de abordagens de IA/AM distribuídas, nas quais o equipamento do usuário realiza parte das tarefas de IA/AM, evitando que dados privados sejam enviados. Aprendizado federado é um exemplo de IA/AM distribuída que vem sendo amplamente utilizada [359]. Embora esse tipo de abordagem possa ser usada também para IA/AM em sistemas de nuvem, há vantagens no uso de EI. Por exemplo, EI permite que as decisões relacionadas à comunicação levem em conta aspectos específicos da abordagem IA/AM utilizada. Para ilustrar, é possível escolher *AirComp* [370], ao invés da OFDMA, para prover acesso a usuários que estejam utilizando Aprendizado Federado. Conforme descrito em [369], usando Aprendizado Federado, o sistema de borda requer a média de atualizações do modelo, em vez de seus valores individuais e, portanto, *AirComp* pode reduzir drasticamente a latência de acesso múltiplo por um fator igual ao número de usuários.

Redes 5G é um passo importante na adoção de soluções baseadas em IA/AM para abordar questões relacionadas à comunicação de dados. As especificações mais recentes apresentadas por entidades como 3GPP, ITU, ETSI e O-RAN *Alliance* permitem que a grande quantidade de soluções baseadas em IA/AM, gerada pela comunidade científica, possa ser efetivamente usada em sistemas de comunicação móvel. Conforme já apresentado em seções anteriores, essas soluções baseadas em IA/AM abrangem todas as camadas (i.e., física [74], enlace [3], rede [371] e transporte [372]) e todas as partes de um sistema de comunicação móvel celular (i.e., núcleo [373], rede de transporte [374] e rede de acesso [375]). No entanto, a maior parte dessas soluções visam melhorar o desempenho da comunicação via IA/AM centralizado, desconsiderando a latência adicional induzida pela inferência do aprendizado [366]. Além disso, esses trabalhos geralmente assumem que os modelos de IA/AM possuem um grande número de amostras de dados para serem treinados e não levam em conta a latência de treinamento desses modelos.

Por outro lado, para utilizar EI, a latência e a confiabilidade de IA/AM devem ser avaliadas em relação à comunicação e aos processos de treinamento e inferência [368], os quais podem ocorrer de maneira descentralizada.

Conforme descrito, EI é um assunto amplo e complexo, o qual abrange muitos conceitos e tecnologias interligados. Atualmente, não existe uma definição formal que seja consenso sobre EI, mas alguns trabalhos tentam estruturar a literatura sobre esse assunto. Por exemplo, os autores de [376] acreditam que o escopo de EI não deve se restringir à execução de modelos de IA/AM apenas em servidores ou dispositivos de borda, mas em colaboração de borda e nuvem. Eles definem seis níveis de inteligência de borda, de co-inferência na borda-nuvem (nível 1) a tudo no dispositivo (nível 6). Em [377], os autores propõem distinguir EI em IA para borda e IA na borda. IA para borda é identificada como uma direção de pesquisa com foco em fornecer uma solução melhor para problemas de otimização restrita em computação de borda com a ajuda de tecnologias eficazes de IA/AM. IA na borda investiga como executar modelos de IA/AM na borda, formando um arcabouço para execução de treinamento e inferência de modelos de IA/AM com sinergia dispositivo-borda-nuvem. Esse arcabouço visa atender os requisitos de desempenho dos algoritmos de IA/AM, dentre outros aspectos (e.g., custo, privacidade, confiabilidade, eficiência, etc.), usando o grande volume de dados disponível na borda e de maneira distribuída.

Os autores de [378] identificam as principais questões que afetam o aprendizado distribuído sobre redes sem fio que é uma das características importantes de EI e, portanto, amplamente abordadas na literatura recente sobre o assunto. A seguir, essas questões são brevemente descritas:

- Escassez de dados: Ao migrar da nuvem para a borda, ocorre uma redução significativa na quantidade de dados usados em IA/AM por múltiplas razões. Naturalmente, a capacidade de armazenar e processar os dados é sensivelmente menor nos sistema de borda, em comparação com a nuvem. A quantidade de fontes de dados que alimentam cada servidor de borda também tende a ser restrita. Para superar a escassez de dados, abordagens IA/AM robustas e colaborativas são importantes.

- Dados não-independentes e/ou não-identicamente distribuídos: É comum que EI obtenha dados de fontes variadas, incluindo os equipamentos dos usuários finais e, portanto, que a propriedade de ser independentes e identicamente distribuídos não seja garantida em muitos casos. Quando os dados obtidos são não-independentes e/ou não-identicamente distribuídos, é comum que a precisão e a velocidade de convergência do processo de aprendizado de máquina sejam significativamente afetadas.

- Privacidade dos dados: Os dados coletados podem conter informações sensíveis aos usuários e, portanto, é comum adotar abordagens (como Aprendizado Federado) que utilizam algum tipo de representação dos dados, e.g., parâmetros de um modelo de AM, em vez de dados propriamente ditos na aprendizagem distribuída. Ainda assim, há preocupação em que pelo menos parte dos dados possam ser reversamente obtidos a partir das informações enviadas, afetando portanto a privacidade dos usuários. Para minimizar esse risco, é possível adotar medidas como codificação extra, introdução de ruído em parâmetros compartilhados e troca de informações redundantes. No entanto, essas soluções trazem problemas adicionais, por exemplo, aumento no atraso de processamento com codificação extra, perda de precisão de inferência devido ao excesso de ruído e atrasos de comunicação extras com informações redundantes.

- Limitação de recursos de computação: Em geral, o treinamento e o uso de modelos de IA/AM requerem grande capacidade de processamento e memória, o que não está normalmente disponível nos equipamentos da borda. Embora os servidores de borda possam aliviar o trabalho dos dispositivos móveis, eles formam uma capacidade computacional notadamente inferior à encontrada na nuvem. Portanto, é importante que considerar abordagens de IA/AM que sejam mais eficientes em termos computacionais e também energéticos, por conta dos dispositivos móveis, ainda que a acurácia seja eventualmente afetada.

- Limitação de recursos de comunicação: Como EI conta com a participação dos dispositivos sem fio, inicialmente para coleta de dados, mas eventualmente também para realizar parte do processamento relacionado a IA/AM, o enlace sem fio se torna um componente crítico. O problema é que o enlace sem fio serve para os dispositivos realizarem todas as suas comunicações, ou seja, para atender todos os serviços. Apesar de redes 5G (e posteriores) adicionarem bastante capacidade ao enlace sem fio, em especial com mais espectro, por exemplo, ondas milimétricas em 5G e THz em 6G, há também uma expectativa de uso intenso da rede com as novas aplicações. Além disso, essas tecnologias tendem a aumentar sensivelmente o consumo de energia dos dispositivos móveis, os quais dependem de baterias. Portanto, o gerenciamento de recursos de comunicação é um aspecto fundamental para a realização do aprendizado distribuído.

- Condições ruins de canal (sem fio): Uma vez que o enlace sem fio é um componente crítico para EI, vale ressaltar que esse tipo de canal de comunicação eventualmente sofre condições ruins devido à exposição do sinal eletromagnético, especialmente devido à mobilidade dos dispositivos. Naturalmente, quando um canal sem fio passa por períodos de condições ruins, o aprendizado distribuído é afetado, por exemplo, aumentando a latência no treinamento e reduzindo da inferência. É esperado que técnicas tradicionais sejam usadas para lidar com o problema, tais como escalonamento, codificação, quantização, retransmissão, gerenciamento de interferência, etc., mas também é possível adotar novas abordagens que explorem os aspectos do aprendizagem distribuído, por exemplo, garantias de latência de treinamento, precisão, confiabilidade e robustez.

- Topologia de rede variável com o tempo: A partição dos dispositivos móveis na EI também introduz dinamicidade na topologia da rede utilizada no aprendizado distribuído. Com redes que variam no tempo, o aprendizado distribuído é afetado por vários fatores como perda de conectividade, colaboração inconsistente e assíncrona, incompatibilidades frequentes de modelos e a tendência de ter dados e modelos desatualizados. Soluções passam por técnicas preditivas/proativas e pela reformulação das interações entre os agentes envolvidos no aprendizado distribuído para que esses utilizem modelos estatísticos simplificados.

A seguir, é apresentada uma coletânea de desafios e oportunidades de pesquisa identificados na literatura recente sobre EI:

- Soluções de rede cientes de computação [376]: Para EI, as aplicações baseadas em IA/AM computacionalmente intensivas são normalmente executadas em um ambiente de computação de borda distribuído. Como resultado, soluções de rede avançadas com ciência dos recursos de computação são altamente desejáveis, de modo que os resultados de computação e os dados possam ser compartilhados de forma eficiente entre diferentes nós de

borda. Nesse contexto, é promissor integrar URLLC com computação de borda para fornecer serviços de EI de baixa latência e alta confiabilidade. Além disso, rede definida por *software* e virtualização de função de rede oferecem o controle flexível sobre os recursos de rede para suportar interconexões sob demanda em diferentes nós de borda para aplicações de IA/AM que demandam computação intensiva. Por outro lado, o projeto de mecanismos de rede autônomos é importante para fornecer o provisionamento de serviço de EI e de maneira eficiente sob coexistência de rede heterogênea dinâmica (por exemplo, 5G/*WiFi*/*LoRa*). Assim, é possível que nós de borda e dispositivos móveis recém-adicionados se autoconfigurem de maneira autônoma.

- Projeto de aplicações baseadas em DNN com múltiplas métricas de desempenho [376]: Para uma aplicação baseada em DNN com uma missão específica, geralmente há uma série de candidatos a modelo que são capazes de concluir a tarefa. No entanto, no contexto de EI, é difícil para os desenvolvedores de software escolherem um modelo DNN apropriado porque os indicadores de desempenho padrão, e.g., k mais acurados ou precisão média, falham em refletir o desempenho do tempo de execução da inferência do modelo DNN em dispositivos de borda. Fatores adicionais, tais como a velocidade de inferência e o uso de recursos também se tornam métricas importantes no contexto de EI. Portanto, é necessário explorar os compromissos entre as diferentes métricas, buscando melhorar a eficiência de aplicações que dependam de EI.

- Serviço inteligente e gerenciamento de recursos [376]: Dada a natureza distribuída da computação de borda, implica que a funcionalidade de EI também está naturalmente dispersa em diversas localizações. Ou seja, diferentes dispositivos de borda podem executar diferentes modelos de IA e oferecer diferentes serviços relacionados. Portanto, é importante projetar soluções eficientes para descoberta de serviços, de modo que os usuários possam identificar e localizar os provedores de serviços de EI relevantes para atender às suas necessidades em tempo hábil. Além disso, os serviços de EI precisam lidar com ambientes altamente dinâmicos, tanto em termos de recursos quanto de demanda. Isso exige que a orquestração e o provisionamento de recursos de borda ocorram de maneira *online* para acomodar de forma eficiente o uso intensivo de EI. Ou seja, é necessário realizar otimização conjunta e em tempo real de recursos de computação heterogênea, comunicação e parâmetros de sistema (por exemplo, escolher o treinamento de modelo adequado e técnicas de inferência) adaptados para as diversas demandas.

- Equilíbrio entre otimalidade e eficiência [377]: Potencialmente, as tecnologias de IA conseguem fornecer soluções ótimas, porém isso pode ter impacto na eficiência no uso de recursos limitados, i.e., os recursos de borda. Nesse contexto, há o desafio de melhorar a usabilidade e a eficiência dos sistemas de computação de borda para diferentes cenários de aplicação com tecnologias de IA incorporadas. É importante ressalta que o compromisso entre a otimalidade e a eficiência deve ser realizada com base nas características dos requisitos que mudam dinamicamente na QoE e na estrutura de recursos da rede. Portanto, há um conflito entre a busca pela solução que atende de maneira ótima um serviço e a quantidade de recursos exigidos para alcançar esse resultado, havendo outros serviços e seus cálculos de soluções competindo pelos recursos.

- Disponibilidade dos dados [377]: A disponibilidade e a usabilidade dos dados brutos para treinamento são questões tradicionais em qualquer aplicação de IA. No contexto de EI, o

fato dos dados pertencerem predominantemente a terceiros (i.e., usuários móveis) e algumas características (e.g., viés) trazem algumas especificidades ao problema. Inicialmente, é necessário ter algum tipo de estratégia (e.g., um mecanismo de incentivo) para obter os dados de usuários móveis. Caso contrário, pode haver uma escassez significativa de dados para treinamento e inferência do modelo. Além disso, é importante assumir que os dados de vários dispositivos finais possuem problemas com viés, o que pode afetar muito o desempenho do aprendizado e que, portanto, precisam ser tratados. Embora o Aprendizado Federado possa superar o problema causado por dados não-independentes e/ou não-identicamente distribuídos em certa medida, o procedimento de treinamento ainda enfrenta dificuldades no projeto de protocolo de comunicação robusto. Em vários casos, o Aprendizado Federado não é suficiente ou não pode ser utilizado, exigindo que outras estratégias precisem ser desenvolvidas.

- Mecanismo de coordenação [377]: Os métodos propostos na adaptação de modelos de IA podem não ser amplamente úteis no contexto de EI porque pode haver uma grande diferença no poder de computação e recursos de comunicação entre os dispositivos de borda. Isso pode fazer com que o mesmo método alcance resultados de aprendizagem diferentes para diferentes grupos de dispositivos móveis. Portanto, há a necessidade de compatibilidade e coordenação entre os dispositivos de borda heterogêneos para mitigar o problema. Isso pode ser alcançado através de um mecanismo de coordenação flexível entre nuvem, borda e dispositivos móveis, levando em conta tanto hardware quanto software. Esse tipo de mecanismo precisa de uma *Application Programming Interface* (API) uniforme para aprendizado no contexto de EI que seja suportada de maneira ampla pelos dispositivos de borda. É importante que tanto a API quanto o projeto do mecanismo de coordenação sejam abertos para motivar a adoção e promover a compatibilidade entre diferentes implementações.

- Generalização para outras arquiteturas de aprendizado na borda [369]: Além do Aprendizado Federado, também é interessante generalizar o acesso múltiplo orientado ao aprendizado para outras arquiteturas, onde o servidor de borda precisa realizar cálculos mais sofisticados sobre os dados recebidos, i.e., não apenas uma média simples como no Aprendizado Federado. O principal desafio na generalização é a maneira de explorar a propriedade de superposição de um canal multiacesso para computar funções mais complexas.

- Extração de características ciente do canal (sem fio) [369]: A abordagem tradicional para processamento de sinais ciente de canal também pode ser projetada em conjunto com a extração de características em sistemas de aprendizado de borda. Essa pode ser considerada uma nova área de pesquisa que tem como objetivo explorar propriedades físicas do canal para realizar extração eficiente de características. O potencial desse tipo de abordagem já foi ilustrado em alguns trabalhos, por exemplo, em [379], os autores comparam o desempenho de aprendizagem de um sistema de aprendizado de borda centralizado usando a técnica de codificação analógica de Grassmann para transmissão analógica rápida de dados com um sistema baseado em dois esquemas coerentes de alta taxa: transmissão MIMO digital e analógica. Um ganho significativo de desempenho foi observado em cenários de alta mobilidade com a técnica de codificação analógica de Grassmann.

7 Conclusão

Este relatório apresentou uma ampla revisão do estado-da-arte de técnicas de IA para 6G.

Referências

[1] I. Qualcomm Technologies, "NR designing a unified, more capable 5G air interface," 2018. [Online]. Available: https://www.qualcomm.com/media/documents/files/the-3gpp-release-15-5g-nr-design.pdf

[2] H. Yin, X. Guo, P. Liu, X. Hei, and Y. Gao, "Predicting Channel Quality Indicators for 5G Downlink Scheduling in a Deep Learning Approach," 2020.

[3] J. Zhang, X. Xu, K. Zhang, B. Zhang, X. Tao, and P. Zhang, "Machine Learning Based Flexible Transmission Time Interval Scheduling for eMBB and uRLLC Coexistence Scenario," *IEEE Access*, vol. 7, pp. 65 811–65 820, 2019.

[4] K. V. Cardoso, C. B. Both, L. R. Prade, C. J. Macedo, and V. H. L. Lopes, "A softwarized perspective of the 5G networks," *arXiv preprint arXiv:2006.10409*, 2020.

[5] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The Road Towards 6G: A Comprehensive Survey," *IEEE Open Journal of the Communications Society*, 2021, publisher: IEEE.

[6] F. A. Aoudia and J. Hoydis, "End-to-end learning of communications systems without a channel model," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2018, pp. 298–303.

[7] T. Jian, B. C. Rendon, E. Ojuba, N. Soltani, Z. Wang, K. Sankhe, A. Gritsenko, J. Dy, K. Chowdhury, and S. Ioannidis, "Deep Learning for RF Fingerprinting: A Massive Experimental Study," *IEEE Internet of Things Magazine*, vol. 3, no. 1, pp. 50–57, 2020.

[8] X. N. Fernando, *Radio over fiber for wireless communications: From fundamentals to advanced topics*. John Wiley & Sons, 2014.

[9] V. Ziegler, H. Viswanathan, H. Flinck, M. Hoffmann, V. Räisänen, and K. Hätönen, "6G architecture to connect the worlds," *IEEE Access*, vol. 8, pp. 173 508–173 520, 2020.

[10] X. Huang, J. A. Zhang, R. P. Liu, Y. J. Guo, and L. Hanzo, "Airplane-Aided Integrated Networking for 6G Wireless: Will It Work?" *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 84–91, 2019.

[11] F. de Brito, I. Nascimento, L. Gonçalves, S. Lins, N. Linder, and A. Klautau, "Signal compression for efficient partitioning of deep neural networks," in *XXXVIII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT 2020) (SBrT 2020)*, Florianópolis, Brazil, Nov. 2020.

[12] L. Silva, F. Brito, L. Ramalho, and A. Klautau, "Compressão dos Sinais de Ativação de Redes Neurais Profundas Particionadas," in *X Conferência Nacional em Comunicações, Redes e Segurança da Informação*, 2020.

[13] L. U. Khan, S. R. Pandey, N. H. Tran, W. Saad, Z. Han, M. N. Nguyen, and C. S. Hong, "Federated learning for edge networks: Resource optimization and incentive mechanism," *IEEE Communications Magazine*, vol. 58, no. 10, pp. 88–93, 2020.

[14] Y. Xiao, G. Shi, and M. Krunz, "Towards ubiquitous AI in 6G with federated learning," *arXiv preprint arXiv:2004.13563*, 2020.

[15] G. Evans, J. Miller, M. I. Pena, A. MacAllister, and E. Winer, "Evaluating the microsoft hololens through an augmented reality assembly application," in *Degraded environments: sensing, processing, and display 2017*, vol. 10197. International Society for Optics and Photonics, 2017, p. 101970V.

[16] Huawei, "Cloud VR network solution white paper. Shenzhen, China. Whiter Paper," 2018.

[17] M. Maier, M. Chowdhury, B. P. Rimal, and D. P. Van, "The tactile internet: vision, recent progress, and open challenges," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 138–145, 2016.

[18] I. Network, "2030-a blueprint of technology, applications and market drivers towards the year 2030 and beyond."

[19] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6g: Ai empowered wireless networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, 2019.

[20] Y. Hu and V. O. Li, "Satellite-based internet: a tutorial," *IEEE Communications Magazine*, vol. 39, no. 3, pp. 154–162, 2001.

[21] Cisco VNI, "Cisco visual networking index: global mobile data traffic forecast update, 2017–2022," Cisco Visual Networking Index, White Paper, February 2019. [Online]. Available: https://s3.amazonaws.com/media.mediapost.com/uploads/CiscoForecast.pdf

[22] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher, "Multiple-antenna techniques for wireless communications-a comprehensive literature survey," *IEEE communications surveys & tutorials*, vol. 11, no. 2, pp. 87–105, 2009.

[23] H. Yang, M. H. Herben, I. J. Akkermans, and P. F. Smulders, "Impact analysis of directional antennas and multiantenna beamformers on radio transmission," *IEEE transactions on vehicular technology*, vol. 57, no. 3, pp. 1695–1707, 2008.

[24] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Communications magazine*, vol. 55, no. 9, pp. 134–141, 2017.

[25] J. Mo, B. L. Ng, S. Chang, P. Huang, M. N. Kulkarni, A. AlAmmouri, J. C. Zhang, J. Lee, and W.-J. Choi, "Beam codebook design for 5G mmWave terminals," *IEEE Access*, vol. 7, pp. 98 387–98 404, 2019.

[26] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 949–973, 2015.

[27] J. Kim and A. F. Molisch, "Fast millimeter-wave beam training with receive beamforming," *Journal of Communications and Networks*, vol. 16, no. 5, pp. 512–522, 2014.

[28] B. Zong, C. Fan, X. Wang, X. Duan, B. Wang, and J. Wang, "6g technologies: Key drivers, core requirements, system architectures, and enabling technologies," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 18–27, 2019.

[29] H. Elayan, O. Amin, R. M. Shubair, and M.-S. Alouini, "Terahertz communication: The opportunities of wireless technology beyond 5G," in *2018 International Conference on Advanced Communication Technologies and Networking (CommNet)*. IEEE, 2018, pp. 1–5.

[30] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE communications magazine*, vol. 52, no. 2, pp. 106–113, 2014.

[31] H. Zhao, R. Mayzus, S. Sun, M. Samimi, J. K. Schulz, Y. Azar, K. Wang, G. N. Wong, F. Gutierrez, and T. S. Rappaport, "28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city," in *2013 IEEE international conference on communications (ICC)*. IEEE, 2013, pp. 5163–5167.

[32] G. R. MacCartney, S. Deng, S. Sun, and T. S. Rappaport, "Millimeter-wave human blockage at 73 GHz with a simple double knife-edge diffraction model and extension for directional antennas," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*. IEEE, 2016, pp. 1–6.

[33] M. Abouelseoud and G. Charlton, "The effect of human blockage on the performance of millimeter-wave access link for outdoor coverage," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*. IEEE, 2013, pp. 1–5.

[34] A. Rosebrock, *Deep Learning for Computer Vision with Python*. pyimagesearch, 2018. [Online]. Available: https://books.google.com.br/books?id=60IvygEACAAJ

[35] C. Jiang, H. Zhang, Y. Ren, Z. Han, K. C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 98–105, 2017.

[36] F. Tang, Y. Kawamoto, N. Kato, and J. Liu, "Future Intelligent and Secure Vehicular Network Toward 6G: Machine-Learning Approaches," *Proceedings of the IEEE*, vol. 108, no. 2, pp. 292–307, 2020.

[37] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, "Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions," *IEEE Access*, vol. 7, pp. 137 184–137 206, 2019.

[38] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 173–196, 2018.

[39] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on selected areas in communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[40] Y. Wang, A. Klautau, M. Ribero, A. C. Soong, and R. W. Heath, "MmWave vehicular beam selection with situational awareness using machine learning," *IEEE Access*, vol. 7, pp. 87 479–87 493, 2019.

[41] S. Rezaie, C. N. Manchón, and E. de Carvalho, "Location- and Orientation-Aided Millimeter Wave Beam Selection Using Deep Learning," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[42] C. Antón-Haro and X. Mestre, "Learning and data-driven beam selection for mmWave communications: An angle of arrival-based approach," *IEEE Access*, vol. 7, pp. 20 404–20 415, 2019.

[43] D. Li, S. Wang, H. Zhao, and X. Wang, "Context-and-Social-Aware Online Beam Selection for mmWave Vehicular Communications," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8603–8615, 2021.

[44] Y. Yang, Z. Gao, Y. Ma, B. Cao, and D. He, "Machine learning enabling analog beam selection for concurrent transmissions in millimeter-wave v2v communications," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 9185–9189, 2020.

[45] W. Xu, F. Gao, S. Jin, and A. Alkhateeb, "3D Scene-Based Beam Selection for mmWave Communications," *IEEE Wireless Communications Letters*, vol. 9, no. 11, pp. 1850–1854, 2020.

[46] C.-H. Lin, W.-C. Kao, S.-Q. Zhan, and T.-S. Lee, "BsNet: A Deep Learning-Based Beam Selection Method for mmWave Communications," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, 2019, pp. 1–6.

[47] H. Echigo, Y. Cao, M. Bouazizi, and T. Ohtsuki, "A Deep Learning-Based Low Overhead Beam Selection in mmWave Communications," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 1, pp. 682–691, 2021.

[48] D. Jagyasi and M. Coupechoux, "DNN Based Beam Selection in mmW Heterogeneous Networks," 2021.

[49] M. Alrabeiah and A. Alkhateeb, "Deep learning for mmWave beam and blockage prediction using Sub-6 GHz channels," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5504–5518, 2020.

[50] M. Alrabeiah, A. Hredzak, and A. Alkhateeb, "Millimeter wave base stations with cameras: Vision-aided beam and blockage prediction," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*. IEEE, 2020, pp. 1–5.

[51] G. H. Sim, S. Klos, A. Asadi, A. Klein, and M. Hollick, "An online context-aware machine learning algorithm for 5G mmWave vehicular communications," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2487–2500, 2018.

[52] W. Shafik, M. Ghasemzadeh, and S. M. Matinkhah, "A Fast Machine Learning for 5G Beam Selection for Unmanned Aerial Vehicle Applications," *Journal of Information Systems and Telecommunication (JIST)*, vol. 4, no. 28, p. 262, 2020.

[53] V. Va, T. Shimizu, G. Bansal, and R. W. Heath, "Online Learning for Position-Aided Millimeter Wave Beam Training," *IEEE Access*, vol. 7, pp. 30 507–30 526, 2019.

[54] JuSang-Lim, KimNam-il, and KimKyung-Seok, "Machine-Learning-Based User Group and Beam Selection for Coordinated Millimeter-wave Systems," *International journal of advanced smart convergence*, vol. 9, no. 4, pp. 156–166, 12 2020.

[55] Yang, Yang and He, Yu and He, Dazhong and Gao, Zhen and Luo, Yihao, "Machine Learning based Analog Beam Selection for 5G mmWave Small Cell Networks," in *2019 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2019, pp. 1–5.

[56] M. Dias, A. Klautau, N. González-Prelcic, and R. W. Heath, "Position and LIDAR-aided mmWave beam selection using deep learning," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2019, pp. 1–5.

[57] C. Jia, H. Gao, N. Chen, and Y. He, "Machine learning empowered beam management for intelligent reflecting surface assisted MmWave networks," *China Communications*, vol. 17, no. 10, pp. 100–114, 2020.

[58] A. Klautau, P. Batista, N. González-Prelcic, Y. Wang, and R. W. Heath, "5G MIMO data for machine learning: Application to beam-selection using deep learning," in *2018 Information Theory and Applications Workshop (ITA)*. IEEE, 2018, pp. 1–9.

[59] M. Alrabeiah, Y. Zhang, and A. Alkhateeb, "Neural Networks Based Beam Codebooks: Learning mmWave Massive MIMO Beams that Adapt to Deployment and Hardware," *arXiv preprint arXiv:2006.14501*, 2020.

[60] K. Bhogi, C. Saha, and H. S. Dhillon, "Learning on a Grassmann Manifold: CSI Quantization for Massive MIMO Systems," *arXiv preprint arXiv:2005.08413*, 2020.

[61] J. Jiang, X. Wang, W.-J. Wang, L. Zhen, and J. Wang, "Deep Clustering-Based Codebook Design for Massive MIMO Systems," *IEEE Access*, vol. 7, pp. 172 654–172 664, 2019.

[62] Y. Zhang, M. Alrabeiah, and A. Alkhateeb, "Learning beam codebooks with neural networks: Towards environment-aware mmWave MIMO," in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2020, pp. 1–5.

[63] J.-C. Chen, "Efficient codebook-based beamforming algorithm for millimeter-wave massive MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 7809–7817, 2017.

[64] Y. Zhang, M. Alrabeiah, and A. Alkhateeb, "Reinforcement Learning of Beam Codebooks in Millimeter Wave and Terahertz MIMO Systems," *arXiv preprint arXiv:2102.11392*, 2021.

[65] J. Jiang, X. Wang, G. A. S. Sidhu, L. Zhen, and R. Gao, "Clustering-based codebook design for MIMO communication system," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE, 2019, pp. 1–6.

[66] H. Lee, M. Girnyk, and J. Jeong, "Deep reinforcement learning approach to MIMO precoding problem: Optimality and Robustness," *arXiv preprint arXiv:2006.16646*, 2020.

[67] A. Balatsoukas-Stimming, O. Castañeda, S. Jacobsson, G. Durisi, and C. Studer, "Neural-network optimized 1-bit precoding for massive MU-MIMO," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2019, pp. 1–5.

[68] S. Takabe, M. Imanishi, T. Wadayama, R. Hayakawa, and K. Hayashi, "Trainable projected gradient detector for massive overloaded MIMO channels: Data-driven tuning approach," *IEEE Access*, vol. 7, pp. 93 326–93 338, 2019.

[69] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "A model-driven deep learning network for MIMO detection," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2018, pp. 584–588.

[70] G. Liu, H. Deng, K. Yang, Z. Zhu, J. Liu, and H. Dong, "A New Design of Codebook for Hybrid Precoding in Millimeter-Wave Massive MIMO Systems," *Symmetry*, vol. 13, no. 5, p. 743, 2021.

[71] W. Ma, C. Qi, Z. Zhang, and J. Cheng, "Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO," *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 2838–2849, 2020.

[72] A. M. Elbir and A. K. Papazafeiropoulos, "Hybrid precoding for multiuser millimeter wave massive MIMO systems: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 552–563, 2019.

[73] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 3027–3032, 2019.

[74] F. Sohrabi, K. M. Attiah, and W. Yu, "Deep learning for distributed channel feedback and multiuser precoding in fdd massive mimo," *IEEE Transactions on Wireless Communications*, 2021.

[75] M. S. Aljumaily and H. Li, "Machine Learning Aided Hybrid Beamforming in Massive-MIMO Millimeter Wave Systems," in *2019 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*. IEEE, 2019, pp. 1–6.

[76] Y. Sun, Z. Gao, H. Wang, and D. Wu, "Machine learning based hybrid precoding for mmWave MIMO-OFDM with dynamic subarray," *arXiv preprint arXiv:1809.03378*, 2018.

[77] J. Kang, J. H. Lee, and W. Choi, "Machine Learning-Based Dimension Optimization for Two-Stage Precoder in Massive MIMO Systems with Limited Feedback," *Applied Sciences*, vol. 9, no. 14, p. 2894, 2019.

[78] K. M. Attiah, F. Sohrabi, and W. Yu, "Deep learning approach to channel sensing and hybrid precoding for tdd massive mimo systems," *arXiv preprint arXiv:2011.10709*, 2020.

[79] X. Li, Y. Huang, W. Heng, and J. Wu, "Machine Learning-Inspired Hybrid Precoding for mmWave MU-MIMO Systems with Domestic Switch Network," *Sensors*, vol. 21, no. 9, p. 3019, 2021.

[80] H. Hojatian, J. Nadal, J.-F. Frigon, and F. Leduc-Primeau, "Unsupervised Deep Learning for Massive MIMO Hybrid Beamforming," *arXiv preprint arXiv:2007.00038*, 2020.

[81] T. Jiang, H. V. Cheng, and W. Yu, "Learning to beamform for intelligent reflecting surface with implicit channel estimate," *arXiv preprint arXiv:2009.14404*, 2020.

[82] S. Nayak and R. Patgiri, "6G Communication: Envisioning the Key Issues and Challenges," *EAI Endorsed Transactions on Internet of Things*, vol. 6, no. 24, p. 166959, Feb 2021. [Online]. Available: http://dx.doi.org/10.4108/eai.11-11-2020.166959

[83] 3GPP TS 38.214 version 15.3.0 Release 15, *Physical layer procedures for data*, Oct 2018. [Online]. Available: http://www.etsi.org

[84] 3GPP TS 38.212 version 15.2.0 Release 15, *Multiplexing and channel coding*, Jul 2019. [Online]. Available: http://www.etsi.org

[85] H. Haggui, S. Affes, and F. Bellili, "FPGA-SDR Integration and Experimental Validation of a Joint DA ML SNR and Doppler Spread Estimator for 5G Cognitive Transceivers," *IEEE Access*, vol. 7, pp. 69 464–69 480, 2019.

[86] T. Ngo, B. Kelley, and P. Rad, "Deep Learning Based Prediction of Signal-to-Noise Ratio (SNR) for LTE and 5G Systems," in *2020 8th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 2020, pp. 1–6.

[87] J. Liu, K. Mei, D. Ma, and J. Wei, "Deep neural network aided scenario identification in wireless multi-path fading channels," 2018.

[88] Y. Ding and H. Kwon, "Doppler Spread Estimation for 5G NR with Supervised Learning," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–7.

[89] Y. Sun, C. Wang, H. Cai, C. Zhao, Y. Wu, and Y. Chen, "Deep Learning Based Equalizer for MIMO-OFDM Systems with Insufficient Cyclic Prefix," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, 2020, pp. 1–5.

[90] Y. Yang, Y. Li, W. Zhang, F. Qin, P. Zhu, and C.-X. Wang, "Generative-adversarial-network-based wireless channel modeling: Challenges and opportunities," *IEEE Communications Magazine*, vol. 57, no. 3, pp. 22–27, 2019.

[91] W. Jiang and H. D. Schotten, "Deep learning for fading channel prediction," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 320–332, 2020.

[92] M. Soltani, V. Pourahmadi, and H. Sheikhzadeh, "Pilot Pattern Design for Deep Learning-Based Channel Estimation in OFDM Systems," *IEEE Wireless Communications Letters*, vol. 9, no. 12, pp. 2173–2176, 2020.

[93] E. Ghadimi, F. Davide Calabrese, G. Peters, and P. Soldati, "A reinforcement learning approach to power control and rate adaptation in cellular networks," in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–7.

[94] H. Park and Y. Lim, "Adaptive power control using reinforcement learning in 5g mobile networks," in *2020 International Conference on Information Networking (ICOIN)*, 2020, pp. 409–414.

[95] Y. V. L. de Melo, R. L. Batista, T. F. Maciel, C. F. M. e. Silva, J. M. B. da Silva, and F. R. P. Cavalcanti, "Power control with variable target sinr for d2d communications underlying cellular networks," in *European Wireless 2014; 20th European Wireless Conference*, 2014, pp. 1–6.

[96] S. Nie, Z. Fan, M. Zhao, X. Gu, and L. Zhang, "Q-learning based power control algorithm for d2d communication," in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2016, pp. 1–6.

[97] C. Yang, J. Han, and X. Xu, "How d2d communication influences energy efficiency of small cell network with sleep scheme," in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, 2014, pp. 1690–1695.

[98] H. Yazar, Arslan, "Reliability enhancement in multi-numerology-based 5G new radio using INI-aware scheduling," 2019.

[99] M. Eisen, C. Zhang, L. F. Chamon, D. D. Lee, and A. Ribeiro, "Learning optimal resource allocations in wireless systems," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2775–2790, 2019.

[100] S. Ali, W. Saad, N. Rajatheva, K. Chang, D. Steinbach, B. Sliwa, C. Wietfeld, K. Mei, H. Shiri, H.-J. Zepernick, T. M. C. Chu, I. Ahmad, J. Huusko, J. Suutala, S. Bhadauria, V. Bhatia, R. Mitra, S. Amuru, R. Abbas, B. Shao, M. Capobianco, G. Yu, M. Claes, T. Karvonen, M. Chen, M. Girnyk, and H. Malik, "6G White Paper on Machine Learning in Wireless Communication Networks," *arXiv:2004.13875 [cs, eess, math]*, Apr. 2020.

[101] E. Björnson and E. Jorswieck, *Optimal resource allocation in coordinated multi-cell systems*. Now Publishers Inc, 2013.

[102] A. Zaidi, F. Athley, J. Medbo, U. Gustavsson, G. Durisi, and X. Chen, *5G Physical Layer: Principles, Models and Technology Components*. Elsevier Science, 2018. [Online]. Available: https://books.google.com.br/books?id=mtJKDwAAQBAJ

[103] A. Ghosh *et al.*, "5G Evolution: A View on 5G Cellular Technology Beyond 3GPP Release 15," *IEEE Access*, vol. 7, pp. 127 639–127 651, 2019.

[104] Y. Kim, Y. Kim, J. Oh, H. Ji, J. Yeo, S. Choi, H. Ryu, H. Noh, T. Kim, F. Sun, Y. Wang, Y. Qi, and J. Lee, "New Radio (NR) and its Evolution toward 5G-Advanced," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 2–7, 2019.

[105] R. Liu, G. Yu, J. Yuan, and G. Y. Li, "Resource Management for Millimeter-Wave Ultra-Reliable and Low-Latency Communications," *IEEE Transactions on Communications*, 2020.

[106] M. Elsayed and M. Erol-Kantarci, "Radio resource and beam management in 5G mmWave using clustering and deep reinforcement learning," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

[107] A. Avranas, M. Kountouris, and P. Ciblat, "Deep reinforcement learning for wireless scheduling with multiclass services," *arXiv preprint arXiv:2011.13634*, 2020.

[108] B. Khodapanah, A. Awada, I. Viering, A. noll Barreto, M. Simsek, and G. Fettweis, "Framework for Slice-Aware Radio Resource Management Utilizing Artificial Neural Networks," *IEEE Access*, vol. 8, pp. 174 972–174 987, 2020, publisher: IEEE.

[109] Y. Liu, X. Wang, J. Mei, G. Boudreau, H. Abou-Zeid, and A. B. Sediq, "Situation-Aware Resource Allocation for Multi-Dimensional Intelligent Multiple Access: A Proactive Deep Learning Framework," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 116–130, Jan. 2021.

[110] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 957–975, 2020.

[111] X. You, C.-X. Wang, J. Huang, X. Gao, Z. Zhang, M. Wang, Y. Huang, C. Zhang, Y. Jiang, J. Wang *et al.*, "Towards 6G wireless communication networks: Vision, enabling technologies, and new paradigm shifts," *Science China Information Sciences*, vol. 64, no. 1, pp. 1–74, 2021.

[112] H. Mossalam, Y. M. Assael, D. M. Roijers, and S. Whiteson, "Multi-objective deep reinforcement learning," *arXiv preprint arXiv:1610.02707*, 2016.

[113] T. Zhou, K. Xu, X. Xia, W. Xie, and J. Xu, "Achievable Rate Optimization for Aerial Intelligent Reflecting Surface-aided Cell-Free Massive MIMO System," *IEEE Access*, 2020.

[114] T. X. Vu, S. Chatzinotas, S. ShahbazPanahi, and B. Ottersten, "Joint power allocation and access point selection for cell-free massive MIMO," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.

[115] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint power allocation and load balancing optimization for energy-efficient cell-free massive MIMO networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6798–6812, 2020.

[116] T. Wang, S. Wang, and Z.-H. Zhou, "Machine learning for 5G and beyond: From model-based to data-driven mobile wireless networks," *China Communications*, vol. 16, no. 1, pp. 165–175, 2019, publisher: IEEE.

[117] R. Dong, C. She, W. Hardjawana, Y. Li, and B. Vucetic, "Deep learning for radio resource allocation with diverse quality-of-service requirements in 5g," *IEEE Transactions on Wireless Communications*, 2020.

[118] C. She, R. Dong, Z. Gu, Z. Hou, Y. Li, W. Hardjawana, C. Yang, L. Song, and B. Vucetic, "Deep learning for ultra-reliable and low-latency communications in 6g networks," *IEEE Network*, vol. 34, no. 5, pp. 219–225, 2020.

[119] R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Estrada-Solano, and O. M. Caicedo, "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," *Journal of Internet Services and Applications*, vol. 9, no. 1, pp. 1–99, 2018.

[120] F. D. Calabrese, L. Wang, E. Ghadimi, G. Peters, L. Hanzo, and P. Soldati, "Learning radio resource management in rans: Framework, opportunities, and challenges," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 138–145, 2018.

[121] F. Al-Tam, N. Correia, and J. Rodriguez, "Learn to schedule (leasch): A deep reinforcement learning approach for radio resource scheduling in the 5g mac layer," *IEEE Access*, vol. 8, pp. 108 088–108 101, 2020.

[122] R. Barazideh, O. Semiari, S. Niknam, and B. Natarajan, "Reinforcement learning for mitigating intermittent interference in terahertz communication networks," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2020, pp. 1–6.

[123] M. Yuan, Q. Cao, M.-o. Pun, and Y. Chen, "Towards user scheduling for 6g: A fairness-oriented scheduler using multi-agent reinforcement learning," *arXiv preprint arXiv:2012.15081*, 2020.

[124] D. Zhao, H. Qin, B. Song, Y. Zhang, X. Du, and M. Guizani, "A reinforcement learning method for joint mode selection and power adaptation in the v2v communication network in 5g," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 2, pp. 452–463, 2020.

[125] N. Kato, B. Mao, F. Tang, Y. Kawamoto, and J. Liu, "Ten Challenges in Advancing Machine Learning Technologies toward 6G," *IEEE Wireless Communications*, vol. 27, no. 3, pp. 96–103, Jun. 2020.

[126] R. Minerva, G. M. Lee, and N. Crespi, "Digital twin in the iot context: a survey on technical features, scenarios, and architectural models," *Proceedings of the IEEE*, vol. 108, no. 10, pp. 1785–1824, 2020.

[127] B. R. Barricelli, E. Casiraghi, and D. Fogli, "A survey on digital twin: Definitions, characteristics, applications, and design implications," *IEEE Access*, vol. 7, pp. 167 653–167 671, 2019.

[128] T. Deng, K. Zhang, and Z.-J. M. Shen, "A systematic review of a digital twin city: A new pattern of urban governance toward smart cities," *Journal of Management Science and Engineering*, 2021.

[129] A. Klautau, N. González-Prelcic, and R. W. Heath, "LIDAR Data for Deep Learning-Based mmWave Beam-Selection," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 909–912, Jun. 2019, conference Name: IEEE Wireless Communications Letters.

[130] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of sumo-simulation of urban mobility," *International journal on advances in systems and measurements*, vol. 5, no. 3&4, 2012.

[131] M. Gschwandtner, R. Kwitt, A. Uhl, and W. Pree, "Blensor: Blender sensor simulation toolbox," in *International Symposium on Visual Computing*. Springer, 2011, pp. 199–208.

[132] P.-C. Chen, Y.-C. Chen, W.-H. Huang, C.-W. Huang, and O. Tirkkonen, "DDPG-Based Radio Resource Management for User Interactive Mobile Edge Networks," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*. IEEE, 2020, pp. 1–5.

[133] I. T. Union. (2121, Apr.) About International Telecommunication Union. [Online]. Available: https://www.itu.int/en/about/Pages/default.aspx

[134] U. Nations. (2121, Apr.) About the United Nations. [Online]. Available: https://www.un.org/en/

[135] A. N. de Telecomunicações. (2121, Apr.) Radiofrequência. [Online]. Available: https://www.gov.br/anatel/pt-br/regulado/radiofrequencia

[136] L. Chettri and R. Bera, "A comprehensive survey on Internet of things (IoT) toward 5G wireless systems," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 16–32, 2020.

[137] J. M. C. Brito, L. L. Mendes, and J. G. S. Gontijo, "Brazil 6G project - an approach to build a national-wise framework for 6G networks," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*, 2020, pp. 1–5.

[138] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal," in *Proceedings of the first international workshop on Technology and policy for accessing spectrum*, 2006, pp. 1–es.

[139] M. Lopez-Benitez, F. Casadevall, A. Umbert, J. Perez-Romero, R. Hachemani, J. Palicot, and C. Moy, "Spectral occupation measurements and blind standard recognition sensor for cognitive radio networks," in *2009 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, 2009, pp. 1–9.

[140] H. T. Madan and P. I. Basarkod, "A survey on efficient spectrum utilization for future wireless networks using cognitive radio approach," in *2018 4th International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, 2018, pp. 47–53.

[141] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical communication*, vol. 4, no. 1, pp. 40–62, 2011.

[142] X. Chen, H.-H. Chen, and W. Meng, "Cooperative communications for cognitive radio networks — from theory to applications," *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1180–1192, 2014.

[143] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.

[144] B. Nadler, F. Penna, and R. Garello, "Performance of eigenvalue-based signal detectors with known and unknown noise level," in *2011 IEEE Int. Conf. on Commun. (ICC)*, 2011, pp. 1–5.

[145] A. R, S. Y. Kulkarni, and S. N. Prasad, "Comparative study of narrowband and wideband opportunistic spectrum access techniques," in *2019 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, 2019, pp. 1–5.

[146] P. P. Jain, P. R. Pawar, P. Patil, and D. Pradhan, "Narrowband spectrum sensing in cognitive radio: Detection methodologies," *International Journal of Computer Sciences and Engineering*, 2019.

[147] Z. Tian and G. B. Giannakis, "A wavelet approach to wideband spectrum sensing for cognitive radios," in *2006 1st International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, 2006, pp. 1–5.

[148] R. Qiu, N. Guo, H. Li, Z. Wu, V. Chakravarthy, Y. Song, Z. Hu, P. Zhang, and Z. Chen, "A unified multi-functional dynamic spectrum access framework: tutorial, theory and multi-GHz wideband testbed," *Sensors*, vol. 9, no. 8, pp. 6530–6603, 2009.

[149] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1128–1140, 2009.

[150] F. B. de Carvalho, W. T. Lopes, and M. S. Alencar, "Performance of cognitive spectrum sensing based on energy detector in fading channels," *Procedia Computer Science*, vol. 65, pp. 140–147, 2015.

[151] X. Liu, B. G. Evans, and K. Moessner, "Comparison of reliability, delay and complexity for standalone cognitive radio spectrum sensing schemes," *IET Communications*, vol. 7, no. 9, pp. 799–807, 2013.

[152] H. Sun, A. Nallanathan, C.-X. Wang, and Y. Chen, "Wideband spectrum sensing for cognitive radio networks: a survey," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 74–81, 2013.

[153] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 1136–1159, 2013.

[154] A. K. Singh, H. Katiyar, R. Kumar, and S. Dixit, "Revolutionizing 5G: Cognitive machine learning," in *2020 International Conference on Electrical and Electronics Engineering (ICE3)*, 2020, pp. 17–20.

[155] H. A. Shah and I. Koo, "Reliable machine learning based spectrum sensing in cognitive radio networks," *Wireless Communications and Mobile Computing*, vol. 2018, 2018.

[156] K. M. Thilina, K. W. Choi, N. Saquib, and E. Hossain, "Machine learning techniques for cooperative spectrum sensing in cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2209–2221, 2013.

[157] M. Bkassiny, S. K. Jayaweera, Y. Li, and K. A. Avery, "Blind cyclostationary feature detection based spectrum sensing for autonomous self-learning cognitive radios," in *2012 IEEE International Conference on Communications (ICC)*, 2012, pp. 1507–1511.

[158] W. Lee, M. Kim, and D.-H. Cho, "Deep cooperative sensing: Cooperative spectrum sensing based on convolutional neural networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 3005–3009, 2019.

[159] J. Gao, X. Yi, C. Zhong, X. Chen, and Z. Zhang, "Deep learning for spectrum sensing," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1727–1730, 2019.

[160] Y. Molina-Tenorio, A. Prieto-Guerrero, and R. Aguilar-Gonzalez, "A novel multiband spectrum sensing method based on wavelets and the higuchi fractal dimension," *Sensors*, vol. 19, no. 6, p. 1322, 2019.

[161] Y. Molina-Tenorio, A. Prieto-Guerrero, R. Aguilar-Gonzalez, and S. Ruiz-Boqué, "Machine learning techniques applied to multiband spectrum sensing in cognitive radios," *Sensors*, vol. 19, no. 21, p. 4715, 2019.

[162] J. Tian, P. Cheng, Z. Chen, M. Li, H. Hu, Y. Li, and B. Vucetic, "A machine learning-enabled spectrum sensing method for OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 11 374–11 378, 2019.

[163] Z. Shi, W. Gao, S. Zhang, J. Liu, and N. Kato, "Machine learning-enabled cooperative spectrum sensing for non-orthogonal multiple access," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 5692–5702, 2020.

[164] Y. Xu, P. Cheng, Z. Chen, Y. Li, and B. Vucetic, "Mobile collaborative spectrum sensing for heterogeneous networks: A bayesian machine learning approach," *IEEE Transactions on Signal Processing*, vol. 66, no. 21, pp. 5634–5647, 2018.

[165] Q. Peng, A. Gilman, N. Vasconcelos, P. C. Cosman, and L. B. Milstein, "Robust deep sensing through transfer learning in cognitive radio," *IEEE Wireless Communications Letters*, vol. 9, no. 1, pp. 38–41, 2019.

[166] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 168–179, 2018.

[167] K. Davaslioglu and Y. E. Sagduyu, "Generative adversarial learning for spectrum sensing," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–6.

[168] T. Erpek, Y. E. Sagduyu, and Y. Shi, "Deep learning for launching and mitigating wireless jamming attacks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 1, pp. 2–14, 2018.

[169] W. Lee, M. Kim, and D.-H. Cho, "Deep cooperative sensing: Cooperative spectrum sensing based on convolutional neural networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 3005–3009, 2019.

[170] M. R. Vyas, D. K. Patel, and M. Lopez-Benitez, "Artificial neural network based hybrid spectrum sensing scheme for cognitive radio," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–7.

[171] Z. Ye, A. Gilman, Q. Peng, K. Levick, P. Cosman, and L. Milstein, "Comparison of neural network architectures for spectrum sensing," in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.

[172] K. Cichoń, H. Bogucka, G. Molis, J. Adamonis, and T. Krilavičius, "Learning and detection mechanisms of spectral-activity information towards energy efficient 5G communication," in *2018 Baltic URSI Symposium (URSI)*, 2018, pp. 273–277.

[173] X. Liu, Q. Sun, W. Lu, C. Wu, and H. Ding, "Big-data-based intelligent spectrum sensing for heterogeneous spectrum communications in 5G," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 67–73, 2020.

[174] T. Xu, T. Zhou, J. Tian, J. Sang, and H. Hu, "Intelligent spectrum sensing: When reinforcement learning meets automatic repeat sensing in 5G communications," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 46–53, 2020.

[175] A. Nasser, H. Al Haj Hassan, J. Abou Chaaya, A. Mansour, and K.-C. Yao, "Spectrum sensing for cognitive radio: Recent advances and future challenge," *Sensors*, vol. 21, no. 7, p. 2408, 2021.

[176] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133 995–134 030, 2020.

[177] S. J. Nawaz, S. K. Sharma, S. Wyne, M. N. Patwary, and M. Asaduzzaman, "Quantum machine learning for 6G communication networks: State-of-the-art and vision for the future," *IEEE Access*, vol. 7, pp. 46 317–46 350, 2019.

[178] H. Yang, A. Alphones, Z. Xiong, D. Niyato, J. Zhao, and K. Wu, "Artificial-intelligence-enabled intelligent 6G networks," *IEEE Network*, vol. 34, no. 6, pp. 272–280, 2020.

[179] H. Song, J. Bai, Y. Yi, J. Wu, and L. Liu, "Artificial intelligence enabled internet of things: Network architecture and spectrum access," *IEEE Computational Intelligence Magazine*, vol. 15, no. 1, pp. 44–51, 2020.

[180] F. Tariq, M. R. A. Khandaker, K.-K. Wong, M. A. Imran, M. Bennis, and M. Debbah, "A speculative study on 6G," *IEEE Wireless Communications*, vol. 27, no. 4, pp. 118–125, 2020.

[181] L. Mendes and et al, "Casos de Uso e Requisitos para Redes 6G)," Projeto Brasil 6G, Relatório Técnico, 2021, version 0.5.0.

[182] M. Leinonen, M. Codreanu, and G. B. Giannakis, "Compressed sensing with applications in wireless networks," *Foundations and Trends® in Signal Processing*, vol. 13, no. 1-2, pp. 1–282, 2019. [Online]. Available: http://dx.doi.org/10.1561/2000000107

[183] J. Liang, L. Li, and C. Zhao, "A transfer learning approach for compressed sensing in 6g-iot," *IEEE Internet of Things Journal*, vol. Early Access, pp. 1–1, 2021.

[184] R. Zhang, H. Zhao, and J. Zhang, "Distributed compressed sensing aided sparse channel estimation in fdd massive mimo system," *IEEE Access*, vol. 6, pp. 18 383–18 397, 2018.

[185] Y. Wang, N. J. Myers, N. González-Prelcic, and R. W. H. Jr., "Site-specific online compressive beam codebook learning in mmWave vehicular communication," 2020.

[186] P. de Souza and et al, "Compressive Learning in Communication Systems: A Neural Network Receiver for Detecting Compressed Signals in OFDM Systems," *IEEE Access*, p. under review, 2021.

[187] Q. Wang, R. Zhang, L.-L. Yang, and L. Hanzo, "Non-orthogonal multiple access: A unified perspective," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 10–16, 2018.

[188] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.

[189] Z. Wei, J. Yuan, D. W. K. Ng, M. Elkashlan, and Z. Ding, "A survey of downlink non-orthogonal multiple access for 5G wireless communication networks," *ZTE Commun.*, vol. 14, no. 4, pp. 17–25, 10 2016.

[190] Z. Ding, M. Peng, and H. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.

[191] Z. Ding, F. Adachi, and H. V. Poor, "Performance of MIMO-NOMA downlink transmissions," in *2015 IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–6.

[192] S. Alavi, K. Cumanan, Z. Ding, and A. Burr, "Beamforming techniques for non-orthogonal multiple access in 5G cellular networks," *IEEE Transactions on Vehicular Technology*, Jul. 2018.

[193] Y. Chen, J. Schaepperle, and T. Wild, "Comparing IDMA and NOMA with superimposed pilots based channel estimation in uplink," in *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2015, pp. 89–94.

[194] H. Nikopour and H. Baligh, "Sparse code multiple access," *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 332–336, 2013.

[195] Z. Yuan, G. Yu, W. Li, Y. Yuan, X. Wang, and J. Xu, "Multi-user shared access for internet of things," in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, 2016, pp. 1–5.

[196] X. Dai, Z. Zhang, B. Bai, S. Chen, and S. Sun, "Pattern Division Multiple Access: A New Multiple Access Technology for 5G," *IEEE Wireless Communications*, vol. 25, pp. 54–60, 04 2018.

[197] X. Dai, S. Chen, S. Sun, S. Kang, Y. Wang, Z. Shen, and J. Xu, "Successive interference cancelation amenable multiple access (SAMA) for future wireless communications," in *2014 IEEE International Conference on Communication Systems*, 2014, pp. 222–226.

[198] M. Liaqat, K. Noordin, T. Latef, and K. Dimyati, "Power-domain non orthogonal multiple access (PD-NOMA) in cooperative networks: an overview," *Wireless Networks*, vol. 26, 01 2020.

[199] G. Mazzini, "Power division multiple access," in *ICUPC '98. IEEE 1998 International Conference on Universal Personal Communications. Conference Proceedings*, vol. 1, 1998, pp. 543–546.

[200] C. Lin, Q. Chang, and X. Li, "A deep learning approach for MIMO-NOMA downlink signal detection," *Sensors*, vol. 19, no. 11, p. 2526, Jun. 2019.

[201] J.-M. Kang, I.-M. Kim, and C.-J. Chun, "Deep learning-based MIMO-NOMA with imperfect SIC decoding," *IEEE Systems Journal*, vol. 14, no. 3, pp. 3414–3417, 2020.

[202] Z. Ding, F. Adachi, and H. V. Poor, "The application of mimo to non-orthogonal multiple access," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 537–552, 2016.

[203] G. Aquino, T. Barbosa, M. Chafii, L. Mendes, and A. K. Gizzini, "MUSA grant-free access framework and blind detection receiver," *Journal of Communication and Information Systems*, 2021, under review.

[204] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8440–8450, 2018.

[205] J. Luo, J. Tang, D. K. C. So, G. Chen, K. Cumanan, and J. A. Chambers, "A deep learning-based approach to power minimization in multi-carrier NOMA with SWIPT," *IEEE Access*, vol. 7, pp. 17 450–17 460, 2019.

[206] Y. Zhang, X. Wang, and Y. Xu, "Energy-efficient resource allocation in uplink NOMA systems with deep reinforcement learning," in *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2019, pp. 1–6.

[207] J. Pan, N. Ye, A. Wang, and X. Li, "A deep learning-aided detection method for ftn-based noma," *Wireless Communications and Mobile Computing*, vol. 2020, pp. 1–11, 01 2020.

[208] M. Kim, N.-I. Kim, W. Lee, and D.-H. Cho, "Deep learning-aided SCMA," *IEEE Communications Letters*, vol. 22, no. 4, pp. 720–723, 2018.

[209] S. Simsir and N. Taspınar, "Channel estimation using neural network in orthogonal frequency division multiplexing-interleave division multiple access (OFDM-IDMA) system," in *2014 International Telecommunications Symposium (ITS)*, 2014, pp. 1–5.

[210] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless communications*, vol. 18, no. 3, pp. 10–21, 2011.

[211] M. Tayyab, X. Gelabert, and R. Jäntti, "A survey on handover management: From LTE to NR," *IEEE Access*, vol. 7, pp. 118 907–118 930, 2019.

[212] F. Du, G. Chen, and L. Qiu, "The Analysis of Mobility Performance in MmWave-μWave Heterogeneous Networks," in *2019 IEEE/CIC International Conference on Communications in China (ICCC)*. IEEE, 2019, pp. 908–913.

[213] B. Van Quang, R. V. Prasad, and I. Niemegeers, "A survey on handoffs—Lessons for 60 GHz based wireless systems," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 1, pp. 64–86, 2010.

[214] L. Sun, J. Hou, and T. Shu, "Optimal handover policy for mmwave cellular networks: A multi-armed bandit approach," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.

[215] H. Okamoto, T. Nishio, M. Morikura, K. Yamamoto, D. Murayama, and K. Nakahira, "Machine-learning-based throughput estimation using images for mmWave communications," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*. IEEE, 2017, pp. 1–6.

[216] Y. Koda, K. Yamamoto, T. Nishio, and M. Morikura, "Reinforcement learning based predictive handover for pedestrian-aware mmWave networks," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2018, pp. 692–697.

[217] Y. Koda, K. Nakashima, K. Yamamoto, T. Nishio, and M. Morikura, "End-to-end learning of proactive handover policy for camera-assisted mmWave networks using deep reinforcement learning," *arXiv preprint arXiv:1904.04585*, 2019.

[218] ——, "Handover management for mmwave networks with proactive performance prediction using camera images and deep reinforcement learning," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 2, pp. 802–816, 2019.

[219] ——, "Cooperative sensing in deep RL-based image-to-decision proactive handover for mmWave networks," in *2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2020, pp. 1–6.

[220] Y. Koda, J. Park, M. Bennis, K. Yamamoto, T. Nishio, and M. Morikura, "Distributed heteromodal split learning for vision aided mmWave received power prediction," *arXiv preprint arXiv:2007.08208*, 2020.

[221] A. Alkhateeb, I. Beltagy, and S. Alex, "Machine learning for reliable mmwave systems: Blockage prediction and proactive handoff," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2018, pp. 1055–1059.

[222] F. B. Mismar and B. L. Evans, "Partially blind handovers for mmWave new radio aided by sub-6 GHz LTE signaling," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2018, pp. 1–5.

[223] M. Alrabeiah, J. Booth, A. Hredzak, and A. Alkhateeb, "Viwi vision-aided mmwave beam tracking: Dataset, task, and baseline solutions," *arXiv preprint arXiv:2002.02445*, 2020.

[224] L. Yan, H. Ding, L. Zhang, J. Liu, X. Fang, Y. Fang, M. Xiao, and X. Huang, "Machine learning-based handovers for Sub-6 GHz and mmWave integrated vehicular networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 10, pp. 4873–4885, 2019.

[225] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation," *IEEE Communications Letters*, vol. 23, no. 4, pp. 652–655, 2019.

[226] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive mimo 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 232–257, 2020.

[227] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive mimo networks: Spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[228] T. Schenk, *RF imperfections in high-rate wireless systems: impact and digital compensation*. Springer Science & Business Media, 2008.

[229] Özlem Tugfe Demir and E. Björnson, "Channel estimation in massive mimo under hardware non-linearities: Bayesian methods versus deep learning," 2020.

[230] S. Ali, W. Saad, N. Rajatheva, K. Chang, D. Steinbach, B. Sliwa, C. Wietfeld, K. Mei, H. Shiri, H.-J. Zepernick, T. M. C. Chu, I. Ahmad, J. Huusko, J. Suutala, S. Bhadauria, V. Bhatia, R. Mitra, S. Amuru, R. Abbas, B. Shao, M. Capobianco, G. Yu, M. Claes, T. Karvonen, M. Chen, M. Girnyk, and H. Malik, "6g white paper on machine learning in wireless communication networks," 2020.

[231] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in ofdm systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, 2018.

[232] ——, "Power of deep learning for channel estimation and signal detection in ofdm systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, 2017.

[233] H. He, S. Jin, C.-K. Wen, F. Gao, G. Y. Li, and Z. Xu, "Model-driven deep learning for physical layer communications," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 77–83, 2019.

[234] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5g and beyond," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 3, pp. 615–637, 2021.

[235] J.-P. Hong, W. Choi, and B. D. Rao, "Sparsity controlled random multiple access with compressed sensing," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 998–1010, 2015.

[236] Y. Qiang, X. Shao, and X. Chen, "A model-driven deep learning algorithm for joint activity detection and channel estimation," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2508–2512, 2020.

[237] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5g: how to empower son with big data for enabling 5g," *IEEE network*, vol. 28, no. 6, pp. 27–33, 2014.

[238] W. Xia, G. Zheng, Y. Zhu, J. Zhang, J. Wang, and A. P. Petropulu, "A deep learning framework for optimization of miso downlink beamforming," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1866–1880, 2019.

[239] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in ofdm systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, 2017.

[240] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017, publisher: IEEE.

[241] L. J. Wong, W. H. C. I. au2, B. Flowers, R. M. Buehrer, A. J. Michaels, and W. C. Headley, "The RFML Ecosystem: A Look at the Unique Challenges of Applying Deep Learning to Radio Frequency Applications," 2020.

[242] G. Baldini, R. Giuliani, and C. Gentile, "An assessment of the impact of IQ imbalances on the physical layer authentication of IoT wireless devices," in *2019 Global IoT Summit (GIoTS)*, 2019, pp. 1–6.

[243] L. J. Wong, W. C. Headley, and A. J. Michaels, "Specific Emitter Identification Using Convolutional Neural Network-Based IQ Imbalance Estimators," *IEEE Access*, vol. 7, pp. 33 544–33 555, 2019.

[244] C. G. Wheeler and D. R. Reising, "Assessment of the impact of CFO on RF-DNA fingerprint classification performance," in *2017 International Conference on Computing, Networking and Communications (ICNC)*, 2017, pp. 110–114.

[245] S. S. Hanna and D. Cabric, "Deep learning based transmitter identification using power amplifier nonlinearity," in *2019 International Conference on Computing, Networking and Communications (ICNC)*, 2019, pp. 674–680.

[246] S. Yun, J.-M. Kang, I.-M. Kim, and J. Ha, "Deep artificial noise: Deep learning-based precoding optimization for artificial noise scheme," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 3465–3469, 2020.

[247] D. He, C. Liu, T. Q. S. Quek, and H. Wang, "Transmit antenna selection in MIMO wiretap channels: A machine learning approach," *IEEE Wireless Communications Letters*, vol. 7, no. 4, pp. 634–637, 2018.

[248] G. Szabó, S. Rácz, N. Reider, H. A. Munz, and J. Pető, "Digital twin: Network provisioning of mission critical communication in cyber physical production systems," in *2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*. IEEE, 2019, pp. 37–43.

[249] H. X. Nguyen, R. Trestian, D. To, and M. Tatipamula, "Digital twin for 5g and beyond," *IEEE Communications Magazine*, vol. 59, no. 2, pp. 10–15, 2021.

[250] E. Blogs, "Digital Twins catalyst reflections from Digital Transformation World," https://www.ericsson.com/en/blog/2019/6/digital-twins-catalyst-booth-reflections-from-digital-transformation-world, 2020, [*Online*; acessado em 07 de julho de 2021].

[251] Huawei, "Huawei launches industry's first site Digital Twins based 5G digital engineering solution," https://www.huawei.com/en/press-events/news/2020/2/site-digital-twins-based-5g-digital-engineering-solution, 2020, [*Online*; acessado em 07 de julho de 2021].

[252] E. Blogs, "5G simulation models with NVIDIA Omniverse platform," https://www.ericsson.com/en/blog/2021/4/5g-simulation-omniverse-platform, 2021, [*Online*; acessado em 07 de julho de 2021].

[253] P. Wilson, "State of smart cities in uk and beyond," *IET Smart Cities*, vol. 1, no. 1, pp. 19–22, 2019.

[254] Spirent Communications, "Simplifying 5G with the Network Digital Twin," White Paper, 2019.

[255] S. Rougerie and J. Israel, "Land mobile propagation satellite model based on 360° images," in *2021 15th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2021, pp. 1–4.

[256] F. Bastos, A. Oliveira, J. Borges, and A. Klautau, "Effects of environment model complexity in ray-tracing simulation for uav channels," *X Conferência Nacional em Comunicações, Redes e Segurança da Informação*, 2020.

[257] B. Li, Y. Shi, Z. Qi, and Z. Chen, "A survey on semantic segmentation," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018, pp. 1233–1240.

[258] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," *International Journal of Multimedia Information Retrieval*, vol. 7, pp. 87–93, 2018.

[259] A. Klautau, A. Oliveira, I. Pamplona, and W. Alves, "Generating MIMO Channels For 6G Virtual Worlds Using Ray-Tracing Simulations," in *IEEE Statistical Signal Processing Workshop*, 2021.

[260] M. C. V. team. Ade20k dataset. [Online]. Available: http://groups.csail.mit.edu/vision/datasets/ADE20K/

[261] N. Kato, B. Mao, F. Tang, Y. Kawamoto, and J. Liu, "Ten Challenges in Advancing Machine Learning Technologies toward 6G," *IEEE Wireless Communications*, vol. 27, no. 3, pp. 96–103, 2020.

[262] G.-T. ITU-T, "Transport network support of imt-2020/5G," 2018.

[263] M. Fiorani *et al.*, "Transport abstraction models for an SDN-controlled centralized RAN," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1406–1409, 2015.

[264] G.-T. ITU-T, "Application of optical transport network recommendations to 5G transport - serir g supplement 67," 2019.

[265] L. M. P. Larsen, A. Checko, and H. L. Christiansen, "A Survey of the Functional Splits Proposed for 5G Mobile Crosshaul Networks," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 146–172, 2019.

[266] L. Contreras *et al.*, "ifusion: Standards-based sdn architecture for carrier transport network," in *2019 IEEE Conference on Standards for Communications and Networking (CSCN)*. IEEE, 2019, pp. 1–7.

[267] P. Sehier *et al.*, "Transport evolution for the ran of the future," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 11, no. 4, pp. B97–B108, 2019.

[268] F. W. Murti *et al.*, "An Optimal Deployment Framework for Multi-Cloud Virtualized Radio Access Networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2251–2265, 2021.

[269] A. Garcia-Saavedra *et al.*, "Fluidran: Optimized vRAN/MEC orchestration," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 2366–2374.

[270] N. Molner *et al.*, "Optimization of an integrated fronthaul/backhaul network under path and delay constraints," *Ad Hoc Networks*, vol. 83, pp. 41–54, 2019.

[271] F. W. Murti, S. Ali, and M. Latva-aho, "Deep Reinforcement Based Optimization of Function Splitting in Virtualized Radio Access Networks," *arXiv preprint arXiv:2105.14731*, 2021.

[272] F. Musumeci *et al.*, "Latency-aware CU placement/handover in dynamic WDM access-aggregation networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 11, no. 4, pp. B71–B82, 2019.

[273] Y. Zhang, *Network Function Virtualization: Concepts and Applicability in 5G Networks*. John Wiley & Sons, 2018.

[274] F. Z. Morais *et al.*, "PlaceRAN: Optimal Placement for the Virtualized Next-Generation RAN," *arXiv preprint arXiv:2102.13192*, 2021.

[275] M. Masdari *et al.*, "An overview of virtual machine placement schemes in cloud computing," *Journal of Network and Computer Applications*, vol. 66, pp. 106–127, 2016.

[276] B. Yi *et al.*, "A comprehensive survey of network function virtualization," *Computer Networks*, vol. 133, pp. 212–262, 2018.

[277] A. Laghrissi and T. Taleb, "A survey on the placement of virtual resources and virtual network functions," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1409–1434, 2018.

[278] T. M. Moerland, J. Broekens, and C. M. Jonker, "Model-based reinforcement learning: A survey," *arXiv preprint arXiv:2006.16712*, 2020.

[279] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, "6g wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28–41, 2019.

[280] N. Chen and M. Okada, "Toward 6G Internet of Things and the Convergence With RoF System," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8719–8733, 2021.

[281] L. Yu, J. Wu, A. Zhou, E. G. Larsson, and P. Fan, "Massively distributed antenna systems with nonideal optical fiber fronthauls: A promising technology for 6g wireless communication systems," *IEEE Vehicular Technology Magazine*, vol. 15, no. 4, pp. 43–51, 2020.

[282] J. Wang, C. Liu, M. Zhu, A. Yi, L. Cheng, and G.-K. Chang, "Investigation of Data-Dependent Channel Cross-Modulation in Multiband Radio-Over-Fiber Systems," *Journal of Lightwave Technology*, vol. 32, no. 10, pp. 1861–1871, 2014.

[283] H. Yang, J. Zeng, Y. Zheng, H.-D. Jung, B. Huiszoon, J. Van Zantvoort, E. Tangdiongga, and A. Koonen, "Evaluation of effects of MZM nonlinearity on QAM and OFDM signals in RoF transmitter," in *2008 International Topical Meeting on Microwave Photonics jointly held with the 2008 Asia-Pacific Microwave Photonics Conference*. IEEE, 2008, pp. 90–93.

[284] W. Chen and W. I. Way, "Multichannel single-sideband SCM/DWDM transmission systems," *Journal of lightwave technology*, vol. 22, no. 7, p. 1679, 2004.

[285] C. Lim, A. T. Nirmalathas, K.-L. Lee, D. Novak, and R. Waterhouse, "Intermodulation distortion improvement for fiber–radio applications incorporating OSSB+ C modulation in an optical integrated-access environment," *Journal of lightwave technology*, vol. 25, no. 6, pp. 1602–1612, 2007.

[286] P. Horvath and I. Frigyes, "Effects of the nonlinearity of a mach-zehnder modulator on OFDM radio-over-fiber transmission," *IEEE communications letters*, vol. 9, no. 10, pp. 921–923, 2005.

[287] J. James, P. Shen, A. Nkansah, X. Liang, and N. J. Gomes, "Nonlinearity and noise effects in multi-level signal millimeter-wave over fiber transmission using single and dual wavelength modulation," *IEEE transactions on microwave theory and techniques*, vol. 58, no. 11, pp. 3189–3198, 2010.

[288] X. Zhang, "Broadband linearization for 5G fronthaul transmission," *Frontiers of Optoelectronics*, vol. 11, no. 2, pp. 107–115, 2018.

[289] T. Ismail, C.-P. Liu, J. E. Mitchell, and A. J. Seeds, "High-Dynamic-Range Wireless-Over-Fiber Link Using Feedforward Linearization," *Journal of Lightwave Technology*, vol. 25, no. 11, pp. 3274–3282, 2007.

[290] S. Korotky and R. de Ridder, "Dual parallel modulation schemes for low-distortion analog optical transmission," *IEEE Journal on Selected Areas in Communications*, vol. 8, no. 7, pp. 1377–1381, 1990.

[291] B. Masella, B. Hraimel, and X. Zhang, "Enhanced Spurious-Free Dynamic Range Using Mixed Polarization in Optical Single Sideband Mach–Zehnder Modulator," *Journal of Lightwave Technology*, vol. 27, no. 15, pp. 3034–3041, 2009.

[292] G.-W. Lee and S.-K. Han, "Linear dual electroabsorption modulator for analog optical transmission," *Microwave and Optical Technology Letters*, vol. 22, no. 6, pp. 369–373, 1999.

[293] S.-H. Lee, J.-M. Kang, I.-H. Choi, and S.-K. Han, "Linearization of DFB laser diode by external light-injected cross-gain modulation for radio-over-fiber link," *IEEE Photonics Technology Letters*, vol. 18, no. 14, pp. 1545–1547, 2006.

[294] X. Zhang, S. Saha, R. Zhu, T. Liu, and D. Shen, "Analog pre-distortion circuit for radio over fiber transmission," *IEEE Photonics Technology Letters*, vol. 28, no. 22, pp. 2541–2544, 2016.

[295] C. Yin, J. Li, H. Chen, Q. Lv, Y. Fan, F. Yin, Y. Dai, and K. Xu, "Behavioral modeling and digital compensation of nonlinearity in multi-band externally-modulated radio-over-fiber links," in *2016 25th Wireless and Optical Communication Conference (WOCC)*. IEEE, 2016, pp. 1–4.

[296] J. Li, C. Yin, H. Chen, F. Yin, Y. Dai, and K. Xu, "Behavioral modeling and digital compensation of nonlinearity in DFB lasers for multi-band directly modulated radio-over-fiber systems," in *Semiconductor Lasers and Applications VI*, vol. 9267. International Society for Optics and Photonics, 2014, p. 92670K.

[297] A. Hekkala, M. Hiivala, M. Lasanen, J. Perttu, L. C. Vieira, N. J. Gomes, and A. Nkansah, "Predistortion of Radio Over Fiber Links: Algorithms, Implementation, and Measurements," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 3, pp. 664–672, 2012.

[298] M. Noweir, Q. Zhou, A. Kwan, R. Valivarthi, M. Helaoui, W. Tittel, and F. M. Ghannouchi, "Digitally linearized radio-over fiber transmitter architecture for cloud radio access network's downlink," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 7, pp. 3564–3574, 2018.

[299] R. M. Borges, L. A. M. Pereira, H. R. D. Filgueiras, A. C. Ferreira, M. S. B. Cunha, E. R. Neto, D. H. Spadoti, L. L. Mendes, and A. Cerqueira, "DSP-based flexible-waveform and multi-application 5G fiber-wireless system," *Journal of Lightwave Technology*, vol. 38, no. 3, pp. 642–653, 2019.

[300] J. He, J. Lee, S. Kandeepan, and K. Wang, "Machine Learning Techniques in Radio-over-Fiber Systems and Networks," in *Photonics*, vol. 7, no. 4. Multidisciplinary Digital Publishing Institute, 2020, p. 105.

[301] F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1383–1408, 2018.

[302] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An Optical Communication's Perspective on Machine Learning and Its Applications," *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 493–516, 2019.

[303] A. C. Najarro and S.-M. Kim, "Nonlinear compensation using artificial neural network in radio-over-fiber system," *Journal of information and communication convergence engineering*, vol. 16, no. 1, pp. 1–5, 2018.

[304] M. U. Hadi, M. Awais, M. Raza, K. Khurshid, and H. Jung, "Neural network DPD for aggrandizing SM-VCSEL-SSMF-Based radio over fiber link performance," in *Photonics*, vol. 8, no. 1. Multidisciplinary Digital Publishing Institute, 2021, p. 19.

[305] L. A. M. Pereira, P. H. F. Santos, R. M. Borges, L. L. Mendes, C. J. A. Bastos Filho, and A. C. S. Junior, "Sistema Rádio sobre Fibra assistido por Inteligência Artificial para aplicações 5G/6G," *Brazilian Journal of Development*, vol. 7, no. 5, pp. 48 948–48 958, 2021.

[306] Y. Cui, M. Zhang, D. Wang, S. Liu, Z. Li, and G.-K. Chang, "Bit-based support vector machine nonlinear detector for millimeter-wave radio-over-fiber mobile fronthaul systems," *Optics express*, vol. 25, no. 21, pp. 26 186–26 197, 2017.

[307] S. Liu, M. Xu, J. Wang, F. Lu, W. Zhang, H. Tian, and G.-K. Chang, "A multilevel artificial neural network nonlinear equalizer for millimeter-wave mobile fronthaul systems," *Journal of Lightwave Technology*, vol. 35, no. 20, pp. 4406–4417, 2017.

[308] E. Liu, Z. Yu, C. Yin, and K. Xu, "Nonlinear Distortions Compensation Based on Artificial Neural Networks in Wideband and Multi-Carrier Systems," *IEEE Journal of Quantum Electronics*, vol. 55, no. 5, pp. 1–5, 2019.

[309] S. Liu, Y. M. Alfadhli, S. Shen, M. Xu, H. Tian, and G.-K. Chang, "A novel ANN equalizer to mitigate nonlinear interference in analog-RoF mobile fronthaul," *IEEE Photonics Technology Letters*, vol. 30, no. 19, pp. 1675–1678, 2018.

[310] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, "6G Wireless Networks: Vision, Requirements, Architecture, and Key Technologies," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28–41, 2019.

[311] X. You, C.-X. Wang, J. Huang, X. Gao, Z. Zhang, M. Wang, Y. Huang, C. Zhang, Y. Jiang, J. Wang, M. Zhu, B. Sheng, D. Wang, Z. Pan, P. Zhu, Y. Yang, Z. Liu, P. Zhang, X. Tao, S. Li, Z. Chen, X. Ma, C.-L. I, S. Han, K. Li, C. Pan, Z. Zheng, L. Hanzo, X. S. Shen, Y. J. Guo, Z. Ding, H. Haas, W. Tong, P. Zhu, G. Yang, J. Wang, E. G. Larsson, H. Q. Ngo, W. Hong, H. Wang, D. Hou, J. Chen, Z. Chen, Z. Hao, G. Y. Li, R. Tafazolli, Y. Gao, H. V. Poor, G. P. Fettweis, and Y.-C. Liang, "Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts," *Science China Information Sciences*, vol. 64, no. 1, p. 110301, Nov 2020. [Online]. Available: https://doi.org/10.1007/s11432-020-2955-6

[312] G.-H. Kim, I. Mahmud, and Y.-Z. Cho, "Hello-Message Transmission-Power Control for Network Self-Recovery in FANETs," in *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, 2018, pp. 546–548.

[313] H. Shakhatreh, A. H. Sawalmeh, A. Al-Fuqaha, Z. Dou, E. Almaita, I. Khalil, N. S. Othman, A. Khreishah, and M. Guizani, "Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges," *IEEE Access*, vol. 7, pp. 48 572–48 634, 2019.

[314] P. Zhou, X. Fang, Y. Fang, R. He, Y. Long, and G. Huang, "Beam Management and Self-Healing for mmWave UAV Mesh Networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1718–1732, 2019.

[315] 3GPP-TR21.915, "Technical Specification Group Services and System Aspects; Release 15 Description," 3rd Generation Partnership Project (3GPP), Technical Report, 2018-12, version 0.5.0.

[316] Citrix Systems, Inc., "What is a load balancer? - load balancing definition - citrix," https://www.citrix.com/glossary/load-balancing.html, 2020, acessado em 15/06/2020.

[317] A. M. Alakeel *et al.*, "A guide to dynamic load balancing in distributed computer systems," *International Journal of Computer Science and Information Security*, vol. 10, no. 6, pp. 153–160, 2010.

[318] NGINX, "What is load balancing? how load balancers work," https://www.nginx.com/resources/glossary/load-balancing/, 2020, acessado em 15/06/2020.

[319] T. V. K. Buyakar, *et al.*, "Prototyping and Load Balancing the Service Based Architecture of 5G Core Using NFV," in *2019 IEEE Conference on Network Softwarization (NetSoft)*. França: IEEE, 2019, pp. 228–232.

[320] I. Alawe *et al.*, "On the scalability of 5G Core network: the AMF case," in *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, IEEE. França: IEEE, 2018, pp. 1–6.

[321] N. Salhab, R. Rahim, and R. Langar, "NFV Orchestration Platform for 5G over On-the-fly Provisioned Infrastructure," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, IEEE. França: IEEE, 2019, pp. 971–972.

[322] C. H. T. Arteaga *et al.*, "A scaling mechanism for an evolved packet core based on network functions virtualization," *IEEE Transactions on Network and Service Management*, vol. 17, no. 2, pp. 779–792, 2019.

[323] A. Banerjee *et al.*, "Scaling the lte control-plane for future mobile access," in *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*. USA: ACM, 2015, pp. 1–13.

[324] ETSI GS NFV-MAN, "Network Functions Virtualisation (NFV); Management and Orchestration," Dec. 2014.

[325] NGMN Alliance, "Description of Network Slicing Concept," 2016.

[326] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network Slicing in 5G: Survey and Challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[327] F. Debbabi, R. Jmal, L. C. Fourati, and A. Ksentini, "Algorithmics and Modeling Aspects of Network Slicing in 5G and Beyonds Network: Survey," *IEEE Access*, vol. 8, pp. 162 748–162 762, 2020.

[328] 3GPP, "Telecommunication Management; Study on Man- agement and Orchestration of Network Slicing for Next Generation Network," 2018.

[329] D. M. Gutierrez-Estevez, M. Gramaglia, A. D. Domenico, G. Dandachi, S. Khatibi, D. Tsolkas, I. Balan, A. Garcia-Saavedra, U. Elzur, and Y. Wang, "Artificial Intelligence for Elastic Management and Orchestration of 5G Networks," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 134–141, 2019.

[330] R. Su, D. Zhang, R. Venkatesan, Z. Gong, C. Li, F. Ding, F. Jiang, and Z. Zhu, "Resource Allocation for Network Slicing in 5G Telecommunication Networks: A Survey of Principles and Models," *IEEE Network*, vol. 33, no. 6, pp. 172–179, 2019.

[331] S. Haeri and L. Trajković, "Virtual Network Embedding via Monte Carlo Tree Search," *IEEE Transactions on Cybernetics*, vol. 48, no. 2, pp. 510–521, 2018.

[332] P. T. A. Quang, Y. Hadjadj-Aoul, and A. Outtagarts, "A Deep Reinforcement Learning Approach for VNF Forwarding Graph Embedding," *IEEE Transactions on Network and Service Management*, vol. 16, no. 4, pp. 1318–1331, 2019.

[333] H. Bin and H. D. SCHOTTEN, "Machine Learning for Network Slicing Resource Management: A Comprehensive Survey," *ZTE COMMUNICATIONS*, vol. 17, no. 4, pp. 27–32, 2019.

[334] B. Han, D. Feng, and H. D. Schotten, "A Markov Model of Slice Admission Control," *IEEE Networking Letters*, vol. 1, no. 1, pp. 2–5, 2019.

[335] D. Bega, M. Gramaglia, A. Banchs, V. Sciancalepore, K. Samdanis, and X. Costa-Perez, "Optimising 5G infrastructure markets: The business of network slicing," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, 2017, pp. 1–9.

[336] L. Zanzi, V. Sciancalepore, A. Garcia-Saavedra, H. D. Schotten, and X. Costa-Pérez, "LACO: A Latency-Driven Network Slicing Orchestration in Beyond-5G Networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 667–682, 2021.

[337] D. Bega, M. Gramaglia, M. Fiore, A. Banchs, and X. Costa-Perez, "DeepCog: Cognitive Network Management in Sliced 5G Networks with Deep Learning," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019, pp. 280–288.

[338] B. Han, J. Lianghai, and H. D. Schotten, "Slice as an Evolutionary Service: Genetic Optimization for Inter-Slice Resource Management in 5G Networks," *IEEE Access*, vol. 6, pp. 33 137–33 147, 2018.

[339] S. D'Oro, L. Galluccio, P. Mertikopoulos, G. Morabito, and S. Palazzo, "Auction-based resource allocation in OpenFlow multi-tenant networks," *Computer Networks*, vol. 115, pp. 29–41, 2017.

[340] R. Li, Z. Zhao, Q. Sun, C.-L. I, C. Yang, X. Chen, M. Zhao, and H. Zhang, "Deep Reinforcement Learning for Resource Management in Network Slicing," *IEEE Access*, vol. 6, pp. 74 429–74 441, 2018.

[341] B. Han, A. DeDomenico, G. Dandachi, A. Drosou, D. Tzovaras, R. Querio, F. Moggio, O. Bulakci, and H. D. Schotten, "Admission and Congestion Control for 5G Network Slicing," in *2018 IEEE Conference on Standards for Communications and Networking (CSCN)*, 2018, pp. 1–6.

[342] C. Benzaid and T. Taleb, "AI-Driven Zero Touch Network and Service Management in 5G and Beyond: Challenges and Research Directions," *IEEE Network*, vol. 34, no. 2, pp. 186–194, 2020.

[343] M. Moradi *et al.*, "Skycore: Moving Core to the Edge For Untethered and Reliable UAV-Based LTE Networks," *GetMobile: Mobile Computing and Communications*, vol. 23, no. 1, pp. 24–29, 2019.

[344] J. Košmerl and A. Vilhar, "Base stations placement optimization in wireless networks for emergency communications," in *2014 IEEE International Conference on Communications Workshops (ICC)*, 2014, pp. 200–205.

[345] I. Bor-Yaliniz, M. Salem, G. Senerath, and H. Yanikomeroglu, "Is 5G ready for drones: A look into contemporary and prospective wireless networks from a standardization perspective," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 18–27, 2019.

[346] U. Fattore, M. Liebsch, and C. J. Bernardos, "Upflight: An enabler for avionic mec in a drone-extended 5G mobile network," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–7.

[347] ITU-T Rec. Y.3172, "Architectural framework for machine learning in future networks including IMT-2020," 2019.

[348] ETSI, "European Telecommunications Standards Institute," . https://www.etsi.org/, 2021, acesso: 09-Maio-2021.

[349] 3GPP-TR21.916, "Technical Specification Group Services and System Aspects; Release 16 Description," 3rd Generation Partnership Project (3GPP), Technical Report, 2020-09, version 0.6.0.

[350] O-RAN, "O-RAN Alliance," . https://www.o-ran.org/, 2021, acesso: 09-Maio-2021.

[351] Y. Wang *et al.*, "From Design to Practice: ETSI ENI Reference Architecture and Instantiation for Network Management and Orchestration Using Artificial Intelligence," *IEEE Communications Standards Magazine*, vol. 4, no. 3, pp. 38–45, 2020.

[352] ETSI GS ENI, "Experiential Networked Intelligence (ENI); System Architecture," 2019.

[353] Q. Duan, "Intelligent and Autonomous Management in Cloud-Native Future Network Survey on Related Standards from an Architectural Perspective," in *Future Internet*, vol. 13, no. 2, 2021.

[354] ITU-T FG-ML5G, "Unified Architecture for Machine Learning in 5G and Future Network," 2019.

[355] O-RAN Alliance, "O-RAN working group 2: AI/ML workflow description and requirements," 2019.

[356] R. Shafin *et al.*, "Artificial Intelligence-Enabled Cellular Networks: A Critical Path to Beyond-5G and 6G," *IEEE Wireless Communications*, vol. 27, no. 2, pp. 212–217, 2020.

[357] ETSI. GS ZSM, "Zero-touch network and Service Management (ZSM); Reference Architecture," 2019.

[358] N. Kato, B. Mao, F. Tang, Y. Kawamoto, and J. Liu, "Ten Challenges in Advancing Machine Learning Technologies toward 6G," *IEEE Wireless Communications*, vol. 27, no. 3, pp. 96–103, 2020.

[359] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities, and challenges," *IEEE Communications Magazine*, vol. 58, no. 6, pp. 46–51, 2020, publisher: IEEE.

[360] J. Konečnỳ, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," *arXiv preprint arXiv:1610.02527*, 2016.

[361] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 1273–1282.

[362] H. B. McMahan, D. Ramage, K. Talwar, and L. Zhang, "Learning differentially private recurrent language models," *arXiv preprint arXiv:1710.06963*, 2017.

[363] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konečnỳ, S. Mazzocchi, H. B. McMahan *et al.*, "Towards federated learning at scale: System design," *arXiv preprint arXiv:1902.01046*, 2019.

[364] Y. Liu, X. Yuan, Z. Xiong, J. Kang, X. Wang, and D. Niyato, "Federated learning for 6G communications: Challenges, methods, and future directions," *China Communications*, vol. 17, no. 9, pp. 105–118, 2020.

[365] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[366] J. Park, S. Samarakoon, M. Bennis, and M. Debbah, "Wireless Network Intelligence at the Edge," *Proceedings of the IEEE*, vol. 107, no. 11, pp. 2204–2239, 2019.

[367] Y. Shi, K. Yang, T. Jiang, J. Zhang, and K. B. Letaief, "Communication-Efficient Edge AI: Algorithms and Systems," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2167–2191, 2020.

[368] Y. Xiao, G. Shi, Y. Li, W. Saad, and H. V. Poor, "Toward Self-Learning Edge Intelligence in 6G," *IEEE Communications Magazine*, vol. 58, no. 12, pp. 34–40, Dec. 2020.

[369] G. Zhu, D. Liu, Y. Du, C. You, J. Zhang, and K. Huang, "Toward an Intelligent Edge: Wireless Communication Meets Machine Learning," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 19–25, 2020.

[370] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated Learning via Over-the-Air Computation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2022–2035, 2020.

[371] R. Alvizu, S. Troia, G. Maier, and A. Pattavina, "Matheuristic With Machine-Learning-Based Prediction for Software-Defined Mobile Metro-Core Networks," *J. Opt. Commun. Netw.*, vol. 9, no. 9, pp. D19–D30, Sep 2017.

[372] T. Zhang and S. Mao, "Machine Learning for End-to-End Congestion Control," *IEEE Communications Magazine*, vol. 58, no. 6, pp. 52–57, 2020.

[373] I. Alawe, A. Ksentini, Y. Hadjadj-Aoul, and P. Bertin, "Improving Traffic Forecasting for 5G Core Network Scalability: A Machine Learning Approach," *IEEE Network*, vol. 32, no. 6, pp. 42–49, 2018.

[374] Y. Zhao, B. Yan, D. Liu, Y. He, D. Wang, and J. Zhang, "SOON: self-optimizing optical networks with machine learning," *Opt. Express*, vol. 26, no. 22, pp. 28 713–28 726, Oct 2018.

[375] Y. Liu, X. Wang, J. Mei, G. Boudreau, H. Abou-Zeid, and A. B. Sediq, "Situation-Aware Resource Allocation for Multi-Dimensional Intelligent Multiple Access: A Proactive Deep Learning Framework," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 116–130, Jan. 2021, conference Name: IEEE Journal on Selected Areas in Communications.

[376] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1738–1762, 2019.

[377] S. Deng, H. Zhao, W. Fang, J. Yin, S. Dustdar, and A. Y. Zomaya, "Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7457–7469, 2020.

[378] J. Park, S. Samarakoon, A. Elgabli, J. Kim, M. Bennis, S.-L. Kim, and M. Debbah, "Communication-Efficient and Distributed Learning Over Wireless Networks: Principles and Applications," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 796–819, 2021.

[379] Y. Du and K. Huang, "Fast Analog Transmission for High-Mobility Wireless Data Acquisition in Edge Learning," *IEEE Wireless Communications Letters*, vol. 8, no. 2, pp. 468–471, 2019.