Inatel

OpenRAN
A CONEXÃO DO FUTURO

Antônio Marcos Alberti
Daniely Gomes Silva
Felipe Augusto Pereira de Figueiredo
Francisco de Assis Silva do Carmo
Guilherme Pedro Aquino
José Marcos Camara Brito

Instituto Nacional de Telecomunicações

ÍNDICE

1. INTRODUÇÃO

A QUINTA GERAÇÃO DAS REDES DE COMUNICAÇÕES MÓVEIS (5G) ESTÁ SENDO APONTADA COMO A PRÓXIMA GRANDE REVOLUÇÃO DOS SISTEMAS DE COMUNICAÇÃO.

Uma nova geração de redes de comunicações móveis surge, aproximadamente, a cada dez anos. A evolução da segunda geração (2G), a primeira geração digital, que surgiu na década de 1990, para a terceira geração (3G), e desta para a quarta geração (4G), sempre teve como principal requisito de desempenho a oferta de maiores taxas de transmissão para os usuários e, consequentemente, a possibilidade de implementação de novos serviços.

A quinta geração das redes de comunicações móveis (5G) está sendo apontada como a próxima grande revolução dos sistemas de comunicação. No 5G o cenário de oferta de maiores taxas de transmissão também está presente, mas agora outros importantes cenários de uso, com requisitos de desempenho específicos, estão sendo contemplados.

1.1. CENÁRIOS DE USO DO 5G

Comumente três cenários de uso são associados ao 5G: ***Enhanced Mobile Broadband*** (eMBB), ***Massive Machine-Type Communication*** (mMTC) e ***Ultra-Reliable and Low Latency Communications*** (URLLC). Um quarto cenário de uso, particularmente importante para o Brasil, é o cenário ***Enhanced Remote Area Communications*** (eRAC). Cada um destes quatro cenários está brevemente descrito e caracterizado, em termos de requisitos de desempenho e aplicações, nas seções 1.1.1 a 1.1.4, a seguir.

OS CENÁRIOS DE USO PREVISTOS PARA O 5G POSSUEM APLICAÇÕES NOS MAIS DIVERSOS SETORES DA ECONOMIA, DENOMINADOS DE VERTICAIS DE MERCADO

Os cenários de uso previstos para o 5G possuem aplicações nos mais diversos setores da economia, denominados de verticais de mercado, tais como: educação, saúde, agricultura e pecuária, indústria, entretenimento, comércio, cidades inteligentes, construção civil, indústria automotiva, logística e transporte, segurança, mineração, ***utilities***, financeiro e governo.

As novas aplicações viabilizadas pelo 5G terão enorme impacto econômico e social. Por exemplo, o estudo ***5G Economy: How 5G Technology will Contribute to the Global Economy***, realizado pelo ***Information Handling Services*** (IHS), aponta que o 5G irá gerar receitas anuais de até US$ 3,5 trilhões e contribuir na produção de bens e serviços no valor de até US$ 12,3 trilhões no ano de 2035. Para tal, a cadeia de valor da rede 5G deverá investir US$ 200 bilhões por ano [1].

5G IRÁ GERAR RECEITAS ANUAIS DE ATÉ US$ 3,5 TRILHÕES E CONTRIBUIR NA PRODUÇÃO DE BENS E SERVIÇOS NO VALOR DE ATÉ US$ 12,3 TRILHÕES NO ANO DE 2035

Os números acima, como vários outros apresentados em diferentes estudos, apontam para o potencial transformador da tecnologia 5G e para sua capacidade de geração de riquezas, além do potencial de contribuir para reduzir problemas sociais, rompendo com a barreira para o acesso à informação existente no mundo. Assim, é incontestável a relevância das redes 5G no desenvolvimento global no futuro próximo e, como consequência, a necessidade de que o Brasil invista na viabilização desta tecnologia para se manter competitivo no cenário mundial.

1.1.1. ENHANCED MOBILE BROADBAND

Este é o cenário voltado para aplicações que demandam a oferta de maiores taxas de transmissão de dados para os usuários e, consequentemente, uma maior capacidade da rede para escoar tráfego. Trata-se do cenário comumente associado à evolução das redes de comunicações móveis.

AS ALTAS TAXAS DE TRANSMISSÃO OFERTADAS PELO CENÁRIO EMBB VIABILIZARÃO UMA SÉRIE DE NOVAS APLICAÇÕES, COMO: IMAGENS 3D, HOLOGRAMAS, ***STREAMING*** DE VÍDEO DE ALTÍSSIMA RESOLUÇÃO, REALIDADE AUMENTADA, REALIDADE VIRTUAL, PRESENÇA VIRTUAL, ROBÔS COLABORATIVOS E ***CLOUD ROBOTICS***.

Os requisitos de desempenho definidos pelo ***International Telecommunication Union*** (ITU) para este cenário são: taxa de dados experimentada pelo usuário de 100 Mbps, em áreas urbanas e suburbanas, e de 1 Gbps em ***hotspots***, com uma taxa de pico de 20 Gbps (20 vezes superior à taxa de pico de 1 Gbps das redes 4G); e capacidade de tráfego por área de 10 Mbps/m^2 (100 vezes maior que a capacidade das redes 4G) [2].

As altas taxas de transmissão ofertadas pelo cenário eMBB viabilizarão uma série de novas aplicações, como: imagens 3D, hologramas, ***streaming*** de vídeo de altíssima resolução, realidade aumentada, realidade virtual, presença virtual, robôs colaborativos e ***cloud robotics***.

Estudo apresentado pela ***Global System for Mobile Association*** (GSMA), sob análises conservadoras, aponta que as comunicações em ondas milimétricas nas redes 5G (cenário eMBB) devem resultar, em 2034, em um incremento de US$ 565 bilhões no Produto Interno Bruto (PIB) mundial. Para a América Latina, o estudo aponta que este aumento será de US$ 20,8 bilhões, representando um aumento de 1,2% no PIB da região, tendo o Brasil, México e Colômbia como os principais impulsionadores deste aumento, contribuindo, respectivamente, com 47%, 29% e 9% do aumento de 1,2% no PIB. Este mesmo estudo aponta que os principais benefícios trazidos pelas aplicações de 5G voltadas para o cenário eMBB são: melhorias na saúde e maior expectativa de vida da população; incremento na independência e autonomia das pessoas; redução na poluição; incremento no acesso à educação; incremento no acesso aos cuidados com a saúde; melhoria na segurança pública e na resposta em situações de emergência; e melhoria da mobilidade urbana, com tempos de deslocamento menores [3].

As aplicações vislumbradas para o cenário eMBB estão normalmente associadas às áreas urbanas densamente povoadas, ambientes internos (por exemplo, ***shopping centers*** e prédios de escritórios) e eventos com grandes multidões, como eventos esportivos e ***shows*** em estádios.

1.1.2.

MASSIVE MACHINE-TYPE COMMUNICATION

O 5G É CONSIDERADO COMO A SOLUÇÃO QUE VIABILIZARÁ A IOT DE FORMA MASSIVA.

Este é o cenário voltado para viabilizar as aplicações de Internet das Coisas (do inglês, ***Internet of Things*** - IoT) de forma massiva. Aplicações envolvendo comunicações ***Machine to Machine*** (M2M) e aplicações não críticas de comunicações ***Vehicle to Vehicle*** (V2V) também são relacionadas a este cenário.

O 5G é considerado como a solução que viabilizará a IoT de forma massiva. Diversas previsões apontam que as aplicações de IoT evoluirão para um cenário onde haverá um enorme número de terminais. Por exemplo, o 5G ***Public Private Partnership*** (5GPPP) prevê que o número de terminais IoT conectados às redes 5G pode chegar a 1 trilhão de terminais [4].

Outra questão relevante para este cenário é a eficiência energética, uma vez que algumas aplicações de IoT demandam que o tempo de duração da bateria dos dispositivos seja muito longo, como, por exemplo, 10 anos. O número massivo de terminais e a questão energética estão refletidos nos requisitos de desempenho definidos pelo ITU para este cenário: uma densidade de conexões de 10^6 dispositivos/km^2 (10 vezes maior que o definido para o 4G) e uma eficiência energética 100 vezes maior que a definida para o 4G [2].

O mMTC é o segundo cenário com maior impacto potencial no Brasil nos curto e médio prazos, com as aplicações de IoT em diversas verticais de mercado. O Plano Nacional de IoT, definido para o país, priorizou quatro verticais de mercado, em função dos impactos que podem ter na economia e na sociedade brasileira, a saber: cidades, indústria, rural e saúde. Destas verticais, três estão principalmente associadas ao cenário mMTC, sendo que a vertical rural se relaciona mais fortemente com o cenário eRAC.

PARA A ÁREA DA SAÚDE, ESTIMA-SE QUE AS APLICAÇÕES DE IOT POSSAM TRAZER UM GANHO ECONÔMICO POTENCIAL PARA O PAÍS, ATÉ 2025, ENTRE US$ 5 E US$ 39 BILHÕES.

Para a área da saúde, estima-se que as aplicações de IoT possam trazer um ganho econômico potencial para o país, até 2025, entre US$ 5 e US$ 39 bilhões. Estes ganhos não estão relacionados apenas com aplicações voltadas para o controle, tratamento e prevenção de doenças, mas também para aplicações relacionadas com a gestão do sistema de saúde e com a interação do usuário com o sistema de saúde [5].

A aplicação da IoT no ambiente das cidades também pode resultar em ganhos econômicos e sociais significativos. Para o Brasil, estima-se que, em 2025, o ganho econômico será de US$ 27 bilhões. As aplicações com maior potencial de impacto no ambiente das cidades estão relacionadas com eficiência energética, saneamento e gestão de recursos hídricos, monitoramento de tráfego e mobilidade urbana, e na área da segurança pública [6].

O NÚMERO DE CONEXÕES IOT NA AMÉRICA LATINA DEVE EVOLUIR DE 444 MILHÕES DE CONEXÕES EM 2017 PARA 1,31 BILHÕES DE CONEXÕES EM 2025.

O número de conexões IoT na América Latina deve evoluir de 444 milhões de conexões em 2017 para 1,31 bilhões de conexões em 2025. As verticais e aplicações que mais contribuirão para este crescimento são: casas inteligentes (27%), construções inteligentes (18%), eletrônica de consumo (16%), ***smart utilities*** (10%), outras aplicações empresariais (10%), vendas no varejo (4%), ***wearables*** (4%), saúde (3%), fábricas inteligentes (2%), veículos inteligentes (2%) e cidades inteligentes (2%) [7].

Um estudo do IHS aponta que os setores que serão mais impactados pelas aplicações do cenário mMTC são: manufatura; mineração; serviços públicos; transporte e armazenamento; vendas de atacado e varejo. Ainda, este estudo aponta que o cenário mMTC contribuirá com US$ 3,6 trilhões (29,3%) dos US$ 12,3 trilhões que o 5G irá gerar em 2035 [1].

1.1.3. ULTRA-RELIABLE AND LOW LATENCY COMMUNICATIONS

Este cenário está associado a aplicações que demandam latência muito baixa e alta confiabilidade da rede, como, dentre outras, aplicações de Internet Tátil. Os requisitos de desempenho especificados para este cenário são uma latência fim-a-fim máxima de 1 ***ms*** e probabilidade de indisponibilidade menor ou igual a 10^{-7} (também denominado de sistema com confiabilidade de sete 9s). A título de comparação, a latência fim-a-fim das redes 4G é de aproximadamente 20 ***ms*** [8].

Assim como para os cenários anteriores, o cenário URLLC também possui aplicações nas mais diversas verticais. Por exemplo, a ITU identificou aplicações de Internet Tátil nas seguintes áreas: robótica, telepresença, realidade aumentada, educação, indústria, realidade virtual, controle de tráfego, jogos (de entretenimento, educativos e simuladores), cultura, ***smart grids*** e aplicações de saúde [8]. Além das aplicações de Internet Tátil, o cenário URLLC também se relaciona a outras aplicações que demandam comunicações ultra confiáveis, como comunicações V2V para carros autônomos.

AS VERTICAIS QUE TERÃO MAIORES IMPACTOS COM O CENÁRIO URLLC SÃO: A INDÚSTRIA AUTOMOTIVA, COM OS CARROS AUTÔNOMOS; A ÁREA DA SAÚDE, COM APLICAÇÕES SOFISTICADAS DE DIAGNÓSTICO, CIRURGIA E REABILITAÇÃO REMOTAS; A ÁREA DA INDÚSTRIA, COM O USO DE GRUPOS DE ROBÔS TRABALHANDO DE FORMA COOPERATIVA E EM TEMPO REAL; A ÁREA DO ENTRETENIMENTO, COM JOGOS EM TEMPO REAL; E AS ÁREAS DE LOGÍSTICA E AMBIENTAL, COM REDES DE DRONES CONECTADOS PROVENDO APLICAÇÕES DE ENTREGA E DE MONITORAMENTO AMBIENTAL.

As verticais que terão maiores impactos com o cenário URLLC são: a indústria automotiva, com os carros autônomos; a área da saúde, com aplicações sofisticadas de diagnóstico, cirurgia e reabilitação remotas; a área da indústria, com o uso de grupos de robôs trabalhando de forma cooperativa e em tempo real; a área do entretenimento, com jogos em tempo real; e as áreas de logística e ambiental, com redes de drones conectados provendo aplicações de entrega e de monitoramento ambiental.

EM PARTICULAR PARA A ÁREA DA SAÚDE, A INTERNET TÁTIL PODE CAUSAR UMA REVOLUÇÃO, PERMITINDO QUE SOFISTICADOS SERVIÇOS POSSAM SER OFERTADOS PARA QUALQUER UM, A QUALQUER INSTANTE E EM QUALQUER LUGAR

Em particular para a área da saúde, a Internet Tátil pode causar uma revolução, permitindo que sofisticados serviços possam ser ofertados para qualquer um, a qualquer instante e em qualquer lugar, como, por exemplo: diagnósticos remotos por especialistas médicos, viabilizados por sofisticados sistemas de telediagnóstico que incluem capacidade de exame físico (mesmo por toque) por meio de robôs controlados remotamente que oferecem realimentação audiovisual e tátil; intervenções cirúrgicas remotas; e serviços de reabilitação por meio de exoesqueleto remotamente controlado por terapeutas com realimentação tátil [8].

1.1.4. ENHANCED REMOTE AREA COMMUNICATIONS

O POTENCIAL REVOLUCIONÁRIO DAS REDES 5G NÃO PODE FICAR RESTRITO AOS GRANDES CENTROS URBANOS, SOB PENA DE AUMENTAR AS DIVISÕES DIGITAL E SOCIAL JÁ EXISTENTES NO PAÍS E DE DIMINUIR A COMPETITIVIDADE DE SETORES ESTRATÉGICOS PARA A ECONOMIA DO PAÍS, COMO O AGRONEGÓCIO.

Este é um cenário de grande interesse para o Brasil, quer seja por razões econômicas, quer seja por razões sociais. O potencial revolucionário das redes 5G não pode ficar restrito aos grandes centros urbanos, sob pena de aumentar as divisões digital e social já existentes no país e de diminuir a competitividade de setores estratégicos para a economia do país, como o agronegócio. Embora de particular interesse para o Brasil, este cenário também é de interesse de um grande número de nações ao redor do Globo. Por exemplo, alguns estudos estimam que haja aproximadamente 4 bilhões de pessoas desconectadas no mundo, vivendo em áreas sem cobertura ou em áreas com cobertura deficitária [9].

Em particular para o Brasil, o cenário eRAC assume importância especial e, por isso, tem sido escolhido como o cenário prioritário em diversas iniciativas, como no Projeto 5G Brasil e no Plano Nacional de IoT, que definiu como uma de suas verticais prioritárias as aplicações em área rural, a qual demanda a existência do cenário eRAC para sua efetiva implementação. Embora as redes de comunicações móveis atendam um percentual significativo da população no Brasil, ainda há grandes problemas de conectividade com áreas remotas e áreas rurais, com distritos distantes da sede dos municípios desatendidos e fazendas localizadas em áreas remotas desprovidas de soluções de comunicação.

EMBORA AS REDES DE COMUNICAÇÕES MÓVEIS ATENDAM UM PERCENTUAL SIGNIFICATIVO DA POPULAÇÃO NO BRASIL, AINDA HÁ GRANDES PROBLEMAS DE CONECTIVIDADE COM ÁREAS REMOTAS E ÁREAS RURAIS, COM DISTRITOS DISTANTES DA SEDE DOS MUNICÍPIOS DESATENDIDOS E FAZENDAS LOCALIZADAS EM ÁREAS REMOTAS DESPROVIDAS DE SOLUÇÕES DE COMUNICAÇÃO.

Como definido pelo projeto *Remote area Access Network for the 5th Generation* (5G-RANGE), que teve por objetivo desenvolver um transceptor 5G otimizado para o cenário eRAC, o principal requisito de desempenho para este cenário é oferecer uma taxa de 100 Mbps disponível a uma distância de 50 km da estação rádio base (ERB) [10].

PARA O BRASIL, EM TERMOS ECONÔMICOS, AS APLICAÇÕES NA VERTICAL DO AGRONEGÓCIO TALVEZ SEJAM AS MAIS IMPORTANTES DO CENÁRIO ERAC. ESTA IMPORTÂNCIA É CORROBORADA PELA RELEVÂNCIA DO AGRONEGÓCIO PARA O PAÍS, QUE RESPONDEU POR 44% DAS EXPORTAÇÕES BRASILEIRAS [11] E 21,6% DO PIB DO PAÍS EM 2017

O cenário eRAC tem grande impacto econômico e social em diversas verticais, como agronegócio, saúde, educação, logística, indústria, cultura e mineração.

Para o Brasil, em termos econômicos, as aplicações na vertical do agronegócio talvez sejam as mais importantes do cenário eRAC. Esta importância é corroborada pela relevância do agronegócio para o país, que respondeu por 44% das exportações brasileiras [11] e 21,6% do PIB do país em 2017 [12].

As aplicações relacionadas com IoT no agronegócio incluem: monitoramento do ar e da água, de dados de umidade, de temperatura, de condutividade e potencial hidrogeniônico (pH) do solo, de incêndio, de dados meteorológicos, de saúde e posicionamento de rebanhos, de peso e alimentação de animais; rastreabilidade de vacinas, medicamentos e insumos; irrigação inteligente; controle de pulverizações; emprego de maquinário autônomo em lavouras; ***drones*** e redes de ***drones*** para monitoramento, mapeamento de terreno e aplicação de pesticidas em lavouras; gestão da produção e do desempenho de máquinas; aplicações na logística da produção rural. Estudo publicado pelo Banco Nacional de Desenvolvimento Econômico e Social (BNDES) aponta um impacto de até US$ 21,1 bilhões do uso de IoT no ambiente rural em 2025 [13].

1.2. ALGUNS ASPECTOS E TENDÊNCIAS TECNOLÓGICAS ASSOCIADAS AO 5G

Os diversos cenários de uso com diferentes requisitos de desempenho resultam em uma certa complexidade para as redes 5G e definem determinadas soluções e tendências tecnológicas para a implementação destas redes. Algumas soluções são de particular interesse para este trabalho e estão resumidas nas subseções 1.2.1 a 1.2.4, a seguir.

1.2.1.

USO DE *MULTIPLE INPUT – MULTIPLE OUTPUT* (MIMO) MASSIVO E MIMO 3D

Em um sistema MIMO utilizam-se múltiplas antenas de transmissão e múltiplas antenas de recepção. Já no MIMO massivo, um número muito grande de antenas é utilizado na estação rádio base (por exemplo, centenas de antenas) para servir simultaneamente vários terminais utilizando o mesmo recurso (tempo e frequência) [14]. Além disso, usando arranjos ativos de antenas, é possível controlar o feixe irradiado pelas antenas tanto na vertical, quanto na horizontal (o que denominamos MIMO 3D), o que permite aumentar a setorização das células, levando a uma maior capacidade de escoamento de tráfego na rede. Por fim, controlando o feixe irradiado pela antena pode-se aumentar a relação sinal/ruído no enlace, o que resulta no aumento da capacidade de transmissão do enlace ou na redução da potência de transmissão, poupando energia [15].

O uso de sistemas MIMO massivo está diretamente relacionado com dois outros aspectos tecnológicos relevantes para as redes 5G: o uso de comunicações em ondas milimétricas e o uso de redes heterogêneas e redes ultradensas.

A operação em ondas milimétricas facilita a implementação de sistemas MIMO massivos, uma vez que em frequências mais altas o arranjo de antenas pode ser miniaturizado.

O uso de ondas milimétricas, por sua vez, leva à necessidade de operação com células muito pequenas, com diâmetro de cobertura da ordem de poucas centenas de metros, o que leva ao aumento significativo do número de células na rede, aumentando a capacidade da rede de escoar tráfego, ainda que aumentando sua complexidade.

Embora o uso de MIMO massivo traga várias vantagens e seja mandatório para se alcançar os requisitos de desempenho do cenário eMBB, é preciso ressaltar que estes sistemas são de grande complexidade tecnológica, o que pode impactar nas decisões sobre a arquitetura a se utilizar nas redes de acesso de rádio (do inglês, ***Radio Access Networks*** - RANs).

O USO DE SISTEMAS MIMO MASSIVO ESTÁ DIRETAMENTE RELACIONADO COM DOIS OUTROS ASPECTOS TECNOLÓGICOS RELEVANTES PARA AS REDES 5G: O USO DE COMUNICAÇÕES EM ONDAS MILIMÉTRICAS E O USO DE REDES HETEROGÊNEAS E REDES ULTRADENSAS.

1.2.2. OPERAÇÃO EM ONDAS MILIMÉTRICAS

Para atender aos requisitos de desempenho do cenário eMBB é necessário o uso de frequências em ondas milimétricas (maiores ou iguais a 30 GHz, embora frequências em torno de 24, 26 e 28 GHz também estejam sendo consideradas e denominadas como milimétricas), nas quais é possível operar com maiores larguras de faixa, resultando em maiores taxas de transmissão disponíveis.

PARA ATENDER AOS REQUISITOS DE DESEMPENHO DO CENÁRIO EMBB É NECESSÁRIO O USO DE FREQUÊNCIAS EM ONDAS MILIMÉTRICAS (MAIORES OU IGUAIS A 30 GHZ, EMBORA FREQUÊNCIAS EM TORNO DE 24, 26 E 28 GHZ TAMBÉM ESTEJAM SENDO CONSIDERADAS E DENOMINADAS COMO MILIMÉTRICAS), NAS QUAIS É POSSÍVEL OPERAR COM MAIORES LARGURAS DE FAIXA, RESULTANDO EM MAIORES TAXAS DE TRANSMISSÃO DISPONÍVEIS.

Como salientado na Seção 1.2.1, o uso de ondas milimétricas também é importante para viabilizar o uso de sistemas MIMO massivos, com os benefícios já descritos.

A desvantagem principal da operação em ondas milimétricas é uma maior atenuação sofrida pelo sinal. Em frequências mais altas a atenuação por propagação em espaço livre aumenta; além disto, o sinal sofre mais atenuação devido a gases atmosféricos, maior atenuação por chuva e maior atenuação por absorção em materiais utilizados nas construções [16].

A maior atenuação experimentada pelo sinal pode ser compensada, como já mencionado, utilizando células muito pequenas, com cerca de 200 metros de diâmetro ou menos, levando ao aumento significativo no número de células necessárias na rede no cenário eMBB.

1.2.3. REDES ULTRADENSAS E REDES HETEROGÊNEAS

Como descrito na Seção 1.2.2, na operação em ondas milimétricas o diâmetro da célula é muito pequeno e o número de células na rede será muito maior que nas gerações anteriores. Por exemplo, nas redes 3G a densidade de estações rádio base é de aproximadamente 4 a 5 estações por km^2 e nas redes 4G de 8 a 10 estações por km^2; por sua vez, a densidade esperada para as redes 5G é de 40 a 50 estações rádio base por km^2 [17].

NAS REDES 3G A DENSIDADE DE ESTAÇÕES RÁDIO BASE É DE APROXIMADAMENTE 4 A 5 ESTAÇÕES POR KM² E NAS REDES 4G DE 8 A 10 ESTAÇÕES POR KM²; POR SUA VEZ, A DENSIDADE ESPERADA PARA AS REDES 5G É DE 40 A 50 ESTAÇÕES RÁDIO BASE POR KM²

A ideia de densificação das redes pode ser aplicada tanto na direção horizontal, com mais estações colocadas ao longo das ruas, por exemplo, quanto na direção vertical, com estações rádio base posicionadas nos andares dos prédios. As células nestas redes densas podem ser classificadas em pico-células ou femto-células. Por exemplo, uma pico-célula pode ser utilizada para cobrir algumas dezenas até poucas centenas de metros, servindo dezenas de usuários ativos, enquanto uma femto-célula pode ser instalada para cobrir áreas internas (como residências, escritórios, salas de reuniões), atendendo um pequeno número de usuários [18].

O uso de redes heterogêneas, com macro-células, pico-células e femto-células é apontada como uma tendência na implementação das redes 5G, com as pico e femto-células utilizadas para escoar o tráfego dos usuários no cenário eMBB, e as macro-células utilizadas para implementar o plano de controle das redes e também para escoar o tráfego dos usuários em cenários em que se utilizem frequência de operação mais baixas (como frequências abaixo de 1 GHz e entre 1 GHz e 6 GHz) [19,20,21].

1.2.4. SOFTWARIZATION E CLOUD RADIO ACCESS NETWORK

AS REDES DEFINIDAS POR SOFTWARE (DO INGLÊS, SOFTWARE DEFINED NETWORKS - SDNS), A VIRTUALIZAÇÃO DAS FUNÇÕES DE REDES (DO INGLÊS, NETWORK FUNCTION VIRTUALIZATION - NFV) E O USO DE CLOUD, EDGE E FOG COMPUTING SÃO EXEMPLOS DO QUE CHAMAMOS DE "SOFTWARIZAÇÃO"

O termo "***softwarization***" das telecomunicações foi cunhado pela iniciativa ***Institute of Electrical and Electronics Engineers – Software Defined Networks*** (IEEE SDN) para agrupar diferentes dimensões de uma tendência global das redes de telecomunicações, que é a implementação de diversas funções e aspectos das redes por meio de ***software***. As Redes Definidas por Software (do inglês, Software Defined Networks - SDNs), a Virtualização das Funções de Redes (do inglês, Network Function Virtualization - NFV) e o uso de cloud, edge e fog computing são exemplos do que chamamos de "softwarização" [22].

Um dos mais importantes ativos das operadoras de redes de comunicações móveis é a rede de acesso de rádio. Particularmente no cenário eMBB das redes 5G, com operações em ondas milimétricas, no qual o número de estações rádio base é muito maior do que nas redes de gerações anteriores, a RAN torna-se ainda mais relevante. Assim, neste cenário, a "***softwarização***" da RAN pode trazer grandes benefícios para as operadoras, levando ao conceito de Rede de Acesso de Rádio na Nuvem (do inglês, ***Cloud Radio Access Network*** - C-RAN) e ao conceito de Rede de Acesso de Rádio Virtual (do inglês, ***Virtual Radio Access Network*** - V-RAN).

UM DOS MAIS IMPORTANTES ATIVOS DAS OPERADORAS DE REDES DE COMUNICAÇÕES MÓVEIS É A REDE DE ACESSO DE RÁDIO.

A "SOFTWARIZAÇÃO" DA RAN PODE TRAZER GRANDES BENEFÍCIOS PARA AS OPERADORAS, LEVANDO AO CONCEITO DE REDE DE ACESSO DE RÁDIO NA NUVEM (DO INGLÊS, CLOUD RADIO ACCESS NETWORK - C-RAN) E AO CONCEITO DE REDE DE ACESSO DE RÁDIO VIRTUAL (DO INGLÊS, VIRTUAL RADIO ACCESS NETWORK - V-RAN).

Os conceitos de C-RAN e V-RAN e o uso de SDN e NVF nas redes 5G serão detalhados nas Seções 2 e 3 deste documento.

1.2.5. *NETWORK SLICING* E INTELIGÊNCIA ARTIFICIAL

PARA VENCER O DESAFIO DE IMPLEMENTAR APLICAÇÕES EM DIFERENTES CENÁRIOS, COM REQUISITOS DE DESEMPENHO DIVERSOS, EM UMA MESMA INFRAESTRUTURA DE REDE, SURGE O CONCEITO DE FATIAMENTO DE REDE (DO INGLÊS, NETWORK SLICING).

Como já mencionado, as redes 5G permitirão a oferta de diversas novas aplicações em diferentes cenários de uso, com requisitos de desempenho às vezes conflitantes, que demandam o uso de diversas tecnologias e a operação em diferentes faixas de frequências, desde frequências abaixo de 1 GHz até operação em ondas milimétricas.

NO FATIAMENTO DE REDE VÁRIAS REDES LÓGICAS SÃO ESTABELECIDAS SOBRE UMA MESMA INFRAESTRUTURA DE REDE COMPARTILHADA, COM CADA REDE LÓGICA DEFINIDA PARA ATENDER UMA APLICAÇÃO ESPECÍFICA E SEUS REQUISITOS DE DESEMPENHO FIM-A-FIM.

Para vencer o desafio de implementar aplicações em diferentes cenários, com requisitos de desempenho diversos, em uma mesma infraestrutura de rede, surge o conceito de fatiamento de rede (do inglês, network slicing). No fatiamento de rede várias redes lógicas são estabelecidas sobre uma mesma infraestrutura de rede compartilhada, com cada rede lógica definida para atender uma aplicação específica e seus requisitos de desempenho fim-a-fim.

O uso de fatiamento de rede, ao mesmo tempo que traz a flexibilidade necessária para a implementação das diversas novas aplicações previstas para as redes 5G, resulta em uma rede em que o gerenciamento e a otimização se tornam muito complexos, com um grande número de parâmetros a serem definidos para otimizar os diversos indicadores chaves de desempenho (do inglês, ***Key Performance Indicators*** - KPIs) dos vários cenários de uso da rede. As soluções tradicionais de otimização da rede podem não ser satisfatórias neste novo contexto e o uso de inteligência artificial (IA) e aprendizado de máquina (do inglês, ***Machine Learning*** - ML) tornam-se fundamentais para a plena implementação das redes 5G. O uso de IA nas redes 5G será detalhado na Seção 4 deste documento.

AS SOLUÇÕES TRADICIONAIS DE OTIMIZAÇÃO DA REDE PODEM NÃO SER SATISFATÓRIAS NESTE NOVO CONTEXTO E O USO DE INTELIGÊNCIA ARTIFICIAL (IA) E APRENDIZADO DE MÁQUINA (DO INGLÊS, MACHINE LEARNING - ML) TORNAM-SE FUNDAMENTAIS PARA A PLENA IMPLEMENTAÇÃO DAS REDES 5G.

1.2.6. ASPECTOS DE SEGURANÇA

Nas gerações anteriores das redes de comunicações móveis (2G até 4G) os principais aspectos de segurança considerados estavam voltados para garantir a operação adequada do sistema de bilhetagem, a autenticação dos usuários e a integridade dos dados na interface aérea, por meio do uso de criptografia [23]. As redes 5G trazem novos e grandes desafios nos aspectos de segurança. Um leque muito maior de aplicações, algumas consideradas de missão crítica, como carros autônomos e cirurgias remotas, associado a um número muito maior de usuários, principalmente devido ao cenário mMTC, a heterogeneidade dos dispositivos conectados à rede, com diferentes capacidades de armazenamento e processamento de dados, e a tendência de ***softwarização*** das redes, tornam os aspectos de segurança e privacidade muito mais críticos nas redes 5G do que nas gerações anteriores. A Seção 5 deste documento detalha os desafios e possíveis soluções relacionados aos aspectos de segurança nas redes 5G.

AS REDES 5G TRAZEM NOVOS E GRANDES DESAFIOS NOS ASPECTOS DE SEGURANÇA. UM LEQUE MUITO MAIOR DE APLICAÇÕES, ALGUMAS CONSIDERADAS DE MISSÃO CRÍTICA, COMO CARROS AUTÔNOMOS E CIRURGIAS REMOTAS, ASSOCIADO A UM NÚMERO MUITO MAIOR DE USUÁRIOS, PRINCIPALMENTE DEVIDO AO CENÁRIO MMTC, A HETEROGENEIDADE DOS DISPOSITIVOS CONECTADOS À REDE, COM DIFERENTES CAPACIDADES DE ARMAZENAMENTO E PROCESSAMENTO DE DADOS, E A TENDÊNCIA DE SOFTWARIZAÇÃO DAS REDES, TORNAM OS ASPECTOS DE SEGURANÇA E PRIVACIDADE MUITO MAIS CRÍTICOS NAS REDES 5G DO QUE NAS GERAÇÕES ANTERIORES.

1.3. OPENRAN

OPENRAN É UM MOVIMENTO DO ECOSSISTEMA DE REDES DE COMUNICAÇÕES MÓVEIS COM O OBJETIVO DE DEFINIR UMA ESTRUTURA FLEXÍVEL E INTEROPERÁVEL PARA AS REDES DE ACESSO, COM DESAGREGAÇÃO ENTRE O HARDWARE E O SOFTWARE DA REDE E O USO DE INTERFACES ABERTAS ENTRE OS DIVERSOS COMPONENTES DA RAN.

OpenRAN é um movimento do ecossistema de redes de comunicações móveis com o objetivo de definir uma estrutura flexível e interoperável para as redes de acesso, com desagregação entre o ***hardware*** e o ***software*** da rede e o uso de interfaces abertas entre os diversos componentes da RAN. O movimento OpenRAN[1] se utiliza da ideia de "***softwarização***" das redes de telecomunicações, utilizando os conceitos de SDN, NFV, C-RAN e V-RAN.

Duas iniciativas importantes para a definição de soluções OpenRAN para as redes de comunicações móveis são a iniciativa OpenRAN do ***Telecom Infra Project*** (TIP) e a iniciativa da O-RAN ***Alliance***, que é uma associação de empresas do ecossistema de comunicações móveis (tais como fabricantes, operadoras, integradores, etc.) que objetiva a definição de especificações que permitam a implantação de soluções OpenRAN interoperáveis. É importante comentar que recentemente houve um acordo entre O-RAN Alliance e TIP no intuito de compartilharem informações e especificações e, também, realizarem conjuntamente testes de integração de soluções OpenRAN.

A Seção 6 deste documento traz o detalhamento da arquitetura para OpenRAN proposta no O-RAN Alliance, bem como uma análise de outros aspectos relevantes e das vantagens e desvantagens do uso de soluções OpenRAN nas redes 5G.

1 - Além de OpenRAN como o movimento descrito, temos também as soluções de redes de acesso que seguem os preceitos definidos pelo OpenRAN. Assim, neste documento utilizaremos "o OpenRAN" quando estivermos nos referindo ao movimento e "a OpenRAN" quando estivermos nos referindo à implementação da rede de acesso em si.

2.
C-RAN E V-RAN

2.1. INTRODUÇÃO

A CRESCENTE DEMANDA POR SERVIÇOS DE ALTA QUALIDADE E VELOCIDADE NAS REDES MÓVEIS, ALÉM DO CONSIDERÁVEL AUMENTO DO NÚMERO DE USUÁRIOS, TEM IMPULSIONADO SIGNIFICATIVAMENTE A EVOLUÇÃO DA INFRAESTRUTURA DAS REDES ATUAIS.

A crescente demanda por serviços de alta qualidade e velocidade nas redes móveis, além do considerável aumento do número de usuários, tem impulsionado significativamente a evolução da infraestrutura das redes atuais [24].

A C-RAN é um tipo de arquitetura de rede capaz de atender este amplo crescimento de forma mais flexível e com um menor número de ERBs. Isso é factível por meio da virtualização e centralização das ERBs, possibilitando a otimização do uso dos recursos disponíveis e até mesmo uma solução cooperativa entre várias operadoras. Essa tecnologia destaca-se por apresentar eficiência energética elevada, melhor utilização do espectro de frequências, redução de interferências co-canal e entre canais, serviços com alta taxa de transferência e possibilidade de novos modelos de negócios. Também apresenta redução das despesas de capital (do inglês, ***capital expenditure*** - CAPEX) e despesas operacionais (do inglês, ***operational expenditure*** - OPEX) [25,26,27].

A C-RAN É UM TIPO DE ARQUITETURA DE REDE CAPAZ DE ATENDER ESTE AMPLO CRESCIMENTO DE FORMA MAIS FLEXÍVEL E COM UM MENOR NÚMERO DE ERBS. ISSO É FACTÍVEL POR MEIO DA VIRTUALIZAÇÃO E CENTRALIZAÇÃO DAS ERBS, POSSIBILITANDO A OTIMIZAÇÃO DO USO DOS RECURSOS DISPONÍVEIS E ATÉ MESMO UMA SOLUÇÃO COOPERATIVA ENTRE VÁRIAS OPERADORAS.

NA ARQUITETURA MÓVEL C-RAN, O PROCESSAMENTO DE SINAIS EM BANDA BASE E DE CANAL DE RADIOFREQUÊNCIA (RF) É VIRTUALIZADO E COMPARTILHADO COM ERBS CENTRALIZADAS.

Na arquitetura móvel C-RAN, o processamento de sinais em banda base e de canal de radiofrequência (RF) é virtualizado e compartilhado com ERBs centralizadas. A centralização e o compartilhamento permitem um tráfego de dados mais dinâmico e uma melhor alocação de recursos tanto computacionais quanto de rádio. Portanto, essa arquitetura apresenta o potencial de reduzir as despesas da rede, uma vez que as ERBs são virtualizadas e não se encontram implantadas fisicamente em áreas diferentes, muitas vezes de difícil acesso. Além disso, ela reduz o consumo de energia quando comparada às redes tradicionais, pois as ERBs estarão localizadas no mesmo dispositivo físico centralizado (ou seja, um servidor) [28].

2.2.
ARQUITETURA C-RAN

Os principais componentes da arquitetura C-RAN são as unidades de banda base (do inglês, ***Baseband Units*** - BBUs) localizadas na nuvem, as cabeças de rádio remotas (do inglês, ***Remote Radio Heads*** - RRHs) e a rede de transporte (do inglês, ***fronthaul***) que interconecta as BBUs e as RRHs, conforme mostrado na **Figura 1**.

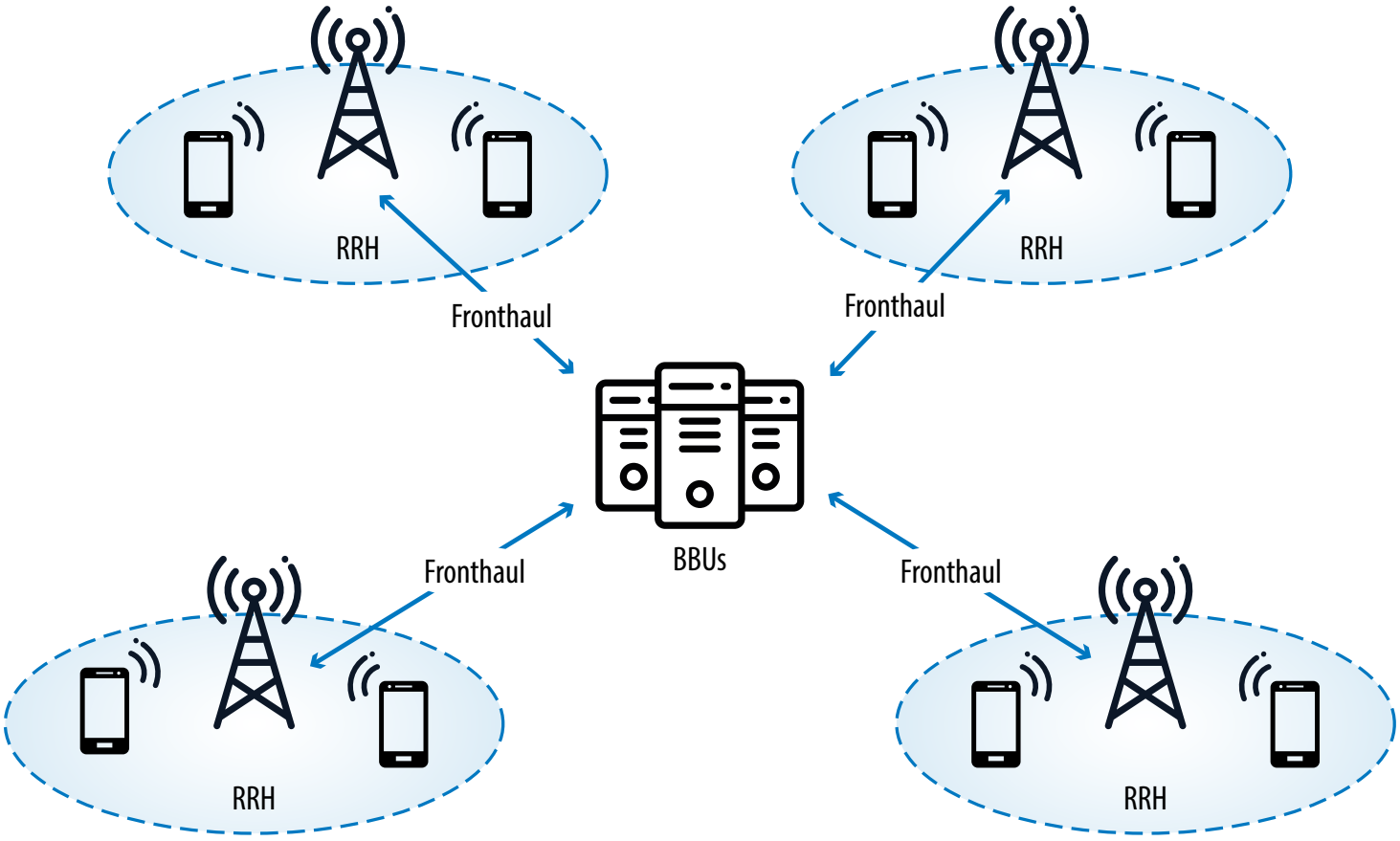

Figura 1: Componentes da arquitetura C-RAN.

Na arquitetura C-RAN, diferentes BBUs são agrupadas e centralizadas na nuvem, por meio do compartilhamento de recursos de rádio. Essa estrutura é denominada de conjunto (do inglês, ***pool***) de BBUs, apresentando elevada capacidade computacional e de armazenamento. Essas BBUs são responsáveis por processar sinais usando recursos compartilhados, alocando-os dinamicamente às RRHs, com base nas necessidades atuais da rede. Ou seja, o número de BBUs pode ser dinamicamente alterado ao longo do tempo.

As RRHs encontram-se distribuídas na rede e correspondem às estruturas responsáveis por transmitirem os sinais de rádio das BBUs para os usuários da rede (do inglês, ***downlink***) e encaminharem os sinais de banda base dos usuários para o processamento na nuvem (do inglês, ***uplink***). As principais funções das RRHs são a amplificação de radiofrequência, elevação e rebaixamento de frequência, filtragem, processamento digital, conversão analógico-digital e digital-analógico.

A rede de transporte realiza a conexão entre o conjunto de BBUs e as RRHs por meio de ligações que suportam elevada largura de banda necessária para lidar com os requisitos do sistema. Para alcançar esse propósito, padrões industriais de interface de conexão, como a Interface de Rádio Pública Comum (do inglês, ***Common Public Radio Interface*** - CPRI) e a Interface de ***Fronthaul*** de Próxima Geração (do inglês, ***Next Generation Fronthaul Interface*** - NGFI) evoluíram para habilitar essas novas interfaces entre BBUs e RRHs. A rede de transporte pode ser implementada utilizando diferentes tecnologias de comunicação, como fibra óptica, redes sem fio ponto-a-ponto nas faixas de micro-ondas ou ondas milimétricas. Para a arquitetura C-RAN, comunicação por fibra óptica é considerada a mais adequada por suportar alta capacidade de transmissão. Entretanto, essa solução apresenta elevado custo e implantação não flexível. A comunicação sem fio por micro-ondas ou ondas milimétricas é mais barata e flexível, mas, por outro lado, tem capacidade reduzida, maior latência e menor confiabilidade quando comparada à fibra óptica [29].

A centralização das BBUs na arquitetura C-RAN apresenta algumas vantagens em relação às redes celulares tradicionais, cujas estruturas encontram-se distribuídas [26]. As BBUs estão localizadas em centros de dados que possibilitam trocas de informações e execução de cálculos complexos, não realizáveis nas redes tradicionais. Como resultado, a implementação de tecnologias avançadas, com alto poder de processamento, se torna possível com a arquitetura C-RAN.

Além disso, com várias BBUs alocadas em nuvem, a virtualização de recursos e o compartilhamento cooperativo de recursos de rádio se tornam factíveis. Consequentemente, a alocação desses recursos pode ser mais flexível e também sob demanda, ao contrário das redes atuais. Isso resulta em melhorias na utilização de serviços, menor consumo de energia e maior satisfação do usuário.

COM VÁRIAS BBUS ALOCADAS EM NUVEM, A VIRTUALIZAÇÃO DE RECURSOS E O COMPARTILHAMENTO COOPERATIVO DE RECURSOS DE RÁDIO SE TORNAM FACTÍVEIS. CONSEQUENTEMENTE, A ALOCAÇÃO DESSES RECURSOS PODE SER MAIS FLEXÍVEL E TAMBÉM SOB DEMANDA, AO CONTRÁRIO DAS REDES ATUAIS.

Também é importante mencionar que na arquitetura C-RAN, cujos servidores possuem elevado poder computacional, os serviços podem ser implantados na borda da rede e não somente na parte central. Dessa forma, os serviços ficarão mais próximos do usuário, alcançando respostas mais rápidas (ou seja, latências menores), aumentando assim a eficiência do sistema.

2.2.1.
ESTRUTURAS DO SISTEMA C-RAN

A estrutura de um sistema C-RAN pode ser classificada como totalmente centralizada, parcialmente centralizada ou híbrida. Essa classificação depende do local onde as funções das camadas física, de Controle de Acesso ao Meio (do inglês, ***Media Access Control*** - MAC) e de rede serão tratadas. As funções da camada física incluem o processamento em banda base, enquanto as funções das camadas de MAC e de rede correspondem aos mecanismos de controle de acesso e de roteamento, respectivamente [30].

Em uma estrutura totalmente centralizada, praticamente todas as funcionalidades das camadas física, MAC e de rede ocorrem nas BBUs. As BBUs são responsáveis por todas as funções de gerenciamento e recursos de processamento. Consequentemente, essa estrutura pode se beneficiar da facilidade na operação e manutenção dessas funcionalidades. Entretanto, a sobrecarga de comunicação entre as BBUs e as RRHs impacta no sistema, uma vez que o desempenho será limitado pela capacidade do ***fronthaul*** [28].

Por outro lado, em uma estrutura parcialmente centralizada, as funções da camada física são realizadas nas RRHs, enquanto as funções das camadas de rede e MAC são realizadas nas BBUs [31]. Isso contribui para reduzir significativamente a sobrecarga da comunicação entre RRHs e BBUs, pois a camada física assume uma grande carga computacional da C-RAN. No entanto, essas estruturas apresentam maior complexidade na comunicação entre as camadas física e MAC, não oferecendo suporte ao compartilhamento de recursos da camada física entre diferentes RRHs. Dessa forma, técnicas de mitigação de interferência e aumento da eficiência espectral, como Multiponto Coordenado (do inglês, ***Coordinated Multi-Point*** - CoMP) e Coordenação de Interferência Inter-celular (do inglês, ***Inter-cell Interference Coordination*** - ICIC) não podem ser implementadas de forma eficiente [32].

Em uma estrutura híbrida, parte das funções da camada física é realizada nas BBUs, enquanto outras são realizadas nas RRHs. Essa estrutura pode ser a mais flexível no compartilhamento de recursos, além de apresentar potencial para a redução do consumo de energia e da sobrecarga na comunicação com as BBUs [33].

2.3. VIRTUALIZAÇÃO DA ARQUITETURA C-RAN

A VIRTUALIZAÇÃO DE REDE CONSISTE EM CRIAR COMPONENTES VIRTUAIS PARA A INFRAESTRUTURA DE REDE IMPLANTADOS NA MESMA ESTRUTURA FÍSICA. DESSA FORMA, ESSA TECNOLOGIA FACILITA O ISOLAMENTO LÓGICO DE RECURSOS, ENQUANTO OS RECURSOS FÍSICOS SÃO COMPARTILHADOS DE FORMA DINÂMICA E ESCALÁVEL

A virtualização de rede consiste em criar componentes virtuais para a infraestrutura de rede implantados na mesma estrutura física. Dessa forma, essa tecnologia facilita o isolamento lógico de recursos, enquanto os recursos físicos são compartilhados de forma dinâmica e escalável [34].

Na arquitetura C-RAN, a virtualização da rede é feita a nível de conjunto de BBUs. Cada BBU é um nó virtual e a comunicação entre eles é realizada por enlaces virtuais. O conjunto de BBUs é implementado em máquina virtuais hospedadas em uma máquina física com o compartilhamento de recursos da Unidade Central de Processamento (do inglês, ***Central Process Unit*** - CPU), memória e rede.

Essa tecnologia apresenta muitas vantagens, como mecanismos de controle flexível, recursos eficientes, redução dos custos de implementação, minimização do tempo necessário para comunicação entre as BBUs, escalabilidade e facilidade para adicionar ou remover BBUs [35].

2.4. DESAFIOS DA ARQUITETURA C-RAN

Na prática, a implantação da arquitetura C-RAN apresenta alguns pontos desafiadores. Asrincipais dificuldades envolvem: i) a necessidade de ***fronthaul*** com alta capacidade, ii) cooperação entre BBUs, iii) agrupamento de células, iv) virtualização do sistema e v) segurança cibernética. Esses desafios são comentados a seguir.

O enlace de ***fronthaul*** entre BBUs e RRHs deve apresentar alta largura de banda com baixos valores de atraso e custo. Conforme discutido na Seção 2.2, a abordagem totalmente centralizada é a estrutura mais adotada na arquitetura C-RAN. Esse tipo de estrutura impõe uma elevada sobrecarga de comunicação no ***fronthaul***. Portanto, a necessidade de uma elevada largura de banda no ***fronthaul*** deve ser prevista. Consequentemente, pode haver dificuldade de utilizar comunicação sem fio no ***fronthaul***. A comunicação por meio de fibra óptica pode fornecer a largura de banda necessária para contornar esse problema. No entanto, a rede óptica apresenta um elevado custo, dificultando a sua implantação pela maioria dos provedores de serviços móveis celulares. Dessa forma, uma relação de compromisso entre atraso, largura de banda e custo deve ser estabelecida.

A cooperação entre as BBUs localizadas no mesmo conjunto possui importância fundamental para viabilizar o compartilhamento de dados de usuários, agendamento e coleta de ***feedback*** do canal. Essa cooperação deve ser estabelecida, apresentando desafios em relação à privacidade do usuário, necessidade de largura de banda elevada e baixa latência na comunicação entre BBUs.

O agrupamento ideal de células da rede e a atribuição do conjunto de BBUs com sobrecarga mínima é um desafio em sistemas C-RAN. Um conjunto de BBUs deve alcançar o número máximo de canais de envio e recebimento, minimizando o atraso no ***fronthaul*** e a sobrecarga. Além disso, uma BBU deve oferecer suporte a várias localizações geográficas distribuídas.

Conforme discutido na Seção 2.3, a virtualização do sistema possibilita o processamento distribuído e o compartilhamento de recursos entre várias BBUs. Entretanto, esse processamento deve ser em tempo real e dinâmico para suportar a mudança de cargas das células. Adicionalmente, a infraestrutura em nuvem utilizada para a implementação das BBUs é diferente da infraestrutura em nuvem comumente utilizada em Tecnologia da Informação (TI), sendo assim necessário modificá-la para atender aos novos requisitos [36].

A segurança em termos de privacidade do usuário e confiabilidade das partes constituintes do sistema deve ser assegurada em uma arquitetura C-RAN. Seus recursos são compartilhados entre BBUs, especialmente em uma arquitetura distribuída, e um grande número de usuários podem fazer uso desse sistema. Portanto, falhas de segurança, como a quebra de privacidade, podem representar pontos de vulnerabilidade para a arquitetura C-RAN [37].

2.5. V-RAN

A evolução das redes C-RAN culminou em uma nova arquitetura de rede conhecida como V-RAN [38], que nada mais é que a C-RAN virtualizada.

A virtualização de rede proposta pela arquitetura V-RAN apresenta vantagens e soluções para as dificuldades apresentadas nas redes C-RAN. Implantações práticas com controle flexível, escalabilidade, baixo custo e uso eficiente de recursos são atributos dessa rede virtual [27].

O desacoplamento das funcionalidades de ***software*** e ***hardware*** da V-RAN e sua natureza centralizada facilitam uma maior diversificação de aplicações, destacando-se em serviços propostos pelas redes 5G e de sexta geração (6G), como acesso massivo de máquinas, Internet tátil e outros [39]. Na rede V-RAN, o ***hardware*** do rádio proprietário é igual ao utilizado nas redes C-RAN. Entretanto, a BBU monolítica é substituída por dois itens desagregados: i) um servidor com ***hardware*** comercial de prateleira (do inglês, ***commercial off-the-shelf*** - COTS) e ii) o ***software*** que implementa as funções da BBU. Já as interfaces proprietárias entre os rádios e BBUs, baseadas em COTS, mantêm-se como na arquitetura RAN convencional [40]. Portanto, a proposta dessa rede virtual sem fio promove o compartilhamento de recursos de rádio e BBUs entre as RRHs. Isso diminui os custos de investimentos e operacionais, promovendo uma maior eficiência energética da rede [41]. Dessa forma, incentivam-se inovações e novos fornecedores são impulsionados a entrarem no mercado [42].

Diferentemente da abordagem de rede em nuvem adotada pela C-RAN, a rede V-RAN apresenta requisitos diferentes, principalmente aqueles relacionados à virtualização, à orquestração e ao dimensionamento de recursos, que podem ser baseados em hipervisores e em contêineres [43,44]. Na virtualização baseada em hipervisor, a máquina virtual executa um sistema operacional completo, levando ao uso intenso de recursos computacionais. Isso resulta em implantações demoradas e inicialização lenta do sistema operacional [45]. Porém, trata-se de uma alternativa mais segura devido ao alto nível de isolamento entre as funções de redes implementadas entre duas máquinas virtuais distintas [45]. Por outro lado, a virtualização baseada em contêiner potencializa um desenvolvimento rápido. Esse tipo de virtualização promove a execução e a implantação de aplicativos de ***software*** em ambientes denominados contêineres. Cada contêiner é uma instância em execução de uma imagem reduzida do sistema operacional. Dessa forma, tem-se um modelo leve e rápido que define e gerencia os processos do sistema. Porém, o menor nível de isolamento entre dois contêineres distintos pode promover um menor nível de segurança dentro do ambiente virtual [46].

Experimentos recentes mostraram que a solução V-RAN executada em contêineres é a mais adequada do ponto de vista de desempenho para a virtualização da RAN, por apresentar baixo tempo de implantação e menor probabilidade de falha do sistema ao realizar uma reconfiguração [47]. Além disso, as V-RANs baseadas em contêineres são mais leves, impondo menor sobrecarga de recursos, e não precisam depender de estruturas pesadas para orquestração e gerenciamento do sistema. Entretanto, pelo ponto de vista da segurança, deve-se ter um maior cuidado com relação à implantação dos diferentes contêineres, pois o menor nível de isolamento entre eles pode aumentar a superfície de ataques cibernéticos.

A ARQUITETURA V-RAN APRESENTA LIMITAÇÕES SIGNIFICATIVAS EM RELAÇÃO À ESCALABILIDADE E LATÊNCIA. ESSES PROBLEMAS PODEM SER RESOLVIDOS DESCENTRALIZANDO A V-RAN COM A UTILIZAÇÃO DA COMPUTAÇÃO DE BORDA MÓVEL (DO INGLÊS, MOBILE EDGE COMPUTING - MEC).

A arquitetura V-RAN apresenta limitações significativas em relação à escalabilidade e latência. Esses problemas podem ser resolvidos descentralizando a V-RAN com a utilização da Computação de Borda Móvel (do inglês, ***Mobile Edge Computing*** - MEC) [48]. Na MEC, os recursos de computação virtualizáveis são deslocados para uma maior proximidade em relação aos assinantes, promovendo um acesso de baixa latência e alta largura de banda para conteúdos, serviços de rede e aplicações dos usuários. A natureza distribuída da arquitetura MEC também a torna atrativa para suportar grandes volumes de dispositivos conectados, sendo assim considerada estratégica para o uso nas redes sem fio 5G e 6G [49].

A abordagem de virtualização de rede empregada na V-RAN faz com que a virtualização e a orquestração sejam mais complexas em relação à virtualização convencional [50]. Ou seja, novos métodos, algoritmos inteligentes e mecanismos de gerenciamento e orquestração devem ser empregados para alocar os recursos computacionais do sistema de forma justa e otimizada entre as diferentes RRHs. Particularidades como capacidade do canal, confiabilidade e fenômenos relacionados a um sistema sem fio também precisam ser considerados. Portanto, para a implantação comercial dessa rede torna-se necessário uma melhor investigação dos desafios técnicos e de gerenciamento que envolvem essa rede [51].

Na tentativa de solucionar as dificuldades da rede, a arquitetura V-RAN está evoluindo do conceito de C-RAN para OpenRAN, cuja proposta dessa nova arquitetura de rede baseia-se em dois pilares fundamentais, a abertura e a inteligência da rede [52]. A partir da Seção 6.1 serão detalhados todos os atributos relacionados à essa nova filosofia de rede.

3. NFV E SDN EM REDES 5G

O 5G FOI DESENVOLVIDO PARA FORNECER UMA REDE FLEXÍVEL, ESCALÁVEL, ÁGIL E PROGRAMÁVEL, SOBRE A QUAL DIFERENTES SERVIÇOS COM REQUISITOS VARIADOS PODEM SER ESTABELECIDOS.

As redes móveis de quinta geração trouxeram uma mudança de paradigma no conceito das arquiteturas de redes móveis até então utilizadas. Entre os principais pontos dessa mudança, destaca-se que o 5G foi desenvolvido para fornecer uma rede flexível, escalável, ágil e programável, sobre a qual diferentes serviços com requisitos variados podem ser estabelecidos. Esses requisitos se mostram ambiciosos e altamente desafiadores, tendo implicações tanto na RAN quanto no núcleo da rede e, portanto, exigiram uma mudança na arquitetura e nas tecnologias que compõem essa nova geração das redes de telecomunicações móveis. Vários conceitos e técnicas inovadoras foram introduzidos pelo 5G. No centro deste desenvolvimento está a NFV e a tecnologia SDN, que são reconhecidos como sendo dois dos principais habilitadores para as redes 5G. Essa seção fornece uma visão geral de ambas tecnologias com referência às redes 5G.

A **Figura 2** ilustra uma visão geral da rede 5G definida em três camadas. No nível mais baixo, são mostrados recursos físicos e ativos, como recursos de computação, rede, armazenamento, e recursos de acesso via rádio. Basicamente, os recursos do primeiro nível são agrupados em diferentes nuvens: i) a nuvem de borda, ii) a nuvem central ou, também, ***data center*** regional e, por fim, iii) o ***data center*** principal (centro de dados). Os recursos físicos no primeiro nível são abstraídos para criar um segundo nível, onde as funções de rede são ativadas como entidades virtualizadas. O nível superior consiste em serviços heterogêneos que devem consumir as aplicações das entidades virtualizadas no segundo nível, a fim de que elas forneçam seus respectivos serviços isolados entre si e de forma transparente. Resumidamente, na camada de criação de valor podem-se ter as soluções implantadas pela operadora que serão de utilidade para a sociedade como, por exemplo, soluções para telemedicina, para transporte e logística, entretenimento, entre outras. A plataforma que irá suportar todas essas soluções precisa implementar questões da teoria de decisão e inferência, controle de sistemas e transformação digital que têm como base um sistema de processamento de grandes quantidades de informações (***Big Data***). Na camada de habilitação de valor têm-se os ativos que fazem, por exemplo, o monitoramento das instalações da operadora e alguns controles de máquinas virtuais. Além disso, as funções de serviços de valor agregado (do inglês, ***Value Added Services*** - VAS) fazem parte dessa camada e compreendem, por exemplo, serviços de entrega e análise de vídeo de fluxo contínuo (***streaming***). As funções de redes, tanto de acesso, transporte e núcleo, bem como núcleo específico para tratamento de dados multimídia (IMS – ***Internet protocol multimedia subsystem***) e, também, de segurança cibernética são implementadas dentro dessa camada e são usadas pelas operadoras para entrega de diferentes serviços aos usuários.

COM O CONCEITO DE FATIAMENTO DE REDE, DIFERENTES SERVIÇOS COM DIFERENTES REQUISITOS, PODEM SER FORNECIDOS POR DIFERENTES FATIAS DE REDE QUE SÃO ESTABELECIDAS SOB UMA MESMA INFRAESTRUTURA FÍSICA, LEVANDO A UMA MAIOR ECONOMIA DO PONTO DE VISTA DE INVESTIMENTO EM RECURSOS TECNOLÓGICOS.

A visão das redes 5G como mostrada na **Figura 2** leva a um conceito muito importante de fatiamento (do inglês, ***slicing***) que se tornou um tema central nas redes 5G. Com o conceito de fatiamento de rede, diferentes serviços com diferentes requisitos, podem ser fornecidos por diferentes fatias de rede que são estabelecidas sob uma mesma infraestrutura física, levando a uma maior economia do ponto de vista de investimento em recursos tecnológicos.

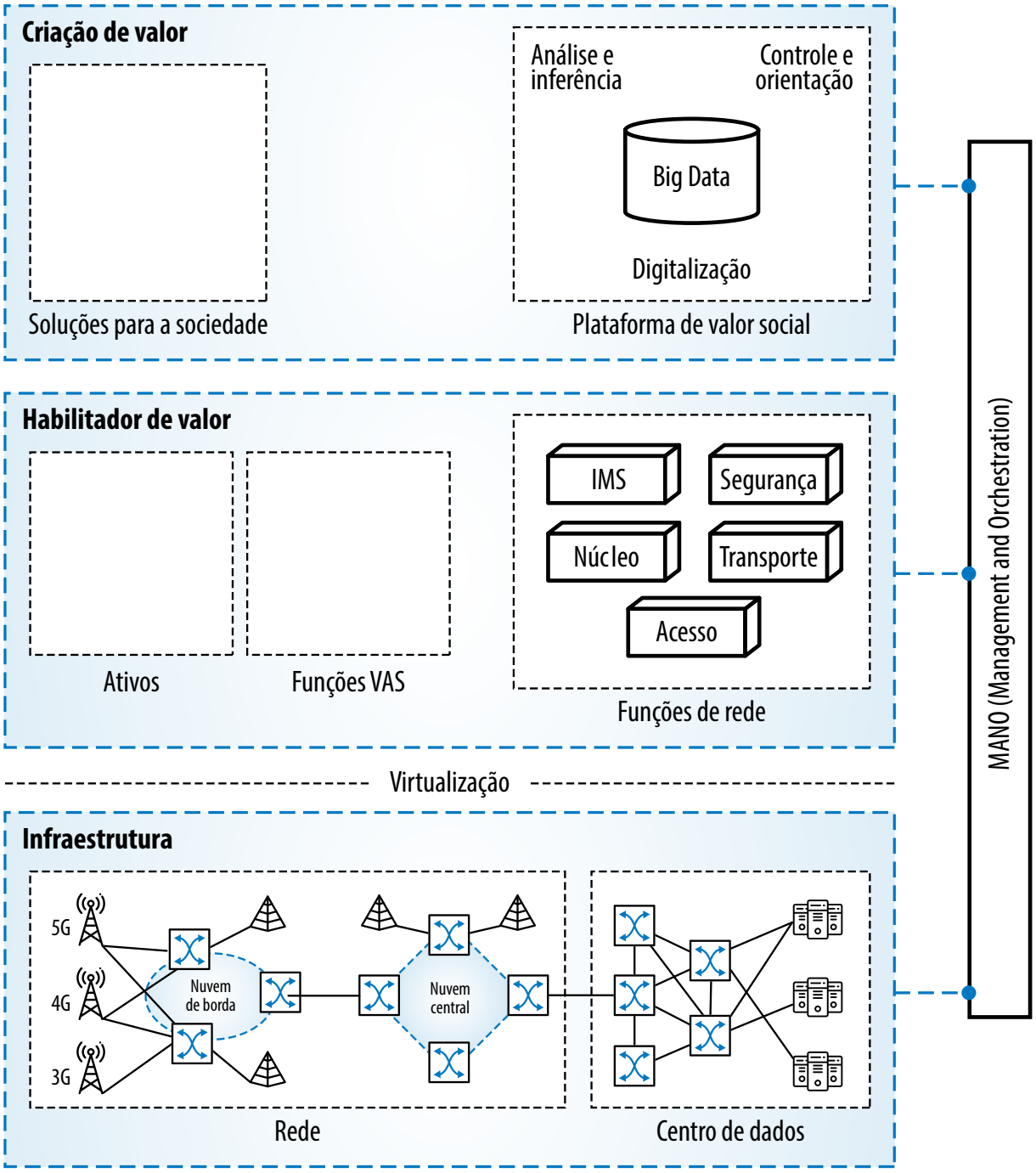

Figura 2: Visão geral em três camadas de uma rede de quinta geração.

A TECNOLOGIA NFV É PROJETADA ESPECIFICAMENTE PARA ATENDER OS REQUISITOS DE FLEXIBILIDADE, AGILIDADE E ESCALABILIDADE DA INFRAESTRUTURA VIRTUALIZADA. ENQUANTO QUE A SDN É DESENVOLVIDA PARA TORNAR OS SERVIÇOS DE CONECTIVIDADE FORNECIDOS POR REDES 5G MAIS PROGRAMÁVEIS, FAZENDO COM QUE OS FLUXOS DE TRÁFEGO POSSAM SER CONTROLADOS DINAMICAMENTE PARA OBTER O MÁXIMO DESEMPENHO DA REDE FÍSICA.

A **Figura 3** traz uma ilustração que representa o conceito de fatiamento de rede, onde os recursos presentes na infraestrutura física são dimensionados para criar várias instâncias de fatia de recursos, que são usadas para serviços específicos. As fatias de rede precisam ser implantadas e gerenciadas durante todo o período em que elas existirem. Portanto, essa forma de funcionamento da rede 5G traz alguns desafios chaves, onde se destacam a necessidade de um gerenciamento contínuo e flexível de recursos físicos e virtualizados nas três camadas mostradas na **Figura 2**, além de uma orquestração ágil de serviços de rede. Considerando esses desafios, duas tecnologias principais estão sendo desenvolvidas para atender os requisitos das redes móveis 5G, NFV e SDN.

A tecnologia NFV é projetada especificamente para atender os requisitos de flexibilidade, agilidade e escalabilidade da infraestrutura virtualizada. Enquanto que a SDN é desenvolvida para tornar os serviços de conectividade fornecidos por redes 5G mais programáveis, fazendo com que os fluxos de tráfego possam ser controlados dinamicamente para obter o máximo desempenho da rede física. As próximas subseções são destinadas a essas duas tecnologias, NFV e SDN.

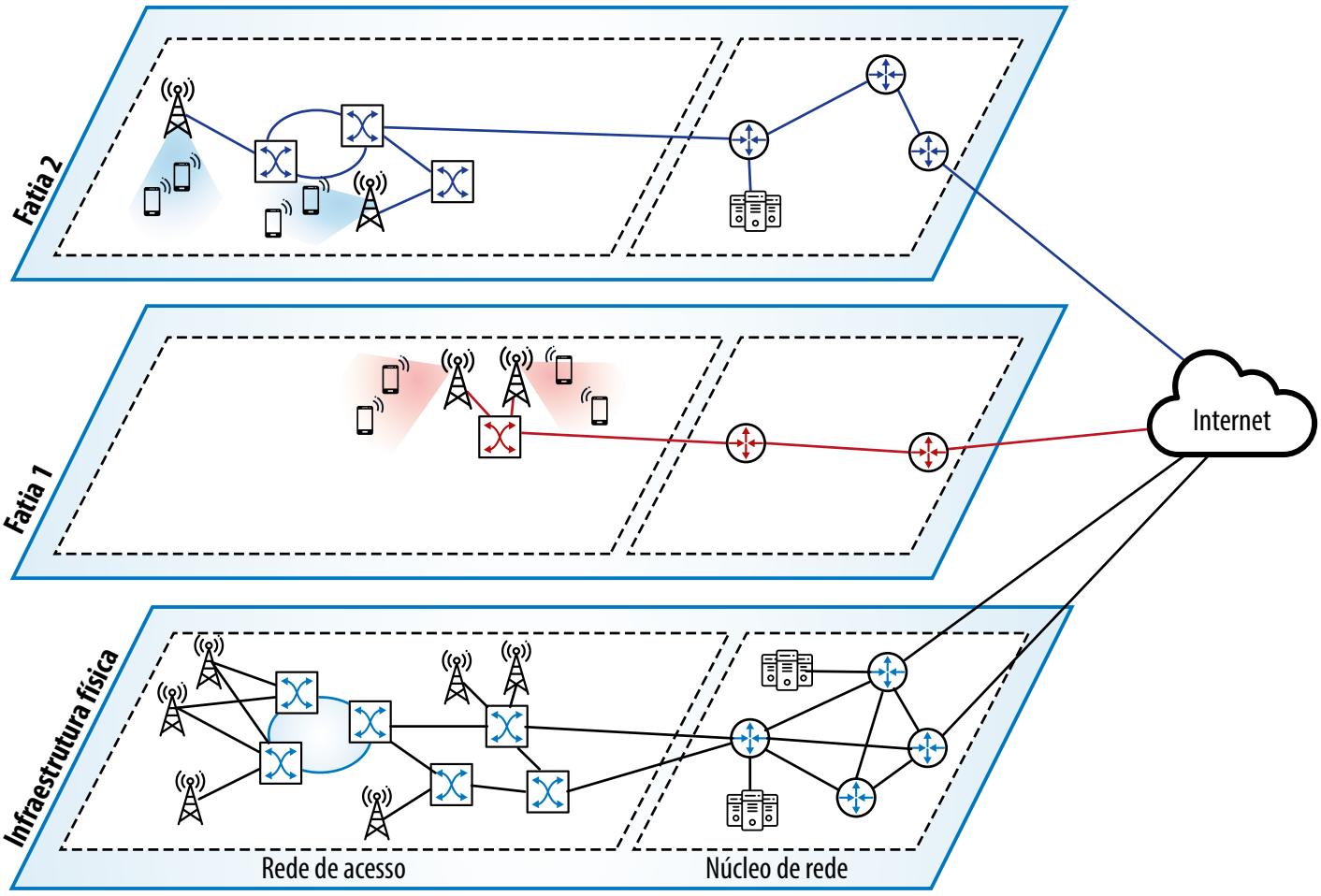

Figura 3: Ilustração sobre a técnica de fatiamento de rede.

3.1. VIRTUALIZAÇÃO DE FUNÇÃO DE REDES E SISTEMA DE ORQUESTRAÇÃO E GERENCIAMENTO

AS VNFS PODEM EMULAR FUNÇÕES DE REDE DESDE UM FIREWALL ATÉ MESMO FUNÇÕES MAIS COMPLEXAS, COMO UM EPC (EVOLVED PACKET CORE)

Tradicionalmente os sistemas NFV foram desenvolvidos para converter aplicativos monolíticos, dedicados a tratar alguma função de rede, em aplicações que funcionassem em máquinas virtuais (do inglês, ***Virtual Machine*** - VM). Portanto, cada VM pode abrigar uma ou mais Funções de Redes Virtualizadas (do inglês, ***Virtualized Network Function*** - VNF). Sendo assim, as VNFs podem emular funções de rede desde um ***firewall*** até mesmo funções mais complexas, como um EPC (***Evolved Packet Core***). Abordagens mais recentes visam um esquema de virtualização mais adequado à computação em nuvem e com maior desempenho, não sendo executado em máquinas virtuais, mas em soluções de contêineres.

Os sistemas de Gerenciamento e Orquestração (do inglês, ***Management and Orchestration*** - MANO) de NFV precisam gerenciar a infraestrutura virtualizada, infraestrutura de comunicação e rede, entidades NFV e os vários ciclos de vida de todos esses componentes. Tendo em vista a complexidade da rede móvel 5G e, principalmente o conceito de ***slicing***, é preciso oferecer uma plataforma MANO que gerencie e orquestre de forma eficaz os recursos físicos e virtuais da rede e que, ao mesmo tempo, seja sensível aos rígidos requisitos dos diferentes cenários que compõem essa nova abordagem.

A estrutura MANO adotada pelo 3GPP (***Third Generation Partnership Project***) para redes 5G é baseada no modelo especificado pela ETSI (***European Telecommunications Standards Institute***) e é representada na **Figura 4** [53]. Basicamente, a estrutura tem três blocos principais: i) o orquestrador NFV (do inglês, ***Network Function Virtualization Orchestration*** - NFVO), ii) o gerenciador de VNFs (do inglês, ***Virtualized Network Function Manager*** - VNFM) e, por fim, iii) o gerenciador da infraestrutura virtualizada (do inglês, ***Virtualized Infraestructure Manager*** - VIM).

De forma sucinta, o bloco VIM gerencia os recursos da infraestrutura NFV (do inglês, ***Network Functions Virtualization Infrastructure*** - NFVI) que é composta pelos componentes de ***hardware*** e ***software*** nos quais as redes virtuais são construídas. Já o VNFM cuida do gerenciamento de cada VNF individualmente. Sendo assim, o VNFM pode ser conectado diretamente aos EMs (***Element Managers***) e às VNFs para realizar ações, como iniciar e configurar as entidades relacionadas. É importante salientar que a estrutura MANO executa o gerenciamento do ciclo de vida de uma fatia da rede gerenciando os VNFs individuais que fazem parte dessa fatia da rede. O NFVO coordena as combinações de VNFs, PNFs (***Physical Network Functions***) e serviços de rede, integrando-os de forma coerente em um grafo de serviços. Além disso, o NFVO se conecta a sistemas externos como OSS (***Operating Support System***) e BSS (***Business Support System***).

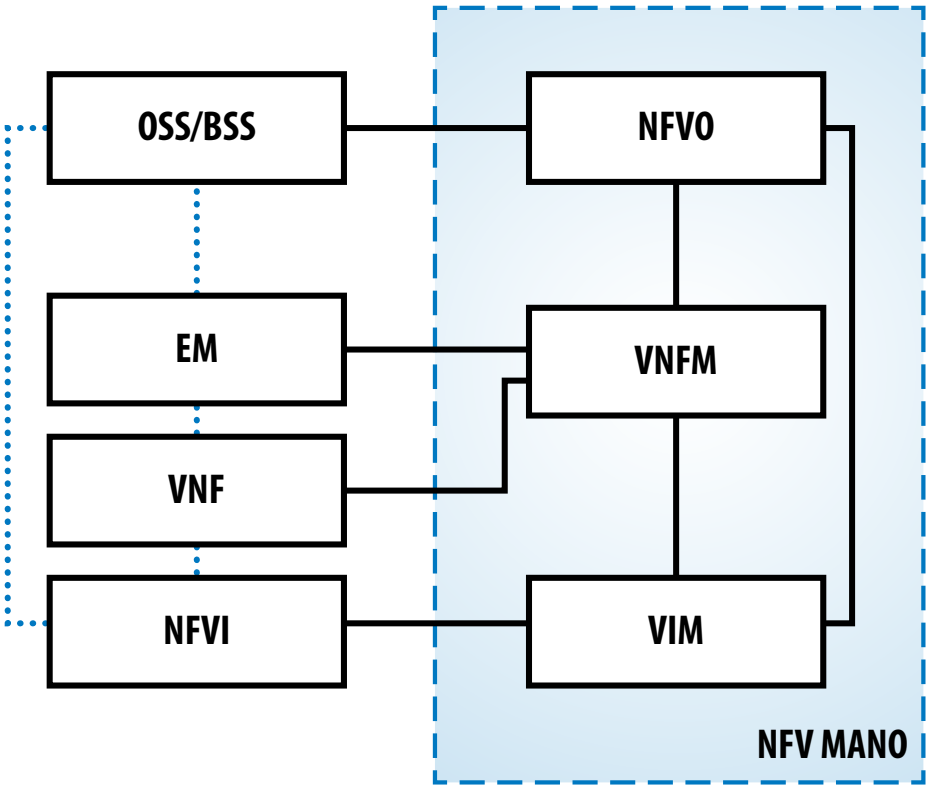

Figura 4: Estrutura ETSI para NFV MANO.

Outro proeminente projeto de uma estrutura MANO é denominado ONAP (***Open Network Automation Platform***). O ONAP, que é amparado pela ***Linux Foundation***, é uma plataforma de orquestração, gerenciamento e automação de redes e serviços de computação na borda. A abordagem da ONAP é semelhante à ETSI e sua implementação pode ser facilmente traduzida em descritores de serviço de rede definidos pela especificação ETSI [54]. A O-RAN ***Alliance*** publica de tempos em tempos versões de seus ***softwares*** de referência para a arquitetura O-RAN. Nessas publicações, a O-RAN utiliza o ONAP como base para desenvolvimento do seu gerenciamento e orquestração de serviços (do inglês, ***Service Management and Orchestration*** - SMO) que é um dos principais elementos dentro da arquitetura O-RAN [55].

3.2. REDES DEFINIDAS POR *SOFTWARE*

A PRINCIPAL MOTIVAÇÃO PARA DESENVOLVIMENTO DAS SDNS É CRIAR UMA SEPARAÇÃO ENTRE O PLANO DE CONTROLE E O HARDWARE DE REDE E, ASSIM, PERMITIR QUE O CONTROLE DA REDE SEJA FEITO DE FORMA EXTERNA POR MEIO DE UMA ENTIDADE DE SOFTWARE CHAMADA DE CONTROLADOR.

A principal motivação para desenvolvimento das SDNs é criar uma separação entre o plano de controle e o ***hardware*** de rede e, assim, permitir que o controle da rede seja feito de forma externa por meio de uma entidade de ***software*** chamada de controlador. O controlador, que irá gerenciar o controle do fluxo de pacotes, fica situado em ***data center*** e proporciona um controle programável dessa rede. Com o controlador SDN, os administradores de rede são capazes de gerenciar a rede 5G e introduzir novos serviços ou mudanças de forma dinâmica e programável, podendo obter maior desempenho dessa rede.

De acordo com a ONF (***Open Network Foundation***), uma arquitetura SDN deve-se basear nos seguintes pontos chaves [56]: i) dissociação do encaminhamento e processamento de tráfego do controle da rede, ii) controle logicamente centralizado e iii) serviços de rede habilitados por meio de programação. A arquitetura ilustrada na **Figura 5** é proposta pela ONF e, portanto, se baseia nesses pontos chaves [56]. Pode-se notar que os controladores SDN estão no centro da arquitetura. Eles são responsáveis pelo provisionamento, gerenciamento e controle de serviços e recursos relacionados. As interfaces que interligam o controlador SDN aos clientes são denominadas A-CPIs (***Applications-Controller Plane Interfaces***), enquanto que as interfaces que interligam esse mesmo controlador aos recursos são denominadas D-CPIs (***Data-controller Plane Interfaces***). Usando essas interfaces, os usuários e aplicativos têm a capacidade de interagir diretamente com a rede. Usando a A-CPI, as aplicações autorizadas estabelecem as sessões de controle para acessar serviços de controle ou para alterar o estado de recursos via D-CPIs. Os recursos incluem explicitamente: armazenamento, processamento e encaminhamento, ou seja, tudo o que é necessário para implementar uma aplicação do cliente. O conceito de recurso também inclui funções de rede que podem ser fornecidas por NFV.

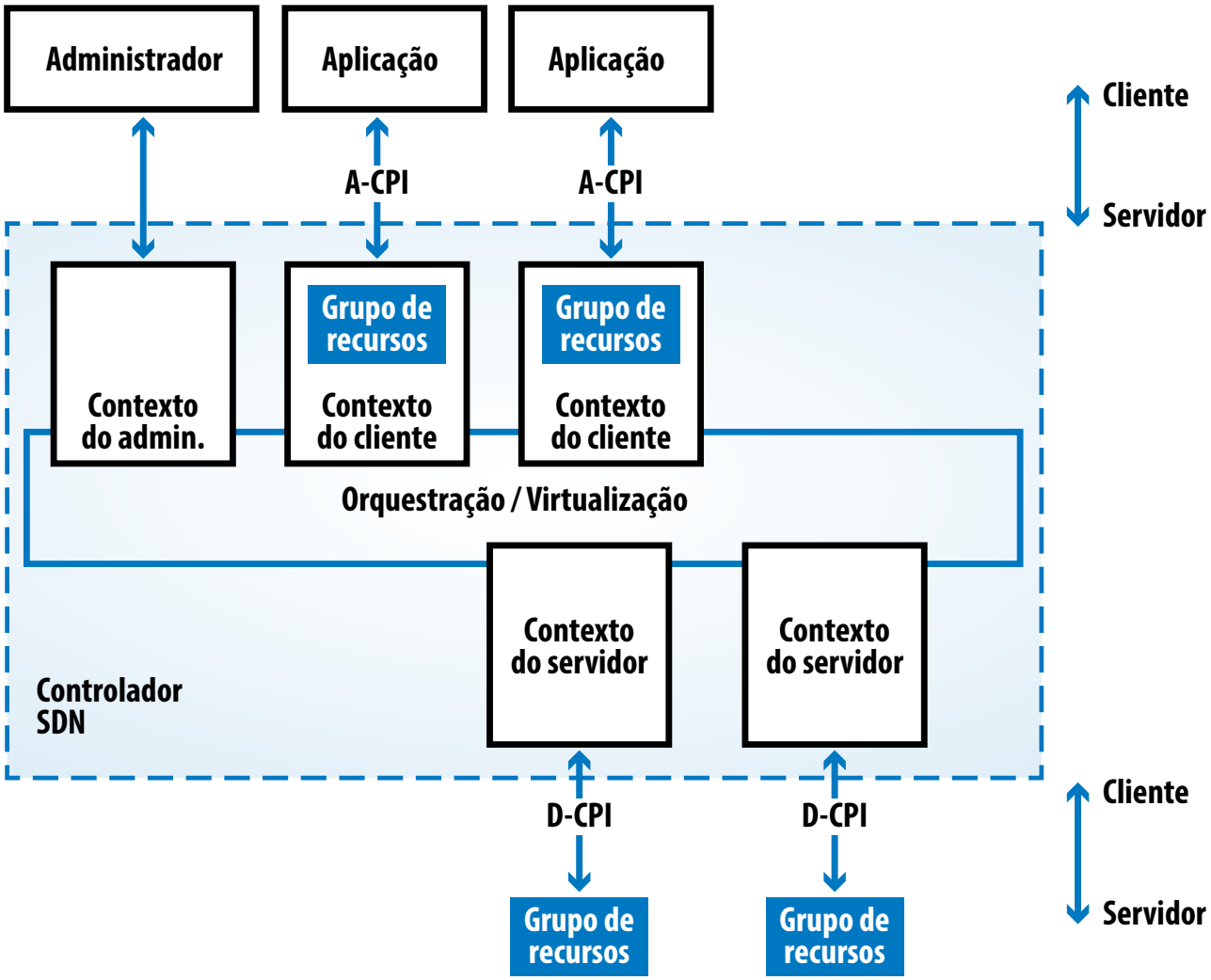

Figura 5: Arquitetura SDN proposta pela ONF.

O bloco denominado administrador é responsável por criar e manter o ambiente necessário para fornecer serviços aos clientes. Ele tem autoridade para configurar o controlador SDN, bem como para criar e gerenciar contextos de cliente e servidor. Para este fim, a configuração de um controlador SDN inclui a instalação e modificação das suas políticas internas e a instalação e configuração de recursos reais e aplicações de controle.

Outro ator importante dentro dessa estrutura é o protocolo de controle da SDN. O OF (***Open Flow***) teve um impacto significativo na evolução das redes de computadores e deu início à evolução rumo ao SDN. O OF permite a comunicação entre os elementos da infra-estrutura de rede e as entidades de controle que são baseadas em ***software***. O OF é mantido pela ONF e hoje é implementado por todos os principais fornecedores de equipamentos de rede. Atualmente, a linguagem de programação P4 (***Programming Protocol-independent Packet Processors***) promete impulsionar essa inovação ainda mais. O P4 trabalha em conjunto com protocolos de controle SDN como o OF, mas permite que novos protocolos de controle sejam criados, bem como novas estruturas para comutadores programáveis. O P4 faz parte da chamada NG-SDN (***Next Generation Software Defined Network***).

A arquitetura SDN fornece o conjunto abstrato completo de recursos e lógica de controle que constitui uma fatia de rede. Segundo a ONF [57], o conceito da fatia de rede 5G é muito similar ao contexto de cliente SDN. O contexto do cliente, como o nome sugere, oferece funções para que o cliente possa gerenciar e controlar os recursos das fatias de rede, incluindo funções relacionadas à Operação, Administração e Gerência (do inglês, ***Operations, Administration and Maintenance*** - OAM). Portanto, é possível verificar que cada contexto de cliente representa um conjunto de recursos gerenciados e controlados por um controlador e é diretamente aplicável ao fatiamento 5G.

As tecnologias de NFV e SDN são elementos centrais para as arquiteturas OpenRAN. A Seção 6.1 traz uma visão mais detalhada da arquitetura proposta pela O-RAN ***Alliance*** e como seus elementos utilizam e interagem com as tecnologias NFV e SDN.

4. INTELIGÊNCIA ARTIFICIAL NO 5G

4.1. INTRODUÇÃO

AS REDES 5G POSSUEM GRANDE COMPLEXIDADE E TRAZEM VÁRIOS NOVOS DESAFIOS, SENDO A INTEGRAÇÃO DE ALGORITMOS DE IA ÀS REDES 5G UMA FORMA DE ENDEREÇÁ-LOS.

IA É DEFINIDA COMO O CAMPO QUE ESTUDA "AGENTES INTELIGENTES", ISTO É, QUALQUER DISPOSITIVO QUE PERCEBE SEU AMBIENTE E EXECUTA AÇÕES QUE MAXIMIZAM SUA CHANCE DE ATINGIR SEUS OBJETIVOS COM SUCESSO

IA é definida como o campo que estuda "agentes inteligentes", isto é, qualquer dispositivo que percebe seu ambiente e executa ações que maximizam sua chance de atingir seus objetivos com sucesso [58]. A IA é dividida em várias subáreas (ou campos), as quais são baseadas em considerações técnicas, como objetivos específicos (por exemplo, robótica ou aprendizado de máquina), o uso de ferramentas específicas (lógica ou redes neurais artificiais) ou profundas diferenças filosóficas [58,59]. ML (***Machine Learning***) é uma das várias subáreas da IA e tem como principal objetivo a criação de algoritmos que melhorem automaticamente seu desempenho por meio da experiência [60].

Como apresentado na Seção 1 deste documento, as redes 5G possuem grande complexidade e trazem vários novos desafios, sendo a integração de algoritmos de IA às redes 5G uma forma de endereçá-los. As redes 5G permitirão que operadoras forneçam uma grande variedade de serviços, porém, a flexibilidade e a riqueza do padrão 5G tornam sua otimização e gerenciamento mais complexos, com uma grande variedade de métricas de desempenho para otimizar.

SOLUÇÕES AUXILIADAS POR IA TÊM O POTENCIAL PARA GERENCIAR ESSA ESCALADA DE COMPLEXIDADE POR MEIO DE RECURSOS COMO REDES AUTO-CONFIGURÁVEIS, AUTO-GERENCIÁVEIS E AUTO-CURÁVEIS QUE USAM TECNOLOGIAS BASEADAS EM ML PARA AUTOMATIZAR FUNÇÕES DA REDE E REDUZIR O OPEX

Os meios tradicionais, isto é, baseados em operadores humanos, usados para implantar, otimizar e operar redes móveis podem não ser capazes de atingir o nível de otimização necessário no 5G. Soluções auxiliadas por IA têm o potencial para gerenciar essa escalada de complexidade por meio de recursos como redes auto-configuráveis, auto-gerenciáveis e auto-curáveis que usam tecnologias baseadas em ML para automatizar funções da rede e reduzir o OPEX. Esta rede "inteligente" deve ser capaz de sensoriar os contextos ambiental e de aplicação, bem como interpretar e agir sobre as informações contextuais em tempo real, de forma extremamente eficiente.

Na sequência, apresentam-se alguns dos benefícios que a adoção de algoritmos de IA/ML trará às redes 5G, além de listar alguns casos de uso e discutir desvantagens e desafios relacionados ao seu emprego.

4.2. BENEFÍCIOS DO USO DE AI/ML EM REDES 5G

Técnicas de IA/ML trarão vários benefícios para redes 5G, entre os mais importantes, pode-se citar [61]:

- Redução de CAPEX. Por exemplo, a infraestrutura das redes atuais é superdimensionada para oferecer redundância e garantir que falhas de hardware não causem interrupções nos serviços. Porém, isso aumenta o CAPEX. Modelos de IA/ML podem fornecer mecanismos para se lidar de forma mais inteligente e previsível com essas falhas. A grande quantidade de dados de telemetria disponíveis pode ser analisada em tempo real por esses modelos para se analisar a integridade da rede, prever falhas e sugerir ações corretivas antes da sua ocorrência.

- Otimização do desempenho da rede. Algoritmos AI/ML podem ser executados continuamente para monitorar desvios de desempenho e corrigi-los, automaticamente. Algoritmos de IA/ML poderão gerenciar recursos de rádio de forma eficiente e otimizada por meio de controle de malha fechada com o intuito de aprimorar o desempenho da rede e a experiência dos usuários, isto é, qualidade de serviço (do inglês, Quality of Service - QoS) e qualidade de experiência (do inglês, Quality of Experience - QoE).

- Criação de novos fluxos de receita. Um dos grandes potenciais para a geração de novas receitas para operadoras e fabricantes está na digitalização da indústria por meio do uso de 5G e IoT. Soluções baseadas nestas tecnologias dão às operadoras e fabricantes a oportunidade de construir fábricas inteligentes que aproveitem as vantagens de tecnologias como automação, IA e realidade aumentada para solução de problemas e criação de novas aplicações e serviços que possam trazer novas receitas.

- IA será vital para melhorar o atendimento e aprimorar a experiência do cliente. Espera-se que modelos de IA ajudem as operadoras a melhorar ainda mais a experiência do cliente de várias maneiras, incluindo melhorar a qualidade da rede e fornecer serviços personalizados.

- Modelos de IA/ML ajudarão a recuperar os investimentos que as operadoras estão fazendo em suas redes para a implantação do 5G. Reduzir os custos operacionais e garantir o retorno sobre os investimentos na rede são as principais prioridades que as operadoras procuram alcançar usando IA/ML. Grande parte das operadoras acredita que os maiores retornos potenciais da adoção de IA/ML estarão no planejamento e gerenciamento de desempenho de rede.

4.3. PAPÉIS DE ALGORITMOS DE IA/ML EM REDES 5G

A INCORPORAÇÃO DE TÉCNICAS DE IA/ML À REDE PODE, POR EXEMPLO, AJUDAR A CRIAR ESTRATÉGIAS DE FATIAMENTO E RESOLVER FALHAS DE MANEIRA INTELIGENTE E OTIMIZAR O DESEMPENHO AUTOMATICAMENTE, DE MODO A OBTER UMA ALOCAÇÃO INTELIGENTE DE RECURSOS E FORNECER UMA CONFIGURAÇÃO ÓTIMA.

Existem vários papéis importantes que algoritmos de IA/ML podem ter em redes 5G, entre eles alguns são listados na sequência [62,63,64,65].

- **Fatiamento de recursos de rede**: o fatiamento de recursos é crítico para diversos casos de uso e para maximizar a flexibilidade das redes 5G. A incorporação de técnicas de IA/ML à rede pode, por exemplo, ajudar a criar estratégias de fatiamento e resolver falhas de maneira inteligente e otimizar o desempenho automaticamente, de modo a obter uma alocação inteligente de recursos e fornecer uma configuração ótima. Além disso, técnicas de IA/ML podem detectar anomalias e falhas relacionadas ao fatiamento, usar o aprendizado para melhorar as estratégias de fatiamento, podendo também garantir soluções de fatiamento de recursos mais econômicas e de alta qualidade.

- **Planejamento para implantação de redes**: técnicas de IA/ML podem ser usadas para planejar a implantação de redes 5G. A implementação inicial de redes 5G enfrentará vários desafios associados à alocação/distribuição das células, sistemas variados, várias bandas de frequência e configuração das funcionalidades de rede. Tradicionalmente, a distribuição das células em uma rede depende de simulações, *drive-tests* e outros mecanismos que se utilizam da experiência de especialistas e que também requerem uma quantidade substancial de mão de obra. No entanto, conforme o 5G é introduzido, o modo de implantação existente enfrentará várias restrições e desafios, especialmente para cenários híbridos que compreendem várias tecnologias, diversas frequências e várias células. Técnicas de IA/ML podem ser usadas no planejamento dessas redes com base em conhecimentos relacionados à cobertura, distribuição das células, gerenciamento de tráfego, histórico de reclamações e análise de parâmetros de rede para melhorar o planejamento e aumentar a capacidade da rede. Dessa forma, as operadoras conseguem tornar o planejamento de infraestrutura mais próximo do ideal teórico e podem reduzir significativamente o custo com mão de obra tanto no planejamento quanto na implantação da rede.

- ***Handover* dinâmico em comunicações V2X (*Vehicle-to-Everything*)**: o gerenciamento de comunicações V2X pode trazer inúmeros benefícios, como maior segurança nas estradas, redução de emissões e redução do tempo de viagem. A tecnologia V2X é baseada na troca de mensagens contendo as identificações das células e medições do estado do canal de rádio. Conforme os veículos trafegam ao longo de uma rodovia, sequências de *handover* subótimas e algumas anomalias podem prejudicar substancialmente a conectividade e, portanto, o desempenho do sistema V2X. Assim, a mitigação desses problemas é de grande importância para este tipo de comunicação. O uso de técnicas de IA/ML permite que informações sobre o estado do canal e informações de *handover* sejam utilizadas para treinar modelos que preveem anomalias e podem, com isto, realizar *handovers* proativos, diminuindo os problemas de perda de conectividade.

- **Setorização dinâmica**: em uma rede de telecomunicações, a experiência do usuário pode ser melhorada adaptando-se às fronteiras dos setores das células às mudanças de tráfego ao longo do dia. A setorização dinâmica maximiza a cobertura da célula, reduz a interferência entre células e setores e o número de *handovers* necessários. Algoritmos de IA/ML podem ser utilizados para aprender o comportamento do tráfego ao longo do dia e ajustar automaticamente os setores da célula, enquanto reduzem a interferência intra- e inter- celular.

- **MIMO massivo**: MIMO massivo é uma das tecnologias chaves do 5G. Para aproveitar as vantagens da tecnologia, atender às necessidades de cobertura e fornecer experiência de usuário ideal, os feixes de transmissão altamente direcionados (do inglês, *beamforming*) criados pelo MIMO massivo precisam corresponder à distribuição dos usuários ao redor da célula e minimizar a interferência entre células vizinhas. Técnicas de IA/ML podem ser usadas para modelagem e estimação de canal além de codificação e detecção de sinais para a tecnologia MIMO massivo. Esses algoritmos podem ser aplicados, por exemplo, aos parâmetros de *beamforming* para ajustar a cobertura da célula em diferentes cenários. Além disso, mecanismos inteligentes para o cálculo de interferência e otimização adaptativa em redes intra- e inter- celulares resultarão em maior eficiência espectral e melhor experiência de usuário.

- **Circuitos de RF:** técnicas de IA/ML podem ser usadas para prever a necessidade da agregação de portadora entre nós da rede 5G e com isso melhorar a qualidade do serviço oferecido aos usuários. Além disso, a aplicação desses algoritmos ao projeto e fabricação de circuitos de RF para a faixa de ondas milimétricas pode melhorar o desempenho de tais circuitos e reduzir o tempo de chegada ao mercado.

- **Manutenção preditiva**: a análise preditiva baseada em modelos de IA/ML pode ajudar as operadoras a prever resultados futuros com base em dados históricos e assim prestar um serviço melhor aos usuários. Isso significa que as operadoras podem usar informações aprendidas por meio dos dados para monitorar o estado dos equipamentos, antecipar falhas com base em padrões e corrigir proativamente problemas de *hardware*, como aqueles que acometem torres de celular, linhas de energia, servidores de *data center* e até mesmo equipamentos nas casas dos clientes. Desta forma, as operadoras podem corrigir problemas antes que os usuários sofram um impacto negativo em seus serviços.

- **Otimização da rede**: o padrão 5G é muito complexo para se usar o mesmo tipo de otimização estática e manual usada pelas gerações anteriores de redes móveis. As redes 5G mudam sua topologia dinamicamente, respondendo às mudanças no tráfego. Quanto melhor a otimização, mais eficiente será o desempenho da rede em termos de espectro e uso de energia. Algoritmos de IA/ML são essenciais para ajudar as operadoras a construir redes de otimização automática (do inglês, *Self-Optimizing Networks* - SONs), as quais oferecem às operadoras a capacidade de otimizar automaticamente o desempenho de componentes com base nas informações de tráfego, por exemplo.

ALGORITMOS DE IA/ML SÃO ESSENCIAIS PARA AJUDAR AS OPERADORAS A CONSTRUIR REDES DE OTIMIZAÇÃO AUTOMÁTICA (DO INGLÊS, SELF-OPTIMIZING NETWORKS - SONS), AS QUAIS OFERECEM ÀS OPERADORAS A CAPACIDADE DE OTIMIZAR AUTOMATICAMENTE O DESEMPENHO DE COMPONENTES COM BASE NAS INFORMAÇÕES DE TRÁFEGO, POR EXEMPLO.

- **Suporte ao cliente**: algoritmos de IA podem ser usados na criação de plataformas de atendimento aos clientes. Essas plataformas, também conhecidas como assistentes virtuais, aprendem a conversar com os clientes de forma bastante eficiente. As operadoras têm recorrido a assistentes virtuais auxiliados por IA para ajudar a lidar com o grande número de solicitações de suporte para instalação, configuração, solução de problemas e manutenção, que muitas vezes sobrecarregam os centros de atendimento ao usuário.

- **Aumento da receita**: para aumentar a receita, o serviço apropriado deve ser oferecido ao cliente certo, no momento correto. Ferramentas de vendas baseadas em IA podem ajudar os departamentos de vendas das operadoras a fazer exatamente isso. Um modelo de IA construído em torno dos serviços das operadoras, de seus clientes e comportamentos pode oferecer recomendações aos vendedores para as próximas ofertas. Os modelos podem estimar a probabilidade de um cliente aceitar uma determinada oferta, prever o melhor canal para contatá-lo e identificar a melhor mensagem para ser enviada, ajudando assim as operadoras a alcançar os usuários com maior probabilidade de venda do serviço.

- **Detecção de fraudes**: o fluxo de dados que trafega através das redes das operadoras está crescendo a uma taxa sem precedentes. Com isso, o número de indivíduos mal-intencionados que querem explorar ou fraudar esses dados tem também aumentado. Portanto, não se pode esquecer dos aspectos de segurança que são um grande desafio. Diferentes tipos de fraudes e ameaças significam que há uma necessidade urgente de ajustar os sistemas de gerenciamento de fraudes para combatê-las, de preferência em tempo real. Como a implantação de novos serviços e produtos sempre envolve risco de fraude e explorações mal-intencionadas, ações preventivas devem ser tomadas com antecedência, a fim de identificar possíveis abusos a tempo. A solução é implementar um sistema de gerenciamento de fraudes e ameaças inteligente. Algoritmos de IA/ML são capazes de encontrar relacionamentos complexos entre muitos atributos, o que é difícil para sistemas baseados em regras. O treinamento periódico dos modelos permite que eles levem em consideração as mudanças no comportamento do fraudador. Utilizar ML significa que não há necessidade de se criar manualmente regras muito complicadas, porque os modelos aprendem com base em exemplos conhecidos.

- **Computação na borda/fronteira da rede** (do inglês, *edge computing*): traz os recursos de computação e armazenamento de dados para mais próximo do local onde são necessários (isto é, usuários), para melhorar os tempos de resposta, aumentar a segurança e economizar largura de banda, o que acarreta em economia de custos no *backhaul*. A combinação de 5G e IA faz surgir novas possibilidades como a melhoria das operações de rede, especialmente em sua borda e a criação de novas oportunidades de serviço. Esta junção tem sido vista como uma plataforma para que as operadoras forneçam serviços de borda abertos e para que desenvolvedores criem aplicações que suportem consumidores, empresas e vários setores verticais [62]. A tendência mais recente nessa direção é a adição de inteligência na borda da rede, utilizando arquiteturas distribuídas para sistemas de IA, onde os dispositivos na borda são configurados para tomar decisões urgentes, enquanto a computação em nuvem é usada para treinamento e ajuste fino dos modelos de IA. Além disso, a sensibilidade à privacidade dos dados também está resultando em dispositivos nas extremidades da rede desempenhando um papel maior no treinamento de modelos de IA/ML, melhorando suas capacidades de inferência. A computação de borda será particularmente importante para o aprendizado de máquina e outras formas de inteligência artificial, como reconhecimento de imagem, análise de fala e uso de sensores em larga escala. Casos de uso específicos incluem vídeo vigilância e segurança, realidades virtual e aumentada, serviços de localização, teletriagem, controle de veículos autônomos, robôs industriais conectados, fluxo do tráfego de dados e previsão de congestionamento para cidades inteligentes e assim por diante [62].

- **Computação distribuída e inteligente**: os dados de treinamento usados por algoritmos de IA/ML, incluindo imagens de vídeo, medições relacionadas à saúde, estatísticas de tráfego e muito mais, estão normalmente localizados em dispositivos de borda da rede sem fio. Transferir esses dados locais para a nuvem para o treinamento de modelos de IA/ML incorre em custos significativos de comunicação, atrasos de processamento e questões de privacidade. Desta forma, abordagens de IA/ML distribuídas através da rede sem fio têm ganhado bastante popularidade. A motivação aqui é usar a borda da rede sem fio como um mecanismo de comunicação/computação integrado para aprendizagem distribuída. Dada a onipresença da infraestrutura da rede sem fio, pode-se imaginar uma tendência crescente de usar a borda da rede sem fio para treinamento de modelos de IA/ML, indo além das tarefas de inferência e *offload* computacional. A tendência para o treinamento na borda surge no modelo de Aprendizado Federado, onde um servidor central orquestra o treinamento local de algoritmos de IA/ML em um grande número de clientes e, em seguida, agrega os modelos locais para desenvolver um modelo global (compartilhado) mais sofisticado sem exigir que os clientes compartilhem seus dados privados [66,67]. Mais especificamente, o Aprendizado Federado funciona assim: o dispositivo do cliente baixa o modelo atual, melhora-o, aprendendo com os dados locais e, em seguida, resume as alterações como uma pequena atualização. Apenas essa atualização do modelo é enviada para o servidor central (nuvem), usando comunicação criptografada, onde é imediatamente calculada a média com outras atualizações dos clientes para melhorar o modelo compartilhado. Todos os dados de treinamento permanecem no dispositivo dos clientes e nenhuma atualização individual é armazenada na nuvem. Espera-se que a necessidade de treinamento local continue a crescer à medida que os modelos de aprendizagem mudam para uma abordagem de aprendizagem adaptativa, *online* e em tempo real, com o intuito de atender à necessidade de serviços em tempo real emergentes.

- **Alocação dinâmica de recursos para veículos aéreos não tripulados**: a aplicação de veículos aéreos não tripulados (do inglês, *Unmanned Aerial Vehicles* - UAVs) tem desempenhado um grande papel em aplicações de proteção de plantações, inspeção de redes de transmissão, fiscalização de fronteiras, exploração geológica, monitoramento ambiental, etc. Entretanto, com mastros e antenas implantados para a cobertura de dispositivos localizados no solo, a cobertura para UAVs pode ser irregular ou inexistente, o que pode levá-los a pousarem ou a retornarem à base. Modelos de IA/ML treinados com medições relacionadas aos UAVs, com base em relatórios de medição dos equipamentos de usuário (por exemplo, condição do canal de rádio, informações de trajetória de voo, clima, áreas proibidas para voo, uso do espaço aéreo e outras medições), podem realizar, considerando as várias informações coletadas pelo sistema, a alocação dinâmica de recursos de rádio para cobertura sob demanda para UAVs.

4.4. DESAFIOS REFERENTES À ADOÇÃO DE TÉCNICAS DE IA/ML EM REDES 5G

IA/ML E 5G REPRESENTAM UMA COMBINAÇÃO PODEROSA DE TECNOLOGIAS, NO ENTANTO, A ADOÇÃO E A INTERAÇÃO DE AMBAS AS TECNOLOGIAS TÊM ALGUNS DESAFIOS ASSOCIADOS.

IA/ML e 5G representam uma combinação poderosa de tecnologias, no entanto, a adoção e a interação de ambas as tecnologias têm alguns desafios associados.

Algoritmos de IA/ML estão impactando profundamente os processos comerciais e industriais, onde as máquinas estão assumindo tarefas anteriormente realizadas apenas por humanos. Portanto, existe uma urgência em aprimorar e/ou capacitar as equipes de operação das redes em preparação para o deslocamento de trabalho que resultará da interseção entre AI/ML e 5G, pois juntos eles oferecem um nível de produtividade e eficiência que os humanos não podem igualar. Portanto, o foco do aprimoramento e capacitação deve ser nas habilidades humanas que as máquinas não podem modelar. Outro grande desafio está relacionado à falta de recursos humanos dedicados exclusivamente ao treinamento dos modelos de IA/ML.

IA quase sempre interage direta ou indiretamente com os humanos e essa interação dá origem a várias questões sociais, estratégicas, de segurança e éticas. O rápido desenvolvimento de técnicas de IA trazem enormes benefícios potenciais. No entanto, é necessário explorar todos os aspectos éticos, sociais e legais dos sistemas de IA quando deseja-se evitar consequências negativas e riscos decorrentes da implementação de IA na sociedade. Portanto, um trabalho importante será a inserção de regras, regulamentos, observação de questões éticas, etc., relacionadas à integração de IA em redes de telecomunicações. Ao longo dos próximos anos, o 5G irá, sem dúvida, permitir o uso de IA em várias partes das redes de telecomunicações, mas será responsabilidade das empresas garantir que isso seja realizado de forma estratégica, segura e ética [68].

Capacitar a força de trabalho para operar essa nova tecnologia será um grande desafio para as operadoras atacarem sozinhas. Elas devem se planejar com bastante antecedência ou correm o risco de não colherem os resultados do uso delas. A competência dessas equipes terá grandes implicações no desempenho da rede.

Com várias preocupações quanto à privacidade e segurança dos dados dos usuários, uma questão importante levantada é quanto à segurança dos algoritmos de IA/ML quanto ao vazamento de dados privados. Uma pergunta ainda sem resposta definitiva é se eles podem ser considerados seguros para uso? É necessário se criar um conjunto de requisitos que vão desde os usuários até requisitos de integridade funcional da rede. Existem vários modelos de segurança criados pela ISO (***International Organization for Standardization***)/IEC (***International Electrotechnical Commission***) que focam na segurança dentro de uma organização e que podem auxiliar a lidar com esses desafios [69]. Além das questões de segurança, outro fator que pode impedir a adoção de algoritmos de IA/ML pode ser simplesmente a falta de confiança nos resultados gerados por eles. Vencer esses desafios exigirá muitos testes juntamente com medições contínuas da precisão/desempenho de tais algoritmos.

A integração de modelos de IA/ML em implantações reais apresenta vários desafios relacionados com a arquitetura, procedimentos adotados pelas redes e estabilidade das soluções. Isso ocorre principalmente porque dados em tempo real da rede precisam ser coletados e agregados para permitir as operações de treinamento e inferência dos modelos. Portanto, necessita-se de uma arquitetura que apresente formas para coletar e agregar os dados para que modelos de IA/ML possam operar. Além disso, outro desafio importante é a estabilidade das soluções baseadas em IA. Por exemplo, pode-se ter um período transitório de acomodação da tomada de decisão, ou seja, do treinamento do modelo, o que pode levar a instabilidades e consequente degradação do QoS e QoE entregues pela rede.

Manter modelos de IA/ML para milhares de células será algo bastante complexo. Esses modelos geralmente não são eficientes para execução em processadores de uso geral (do inglês, ***General Purpose Processors*** – GPPs) e podem exigir a utilização de GPUs (***Graphics Processing Units***), o que, por sua vez, pode acarretar em despesas maiores. Por outro lado, tem-se hoje desenvolvimentos em computação neuromórfica com cerca de 1000 vezes menos gasto energético [70,71]. Tem-se ainda um avanço recente em IA fotônica, que pode ser outro ***breakthrough*** para 5G/6G nos próximos anos [72]. Tais avanços permitirão a execução de IA em ***hardware*** com reduzido gasto energético.

Outro desafio são os altos custos relacionados com o rotulamento e preparação (isto é, pré-processamento) dos dados para algoritmos de ML supervisionados. Este é um processo intrincadamente complexo e que necessita de mão de obra especializada (isto é, cientistas e engenheiros de dados), o que o torna um processo caro. Na maioria dos casos, as tarefas de rotulagem e preparação dos dados são feitas por humanos, o que torna o processo sujeito a erros e que consequentemente afeta o desempenho e a precisão dos modelos de IA/ML. Além disso, como os dados estão sempre sendo gerados, a mão de obra humana envolvida na rotulagem pode ficar presa nesse processo por tempo indeterminado o que pode travar os esforços de criação e desenvolvimento de algoritmos e modelos de IA/ML para promover os objetivos de negócios. Uma possível alternativa para este desafio pode ser a terceirização e/ou ***crowdsourcing*** destas tarefas.

A aplicação de modelos de IA/ML para computação de borda promete flexibilidade, escalabilidade, reutilização de ***software***/***hardware*** e ***design*** de sistemas automatizados. No entanto, existem vários desafios ao se aplicar IA/ML a sistemas de computação de borda em redes sem fio. O mais importante é a capacidade dos modelos de IA/ML de se adaptarem em tempo real à dinâmica de mudanças rápidas que ocorrem na borda das redes sem fio, e normalmente com quantidade limitada de dados. Permitir uma aprendizagem confiável, com quantidade limitada de dados e em tempo real na borda de redes sem fio exigirá uma abordagem interdisciplinar, capaz de compreender a teoria fundamental por trás das técnicas de IA/ML, adaptando algoritmos de IA/ML para aplicações sem fio e compreendendo a natureza incerta e dinâmica da aprendizagem por canais sem fio.

AS PRIMEIRAS OPERADORAS QUE VENCEREM ESSES E OUTROS DESAFIOS RELACIONADOS COM A ADOÇÃO DE ALGORITMOS DE IA/ML EM REDES 5G TERÃO UMA CLARA VANTAGEM SOBRE AS OUTRAS

A distribuição ideal de inteligência é um desafio importante para sistemas 5G que devem servir a uma grande variedade de sistemas autônomos distintos. Os sistemas autônomos atuais dependem principalmente de políticas estáticas definidas por humanos. Essa abordagem funciona bem com os sistemas mais simples de hoje, mas quando se trata do controle aninhado de sistemas autônomos díspares e com restrições de latência rigorosas, técnicas avançadas, como políticas de auto aprendizado e controle autônomo, se tornam essenciais [62].

As limitações das redes sem fio e dos dispositivos devem ser tratadas de forma a permitir a integração total de abordagens de IA/ML (i.e., treinamento e inferência) na borda da rede. Conectividade não confiável e que muda dinamicamente devido às condições do canal sem fio, juntamente com a mobilidade dos dispositivos de borda, cria desafios para a distribuição de cargas de trabalho de aprendizagem para a nuvem ou outros dispositivos de borda. Além disso, os dispositivos sem fio são heterogêneos no sentido que recursos de computação, armazenamento e geração de dados de diferentes dispositivos são díspares e refletem apenas observações parciais. Isso torna difícil para os modelos de aprendizado convergirem para o modelo verdadeiro com base apenas em conjuntos de dados locais, ao realizar o treinamento de modelos de IA/ML na borda da rede. Portanto, há a necessidade de se desenvolver dispositivos que suportem as técnicas necessárias visando permitir aprendizado confiável que leva em conta as limitações envolvidas na comunicação sem fio [62,66,67,73].

Assim, pode-se concluir que as primeiras operadoras que vencerem esses e outros desafios relacionados com a adoção de algoritmos de IA/ML em redes 5G terão uma clara vantagem sobre as outras. Tais algoritmos terão o potencial de criar oportunidades interessantes para elas, pois eles poderão ajudar a gerenciar os custos de implantação e manutenção de redes, customizar a infraestrutura e serviços mais facilmente, realizar manutenção preditiva e oferecer qualidade de serviço e experiência muito melhores do que as oferecidas atualmente.

TAIS ALGORITMOS TERÃO O POTENCIAL DE CRIAR OPORTUNIDADES INTERESSANTES PARA ELAS, POIS ELES PODERÃO AJUDAR A GERENCIAR OS CUSTOS DE IMPLANTAÇÃO E MANUTENÇÃO DE REDES, CUSTOMIZAR A INFRAESTRUTURA E SERVIÇOS MAIS FACILMENTE, REALIZAR MANUTENÇÃO PREDITIVA E OFERECER QUALIDADE DE SERVIÇO E EXPERIÊNCIA MUITO MELHORES DO QUE AS OFERECIDAS ATUALMENTE.

Outra conclusão importante é que a adoção na área de telecomunicações ainda está em um estágio inicial. Os próximos 5 a 10 anos serão cruciais para a transformação das redes de telecomunicações. O 5G oferece às operadoras e fabricantes uma oportunidade sem precedentes de otimizar o desempenho das redes em tempo real e extrair os benefícios que as novas tecnologias oferecem. Algoritmos de IA/ML são ferramentas fundamentais para otimização e automação das redes de próxima geração.

Com sua maturidade gradual, técnicas de IA/ML serão introduzidas em vários cenários para ajudar fabricantes e operadoras a alcançarem uma transformação inteligente na operação e manutenção das redes e atingir os níveis de automação e otimização necessários para a realização de todo o potencial do 5G. Além disso, técnicas de IA/ML criarão oportunidades interessantes para todo o setor de telecomunicações, pois elas poderão ser utilizadas para criar abordagens mais pessoais para os clientes, ao mesmo tempo que ajudarão a gerenciar os custos de implantação e manutenção de redes. Portanto, a hora de começar a planejar a adoção de tais técnicas é agora.

5. ASPECTOS DE SEGURANÇA NAS REDES 5G

AS REDES 5G TERÃO QUE ENFRENTAR OS MESMOS PROBLEMAS DE SEGURANÇA ASSOCIADOS ÀS GERAÇÕES ANTERIORES E, ALÉM DISSO, TAMBÉM TERÃO QUE ENFRENTAR NOVOS DESAFIOS DE SEGURANÇA PRINCIPALMENTE DEVIDO À EXPECTATIVA DE UMA MASSIVA CONEXÃO DE NOVOS DISPOSITIVOS NO SISTEMA E, TAMBÉM, DEVIDO AO USO DE AMBIENTES VIRTUALIZADOS NAS ESTRUTURAS DAS REDES.

Tradicionalmente, as redes de comunicações móveis apresentaram algum tipo de risco para a segurança cibernética. Todas as tecnologias das várias gerações das comunicações móveis apresentaram ou apresentam alguma vulnerabilidade que pode ser explorada com intuito de causar danos à segurança. Uma maior preocupação com relação à introdução de mecanismos de segurança nessas redes fica evidente, mais especificamente a partir das redes 3G, devido ao aumento efetivo de conexões à Internet feitas por meio de ***smartphones***. Os sistemas 4G enfrentam basicamente os mesmos problemas de segurança presentes nas redes tradicionais de computadores aplicadas em ambientes de tecnologia da informação. Pode-se destacar vulnerabilidades relacionadas a ataques de negação de serviços (do inglês, ***Denial of Service*** - DoS), ataques de falsificação de endereço de Protocolo de Internet (do inglês, ***Internet Protocol*** - IP), roubo de identificações de usuários, roubo de serviços, ataques de intrusão em dispositivos, entre outros problemas [23].

TODAS AS TECNOLOGIAS DAS VÁRIAS GERAÇÕES DAS COMUNICAÇÕES MÓVEIS APRESENTARAM OU APRESENTAM ALGUMA VULNERABILIDADE QUE PODE SER EXPLORADA COM INTUITO DE CAUSAR DANOS À SEGURANÇA.

As redes 5G terão que enfrentar os mesmos problemas de segurança associados às gerações anteriores e, além disso, também terão que enfrentar novos desafios de segurança principalmente devido à expectativa de uma massiva conexão de novos dispositivos no sistema e, também, devido ao uso de ambientes virtualizados nas estruturas das redes. Nas redes 5G alguns aspectos de segurança já estão mapeados e sendo delineados no âmbito de alguns órgãos de padronização internacionais. Os principais aspectos de segurança listados por esses órgãos são: i) confidencialidade e integridade dos dados, ii) autenticidade, iii) políticas de segurança e iv) disponibilidade da rede [74]. Com relação à confidencialidade e integridade, é importante notar que nas redes 5G a criptografia é obrigatória para o tráfego de dados, garantindo a proteção dos dados e evitando ataques de espionagem por pessoas não autorizadas. Além disso, há uma maior preocupação quanto à garantia da integridade dos dados em todos os domínios da rede, seja durante o envio, armazenamento e processamento desses dados. Com relação à questão da autenticidade, é importante mencionar que o 5G reforça os mecanismos de autenticação mútua para evitar que usuários se conectem em redes falsas e, também, para que redes autênticas possam verificar as identidades dos usuários a elas conectadas. A respeito das políticas de segurança centralizada, é importante salientar que as operadoras de redes móveis (do inglês, ***Mobile Network Operators*** - MNOs) poderão se beneficiar das tecnologias de virtualização e, portanto, muitos de seus serviços poderão estar fora de seus domínios e alocados, por exemplo, em nuvens públicas. Isso caracteriza uma grande mudança em relação aos controles de segurança que as MNOs precisarão implementar. Por fim, a disponibilidade está relacionada ao fato de

garantir que os recursos de rede estejam acessíveis sempre que solicitados, e que deverão estar disponíveis para usuários legítimos sem comprometer a eficácia da rede. A disponibilidade também será um fator de medição da sustentabilidade da rede contra os ataques já conhecidos nas redes de computadores, como por exemplo, ataques DoS. A implementação de mecanismos de segurança que atendam esses requisitos básicos mencionados acima fará com que as redes 5G ofereçam ao usuário uma maior proteção do que as redes legadas existentes. Essa subseção tem como objetivo apresentar os principais mecanismos relacionados à segurança cibernética que são padronizados pelo 3GPP para as redes 5G.

O padrão 3GPP define duas diferenças entre implementações das redes 5G que são importantes do ponto de vista da segurança cibernética. Na primeira forma de implementação, denominada 5G ***Non-Standalone*** (5G-NSA), o núcleo de rede EPC é usado tanto para conexões 4G quanto para as conexões 5G. Já na forma conhecida como 5G ***Standalone*** (5G-SA) as conexões 4G são feitas usando o EPC enquanto que as conexões 5G são feitas usando um novo núcleo de rede denominado 5G ***Core*** (5GC) [75]. Nas implementações 5G-NSA, a rede 5G herda todas as vulnerabilidades já conhecidas do 4G, como por exemplo, i) rastreamento do ***International Mobile Subscribe Identity*** (IMSI), ii) ataque do homem no meio (do inglês, MITM - ***Man in the Middle***) e iii) problemas de integridade e confidencialidade quando usuários estão em ***roaming***. Já as implementações 5G-SA aperfeiçoam os mecanismos de segurança e resolvem ou mitigam as vulnerabilidades já conhecidas das redes legadas. Portanto, os problemas relacionados a rastreamento de IMSI e de falta de confidencialidade e integridade em processos de ***roaming*** são tratados nas versões 5G-SA. No entanto, aspectos como a introdução de ambientes virtualizados trazem consigo novos desafios às MNOs que operarem as redes 5G-SA. A **Figura 6** traz uma ilustração dos conceitos abordados acima e que serão melhores explorados a partir desse ponto.

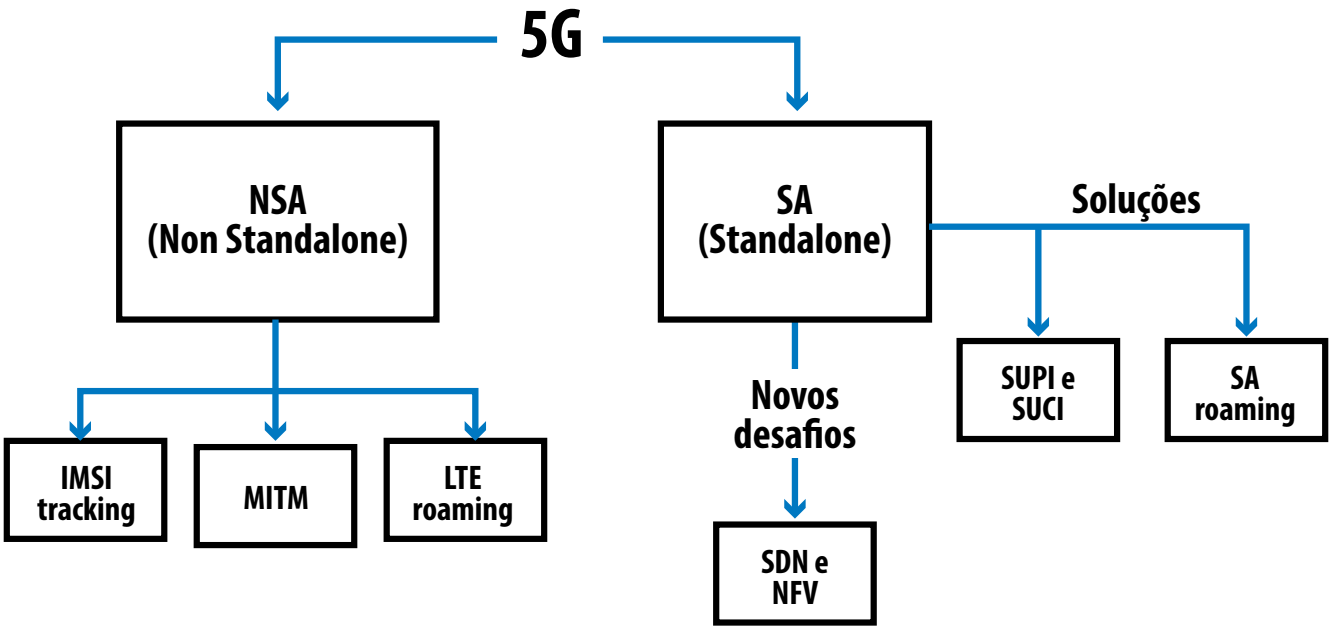

Figura 6: Principais diferenças relacionadas à segurança para infraestruturas 5G.

Nas redes móveis anteriores ao 5G-SA, o UE envia o IMSI em texto claro (sem criptografá-lo) durante alguns procedimentos de sinalização e autenticação. De posse de um Rádio Definido por ***Software*** (do inglês, ***Software Defined Radio*** - SDR), um terceiro malicioso poderá ter acesso ao IMSI no momento que esse é enviado à rede. Esse ataque é conhecido como rastreamento de IMSI e, por se tratar de uma identidade única atribuída internacionalmente a usuários da rede móvel, isso viola o princípio da privacidade, expondo, por exemplo, locais frequentados por determinados usuários. Além disso, nessas redes legadas, a integridade do tráfego do plano de usuário não é adequadamente protegida por padrão [74,76]. Isso abre brechas para que o fluxo de dados seja interceptado por ataques do tipo MITM. Nesses ataques, um terceiro malicioso pode manipular mensagens entre o móvel e a rede. Por fim, algumas interfaces do plano de controle das redes legadas utilizam protocolos de sinalização que são conhecidamente vulneráveis, como é o caso dos protocolos SS#7 e ***Diameter*** [77]. O SS#7 ainda é usado em algumas interfaces de interconexão de redes que não suportam o protocolo ***Session Initiation Protocol*** (SIP), enquanto que o ***Diameter*** é definido por padrão para funções de autenticação, autorização e ***Policy Charging and Control*** (PCC). Esses dois protocolos são, na maior parte, usados para estabelecer conexões de usuários em ***roaming*** e apresentam vulnerabilidade a ataques de espionagem. Nesses casos, terceiros maliciosos podem obter informações de detalhamento das chamadas de voz, mensagens de texto e, também, é possível fazer rastreamento do número de telefone de um usuário. As redes 5G-NSA irão herdar todos esses problemas das redes legadas. Porém, a padronização do 5G-SA possui mecanismos de segurança que resolvem ou mitigam essas vulnerabilidades.

No 5G-SA o problema do rastreamento de IMSI e a vulnerabilidade ao ataque MITM são resolvidos, pois, por padrão, todo o tráfego de dados na interface de rádio 5G é criptografado e sujeito a autenticação mútua. No 5G-SA o IMSI passa a ser denominado ***Subscriber Permanet Identifier*** (SUPI) e é criptografado e transmitido como ***Subscriber Concealed Identifier*** (SUCI) sempre que precisa ser trocado pela rede [63]. Novos processos de segurança relacionados ao ***Embed Subscriber Indentity Module*** (eSIM) substituem o cartão ***Subscriber Indentity Module*** (SIM) até então usado por usuários de redes legadas e trazem consigo novas características de segurança como, uso de infraestrutura de chave pública e certificados digitais para autenticação mútua. Com esses novos mecanismos, na rede 5G-SA não será mais possível rastrear o IMSI do UE usando as metodologias de ataque atuais [78].

As vulnerabilidades encontradas em processos de ***roaming*** também são tratadas pelo 3GPP na padronização do 5G-SA. Basicamente, é introduzido na rede um elemento denominado ***Security Edge Protection Proxy*** (SEPP) que protege a ***Home Public Land Mobile Network*** (HPLMN) nas interconexões com outras redes usadas durante o ***roaming***, como a ***Visited Public Land Mobile Network*** (VPLMN). O SEPP age como um ***proxy*** de segurança por onde são estabelecidas conexões criptografadas e autenticadas. Além disso, o 5G-SA substitui protocolos vulneráveis, como ***Diameter,*** por protocolos mais atuais e intrinsicamente seguros, como o ***Hypertext Transfer Protocol*** 2.0 (HTTP/2) e o ***Transport Layer Security*** (TLS). Sendo assim, esses novos mecanismos de segurança aumentam a proteção contra-ataques de espionagem e também conseguem evitar vários tipos de fraude de ***roaming***.

É IMPORTANTE TRABALHAR COM TÉCNICAS DE ISOLAMENTO DE FATIAS DE REDE PARA QUE UM DETERMINADO DADO QUE ESTEJA SENDO TRATADO POR UMA FATIA NÃO SOFRA NENHUM ATAQUE DE ESPIONAGEM OU ACESSO INDEVIDO POR OUTRAS FATIAS DE REDE QUE ESTEJAM COMPARTILHANDO A MESMA INFRAESTRUTURA.

A virtualização introduzida pelas redes 5G-SA traz consigo novos desafios relacionados à segurança cibernética. Nos ambientes virtualizados, diferentes VNFs podem compartilhar os mesmos recursos computacionais e, além disso, com a introdução do fatiamento de rede, diferentes fatias podem também estar dividindo os mesmos recursos da infraestrutura física da rede. Além disso, a rede 5G é projetada para ser amigável à computação em nuvem e, portanto, práticas seguras devem ser seguidas para garantir que dados não sejam vazados ou que outros usuários da nuvem não possam explorar a rede da MNO por meio da estrutura da nuvem. Esses compartilhamentos de recursos exigem uma maior preocupação com relação ao isolamento das informações compartilhadas. Para um maior isolamento entre as VNFs, questões como isolamento dentro de ambientes de sistemas operacionais, hipervisores e contêineres são pontos centrais da segurança. Além disso, é importante trabalhar com técnicas de isolamento de fatias de rede para que um determinado dado que esteja sendo tratado por uma fatia não sofra nenhum ataque de espionagem ou acesso indevido por outras fatias de rede que estejam compartilhando a mesma infraestrutura. O isolamento adequado permitirá a proteção da integridade e da confidencialidade dentro da rede [79]. Esses aspectos de segurança mais voltados às questões de isolamento em ambientes virtualizados são melhores explorados na próxima subseção.

5.1. ASPECTOS DE SEGURANÇA EM AMBIENTE VIRTUALIZADOS

As redes 5G serão, em muitos casos, constituídas a partir de ambientes virtualizados. VMs que estão instaladas em um mesmo ***hardware*** podem ser usadas para controlar diferentes serviços dentro desses ambientes. Nesses casos, o hipervisor é responsável pela gerência das VMs e pelo controle de acesso delas ao ***hardware***. Ataques de canal lateral podem acontecer nesses cenários e ignoram os controles de acesso e auditorias implementadas pelo hipervisor para violar o isolamento de recursos e, assim, obter algum tipo de informação acerca das VMs vizinhas.

Uma das vantagens de se trabalhar com ambientes virtualizados é a facilidade em se fazer um balanceamento de carga dos servidores envolvidos nos tratamentos de dados da rede. Por exemplo, se um determinado servidor está com uma carga de trabalho muito alta, pode-se aliviar essa carga migrando algumas de suas VMs para outros servidores que estejam em regime de trabalho mais leve. Com as atuais tecnologias, essa migração de VMs pode ser feita sem que haja impactos significativos na continuidade dos serviços prestados pelo sistema. Contudo, o procedimento de migração de VM deve ser feito de forma segura para evitar que atacantes consigam interceptar essas VMs e, consequentemente, ter acesso aos dados presentes dentro delas. Basicamente, deve-se prezar pela confidencialidade dos dados dentro das VMs e, também, pela integridade desses dados, a fim de evitar que alterações indevidas possam prejudicar o funcionamento dessas VMs ou até mesmo dos servidores que as receberão. Soluções como o uso do TLS podem disponibilizar mecanismos de criptografia, integridade e autenticação mútua para a migração segura da VM [80]. Porém, podem deixar a migração mais lenta e impactar na continuidade do serviço dependendo do tamanho da VM a ser transmitida. Outras tecnologias como ***Live Migration Defense Framework*** (LMDF) e Intel TxT oferecem serviços de segurança à migração, levando-se em consideração a continuidade de serviços [81,82].

Algumas outras preocupações devem ser levadas em consideração para a segurança em ambientes virtualizados. Configurações maliciosas feitas por administradores mal-intencionados ou mal capacitados podem permitir que determinadas VNFs consigam alterar seus privilégios e ter controles em funções específicas do hipervisor ou do ***hardware***. Assim, podendo desferir ataques de canal lateral ou de migração com maior facilidade. Falhas de ***software*** das VNFs podem levar o sistema a ter comportamentos indesejados, como, por exemplo, contornar ou travar funções de ***firewall*** ou causar estouro de ***buffer*** em algum ponto do sistema. Essas VNFs danificadas podem expor falhas do sistema para atacantes que irão explorar a rede a fim de obter vantagens ilícitas. Além disso, administradores da rede conseguem ter acesso aos hipervisores e executar operações de pesquisa para obtenção de senhas, chaves de conexão a aplicativos, dentre outros. Desta forma, o próprio administrador irá violar a privacidade dos usuários e a confidencialidade dos dados presentes na rede.

UM ATACANTE QUE CONSEGUIR ACESSO A UMA DETERMINADA FATIA DE REDE, PODERÁ EXECUTAR UM ATAQUE NO CONJUNTO DE FATIAS QUE FAZEM USO DO MESMO HARDWARE

Aspectos de segurança também devem ser observados quando existem fatias de rede sendo usadas para diferentes serviços dentro da mesma infraestrutura. Tanto os orquestradores, as plataformas de computação em nuvem e as aplicações envolvidas precisam de mecanismos de segurança que forneçam controles específicos para a correta criação de cada uma das fatias. Após a implantação, deve haver um monitoramento contínuo para verificar se os aspectos de segurança estão sendo mantidos durante todo o ciclo de vida da fatia de rede. As fatias devem ser criadas visando o isolamento do tráfego de dados entre si e, também, visando uma segregação física e lógica relacionada à infraestrutura. Ou seja, fatias que têm requisitos mais exigentes de segurança não devem ser alocadas compartilhando os mesmos elementos físicos e lógicos de fatias que possuem baixos requisitos de segurança. Por exemplo, uma fatia de rede usada para cirurgia remota deve considerar mecanismos de integridade e autenticação mútua constantemente para combater qualquer ameaça de interceptação da fatia, porém, uma fatia destinada a jogos não exigirá o mesmo nível de segurança.

O 3GPP especifica em seus padrões o uso de mecanismos como o TLS e o Oauth2 para criptografia, autorização e autenticação mútua entre entidades que gerenciam as fatias de rede [83]. Esses controles usados de maneira adequada reduzem o risco de vazamento de dados das fatias e podem conter o espalhamento de ameaças de códigos maliciosos. No entanto, vulnerabilidades conhecidas de alguns ***chipsets***, denominadas ***Spectre*** e ***Meltdown*** [84], podem quebrar o isolamento das fatias de redes dando acesso às informações para terceiros maliciosos. O ataque de canal lateral também tem de ser observado no contexto das fatias de rede, pois quando as fatias de rede compartilham um determinado ***hardware***, qualquer comprometimento de isolamento deste ***hardware*** irá comprometer a segurança das fatias envolvidas. Diante deste cenário, um atacante que conseguir acesso a uma determinada fatia de rede, poderá executar um ataque no conjunto de fatias que fazem uso do mesmo ***hardware*** [85]. Além disso, ataques à privacidade também são críticos, uma vez que provedores de infraestrutura ou fornecedores de VNFs podem se aproveitar de fatias de rede com menor nível de segurança para capturar informações dentro de fatias que compartilham um mesmo ponto na infraestrutura [86].

FATIAS QUE TÊM REQUISITOS MAIS EXIGENTES DE SEGURANÇA NÃO DEVEM SER ALOCADAS COMPARTILHANDO OS MESMOS ELEMENTOS FÍSICOS E LÓGICOS DE FATIAS QUE POSSUEM BAIXOS REQUISITOS DE SEGURANÇA. POR EXEMPLO, UMA FATIA DE REDE USADA PARA CIRURGIA REMOTA DEVE CONSIDERAR MECANISMOS DE INTEGRIDADE E AUTENTICAÇÃO MÚTUA CONSTANTEMENTE PARA COMBATER QUALQUER AMEAÇA DE INTERCEPTAÇÃO DA FATIA, PORÉM, UMA FATIA DESTINADA A JOGOS NÃO EXIGIRÁ O MESMO NÍVEL DE SEGURANÇA.

Em todas as camadas existentes na rede 5G, desde a camada mais inferior, a de infraestrutura, até a camada mais alta, a de virtualização de funções de rede, a segurança deve estar presente e ser um ponto focal. Mesmo com as orientações e as melhorias implementadas pelos órgãos de padronização, ainda há alguns desafios de segurança a serem superados dentro das redes 5G. Novos desafios também surgem quando a implementação da rede 5G segue os conceitos do OpenRAN. Esses pontos são melhor explorados na Subseção 6.3.

EM TODAS AS CAMADAS EXISTENTES NA REDE 5G, DESDE A CAMADA MAIS INFERIOR, A DE INFRAESTRUTURA, ATÉ A CAMADA MAIS ALTA, A DE VIRTUALIZAÇÃO DE FUNÇÕES DE REDE, A SEGURANÇA DEVE ESTAR PRESENTE E SER UM PONTO FOCAL.

6.
OPENRAN

HÁ UM MOVIMENTO DO ECOSSISTEMA DE REDES DE COMUNICAÇÕES MÓVEIS QUE OBJETIVA DEFINIR UMA ESTRUTURA FLEXÍVEL E INTEROPERÁVEL PARA AS REDES DE ACESSO, COM DESAGREGAÇÃO ENTRE O HARDWARE E O SOFTWARE DA REDE E O USO DE INTERFACES ABERTAS. ESTE MOVIMENTO DENOMINA-SE OPENRAN

Como mencionado na Seção 1 deste documento, há um movimento do ecossistema de redes de comunicações móveis que objetiva definir uma estrutura flexível e interoperável para as redes de acesso, com desagregação entre o ***hardware*** e o ***software*** da rede e o uso de interfaces abertas. Este movimento denomina-se OpenRAN e será detalhado nesta seção, incluindo a arquitetura O-RAN e aspectos relacionados à segurança e ao uso de inteligência artificial, bem como suas vantagens e desvantagens.

6.1. ARQUITETURA OPENRAN

A arquitetura OpenRAN se apoia em diversos ingredientes provenientes de avanços nas áreas de Tecnologia da Informação e Comunicações (TIC) e comunicações móveis (padronizados ou não), tais como, C-RAN, virtualização de funções de rede, redes programáveis, orientação a serviços (do inglês, ***Everything as a Service*** - XaaS), inteligência artificial, MEC, ***big data***, dentre muitos outros.

Dentre as arquiteturas de OpenRAN, temos a arquitetura definida pela O-RAN ***Alliance,*** aqui denominada O-RAN, que é tratada com mais detalhe neste documento, e as arquiteturas NG-RAN (***Next Generation Radio Access Network***) do ETSI [87], SD-RAN (***Software Defined*** RAN) da ONF [88,89] e OpenRAN do TIP [90]. O documento TS 38.401 [87] descreve a arquitetura geral da NG-RAN, incluindo interfaces 3GPP de nova geração (do inglês, ***New Generation*** - NG), interfaces do tipo Xn e F1, bem como interação dessas com a interface de rádio.

A ARQUITETURA OPENRAN SE APOIA EM DIVERSOS INGREDIENTES PROVENIENTES DE AVANÇOS NAS ÁREAS DE TECNOLOGIA DA INFORMAÇÃO E COMUNICAÇÕES (TIC) E COMUNICAÇÕES MÓVEIS (PADRONIZADOS OU NÃO), TAIS COMO, CLOUD-RAN, VIRTUALIZAÇÃO DE FUNÇÕES DE REDE, REDES PROGRAMÁVEIS, ORIENTAÇÃO A SERVIÇOS (DO INGLÊS, EVERYTHING AS A SERVICE - XAAS), INTELIGÊNCIA ARTIFICIAL, MEC, BIG DATA, DENTRE MUITOS OUTROS.

OS PRINCÍPIOS QUE GUIARAM O DESENVOLVIMENTO DA ARQUITETURA O-RAN FORAM:

- **LIDERAR A INDÚSTRIA NA DIREÇÃO DE RAN ABERTA, COM INTERFACES INTEROPERÁVEIS, BASEADA EM VIRTUALIZAÇÃO, COM SUPORTE DE BIG DATA E IA.**
- **MAXIMIZAR O USO DE HARDWARE COTS MINIMIZANDO EQUIPAMENTOS PROPRIETÁRIOS.**
- **ESPECIFICAR INTERFACES E APIS (APPLICATION PROGRAMMING INTERFACES) APROPRIADAS.**
- **EXPLORAR OPEN SOURCE ONDE FOR ADEQUADO.**

Já a arquitetura OpenRAN do TIP [87] visa definir e construir soluções de rede de acesso de rádio para 2G até 5G, com neutralidade de fornecedor, baseadas em ***hardware*** de propósito geral e com interfaces abertas. O objetivo é criar produtos e soluções interoperáveis a partir de múltiplos fornecedores. Dentre os pilares dessa arquitetura OpenRAN TIP tem-se a desagregação de funções, o uso de interfaces abertas, diversas opções de divisão funcional na RAN, flexibilidade multifornecedor, uso de soluções de virtualização ***bare metal*** ou contêineres e adoção de IA/ML.

A arquitetura SD-RAN da ONF [87,88] visa construir componentes ***open source*** para o ecossistema de RAN, complementando a proposta O-RAN e testando componentes em campo. O foco é criar soluções RAN multifornecedor, fomentando a inovação no ambiente de RAN. Esta iniciativa está desenvolvendo componentes nativos em ***data center***, apoiados sobre o controlador SDN ONOS (***Open Network Operating System***). De acordo com [89], SD-RAN segue a proposta de O-RAN, enriquecendo-a, alimentando-a com componentes ***open source*** e ampliando a sua visão.

A especificação da arquitetura O-RAN foi feita pelo Grupo de Trabalho 1, sendo a especificação 2.0 (de julho de 2020) a mais recente no ato da escrita deste documento [90]. Essa especificação (O-RAN.WG1.O-RAN-***Architecture-Description***-v02.00) se baseia em especificações técnicas e relatórios do 3GPP, bem como em outros documentos da aliança O-RAN. Vale ressaltar que a arquitetura O-RAN deve ser consistente na medida do possível com a arquitetura 3GPP e suas interfaces. Segundo a especificação 2.0 da O-RAN, os princípios que guiaram o desenvolvimento da arquitetura O-RAN foram:

- Liderar a indústria na direção de RAN aberta, com interfaces interoperáveis, baseada em virtualização, com suporte de *big data* e IA.
- Maximizar o uso de *hardware* COTS minimizando equipamentos proprietários.
- Especificar interfaces e APIs (*Application Programming Interfaces*) apropriadas.
- Explorar *open source* onde for adequado.

A **Figura 7** ilustra a arquitetura O-RAN para o caso de implantação mais típico [90], conciliando a visão arquitetural da especificação 2.0 com o contexto de fatiamento de recursos de rede via NFV e SDN. Os componentes da arquitetura são geograficamente distribuídos em diversas regiões, que vão desde nacionais/internacionais, até regionais, de borda e locais.

No domínio local encontram-se os UEs. No ***site*** local (onde ficam as células de comunicação) tem-se o componente ***Radio Unit*** (O-RU). No ***data center*** de borda do domínio local (O-***Cloud*** de Borda, quando se utiliza virtualização) tem-se o componente O-DU (***Open Distributed Unit***). No ***data center*** regional ficam as funções de rede O-RAN, nomeadamente Near-RT RIC (***Near Real-Time*** RAN ***Intelligent Controller***), O-DU e O-CU (***Open Centralized Unit***). Elas podem ser implementadas como VNFs dentro do conceito de NFV e/ou em ***hardware*** otimizado. No caso de implementação em ***hardware*** especializado, são chamadas de PNFs. O mesmo vale para as O-DUs, conforme o documento [91].

Segundo [90], o caso mais típico de implantação dessas funções seria na forma de VNFs. As funções ***Near***-RT RIC e O-CU são implementadas em diversas O-***Clouds*** regionais, enquanto a função O-DU é implementada em O-***Clouds*** de borda. Para cada O-***Cloud*** regional esperam-se várias instâncias de O-***Clouds*** de borda. Uma O-***Cloud*** é uma infraestrutura de TIC (no ***data center***) que segue o padrão O-RAN, compatível com o padrão NFV ETSI. Portanto, a infraestrutura física de TIC (servidores, discos, comutadores, etc.) é virtualizada pelo componente NFVI visando suportar assim as VNFs de O-RAN. Isso permite que as funções de rede relativas à estação rádio base O-eNB (Open ***Evolved Node*** B) também sejam virtualizadas na O-***Cloud*** regional. O controlador ***Near***-RT possui várias instâncias de aplicações (xApps) de terceiros, que complementam as suas funções.

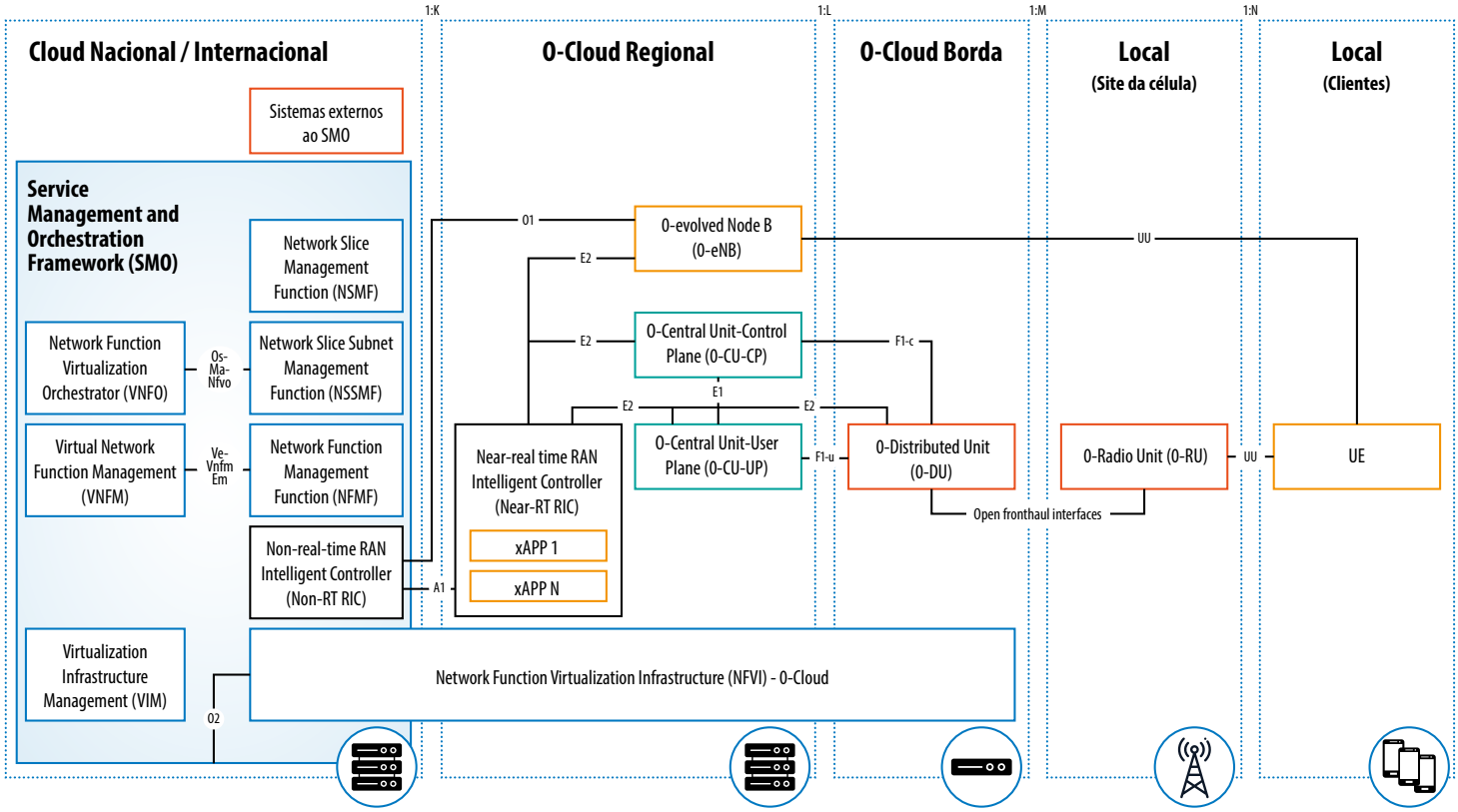

Figura 7: Ilustração da arquitetura O-RAN.

No lado da ***cloud*** nacional/internacional, existem outros componentes das arquiteturas NFV e O-RAN. O NFVI também suporta as funções de O-RAN voltadas para gerenciamento e controle, dentre elas o controlador ***Non***-RT RIC (***Non Real-Time*** RAN ***Intelligent Controller***), que é implementado como uma VNF. Embora não esteja mostrado na **Figura 7**, controladores SDN (padrão ***OpenFlow*** da ONF [87]) também podem estar presentes na ***cloud*** nacional ou regional, o que é preferível devido ao atraso no plano de controle.

Para gerenciar a NFVI tem-se o VIM. Ainda com relação à arquitetura ETSI NFV, existem os componentes VNFM e VNFO (***Virtual Network Function Orquestrator***). Todos esses componentes podem fazer parte de um componente O-RAN chamado SMO ***Framework***, embora alguns deles possam também ser executados fora desse componente. A arquitetura O-RAN prevê ainda outros sistemas externos ao SMO, como por exemplo o OSS e o BSS.

O ***Non***-RT RIC é responsável por juntar dados de longo prazo relacionados a fatias de rede por meio da interação com o SMO e aplicar IA/ML no ***Near***-RT RIC visando prover casos de uso inovadores na RAN. Esse componente deve estar ciente das fatias existentes e seus respectivos SLAs (***Service Level Agreements***) por meio do SMO. Também obtém métricas específicas, parâmetros de configuração e atributos de fatias da infraestrutura via SMO. Pode ainda obter informações enriquecidas de terceiros para aperfeiçoar o fatiamento de recursos. Constrói modelos de IA/ML a serem implantados no ***Near***-RT RIC. Efetiva otimizações quase em tempo real para alimentar parâmetros específicos do ***Near***-RT RIC, O-CU, O-DU e O-RU via interface O1 do SMO (passando por ele). Dados obtidos são usados para guiar políticas e assistir o Near-RT RIC via interface A1 criando um laço fechado.

O ***Near***-RT RIC é responsável pela execução de xApps de terceiros visando otimizar fatias de recursos que podem se estender desde o núcleo da rede até o terminal do usuário, cobrindo diversas funções de rede. xApps podem usar IA/ML ou outras técnicas de controle para aplicar políticas enviadas para esse componente pela interface A1. O ***Near***-RT RIC deve ter o conhecimento das fatias existentes e seus requisitos. Isso é informado pelo SMO via interface O1 durante o provisionamento. A configuração dos recursos das fatias nos nós acessíveis via interface E2 é feita por meio de dois ***loops***: lento (via O1) e rápido (via E2). O ***Near***-RT RIC pode usar o mecanismo rápido para fatiamento de recursos de rádio na RAN. Também executa monitoramento de desempenho da fatia no sentido reverso da E2.

O componente O-CU implementa a parte alta da pilha de protocolos da RAN. Deve suportar os recursos de fatiamento previstos pelo 3GPP na RAN. Dependendo dos requisitos de cada fatia, O-CU-UP pode ser compartilhado por várias fatias simultaneamente. O-RAN aprimora o fatiamento provido pelo 3GPP por meio do uso da interface E2 e da aplicação de técnicas de otimização quase em tempo real via a interface O1. O objetivo é suportar parâmetros adicionais de configuração. Isso permite que as camadas superiores da pilha RAN sejam cientes da fatia que estão atendendo, utilizando técnicas específicas de alocação e isolamento. Essas pilhas são inicialmente configuradas via interface O1 e depois atualizadas pelo ***Near***-RT RIC via E2. O O-CU responde a solicitações de monitoramento de desempenho e garantia de SLA (para fatias instaladas na infraestrutura) quando solicitado pelo SMO e pelo ***Near***-RT RIC, via interfaces O1 e E2, respectivamente.

Já o componente O-DU implementa a parte baixa da pilha de protocolos da RAN. Ele também deve suportar estratégias de alocação de recursos por fatia, incluindo a camada MAC. A MAC 5G aloca e isola blocos de recursos físicos (do inglês, ***Physical Resource Blocks*** - PRBs) para as fatias. Os níveis de alocação de recursos são configurados a partir de: i) configuração estática advinda da interface O1; ii) diretivas O-CU recebidas via interface F1; e iii) guiamento dinâmico recebido via E2. Baseado em pedidos de gerência de desempenho enviados pelo SMO e ***Near***-RT RIC, o O-DU pode gerar e enviar objetos de monitoramento de desempenho (do inglês, ***Performance Monitoring Objects*** - PMOs) através das interfaces O1 e E2, respectivamente.

Por fim, o componente O-RU implementa o ***front end*** de rádio que é hospedado em ***site*** proprietário. A O-RU pode utilizar tecnologia de SDR, que permite alterar configurações e até mesmo funcionalidades do rádio através de ***software*** de controle. Em [92], algumas opções de SDR para O-RU são exploradas, como por exemplo ***bladeRF*** e USRP X310. Ainda segundo [92], os desenvolvimentos correntes em O-RAN não incluem ***software open source*** para os ***front ends*** de RF.

A inter-relação entre NFV ETSI e O-RAN é intensa. A arquitetura suporta componentes voltados para o fatiamento da porção de RAN (também como VNFs), nomeadamente: CSMF (***Communication Service Management Function***), NFMF (***Network Function Management Function***), NSSMF (***Network Slice Subnet Management Function***) e NSMF (***Network Slice Management Function***) [54,93]. A função CSMF traduz o requisito relacionado ao serviço de comunicação em requisitos relacionados às fatias. A função NSMF é responsável pelo gerenciamento e orquestração das instâncias de fatias (em inglês, ***Network Slice Instances*** - NSIs). Já a função NSSMF é responsável pelo gerenciamento e orquestração das sub-redes de uma fatia (em inglês, ***Network Slice Subnet Instances*** - NSSIs). Por fim, a função NFMF gerencia uma ou mais VNFs, provendo exposição e governança de dados.

Quatro tipos de fatias são definidos para cada tipo de serviço 5G, respectivamente: eMBB, URLLC, MIoT (***Massive Internet of Things***) e V2X. De acordo com o documento 3GPP TR 28.801 [94], uma fatia de rede pode conter uma ou mais sub-redes e cada uma delas inclui uma ou mais funções de rede (físicas ou virtuais). O número máximo de fatias simultâneas é de 8 por UE. Entretanto, uma única conexão de sinalização é usada independentemente do número de fatias. O documento [54] apresenta os procedimentos para a instanciação de uma subrede de uma fatia. O ciclo de vida de uma fatia se inicia pela função NSMF, que requisita para a função NSSMF a criação de uma subrede na RAN. A função NSSMF, por sua vez, solicita a instanciação de VNFs (O-CU e O-DU) e configuração de PNF (O-RU) para essa fatia via orquestrador de NFV do ETSI (NFVO). Na sequência, o NFVO aciona a função VNFM para efetivar a criação das VNFs sobre a NFVI. Ao final da instanciação da O-CU e O-DU, a VNFM notifica a função NFMF; também notifica o orquestrador NFVO, que por sua vez avisa a função NSSMF. Por fim, uma notificação é enviada da função NSSMF para a função NSMF completando a instanciação das funções relativas a uma fatia na RAN. Esse processo pode fazer parte de um processo maior que envolve também funções de núcleo da rede 5G.

A O-RAN permite diversidade de implementações para cada componente da arquitetura. A interoperabilidade entre as funções é suportada por interfaces padronizadas na O-RAN. A interface A1 é a interface entre ***Non***-RT RIC e ***Near***-RT RIC, ou seja, uma interface entre os controladores. Ela suporta gerenciamento de políticas, de modelos de IA/ML, e de serviços de enriquecimento (contextualização) de informações. Essa interface é utilizada para enviar políticas específicas que guiam alocações e controle em fatias (por fatia) e receber ***feedback*** específico sobre as políticas implantadas. É utilizada para troca de informações enriquecidas da RAN com atores externos. As políticas trocadas pela A1 não são persistentes (perdidas em caso de reiniciar o ***Near***-RT RIC). Elas têm precedência em relação às recebidas pela interface O1, mas não devem desviar significativamente dessas.

A ARQUITETURA O-RAN PERMITE GRANDE FLEXIBILIDADE DE ATORES, IMPLEMENTAÇÕES E IMPLANTAÇÕES. A RECONFIGURABILIDADE DOS COMPONENTES EM TEMPO DE OPERAÇÃO É UMA DAS VANTAGENS, ENDEREÇANDO UMA DAS CRÍTICAS COMUMENTE FEITAS ÀS SOLUÇÕES TRADICIONAIS BASEADAS EM CAIXAS PRETAS DE UM ÚNICO FORNECEDOR

A interface E2 é a interface entre ***Near***-RT RIC (controlador) e Nós Conectados via E2 (Nós E2, os controlados). Ela suporta as primitivas: ***Report***, ***Insert***, ***Control*** e ***Policy*** para controlar dispositivos expostos através de nós E2. Essas primitivas também são usadas por xApps para configurar fatias nos nós E2, incluindo configurações e alocações de recursos de rádio, políticas de escalonamento, e outros parâmetros da RAN. Ela também é usada para configurar e receber relatórios por fatia e dados de desempenho dos nós E2. Inclui PM (***Performance Management***) do 3GPP: utilização de PRBs, atraso médio, etc.

A interface entre o gerente e os elementos gerenciados na O-RAN é a O1. Ela é usada para configurar parâmetros específicos de fatia em nós O-RAN com base nos requisitos de serviço de cada fatia e métricas específicas de monitoramento de desempenho e falhas de nós O-RAN. Essa interface segue a TS 28.541 [95], incluindo atributos da política de gerência de recursos de rádio (em inglês, ***Radio Resource Management*** - RRM) para dividir PRBs entre fatias. Estende modelos 3GPP para definir perfis e configurações específicas de fatias.

A interface entre o SMO e a O-Cloud é a interface O2. Ela é usada para gerência do ciclo de vida de funções virtuais de rede. Ou seja, a instanciação, configuração, operação e encerramento de VNFs. O componente NSSMF cria, provisiona, modifica e deleta NSSIs através de interação com o SMO. NSSIs são sub-redes que atendem uma fatia em parte do trajeto ou em todo ele. Para tanto, o NSSMF dispara a instanciação das funções de rede O-RAN de acordo com os requisitos das fatias.

A arquitetura O-RAN permite grande flexibilidade de atores, implementações e implantações. A reconfigurabilidade dos componentes em tempo de operação é uma das vantagens, endereçando uma das críticas comumente feitas às soluções tradicionais baseadas em caixas pretas de um único fornecedor [92]. A demanda é pela programabilidade total de granularidade fina, pela adaptabilidade dos componentes, pela decomposição flexível e ajustável de funcionalidades, pela agilidade das ações. Tudo isso visando atender requisitos conflitantes, otimização de funcionalidades, alta qualidade de serviços, oscilações de demanda e uso compartilhado e eficiente de recursos na RAN.

EMBORA AS INTERFACES SEJAM ABERTAS E PADRONIZADAS, DIFERENÇAS DE INTERPRETAÇÃO DE FUNCIONALIDADES PODEM LEVAR A INTERAÇÕES NÃO CLARAS ENTRE COMPONENTES DE DIFERENTES ORIGENS.

Entretanto, embora as interfaces sejam abertas e padronizadas, diferenças de interpretação de funcionalidades podem levar a interações não claras entre componentes de diferentes origens. Em outras palavras, mesmo usando uma interface padrão, existe a dependência da interoperabilidade de dados que trafegam pelas interfaces e implementações de funções em componentes. É preciso que cada componente implemente de forma correta e completa o que é previsto em cada posição, independentemente de quem está realizando tal desenvolvimento. Nesse contexto, a arquitetura O-RAN precisará evoluir mecanismos que facilitem a verificação de escopo, interoperabilidade, adequabilidade e funcionamento conjunto de componentes. A avaliação adequada de desempenho de funções físicas e virtuais é outra demanda em aberto. Como avaliar de forma apropriada e reproduzível o desempenho de VNFs que implementam componentes da arquitetura O-RAN?

A ARQUITETURA O-RAN PRECISARÁ EVOLUIR MECANISMOS QUE FACILITEM A VERIFICAÇÃO DE ESCOPO, INTEROPERABILIDADE, ADEQUABILIDADE E FUNCIONAMENTO CONJUNTO DE COMPONENTES.

Uma primeira resposta a essa pergunta é dada em [96]. Os autores desse Internet ***draft*** descrevem uma metodologia automatizada para avaliar o desempenho de VNFs. O objetivo é permitir a verificação de alta fidelidade do desempenho de VNFs de acordo com variações de configuração interna, de infraestrutura, de ambiente de execução, de conectividade e de carga de trabalho. Segundo os autores, a avaliação de desempenho manual é ineficaz (devido à grande quantidade de configurações possíveis) e custosa (demanda incontáveis horas de trabalho). Portanto soluções automatizadas, possivelmente com uso de IA, serão demandadas. Tais soluções devem se apoiar em quatro princípios básicos de teste de VNFs: comparabilidade, reprodutibilidade, configurabilidade e interoperabilidade. No momento da escrita desse texto, o ***benchmarking*** em escala de funções da arquitetura O-RAN ainda é um problema em aberto. O mesmo ocorre com o desenvolvimento de metodologias ***DevOps*** ágeis focadas em desempenho de VNFs, incluindo as previstas em O-RAN e ETSI NFV. O documento também apresenta um ***software open source*** chamado de ***Gym***, que já foi utilizado para analisar o desempenho de um comutador virtual chamado OVS (***Open vSwitch***), bem como componentes de IMS (IP ***Multimedia Subsystem***).

Outro aspecto importante da arquitetura O-RAN, que deriva do uso de NFV e SDN, é a elasticidade da solução para atender o aumento ou a redução da demanda de serviços na RAN. Conforme [97], a maior parte das soluções de C-RAN possui apenas elasticidade vertical na alocação de recursos para processamento de sinais em BBU. Isso significa que as soluções escalam via criação de novas VMs em um único servidor. Utilizam ainda uma granularidade fixa para as ações de elasticidade. Em [97], é proposta a solução ***Elastic***-RAN para o problema de processamento de sinais de banda base. ***Elastic***-RAN possui solução de elasticidade multinível, isto é, tanto na vertical (acréscimo e decréscimo de VMs), quanto na horizontal (acréscimo e decréscimo de máquinas físicas). ***Elastic***-RAN utiliza ainda granularidade adaptativa, o que permite a tomada de decisão de elasticidade de forma dinâmica conforme a variação da carga de processamento nas VNFs. Isso levando em conta o gasto de CPU, memória, disco, conectividade e energia. Tal solução se aproxima mais da demanda de elasticidade que enxergamos para a O-***Cloud***, tirando assim maior proveito do uso das tecnologias NFV e SDN.

MUITOS ENXERGAM QUE UM MERCADO DE FUNÇÕES DE REDE O-RAN PODERÁ EMERGIR, SUPORTANDO EM UM EXTREMO SOLUÇÕES OPEN SOURCE GRATUITAS E EM OUTRO, SOLUÇÕES PAGAS DE ALTO CUSTO

Muitos enxergam que um mercado de funções de rede O-RAN poderá emergir, suportando em um extremo soluções ***open source*** gratuitas e em outro, soluções pagas de alto custo. Os requisitos de implantação e os clientes definirão quais são as melhores opções em cada posição. O objetivo que muitos enxergam é a migração de soluções inflexíveis, rígidas, monolíticas, opacas, não reconfiguráveis para soluções flexíveis, modulares, transparentes, reconfiguráveis (programáveis) e adaptáveis. O argumento é que as soluções tradicionais têm programabilidade e controlabilidade limitadas, o que dificulta adaptar e fatiar a rede para diferentes cenários de aplicação, requisitos funcionais e condições de tráfego [92]. As soluções tradicionais teriam gerência de recursos ineficiente, inabilidade para implementar conectividade programável e compartilhada. A granularidade e abrangência dos controles e gerência é uma demanda importante do 5G.

A ARQUITETURA O-RAN PERMITE O COMPARTILHAMENTO E A DIVERSIDADE DE RECURSOS DE INFRAESTRUTURA E SOFTWARE, ALAVANCANDO NOVOS MODELOS DE OPERAÇÃO, GERÊNCIA E NEGÓCIOS, PERMITINDO A MIGRAÇÃO DE SISTEMAS EM SILOS PROPRIETÁRIOS, PARA ECOSSISTEMAS MULTIFORNECEDOR, COM IMPLEMENTAÇÕES PROGRAMÁVEIS DE TODO TIPO DE ORIGEM (OPEN SOURCE, LICENCIADAS, PROPRIETÁRIAS, ETC.)

A arquitetura O-RAN permite o compartilhamento e a diversidade de recursos de infraestrutura e ***software***, alavancando novos modelos de operação, gerência e negócios, permitindo a migração de sistemas em silos proprietários, para ecossistemas multifornecedor, com implementações programáveis de todo tipo de origem (***open source***, licenciadas, proprietárias, etc.). Permite agilizar o ciclo de desenvolvimento de soluções, feito a muitas mãos entre clientes, fornecedores, prestadores de serviços, etc. [92]. A comunidade de pesquisa pode entrar no ***loop*** de desenvolvimento de soluções, utilizando ***setups*** do mundo real em suas avaliações. Operadores de rede podem interagir e controlar componentes simulando cenários operacionais junto de fornecedores, integradores e clientes finais.

A arquitetura O-RAN habilita a co-criação de soluções a muitas mãos, facilitando a evolução de soluções e a reprodutibilidade de testes. Entretanto, o desenvolvimento do conjunto de soluções por muitas pessoas, incluindo ***hardware*** e ***software***, ainda é um grande desafio para o atual modelo organizacional das instituições envolvidas. Atingir a maturidade, completude e robustez de soluções nesse contexto são desafios à parte. A arquitetura O-RAN visa justamente permitir a integração suave de componentes de novos participantes. Nesse contexto, emergem inúmeras questões relacionadas à segurança e privacidade, que serão exploradas na Seção 6.3 deste documento. Soma-se ainda o ***debug*** e o escrutínio de código dos componentes da O-RAN, incluindo xApps.

A criação de consórcios e ecossistemas facilita a obtenção de dados para desenvolvimento e treinamento de algoritmos inovadores, em especial dos algoritmos Non-RT RIC e Near-RT RIC. Tal abordagem ataca justamente a questão de como fornecedores isolados poderão ter acesso a dados de qualidade para treinar suas IAs sem a formação de consórcios. A arquitetura O-RAN foi desenhada para fornecer caminho para essa questão. De acordo com [92], a otimização tradicional e a configuração manual das redes se tornam impossíveis em 5G. O-RAN vai na direção de redes auto organizadas, tendo os controladores RIC como um pilar importante de implementação dessa ideia.

O controlador Near-RT RIC permite que xApps desenvolvidas por terceiros executem funções de controle e otimização da rede. A arquitetura permite que seja criado um mercado dinâmico comum dessas aplicações. Conforme [98], um operador pode contar com xApps de terceiros para controlar o processo de ***handover*** de usuários, fazer o balanceamento de carga, otimizar as políticas de escalonamento. Embora possível do ponto de vista técnico, essa abordagem esbarra na confiança necessária para delegar o controle de funções críticas de rede a ***software*** de terceiros. A aliança O-RAN estabeleceu uma comunidade de ***software*** junto com a Fundação Linux. A qualidade do ***software*** desenvolvido tanto para os componentes da arquitetura, quanto das xApps é um ponto crítico. Ainda, para tirar proveito desse modelo, as operadoras de telecomunicações precisarão modificar seu modo de operação, permitindo maior colaboração com desenvolvedores terceirizados.

Outro aspecto relevante é a relação da arquitetura O-RAN com SDN ou até mesmo com NG-SDN [88]. Muito se fala sobre essa relação, mas pouco se comenta sobre como de fato os controladores SDN (ou NG-SDN) se integram a tudo o que já discutimos. Tipicamente, a integração de SDN com NFV se dá via interface(s) norte nos controladores SDN. Ou seja, orquestradores NFV (ou outros componentes da arquitetura NFV) interagem com os controladores SDN, refletindo nesses as demandas de conectividade do grafo de serviços (conexões entre VNFs) que implementam para cada fatia de rede. Portanto, a configuração de encaminhamento de tráfego nos componentes da O-RAN poderia ser efetivada através de controladores SDN utilizando protocolo ***OpenFlow***. Em NG-SDN, tanto o ***firmware*** dos comutadores físicos ou virtuais, quanto o protocolo controlador-controlado podem ser implementados conforme demanda da arquitetura.

6.2. INTELIGÊNCIA ARTIFICIAL NO OPENRAN

O PADRÃO 5G EXIGE NOVOS NÍVEIS DE FLEXIBILIDADE NA ARQUITETURA, DIMENSIONAMENTO E IMPLANTAÇÃO DE REDES DE TELECOMUNICAÇÕES. A TECNOLOGIA DE COMPUTAÇÃO EM NUVEM OFERECE ALTERNATIVAS INOVADORAS PARA FLEXIBILIZAR A ARQUITETURA E FACILITAR AS IMPLANTAÇÕES DAS REDES 5G, ESPECIALMENTE PARA AS RANS, COMPLEMENTANDO AS SOLUÇÕES JÁ EXISTENTES E COMPROVADAS (ISTO É, TESTADAS).

O padrão 5G exige novos níveis de flexibilidade na arquitetura, dimensionamento e implantação de redes de telecomunicações. A tecnologia de computação em nuvem oferece alternativas inovadoras para flexibilizar a arquitetura e facilitar as implantações das redes 5G, especialmente para as RANs, complementando as soluções já existentes e comprovadas (isto é, testadas).

A RAN é a parte de um sistema de telecomunicações que conecta dispositivos a outras partes de uma rede por meio de conexões de rádio (i.e., sem fio). A RAN reside entre o equipamento do usuário, como um telefone celular, um computador ou qualquer máquina ou dispositivo controlado remotamente, e fornece a conexão com o ***core*** da rede. A RAN fornece acesso e coordena o gerenciamento de recursos de rádio. Os componentes da RAN incluem a estação rádio base, a qual pode ser dividida em unidades de processamento de banda base e de radiofrequência, antenas e um controlador RAN.

Como já apresentado, OpenRAN é um termo da indústria para uma arquitetura de RAN aberta com interfaces interoperáveis também abertas e desagregação de ***software*** e ***hardware***, que permitem abordagens de ***big data*** e inovações baseadas em técnicas de IA/ML no domínio da RAN. Para realizar isso em larga escala, aproveita-se das tecnologias de computação em nuvem e dos paradigmas de escalonamento trazidos do conceito de XaaS, onde tudo no sistema 5G, por exemplo, pode ser tratado como um serviço, que nos levam ao C-RAN. C-RAN refere-se à realização de funções da RAN em uma plataforma de computação genérica em vez de uma plataforma de ***hardware*** construída para um propósito específico; centralização do processamento de banda base; e ao gerenciamento da virtualização de serviços de rede de acesso usando princípios nativos da computação em nuvem [99].

As tecnologias de IA/ML e computação em nuvem estão sendo usadas também no ***core*** das redes de telecomunicações e, desta forma, é um processo natural que a RAN evolua e se integre ao resto da infraestrutura da rede. Desta forma, usando os mesmos ingredientes (IA/ML e ***Cloud***), tanto no ***core*** da rede quanto na RAN, criam-se ambientes homogêneos em todas as regiões da rede, trazendo várias vantagens ao desenvolvimento, integração e implantação do 5G.

Portanto, essas três tecnologias se complementam de forma a criar redes virtualizadas, abertas, interoperáveis, flexíveis e escalonáveis que facilitarão a automação, gerenciamento, otimização e o uso de IA/ML de forma mais ampla e transparente, além de atender à grande demanda por dados que a próxima geração de redes móveis sem fio espera.

Nesta subseção discute-se o emprego de IA na filosofia de redes abertas e virtualizadas introduzidas pela O-RAN.

6.2.1. UM PARADIGMA PARA HABILITAR A APLICAÇÃO DE IA NAS REDES DE TELECOMUNICAÇÕES

As operadoras de telecomunicações têm adotado técnicas de IA/ML para criar redes auto-configuráveis, auto-gerenciáveis, auto-curáveis. Entretanto, a integração de IA/ML em implantações reais apresenta vários desafios relacionados com a arquitetura e procedimentos adotados pelas redes. Isso ocorre principalmente porque dados da rede em tempo real precisam ser coletados e agregados para permitir as operações de treinamento e inferência. Portanto, necessita-se de uma arquitetura que apresente formas para coletar e agregar os dados gerados pelas redes em tempo real para que modelos de IA/ML possam operar corretamente.

Como apresentado na Subseção 6.1, a arquitetura O-RAN prevê a função RIC (***RAN Intelligent Controller***) com interfaces com gNBs (***gNodeBs***) e/ou eNBs (***eNodeBs***), para monitorar, aprender e realizar atuações em malha-fechada, ou seja, com realimentação. Esse controlador inteligente é uma das funções principais introduzidas na arquitetura O-RAN e têm como objetivo aprimorar as funções de rede tradicionais através da incorporação de inteligência. Desta forma, as redes de telecomunicações poderão aprender e otimizar seus processos de forma conjunta e contínua [100,101,102].

O RIC interage com as funções de RAN existentes, onde algoritmos personalizados, habilitados pela adoção de técnicas de IA/ML, podem ser projetados e implementados como aplicações nativas em nuvem. Redes 5G com essa camada de inteligência adicionada ao RIC devem ter um desempenho muito melhor e reagir de forma mais dinâmica às mudanças da rede. O RIC é dividido em duas outras funções, ***Non***-RT RIC e ***Near***-RT RIC.

A função ***Non***-RT RIC faz parte da camada de orquestração e gerenciamento e tem como tarefa recuperar as configurações necessárias, indicadores de desempenho, relatórios de medição e outros dados do gerenciamento de configuração e desempenho com a finalidade de construir/treinar modelos IA/ML. Além disso, outras informações da camada de aplicação, por exemplo, localização GPS (***Global Positioning System***) dos usuários e previsões de padrão de mobilidade podem ser usadas para enriquecer o modelo. Em seguida, o modelo IA/ML treinado é transferido para o ***Near***-RT RIC e pode ser usado para inferir a melhor distribuição das células, prever a configuração e alocação ideal de recursos de rádio e a configuração de parâmetros para sistemas MIMO massivo para cada célula de acordo com um objetivo de otimização global definido pelo operador da rede [103,104]. A O-RAN ***Alliance*** tem se esforçado para orientar a indústria em direção ao desenvolvimento de RICs habilitados para IA/ML [27].

6.2.2. ETAPAS PARA O USO DE SOLUÇÕES ASSISTIDAS POR IA/ML NO OPENRAN

Com base nos requisitos da O-RAN ***Alliance*** [105], existem várias etapas que precisam ser realizadas por qualquer solução assistida por IA/ML. Estas etapas são descritas a seguir.

1. A primeira etapa é a *consulta de capacidades e propriedades do modelo*, que é realizada pela camada SMO quando o modelo for executado pela primeira vez ou atualizado. Essas capacidades e propriedades incluem: *i)* poder de processamento de *hardware* onde o modelo de IA/ML será executado (por exemplo, recursos como CPU/GPU e memória que podem ser alocados para treinamento e inferência do modelo); ii) propriedades como formatos de modelo e mecanismos de IA/ML suportados, por exemplo, Protobuf, JSON (*JavaScript Object Notation*) ou quaisquer formatos de dados VES (*Virtual Event Streaming*) específicos do ONAP; e iii) fontes de dados disponíveis para execução do *pipeline*[2] de IA/ML (por exemplo, suporte para fluxos de dados, *data lake*[3] ou qualquer acesso específico a banco de dados).

2 - O pipeline de IA/ML é o conjunto de funcionalidades, funções ou entidades funcionais específicas para uma solução assistida por IA/ML.

3 - Um *data lake* é um sistema ou repositório de dados armazenados em seu formato natural/bruto, geralmente como BLOBs (*Binary Large OBjects*), que são coleções de dados binários armazenados como uma única entidade em um sistema de gerenciamento de banco de dados.

2. A próxima etapa é a *seleção e o treinamento do modelo*, em que o *host* de treinamento do modelo de IA/ML seleciona e treina um modelo de IA/ML em relação a uma solução específica assistida por IA/ML (dependendo do caso de uso) a ser executada. Após o treinamento, o *host* de treinamento envia o modelo de volta ao *Non*-RT RIC no SMO para sua implantação. Depois que o modelo é treinado e validado, ele é publicado no catálogo de modelos do SMO.

3. Na etapa de *implantação e inferência do modelo de IA/ML*, o *host* de inferência é então configurado com o arquivo de descrição do modelo e os dados em tempo real são usados para inferência. O *host* de inferência é a função de rede que hospeda o modelo de IA/ML durante o modo de inferência (que inclui tanto a execução do modelo quanto qualquer aprendizado *online*, se aplicável). O resultado da inferência é enviado para o *Near*-RT RIC, de onde uma política é gerada para tomar ações corretivas, por exemplo.

4. Dependendo do resultado da inferência, as ações correspondentes são executadas usando os atores relacionados. Um ator é uma entidade que hospeda uma solução assistida por IA/ML usando a inferência do modelo de IA/ML para tomar decisões. Com base na localização do modelo de inferência, nos atores e nos tipos de ações possíveis, diferentes interfaces (i.e., O1, A1 e E2) são utilizadas.

5. Finalmente, ao *monitorar o desempenho do modelo*, o *host* de inferência envia o desempenho do modelo para o *host* de treinamento para fins de *reimplantação ou atualização do modelo*. Com base no *feedback* de avaliação de desempenho, alguma orientação pode ser fornecida para que se use um modelo diferente no *host* de inferência ou uma notificação pode ser enviada indicando a necessidade de se treinar novamente o modelo. O mecanismo de *feedback* e como a troca do modelo pode ocorrer é deixada para estudos posteriores.

6.2.3. CASOS DE USO

Como descrito na Seção 4, existem vários exemplos de casos de uso da utilização de modelos de IA/ML no 5G, como por exemplo o gerenciamento de ***handover*** dinâmico baseado em contexto para comunicação V2X, otimização de QoE e QoS dos usuários, alocação dinâmica de recursos para UAVs. Todos esses casos de uso são habilitados pelo paradigma do OpenRAN e demonstram a aplicabilidade prática da arquitetura O-RAN. Além dos principais casos de uso descritos na Seção 4, outros casos de uso possíveis e habilitados pelo paradigma do OpenRAN são listados na sequência [98,106,107,108].

- **Controle de escalonamento em redes 5G fatiadas**: Uma questão importante no fatiamento de recursos da rede é a política de escalonamento (*scheduling*), que aloca recursos limitados dinamicamente para usuários com vários requisitos de QoS de acordo com mudanças no perfil de tráfego e estado da rede. Portanto, neste caso de uso, um modelo IA/ML em execução em um xApp no *Near*-RT RIC seleciona uma política de escalonamento de melhor desempenho para cada fatia de recursos da RAN. Esta abordagem pode otimizar a alocação de recursos da rede de acordo com as demandas e dinâmica da própria rede.

- **Direcionamento de tráfego**: o controle de tráfego de redes tradicionais é frequentemente centrado na célula e, portanto, elas perdem a oportunidade de ajustar o tráfego às diferentes condições de rádio da sua área de cobertura, tratando a maioria dos dispositivos da mesma maneira, ou seja, usando valores médios da célula. A arquitetura O-RAN foi projetada para melhorar essa situação, personalizando estratégias de controle de tráfego centradas nos dispositivos e fornecendo otimizações proativas ao prever as condições da rede. Isso pode ser realizado através da aplicação de estratégias de controle de tráfego utilizando técnicas de AI/ML nos componentes *Non*-RT e *Near*-RT RIC a partir dos dados coletados através da interface O1 dos módulos O-CUs e O-DUs.

- **Escalonamento elástico da RAN virtualizada:** no contexto de uma RAN virtualizada, a capacidade de processamento é centralizada no *data center* (ou seja, um *cluster* de unidades de banda base, também conhecido como *pool* de BBUs); enquanto a cobertura é distribuída nos pontos de rádio. Para obter o máximo de economia, várias BBUs são agrupadas para trabalharem juntas na mesma infraestrutura virtualizada. O *software* em execução nas BBUs será arquitetado de modo que possa ser escalonado de acordo com a demanda. A RAN virtualizada também será compatível com NFV e a estrutura de orquestração NFV será capaz de escalonar elasticamente o número de entidades com base em políticas baseadas na hora

do dia ou políticas baseadas no uso de processador/memória, por exemplo. Técnicas de IA/ML podem ser empregadas para construir modelos de como a demanda varia no tempo e encontrar ciclos de crescimento e redução da demanda ao longo do dia. O orquestrador NFV pode então consultar esse modelo em vez de uma política estática e escalonar os recursos da rede de forma mais eficiente com base em uma previsão de demanda mais precisa.

- **Predição de localização**: técnicas de IA/ML podem ajudar a localizar geograficamente, com precisão, um dispositivo de usuário (i.e., um telefone celular), com base nas condições de sinal experimentadas por esse dispositivo, da seguinte forma: em uma RAN virtualizada, métricas como a resposta ao impulso do canal medida em vários pontos e potência relativa de cada dispositivo vista por diferentes pontos de rádio podem ser usadas como entradas de um modelo para localizar o dispositivo. Técnicas de IA/ML podem ser empregadas para fazer rastreamento de localização *indoor* com precisão com base nas medidas discutidas acima. Isso permite determinar a localização precisa do usuário e, usando IA/ML, pode-se construir um banco de dados espaço-temporal de padrões de densidade dos usuários, padrões de movimento e padrões de tráfego de dados. Essas informações serão muito úteis para uma operadora ajustar sua rede, bem como para permitir serviços de valor agregado aos proprietários dos estabelecimentos onde a localização é feita. Por exemplo, técnicas que extraem o padrão de movimento de multidões podem ser usadas para disparar alarmes quando os padrões de movimento se desviam da média, o que pode servir como indicador de uma emergência no local. Em um ambiente de varejo, um *shopping* por exemplo, a extração do padrão de movimento pode ser muito útil para entender quais lojas e em qual sequência as pessoas costumam visitar. Da mesma forma, conhecer os padrões de movimento dos clientes ajudará os lojistas a colocar seus anúncios e ofertas nos locais de maior movimento. Entretanto, aqui existe um problema com a Lei Geral de Proteção de Dados Pessoais (LGPD). Os dados poderão ser oferecidos aos lojistas somente através de um serviço compatível com a LGPD.

6.2.4. VANTAGENS DA APLICAÇÃO DE IA/ML NO PARADIGMA OPENRAN

O PARADIGMA DO OPENRAN FOI PENSADO PARA APLICAR ALGORITMOS DE IA/ML COM O INTUITO DE MELHORAR SIGNIFICATIVAMENTE O DESEMPENHO E AUTOMAÇÃO DE OPERAÇÃO DAS REDES DE ACESSO POR MEIO DE ALGORITMOS QUE APRENDEM COM A EXPERIÊNCIA E APRIMORAM CONTINUAMENTE O SISTEMA.

Além dos benefícios do uso de técnicas de IA/ML discutidos na Seção 4, a mudança de paradigma introduzida pelo OpenRAN traz algumas vantagens importantes.

O paradigma do OpenRAN foi pensado para aplicar algoritmos de IA/ML com o intuito de melhorar significativamente o desempenho e automação de operação das redes de acesso por meio de algoritmos que aprendem com a experiência e aprimoram continuamente o sistema. Muitos recursos da RAN, que tradicionalmente são configurados manualmente, contarão com recursos de IA/ML para lidar com a complexidade crescente das redes atuais e futuras. As especificações da O-RAN ***Alliance*** fornecem uma estrutura para usar IA/ML visando otimizar a forma como os recursos de rádio são gerenciados em uma rede 5G. A estrutura fornece interfaces abertas e desagrega várias funções, o que também simplifica os requisitos de ***hardware*** no ***site*** da célula. Isso é feito por meio de aplicações hospedadas nas funções lógicas RIC que podem ser implementadas para controle quase em tempo real (***Near***-RT RIC, conectado ao CU (***Centralized Unit***) através da interface aberta E2) e para controle com maior atraso (***Non***-RT RIC, conectado ao ***Near***-RT RIC por meio da interface A1). Algoritmos de IA/ML, executados como aplicações nas funções lógicas RIC, podem ajudar as operadoras a melhorar a utilização da rede e a experiência do cliente por meio da automação. Esses algoritmos são treinados com dados reais da rede e, a seguir, inferem como gerenciar os recursos da rede de maneira otimizada. Desta forma, as operadoras poderão escalar suas redes sem aumentar as despesas de operação desta na mesma taxa. O objetivo final é reduzir custos e alcançar operações enxutas com gerenciamento de rede totalmente automatizado e sem intervenção humana, além da orquestração de serviços usando automação em malha fechada baseada em algoritmos de IA/ML.

Para redes 4G, as operadoras de telefonia móvel abordaram o gerenciamento da RAN como um processo manual e mecânico que requer intervenção humana para gerenciamento, configuração e operações da rede. Porém, desde o 4G, as operadoras de telefonia móvel já buscavam formas de virtualizar suas redes e adicionar automação e inteligência para reduzir despesas operacionais. Desta forma, a O-RAN ***Alliance*** propôs uma nova arquitetura que introduziu as funções lógicas RIC para trazer automação e inteligência para as redes móveis de próxima geração. O RIC atua para tomar decisões com base nas métricas coletadas dos vários nós da RAN. Essas decisões são orientadas por algoritmos sofisticados, baseados em IA/ML, integrados a várias aplicações chamadas de xApps. As xApps podem ser usadas para um caso de uso específico, uma otimização específica ou uma característica específica. Podem haver muitos xApps sendo executados no ***Near***-RT RIC para resolver problemas relacionados a qualidade de serviço, qualidade de experiência e conectividade perene dos usuários, fornecendo serviços, experiências e conectividade muito superiores. Usuários estão familiarizados com as ***App Stores*** para dispositivos móveis. Eles podem baixar e usar diferentes aplicativos para diversos fins. As xApps serão hospedadas no RIC e, prevê-se que, haverá uma "Loja de xApps" para a rede móvel. Vários fornecedores enviarão seus próprios algoritmos para a loja. Por exemplo, uma empresa pode desenvolver um aplicativo que se concentre apenas no balanceamento de carga ou apenas no controle de mobilidade. As operadoras de telefonia móvel poderão selecionar quais aplicativos desejam usar no RIC. Com essa camada de inteligência adicionada ao RIC, as redes 5G poderão ter um desempenho muito melhor e reagir de forma mais dinâmica às mudanças da rede [109]. Essa mudança de paradigma fará com que surjam novos participantes do ecossistema especializados nesses algoritmos. Haverá uma maior disrupção da RAN à medida que esses novos ***players*** atendam a estas demandas de forma eficiente e mudem o cenário das RANs para sempre.

6.2.5. DESAFIOS E PROBLEMAS ABERTOS

A implementação de técnicas de IA/ML em redes abertas e virtualizadas traz diversos desafios. Nesta subseção estão descritos alguns dos desafios que são considerados mais importantes.

- **Gestão do ciclo de vida do modelo**: modelos de IA/ML são ativos valiosos e em constante evolução. Como tal, uma cadeia de engenharia de *software* sofisticada e controlada é necessária para a gestão do ciclo de vida dos modelos. Quando um modelo é atualizado por meio de um retreinamento fora de seu contexto de uso inicial, por exemplo, muitos desafios significativos são introduzidos em termos de rastreabilidade, estabilidade e a possibilidade de identificar falhas. Com potencialmente muitas instâncias de um modelo implantadas na célula, agrupamentos de células ou mesmo a nível de usuário, este desafio é agravado [109].

- **Aprimorando o aprendizado de máquina na RAN com simulação:** apesar do trabalho contínuo para melhorar a coleta de dados de alta fidelidade na RAN, sua natureza distribuída significa que é provável que os dados continuem a ser esparsos e que a coleta de dados continue a ser uma tarefa que consumirá muitos recursos. Ao mesmo tempo, o aprendizado por meio da exploração pode ser difícil em sistemas ativos, onde a exploração pode degradar ou travar o sistema. Uma abordagem para mitigar esses desafios é integrar a simulação em *loop* ao sistema, fornecendo os meios para uma geração de dados muito mais rica, bem como o uso de técnicas exploratórias de ML. Permitir a simulação em *loop* cria novas possibilidades para a exploração de cenários complexos e em várias etapas, sem comprometer a integridade do sistema. Olhando para o futuro, a automação baseada em técnicas de IA/ML mais sofisticada na RAN exigirá representações digitais dos sistemas (os chamados *digital twins*) implantados para apoiar a tomada de decisões complexas.

- **Algoritmos para RAN**: a otimização de algoritmos da RAN é vital para a criação de um controlador baseado em IA/ML único e abrangente que englobe toda a hierarquia de controle. O benefício de tal controlador seria a capacidade inerente de otimizar vários parâmetros de transmissão entre camadas, simultaneamente. A criação de um controlador com a capacidade de aprender diretamente por meio da exploração do espaço de estados removeria os limites impostos por algoritmos projetados por humanos, tornando possível identificar melhores combinações de parâmetros de transmissão dentro de uma camada e/ou entre camadas. Além disso, um controlador com a capacidade de aprender a partir dos dados seria inerentemente ajustado ao ambiente e livre de hiperparâmetros de rede[4], o que levaria à simplificação da pilha de *software*. Entretanto, a substituição dos algoritmos das camadas 1 até 3, por exemplo, por um único controlador baseado em IA/ML apresenta vários desafios. Um desafio é que mudanças rápidas de parâmetros introduzem o problema de transientes e, portanto, exigem que o controlador baseado em IA/ML preveja a evolução do estado de curto prazo do sistema devido a mudanças de canal e tráfego, por exemplo, bem como as ações que o próprio controlador submete ao sistema. Outro desafio é a necessidade de redefinir o problema de acesso por rádio de forma que possibilite o aprendizado por meio da interação com o ambiente RAN. A abordagem atual de dividir para conquistar com o intuito de fornecer acesso de rádio aos dispositivos de usuários, dividindo o problema em muitos subproblemas de complexidade gerenciável e projetando soluções específicas para cada subproblema, é difícil de se aplicar utilizando controladores baseados em IA/ML. Em outras palavras, permanecer dentro de uma estrutura RAN fragmentada com diferentes controladores baseados em IA/ML, cada um tentando otimizar um recurso RAN diferente enquanto aprende por meio da interação com o ambiente da RAN, impediria o sistema de aprender e prejudicaria o desempenho da rede [110]. Outro aspecto relevante nesse contexto são as interações ainda não conhecidas a fundo entre os controladores RIC e os controladores SDN, bem como os inter-relacionados com os orquestradores NFV. A estabilidade na hierarquia de controle distribuído baseado em IA/ML ainda precisa ser validada.

4 - Em ML, um hiperparâmetro é um parâmetro cujo valor é usado para controlar o processo de aprendizado. O termo hiperparâmetro é utilizado para diferenciá-lo de um parâmetro que é encontrado durante o processo de aprendizado.

A INTELIGÊNCIA ESTÁ RAPIDAMENTE SE TORNANDO UMA NECESSIDADE PARA A IMPLANTAÇÃO, AUTOMAÇÃO, GERENCIAMENTO E OTIMIZAÇÃO DE REDES SEM FIO DE PRÓXIMA GERAÇÃO.

Sendo assim, conclui-se que IA e ML oferecem técnicas bem estabelecidas para resolver problemas de otimização multiparamétrica historicamente muito complexos. Usadas corretamente, essas técnicas têm um potencial tremendo para superar desafios complexos envolvidos na operação de RANs. A inteligência está rapidamente se tornando uma necessidade para a implantação, automação, gerenciamento e otimização de redes sem fio de próxima geração. Isso se deve principalmente à crescente complexidade das redes sem fio 5G e além, em resposta à necessidade de lidar com os exigentes requisitos de serviço. Portanto, O-RAN ***Alliance*** propôs um novo paradigma arquitetural que tem como objetivo alavancar técnicas de IA/ML para incorporar inteligência a cada camada da arquitetura RAN. A inteligência incorporada às camadas da rede permite a alocação dinâmica de recursos de rádio e tem a capacidade de otimizar a eficiência de toda a rede. Desta forma, o paradigma OpenRAN oferece grande potencial para se estabelecer como uma arquitetura aberta, colaborativa e assistida por IA/ML, que irá apoiar a evolução em direção à próxima geração de redes de comunicação sem fio. Entretanto, muitos desafios ainda precisam ser vencidos para que a inteligência de máquina de fato implemente com estabilidade o plano de controle e gerência de 5G no acesso e *core*.

O PARADIGMA OPENRAN OFERECE GRANDE POTENCIAL PARA SE ESTABELECER COMO UMA ARQUITETURA ABERTA, COLABORATIVA E ASSISTIDA POR IA/ML, QUE IRÁ APOIAR A EVOLUÇÃO EM DIREÇÃO À PRÓXIMA GERAÇÃO DE REDES DE COMUNICAÇÃO SEM FIO.

O-RAN ALLIANCE PROPÔS UM NOVO PARADIGMA ARQUITETURAL QUE TEM COMO OBJETIVO ALAVANCAR TÉCNICAS DE IA/ML PARA INCORPORAR INTELIGÊNCIA A CADA CAMADA DA ARQUITETURA RAN.

6.3. ASPECTOS DE SEGURANÇA DO OPENRAN

O OPENRAN PRECISA LIDAR COM NOVOS DESAFIOS DA SEGURANÇA CIBERNÉTICA. QUESTÕES CENTRAIS RELACIONADAS À CADEIA CONFIÁVEL DE FORNECEDORES E A PADRONIZAÇÃO E TESTES DE CONFORMIDADE DOS DIFERENTES HARDWARES DE PROPÓSITO GERAL E SOFTWARES DE CÓDIGO ABERTO SE TORNAM IMPORTANTES NO CONTEXTO DA SEGURANÇA DAS REDES OPENRAN.

Da mesma forma que a segurança é um ponto central das redes 5G em geral, nas redes estruturadas pelo conceito OpenRAN a segurança também deve ser tratada com extremo cuidado. Por se basearem no padrão do 3GPP e, também, por fomentar a questão de redes virtualizadas, o OpenRAN possuirá os mesmos problemas de segurança já descritos na Seção 5. No entanto, por ter como um dos principais pilares a questão de fomentar uma rede desagregada com equipamentos e ***softwares*** de múltiplos fornecedores, o OpenRAN precisa lidar com novos desafios da segurança cibernética. Questões centrais relacionadas à cadeia confiável de fornecedores e a padronização e testes de conformidade dos diferentes ***hardwares*** de propósito geral e ***softwares*** de código aberto se tornam importantes no contexto da segurança das redes OpenRAN. Essa subseção aborda de forma mais detalhada os principais aspectos de segurança aplicados às redes OpenRAN.

Analisando as arquiteturas propostas da O-RAN, em um primeiro momento podem surgir preocupações com as primeiras camadas da rede que tratam da transmissão de sinais pela interface aérea e pelo ***fronthaul***. Com relação à interface aérea, mecanismos de segurança que ofereçam proteção à confidencialidade, integridade e autenticação mútua já são definidos pelo padrão 3GPP e, por consequência, pelo padrão da O-RAN ***Alliance***. No entanto, a arquitetura da O-RAN introduz uma nova interface denominada ***open fronthaul*** que difere do padrão do 3GPP. Sendo assim, uma nova interface aumenta a superfície de ataque da rede O-RAN em comparação com o 3GPP. Segundo a própria O-RAN ***Alliance***, mecanismos de segurança para confidencialidade, integridade e autenticação mútua já estão incorporados na padronização da interface ***open fronthaul*** nos planos de gerência e de usuário [111]. No entanto, os planos de controle e sinalização ainda carecem de mecanismos que não foram incorporados na padronização existente [111]. Portanto, com a padronização existente, a rede O-RAN pode ser vulnerável a ataques MITM dentro da interface open ***fronthaul***. Um terceiro malicioso poderia manipular o tráfego de gerenciamento e controle que ocorre entre os componentes O-DU e O-RU e, até mesmo, poderia conseguir acesso a controles do próprio SMO por meio da

interface O1. O grupo de trabalho da O-RAN ***Alliance*** denominado ***Security Task Group*** (STG) está trabalhando para incorporar nas próximas versões de sua padronização, os mecanismos de segurança apropriados para o plano de gerência, controle e sinalização do open ***fronthaul***, baseados no TLS e em certificados digitais X.509 [112,118] para autenticação mútua [111]. O mesmo grupo aponta que está estudando formas de proteger o SMO contra-ataques iniciados no ***fronthaul***.

Outra diferença significativa entre os padrões 3GPP e O-RAN ***Alliance*** está relacionada à introdução dos blocos ***Non***-RT RIC e ***Near***-RT RIC na rede O-RAN. Esses blocos precisam receber e analisar conjuntos de informações, muitas vezes sensíveis, sobre a rede ou até mesmo sobre usuários da rede, para tomar decisões que envolvam a otimização dos recursos da rede de rádio. Além disso, de acordo com as premissas da própria O-RAN ***Alliance***, esses blocos são constituídos de xApps que podem ser desenvolvidos por diferentes fabricantes de ***software***, pela própria MNO ou por comunidades de ***softwares*** de código aberto. Sendo assim, é importante que a O-RAN ***Alliance*** desenvolva procedimentos para assegurar que o funcionamento dos xApps não tenha nenhuma falha, ou configuração maliciosa, que permita vazamento de dados sensíveis, perda de isolamento entre VMs, ou, até mesmo, que comprometa a disponibilidade da rede. Uma vez que, xApps diferentes podem tomar decisões antagônicas a respeito de um determinado ajuste, causando a indisponibilidade da rede. Além disso, controles rigorosos devem ser mantidos para impedir que esses xApps introduzam vulnerabilidades na rede que possam ser exploradas por um atacante externo. Nesse sentido, o STG está trabalhando para implementar procedimentos que visam construir uma cadeia confiável de provedores de xApps, que sigam as boas práticas de segurança ligadas a essa indústria. Além disso, o ***Test and Integration Focus Group*** (TIFG) tem seu foco em garantir a integração segura dos diferentes xApps, a fim de evitar que aplicações diferentes entrem em conflito e causem indisponibilidade, ou funcionamento não otimizado, de algum setor da rede de rádio. Contudo, as ações desses dois grupos não mostraram nenhum avanço significativo para sanar os problemas relacionados ao conceito de múltiplos fornecedores de xApps. Ainda não se tem uma decisão concreta sobre processos de validação de assinatura de ***softwares*** e validação de padrões de segurança para os xApps de código aberto. Apesar de o STG apontar que terá concluído, até o final de 2021, as ações de segurança relacionados à isolação e assinaturas de códigos de xApps e, também, relacionados à processos de construção de uma cadeia confiável de fornecedores, alguns trabalhos apontam que o desenvolvimento de padrões aceitáveis de validação de segurança neste universo poderá não estar disponível antes de 2025 [52].

Do ponto de vista da operação e manutenção da rede O-RAN, é importante mencionar que o aumento da complexidade operacional com vários fornecedores pode trazer riscos para a disponibilidade e integridade da rede. Capacitação de equipes de operação e integração em tempo hábil para operar o sistema OpenRAN deve ser uma preocupação durante a tomada

de decisão sobre a adoção do novo sistema, e isso não deve ser feito sem o acompanhamento da equipe de segurança da MNO. Sem um processo criterioso de análise de segurança, é possível que a OpenRAN sofra com a presença de inúmeras vulnerabilidades nas diversas camadas de ***hardware***, ***firmware*** e ***software*** e até mesmo possa ser alvo dos mais diversos tipos de ataques. Dada a escolha pelo sistema OpenRAN, é fundamental que haja um processo de capacitação de equipes de profissionais em curto prazo, para atuarem em correções de falhas com agilidade e assertividade. Em contrapartida, a adoção de uma rede baseada em soluções proprietárias de um único fornecedor carrega consigo a possibilidade de se beneficiar das equipes de resposta a incidentes que muitos dos grandes fornecedores mantêm em seu quadro de operações. Essas equipes já possuem capacitação antecipada e tendem a conhecer todo o sistema desenvolvido internamente pelo fabricante. Assim, podem fornecer correções de falhas mais assertivas e ágeis. Além disso, fabricantes de soluções 3GPP submetem suas soluções a processos já estabelecidos de auditoria, como o ***Network Equipment Security Assurance Scheme*** (NESAS) que é mantido pelo próprio 3GPP em conjunto com a GSMA [114], uma das maiores associações de testes e validação de soluções de segurança para o mercado de comunicações móveis. Contudo, na outra ponta, os modelos de segurança analisados pela O-RAN ***Alliance*** ainda carecem de desenvolvimento de soluções que estão sendo realizadas em conjunto com as diversas áreas envolvidas no projeto. Fornecedores e parceiros de integração de sistemas esperam aprimorar a segurança e a confiabilidade nas redes ao longo deste processo. No entanto, ainda há um tempo para que testes de interoperabilidade e segurança de ponta-a-ponta mostrem que as redes OpenRAN sejam seguras e que haja confiança por parte das MNOs para implantarem essa solução em maior escala.

DO PONTO DE VISTA DA OPERAÇÃO E MANUTENÇÃO DA REDE O-RAN, É IMPORTANTE MENCIONAR QUE O AUMENTO DA COMPLEXIDADE OPERACIONAL COM VÁRIOS FORNECEDORES PODE TRAZER RISCOS PARA A DISPONIBILIDADE E INTEGRIDADE DA REDE

SEM UM PROCESSO CRITERIOSO DE ANÁLISE DE SEGURANÇA, É POSSÍVEL QUE A OPENRAN SOFRA COM A PRESENÇA DE INÚMERAS VULNERABILIDADES NAS DIVERSAS CAMADAS DE HARDWARE, FIRMWARE E SOFTWARE E ATÉ MESMO POSSA SER ALVO DOS MAIS DIVERSOS TIPOS DE ATAQUES.

6.4. VANTAGENS E DESVANTAGENS DO OPENRAN

UM DOS PRINCIPAIS ARGUMENTOS PARA A ADOÇÃO DE OPENRAN É A REDUÇÃO DO CUSTO DA INFRAESTRUTURA DA RAN.

A parte conhecida como RAN das redes de comunicações móveis representa um campo fértil para inovações e quebras de paradigmas. A RAN tradicional, usada ainda hoje, é muito parecida com a RAN introduzida pelos sistemas 2G e, basicamente, seus componentes são projetados para promoverem um alto desempenho, com tecnologias proprietárias fornecidas por um pequeno número de fabricantes que dominam o mercado de componentes da RAN. O novo conceito de OpenRAN promete um ecossistema maior e mais competitivo, no qual as tecnologias proprietárias são substituídas por alternativas alinhadas a este novo paradigma e com maior possibilidade de inovações no setor. Contudo, como pode ser visto ao longo desta seção, ainda há um longo caminho a ser percorrido para que uma solução de OpenRAN competitiva e madura esteja disponível para a implementação das redes 5G em larga escala.

As soluções de OpenRAN são desenhadas para remodelar a indústria de componentes RAN em direção a redes móveis mais inteligentes, abertas, virtualizadas e totalmente interoperáveis [115], baseadas em ***softwares*** desagregados que funcionam em ***hardwares*** de uso genérico, como os processadores GPP e os servidores COTS.

AS MNOS SE MOSTRAM INTERESSADAS COM A PERSPECTIVA DA SOLUÇÃO OPENRAN FAVORECER UM MAIOR NÚMERO DE FORNECEDORES E, ALÉM DISSO, PODER SUBSTITUIR O HARDWARE E SOFTWARE MONOLÍTICOS POR COMPONENTES DESAGREGADOS QUE PERMITAM A COMBINAÇÃO DE DIFERENTES EQUIPAMENTOS DE DIFERENTES FABRICANTES. EM PRINCÍPIO, ESTE ECOSSISTEMA AMPLIADO FAVORECE UMA MAIOR COMPETIÇÃO E, CONSEQUENTEMENTE, A REDUÇÃO DE CAPEX DOS COMPONENTES DA RAN.

Nos últimos anos, tanto MNOs quanto fornecedores de tecnologia ligados ao conceito de OpenRAN vêm desenvolvendo soluções e realizando testes de seus ***softwares*** e ***hardwares***. O objetivo dessa seção é explorar os principais resultados já obtidos com esses diversos testes e fazer um paralelo entre as vantagens e desvantagens da OpenRAN, comparando-a às redes tradicionais baseadas em soluções proprietárias.

Um dos principais argumentos para a adoção de OpenRAN é a redução do custo da infraestrutura da RAN [98]. As MNOs se mostram interessadas com a perspectiva da solução OpenRAN favorecer um maior número de fornecedores e, além disso, poder substituir o ***hardware*** e ***software*** monolíticos por componentes desagregados que permitam a combinação de diferentes equipamentos de diferentes fabricantes. Em princípio, este ecossistema ampliado favorece uma maior competição e, consequentemente, a redução de CAPEX dos componentes da RAN. Estudos apontam que a

HÁ AINDA MUITOS QUESTIONAMENTOS QUANTO AOS REAIS GANHOS EM TERMOS DE OPEX QUE A OPENRAN PODERÁ ATINGIR, PRINCIPALMENTE, POR CAUSA DE DOIS PONTOS CRUCIAIS: I) CUSTO DE INTEGRAÇÃO DE COMPONENTES DA RAN E II) DESEMPENHO DA REDE.

adoção de OpenRAN pode gerar economia de 40 a 50 % com relação ao CAPEX em comparação com a implantação de RANs tradicionais [40,116]. Além disso, a Rakuten, que é uma MNO japonesa que aplica o conceito de OpenRAN em sua rede comercial, mostrou que seu CAPEX foi 40% menor na implantação de uma solução OpenRAN se comparado ao CAPEX de implantação das RANs proprietárias [117]. É importante destacar que o resultado apresentado pela Rakuten é relacionado a implantação de redes legadas e não de redes 5G [117]. Resultados relacionados ao CAPEX da Rakuten para implantação de redes 5G em larga escala, até onde foi pesquisado, ainda não estão disponíveis. É importante mencionar que o potencial de redução de CAPEX não é exclusividade das soluções OpenRAN. Soluções proprietárias de C-RAN também podem levar a economia de CAPEX, segundo apontou o estudo da Senza Fili, que mostra uma redução de CAPEX de 49 % quando se adota o C-RAN em comparação com RANs tradicionais [118].

Um segundo argumento favorável à decisão de adotar OpenRAN é que, de acordo com a O-RAN ***Alliance***, o sistema O-RAN é capaz de reduzir o OPEX da rede [98]. Basicamente, a O-RAN ***Alliance*** aponta que a redução do OPEX se dá principalmente por causa da possibilidade de automação da RAN. Ou seja, o O-RAN introduz a inteligência artificial em toda a arquitetura da RAN, alavancando novas tecnologias baseadas em aprendizado de máquina para automatizar os processos e as funções de redes, assim reduzindo a intervenção humana e, portanto, diminuindo os custos operacionais. Além disso, a virtualização de alguns componentes da RAN permite rápida implementação e escalonamento de capacidade, com a maior parte do trabalho sendo feita de forma centralizada e por meio de instalações e atualizações de ***software***. Alguns estudos feitos por empresas de consultoria e investimentos confirmam que o uso de OpenRAN irá levar a um menor OPEX para as MNOs, algo em torno de 30 a 35% de economia [116]. De acordo com [118], também é possível perceber que haverá economia de OPEX para operadoras que adotarem soluções C-RAN, mesmo que proprietárias. Contudo, há ainda muitos questionamentos quanto aos reais ganhos em termos de OPEX que a OpenRAN poderá atingir, principalmente, por causa de dois pontos cruciais: i) custo de integração de componentes da RAN e ii) desempenho da rede. Ainda, há estudos que mostram que o uso de múltiplos fornecedores na implementação de ambientes de redes pode levar a aumento de OPEX, como apresentado a seguir nesta seção. Por fim, deve-se ressaltar que o uso de inteligência artificial e automação de processos também podem ser aplicados nas soluções tradicionais de implementação das RANs, não sendo uma exclusividade das soluções OpenRAN, embora possam ser favorecidas neste ambiente pela possível futura existência de um ecossistema maior de empresas.

Com relação ao custo de integração, em linhas gerais, uma MNO que optar por utilizar OpenRAN terá que lidar com diferentes fornecedores de RUs, COTS e fornecedores de *softwares* de DU (***Distributed Unit***) e CU. Portanto, a integração entre todos esses componentes de diferentes fornecedores irá trazer mais custos operacionais à MNO. A **Figura 8** ilustra a diferença entre a implantação de uma RAN baseada em um único fornecedor e a implantação de OpenRAN baseada em múltiplos fornecedores.

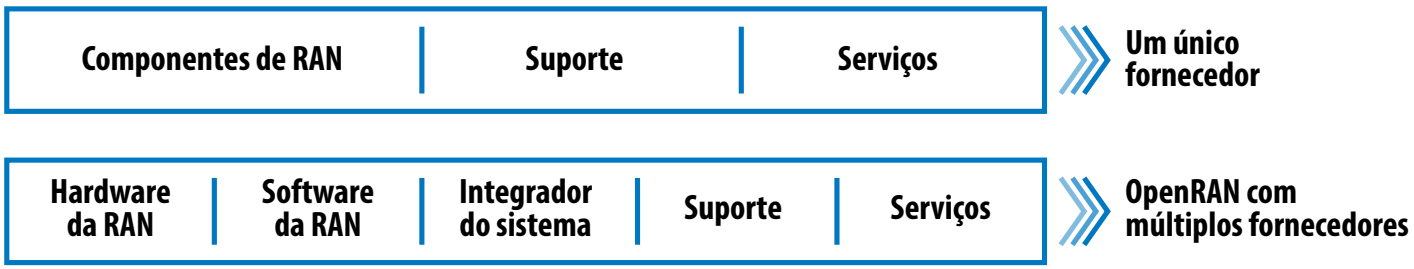

Figura 8: Pilha de custos das diferentes abordagens para implantação da RAN.

As abordagens tradicionais de implantação da RAN apresentam os custos de ***hardware*** e ***softwares*** que compõem a RAN, os custos com suporte e os custos com serviços prestados por parte do fornecedor. Já na abordagem OpenRAN, tem-se de forma separada os custos de ***hardware*** e ***software***, uma vez que é possível adquirir esses componentes de fornecedores diferentes. Além disso, existirão os custos com suporte e serviços, da mesma forma que a rede tradicional. Por fim, no longo prazo existirão custos com o integrador do sistema (do inglês, ***System Integrator*** - SI). É importante mencionar que o SI poderá ser a própria MNO, um fornecedor específico, ou uma empresa terceirizada provedora de serviços de integração. O modelo do SI adotado pela MNO irá definir no longo prazo qual será o custo de integração. Pelo fato de não se ter ao certo valores que modelem os custos com a integração do sistema, principalmente em relação às redes 5G, é difícil saber qual das duas pilhas apresentará maior ou menor custo total.

A POSSIBILIDADE DE HAVER UM ECOSSISTEMA COM VASTA DISPONIBILIDADE DE SOFTWARES E HARDWARES DE DIFERENTES FORNECEDORES, FREQUENTEMENTE APONTADA COMO UM DOS PRINCIPAIS BENEFÍCIOS DA OPENRAN, PODE TAMBÉM SER UM DE SEUS MAIORES OBSTÁCULOS, DEVIDO ÀS QUESTÕES DA INTEGRAÇÃO E INTEROPERABILIDADE.

Além do exposto acima, há também uma incerteza quanto a um problema operacional que pode ocorrer com a OpenRAN com múltiplos fornecedores. Em uma RAN tradicional, as MNOs normalmente têm um único fornecedor que auxilia na solução de eventuais problemas e mal funcionamento de componentes. No entanto, num cenário com múltiplos fornecedores, a MNO pode não ter imediatamente a informação sobre o produto que não funciona de maneira correta. Por exemplo, quando uma falha acontece em um equipamento que é constituído de um servidor COTS, um ***firmware*** e um ***software***, cada um fornecido por um fabricante diferente, a MNO pode demorar para ter a correção da falha, uma vez que os diferentes fabricantes podem divergir entre eles sobre a raiz do problema; situação semelhante pode acontecer com equipamentos de fabricantes distintos interconectados por meio de uma interface aberta, quando surge um problema de interoperabilidade não detectado anteriormente. Portanto, fica difícil entrar em contato com o responsável pela falha ou pelo mal funcionamento. Esses problemas, juntamente com integração deficitária, podem fazer com que a rede trabalhe em regime de ineficiência energética ou até mesmo demandar mais investimentos na infraestrutura para atendimento de KPIs importantes. Estudos apontam também que o uso de múltiplos fornecedores para a implementação de ambientes de rede resulta em alguns problemas e pontos de atenção. Por exemplo, o estudo apresentado em [119] indica que o uso de múltiplos fornecedores resulta em: i) crescimento da complexidade de interoperabilidade, ii) decréscimo da efetividade na resolução de problemas e iii) aumento de custo para remediar problemas operacionais. Além disso, este estudo indica que as economias atingidas em termos de CAPEX são suprimidas pelos custos operacionais ao longo da vida útil da rede. Portanto, a possibilidade de haver um ecossistema com vasta disponibilidade de ***softwares*** e ***hardwares*** de diferentes fornecedores, frequentemente apontada como um dos principais benefícios da OpenRAN, pode também ser um de seus maiores obstáculos, devido às questões da integração e interoperabilidade.

QUANDO UMA FALHA ACONTECE EM UM EQUIPAMENTO QUE É CONSTITUÍDO DE UM SERVIDOR COTS, UM FIRMWARE E UM SOFTWARE, CADA UM FORNECIDO POR UM FABRICANTE DIFERENTE, A MNO PODE DEMORAR PARA TER A CORREÇÃO DA FALHA, UMA VEZ QUE OS DIFERENTES FABRICANTES PODEM DIVERGIR ENTRE ELES SOBRE A RAIZ DO PROBLEMA; SITUAÇÃO SEMELHANTE PODE ACONTECER COM EQUIPAMENTOS DE FABRICANTES DISTINTOS INTERCONECTADOS POR MEIO DE UMA INTERFACE ABERTA, QUANDO SURGE UM PROBLEMA DE INTEROPERABILIDADE NÃO DETECTADO ANTERIORMENTE.

Um importante passo foi dado recentemente com relação aos problemas apontados sobre integração e desempenho. A O-RAN ***Alliance*** e a TIP uniram esforços para criação do PlugFest e, também, de centros de teste e integração (do inglês, ***Open Test and Integration Center*** - OTIC) [115]. Os OTICs possuem objetivos de verificação de componentes OpenRAN, bem como se suas interfaces estão de acordo com o padrão, promovendo a validação das implementações do ecossistema OpenRAN. Por sua vez, o PlugFest serve para que os envolvidos apresentem provas de conceito para demonstração de funcionalidades e interoperabilidade de vários fornecedores de equipamentos OpenRAN. Porém, no momento, não existem ainda padrões ou metodologias de teste bem definidos para garantir a interoperabilidade e desempenho de produtos de ***software*** e ***hardware*** de diferentes fornecedores para o 5G.

Com relação ao desempenho de rede, em alguns casos de uso mais voltados para cobertura de áreas rurais, a OpenRAN mostrou desempenhos satisfatórios em testes comerciais realizados, principalmente pelas MNOs Vodafone e Telefônica. A cobertura de áreas rurais tem um foco diferente da cobertura de áreas urbanas. Nas áreas rurais, as exigências quanto à taxa de transmissão, mobilidade e volume de tráfego são menores. A Vodafone implantou OpenRAN em áreas rurais e de baixa remuneração média por usuário (do inglês, ***Average Revenue per User*** - ARPU) em diversos países, como, Turquia, Índia, República Democrática do Congo, Moçambique, Irlanda e também pelo Reino Unido. Contudo, essas MNOs somente divulgaram informações a respeito do desempenho da OpenRAN aplicada em redes legadas. A Vodafone apresentou informações sobre o desempenho das redes 2G e 4G implantadas na Turquia, onde vários KPIs foram atendidos, mostrando uma qualidade de serviço adequada para o cenário em questão [120]. Com um foco parecido, a Telefônica apoiou o programa Internet para Todos Peru (IpTP) que foi lançado em maio de 2019. O programa IpTP foi constituído usando o conceito de OpenRAN com vários fornecedores de componentes de rede 3G e 4G. Rádios configurados via ***software*** e menor necessidade de otimização feita por humanos permitiu que o IpTP criasse um modelo de negócios para MNOs explorarem a cobertura rural do país de forma economicamente sustentável. É importante citar que, no Brasil, as MNOs Tim, Vivo e Algar Telecom já iniciaram seus testes das redes OpenRAN [121,122].

AS IMPLEMENTAÇÕES DE COMPONENTES DE BANDA BASE ALTAMENTE VIRTUALIZADOS E BASEADOS EM COTS NÃO CONSEGUIRÃO ATINGIR DESEMPENHO SATISFATÓRIO SEM PLATAFORMAS ADICIONAIS DE ACELERADORES DE HARDWARE BASEADAS EM FPGAS (DO INGLÊS, FIELD PROGRAMMABLE GATE ARRAYS) GPUS OU POSSIVELMENTE NOVAS ARQUITETURAS DE CIRCUITOS IMPRESSOS, TODAS COM IMPLICAÇÕES DE MAIOR CUSTO E MAIOR CONSUMO DE ENERGIA ELÉTRICA.

Nos outros casos de uso de OpenRAN para aplicação 4G em áreas mais populosas, o desempenho alcançado pelas RANs baseadas em COTS usando GPPs ficou abaixo do desempenho que uma RAN com ***hardware*** dedicado com circuitos integrados de aplicação específica (do inglês, ***Application Specific Integrated Circuits*** - ASIC) pode alcançar. Além do desempenho maior, os equipamentos dedicados também consomem menor quantidade de energia elétrica, mostrando-se assim equipamentos com maior eficiência energética. Como exemplo, um estudo feito no Japão [123], mostrou que a Rakuten ainda possui um desempenho médio pior que outras três MNOs concorrentes. Os usuários da rede 4G da Rakuten estão experimentando taxas de transferência de dados nos ***downloads*** mais baixas e latências de rede mais altas e, de forma geral, um desempenho médio mais baixo se comparado às demais concorrentes. De acordo com [123], as análises de desempenho foram feitas com base na rede 4G, uma vez que os serviços 5G lançados recentemente pela Rakuten não possuem dados suficientes para análises de desempenho. Porém, espera-se que as diferenças de desempenho em redes 5G sejam ainda maiores, principalmente devido à maior demanda de tráfego, maior complexidade de cenários, maior mobilidade, uso de MIMO massivo e algoritmos de cancelamento ou mitigação de interferências. Uma possível conclusão até agora é que as implementações de componentes de banda base altamente virtualizados e baseados em COTS não conseguirão atingir desempenho satisfatório sem plataformas adicionais de aceleradores de ***hardware*** baseadas em FPGAs (do inglês, ***Field Programmable Gate Arrays***) GPUs ou possivelmente novas arquiteturas de circuitos impressos, todas com implicações de maior custo e maior consumo de energia elétrica.

A segurança cibernética deve ser uma preocupação constante para a tomada de decisão sobre a utilização de OpenRAN nas redes das MNOs. Portanto, uma questão importante a ser analisada é se a OpenRAN é de fato segura. De acordo com [124], a OpenRAN pode ser segura se seguir três critérios: i) estabelecer uma cadeia confiável de fornecedores que estabeleçam boas práticas de segurança cibernética em seus processos e que sejam transparentes com relação a essas práticas, ii) empregar interfaces abertas e padronizadas de acordo com as especificações da O-RAN ***Alliance*** e 3GPP e, por fim, iii) estabelecer testes de segurança cibernética em laboratórios confiáveis. Esses critérios não são muito diferentes dos critérios estipulados pelos fornecedores de componentes proprietários. No entanto, uma RAN monolítica se beneficia de um conceito denominado segurança por obscuridade, por possuir componentes cujos programas não são publicamente conhecidos. A OpenRAN precisa ter critérios de segurança mais rígidos e transparentes, pois muitos dos programas que compõem a arquitetura de padrão aberto são conhecidos publicamente, como por exemplo os sistemas operacionais Linux,

UMA RAN MONOLÍTICA SE BENEFICIA DE UM CONCEITO DENOMINADO SEGURANÇA POR OBSCURIDADE, POR POSSUIR COMPONENTES CUJOS PROGRAMAS NÃO SÃO PUBLICAMENTE CONHECIDOS.

O SUCESSO DA OPENRAN DEPENDE, ENTRE MUITAS COISAS, DA MANUTENÇÃO DE UMA COMUNIDADE VIBRANTE E ATIVA DE FORNECEDORES, COM OPORTUNIDADE DE NEGÓCIOS PARA TODOS ELES DENTRO DO ECOSSISTEMA.

os contêineres Docker, entre outros. Por outro lado, o fato de haver uma comunidade ativa e ética analisando os problemas de segurança dos programas envolvidos pode favorecer o descobrimento de falhas de segurança e suas respectivas soluções e atualizações, ao mesmo tempo em que também pode se tornar um pesadelo para as MNOs. Se um terceiro malicioso descobrir uma falha antes da comunidade, ele poderá explorá-la a fim de obter vantagens ilícitas e causar danos à rede e também à imagem da MNO. Recentemente, a O-RAN ***Alliance*** se pronunciou sobre o assunto de segurança cibernética em seu desenvolvimento dos padrões O-RAN [111]. A O-RAN ***Alliance*** basicamente aponta os aspectos positivos em relação à segurança sob a ótica de seu padrão aberto e com desagregação. Porém, cita também que a desagregação pode aumentar a superfície de ataque cibernético e que o uso de IA em componentes da RAN pode trazer consequências imprevistas durante o funcionamento da rede. Por fim, a O-RAN ***Alliance*** reconhece que ainda há desafios a serem superados quanto à segurança, mas anuncia a criação de um grupo, o STG que irá tratar todos os aspectos de segurança dentro das especificações do padrão. Como conclusão, o STG mostrou que alguns mecanismos de segurança, para autenticação, integridade e confidencialidade já estão disponíveis no padrão, porém, alguns outros aspectos, somente serão finalizados na segunda metade do ano de 2021. Sem dúvida, essa iniciativa é muito importante para o avanço do O-RAN, porém, pode adiar a adoção em massa desse padrão por parte das MNOs.

Por fim, um aspecto importante que se deve levar em consideração é que o sucesso da OpenRAN depende, entre muitas coisas, da manutenção de uma comunidade vibrante e ativa de fornecedores, com oportunidade de negócios para todos eles dentro do ecossistema. No longo prazo, isso pode se tornar um desafio, pois fornecedores menores e instáveis economicamente podem não sobreviver muito tempo no mercado das comunicações móveis. Neste momento, as implantações OpenRAN se concentram mais em redes legadas e em alguns projetos pilotos 5G implantados em alguns nichos, como redes ***indoor*** e de áreas rurais. Para a utilização da OpenRAN em maior escala nas redes 5G, principalmente em redes que façam uso de MIMO massivo, ainda devem ser necessários alguns anos (algo em torno de 5 anos talvez seja uma boa previsão) para que a OpenRAN seja madura e competitiva o suficiente para ser adotada como solução primeira pelas MNOs tradicionais para suas redes 5G. Este tempo para maturação pode ser um problema para a OpenRAN, pois como as principais MNOs já estão implantando suas infraestruturas 5G, é provável que os novos fornecedores de OpenRAN não consigam uma participação de mercado significativa na construção do 5G, ao menos em um primeiro momento. Sem esses fornecedores, a comunidade pode perder força e, até mesmo, fazer com que o mercado de OpenRAN se estabeleça com alguns poucos e grandes fornecedores, algo muito parecido com o que se vê hoje no mercado tradicional de RAN.

PARA A UTILIZAÇÃO DA OPENRAN EM MAIOR ESCALA NAS REDES 5G, PRINCIPALMENTE EM REDES QUE FAÇAM USO DE MIMO MASSIVO, AINDA DEVEM SER NECESSÁRIOS ALGUNS ANOS (ALGO EM TORNO DE 5 ANOS TALVEZ SEJA UMA BOA PREVISÃO) PARA QUE A OPENRAN SEJA MADURA E COMPETITIVA O SUFICIENTE PARA SER ADOTADA COMO SOLUÇÃO PRIMEIRA PELAS MNOS TRADICIONAIS PARA SUAS REDES 5G.

7. CONCLUSÃO

As redes 5G, com seus diversos cenários de uso, possuem potencial para revolucionar os mais diversos setores da sociedade. Além das altíssimas taxas de transmissão disponíveis no cenário eMBB, novas aplicações de IoT e de Internet tátil, além de aplicações de missão crítica, são viabilizadas pelos cenários mMTC e URLLC. Por fim, o cenário de 5G para áreas remotas pode contribuir significativamente para diminuir a divisão digital no país, ao mesmo tempo que viabiliza o conceito de fazendas inteligentes, agregando mais valor ao agronegócio.

A complexidade para implementação das diversas aplicações, com requisitos de desempenho diversos, dos cenários de uso estabelecidos para o 5G, só é possível por meio do conceito de fatiamento de rede (ou ***network slicing***), no qual várias redes lógicas são estabelecidas sobre uma mesma infraestrutura de rede compartilhada, com cada rede lógica atendendo uma aplicação específica. A incorporação de algoritmos utilizando inteligência artificial e aprendizado de máquina é fundamental para que se obtenha a otimização do desempenho da rede. Porém, é importante considerar que a adoção da IA não é tão simples. Ao adotar alguma solução de rede baseada em otimizações via IA, é fundamental possuir uma equipe especializada para configuração, parametrização, pré-processamento e treinamento da própria IA. Caso os operadores não possuam tal capacitação, os resultados da otimização da rede podem não ser alcançados.

Flexibilidade, escalabilidade, agilidade e a capacidade de ser programável são palavras chaves para que as redes 5G possam entregar todos os serviços previstos com qualidade. Estes requisitos dão origem à tendência de "***softwarização***" das redes 5G, com os conceitos de redes definidas por ***software*** e virtualização de funções de rede assumindo papéis importantes.

Outro aspecto de grande relevância nas redes 5G está relacionado à questão da segurança cibernética. Um leque muito maior de aplicações, algumas consideradas de missão crítica, associado a um número muito maior de usuários, principalmente devido ao cenário mMTC, a heterogeneidade dos dispositivos conectados à rede, com diferentes capacidades de armazenamento e processamento, e a tendência de "***softwarização"*** das redes, tornam os aspectos de segurança e privacidade muito mais críticos nas redes 5G do que nas gerações anteriores.

A tendência de "***softwarização***" das redes, que antes estava restrita ao núcleo da rede, começou também a ganhar força na rede de acesso, dando origem ao OpenRAN**,** que pode ser definido como um movimento do ecossistema de redes de comunicações móveis com o objetivo de definir uma estrutura flexível e interoperável para as redes de acesso, com desagregação entre o ***hardware*** e o ***software*** da rede e o uso de interfaces abertas entre os diversos componentes da RAN. O movimento OpenRAN se utiliza da ideia de "***softwarização***" das redes de telecomunicações, utilizando os conceitos de SDN, NFV, C-RAN e V-RAN, baseada em ***softwares*** desagregados que funcionam em ***hardwares*** de uso genérico, como processadores GPP e servidores COTS.

Nos últimos anos, tanto MNOs quanto fornecedores de tecnologia ligados ao conceito de OpenRAN vêm desenvolvendo soluções e realizando testes de seus ***softwares*** e ***hardwares***. As MNOs se mostram interessadas com a perspectiva do OpenRAN favorecer um maior número de fornecedores e, além disso, permitir substituir os ***hardwares*** e ***softwares*** monolíticos por componentes desagregados que permitam a combinação de diferentes equipamentos de diferentes fabricantes. Em princípio, este ecossistema ampliado favorece uma maior competição e, consequentemente, a redução de CAPEX dos componentes da RAN.

A redução do OPEX que pode ser obtida ao se adotar o OpenRAN é um tema bastante sensível que ainda carece de melhor análise. A redução da intervenção humana e a virtualização prometem reduzir consideravelmente os custos de operação e manutenção da RAN. No entanto, há de se considerar um custo adicional às redes OpenRAN devido ao integrador do sistema. Esses custos podem variar de acordo com o tipo de SI adotado e ainda não existem muitos dados sobre o impacto dos custos do SI no OPEX das MNOs. Além disso, é possível perceber que uma integração deficitária pode levar a rede a trabalhar em regime de baixa eficiência energética ou com baixo desempenho de KPIs específicos. Isso poderia demandar mais investimentos em infraestrutura, impactando tanto os custos de CAPEX quanto OPEX.

A segurança cibernética, como já destacado, é uma questão chave para as redes 5G, e a OpenRAN pode também ter algumas desvantagens neste quesito, quando comparada às soluções tradicionais de RAN, devido ao fato de haver uma maior superfície de ataque na OpenRAN e também devido à falta de maturidade desta tecnologia.

Até o momento em que esse documento estava sendo escrito, as principais informações sobre implementações de redes OpenRAN se concentravam em dados obtidos de aplicações em redes indoor e coberturas de áreas rurais. A maior parte dessas implementações utilizavam tecnologias de redes anteriores ao 5G. Os resultados preliminares mostram que nesses nichos o OpenRAN possui um bom potencial tecnológico. Contudo, o mercado ainda carece de mais informações a respeito de implementações do OpenRAN utilizando tecnologias 5G, como MIMO massivo e, principalmente, aplicados a grandes centros urbanos e também a cenários que exijam baixíssima latência. Talvez, dentro de um horizonte de alguns anos, resultados mais consistentes que mostrem a maturidade das redes OpenRAN em cenários mais complexos, possam indicá-las como sendo a primeira solução de implementações 5G das MNOs. Durante esse tempo, o OpenRAN deve ter o cuidado de manter uma comunidade vibrante em torno do tema, criando um cenário favorável aos novos fornecedores dessa tecnologia e, assim, fomentando ainda mais o ecossistema das soluções abertas, seguras, interoperáveis e inteligentes para o mercado de RAN.

ACRÔNIMOS:

2G: segunda geração das redes de comunicações móveis
3G: terceira geração das redes de comunicações móveis
3GPP: *Third Generation Partnership Project*
4G: quarta geração das redes de comunicações móveis
5G: quinta geração das redes de comunicações móveis
5GC: 5G *Core*
5G-NSA: 5G *Non-Standalone*
5GPPP: 5G *Public Private Partnership*
5G-SA: 5G *Standalone*
6G: sexta geração das redes de comunicações móveis
A-CPIs: *Applications-Controller Plane Interfaces*
APIs: *Application Programming Interfaces*
ARPU: remuneração média por usuário - *Average Revenue per User*
ASIC: circuitos integrados de aplicação específica - *Application Specific Integrated Circuits*
BBUs: unidades de banda base - *Baseband Units*
BLOBs: *Binary Large OBjects*
BNDES: Banco Nacional de Desenvolvimento Econômico e Social
BSS: *Business Support System*
CAPEX: despesas de capital - *capital expenditure*
CoMP: Multi-Ponto Coordenado - Coordinated *Multi-Point*
COTS: comercial de prateleira *- commercial off-the-shelf*
CPRI: Interface de Rádio Pública Comum - *Common Public Radio Interface*
CPU: Unidade Central de Processamento - *Central Process Unit*
C-RAN: Rede de Acesso a Rádio na Nuvem - *Cloud Radio Access Network*
CSMF: *Communication Service Management Function*
CU: *Centralized Unit*
D-CPIs: *data-controller plane interfaces*
DoS: negação de serviços - *Denial of Service*
DU: *Distributed Unit*
EMs: *Element Managers*
eMBB: *Enhanced Mobile Broadband*
eNBs: *eNodeBs*
EPC: *Evolved Packet Core*
eRAC: *Enhanced Remote Area Communications*
ERB: estação rádio base
eSIM: *Embed Subscriber Indentity Module*
ETSI: *European Telecommunications Standards Institute*
FPGAs: arranjo de portas programáveis em campo - *Field Programmable Gate Arrays*
gNBs: *gNodeBs*
GPPs: processadores de uso geral - *General Purpose Processors*
GPS: *Global Positioning System*
GPUs: *Graphics Processing Units*
GSMA: *Global System for Mobile Association*
HPLMN: *Home Public Land Mobile Network*
HTTP/2: *Hypertext Transfer Protocolo 2.0*
IA: inteligência artificial
ICIC: Coordenação de Interferência Inter-celular - *Inter-cell Interference Coordination*
IEC: *International Electrotechnical Commission*
IEEE SDN: *Institute of Electrical and Electronics Engineers – Software Defined Networks*
IHS: *Information Handling Services*
IMS: IP *Multimedia Subsystem*
IMSI: *International Mobile Subscribe Identity*
IoT: Internet das Coisas - *Internet of Things*
IP: Protocolo da Internet - *Internet Protocol*
IpTP: Internet para Todos Peru
ISO: *International Organization for Standardization*
ITU: *International Telecommunication Union*
JSON: *JavaScript Object Notation*
KPIs: indicadores chaves de desempenho - *Key Performance Indicators*
LGPD: Lei Geral de Proteção de Dados Pessoais
LMDF: *Live Migration Defense Framework*
M2M: *Machine to Machine*
MAC: Controle de Acesso ao Meio - *Media Access Control*
MANO: Gerenciamento e Orquestração - *Management and Orchestration*
MEC: computação de borda móvel - *Mobile Edge Computing*
MIMO: *Multiple Input – Multiple Output*
MIoT: *Massive Internet of Things*
MITM: *homem no meio - Man in the Middle*
ML: aprendizado de máquina - *Machine Learning*
mMTC: *Massive Machine-Type Communication*
MNOs: operadoras de redes móveis - *Mobile Network Operator*
*Near-***RT RIC:** *Near Real-Time* **RAN** *Intelligent Controller*
NESAS: *Network Equipment Security Assurance Scheme*
NFMF: *Network Function Management Function*
NFV: Virtualização das Funções de Redes - *Network Function Virtualization*

NFVI: *Network Functions Virtualization Infrastructure*

NFVO: *Network Function Virtualization Orchestration*

NG: nova geração - *New Generation*

NGFI: Interface de *Fronthaul* **de Próxima Geração -** *Next Generation Fronthaul Interface*

NG-RAN: *Next Generation Radio Access Network*

NG-SDN: *Next Generation Software Defined Network*

Non-**RT RIC:** *Non Real-Time* **RAN** *Intelligent Controller*

NSIs: orquestração das instâncias de fatias - *Network Slice Instances*

NSMF: *Network Slice Management Function*

NSSIs: gerenciamento e orquestração das sub-redes de uma fatia - *Network Slice Subnet Instances*

NSSMF: *Network Slice Subnet Management Function*

OAM: Operação, Administração e Gerência - *Operations, Administration and Maintenance*

O-CU: *Open Centralized Unit*

O-DU: *Open Distributed Unit*

OF: *Open Flow*

ONAP: *Open Network Automation Platform*

ONF: *Open Network Foundation*

ONOS: *Open Network Operating System*

OPEX: despesas operacionais - *operational expenditure*

OSS: *Operating Support System*

OTIC: centros de teste e integração - *Open Test and Integration Center*

OVS: *Open vSwitch*

P4: *Programming Protocol-independent Packet Processors*

PCC: *Policy Charging and Control*

pH: potencial hidrogeniônico

PIB: Produto Interno Bruto

PM: *Performance Management*

PMOs: objetos de monitoramento de desempenho - *Performance Monitoring Objects*

PNFs: *Physical Network Functions*

PRBs: blocos de recursos físicos - *Physical Resource Blocks*

QoE: qualidade de experiência - *Quality of Experience*

QoS: qualidade de serviço - *Quality of Service*

RANs: *Radio Access Networks*

RF: radiofrequência

RIC: *RAN Intelligent Controller*

RRHs: cabeças de rádio remotas - *Remote Radio Heads*

RRM: gerência de recursos de rádio - *Radio Resource Management*

SDN: Rede Definida por *Software - Software Defined Network*

SDR: Rádio Definido por *Software - Software Defined Radio*

SD-RAN: *Software Defined* **RAN**

SEPP: *Security Edge Protection Proxy*

SI: integrador do sistema - *System Integrator*

SIM: *Subscriber Indentity Module*

SIP: *Session Initiation Protocol*

SLAs: *Service Level Agreements*

SMO: gerenciamento e orquestração de serviços - *Service Management and Orchestration*

SONs: redes de otimização automática - *Self-Optimizing Networks*

STG: *Security Task Group*

SUCI: *Subscriber Concealed Identifier*

SUPI: *Subscriber Permanet Identifier*

TI: Tecnologia da Informação

TIC: Tecnologia da Informação e Comunicações

TIFG: *Test and Integration Focus Group*

TIP: *Telecom Infra Project*

TLS: *Transport Layer Security*

UAVs: veículos aéreos não tripulados - *Unmanned Aerial Vehicles*

UEs: equipamentos do usuário - *User Equipments*

URLLC: *Ultra-Reliable and Low Latency Communications*

V2V: *Vehicle to Vehicle*

V2X: *Vehicle-to-Everything*

VES: *Virtual Event Streaming*

VIM: *Virtualized Infraestructure Manager*

VM: máquinas virtuais - *Virtual Machine*

VNF: Função de Rede Virtualizada - *Virtualized Network Function*

VNFM: *Virtualized Network Function Manager*

VNFO: *Virtual Network Function Orquestrator*

VPLMN: *Visited Public Land Mobile Network*

V-RAN: Rede de Acesso a Rádio Virtual - *Virtual Radio Access Network*

XaaS: orientação a serviços - *Everything as a Service*

xApps: aplicações

REFERÊNCIAS BIBLIOGRÁFICAS:

SEÇÃO 1

[1] IHS, "*IHS Economics & IHS Technology Report: 5G Economy - How 5G Technology will Contribute to the Global Economy*"**.**

[2] ITU-R, "*ITU Radiocommunication Sector - Recommendation M.2083-0*"**, 2015.**

[3] GSMA, "*Study on Socio-Economic Benefits of 5G Services Provided in mmWave Bands*"**, 2018.**

[4] 5GPPP, "*5G Vision*"**, 2015. [Online]. Disponível em:** https://5g-ppp.eu/wpuploads/2015/02/5G-Vision-Brochure-v1.pdf**.**

[5] BNDES, "*Produto 7B: Aprofundamento de Verticais: Saúde*"**, 2017.**

[6] BNDES, "*Produto 7A: Aprofundamento de Verticais: Cidades*"**, 2017.**

[7] GSMA - GSM Association, "*The Mobile Economy - Latin America and the Caribbean 2018*"**, 2018.**

[8] ITU-T, "*The Tactile Internet: ITU-T technology watch report*"**, 2014.**

[9] Philbeck, I., "*Connecting the Unconnected: Working together to achieve Connect 2020 Agenda Targets*"**, International Telecommunications Union, Genebra, 2017.**

[10] 5GRange, "*5G Range: Remote Area Access Networks for 5th Generation*"**, EU-Brazil Joint Call H2020 EUB 2017, 2017.**

[11] Cepea - Centro de Estudos Avançados em Economia Aplicada - ESALQ-USP, "*Índices: Exportação do Agronegócio, 2017*"**, 2017.**

[12] Cepea - Centro de Estudos Avançados em Economina Aplicada - ESALQ-USP, "*Participação do Agronegócio no PIB do Brasil, 2017*"**, 2017.**

[13] BNDES, "*Relatório Produto 7C: Aprofundamento de Verticais - Rural, Dezembro de 2017*"**, 2017.**

[14] Larsson, E. G. et. al., "*Massive MIMO for next generation wireless systems*"**, IEEE Communications Magazine, February 2014, 186-195.**

[15] Yifei, Y. and Longming, Z., "*Application scenarios and enabling technologies for 5G*"**, China Communications, November, 2014, 69-79.**

[16] Wei, L. et al., "*Key elements to enable millimeter wave communications for 5G wireless systems*"**, IEEE Wireless Communications, December 2014, 136-143.**

[17] Ge, X. et al., "*5G ultra-dense cellular networks*"**, IEEE Wireless Communications, Vol. 23, Issue 1, February 2016, 72-79.**

[18] Kamel, M. et. al., "*Ultra-dense networks: a survey*"**, IEEE Communications Surveys & Tutorials, Vol. 18, issue 4, 2016, 2522-2545.**

[19] Jungnicker, V. et. al., "*The role of small cells, coordinated multipoint, and massive MIMO in 5G*"**, IEEE Communications Magazine, May 2014, 44-51.**

[20] Maeder, A. et. al., "*A scalable and flexible radio access network architecture for fifth generation mobile networks*"**, IEEE Communications Magazine, November 2016, 16-23.**

[21] Marsch, P. et. al., "*5G radio access network architecture: design guidelines and key considerations*"**, IEEE Communications Magazine, November 2016, 24-32.**

[22] Manzalini, A. et. al., "*Towards 5G software-defined ecosystems*"**, IEEE SDN White Paper, 2014.**

[23] Khan, R. et. al., "*A Survey on Security and Privacy of 5G Technologies: Potential Solutions, Recent Advancements, and Future Directions*"**, IEEE Communications Surveys and Tutorials, Vol. 22, No 1, First Quarter 2020, 196-248.**

SEÇÃO 2

[24] Chih-Lin, I. et al., "*Toward green and soft: A 5G perspective*"**, IEEE Communications Magazine 52.2 (2014): 66-73.**

[25] Salman, T., "*Cloud RAN: Basics, Advances and Challenges - A Survey of C-RAN Basics Virtualization, Resource Allocation, and Challenges*"**, (2016): 1-16.**

[26] Wu, J. et al., "*Cloud radio access network (C-RAN): a primer*"**, IEEE network 29.1 (2015): 35-41.**

[27] Gavrilovska, L., Rakovic, V. and Denkovski, D., "*From Cloud RAN to Open RAN*"**, Wireless Pers Commun 113, 1523–1539 (2020).** https://doi.org/10.1007/s11277-020-07231-3

[28] Chih-Lin, I. et al., "*Recent progress on C-RAN centralization and cloudification*"**, IEEE Access 2 (2014): 1030-1039.**

[29] Tonini, F. et al., "*Cost-optimal deployment of a C-RAN with hybrid fiber/FSO fronthaul*"**, Journal of Optical Communications and Networking 11.7 (2019): 397-408.**

[30] Fajardo, J. O., Taboada, I., and Liberal, F., "*Radio-aware service-level scheduling to minimize downlink traffic delay through mobile edge computing*"**, International Conference on Mobile Networks and Management. Springer, Cham, 2015.**

[31] Rabia, T., Braham, O., and Pujolle, G., "*Partially centralized C-RAN architecture using TRILL protocol*", 2016 7th International Conference on the Network of the Future (NOF), IEEE, 2016.

[32] Checko, A. et al., "*Cloud Radio Access Network architecture. Towards 5G mobile networks*", Technical University of Denmark, 2016.

[33] Alabbasi, A., Wang, X., and Cavdar, C., "*Optimal processing allocation to minimize energy and bandwidth consumption in hybrid CRAN*", IEEE Transactions on Green Communications and Networking 2.2 (2018): 545-555.

[34] Hoffmann, M., and Staufer, M., "*Network virtualization for future mobile networks: General architecture and applications*", 2011 IEEE international conference on communications workshops (ICC). IEEE, 2011.

[35] Dawson, A. W., Marina, M. K., and Garcia, F. J., "*On the benefits of RAN virtualization in C-RAN based mobile networks*", 2014 Third European Workshop on Software Defined Networks. IEEE, 2014.

[36] Agrawal, R. et al., "*Cloud RAN challenges and solutions*", Annals of Telecommunications 72.7 (2017): 387-400.

[37] Sucasas, V., Mantas, G., and Rodriguez, J., "*Security challenges for cloud radio access networks*", Backhauling/Fronthauling for Future Wireless Systems (2016): 195-211.

[38] Checko, A. et al., "*Cloud RAN for mobile networks - A technology overview*", IEEE Communications surveys and tutorials 17.1 (2014): 405-426.

[39] Ghoreishi, S. E. et al., "*A cost-driven approach to caching-as-a-service in cloud-based 5G mobile networks*", IEEE Transactions on Mobile Computing 19.5 (2019): 997-1009.

[40] Parallel Wireless, "*Everything you need to know about Open Ran*", 2020. [Online]. Disponível em: https://www.parallel-wireless.com/wp-content/uploads/Parallel-Wireless-e-Book-Everything-You-Need-to-Know-about-Open-RAN.pdf

[41] Chaudhary, J. K., Zou, J., and Fettweis, G., "*Cost saving analysis in capacity-constrained C-RAN fronthaul*", 2018 IEEE Globecom Workshops (GC Wkshps). IEEE, 2018.

[42] Rendon Schneir, J. et al., "*A business case for 5G mobile broadband in a dense urban area*", Telecommunications Policy 43.7 (2019): 1-1.

[43] Dzogovic, B., Feng, B., and Van Do, T., "*Building virtualized 5G networks using open source software*", 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). IEEE, 2018.

[44] Chang, H. C. et al., "*Performance evaluation of Open5GCore over KVM and Docker by using Open5GMTC*", NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium. IEEE, 2018.

[45] Lee, C. N. et al., "*A feasible 5G cloud-ran architecture with network slicing functionality*", 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, 2018.

[46] Luo, Y. C. et al., "*A computation workload characteristic study of C-RAN*", 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS). IEEE, 2018.

[47] Gavrilovska, L., Rakovic, V., and Denkovski, D., "*Aspects of resource scaling in 5G-MEC: Technologies and opportunities*", 2018 IEEE Globecom Workshops (GC Wkshps). IEEE, 2018.

[48] Abbas, N. et al., "*Mobile edge computing: A survey*", IEEE Internet of Things Journal 5.1 (2017): 450-465.

[49] Solozabal, R. et al., "*Exploitation of mobile edge computing in 5G distributed mission-critical push-to-talk service deployment*", IEEE Access 6 (2018): 37665-37675.

[50] Xia, W. et al., "*Programmable hierarchical C-RAN: From task scheduling to resource allocation*", IEEE Transactions on Wireless Communications 18.3 (2019): 2003-2016.

[51] Coronado, E. et al., "*5G-EmPOWER: A software-defined networking platform for 5G radio access networks*", IEEE Transactions on Network and Service Management 16.2 (2019): 715-728.

[52] Singh, S. K., Singh, R., and Kumbhani, B., "*The Evolution of Radio Access Network Towards Open-RAN: Challenges and Opportunities*", 2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW). IEEE, 2020.

SEÇÃO 3

[53] 3GPP, "*Technical Specification Group Services and System Aspects; Telecommunication management; Study on network management of virtualized networks*", Release 13, 2015.

[54] O-RAN Technical Report, "*O-RAN Working Group 1; Study on O-RAN Slicing*", O-RAN.WG1.Study-on-O-RAN-Slicing-v02.00, 2020, https://www.o-ran.org/s/O-RANWG1Study-on-O-RAN-Slicing-v0200.pdf

[55] O-RAN Software Community, "*Open & Intelligent Software for the Radio Access Networks*", 2021. [Online]. Available: https://www.o-ran.org/software. [Acesso em 01 2021].

[56] Open Networking Foundation, "*SDN architecture*", 2016.

[57] Open Network Foundation, "*TR-526: Applying SDN architecture to 5G slicing*", 2016.

SEÇÃO 4

[58] **Russell, S., and Norvig, P.,** "*Artificial Intelligence: A Modern Approach*", **Upper Saddle River, NJ: Prentice Hall, 2010.**

[59] **McCorduck, P.,** "*Machines who think: A personal inquiry into the history and prospects of artificial intelligence*", **CRC Press, 2004.**

[60] **Mitchell, T. M.,** "*Machine Learning*", **McGraw Hill, 1997.**

[61] **Mockel, P., and Makala, B.,** "*Artificial Intelligence and 5G Mobile Technology Can Drive Investment Opportunities in Emerging Markets*", **International Financial Corporation (IFC), December 2019.**

[62] **5G Americas,** "*5G At The Edge*", **October 2019. [Online]. Available:** https://www.5gamericas.org/5g-at-the-edge/

[63] **5G Americas,** "*The 5G Evolution: 3GPP Releases 16-17*", **January 2020. [Online]. Available:** https://www.5gamericas.org/5g-evolution-3gpp-releases-16-17

[64] **5G Americas,** "*Precision Planning for 5G Era Networks with Small Cells*", **October 2019. [Online]. Available:** https://www.5gamericas.org/precision-planning-for-5g-era-networks-with-small-cells/

[65] **Churchill, L.,** "*Areas where AI Is Transforming the Telecom Industry*", **TechSee, December 2020. [Online]. Available:** https://techsee.me/blog/artificial-intelligence-in-telecommunications-industry/

[66] **Kone?ný, J. et al.,** "*Federated learning: Strategies for improving communication efficiency*", **arXiv preprint arXiv:1610.05492, 2016.**

[67] **McMahan, B. et al.,** "*Communication-efficient learning of deep networks from decentralized data*", **In Artificial Intelligence and Statistics, pp. 1273-1282. PMLR, 2017.**

[68] **Bird, E. et al.,** "*The ethics of artificial intelligence: Issues and initiatives*", **European Parliamentary Research Service, March 2020.**

[69] **ISO,** "*ISO/IEC TS 27008:2019 Information technology — Security techniques — Guidelines for the assessment of information security controls*", **January 2019. [Online]. Available:** https://www.iso.org/standard/67397.html

[70] **Zhao, W. S. et al.,** "*Nanotube devices based crossbar architecture: toward neuromorphic computing*", **Nanotechnology 21, no. 17 (2010): 175202.**

[71] **Monroe, D.,** "*Neuromorphic computing gets ready for the (really) big time*", **(2014): 13-15.**

[72] **Wu, H., and Dai, Q.,** "*Artificial intelligence accelerated by light*", **Nature, Jan 2021. [Online]. Available:** https://www.nature.com/articles/d41586-020-03572-y

[73] **Yao, M., Sohul, M., Marojevic, V., and Reed, J. H.,** "*Artificial Intelligence Defined 5G Radio Access Networks*", **in** *IEEE Communications Magazine*, **vol. 57, no. 3, pp. 14-20, doi: 10.1109/MCOM.2019.1800629, March 2019.**

SEÇÃO 5

[74] **Zhang, S., Wang, Y., and Zhou, W., "Towards secure 5G networks: A Survey", Comput. Networks, vol. 162, p. 106871, 2019, doi: 10.1016/j.comnet.2019.106871.**

[75] **Huazhang, L., Zhonghao, Z., and Shuai, G.,** "*5G Edge Cloud Networking and Case Analysis*", **Int. Conf. Commun. Technol. Proceedings, ICCT, pp. 617–621, 2019, doi: 10.1109/ICCT46805.2019.8947289.**

[76] **Degefa, F. B., Lee, D., Kim, J., Choi, Y., and Won, D.,** "*Performance and security enhanced authentication and key agreement protocol for SAE/LTE network*", **Comput. Networks, vol. 94, pp. 145–163, 2016, doi: 10.1016/j.comnet.2015.11.014.**

[77] **Bandyopadhyay, P. R.,** "*A novel large slosh-or-spin low-speed underwater propulsor bridges the unsteady and steady propulsion mechanisms of nature and engineering*", **IEEE J. Ocean. Eng., vol. 41, no. 4, pp. 868–881, 2016, doi: 10.1109/JOE.2015.2497879.**

[78] **T. Specification and G. Services,** "*3gpp tr 21.915*", **vol. 0, no. Release 15, 2019.**

[79] **Alhebaishi, N., Wang, L., and Jajodia, S.,** "*Modeling and mitigating security threats in network functions virtualization (NFV)*", **vol. 12122 LNCS. Springer International Publishing, 2020.**

[80] **N. F. Virtualization,** "*NFV Framework and Major Components*", **vol. 12, no. 4, pp. 38–50, 2020.**

[81] **I. Corporation,** "*Intel® Trusted Execution Technology (Intel® TXT) Enabling Guide Intel® Trusted Execution Technology (Intel® TXT) Enabling Guide | 2*", **pp. 1–32, 2015. [Online]. Available: https://software.intel.com/sites/default/files/Intel_R__TXT_Enabling_Guide.pdf.**

[82] **Biedermann, S., Zittel, M., and Katzenbeisser, S.,** "*Improving security of virtual machines during live migrations*", **2013 11th Annu. Conf. Privacy, Secur. Trust. PST 2013, pp. 352–357, 2013, doi: 10.1109/PST.2013.6596088.**

[83] **3GPP TS 33.501 V17.0.0 (2020-12),** "*Technical Specification Group Services and System Aspects; Security architecture and procedures for 5G system (Release 17)*", **2020.**

[84] Dastres, R., and Soori, M., "*Impact of Meltdown and Spectre on CPU Manufacture Security Issues*", International Journal of Engineering and Future Technology, v. 18, n. 2, p. 62-69, 2020.

[85] Olimid, R. F., and Nencioni, G., "*5G Network Slicing: A Security Overview*", IEEE Access, vol. 8, pp. 99999–100009, 2020, doi: 10.1109/ACCESS.2020.2997702.

[86] Henda, N. B., "*Overview on the security in 5G phase 2*", J. ICT Stand., vol. 8, no. 1, pp. 1–14, 2019, doi: 10.13052/jicts2245-800X.811.

SEÇÃO 6

SEÇÃO 6.1

[87] 3GPP Technical Specification 38.401, "*NG-RAN; Architecture description (Release 15)*", 2017.

[88] Open Network Foundation, https://opennetworking.org/.

[89] SD-RAN, "*ONF's Software-Defined RAN Platform Consistent with the O-RAN Architecture*", White Paper, agosto 2020, acessível em: https://opennetworking.org/wp-content/uploads/2020/08/SD-RAN-v2.0.pdf

[90] OpenRAN, "*Telecom Infrastructure Project*", 2020, acessível em: https://telecominfraproject.com/openran/

[91] O-RAN Technical Report, "*Cloud Architecture and Deployment Scenarios for O-RAN Virtualized RAN*", O-RAN-WG6.CAD-V01.00.00, 2019, https://www.o-ran.org/s/O-RAN-WG1OAM-Architecture-v0200.pdf

[92] Bonati, L., Polese, M., D'Oro, S., Basagni, S., and Melodia, T., "*Open, Programmable, and Virtualized 5G Networks: State-of-the-Art and the Road Ahead*", Computer Networks, Volume 182, 2020, 107516, https://doi.org/10.1016/j.comnet.2020.107516.

[93] O-RAN Technical Specification, "*O-RAN Working Group 1; Slicing Architecture*", O-RAN.WG1.Slicing-Architecture-v01.00, 2020, https://www.o-ran.org/s/O--RANWG1Slicing-Architecture-v0100.pdf

[94] 3GPP Technical Report 28.801, "*Telecommunication management; Study on management and orchestration of network slicing for next generation network*", Release 14, 2018, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3091

[95] 3GPP Technical Specification 28.541, "*Management and orchestration; 5G Network Resource Model (NRM); Stage 2 and stage 3*", Release 15, 2020, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3400

[96] Rosa, R. V., Rothenberg, C. E., Peuster, M., and Karl, H., "*Methodology for VNF Benchmarking Automation*", Internet draft-rosa-bmwg-vnfbench-06, outubro de 2020.

[97] da Rosa Righi, R. et al., "*Elastic-RAN: An Adaptable Multi-level Elasticity Model for Cloud Radio Access Networks*", Computer Communications, Volumes 142–143, 2019, Pages 34-47, ISSN 0140-3664, https://doi.org/10.1016/j.comcom.2019.04.012.

[98] O-RAN Alliance White Paper, "*O-RAN use cases and deployment scenarios*", Tech. Rep., Feb. 2020. [Online]. Available: https://www.o-ran.org/resources

SEÇÃO 6.2

[99] Parsons, E., and Foglander, G., "*The four key components of Cloud RAN*", Ericsson, 2020. [Online]. Available: https://www.ericsson.com/en/blog/2020/8/the-four-components-of-cloud-ran

[100] Zhang, C., Patras, P., and Haddadi, H., "*Deep Learning in Mobile and Wireless Networking: A Survey*", IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2224–2287, March 2019.

[101] Luong, N. C. et al., "*Applications of Deep Reinforcement Learning in Communications and Networking: A Survey*", IEEE Communications Surveys & Tutorials, vol. 21, no. 4, pp. 3133–3174, May 2019.

[102] Niknam, S. et al., "*Intelligent O-RAN for Beyond 5G and 6G Wireless Networks*", arXiv preprint arXiv:2005.08374 (2020).

[103] Bonati, L. et al., "*Intelligence and Learning in O-RAN for Data-driven NextG Cellular Networks*", arXiv preprint arXiv:2012.01263 (2020).

[104] Rivera, P. E. I., Mollahasani, S., and Erol-Kantarci, "*Multi Agent Team Learning in Disaggregated Virtualized Open Radio Access Networks (O-RAN)*", arXiv preprint arXiv:2012.04861 (2020).

[105] O-RAN Alliance, "*O-RAN working group 2: AI/ML workflow description and requirements*", Tech. Rep., Mar. 2019.

[106] Small Cell Forum, "*SCF233: Small Cell SON and Orchestration from 4G to 5G*", February 2020. [Online]. Available: https://scf.io/en/documents/233_Small_Cell_SON_and_Orchestration_from_4G_to_5G.php

[107] Yan, M., Feng, G., Zhou, J., Sun, Y., and Liang, Y. C., "*Intelligent resource scheduling for 5G radio access network slicing*", IEEE Transactions on Vehicular Technology, vol. 68, no. 8, pp. 7691– 7703, 2019.

[108] Shenbagaraman, G., "*Who disaggregated my RAN? Part 4: Open RAN: Disaggregated and smart*", RCRWireless, June 2020. [Online]. Available: https://www.rcrwireless.com/20200623/open_ran/

who-disaggregated-my-ran-part-4-open-ran-disaggregated-and-smart

[109] Ekudden, E., "*Artificial intelligence in RAN - a software framework for AI-driven RAN automation*", **Ericsson Reports and Papers, December 2020. [Online]. Available:** https://www.ericsson.com/en/reports--and-papers/ericsson-technology-review/articles/artificial-intelligence-in-ran

[110] Ekudden, E., "*Enhancing RAN performance with AI*", **Ericsson Reports and Papers, January 2020. [Online]. Available:** https://www.ericsson.com/en/reports--and-papers/ericsson-technology-review/articles/enhancing-ran-performance-with-ai

SEÇÃO 6.3

[111] Zwarico, S. J. A., "*The O-RAN ALLIANCE Security Task Group Tackles Security Challenges on All O-RAN Interfaces and Components*", **O-RAN Alliance, 10 2020. [Online]. Available: https://www.o-ran.org/blog/2020/10/24/the-o-ran-alliance-security-task-group-tackles-security-challenges-on-all--o-ran-interfaces-and-components. [Acesso em 01 2021].**

[112] de Carné de Carnavalet, X., and van Oorschot, P. C., "*A survey and analysis of TLS interception mechanisms and motivations*", **2020. [Online]. Available: http://arxiv.org/abs/2010.16388.**

[113] Hawanna, V., Kulkarni, V. Y., Rane, R. A., Mestri, P., and Panchal, S., "*Risk Rating System of X.509 Certificates*", **Procedia Comput. Sci., vol. 89, pp. 152–161, 2016, doi: 10.1016/j.procs.2016.06.027.**

[114] GSM Association, "*Network Equipment Security Assurance Scheme - Overview. FS.13*", **pp. 1–23, 2019. [Online]. Available: https://www.gsma.com/security/wp-content/uploads/2019/11/FS.13-NESAS-Overview-v1.0.pdf.**

SEÇÃO 6.4

[115] O-RAN Alliance. [Online]. Available: https://www.o-ran.org. [Acesso em 01 2021].

[116] iGR, "*OpenRAN Integration: Run with it*", **04 2020. [Online]. Available: https://www.parallelwireless.com/wp-content/uploads/iGR-OpenRAN-Integration-White--Paper.pdf. [Acesso em 01 2021].**

[117] Davies, J., "*Rakuten still sailing smoothly on its disruptive mission*", **telecoms.com, 08 2020. [Online]. Available: https://telecoms.com/506005/rakuten-still-sailing-smoothly-on-its-disruptive-mission/. [Acesso em 01 2021].**

[118] Paolini, M., "*How much can operators save with a Cloud RAN? A TCO model for virtualized and distributed RAN*", **2017.**

[119] Deloitte, "*Multivendor network archtectures, TCO and operational risk*", **2012.**

[120] Brown, G., "*TIP OpenRAN: Toward Disaggregated Mobile Networking*", **2019.**

[121] Nery, C., "*Tim e Vivo deverão liderar estratégia Open RAN no Brasil*", **Tele Síntese, 10 2020. [Online]. Available: https://www.telesintese.com.br/tim-e-vivo-deverao-liderar-estrategia-openran-no-brasil/. [Acesso em 01 2021].**

[122] Convergência digital, "*Algar Telecom testa Open RAN em centro de 5G criado pela IBM, Flex e FIT*", **12 2020. [Online]. Available: https://www.convergenciadigital.com.br/cgi/cgilua.exe/sys/start.htm?UserActiveTemplate=site&infoid=55749&sid=17#.YAHMjUHPyUk. [Acesso em 01 2021].**

[123] Tutela, "*Disruption with O-RAN: Rakuten's mobile experience so far*", **10 2020. [Online]. Available: https://www.tutela.com/blog/disruption-through-o-ran-rakutens-mobile-experience-so-far-2020. [Acesso em 01 2021].**

[124] Wenger, E., "*Security in Open RAN Networks*", **Cisco, 09 2020. [Online]. Available:** https://blogs.cisco.com/gov/security-in-open-ran-networks.

Inatel - Brazil

Radiocommunications Reference Center

CxSC Telecom
Centro de Segurança Cibernética

Inatel