



Programa de Pós-Graduação em
Computação Aplicada
Mestrado/Doutorado Acadêmico

Fernando Zanferrari Morais

PlaceRAN - Uma solução de posicionamento das funções de rádio de acesso móveis virtualizadas de quinta geração

São Leopoldo, 2021

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA
NÍVEL MESTRADO

FERNANDO ZANFERRARI MORAIS

PLACERAN - UMA SOLUÇÃO DE POSICIONAMENTO DAS FUNÇÕES DE
RÁDIO DE ACESSO MÓVEIS VIRTUALIZADAS DE QUINTA GERAÇÃO

SÃO LEOPOLDO
2021

Fernando Zanferrari Morais

PLACERAN - UMA SOLUÇÃO DE POSICIONAMENTO DAS FUNÇÕES DE
RÁDIO DE ACESSO MÓVEIS VIRTUALIZADAS DE QUINTA GERAÇÃO

Dissertação apresentada como requisito
parcial para a obtenção do título de Mestre
pelo Programa de Pós-Graduação em
Computação Aplicada da Universidade do
Vale do Rio dos Sinos — UNISINOS

Orientador:
Prof. Dr. Rodrigo R. Righi

Co-orientador:
Prof. Dr. Cristiano B. Both

São Leopoldo
2021

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)

Morais, Fernando Zanferrari

PlaceRAN - Uma solução de posicionamento das funções de rádio de acesso móveis virtualizadas de quinta geração / Fernando Zanferrari Moraes — 2021.

107 f.: il.; 30 cm.

Dissertação (mestrado) — Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, São Leopoldo, 2021.

“Orientador: Prof. Dr. Rodrigo R. Righi, Unidade Acadêmica de Pesquisa e Pós-Graduação”.

1. redes móveis de quinta geração. 2. NG-RAN. 3. Arquitetura NFV. 4. Posicionamento de funções virtualizadas. 5. *Crosshaul*. 6. Orquestração. I. Título.

CDU 004:519.6

Bibliotecária responsável: (Silvana Dornelles Studzinski — CRB 10/2524

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) - Código de Financiamento 001 / This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

(Esta folha serve somente para guardar o lugar da verdadeira folha de aprovação, que é obtida após a defesa do trabalho. Este item é obrigatório, exceto no caso de TCCs.)

AGRADECIMENTOS

Primeiramente, agradeço à minha esposa por todo o apoio, paciência e dedicação para com a nossa família e o nosso pequeno. Sem ela, nada disso seria possível. Agradeço também aos meus pais, que sempre foram compreensivos e estiveram comigo. Sem eles, nunca estaria neste momento da vida. Ainda, agradeço aos Profs. Dr. Rodrigo da Rosa Righi e Dr. Cristiano Bonato Both, meu orientador e co-orientador, que me proporcionaram conhecimento, empenho, diretrizes e sempre me trouxeram de voltas aos trilhos, quando necessário. Agradeço também aos colegas Júlio Renner e Gustavo Zanatta, ao Prof. Dr. Leizer Pinto, e ao Prof. Dr. Kleber Cardoso e seu orientando Gabriel Faria por toda a cooperação e aprendizagem. Por fim, agradeço à CAPES pela bolsa que me permitiu acesso a pós graduação.

RESUMO

Para alcançar as transformações digitais visionadas e impulsionadas pela demanda de serviços da sociedade, os avanços das redes móveis de quinta geração, especificados pelas organizações padronizadoras em conjunto com a indústria, são amplamente disruptivos quando comparado com as redes precedentes. Para atender a escala demandada, a nova geração de redes de rádio de acesso (NG-RAN) é orientada a dois conceitos: (i) desacoplamento e desagregação das funções de rádio em até três unidades e (ii) imersão nos conceitos de *software*, principalmente, virtualização: resultando assim, na arquitetura vNG-RAN. Neste contexto, o problema do posicionamento das funções de rádio em três unidades sob as redes de transporte e as estratégias de alocação de recursos computacionais dessas funções é definido como um problema *NP-hard*. Além disso, a tomada de decisão entre a desagregação da RAN, o encaminhamento na rede de transporte e as estratégias de alocação de recursos computacionais é um grande desafio de pesquisa e com grande interesse da indústria. Sendo assim, a presente dissertação apresenta o PlaceRAN, uma solução de otimização de posicionamento com foco em planejamento de redes, aliada a um orquestrador para o desenvolvimento da vNG-RAN (RAN virtualizada). Primeiramente, a solução de otimização de posicionamento possui três estágios com os objetivos: (i) maximizar a agregação das funções de rádio e minimizar a utilização de recursos computacionais, (ii) minimizar o número de *Disaggregated RAN Combination* (DRC), (iii) priorizar os DRCs de acordo com a estratégia de posicionamento escolhida. Para o orquestrador, o mesmo é alinhado com a arquitetura *Network Function Virtualization* (NFV) e tem como objetivo: (i) prover a alocação das funções de rádio virtualizadas com base na orientação da solução de otimização e (ii) ter ciência da topologia da rede e recursos computacionais. A avaliação foi realizada com redes e parâmetros reais para a solução de otimização e para o orquestrador, um experimento baseado em uma dessas redes reais foi emulada. Os resultados apresentam que as soluções de posicionamento alcançam até 80% de agregação das unidades de rádio virtualizadas (vRU, vDU e vCU) em menos de 20% dos recursos computacionais com redução no número de DRCs orientada pela estratégia de posicionamento. Assim como, o orquestrador apresenta funcionalidade total na alocação demandada pela solução de otimização sob plataforma de orquestração de contêiner de produção. Obtendo assim, total viabilidade de desenvolvimento da vNG-RAN. Finalmente, a solução PlaceRAN contribui para o âmbito científico para o avanço da virtualização e desagregação da arquitetura NG-RAN, assim como é alinhada com as principais iniciativas da indústria.

Palavras-chave: redes móveis de quinta geração. NG-RAN. Arquitetura NFV. Posicionamento de funções virtualizadas. *Crosshaul*. Orquestração.

ABSTRACT

To achieve the digital transformations envisioned and driven by the demand for services from society, the advances in the fifth-generation mobile networks, specified by the standardizing organizations jointly with the industry, are largely disruptive compared to the previous networks. To meet the demand scale, the new generation of access radio networks (NG-RAN) is guided by two concepts: (i) radio functions decoupling and disaggregation in up to three units and (ii) immersion in software concepts, mainly virtualization. Thus resulting in the vNG-RAN architecture. In this context, placement of the radio functions in three units under the transport networks and the computational resources is defined as an NP-hard problem. Also, the decision-making between the RAN disaggregation, the routing in the transport network, and the strategies for allocating computational resources is an unprecedented research challenge and with high interest from the industry. Therefore, this dissertation presents the PlaceRAN solution: a placement optimization solution focused on network planning, combined with an orchestrator to develop vNG-RAN, i.e., a virtualized RAN. First, the placement optimization solution has three stages with the objectives: (i) maximize the aggregation of radio functions and minimize the use of computational resources, (ii) minimize the number of Disaggregated RAN Combination (DRC), (iii) prioritize DRCs according to the chosen placement strategy. For the orchestrator, it is aligned with the Network Function Virtualization (NFV) architecture. It aims to: (i) provide the allocation of virtualized radio functions based on the placement optimization solution and (ii) be aware of the network topology and computational resources. The evaluation was conducted with two real networks and parameters for the optimization solution, and for the orchestrator, an experiment based on one of these real networks was emulated. The results show that the placement solutions reach up to 80% aggregation of the virtualized radio units (vRU, vDU, and vCU) in less than 20% of the computational resources, reducing the number of DRCs guided by the placement strategy. Likewise, the orchestrator has total functionality in allocating demand for the optimization solution under the production container orchestration platform. Thus obtaining total viability for the development of vNG-RAN. Finally, the PlaceRAN solution contributes to the scientific field for the advancement of virtualization and disaggregation of the NG-RAN architecture and aligns with the main industry initiatives.

Keywords: fifth-generation of mobile networks. NG-RAN. NFV architecture. Placement of virtualized functions. *Crosshaul*. Orchestration.

LISTA DE FIGURAS

1	A evolução da NG-RAN de baseada em <i>hardware</i> para baseada em <i>software</i> é exposta. A Figura 1(a) mostra a arquitetura NG-RAN dividida nos nós CU, DU e RU baseada em <i>hardware</i> . Enquanto que, a Figura 1(b) apresenta os nós da vNG-RAN e o posicionamento em uma infraestrutura computacional (desde a centralização na nuvem até a distribuição na borda) orquestrado por uma função centralizada. . . .	22
2	Representa a evolução da arquitetura RAN, iniciando pela D-RAN (tecnologia LTE) pela sub-figura 2(a), explicitando a evolução da arquitetura C-RAN (em ambientes de computação em nuvem) e a centralização da BBU pela sub-figura 2(b) e terminando no atual desenvolvimento, a arquitetura NG-RAN, que proporciona alta distribuição, flexibilidade e aperfeiçoamento das funções de rádio pela sub-figura 2(c).	25
3	Oito opções de desagregação da pilha de protocolos proposta pela arquitetura NG-RAN para prover flexibilidade na composição dos nós de rede.	29
4	Cenários de arquiteturas suportadas pela NG-RAN, combinando as opções de desagregação da RAN e o posicionamento da rede <i>crosshaul</i> (ITU-T, 2018).	30
5	Arquitetura conceito das redes <i>crosshaul</i> (neste exemplo, formato anel, porém não há restrições para outras topologias) para o desenvolvimento da arquitetura NG-RAN.	31
6	Arquitetura referência NFV ETSI para virtualização de redes para aplicações genéricas. A arquitetura dividida nos blocos funcionais de gerenciamento e orquestração (MANO), sistema de suporte operacional e de negócios (OSS/BSS), funções de rede (VNFs) e infraestrutura computacional (NFVI) virtualizadas.	34
7	Posicionamento e encadeamento das VNFs da arquitetura vNG-RAN, assim como a conexão com a arquitetura NFV e os blocos MANO, NFVI e VNF.	38
8	A solução PlaceRAN é baseada na arquitetura referência NFV ETSI. A figura detalha, além do orquestrador inserido no bloco funcional MANO NFVO, os blocos funcionais MANO VIM e VNFM e os componentes NFVI e VNFs da arquitetura NFV, assim como os componentes das redes móveis CN e UE, além dos dados de entrada do operador de rede e das redes <i>crosshaul</i> . O orquestrador possui quatro blocos funcionais específicos. O primeiro refere-se a base denominada de Catálogo e divide-se nas bases de informações de Topologia <i>Crosshaul</i> e de Recursos NFVI, além das funções virtualizadas da RAN. Os outros três blocos funcionais são o Otimizador responsável pelo processo de otimização do posicionamento, o bloco Gerenciador que realiza toda a interação interna e externa do orquestrador e o bloco Executor que envia o plano de alocação e o encadeamento das VNFs da RAN.	52

9	O diagrama de sequência, apresenta as etapas do processo decisório proposto pela arquitetura PlaceRAN. Desde a entrada do operador de rede até a alocação da função de rádio virtualizada.	56
10	Combinações de desagregação da RAN (DRC - <i>Disaggregated RAN Combination</i>) para o desenvolvimento das funções de rádio virtualizadas.	57
11	Apresenta exemplo do funcionamento dos estágios desenvolvidos para a formulação do problema. A Figura 11a representa a entrada de RUs no Otimizador e a infraestrutura computacional conectada pelas redes <i>crosshaul</i> . A Figura 11b apresenta solução do estágio 1 com a resultante de 01 DRC1, 01 DRC2 e 01 DRC13. A Figura 11c mostra a resultante do estágio 2 com a redução na quantidade de DRCs para 2 (02 DRC1 e 01 DRC13). Por fim, a Figura 11d exibe a resultante final, com a priorização por pesos dos DRCs (02 DRC2 e 01 DRC12). . . .	61
12	Topologias para avaliação conceitual do Otimizador de Posicionamento.	68
13	Protótipo desenvolvido para a solução de Orquestração PlaceRAN. Os sub-blocos Executor e Gerenciador são desenvolvidos sob a funcionalidade do K8S operadores. O Otimizador utiliza as mesmas plataformas e <i>software</i> , porém virtualizada em um contêiner. O sub-bloco Otimizador é desenvolvido sob um contêiner virtualizado. Catálogo RAN VNFs é suportado pela plataforma Docker Registry. Os Catálogo Topologia e Recursos NFVI, assim como a entrada RU, são concebidos como um <i>ConfigMap</i> sob um arquivo JSON. A RAN foi emulada pela ferramenta OAI e o CN pela ferramenta Free5GC. Toda a infraestrutura computacional NFVI é gerenciada pela plataforma K8S no formato nativo de <i>cluster</i>	73
14	Interface de Posicionamento e Função para o novo posicionamento. . .	74
15	Exemplo de objeto do Gerenciador.	75
16	Exemplo de saída do Executor para o K8S	75
17	Análise da relação entre quantidade de RCs utilizados e o nível de agregação das funções de rádio virtualizadas.	80
18	Escolhas dos DRCs do Otimizador para as duas topologias e os três cenários.	82
19	Avaliação dos estágios do modelo versus recursos de rede.	85
20	Exemplo de alocação de um encadeamento de VNFs na ferramenta K8S.	88
21	Consumo médio e desvio padrão da métrica de processamento CPU dos recursos do cluster (25 RCs) em função do tempo.	90
22	Consumo médio e desvio padrão da métrica de memória RAM dos recursos do cluster (25 RCs) em função do tempo.	91
23	Apresenta o consumo do recurso de CPU médio e de pico de todos os nós RCs.	92
24	Consumo incremental dos recursos de RAM de todos os nós RCs. . . .	93

LISTA DE TABELAS

1	Requisitos de referência da rede <i>crosshaul</i> (valores máximos suportados pela RAN e dependentes da demanda da carga da rede) com base na desagregação da arquitetura NG-RAN (3GPP, 2017a)	31
2	Padrão de Orquestradores, adaptado de (MAMUSHIANE et al., 2019).	41
3	Trabalhos relacionados - Soluções de orquestração para a virtualização e desagregação da arquitetura NG-RAN.	42
4	Trabalhos Relacionados - Soluções de otimização do posicionamento da RAN.	45
5	Resumo das notações da formulação do problema	59
6	Cenários para a avaliação da solução de otimização de posicionamento.	68
7	Utilização de CPU por core da pilha de protocolos da RAN (HU; WANG, 2019; KAZUNARI, 2018; YEOH et al., 2016).	69
8	Parâmetros de distância e quantidade de saltos da topologia T2	70
9	Parâmetros das opções de desagregação avaliadas.	71
10	Parâmetros das opções de desagregação avaliadas para a tecnologia LTE (FORUM, 2016).	78
11	Nível de agregação por nó na topologia T1.	81
12	Nível de agregação por nó na topologia T2.	82
13	Análise dos 3 estágios do Modelo do Otimizador da topologia T1 cenário CM.	84
14	Escolhas dos DRCs pelo Otimizador para a avaliação experimental. . .	89
15	Consumo computacional com base na alocação das VNFs da RAN por Pods.	94

LISTA DE SIGLAS

3GPP	3rd Generation Partnership Project
BBU	BaseBand Unit
BS	Base Station
CN	Core Network
C-RAN	Centralized Radio Access Network
CU	Central Unit
DU	Distributed Unit
ETSI	European Telecommunications Standards Institute
ITU-T	International Telecommunications Union - Telecommunication Standardization
LTE	Long Term Evolution
MAC	Medium Access Control
MANO	Management and Orchestration
NFV	Network Function Virtualization
NFVI	Network Function Virtualization Infrastructure
NFVO	Network Function Virtualization Orchestrator
NG-RAN	Next Generation Radio Access Network
PDCP	Packet Data Control Protocol
PHY	Physical Layer
RAN	Radio Access Network
RC	Recurso Computacional
RF	Radio Frequency
RLC	Radio Link Control
RRC	Radio Resource Control
RRH	Remote Radio Head
RU	Remote Unit
SDAP	Service Data Adaptation Protocol
SFC	Service Function Chain
UE	User Equipment
UP	User Plane
VIM	Virtualized Infrastructure Manager
VNF	Virtual Function Network

VNFM Virtual Function Network Manager

SUMÁRIO

1 INTRODUÇÃO	16
1.1 Motivação	18
1.2 Questão de Pesquisa	20
1.3 Objetivos	21
1.4 Organização do Texto	23
2 FUNDAMENTAÇÃO TEÓRICA	24
2.1 Redes de Rádio de Acesso de Próxima Geração	24
2.1.1 Evolução das Redes de Rádio de Acesso	24
2.1.2 Desagregação das Redes de Rádio de Acesso de Próxima Geração	27
2.1.3 Redes de Transporte <i>Crosshaul</i>	29
2.1.4 Requisitos de Desagregação da das Redes de Rádio de Acesso de Próxima Geração	30
2.2 Virtualização das Redes Móveis	32
2.2.1 Virtualização das Redes de Rádio de Acesso	32
2.2.2 Arquitetura NFV	33
2.2.3 Métricas da Arquitetura NFV	35
2.2.4 Posicionamento das Funções de Rede Virtualizadas	36
3 TRABALHOS RELACIONADOS	39
3.1 Metodologia para Seleção de Artigos	39
3.2 Orquestração das funções de rádio virtualizadas	40
3.2.1 Iniciativas <i>Open Source</i>	40
3.2.2 Iniciativas da Academia	42
3.3 Modelos de posicionamento para orquestração das funções de rádio desagregadas e virtualizadas	44
3.4 Análise dos Trabalhos Relacionados e Oportunidades de Pesquisa	46
4 SOLUÇÃO ORQUESTRAÇÃO E DE POSICIONAMENTO PLACERAN	50
4.1 Decisões de Projeto	50
4.2 Orquestração das Funções de Rádio Virtualizadas	51
4.2.1 Componentes da Arquitetura	51
4.2.2 Solução de Orquestração	53
4.3 Modelagem do Otimizador de Posicionamento	55
4.3.1 Contextualização do Otimizador	55
4.3.2 Modelagem do Sistema	58
4.3.3 Formulação do Problema	60
5 METODOLOGIA DE AVALIAÇÃO	67
5.1 Metodologia de Avaliação Conceitual	67
5.2 Metodologia de Avaliação Experimental	72
5.2.1 Protótipo	72
5.2.2 Estudo de Caso	76

6	RESULTADOS	79
6.1	Resultados da análise conceitual do Otimizador de posicionamento	79
6.1.1	Recursos Computacionais versus Nível de Agregação	79
6.1.2	Definição dos DRCs	81
6.1.3	Impacto dos Recursos de Rede	85
6.1.4	Considerações Finais	86
6.2	Resultados do experimento da Solução PlaceRAN	87
6.2.1	Funcionalidade do Estudo de Caso da Solução PlaceRAN	88
6.2.2	Análise do comportamento dos recursos computacionais	89
6.2.3	Considerações Finais	94
7	CONCLUSÃO	96
7.1	Contribuições	98
7.2	Limitações	99
7.3	Trabalhos Futuros	99
7.4	Publicações	100
	REFERÊNCIAS	101

1 INTRODUÇÃO

A quinta geração de redes móveis conta com a atenção do mundo não apenas voltada para os quesitos evolutivos tecnológicos, mas também econômicos e de estratégias nacionais, trazendo profundas transformações para atender as crescentes demandas da sociedade (RADUKIĆ et al., 2019; SHARMA, 2019). A estrutura base das redes móveis, independentemente da geração (quinta geração ou anteriores), é constituída pelo núcleo de rede (CN - *Core Network*), pela rede de acesso de rádio (RAN - *Radio Access Network*), pelas redes de transporte e pelos dispositivos dos usuários (UE - *user equipment*). De forma resumida, o CN é responsável pela gestão dos UEs e tem um posicionamento geográfico centralizado, sendo suas principais funções: a sinalização, o armazenamento dos perfis e das políticas e a conexão dos serviços com as redes exteriores (por exemplo, a Internet). Por outro lado, a RAN possui como essência o gerenciamento da interface aérea (sem fios), através de estações rádio base (BS - *Base Station*) amplamente distribuídas geograficamente. Assim, alcança o objetivo de garantir a conexão sem fio dos UEs e, consequentemente, dos serviços disponibilizados. Por fim, o elo de comunicação entre o CN e a RAN são as redes de transporte, denominadas de redes *backhaul* (AL-DULAIMI et al., 2018; KREHER; GAENGER, 2010).

Para alcançar as transformações digitais visionadas pela quinta geração de redes móveis, as principais organizações padronizadoras de telecomunicações, o *3rd Generation Partnership Project* (3GPP), o *International Telecommunication Union-Telecommunication* (ITU-T) e o *European Telecommunications Standards Institute* (ETSI) especificam e recomendam relevantes disruptões quando comparadas com as duas arquiteturas antecessoras: a RAN distribuída (D-RAN) e a RAN centralizada na nuvem (C-RAN - *Cloud-RAN*) (3GPP, 2017b; ITU-T, 2017), visto que a D-RAN (em funcionamento nas redes móveis) opera com BS monolíticas baseadas em *hardware* proprietários, com todas as funções de rádio em um único dispositivo (a pilha de protocolos inteira de uma BS), limitando as evoluções da interface aérea, a eficiência computacional e os novos serviços. Sob outra perspectiva, a C-RAN propõe a centralização das unidades de banda base (BBU - *Baseband Unit*) e a distribuição da cabeça remota de rádio (RRH - *Remote Radio Head*). No entanto, o desenvolvimento da arquitetura é limitado, principalmente devido à requisição de altas taxas de dados para comunicação entre BBU e RRH com introdução das redes de transporte *fronthaul* (AGRAWAL et al., 2017).

Em consequência das limitações encontradas nas arquiteturas D-RAN e C-RAN, o assunto teve a atenção atraída pela indústria e pelas operadoras de redes móveis, resultando em profundas investigações. Atualmente, as mais promissoras investigações são baseadas em soluções abertas chamadas Open RAN (O-RAN), lideradas pelos grupos Telecom

Infra Project (TIP), O-RAN Alliance e Small Cell Forum (SCF) (ALLIANCE, 2020a). De forma específica, dentre as investigações em andamento, a imersão em conceitos de *software* (por exemplo, virtualização e redes programáveis) possuem alta relevância, pois direcionam para conexão com as melhores práticas de Tecnologia da Informação (TI), tais como: (i) hardware neutro de uso geral, (ii) interoperabilidade, (iii) agilidade no desenvolvimento de novos produtos para o mercado (*time-to-market*), (iv) economia de custo total de propriedade (TCO - *total cost of ownership*), (v) desenvolvimento contínuo (DevOps - *carrier development and operations*) e (iv) automação (ALLIANCE, 2020a; GSMA, 2018).

Fundamentadas nas limitações das arquiteturas predecessoras e nas investigações da indústria, duas relevantes transformações são propostas para alavancar a RAN: (i) a nova geração de redes de rádio de acesso (NG-RAN - *Next Generation Radio Access Network* especificada pelo *Release 15* da 3GPP e refinada pelo ITU-T, e (ii) o desenvolvimento nativo de conceitos definidos por *software*, principalmente a virtualização da RAN (vRAN) em desenvolvimento pela ETSI e pelas iniciativas O-RAN (3GPP, 2017a; GAVRILOVSKA et al., 2020; ITU-T, 2018). Sendo assim, a união da arquitetura NG-RAN e o conceito vRAN direcionam a evolução para uma arquitetura NG-RAN virtualizada (vNG-RAN). A vNG-RAN contribui para (i) uma eficaz e massiva distribuição geográfica, (ii) para a evolução das funções de rádio e (iii) para o compartilhamento e eficiência de recursos de *hardware*, com a imersão em infraestruturas virtualizadas genéricas (GAVRILOVSKA et al., 2020).

Abordando de forma singular a primeira transformação proposta para vNG-RAN, a nova arquitetura tem como prerrogativa a desagregação das BS em três diferentes unidades de funções de rádio (dividindo os protocolos da pilha em três unidades): (i) unidade central (CU - *Central Unit*), (ii) unidade distribuída (DU - *Distributed Unit*) e (iii) unidade de rádio (RU - *Radio Unit*) (BERTENYI et al., 2018; ITU-T, 2018). Dentro do contexto de três unidades, a nova arquitetura possibilita a flexibilização da desagregação em dois pontos distintos da pilha de protocolos (composta por seis protocolos) e oito diferentes opções de divisões entre os protocolos da RAN, proporcionando, assim, diferentes opções de combinações entre as mesmas. Além disso, introduz a agregação das unidades de rádio, em que um grupo de RUs pode ser agregado em uma DU, assim como um grupo de DUs em uma CU. Para a comunicação entre os nós, a transformação trouxe mudanças nas redes de transporte com a criação das redes *midhaul* (comunicação entre CU-DU). Neste sentido, a mesma soma-se às redes já apresentadas, o *fronthaul* (comunicação entre DU-RU, mesmo conceito da C-RAN) e o *Backhaul* (comunicação CN-CU), sendo o conjunto e a integração destas redes denominada de *crosshaul* (ITU-T, 2018; LARSEN et al., 2018).

No que se refere à segunda transformação proposta, a virtualização, o principal avanço

para imersão das redes móveis é a arquitetura de virtualização das funções de rede (NFV - *Network Function Virtualization*) liderada pelo ETSI. A transformação tem como propósito a eliminação da dependência de *hardware* específico e, por consequência, desacopla as funções de rede físicas (PNF - *Physical Network Function*) tradicionais em funções de rede virtualizadas (VNF - *Virtualization Network Functions*). Assim, as VNFs são desenvolvidas sob uma infraestrutura NFV (NFVI - *NFV infrastructure*) genérica que vislumbra o compartilhamento de conjuntos de recursos computacionais, físicos ou virtuais de forma auto-provisionada, distribuída, sob demanda e com consumo de recursos de forma eficiente. Além disso, provê suporte ao desenvolvimento das VNFs na borda da rede (por exemplo, computação na borda ou *edge computing* próximos aos UEs) e centralizados em *data centers* (por exemplo, computação em nuvem ou *cloud computing*) (ETSI, 2014a; ITU-T, 2014). Sendo assim, essa transformação resulta na virtualização das funções de rádio que compõem as unidades de rádio. Dessa maneira, transforma as unidades em virtual CU (vCU), virtual DU (vDU) e virtual RU (vRU) (ou O-CU, O-DU e O-RU pela nomenclatura das iniciativas O-RAN). Em consequência disso, potencializa a administração das funções virtualizadas pela arquitetura de referência NFV com base no bloco funcional de gerenciamento e orquestração (MANO - *Management and Orchestration*) (ETSI, 2018).

1.1 Motivação

Apesar das transformações visionadas, expressivos desafios são presumidos para alcançar o aperfeiçoamento dos serviços tradicionais (por exemplo, banda larga) e o atendimento dos novos serviços baseados no suporte a requisitos de latência extremamente baixa, massivas conexões de dispositivos e elevadas taxas de dados. Sendo assim, no âmbito da arquitetura vNG-RAN a flexibilidade proposta pela desagregação da pilha de protocolos permite que as funções de rádio sejam agregadas, posicionadas (geograficamente) e que componham as unidades de rádio de uma forma que leve em consideração os recursos de rede disponíveis e os requisitos demandados pelo serviço. Sendo assim, o posicionamento das funções de rádio em uma abordagem de gerenciamento de rede, inserido dentro das ações de orquestração, é considerado vital para as redes de quinta geração alcançarem as transformações almejadas. Sem o posicionamento automatizado, alavancar o crescimento de BSs manualmente restringe a evolução dessas transformações. No entanto, o desenvolvimento da arquitetura vNG-RAN é um problema desafiador sem precedentes, uma vez que as redes de transporte *crosshaul* têm diferentes topologias e, taxa de dados e requisitos de latência restritos. Também, os recursos computacionais (RC), atualmente, possuem capacidades distintas na NFVI e a demanda de tráfego do usuário é não-uniforme.

Para o posicionamento das funções de rádio virtualizados da vNG-RAN, os direcionamentos de orquestração seguem as definições do ETSI, com a definição de que orquestrar é automatizar, gerenciar, operar uma infraestrutura computacional (*hardware* e *software*) no formato distribuído. Desse modo, o posicionamento é baseado em VNFs (VNF-P - VNF-*placement*) e coordenado por soluções de orquestração, como, por exemplo, o bloco funcional de gerenciamento de orquestração (MANO - *Management and Orchestration*) pertencente à arquitetura NFV. Atualmente, apesar da relevância da vNG-RAN nas pesquisas de redes de quinta geração, nenhum método de tomada de decisão é especificado para a VNF-P das funções de rádio virtualizadas. Por definição, o VNF-P é baseado em algoritmos de processos decisórios de melhor escolha, normalmente do tipo NP-*Hard* (ótimo ou sub-ótimo). Para execução do processo decisório de VNFs, as principais funcionalidades da VNF-P dizem respeito à alocação das VNFs em RCs e de encadeamento das VNFs, que também denomina-se encadeamento das funções de serviço (SFC - *Service Function Chain*). Ambas as funcionalidades são apoiadas por métricas de rede orientativas, tais como métricas computacionais (processamento, memória e armazenamento) e de rede (capacidade de *links* e latência) (MASDARI et al., 2016).

Para vNG-RAN, as métricas de rede orientativas possuem necessidades embasadas nas redes móveis e na tecnologia da RAN. Nesse sentido, para a melhor decisão do VNF-P, o orquestrador carece de avaliação dos requisitos da desagregação da pilha de protocolo RAN, do encaminhamento das SFCs pelas redes *crosshaul* e das estratégias dos RCs para implementação das funções e unidades de rádio. Portanto, o VNF-P ótimo avalia os requisitos de taxa de dados e de latência para cada opção de divisão entre CU-DU e DU-RU. Além disso, cada opção de divisão resulta em um custo computacional (processamento, memória e armazenamento) a ser avaliado de acordo com a escalabilidade da rede. A VNF-P deve ser ciente da ocupação da carga, encaminhamento das SFCs e alocação de recursos computacionais, elevando o volume de possibilidades para projetar a arquitetura vNG-RAN e, conseqüentemente, a métrica de tempo.

Conforme análise realizada nos trabalhos relacionados e detalhados no capítulo 3, a comunidade científica e a indústria têm empregado esforços para o desenvolvimento da orquestração das VNF-Ps das funções de rádio virtualizadas. Entretanto, majoritariamente, os trabalhos possuem desacoplamento entre a otimização do posicionamento de funções de rádio e a solução de orquestração, resultando no particionamento entre eles. Nesta perspectiva, a avaliação dos dois temas é realizada de forma particular. Para a otimização do posicionamento da vNG-RAN, a literatura apresenta poucas otimizações exatas (AROUK et al., 2018; GARCIA-SAAVEDRA et al., 2018; MURTI et al., 2020a; YUSUPOV et al., 2018) e massivo desenvolvimento de otimizações baseadas em algoritmos aproximativos ou heurísticos que levam a soluções subótimas (AROUK et al., 2017; MASOUDI et al., 2020; MOLNER et al., 2019). Ademais, abordam principalmente es-

estratégias específicas para realizar o posicionamento, tais como: (i) centralização dos nós BBU/CU da arquitetura C-RAN (BHAMARE et al., 2018; GARCIA-SAAVEDRA et al., 2018; MAHAPATRA et al., 2017); (ii) avaliação de custos financeiros (AROUK et al., 2018; MASOUDI et al., 2020); (iii) eficiência das restrições das redes *crosshaul* (majoritariamente, as redes *fronthaul*) (HARUTYUNYAN et al., 2020; MURTI et al., 2020a,b; YUSUPOV et al., 2018); e (iv) eficiência dos recursos (AROUK et al., 2017; SONG et al., 2019).

A exceção do desacoplamento entre a otimização do posicionamento e a solução de orquestração é o trabalho de Matoussi et al. (2020). No entanto, o mesmo desenvolve uma solução de orquestração para arquitetura C-RAN. Ademais, não está alinhado com a arquitetura NFV e iniciativas de orquestração experimentais de código aberto. Dentre as iniciativas de código aberto são atualmente os principais desenvolvedores de soluções virtualizadas para redes móveis. Das iniciativas de código aberto destacam-se os projetos *Open Source MANO* (OSM) e *Open Network Automation Platform* (ONAP) (MAMUSHIANE et al., 2019). No entanto, tais iniciativas ainda carecem de desenvolvimento de VNF-P e desenvolvimento específico para RAN. Outra relevante iniciativa de implementação da RAN é o projeto Mosaic5G. Este por sua vez é uma iniciativa sem fins lucrativos e promotora de contribuição entre a indústria e a academia que desenvolve a virtualização da RAN, mas sem a orquestração de VNF-P (NIKAEIN, 2018).

Fundamentada na pesquisa realizada da literatura, percebe-se uma lacuna e, consequentemente, uma oportunidade de desenvolvimento de pesquisa no tema orquestração de VNF-Ps em relação a: (i) vNG-RAN desagregada com elevadas opções de divisão funcional e desempenho na agregação da pilha de protocolo, (ii) otimização do posicionamento das funções de rádio da vNG-RAN e (iii) orquestração do posicionamento de forma especializada para a vNG-RAN. Conforme o contexto apresentado, a Seção 1.2 aborda o projeto de estudo de pesquisa desenvolvido através da questão de pesquisa a partir das lacunas e oportunidades apresentadas.

1.2 Questão de Pesquisa

Frente aos avanços requisitados das redes de quinta geração, a investigação da arquitetura vNG-RAN é uma realidade no que tange à otimização e desempenho de rede, particularmente em vista da transformação digital em curso na sociedade (RADUKIĆ et al., 2019; RESEARCH, 2019). Tendo em vista essa informação, surge a seguinte questão de pesquisa:

Como poderia ser uma solução de orquestração para o posicionamento das funções de rádio desagregados e virtualizados da rede de acesso de nova geração?

A orquestração do posicionamento de VNFs é um tópico amplamente discutido em ambientes computacionais de forma genérica, nos quais o *software* é desacoplado do *hardware*. Apesar de ampla análise, a tomada de decisão do algoritmo de posicionamento, executada pelo orquestrador, é orientada em conformidade com as características e estratégias para com as funções virtualizadas, pela alocação de recurso na infraestrutura computacional e pela comunicação de rede (ZHANG, 2018). A Figura 1 apresenta o ambiente alvo da pesquisa, bem como expõe de forma comparativa a evolução da arquitetura NG-RAN baseada em *software*. Para a vNG-RAN, a solução de orquestração do posicionamento das funções de rádio virtualizadas deve considerar: (i) a desagregação da pilha de protocolos, (ii) os recursos computacionais disponíveis para a alocação e (iii) as restrições das redes de transporte *crosshaul*.

Dessa forma, para orientar as investigações de forma específica, duas questões de pesquisa específicas são definidas. A primeira questão de pesquisa específica é endereçada para a avaliação das soluções de orquestradores, dos ambientes de implementação e das arquiteturas desenvolvidas para a vNG-RAN. Diferentemente, a segunda questão de pesquisa específica é direcionada para as soluções de modelos de otimizações e suas estratégias de tomada de decisão para com a vNG-RAN. Ambas as perguntas contribuem para atingir os objetivos do trabalho. As perguntas seguem conforme abaixo:

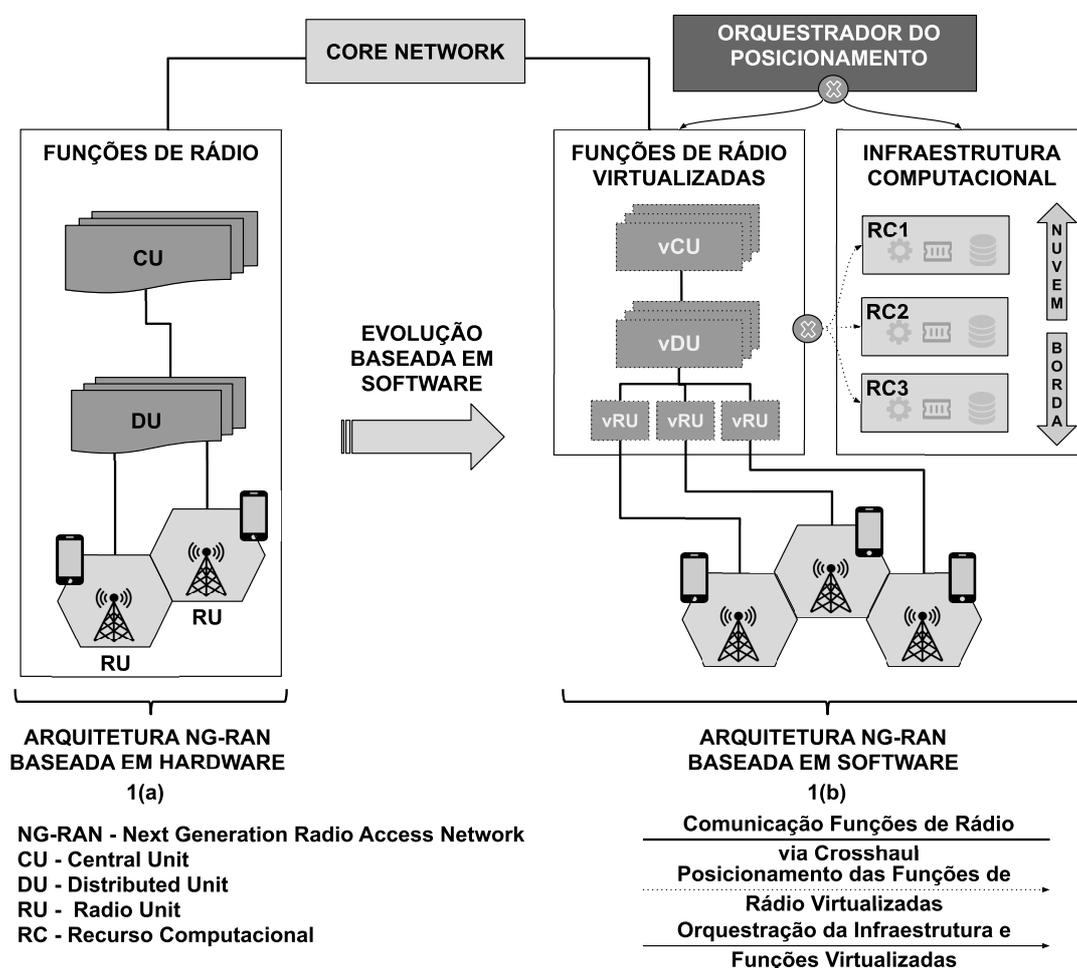
1. *Como modelar um orquestrador para o posicionamento dos rádios desagregados e virtualizados alinhado com os padrões definidos para vNG-RAN?*
2. *Como representar o posicionamento das funções virtuais de uma rede de rádio de acesso desagregada usando o menor número de recursos computacionais?*

1.3 Objetivos

O presente trabalho tem como objetivo geral o desenvolvimento da solução PlaceRAN, composto por duas etapas: (i) um orquestrador especializado para a implementação das funções de rádio virtualizadas para a solução de otimização e (ii) uma solução para a otimização de posicionamento da vNG-RAN. Ambos os objetivos gerais são orientados pelo contexto da evolução das redes de acesso na direção da vNG-RAN, da arquitetura de referência de virtualização NFV e das estratégias das iniciativas O-RAN. Sendo assim, com a finalidade de alcançar os resultados, os objetivos específicos são apresentados abaixo e direcionam a solução PlaceRAN de orquestração da otimização de posicionamento da vNG-RAN.

- Pesquisar o estado da arte no que tange à orquestração e a otimização do posicionamento da arquitetura vNG-RAN com foco na flexibilidade de desagregação e virtualização das funções de rádio;

Figura 1: A evolução da NG-RAN de baseada em *hardware* para baseada em *software* é exposta. A Figura 1(a) mostra a arquitetura NG-RAN dividida nos nós CU, DU e RU baseada em *hardware*. Enquanto que, a Figura 1(b) apresenta os nós da vNG-RAN e o posicionamento em uma infraestrutura computacional (desde a centralização na nuvem até a distribuição na borda) orquestrado por uma função centralizada.



Fonte: Elaborada pelo autor

- Desenvolver um modelo combinatório para as oportunidades de desagregação da vNG-RAN;
- Desenvolver uma solução de otimização de abordagem exata para o posicionamento da vNG-RAN de forma conceitual utilizando prova de conceito matemática;
- Desenvolver um orquestrador experimental para o posicionamento da vNG-RAN considerando o uso de funções de rádio virtualizadas emuladas e ciente da topologia *crosshaul*. O orquestrador também deve seguir as especificações da ETSI NFV e as iniciativas de código aberto para futuras integrações.
- Desenvolver um método de avaliação de desempenho para o PlaceRAN, tanto para

a formulação matemática quanto para o experimento com base em métricas, parâmetros e topologias reais;

- Avaliar os resultados obtidos e compará-los com as especificações almejadas das organizações padronizadoras;

Com base nos objetivos específicos, o modelo tem como foco o planejamento das redes vNG-RAN, para que, nele, possam ser superados os componentes que limitam o desempenho das rede móveis no que se refere às restrições impostas pelo ambiente que envolve a vNG-RAN. Como consequência, a solução PlaceRAN contribui para o desenvolvimento da vNG-RAN e para a transformação digital demandada pela sociedade.

1.4 Organização do Texto

O restante do trabalho é organizado em cinco capítulos. O capítulo 2 apresenta conceitos fundamentais da evolução e virtualização da arquitetura NG-RAN para a compreensão do restante do trabalho. Logo após, o capítulo 3 apresenta os trabalhos relacionados aos temas dessa pesquisa, que englobam perguntas associadas à questão de pesquisa. Esse capítulo objetiva apresentar o que já existe no estado da arte, bem como identificar onde existem lacunas e oportunidades da literatura. O capítulo 4 apresenta a solução PlaceRAN desenvolvida, bem como as decisões de projeto, que busca preencher as lacunas identificadas no capítulo anterior, além de atingir os objetivos do presente trabalho. A metodologia de avaliação da análise conceitual da formulação matemática e da emulação do protótipo experimental é apresentada no capítulo 5. Os resultados obtidos na simulação e emulação são apresentados e discutidos no capítulo 6. Por fim, o capítulo 7 apresenta a conclusão do trabalho, as contribuições da solução PlaceRAN e as possibilidades de trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

No presente capítulo, são caracterizados e discutidos os conceitos fundamentais desta dissertação. Esses conceitos têm como objetivo servir como base para o restante do trabalho. Sendo assim, o capítulo é dividido em duas seções que representam as duas principais transformações da vNG-RAN: a arquitetura NG-RAN e a vRAN. Dessa forma, inicialmente, na seção 2.1 são exploradas as concepções da evolução da arquitetura NG-RAN, enquanto, na Seção 2.2, os conceitos da vRAN são examinados com foco na transformação das funções de rádio.

2.1 Redes de Rádio de Acesso de Próxima Geração

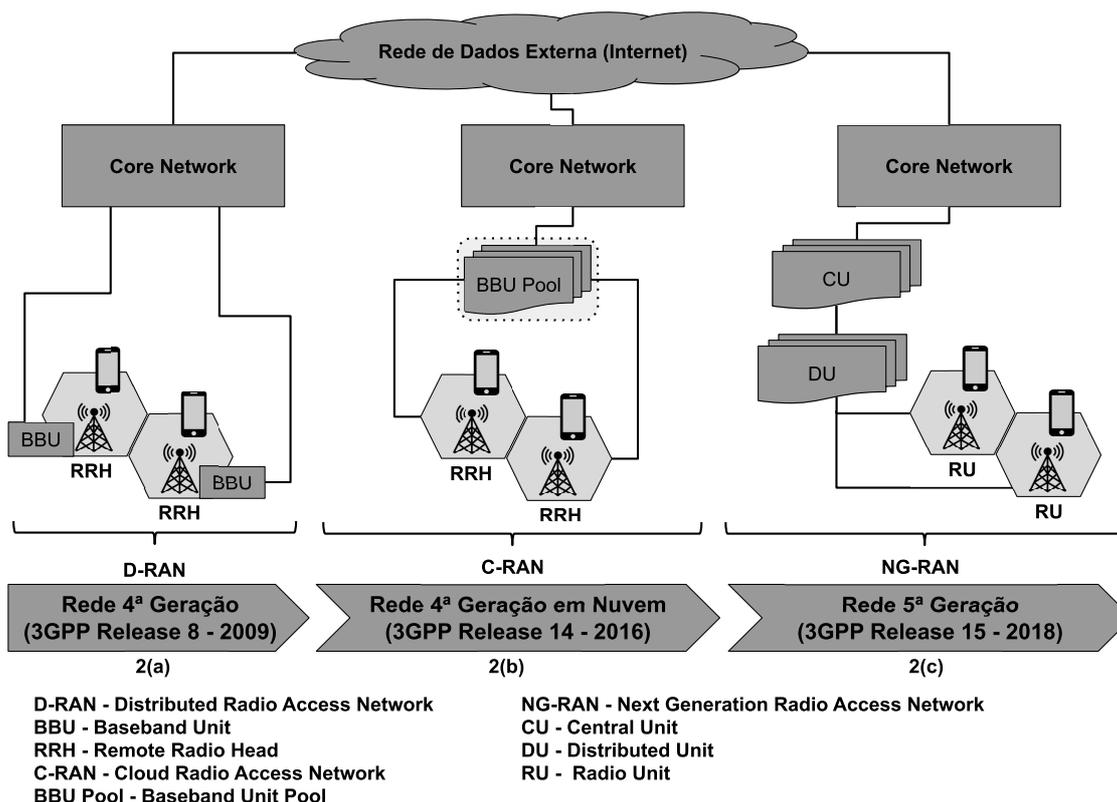
Para alcançar a especificação da arquitetura NG-RAN no *Release 15*, a organização 3GPP concentrou esforços na avaliação da tecnologia em funcionamento *Long Term Evolution* (LTE) e avaliou profundamente o impacto da proposta C-RAN. Nesse sentido, a presente seção explora a evolução da RAN desde a tecnologia LTE até a especificação atual, assim como traz também os conceitos de desagregação da pilha de protocolos, as restrições das redes de transporte *Crosshaul* e os requisitos para o desenvolvimento da arquitetura NG-RAN.

2.1.1 Evolução das Redes de Rádio de Acesso

A evolução da RAN, por meio das especificações padronizadoras do 3GPP para a rede LTE, estabeleceu a partir do *Release 8*, significantes aprimoramentos para a imersão em redes de pacotes estatísticos e a elevação das taxas de dados. A partir disso, denominou as BSs do LTE de *evolved NodeB* (eNB). Uma eNB é composta por dois elementos de rede, a Unidade de Banda Base (BBU - *Baseband Unit*) e a Cabeça de Rádio Remota (RRH - *Remote Radio Head*). BBU é responsável pelas funções de processamento de dados de banda base e normalmente é distribuído e co-localizado com o RRH, ao passo que RRH consiste em unidades de rádio e antenas (AGRAWAL et al., 2017). Contudo, as especificações do *Release 8* foram desenvolvidas com base na arquitetura D-RAN, em padrões de comunicação e capacidade de processamento particulares para os serviços propostos para rede LTE (banda larga e voz), dispondo de suporte para um número fixo de cargas de entrada (CHEN et al., 2015). Conseqüentemente, a recente demanda de novos requisitos exigiu reavaliação da arquitetura RAN, para que seja alcançado o suporte à baixa latência, às altas taxas de dados, à adaptabilidade em tráfego não uniforme, à escalabilidade, à cobertura ultra-densa e ao baixo consumo de energia (AGRAWAL et al.,

2017) A Figura 2 detalha a evolução arquitetônica da RAN, e, em específico, a Figura 2(a) apresenta o *Release 8*.

Figura 2: Representa a evolução da arquitetura RAN, iniciando pela D-RAN (tecnologia LTE) pela sub-figura 2(a), explicitando a evolução da arquitetura C-RAN (em ambientes de computação em nuvem) e a centralização da BBU pela sub-figura 2(b) e terminando no atual desenvolvimento, a arquitetura NG-RAN, que proporciona alta distribuição, flexibilidade e aperfeiçoamento das funções de rádio pela sub-figura 2(c).



Fonte: Elaborada pelo autor

Apesar de não especificado pelo 3GPP, o conceito da RAN centralizada em ambientes de computação em nuvem, a arquitetura C-RAN emergiu como candidata para suprir as deficiências apresentadas pela RAN, a partir do *Release 14* (3GPP, 2017a). Portanto, a arquitetura C-RAN, propõe modificação significativa nas BSs, dado que, as BBUs são desacopladas das RRHs e centralizadas em um ou mais conjuntos de BBUs (*BBU pool*) em uma infraestrutura de computação em nuvem, conforme expresso na Figura 2(b). Esse desacoplamento permite obter alta densidade de RRHs distribuídas geograficamente mais próximas dos UEs (HABIBI et al., 2019; PLIATSIOS et al., 2018). Além de oferecer maior capacidade ao sistema e menor consumo de energia, uma vez que a propagação do sinal de comunicação é concentrada próxima aos UEs (HOSSAIN et al., 2019). Outro ponto relevante da arquitetura C-RAN, refere-se à introdução da nova camada de trans-

porte *fronthaul* responsável pela comunicação entre as RRHs e as BBUs (AGRAWAL et al., 2017), sendo que este apresenta características semelhantes ao *backhaul* no que diz respeito às tecnologias aplicadas.

Embora a arquitetura C-RAN proponha melhorias expressivas no desempenho dos atributos do rádio (eficiência de processamento, mobilidade, interferência e outros), a divisão entre BBU e RRH necessita de altas taxas de dados da rede *fronthaul*. Dessa forma, limita uma alta centralização da BBU pool e conseqüentemente vantagens dessa arquitetura para as funções de rádio. Nesse sentido, o *Release 15* foi especificado para a NG-RAN, o qual adicionou, à evolução da C-RAN, principalmente, a flexibilidade de desagregação (até dois pontos distintos e diferentes grupos de protocolos) e a eficiência de processamento computacional. Ademais, oportuniza o agrupamento das unidades de rádio de menor ordem em um nó de maior ordem (um grupo de RUs em uma DU e um grupo de DUs em uma CU). Além disso, o *Release 15* especificou para a NG-RAN a interoperabilidade entre os sistemas de quarta e quinta geração, introduzindo quatro variações para as BSs independentes do CN em funcionamento. No caso do LTE a BS foi mantida como *Evolved NodeB* (eNB) quando conectada ao CN e *Next-Generation - Evolved NodeB* (ng-eNB) para conexão com o CN da quinta geração. Já para as redes móveis de quinta geração, as BSs foram denominadas de *Next-Generation NodeB* (gNB) para conexão com o CN desta mesma geração e en-gNB (sem acrônimo definido) para o CN (BERTENYI et al., 2018; MARSCH et al., 2018).

A principal transformação da arquitetura NG-RAN refere-se à desagregação e à quantidade de unidades em que a desagregação pode ser realizada em até dois pontos distintos e as funções de rádio são divididas em até três unidades distintas: CU, DU e RU (3GPP, 2017a; BERTENYI et al., 2018), conforme a Figura 2(c). A arquitetura particionada com três unidades divide conceitualmente o elemento do LTE BBU nos elementos gNB-CU e gNB-DU, enquanto mantém a RRH apenas com alteração na nomenclatura para RU. Tal transformação, tem como objetivo possibilitar ampla distribuição de nós RUs e DUs na rede para um eficiente manuseio dos recursos computacionais e, principalmente, para a garantia dos requisitos baseados nos serviços (como por exemplo, o requisito de latência) (AL-DULAIMI et al., 2018). Nas iniciativas O-RAN da indústria, a arquitetura com três unidades também é adotada, mas com a inserção da palavra *Open* resultando nos nós O-CU, O-DU e O-RU, tendo no entanto, os mesmo objetivos, apesar de as iniciativas O-RAN acrescentarem pesquisas relacionadas à inteligência de rede centralizada (ALLIANCE, 2020a). Por outro lado, como já ressaltado, a desagregação dos protocolos da RAN, que ocorre tanto pelas organizações padronizadoras quanto pela indústria, oportuniza a flexibilização da composição dos protocolos que integram as funções de rádio (MARSCH et al., 2018), conforme elucidado a seguir.

2.1.2 Desagregação das Redes de Rádio de Acesso de Próxima Geração

A flexibilização das funções de rádio na literatura é denominada de desagregação da RAN, a qual permite que as funções de rádio não sejam rígidas na composição de seus protocolos e recursos de *hardware*. Isso ocorre, desde que obedecem a composição e sequência da pilha de seis protocolos normativos (MARSCH et al., 2018). Diferentemente da RAN LTE, que apesar de também ser orientada à pilha de protocolos, foi desenvolvida em modelo baseado em *hardware* monolítico em que poucas interações entre nós lógicos são especificadas (resultando em uma arquitetura RAN simples e limitada) (BERTENYI et al., 2018). Sendo assim, o desenvolvimento da NG-RAN vislumbra benefícios adicionais aos existentes, principalmente nas funções de rádio que compõem as unidades CU e DU, conforme abaixo (BERTENYI et al., 2018):

- Implementação flexível e compartilhada de *hardware*, permitindo soluções escaláveis e com eficiência econômica.
- Uma arquitetura dividida propicia a coordenação de desempenho dos recursos, o gerenciamento de carga de trabalho e a otimização de desempenho em tempo real, além de permitir implementações baseadas em *software*.
- A desagregação das funções de rádio permite flexibilidade baseada nas aplicações de distintos casos de uso.

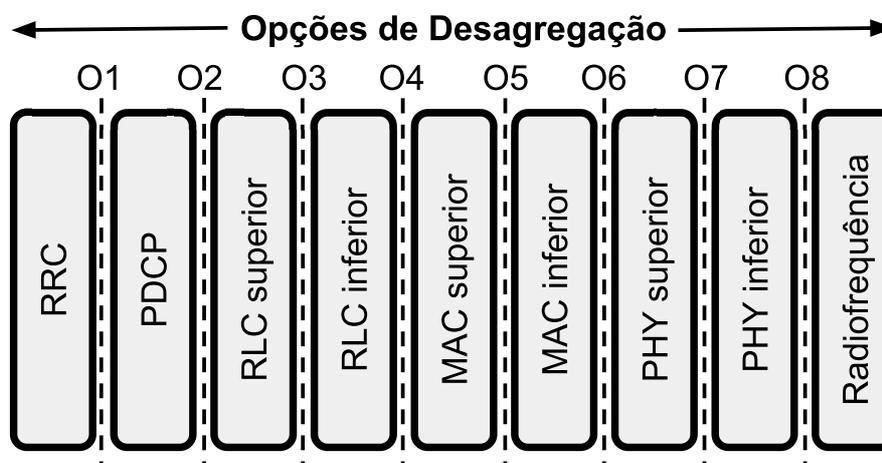
A orientação dos benefícios, viabiliza uma escolha de desagregação das funções de rádio conforme os cenários de implementação da RAN, das suas restrições e de seus serviços previstos, como por exemplo, a necessidade de especificidade de requisitos dos serviços, tais como: baixa latência, alta taxa de transferência, densidade específica do usuário e demanda de carga por determinada área geográfica (MARSCH et al., 2018). Além disso, apresenta a necessidade de interoperar com redes *crosshaul* heterogêneas e com diferentes níveis de desempenho (BERTENYI et al., 2018).

A desagregação das funções de rádio oportuniza orientação flexível, tanto pela demanda por requisitos, quanto por restrições da rede, para determinar a composição dos protocolos que constituem as funções de rádio. Na tecnologia LTE, a RAN tem a pilha de protocolos composta por cinco principais protocolos: camada física (PHY - *Physical Layer*), camada de enlace (MAC - *Media Access Control*), controle de enlace de rádio (RLC - *Radio Link Control*), protocolo de controle de pacotes de dados (PDCP - *Packet Data Control Protocol*) e controle de recursos de rádio (RRC - *Radio Resource Control*). Em relação à recomendação da 3GPP Release 15, a inclusão do protocolo de adaptação aos dados do serviço (SDAP - *Service Data Adaptation Protocol*) é prevista na pilha de protocolos, e são detalhados a seguir (3GPP, 2017a):

- O protocolo PHY é responsável pelo processamento dos sinais digitais e analógicos de comunicação entre UEs e ERBs, sendo baseado em técnicas de modulação e codificação adaptáveis.
- O protocolo MAC fornece controle de baixo nível da camada física, principalmente ao agendamento das transmissões de dados entre UE e ERB. Além disso, manuseia o mapeamento e a seleção dos canais lógicos e dos canais de transporte, a medição baseada em relatórios de tráfego; e a correção de erros através do protocolo de solicitação de repetição automática híbrida (HARQ - *Hybrid Automatic Repeat Request*). A camada MAC, ainda oferece serviços ao RLC por meio de canais lógicos.
- O protocolo RLC garante a entrega confiável de fluxos de dados que precisam chegar intactos à camada MAC. É responsável pela segmentação e manipulação de retransmissão utilizando o protocolo de solicitação de repetição automática (ARQ - *Automatic Repeat Request*). Além disso, oferece serviços ao PDCP através de *bearers* de rádio.
- O protocolo PDCP executa funções de transporte de nível superior relacionadas à compressão de datagramas IP, segurança (criptografia e segurança da informação) e túneis virtuais dos rádio com os UEs.
- O protocolo RRC é incumbido de toda a sinalização e controle do UE com a BS. Esta incumbência inclui estabelecimento, reconfiguração e atualização de *bearers* de rádio, processos de conexão de mobilidade, controle de admissão, monitoramento de localização e controle de energia.
- O protocolo SDAP mapeia a interação entre o pacote de um fluxo com qualidade de serviço associado e o *bearer* de radio do plano de dados (específico para a carga útil do UE), marcando os pacotes de dados do usuário de forma eficiente.

Com base na pilha de protocolos da RAN, oito opções de desagregação são admitidas, de opção 1 (O1) até opção 8 (O8), conforme a Figura 3. A convenção das opções se dá pela fronteira entre os distintos protocolos e pela intra divisão dos protocolos RLC, MAC e PHY, estabelecida em inferior e superior (*low e high*) (3GPP, 2017a). Todas as fronteiras entre os protocolos são passíveis de serem tratadas como uma opção, exceto o SDAP. Os protocolos RRC e PDCP não possuem intra divisões (ITU-T, 2018). De forma pontual, a divisão do protocolo PHY, opção 7, ainda pode ser subdividida nas opções 7-1, 7-2, 7-3 e 7-4. Por fim, a opção 8 é a desagregação admitida com os recursos físicos, antenas e conversores eletro-ópticos, denominados na pilha de protocolos de Rádio Frequência (RF) (LARSEN et al., 2018).

Figura 3: Oito opções de desagregação da pilha de protocolos proposta pela arquitetura NG-RAN para prover flexibilidade na composição dos nós de rede.



Fonte: Adaptado de (3GPP, 2017a).

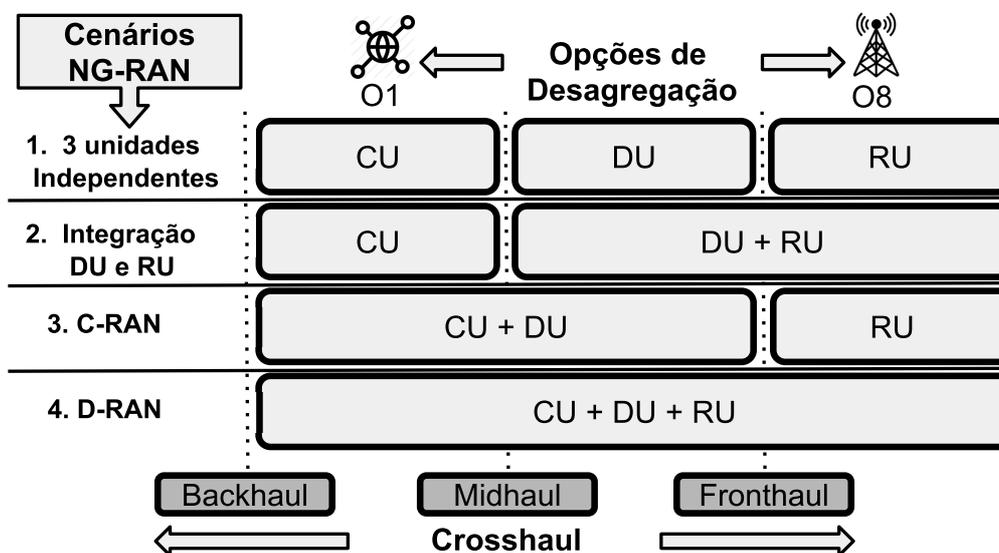
Embora a NG-RAN possua três unidades de rádio (CU, DU e RU), no máximo duas desagregações são admitidas, combinadas às oito opções de desagregação. Na literatura, o mesmo é denominado de três unidades independentes (ITU-T, 2018; MARSCH et al., 2018). Entretanto, também são admitidos cenários de desagregações em um único ponto ou sem desagregação. Para a desagregação em um ponto, dois cenários são admitidos. O primeiro é a integração da RU e CU e o segundo é o C-RAN (ITU-T, 2018). Por fim, a D-RAN opera com a totalidade das unidades em um único dispositivo e em um mesmo ambiente físico sem desagregação. As arquiteturas C-RAN e D-RAN são exploradas na subseção 2.1. A Figura 4 expõe os cenários detalhados, combinando as opções de desagregação e os pontos de atuação da rede *crosshaul*.

2.1.3 Redes de Transporte *Crosshaul*

A desagregação da RAN passa por forte sinergia com a rede *crosshaul* e as redes que a compõe (*backhaul*, *midhaul* e *fronthaul*), sendo esta diretamente associada às redes de quinta geração (ITU-T, 2018; LARSEN et al., 2018). A Figura 5 apresenta a sinergia entre a desagregação da RAN e a rede *crosshaul*, detalhando o posicionamento conceitual. Apesar do modelo conceitual, a alta flexibilidade da arquitetura NG-RAN demanda, do *crosshaul*, possibilidades de diferentes suportes à desagregação, desde que sejam atendidos os requisitos exigidos.

A rede *backhaul* é a precursora das redes de transporte para atendimento às redes móveis e foi introduzida com a tecnologia LTE para prover comunicação entre BS e CN

Figura 4: Cenários de arquiteturas suportadas pela NG-RAN, combinando as opções de desagregação da RAN e o posicionamento da rede *crosshaul* (ITU-T, 2018).



Fonte: Adaptado de (ITU-T, 2018).

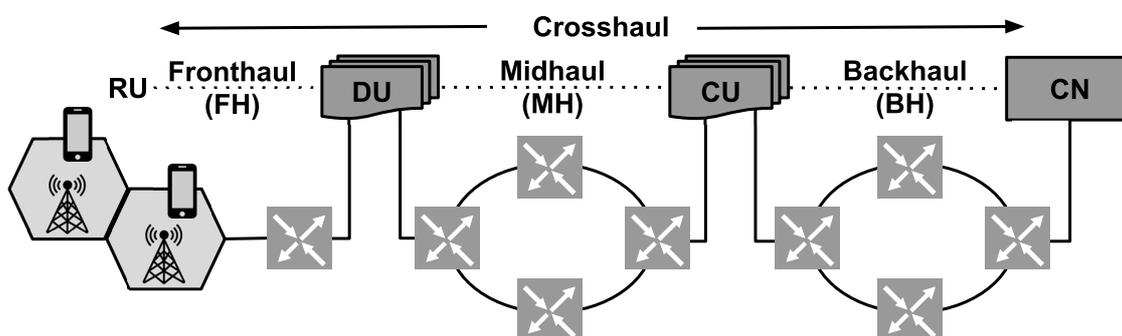
(FIORANI et al., 2015), sendo posicionado entre CU e CN na arquitetura NG-RAN, conforme a Figura 5. No que se refere à desagregação proposta na NG-RAN, o posicionamento da rede é mantido sem alterações. Entretanto, o posicionamento flexível da CU pode acarretar em um tráfego, com características da rede *backhaul*, mais próximos da borda ou do núcleo da rede (ITU-T, 2019).

Em conformidade com a cronologia, a rede *fronthaul* foi introduzida com a arquitetura C-RAN para o desacoplamento dos nós RRH e BBU. Logo, a rede *fronthaul* conecta as RRHs a um *pool* centralizado de BBUs (FIORANI et al., 2015). Entretanto, para a arquitetura NG-RAN, a rede *fronthaul* é posicionada majoritariamente entre os protocolos mais baixos da pilha de protocolo, entre as opções de desagregação que compõem a comunicação entre a RU e a DU (protocolos PHY e MAC). Além disso, a rede *midhaul* foi desenvolvida para ser aplicada na nova arquitetura NG-RAN, estando posicionada entre os nós CU e DU (ITU-T, 2018).

2.1.4 Requisitos de Desagregação da das Redes de Rádio de Acesso de Próxima Geração

Em se tratando de requisitos de rede, independentemente da rede *crosshaul*, os dois principais dizem respeito às métricas de taxa de dados (dada em bits por segundos) e à latência. A Tabela 1 apresenta os valores de referência (valores máximos suportados pela RAN e dependentes da demanda da carga da rede) para taxa de dados e latência requisita-

Figura 5: Arquitetura conceito das redes *crosshaul* (neste exemplo, formato anel, porém não há restrições para outras topologias) para o desenvolvimento da arquitetura NG-RAN.



Fonte: Adaptado de (ITU-T, 2018).

dos para a interação entre cada opção de divisão dos protocolos. Os valores apresentados na Tabela 1 são de referência do 3GPP (3GPP, 2017a) para os canais de rádio de 100MHz largura de banda, 32 antenas, 8 camadas de entrada múltipla e saída múltipla (MIMO - *Multiple Input, Multiple Output*) e modulação de amplitude em quadratura (QAM - *Quadrature Amplitude Modulation*) de 256 símbolos.

Tabela 1: Requisitos de referência da rede *crosshaul* (valores máximos suportados pela RAN e dependentes da demanda da carga da rede) com base na desagregação da arquitetura NG-RAN (3GPP, 2017a)

Opções de Desagregação	Divisão Funcional	Taxa de bits Aproximada	Latência Máxima Sentido único (one way)
Opção 1	RRC-PDCP	DL: 4Gb/s - UL: 3Gb/s	10ms
Opção 2	PDCP-High RLC	DL: 4Gb/s - UL: 3Gb/s	1,5~10ms
Opção 3	Intra RLC	DL: 4Gb/s - UL: 3Gb/s	1,5ms~10ms
Opção 4	Low RLC-High MAC	DL: 4Gb/s - UL: 3Gb/s	~0,1ms
Opção 5	Intra MAC	DL: 4Gb/s - UL: 3Gb/s	< 1ms
Opção 6	Low MAC-High PHY	DL: 4Gb/s - UL: 5Gb/s	0,250ms
Opção 7	Intra PHY	DL:10,1~86,1Gb/s - UL:16,6~86,1Gb/s	0,250ms
Opção 8	Low PHY-RF	DL:157,3Gb/s - UL:157,3Gb/s	0,250ms

Para a taxa de dados, a Tabela 1 é dividida em *downlink* (DL) e *uplink* (UL), apresentando assimetria entre os mesmos, além de exigir do *crosshaul* altíssimas taxas nas opções O7 e O8. No que se refere à latência, a métrica exige da rede *crosshaul* medidas extremamente baixas para atendimento a desagregação, sendo essa, entre as opções O4 e O8, medidas inferiores de 1ms. Tais requisições para latência ocorrem tanto em virtude das características das opções de desagregação, quanto dos novos serviços das redes de quinta geração, uma vez que o suporte para os serviços de altas taxas de dados e os ser-

viços de atraso restrito não devem ser superiores a 4ms e 0,5ms, respectivamente (ITU-T, 2018). Dentro das redes móveis, as recomendações do 3GPP para a latência fim-a-fim (UE até o CN) especificam uma latência de 20ms para sinalização (CP - *Control Plane*) e uma variação entre 1~4ms para a carga útil do UE (UP - *User Plane*) (3GPP, 2017b).

Embora sejam especificadas as taxa de dados para cada uma das opções de desagregação, essencialmente as redes *backhaul* e *midhaul* são responsáveis por agregação de tráfego devido a agrupamento de distintos nós da RAN. Sendo assim, tanto o *backhaul*, quanto o *midhaul* são especificados para suportar centenas de Gbit/s, com a diferença de que o *backhaul* agrega os nós CUs (último nós antes do CN) e o *midhaul* agrega os nós DU (que são agregados pelos nós CUs) (AL-DULAIMI et al., 2018; ITU-T, 2018). Quanto à topologia de rede, o ITU-T prevê uma topologia ponto-a-ponto como base para o *fronthaul* e, por isso, não é prevista agregação de tráfego como nas camadas superiores (ITU-T, 2018). Por fim, para ambas as camadas *midhaul* e *backhaul*, o ITU-T prevê uma topologia em anel ou árvore (ITU-T, 2019).

2.2 Virtualização das Redes Móveis

Essa seção tem como objetivo elucidar os direcionamentos da vRAN para a construção da arquitetura vNG-RAN. Dessa forma, a Subseção 2.2.1 expõe os motivadores e os componentes da vRAN. Na subseção 2.2.2, a principal arquitetura de virtualização o NFV, é explorada, assim como na subseção 2.2.3, as métricas da arquitetura NFV são exibidas. Por fim, a subseção 2.2.4 detalha os conceitos do posicionamento e encadeamento de funções de virtualização, bem como conecta estes conceitos à arquitetura vNG-RAN.

2.2.1 Virtualização das Redes de Rádio de Acesso

Apesar de todos os avanços da arquitetura NG-RAN para desagregação e distribuição das unidades de rádio, decompor as funções de rádio em componentes não monolíticos e agregar as funções para melhorar o desempenho de rádio e processamento são vitais para alavancar as redes móveis de quinta geração. Nesse sentido, a imersão dos conceitos baseados em *software* é proposta de forma nativa para a evolução, principalmente a virtualização. Em imersão a esse conceito, cada unidade da NG-RAN (CU, DU e RU) é considerada uma função de rede virtualizada (vCU, vDU e vRU) composta por funções de rádio desagregada. Sendo assim, estes podem ser executados em até três recursos computacionais distintos (BERTENYI et al., 2018). Cada dispositivo de computação possui um conjunto de protocolos da RAN neles executados (sejam estes simétricos ou assimétricos). Embora RU possa ser virtualizada (vRU) e suportar o compartilhamento de recursos computacionais, o protocolo RF da pilha de protocolos da RAN não possui suporte para

virtualização. Entretanto, para atingir o avanço almejado, é necessário que a vRAN seja realizada de forma a satisfazer as condições impostas pela desagregação da arquitetura NG-RAN (MARSCH et al., 2018).

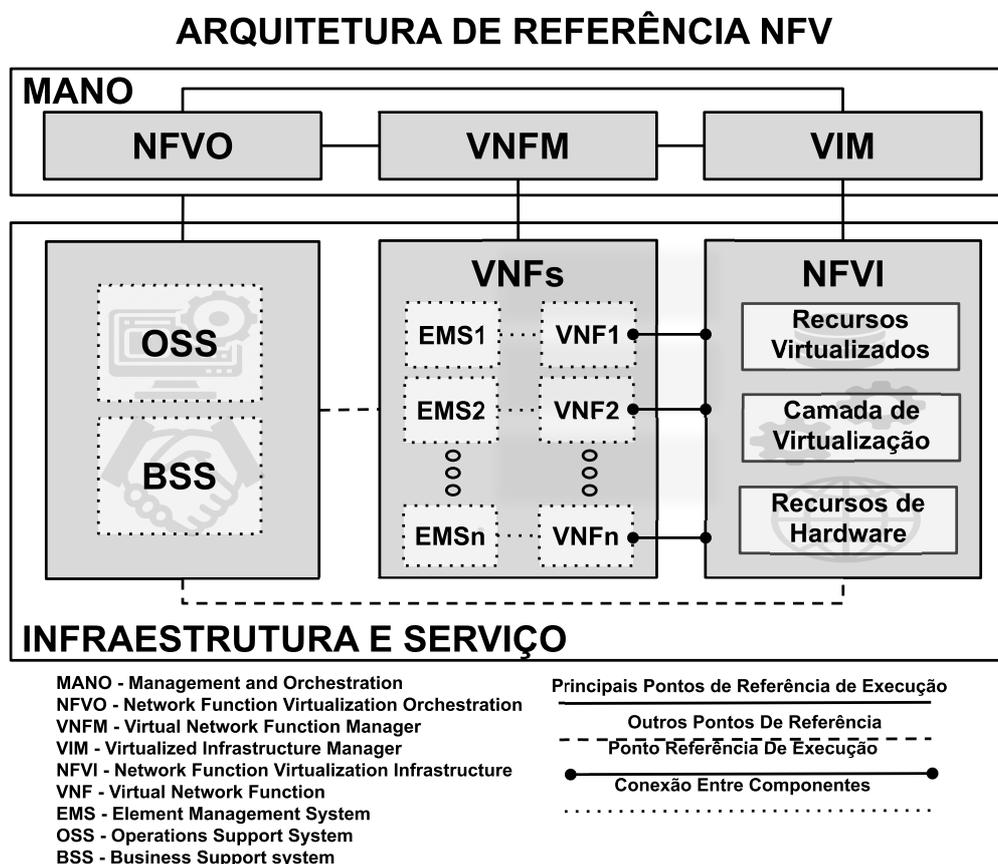
A organização padronizadora ETSI atualmente é a principal referência no processo de virtualização das redes móveis, incluindo a RAN. A mesma define que virtualização é "a eliminação da dependência entre as funções de rede e o *hardware*" (desacoplamento do *hardware* e do *software*) (ETSI, 2013) e, com isso, trabalha de forma contínua no desenvolvimento do conceito e da arquitetura NFV, com o objetivo de oferecer a oportunidade de implementação das funções de rede em *software* que podem ser instanciadas em diferentes locais e *hardware* de rede (GUERZONI et al., 2012). Dessa maneira, propicia a criação de VNFs (no caso da RAN, as unidades vCU, vDU e vRU) separados do *hardware* proprietário subjacente e transfere-as para uma infraestrutura de recursos computacionais compartilhada baseada em *software* (BERNARDOS et al., 2019; ETSI, 2013). É importante ressaltar que as iniciativas O-RAN da indústria estão alinhadas com o ETSI (BONATI et al., 2020).

2.2.2 Arquitetura NFV

Inserido na organização ETSI, o grupo de trabalho *Industry Specification Groups* (ISG) NFV colabora para o desenvolvimento da estrutura e da arquitetura da tecnologia de virtualização para consolidar os diferentes equipamentos de rede que compõem as redes móveis. O resultado da padronização do ISG NFV é uma arquitetura de referência composta por quatro distintos macro blocos: (i) infraestrutura NFVI, que inclui recursos físicos de computação, rede e armazenamento, e componentes de *software* para suportar a execução de VNFs; (ii) VNFs que são instanciadas e gerenciadas de formas individuais no (EMS - *Element Management System*) na NFVI; (iii) o MANO das VNFs e da NFVI (BERNARDOS et al., 2019); e (iv) os sistemas de suporte ao serviço e ao negócio (OS-S/BSS - *Operations Support System and Business Support System*). A Figura 6 detalha a arquitetura proposta pelo grupo ISG NFV da ETSI.

No que se refere ao bloco o MANO, este é subdividido nos três componentes distintos: o gerenciador das VNFs (VNFM - *VNFM Manager*), a Gerência de virtualização de infraestrutura (VIM - *Virtualized Infrastructure Manager*) e o orquestrador NFV (NFVO - *NFV orchestrator*). Nesse sentido, o componente VNFM é responsável pelo gerenciamento do ciclo de vida de uma única ou de várias instâncias de VNFs do mesmo ou de diferentes tipos. Esse processo inclui a configuração e instanciação das VNFs, considerando coordenação de elasticidade dos recursos e coletando informações de desempenho do NFVI, assim como o VIM é o componente que controla e gerencia os recursos associados ao NFVI. Dependendo de sua configuração, um VIM pode ser dedicado ao controle

Figura 6: Arquitetura referência NFV ETSI para virtualização de redes para aplicações genéricas. A arquitetura dividida nos blocos funcionais de gerenciamento e orquestração (MANO), sistema de suporte operacional e de negócios (OSS/BSS), funções de rede (VNFs) e infraestrutura computacional (NFVI) virtualizadas.



Fonte: Adaptado de (ETSI, 2014a).

de um tipo específico de recurso NFVI (computação, rede ou armazenamento) ou pode controlar múltiplos NFVIs.

No que se refere ao bloco NFVO, o mesmo é responsável pela orquestração da arquitetura. Primeiramente, a definição de orquestração é definida pela ITU-T, de forma geral, como conceito de gerenciamento encarregado pelo controle, configuração, geração de relatórios e de eventos entre a infraestrutura das funções de rede e os sistemas de gerenciamento de rede (ITU-T, 2017). De outro modo, define-se a orquestração como o responsável pela coordenação e automação das redes de alta complexidade, sejam estas físicas ou virtuais. Ademais, pode atribuir inteligência inerente ou apenas controle autônomo (ITU-T, 2017). Na mesma direção, o ETSI conceitua que a orquestração é definida como a automação, o gerenciamento e a operação da infraestrutura, sendo esta composta

por elementos e recursos baseados em *hardware* e *software*. Por fim, o ETSI ressalta que o conceito de orquestração habilita a infraestrutura para atuar de forma distribuída. (AFOLABI et al., 2018; ETSI, 2014a).

Com base na contextualização de orquestração pelas organizações padronizadoras, o bloco NFVO é o componente da arquitetura NFV responsável pela orquestração dos recursos NFVI em vários VIMs e pelo gerenciamento do ciclo de vida dos serviços de rede. Para este fim, o bloco NFVO é dividido em duas categorias: (i) orquestração de recursos (RO - *Resources Orchestration*) e (ii) orquestração de serviços de rede (NSO - *Network Service Orchestration*). O primeiro é usado para fornecer o acesso a recursos NFVI de forma abstraída, independentemente de quaisquer VIMs, bem como governança de instâncias VNF compartilhando recursos da infraestrutura NFVI, tais como computação, armazenamento e rede. Já o segundo lida com a criação de serviços fim-a-fim, compondo diferentes VNFs, e com o gerenciamento da topologia das instâncias de serviços de rede. A atuação é satisfeita por meio do fornecimento de funções para lidar com a integração, a instanciação, o escalonamento, a atualização e o encerramento de serviços de rede e qualquer operação em seus SFCs entre VNFs associadas. O NFVO tem acesso aos Catálogos dos Serviços de Rede e VNF, que contêm informações sobre os serviços e funções dos serviços disponíveis.

Atualmente, para a implementação das VNFs nas NFVIs, as duas principais técnicas de instanciação de máquinas virtuais são baseadas em *hypervisor* ou em contêineres (REHMAN et al., 2019). Ambas propõem-se a particionar dos recursos físicos e do sistema operacional (OS - *Operational System*) e, assim, alocar VNFs individualizadas. As principais diferenças entre as duas soluções de máquinas virtuais são que o *hypervisor*, também chamado de monitor da máquina virtual (VMM - *Virtual Machine Monitor*), é um *software* que isola o OS e os recursos das VNFs. Dessa maneira, permite a criação e o gerenciamento das VNFs em função de um OS individualizado sob o *hardware* hospedeiro e com execução como convidado (YI et al., 2018), enquanto que a tecnologia baseada em contêineres não precisa de OS separado para hospedar VNFs. Como diferenciais, o contêiner reduz a sobrecarga em comparação com o *hypervisor*, uma vez que é executado diretamente no OS do hospedeiro e sob um mecanismo embarcado de gerenciamento de contêiner acoplado ao OS (YI et al., 2018).

2.2.3 Métricas da Arquitetura NFV

Garantia de qualidade e experiência são requisitos pertinentes para serviços de telecomunicações, os quais são prestados em contexto de redes móveis (ETSI, 2018). Sendo assim, para o desenvolvimento da arquitetura NFV, tais requisitos emergem de forma relevante para a oferta de melhores serviços. Nesse sentido, a padronização de métricas

produz uma validação de desempenho e confiabilidade da rede. As métricas devem ser especificadas conforme utilidade pretendida para transmitir o significado exato de um valor medido (ETSI, 2014b). Além disso, medição de uma métrica é definida como um conjunto de operações com o objetivo de determinar um valor medido ou resultado de uma medição (ETSI, 2014a). Uma vez que a arquitetura NFV é projetada para abranger tecnologias distintas, diferentes tipos de métricas precisam ser definidos e coletados (ETSI, 2014b).

ETSI NFV define grupos de métricas de qualidade para orquestração (por exemplo, alocação de recursos, configuração e estrutura) e os conecta com as redes *crosshaul* (ou redes transporte baseada em pacotes) (ETSI, 2014a). Em relação à velocidade operacional, a métrica de latência é destacada, tanto para o provisionamento, quanto para as características de rede, tais como: o atraso, a variação do atraso e a taxa de dados. Métricas de acurácia também são exploradas nas definições de orquestração, tendo destaque nas políticas de posicionamento, na diversidade da rede *crosshaul* e na taxa de dados perdidos. Por fim, o grupo de disponibilidade é recomendado com as métricas de interrupção de conectividade, provisionamento para reabilitação e falhas (ETSI, 2014a).

A NFVI também possui recomendação específica para os nós e seus recursos computacionais (armazenamento, computação e comunicação de rede), sendo estes baseados nas métricas de processamento, de rede e de memória (ETSI, 2018). A métrica de processamento tem foco na unidade central de processamento (CPU - *Central Processing Unit*) e sua utilização expressa-se em porcentagem. As métricas de rede são expressadas pelos parâmetros de contadores, pelo tempo de medida dos contadores, pela velocidade e pelo estado da interface. Além disso, a métrica de memória tem como foco sua medida de ocupação medida pela unidade Bytes (ETSI, 2018).

2.2.4 Posicionamento das Funções de Rede Virtualizadas

Apesar da evolução vislumbrada pela virtualização da NG-RAN e sua distribuição, o melhor posicionamento das VNFs ou VNF-P da NG-RAN é considerado um dos principais problemas e desafios (ZHANG, 2018). O problema de posicionamento dos VNFs implica em escolher qual o melhor local para implantar as instâncias VNFs nos recursos de uma rede física (ZHANG, 2018). O posicionamento das VNFs é um problema combinatório desafiador, pois envolve um grande número de decisões de posicionamento discretas. Esse tipo de problema de decisão é conhecido como NP *Hard*, devido à dificuldade de encontrar um cenário ideal ou próximo ao ideal (ZHANG, 2018). Por exemplo, se " n " for o número total de VNFs e " m " for o número total de recursos computacionais, o número de mapeamento possível poderá ser m^n (MASDARI et al., 2016).

O tópico de posicionamento é estudado amplamente em ambientes de computação em

nuvem, uma vez que é considerada uma operação crítica, devido à tomada de decisão para determinar o melhor recurso físico apropriado para hospedar as máquinas virtuais (incluindo VNFs) (LAGHRISSEI; TALEB, 2018; YI et al., 2018). Para a computação em nuvem, os requisitos considerados mais relevantes são a eficiência energética, os recursos computacionais (rede, armazenamento e computação) e o suporte à qualidade de serviço. O problema do posicionamento pode ser dividido em duas partes: a primeira parte é a admissão de novas solicitações para o provisionamento e a alocação nos recursos computacionais, e a segunda parte inclui a otimização (por exemplo, monitoração de desempenho e processos de migração). A complexidade do posicionamento tem relação direta com os requisitos de entrada de uma instanciação, sendo estes, muitas vezes, imprevisíveis (MASDARI et al., 2016).

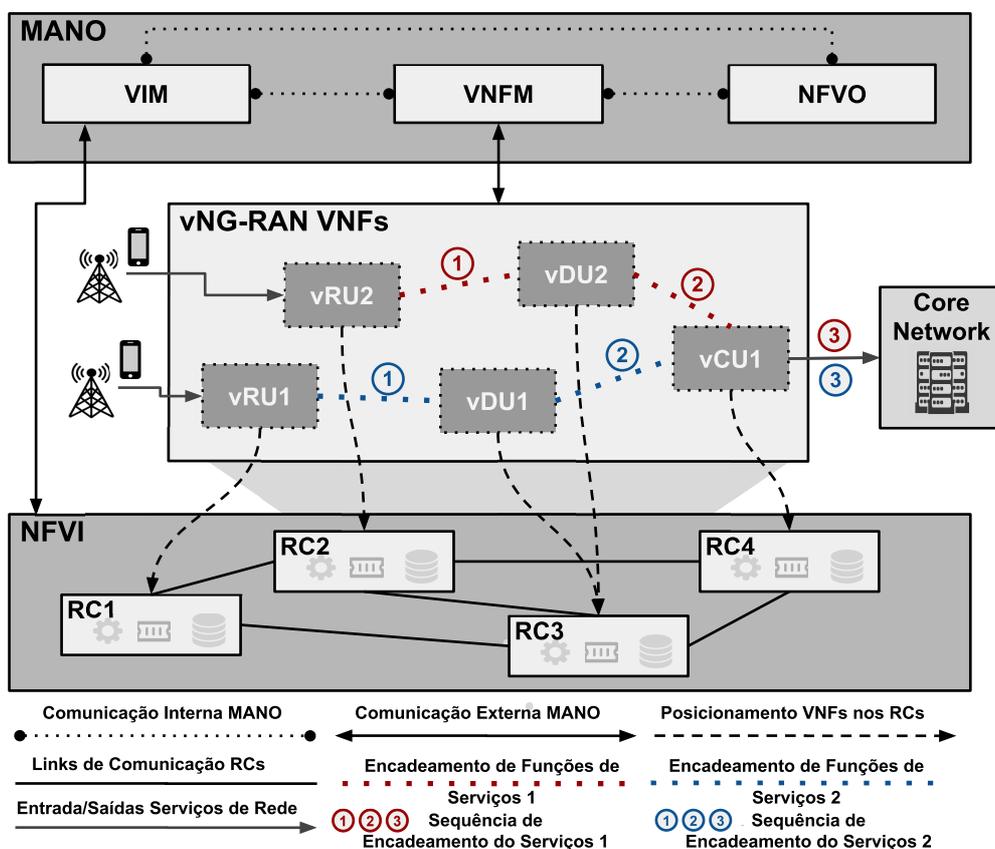
Concentrando-se no problema VNF-P, a entrada consiste em uma cadeia de funções de serviços (SFC - *Service Function Chain*), que é composta por um conjunto de VNFs, enquanto a rede de substrato (recursos de rede e computacionais) fornece as restrições físicas em termos de taxa de dados e capacidade. Nesse contexto, o termo capacidade não está relacionado apenas a recursos computacionais, como número de núcleos de CPU e memória. Pelo contrário, refere-se aos recursos de encaminhamento e processamento de pacotes. Cada cadeia de serviço é um fluxo de pacotes de rede que flui através de uma sequência de VNFs a uma determinada taxa. O tráfego de rede para uma determinada cadeia de serviços deve ser orientado a uma pré-definida sequência de VNFs (LAGHRISSEI; TALEB, 2018).

Cada cadeia de serviço em um problema de posicionamento de VNFs pode ser vista como uma rede virtual a ser mapeada em uma rede física. No entanto, o posicionamento dos VNFs difere ao passo que cada nó (VNF) em uma rede virtual pode ser mapeado para várias instâncias que são colocadas em nós diferentes na rede física. O posicionamento do VNF combina os problemas de mapeamento de cadeias de serviço em redes virtuais e mapeia as redes virtuais resultantes na rede física. Do ponto de vista arquitetural do NFV MANO, a solução do problema de posicionamento dos VNFs deve ser desenvolvida no NFVO, sendo este, o componente responsável por manusear a alocação de recursos na NFVI (ZHANG, 2018).

Com o intuito de conectar os conceitos apresentados na seção 2.2, a Figura 7 apresenta um exemplo do VNF-P e do SFC posicionados na arquitetura NFV representando a alocação das VNFs de rádio (vCU, vDU e vRU) nos recursos computacionais, o encadeamento das VNFs de rádio, a NFVI e o contexto fim-a-fim das redes móveis (desde CN até o UE). Embora não representada na figura, a rede *crosshaul* é responsável pelo encaminhamento do tráfego e pelo consequente encadeamento das VNFs.

De forma a concluir o capítulo 2, a seção 2.2.4 fundamenta o conceito de posicionamento das funções virtualizadas em uma arquitetura de referência NFV. A arquitetura é

Figura 7: Posicionamento e encadeamento das VNFs da arquitetura vNG-RAN, assim como a conexão com a arquitetura NFV e os blocos MANO, NFVI e VNF.



Fonte: Elaborada pelo autor

explorada na seção 2.2. Assim como conectam os conceitos de virtualização apresentados no capítulo com os elementos base da nova arquitetura NG-RAN da Seção 2.1, tais como desagregação da pilha de protocolos, restrições das redes *crosshaul* e virtualização das funções de rádio, estes conceitos são pertinentes para um melhor entendimento do contexto que envolve a evolução das redes móveis de quinta geração, com foco nas redes de rádio de acesso e a solução PlaceRAN. O entendimento também é válido para o seguinte capítulo 3, pois nele são apresentados os trabalhos já desenvolvidos na literatura os quais corroboram com o mesmo contexto da solução PlaceRAN.

3 TRABALHOS RELACIONADOS

Este capítulo está organizado de modo a apresentar a perspectiva de orquestração da arquitetura vNG-RAN, e avaliar as lacunas ainda existentes a partir da determinação do estado da arte em relação ao problema de VNF-P, descrevendo o desenvolvimento das relevantes soluções de orquestração de posicionamento. A pesquisa realizada foi dividida em quatro seções. Na seção 3.1, o método de seleção dos artigos é elucidado, dando base para o desenvolvimento das seções 3.2 e 3.3. A seção 3.2 é diretamente relacionada à primeira questão específica desta pesquisa e examina pesquisas relacionadas a arquiteturas de orquestração da vNG-RAN. A segunda seção 3.3 analisa, na literatura, os modelos de posicionamento para as funções de rádio virtualizadas, sendo fundamentada na segunda questão específica desta pesquisa. Finalmente, a seção 3.4, sintetiza as lacunas e as oportunidades, bem como ressalta as principais ambições do trabalho em desenvolvimento.

3.1 Metodologia para Seleção de Artigos

Para a seleção dos artigos foram realizadas pesquisas nas bases Scopus e *Google Acadêmico*, fundamentadas em tópicos e palavras chaves pré-determinadas para as questões de pesquisa específicas nas quais têm base as seções 3.2 e 3.3. O processo de filtragem dos artigos adotou os seguintes critérios: (i) ano de publicação, com intervalo de 4 anos de 2017 a 2020; (ii) tipo de publicação, *i.e.*, anais científicos ou de conferências de periódicos; (iii) contexto de publicação, permitindo a inclusão de itens no estudo como, por exemplo, pesquisas acadêmicas e ensaios da indústria; e (iv) idioma, com apenas artigos na língua inglesa. A decisão por artigos a partir do ano de 2017 deve-se ao fato de a arquitetura NG-RAN ter sido especificada pelo 3GPP na *Release 15* no ano de 2018.

No método de seleção de artigos, aplicou-se os critérios de exclusão e o processo de filtragem dos artigos relevantes e representativos da pesquisa. Os critérios de exclusão foram aplicados a manuscritos que não abordam os temas específicos de cada seção do capítulo. As etapas do processo de filtragem foram fundamentadas na eliminação de impurezas (artigos fora do contexto pesquisado), descarte de artigos duplicados, análise de título e resumo, e, finalmente, a análise do trabalho por completo. Por último, orientadas pela seleção dos artigos da primeira questão específica de pesquisa, as iniciativas de código aberto (*open source*) foram avaliadas diretamente em seus respectivos repositórios.

3.2 Orquestração das funções de rádio virtualizadas

Esta seção é dividida em duas subseções com o intuito de apresentar as arquiteturas de orquestração da vNG-RAN. Primeiramente, são exploradas as iniciativas *open source* na subseção 3.2.1, as quais são, por sua vez, amplamente ligadas a grandes projetos de pesquisa e a indústria. Por outro lado, a subseção 3.2.2 apresenta os trabalhos acadêmicos pertinentes e atuais. Ambas as subseções estão alinhadas com a primeira questão específica de pesquisa.

3.2.1 Iniciativas *Open Source*

A virtualização atua como um facilitador para a criação de instâncias logicamente isoladas sobre o *hardware* físico, que pode ser compartilhado de maneira dinâmica, eficiente e flexível (MAMUSHIANE et al., 2019), o qual é aliado à desagregação da NG-RAN para habilitar as funções de rádio da vNG-RAN. Atualmente, as duas tecnologias fundamentais de virtualização, *hypervisor* e contêiner, são as propulsoras da convergência entre virtualização e desagregação. Ambas as tecnologias incorporam diferentes meios de virtualização, orquestração e dimensionamento de recursos. Além disso, as duas tecnologias são desenvolvidas em diferentes estruturas de virtualização, tais como: (i) *VMware*, *Linux Kernel-based Virtual Machine* (KVM), *Oracle VirtualBox*, *Hyper-V* para *hypervisor* e (ii) *Linux Containers* (LXC) e *Docker* para contêineres (BONATI et al., 2020; MAMUSHIANE et al., 2019).

Resultados recentes mostram que a solução em contêiner tende a ser a mais apropriada para virtualização de RAN, devido aos requisitos estritos de atraso e escalabilidade de recursos (BONATI et al., 2020; GAVRILOVSKA et al., 2020). No que se refere a orquestração de contêineres em aplicações genéricas, quatro ferramentas lideram as implementações: (i) *Kubernetes* (K8S), (ii) *Docker Swarn*, (iii) *Mesosphere Marathon* e (iv) *Cloudify* (CASALICCHIO, 2019). Apesar das quatro opções, K8S desponta no cenário de virtualização de telecomunicações como a principal ferramenta, devido ao amplo desenvolvimento em TI, às opções de configurações flexíveis (*e.g.*, *threshold* customizados) e ao suporte a contêineres baseados em *Docker* (amplamente difundido) (BONATI et al., 2020; CASALICCHIO, 2019). A ferramenta K8S possui foco em gerenciamento de infraestruturas virtualizadas e orquestração no formato de *clusters* com gestão sob um *pool* de recursos computacionais, sejam estes físicos ou virtuais (KUBERNETES.IO, 2020).

Independentemente da tecnologia de virtualização, a arquitetura NFV com base na orquestração e gerenciamento ETSI MANO, é o modelo orientativo para o desenvolvimento da virtualização da vNG-RAN (ALLIANCE, 2020a; MAMUSHIANE et al.,

2019). Entre os principais projetos de código aberto para o ETSI MANO, destacam-se as iniciativas CORD/XOS (*Central Office Re-architected as a Datacenter/Everything-as-a-Services*), ONAP (*Open Network Automation Platform*), OSM (*Open Source MANO*), Cloudify, Open Baton, Gohan e Tacker, lideradas pelas respectivas organizações ON.lab, Linux Foundation, ETSI, Giga Space, Fraunhofer, NTT Data e OpenStack Foundation (MAMUSHIANE et al., 2019). A Tabela 2 apresenta o direcionamento dos projetos da indústria, sendo massivamente exploradas a orquestração de recursos virtualizados em ambientes *cloud* e a arquitetura NFV. O bloco funcional MANO também é amplamente desenvolvido, assim como a funcionalidade NFVO. Entretanto, as funcionalidades de integração com gerência e modelo de negócios tem baixa investigação (MAMUSHIANE et al., 2019).

Tabela 2: Padrão de Orquestradores, adaptado de (MAMUSHIANE et al., 2019).

Solução de Orquestração	Líder	Orquestrador de recursos		NFV MANO Framework		
		Cloud	NFV	VNFM	NFVO	OSS/BSS
CORD/XOS	ON.lab	X	X	X	X	
ONAP	Linux Foundation	X	X	X	X	X
OSM	ETSI	X	X	X	X	
Cloudify	GigaSpace	X	X	X	X	
OpenBaton	Fraunhofer	X	X	X		
X-MANO	H2020 Vital		X		X	
Gohan	NTT Data	X	X	X	X	X
Tacker	OpenStack Foundation	X	X	X	X	
TeNor	FP7 T-NOVA	X	X	X		X

Os projetos ONAP (FOUNDATION, 2020a) e OSM (OSNM, 2020), atualmente surgem como os de maior relevância para orquestração e desenvolvimento da arquitetura NFV MANO para ambientes de telecomunicações virtualizados genéricos. Para o vNG-RAN em particular, o OSM e o ONAP estão na mesma direção que as principais iniciativas O-RAN e oferecem suporte à virtualização de contêiner (BONATI et al., 2020; GAVRILOVSKA et al., 2020). Outra iniciativa relevante para a implantação do vNG-RAN é o Mosaic5G (MOSAIC5G.IO, 2020). Essa é uma iniciativa sem fins lucrativos que promove contribuições entre a indústria e a academia. A Mosaic5G projetou a plataforma de código aberto denominada de Kube5G com suporte a arquitetura NFV MANO para orquestrar funções de rede containerizadas, incluindo RAN, sob a ferramenta K8S (NIKAEIN, 2018). A Kube5G explora a funcionalidade dos operadores. Para esse contexto, um operador é um controlador específico que amplia as funcionalidades K8S a fim de criar, configurar e gerenciar instâncias de aplicações complexas para um usuário do K8S (DOBIES; WOOD, 2020). Dessa forma, é possível o desenvolvimento de Kube5G-Operator para automatizar a implantação de recursos e serviços e fornecer dinâmica para a divisão de RAN usando o *software* de simulação aberto OpenAirInterface. Esse *software* é mantido pela OpenAirInterface Software Alliance (OSA), um consórcio sem fins

lucrativos que promove uma comunidade internacional de colaboração industrial e de pesquisa para o desenvolvimento de código aberto para redes de celulares. Iniciada pelo EURECOM Alliance, o consórcio promove o desenvolvimento das redes móveis usando *hardware* não proprietários (ALLIANCE, 2020b).

3.2.2 Iniciativas da Academia

No que se refere às iniciativas acadêmicas, é possível constatar que a literatura segue os mesmos preceitos da indústria, conforme apresentado na Tabela 3. Esses preceitos referem-se ao processo de virtualização, aos tipos de tecnologias, às ferramentas de virtualização e às ferramentas utilizadas para orquestração e gerenciamento, apresentando, entretanto, objetivos distintos, mas inerentes ao amplo tema de orquestração. Além disso, pode-se observar na Tabela 3 para a RAN, o endereçamento dos tipos de unidades aplicadas e seu respectivo modelo de virtualização das funções de rádio, assim como sua desagregação.

Tabela 3: Trabalhos relacionados - Soluções de orquestração para a virtualização e desagregação da arquitetura NG-RAN.

Trabalhos Relacionados	Objetivos da Orquestração	Tecnologia de Virtualização	Ferramenta de Orquestração	Virtualização da RAN	Desagregação da RAN
(DZOGOVIC et al., 2019)	Implementação da C-RAN containerizada com orquestração nativa do K8S	Contêiner	K8S	BBU - VNF RRH - VNF	NE
(NOVAES et al., 2020)	Implementação da C-RAN containerizada com orquestração nativa do K8S	Contêiner	K8S	BBU - VNF RRH - VNF	NE
(KONDEPU et al., 2018)	Orquestração visa alta disponibilidade com a rede <i>fronthaul</i> adaptável e RAN desagregada. Aplica a arquitetura NFV.	Contêiner	NE	CU - VNF DU - VNF	O7-1 e O8
(SAHA et al., 2018)	Orquestração baseada nas redes <i>fronthaul</i>	<i>Hypervisor</i>	TOSCA	BBU - VNF RRH - PNF	O2 e O5
(SEUNG-QUE et al., 2018)	Orquestração manual alinhada com a arquitetura NFV, flexibilidade de implementação de VNFs ou PNFs.	<i>Hypervisor</i>	FLEXELL	CU - VNF RU - PNF	O2
(DALLA-COSTA et al., 2020)	Orquestrador orientado ao balanceamento de carga da C-RAN conforme as limitações do <i>fronthaul</i> e capacidade de RCs	Contêiner	Orchestra	BBU - VNF RRH - VNF	O7
(MATOUSSI et al., 2020)	Orquestrador de otimização do posicionamento da RAN com foco em minimizar o processamento e os RCs.	Contêiner	SPLIT HPSO Orchestrator	BBU - VNF RRH - VNF	D-RAN e O7
(MAKRIS et al., 2019)	Orquestração orientada ao desempenho da rede <i>fronthaul</i> sem fios e migração para redes Wifi. Aplica a arquitetura NFV.	NE	NITOS OSM	CU - VNF DU - VNF	O2

NE - Não Especificado

Em relação aos trabalhos desenvolvidos, Dzogovic et al. (2019) explora a virtualização da RAN por contêineres, com a ferramenta K8S, bem como a arquitetura da RAN LTE baseada em BBU e RRH. Entretanto, a orquestração é investigada de forma nativa pelo K8S e com análise apenas do processo de alocação das VNFs. Sendo assim, a pesquisa tem como foco a validação da RAN e do CN em um ambiente virtualizado pela ferramenta Openstack como gerenciador da infraestrutura de computação em nuvem fe-

derada. A pesquisa também tem como objetivo a comunicação entre a RAN e o CN via *switches* virtuais e conclui com sucesso a montagem de todo o ambiente proposto.

A pesquisa liderada por Novaes et al. (2020) investigou a arquitetura C-RAN virtualizada com base em contêineres pelo *software* OAI (unidade BBU e RRH). Entretanto, o objetivo do trabalho é uma análise de desempenho de escala das unidades da RAN e orquestração pela ferramenta K8S para trazer flexibilidade na alocação de recursos. Em uma direção similar, mas com objetivos distintos, o artigo de Kondepu et al. (2018) propõe uma orquestração de contêineres flexível e orientada pela alta disponibilidade da rede de transporte baseada na arquitetura de referência NFV. A pesquisa utiliza dois tipos de divisão da RAN, a O7a e O8. As ferramentas de orquestração não são explicitadas no trabalho, mas oportunizam interação entre as VNFs da RAN (CU e DU) com o controlador da rede de transporte via *shell scripts* do SO Linux.

Em Saha et al. (2018), observa-se um trabalho desenvolvido sob uma extensão do projeto de código aberto CORD/XOS. O trabalho apresenta a plataforma Mobile-CORD (M-CORD) com uma validação fim-a-fim com dispositivos reais de uma rede C-RAN virtualizada pelo *hypervisor* KVM e análise de desempenho entre as opções de desagregação 2 e 5. Entretanto, o cerne do trabalho foi vinculado à orquestração pela análise de controle de fluxo de tráfego visando a estratégias eficientes para as redes *fronthaul*. Em complemento, o estudo de Seung-Que et al (2018) desenvolveu a arquitetura RAN desagregada, em que a CU é virtualizada pelo *hypervisor* Vmware EXSI e a DU em funções de rede físicas PNF. O bloco funcional MANO também foi aplicado ao estudo com as ferramentas FLEXCELL e Vmware Center. No entanto, o objetivo da orquestração diz respeito a alocação das VNFs conforme demanda manual dos operadores de rede.

O trabalho de Dalla et al (2020) propõe um orquestrador para balanceamento de carga da C-RAN conforme as limitações do *fronthaul* com foco na flexibilidade da opção de desagregação 7 (possibilidade de 4 diferentes tipos de desagregação), e consequente posicionamento das funções de rádio virtualizadas, em uma infraestrutura computacional com três níveis: nuvem, regional e borda. Para a implementação da proposta foram utilizadas as ferramentas do projeto FUTEBOL (BOTH et al., 2019) para a gestão de infraestrutura computacional, e, além disso, foi desenvolvido o orquestrador Orchestra. No mesmo sentido, o trabalho Matoussi et al (2020) desenvolveu a arquitetura AgilRAN e o orquestrador SPLIT-HPSO. Tendo a arquitetura C-RAN virtualizada em contêineres do emulador OAI, a desagregação por funções de processamento e controle de fluxo do tráfego para diferenciação de serviços. Neste sentido, com os objetivos específicos de orquestração a redução do processamento computacional e eficiência do *fronthaul*.

Finalmente, o trabalho de Makris et al (2019) utilizou uma versão estendida da ferramenta de código aberto OSM, denominada de NITOS OSM. O emulador OAI foi utilizado para a desagregação entre as unidades de rádio CU e DU (opção 2). As tecnologias

e ferramentas de virtualização não são explicitadas no trabalho necessitando também de maiores informações sobre o orquestrador, apesar de usar a ferramenta OSM para a gestão de infraestrutura computacional e orquestração. Uma análise de desempenho baseada em disponibilidade foi realizada com diferentes tipos de redes *backhaul*.

3.3 Modelos de posicionamento para orquestração das funções de rádio desagregadas e virtualizadas

Com base na imersão das redes móveis nos conceitos de *software*, o posicionamento das funções virtualizadas é amplamente investigadas na literatura acadêmica. Para a RAN, o tópico tem ganhando espaço a partir da desagregação das funções de rádio e virtualização da arquitetura. Entretanto, devido à escala demandada pela RAN, a literatura tem concentrado esforços em soluções próximas de ótimas para o tópico de posicionamento das funções virtualizadas, a partir de métricas impostas pelas características de rede e pelos recursos. Os trabalhos são desenvolvidos sob provas de conceito matemáticos, principalmente modelos matemáticos ou programações lineares inteiras (ILP - *Integer Linear Programming*) e com aplicações de algoritmos aproximativos ou heurísticos. Com as exceções de Arouk et al. (2018), Yusupov et al. (2018), Garcia-Saavedra et al. (2018), Murti et al. (2020a) e Murti et al. (2020b) que obtiveram resultados ótimos dentro de contextos de posicionamento específicos, sendo os dois primeiros com a aplicação da técnica de programa de número inteiro misto restrito quadraticamente (MIQCP - *Mixed-Integer Quadratically Constrained Program*) e os dois últimos com o método de decomposição de Bender's aliados a técnicas de linearização.

De forma geral, os trabalhos concentram-se na centralização da CU ou BBU (dependendo da geração da RAN) para uma distribuição das camadas mais baixas da pilha de protocolos, além de em profunda análise das redes de transporte, principalmente da rede *fronthaul*. A Tabela 4 reúne os principais trabalhos desenvolvidos relacionados ao posicionamento da RAN virtualizada e desagregada. Além disso, a mesma é composta por uma análise do objetivo da pesquisa, pela arquitetura, pelas unidades de rádio e pela quantidade de desagregações da RAN, assim como, pelas topologias de transporte aplicadas a alocação de recursos computacionais.

Primeiramente, os trabalhos de Garcia-Saavedra et al. (2018) e Fonseca et al. (2019) possuem objetivos semelhantes, uma vez que o segundo sugere uma análise comparativa do primeiro. Ambos buscam a centralização da CU, porém Garcia-Saavedra et al. (2018) adiciona a análise de custos para C-RAN e D-RAN, com a análise do *fronthaul* e da computação na borda. Os dois trabalhos utilizam a técnica *Bender Decomposition* e propõem desagregação da RAN em três níveis. Como principal diferença, Fonseca et al. (2019) modela matematicamente com o conceito de programação linear inteira mista

Tabela 4: Trabalhos Relacionados - Soluções de otimização do posicionamento da RAN.

Trabalhos Relacionados	Objetivos de Posicionamento	RAN			Redes Transporte	Recursos Comput.
		Arquitetura	Unidades de Rádio	Desag.		
(FONSECA et al., 2019)	Maximizar a centralização da CU	CU-RU	C-RAN	3	Fronthaul	✗
(GARCIA-SAAVEDRA et al., 2018)	Maximizar a centralização da CU e avaliar a latência na borda	CU-RU	C-RAN	3	Fronthaul	✓
(BHAMARE et al., 2018)	Maximizar a centralização da BBU e minimizar a latência	C-RAN	BBU-RRH	-	Fronthaul	✗
(AROUK et al., 2017)	Minimizar o custo do encadeamento de VNF	C-RAN	BBU-RRH	1	Fronthaul	✓
(MOLNER et al., 2019)	Maximizar a distribuição das DUs sob redes ópticas não dedicadas	NG-RAN	CU-DU	2	Fronthaul Backhaul	✗
(RUIZ et al., 2018)	Maximizar a centralização da BBU para reduzir CPU e garantir BW	C-RAN	BBU-RRH	1	Fronthaul	✓
(MAHAPATRA et al., 2017)	Identificar a posição ideal do centroide para a BBU	C-RAN	BBU-RRH	1	Fronthaul Backhaul	✗
(AROUK et al., 2018)	Minimizar os custos de rede	C-RAN	CU-DU-RU	1	Crosshaul	✓
(MASOUDI et al., 2020)	Minimizar os COTS para o agrupamento de DUs	C-RAN	DU-RU	4	Fronthaul	✓
(HARUTYUNYAN et al., 2020)	CU dinâmica para minimizar o consumo de energia	NG-RAN	CU-DU	4	Fronthaul	✗
(SONG et al., 2019)	Minimizar o uso de recursos computacionais	D-RAN	Monolithic	1	-	✓
(HARUTYUNYAN et al., 2018)	Minimizar a interferência entre células e da rede <i>fronthaul</i>	C-RAN	CU-DU	4	Fronthaul	✗
(YUSUPOV et al., 2018)	Minimizar os nós de rede e latência, e maximizar a taxa de dados	NG-RAN	CU-DU-RU	4	Crosshaul	✓
(MURTI et al., 2020a), (MURTI et al., 2020b)	Minimizar o custo da vRAN e do roteamento, com base no posicionamento centralizado da CU	C-RAN	CU-DU	3	Fronthaul	✓
(MATOUSSI et al., 2020)	Minimizar o uso de recursos computacionais e do <i>fronthaul</i>	C-RAN	BBU-RRH	3	Fronthaul	✓

(MILP - *Mixed Integer Linear Programming*), utilizando variáveis inteiras e não inteiras, enquanto Garcia-Saavedra et al. (2018) utiliza ILP e avalia a solução em três redes reais e uma sintética.

A redução dos custos operacionais em conjunto com a centralização da BBU da arquitetura C-RAN e a minimização da métrica de latência em um ambiente de computação em nuvem distribuído é desenvolvido por Bhamare et al. (2018). O modelo matemático é baseado em ILP e dois algoritmos heurísticos são desenvolvidos para alcançar seus objetivos: (i) a ramificação e limite (BnB - *Branch-and-Bound*) e (ii) a normalização simulada (SA - *Simulated Annealing*). Com distribuição do ambiente de computação em nuvem em três níveis, Arouk et al. (2017) indica, além da centralização, um balanceamento de carga dinâmico, aliado a um algoritmo heurístico denominado posicionamento multi-objetivo (MOP - *Multi-Objective Placement*). Diferentemente dos trabalhos anteriores, além da centralização como objetivo, Molner et al. (2019) avalia profundamente as redes *backhaul* e *fronthaul* com uma topologia de hierarquias em anel, utilizando, no entanto, ambientes computacionais pré-definidos na borda e na nuvem.

A utilização de algoritmos não heurísticos é averiguada em Ruiz et al. (2018) e Mahapatra et al. (2017). No primeiro, o algoritmo genético (GA - *Genetic Algorithm*) foi implementado para validar a centralização da BBU com o propósito de reduzir a CPU e garantir a taxa de dados sob redes *Crosshaul* ópticas. O segundo sugere o algoritmo *K-means* para descoberta do centróide do *cluster* para a centralização ideal da BBU, também sob redes *Crosshaul* ópticas. A métrica custo também é tratada para o posicionamento da RAN. Nesse sentido, os trabalhos de Arouk et al. (2018) e de Masoudi et al. (2020) ende-

reçam suas respectivas investigações. Em Arouk et al. (2018), embora a proposta seja a avaliação geral do custo, o trabalho foi direcionado à escolha de recursos computacionais e limitações dos *links* de transporte para um cenário de planejamento de rede, enquanto que, em Masoudi et al. (2020), o algoritmo é direcionado para a migração da D-RAN para C-RAN, com foco no agrupamento de diversas RUs em uma CU sob uma rede *fronthaul* de tecnologia específica.

Baseado em objetivos específicos, porém distintos, os trabalhos Harutyunyan et al. (2018) e Harutyunyan et al. (2020), e Song et al. (2019), também pesquisam sobre o posicionamento da RAN. No que se refere a Harutyunyan et al. (2018) o alvo da pesquisa é o desenvolvimento da C-RAN com a mitigação da interferência entre células e a redução do *fronthaul*. Para este, foram desenvolvidos dois modelos matemáticos e dois algoritmos heurísticos, sendo um estático e outro dinâmico, respectivamente. Em Harutyunyan et al. (2020), o posicionamento dinâmico da CU foi investigado de forma específica para enlace sem fios. De outro modo, Song et al. (2019) fundamenta sua pesquisa no posicionamento de VNF, bem como no processo de alocação e no encadeamento das funções. Entretanto, não avalia a desagregação da RAN e as redes de transporte.

Outros relevantes trabalhos foram desenvolvidos para análise do problema de posicionamento da RAN, tais como Yusupov et al. (2018), dois trabalhos complementares de Murti et al. (2020a) e Murti et al. (2020b), e Matoussi et al. (2020). No primeiro trabalho, Yusupov et al. (2018) investiga a minimização da latência e dos nós computacionais, somados à maximização da taxa de dados dos *links* sob a arquitetura NG-RAN com quatro opções de desagregação e topologia sintética de baixa escala, além da comparação de dois modelos matemáticos de distintos direcionamentos pelos serviços de latência extremamente baixa e altas taxas de dados. No entanto, o trabalho não leva em consideração o agrupamento das funções de rádio e as características de rádio da RAN. Por outro lado, os trabalhos de Murti et al. (2020a) e Murti et al. (2020b) focam na minimização dos custos operacionais de rede visando o agrupamento de diversas DUs em poucas CUs com base na escolha da melhor desagregação. Entretanto, o trabalho ainda foi desenvolvido sob duas unidades de rádio. Por fim, Matoussi et al. (2020) (trabalho abordado na Subseção 3.2.2 com um experimento de virtualização da C-RAN) apresenta um algoritmo de posicionamento orientado pela estrutura do emulador OAI para reduzir o uso de recursos computacionais, o consumo de energia e o tráfego do *fronthaul*.

3.4 Análise dos Trabalhos Relacionados e Oportunidades de Pesquisa

Esta seção explora as lacunas e assuntos em aberto na literatura relacionados às soluções de orquestração da otimização do posicionamento das funções de rádio da arquitetura vNG-RAN. Primeiramente, os dois tópicos são tratados majoritariamente de forma

separada, conforme detalhado nas Seções 3.2 e 3.3. Portanto, apesar de todo o desenvolvimento em andamento da vNG-RAN para as redes móveis de quinta geração, até onde é de conhecimento do corrente trabalho, ainda há uma lacuna de pesquisa aberta para o desenvolvimento de uma solução específica para a orquestração do posicionamento das funções de rádio virtualizadas.

Nesse sentido, conforme explorado na Seção 3.2, a abordagem das lacunas das soluções é dividida em iniciativas da indústria e da academia. Como principal diferença entre ambas identifica-se o desenvolvimento de soluções genéricas de orquestração da otimização de posicionamento de funções virtualizadas (sem especialização para vNG-RAN) na indústria e, na academia, soluções com abordagens específicas ainda em C-RAN. Abaixo, segue detalhamento das lacunas encontradas:

1. Soluções de Orquestração na Indústria

- (a) Soluções de Orquestração alinhadas com a arquitetura NFV.
- (b) Principais soluções de orquestração para redes móveis são os projetos OSM e ONAP.
- (c) OSM aborda o tema de posicionamento de VNFs via a contribuição da parceira ARCTOS LABs, mas sem especificação para a RAN (ERIKSSON, 2017).
- (d) ONAP possui estreita relação com as iniciativas O-RAN e está desenvolvendo uma plataforma genérica de otimização denominada de *ONAP Optimization Framework* (OOF), ainda sem suporte especializado para a RAN. Ademais, há *roadmap* para ciência de topologia *crosshaul*.
- (e) Projeto Mosaic5G possui uma solução de orquestração para a RAN, porém sem suporte à otimização do posicionamento das funções de rádio virtualizadas.

2. Soluções de Orquestrações na Academia

- (a) Sem trabalhos na academia com orquestradores com suporte a vNG-RAN.
- (b) Soluções são majoritariamente executadas por funções nativas das ferramentas aplicadas (sem otimização de posicionamento), por exemplo, *Openstack* (LUO et al., 2018) e K8S (NOVAES et al., 2020), sendo estas sem especialidade dos requisitos da RAN e da topologia *crosshaul*.
- (c) Desenvolvimento de ferramentas para as soluções próprias de orquestração da RAN, como realizado em Seung-Que et al. (2018), Dalla-Costa et al. (2020) e Matoussi et al. (2020)..

- (d) Em particular, Makris et al. (2019) utilizou uma versão estendida do OSM, mas com orquestração orientada ao *fronthaul*.
- (e) Soluções de orquestração baseadas em balanceamento de carga (DALLA-COSTA et al., 2020).

Sob outra perspectiva, conforme diagnosticado na presente seção, o objetivo de posicionamento da RAN é explorado de forma individual e com baixa conexão com soluções de orquestradores. Sendo assim, o tema de pesquisa otimização do posicionamento, assim como orquestradores, ainda é majoritariamente analisado para a arquitetura C-RAN, tendo como principais objetivos:

- a) A centralização das unidades CUs ou BBU's (BHAMARE et al., 2018; FONSECA et al., 2019; GARCIA-SAAVEDRA et al., 2018; MAHAPATRA et al., 2017; MOLNER et al., 2019; MURTI et al., 2020b; RUIZ et al., 2018)
- b) A otimização do encaminhamento do *backhaul* e do *fronthaul* (FONSECA et al., 2019; GARCIA-SAAVEDRA et al., 2018; HARUTYUNYAN et al., 2018, 2020; MAHAPATRA et al., 2017; MURTI et al., 2020b), assim como da latência (BHAMARE et al., 2018; MOLNER et al., 2019)
- c) A redução do uso de recursos computacionais (BHAMARE et al., 2018; RUIZ et al., 2018; SONG et al., 2019)
- d) Análise das decisões específicas dos requisitos da RAN (HARUTYUNYAN et al., 2018)
- e) A minimização de custos e consequente investimentos financeiros (AROUK et al., 2017; MASOUDI et al., 2020; MURTI et al., 2020b).

Embora corroborem para o desenvolvimento da RAN, os trabalhos discutidos anteriormente carecem de alinhamento com os avanços da NG-RAN e, principalmente, de sintonia com a análise de recursos computacionais, com as redes *crosshaul* e com as altas possibilidades de desagregação das três unidades de rádio (trabalhos atingem no máximo 4 opções de desagregação).

Apesar de a literatura apresentar diferentes segmentos para o desenvolvimento de arquiteturas de orquestração e algoritmos de posicionamento, o trabalho de Matoussi et al. (2020) propõe tal conexão com a arquitetura AgilRAN. Todavia, o mesmo possui lacunas para com os recentes avanços da RAN e, no que se refere à arquitetura, ainda trabalha com a arquitetura C-RAN. Para o desenvolvimento da arquitetura virtualizada, embora utilize contêiner e *Docker*, o trabalho não explora ferramentas de orquestração de

VNFs e propõe um algoritmo externo para tal tomada de decisões. Ainda na arquitetura virtualizada, o estudo faz referência à arquitetura NFV MANO, porém, com apenas uma interface de comunicação, sem detalhamento do seu funcionamento. O trabalho possui ciência de topologia do *fronthaul*, mas a escala e o tipo de topologia não são explorados. Em se tratando do algoritmo, o mesmo é orientado às características particulares de processamento do emulador OAI e utiliza apenas as subdivisões da opção de desagregação 7.

Para concluir, considerando as limitações expostas pelos trabalhos propostos, é possível identificar a necessidade de desenvolvimento de uma solução de orquestração para a otimização do posicionamento das funções de rádio de nova geração virtualizada. Esta por sua vez, deve ser alinhada com os recentes avanços da arquitetura vNG-RAN, com a alta flexibilidade de desagregação, com as características das funções de rádio, com a ciência de topologia *crosshaul* e com a eficiência de recursos computacionais. Além disso, deve também estar orientada à arquitetura de referência NFV, aos recentes avanços de virtualização e às iniciativas da indústria. A partir disso, tais oportunidades são aprofundadas no seguinte capítulo.

4 SOLUÇÃO ORQUESTRAÇÃO E DE POSICIONAMENTO PLACERAN

No presente capítulo, é apresentada a solução de posicionamento e orquestração PlaceRAN para as redes de rádios de acesso de nova geração. O capítulo é dividido em três seções. A seção 4.1 orienta as tomadas de decisão da elaboração da solução. Na seção 4.2 é introduzido o modelo do orquestrador com ênfase na organização dos componentes, no alinhamento com a arquitetura NFV e no funcionamento da solução fim-a-fim, incluindo a otimização do posicionamento. Por fim, a seção 4.3 define o modelo do sistema no que se refere especificamente à otimização do posicionamento da vNG-RAN e apresenta a formulação matemática da solução para o problema de otimização de pesquisa.

4.1 Decisões de Projeto

A solução PlaceRAN busca a orquestração e a otimização do posicionamento das funções virtualizadas de rádio de acesso da nova geração para planejamento de rede, que são desenvolvidas para aplicação nas redes de quinta geração. Para alcançar esse objetivo, o trabalho foi dividido em dois sub-objetivos, sendo eles: (i) o modelo para o orquestrador e (ii) a modelagem matemática para a otimização. Sendo assim, optou-se por utilizar a principal arquitetura de virtualização aplicada para redes móveis, a arquitetura de referência NFV para o desenvolvimento do orquestrador e a modelagem do sistema (ETSI, 2014a). O orquestrador é projetado conforme bloco funcional NFVO da arquitetura NFV com a divisão entre gestão de recursos e gestão de serviços. Para a alocação dos recursos computacionais, as métricas também foram embasadas na arquitetura NFV. A tecnologia de virtualização definida é a baseada em contêineres, seguindo as orientações das iniciativas O-RAN (ALLIANCE, 2020a) e por conta de tecnologia ser a mais indicada para a RAN por ter alta flexibilidade, granularidade, velocidade de provisionamento e baixo consumo de recursos (ETSI, 2016; GAVRILOVSKA et al., 2020). O modelo matemático, segue as prerrogativas de análise combinatória para otimização com abordagens exatas e que visam soluções ótimas ou próximas do ótimo. Dado este contexto, as decisões de projeto tomadas são divididas nas duas sub-divisões apresentadas:

1. Decisões para o Orquestrador

- (a) Foco em planejamento de redes com infraestruturas virtualizadas, tendo, como resultado, uma implementação de uma rede nova.
- (b) Baseado na arquitetura de referência NFV e na tecnologia de contêineres, sem suporte a PNFs.

- (c) Interoperável com as iniciativas O-RAN e com a integração das principais ferramentas de orquestração da quinta geração de redes móveis.
- (d) Ciente da topologia *Crosshaul* com entrada única, pois este suporte é vital para a vNG-RAN.

2. Decisões para a modelagem da otimização

- (a) Baseado em conceitos de otimização de abordagens exatas, busca da solução ótima para o modelo.
- (b) A eficiência temporal não é um requisito do modelo por ter o objetivo de planejamento de rede.
- (c) O modelo não restringe tipos específicos de topologias.
- (d) Nós de RCs e do *crosshaul* são co-localizados no mesmo ambiente geográfico e, com isso, a interligação entre os mesmos não foi inserindo aumento nos valores de taxa de bits e latência.
- (e) Resultado da otimização é único (não há resposta parcial), ou seja, possui ou não uma solução viável.
- (f) As métricas aplicadas ao modelo são : CPU e memória para os recursos computacionais, e latência e taxa de dados para os *links* das redes *crosshaul*.

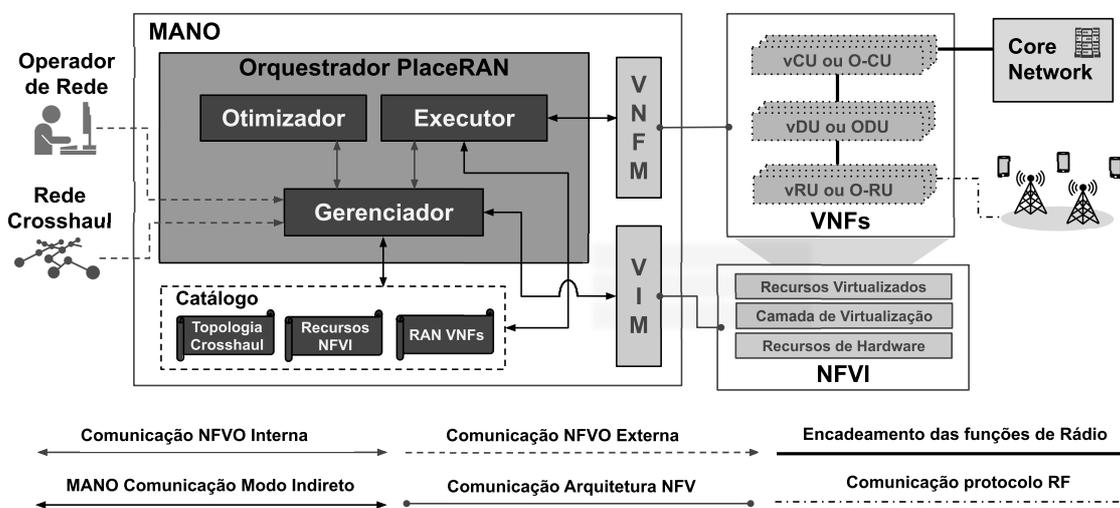
4.2 Orquestração das Funções de Rádio Virtualizadas

O orquestrador tem como objetivo o posicionamento das funções de rádio virtualizada em uma da arquitetura vNG-RAN com base na arquitetura de referência NFV, proposta pela ETSI, conforme indicado na Figura 8. Neste sentido, a arquitetura apresentada abrange todos os componentes da arquitetura NFV, incluindo os blocos MANO NFVO, VNF e NFVI. Além disso, a arquitetura de referência possui suporte ao CN e a UEs, a entrada de dados externas pelo operador de rede e a topologia *crosshaul*. Embora a arquitetura abranja as redes móveis de forma fim-a-fim, o objetivo do trabalho diz respeito a orquestração conforme delimitações apresentadas na seção 4.1. Sendo assim, a presente seção é dividida em duas subseções. A subseção 4.2.1 detalha todos os componentes da arquitetura e as interações com o orquestrador, enquanto a subseção 4.2.2 apresenta o detalhamento e o funcionamento do orquestrador desenvolvido na solução PlaceRAN.

4.2.1 Componentes da Arquitetura

A presente subseção explora a arquitetura e os componentes para o desenvolvimento da solução de orquestração. A seguir, são detalhados os oito componentes que compõem

Figura 8: A solução PlaceRAN é baseada na arquitetura referência NFV ETSI. A figura detalha, além do orquestrador inserido no bloco funcional MANO NFVO, os blocos funcionais MANO VIM e VNFM e os componentes NFVI e VNFs da arquitetura NFV, assim como os componentes das redes móveis CN e UE, além dos dados de entrada do operador de rede e das redes *crosshaul*. O orquestrador possui quatro blocos funcionais específicos. O primeiro refere-se a base denominada de Catálogo e divide-se nas bases de informações de Topologia *Crosshaul* e de Recursos NFVI, além das funções virtualizadas da RAN. Os outros três blocos funcionais são o Otimizador responsável pelo processo de otimização do posicionamento, o bloco Gerenciador que realiza toda a interação interna e externa do orquestrador e o bloco Executor que envia o plano de alocação e o encadeamento das VNFs da RAN.



Fonte: Elaborada pelo autor

a arquitetura para o desenvolvimento do orquestrador.

- (i) **Operador de Rede** - Responsável pela entrada de dados no que se refere a quantidade de vRUs a serem posicionadas por RCs. Posicionamento dos vDUs e vCUs serão abordados conforme seção 4.3.
- (ii) **Rede Crosshaul** - Dado de entrada da arquitetura com as informações de latência e capacidade de taxa de dados dos *link* entre os nós de transporte.
- (iii) **Core Network (CN)** - Último ponto do encadeamento das funções de rádio virtualizadas, além de representar a comunicação fim-a-fim das redes móveis, desde o UE até a conexão com as redes externas, principalmente a Internet.
- (iv) **UE** - Dispositivo com comunicação direta com a interface aérea do rádio da RAN. Realiza a autenticação no CN para a conexão com as redes externas e, consequentemente, gera tráfego baseado nos serviços das redes móveis.

- (v) **NFVI** - Componente da arquitetura NFV responsável pelos RCs e infraestrutura virtualizada, seja esta com base em *hardware* ou em máquinas virtuais.
- (vi) **VNFs** - Componente da arquitetura NFV responsável pelo *software* de virtualização das funções de rádio, bem como pelo encadeamento das funções e, consequentemente, pelo tráfego entre as funções de rádio virtualizadas.
- (vii) **MANO VIM** - Responsável por prover o status das métricas computacionais dos RCs do componente NFVI para o orquestrador diretamente pela função Gerenciador, tais como a disponibilidade de armazenamento, a memória e o processamento.
- (viii) **MANO VNFM** - Tem a finalidade de receber, do orquestrador, pela função Executor, o posicionamento da alocação das funções de rádio virtualizadas, além de executar a alocação sob a infraestrutura virtualizada e prover o encadeamento dessas funções.

4.2.2 Solução de Orquestração

De forma geral, para o posicionamento das funções de rádio virtualizadas, o orquestrador é o responsável por toda a gestão e coordenação do sistema. O mesmo está localizado no componente NFVO do macro bloco MANO da arquitetura NFV. Dentre as principais atribuições, estão: (i) receber a entrada da topologia da rede pelo operador de rede, (ii) analisar os recursos computacionais disponíveis, (iii) otimizar o posicionamento das VNFs e (iv) enviar as informações de execução para a implantação do posicionamento das funções de rádio virtualizadas. O orquestrador possui interação com todo o sistema. Dentro do MANO, sua interação é com os blocos VIM e VNFM para coleta de dados das informações de infraestrutura e execução da escala da rede. A solução não prevê interação entre os blocos VNFM e VIM, pois a mesma adota o modo indireto que implica na comunicação direta entre o bloco NFVO e os blocos VIM e VNFM (ETSI, 2014b), concentrando, assim, todas as ações no orquestrador. Conforme as atribuições descritas, o orquestrador proposto é subdividido em 4 sub-blocos funcionais: (i) Catálogos; (ii) Gerenciador; (iii) Otimizador; e (iv) Executor. Os quatro sub-blocos funcionais são descritos abaixo:

- (i) **Catálogos** - armazena dados sobre os serviços, as funções e os recursos disponíveis na rede. Para o orquestrador, o bloco Catálogo é dividido em três recursos distintos. O primeiro, denominado de RAN VNFs, opera como um repositório das VNFs padrões dos nós vCU, vDU e vRU. A requisição das VNFs é solicitada pela função Executor e enviada diretamente para o bloco VNFM. A partir deste, o bloco

VNFM tem acesso às VNFs para alocação das mesmas conforme orientação da função Executor. Os catálogos Recursos NFVI e Topologia *Crosshaul* também operam como repositórios, porém, nestes casos, com funcionalidade de armazenamento de dados referentes aos recursos (disponíveis e alocados) da NFVI e aos *links* e suas respectivas capacidades (latência e taxa de dados). Para o catálogo Recursos NFVI, a atualização é efetuada com base na comunicação entre o sub-bloco Gerenciador e o bloco VIM. As informações pertinentes para o orquestrador são as disponibilidades de recursos de CPU, a memória e a disponibilidade em disco dos RCs. No que se refere ao catálogo Topologia *Crosshaul*, a entrada de dados é realizada de forma externa também com o sub-bloco Gerenciador. Com base nesta primeira ação, o Gerenciador atualiza as informações no catálogo Topologia *Crosshaul*. Por fim, para a etapa de posicionamento, ambos os catálogos são requisitados pelo Gerenciador, para o processo de otimização do posicionamento das funções de rádio virtualizadas.

- (ii) **Gerenciador** - centraliza e determina as ações da solução de orquestração, é o responsável pelas tomadas de decisões e está posicionado como a funcionalidade de orquestração de recursos do NFVO. Inicialmente, para começar o processo de planejamento da vNG-RAN, o Gerenciador recebe a entrada de dados externos das vRUs pelo operador de rede e da topologia *Crosshaul*. O Gerenciador opera sob demanda de eventos externos. Com base nestas entradas, o Gerenciador requisita ao bloco VIM a atualização dos catálogos de Recursos NFVI. Os dados, o conjunto de entradas externas e os recursos da NFVI são enviados para o sub-bloco Otimizador para a análise do posicionamento ótimo das funções de rádio. O resultado do Otimizador é analisado pela função Gerenciador que envia o plano de alocação e encadeamento de VNFs para o sub-bloco Executor. Outra ação após a finalização do Otimizador é a atualização dos catálogos pelo Gerenciador.
- (iii) **Executor** - atua na aplicação das configurações demandadas pela Otimização conforme as entradas do Gerenciador. Para efetuar as execuções, esse módulo interage com o blocos VNFM e VIM. O sub-bloco está posicionado no NFVO como o orquestrador dos serviços de rede. A interação com o bloco funcional VNFM atribui as atividades de criação e encadeamento das VNFs das funções de rádio. Além disso, a interação com o VIM opera na alocação das VNFs nos recursos dos RCs. Sendo assim, o Executor processa as requisições enviadas pelo Gerenciador, requisita as VNFs do Catálogo VNFs RAN e envia para o VIM e VNFM as alocações das VNFs nos RCs pertinentes.
- (iv) **Otimizador** - responsável pelo processo decisório de otimização do posiciona-

mento das funções de rádio virtualizadas. Com base na arquitetura, o Otimizador recebe requisições e retorna soluções vindas apenas do Gerenciador. Esta interação é efetuada a partir da solicitação do Gerenciador para a análise de otimização de posicionamento da rede. A solicitação é realizada com a entrada do conjunto de informações repassadas pelo operador de rede, pelos catálogos Topologia *Crosshaul*, pelos Recursos NFVI e pelo consumo computacional das VNFs do catálogo VNFs RAN. Como resultado, o Otimizador entrega uma análise de viabilidade ou não do sistema. Caso a solução seja viável, o posicionamento das unidades vRU, vDU e vCU é repassado ao Gerenciador, bem como as opções de desagregação da RAN, os RCs a serem alocados, o encadeamento da funções de rádio e o encaminhamento das redes *crosshaul*.

A Figura 9 apresenta o diagrama de sequência com a sucessão de acontecimentos da solução de orquestração de posicionamento PlaceRAN e da arquitetura proposta. O diagrama de sequência tem início no evento de entrada do operador de rede, a gestão da função Gerenciador sob o sistema e, principalmente, as funções Otimizador, Monitor e Executor. O diagrama de sequência, também exibe a interação das tarefas entre o orquestrador NFVO e os blocos VNFM e VIM, assim como a conclusão da operação (alocação das VNFs nos recursos) e realimentação de execução ou não ao operador.

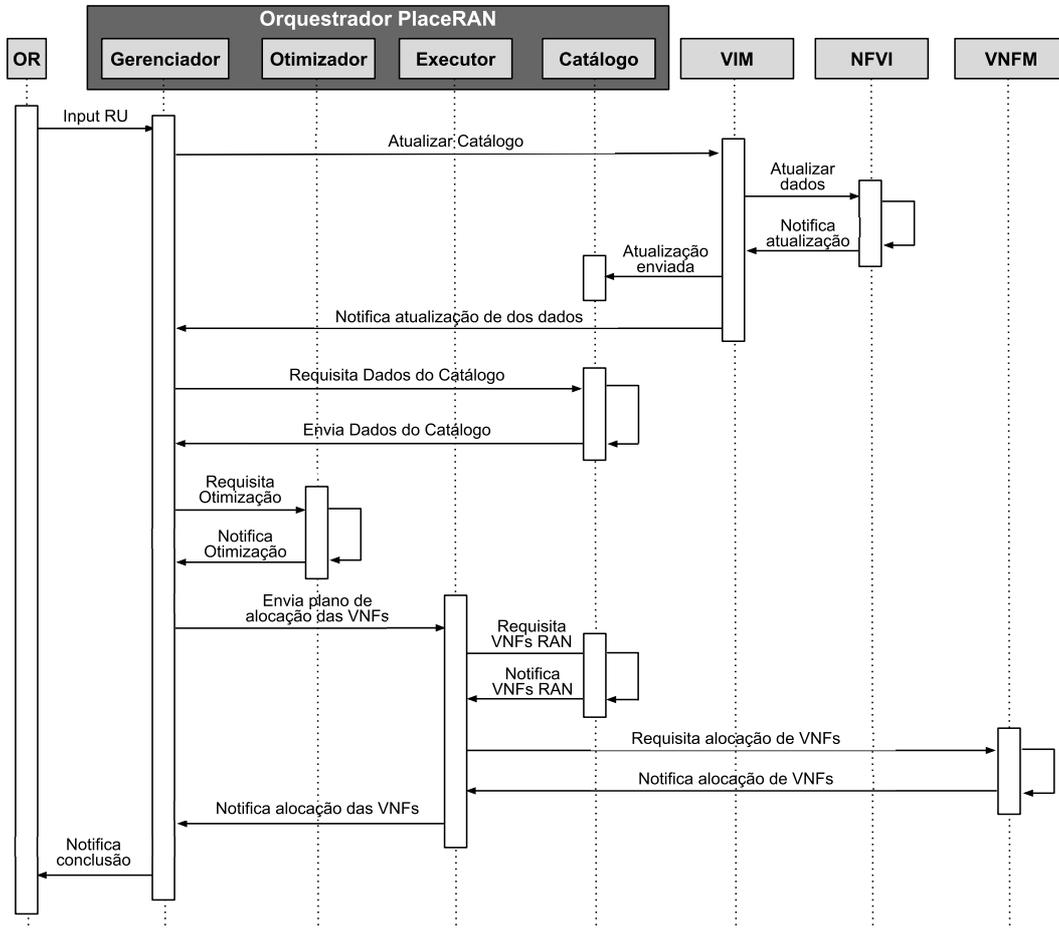
4.3 Modelagem do Otimizador de Posicionamento

Nessa seção, o modelo do sub-bloco Otimizador é explorado, uma vez que ele é o responsável por suportar e orientar as tomadas de decisões para o posicionamento das funções de rádio virtualizadas. Sua função é, também, avaliar todas as restrições impostas pela rede, tendo papel vital para o processo de alocação e encadeamento realizado pelo orquestrador. A seção é subdividida em três partes. A subseção 4.3.1 elucida o contexto da modelagem, enquanto a subseção 4.3.2 explora a modelagem do sistema. Por fim, a subseção 4.3.3 apresenta a formulação para a resolução do problema combinatório de posicionamento das funções virtualizadas de rádio.

4.3.1 Contextualização do Otimizador

A ideia fundamental da desagregação da vNG-RAN é decompor as funções de rádio em unidades virtualizadas que podem ser distribuídas para funcionar em diferentes RCs, ou seja, uma abordagem não monolítica. Portanto, é necessário definir qual a desagregação deve ser escolhida e quais condições devem ser satisfeitas para que haja o alcance dos requisitos. A vNG-RAN desagregada possui duas orientações principais. A primeira

Figura 9: O diagrama de seqüência, apresenta as etapas do processo decisório proposto pela arquitetura PlaceRAN. Desde a entrada do operador de rede até a alocação da função de rádio virtualizada.



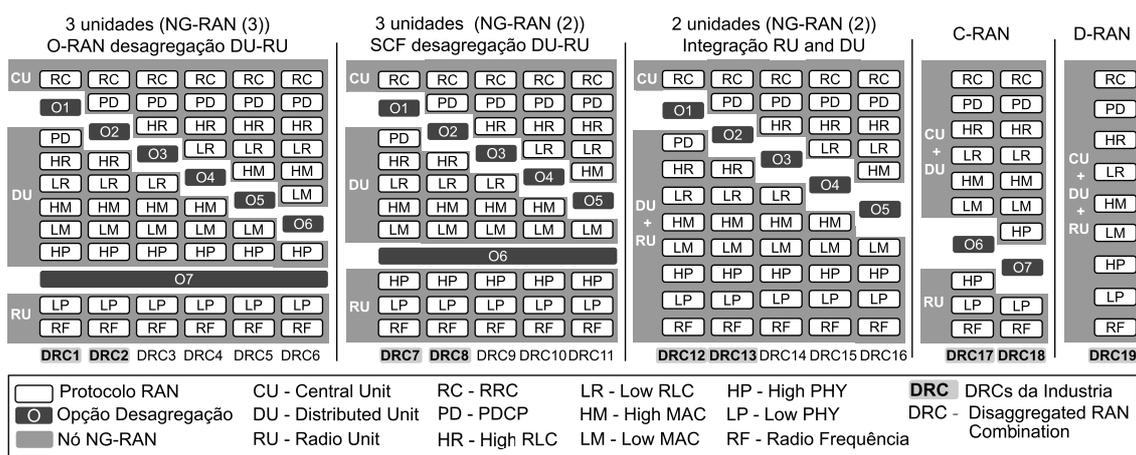
Fonte: Elaborada pelo autor

é a orientação indicada pelos conceitos de oito divisões funcionais que especificam as funções de rádio e consequente unidade de rádio (vCU, vDU e vRU). A segunda é direcionada pelos requisitos de comunicação entre funções de rádio (latência e taxa de dados introduzidos na Tabela 1) conforme as exigências da opção de desagregação determinada (características da pilha de protocolos da RAN) (GAVRILOVSKA et al., 2020; LARSEN et al., 2018).

Conforme elucidado no capítulo 2, cada unidade de rádio (CU, DU e RU) para a vNG-RAN é uma VNF desagregada em três funções independentes. Cada unidade de rádio é composta por um conjunto de protocolos da RAN transformado em VNFs. Sendo assim, há uma análise combinatória que relaciona as unidades/funções virtualizadas e as divisões funcionais guiadas pela pilha de protocolos da RAN. Com base nestes princípios, para desenvolver a otimização do posicionamento da vNG-RAN, o presente trabalho de-

envolveu o conceito de combinação de desagregação da RAN (DRC - *Disaggregated RAN Combination*) conforme apresentado na Figura 10. Os DRCs são elaborados com base nos cenários NG-RAN da seção 2.1.2 (ITU-T, 2018), nas três unidades da NG-RAN (CU, DU e RU) e em sete opções de desagregação (a opção 8 não é explorada, pois ainda não pode ser virtualizada). Cada divisão funcional e grupo de protocolos correspondem a uma VNF de unidade de rádio específico, caracterizada em um DRC. O conceito de DRC desenvolvido no presente trabalho, representa a preservação da ordem de pilha do protocolo durante o processamento dos VNFs. Dezenove DRCs são mapeados considerando as sete opções de divisão, conforme exibido na Figura 10 (cada coluna da Figura 10 é um diferente DRC e divisão funcional). Certos DRCs não são usados na prática porque não são eficazes em termos de custos ou vantagens, como, por exemplo, *midhaul* com latência inferior a 1ms (Tabela 1 opção de desagregação 5), que é altamente restritivo para comunicação em redes de transporte (LANDON, 2018).

Figura 10: Combinações de desagregação da RAN (DRC - *Disaggregated RAN Combination*) para o desenvolvimento das funções de rádio virtualizadas.



Fonte: Elaborada pelo autor

Para o desenvolvimento da otimização, optou-se por utilizar nove DRCs (destacados na Figura 10 em cinza e negrito), pois são efetivamente adotados em implantações de vNG-RAN, cuja escolha é baseada em órgãos de padronização e iniciativas da indústria (3GPP, 2018; ETSI, 2019; ITU-T, 2018). Em ambas as arquiteturas de três unidades independentes (O-RAN e SCF), o foco está nas opções de desagregação O1 e O2 para CU-DU e O6 e O7 para DU-RU (Figura 10 - DRC1, DRC2, DRC7 e DRC8). A indústria considera a opção de desagregação O1 como uma possibilidade do plano de dados descentralizado, enquanto que a opção O2 é consolidada pelas instituições padronizadoras 3GPP e ITU-T por meio da interface F1, que é uma referência da indústria para as iniciativas O-RAN e SCF. Por outro lado, a opção O6 oportuniza uma desagregação de menor impacto de

requisitos, enquanto o O7 (em específico O7-2 para O-RAN) contribui para uma maior eficiência nas funcionalidades de rádio (ALLIANCE, 2020a; ITU-T, 2018; MAVENIR, 2019). Dois DRCs da indústria foram escolhidos para a integração DU e RU (Figura 10 - DRC12 e DRC13), assim como duas opções para o cenário C-RAN (Figura 10 - DRC17 e DRC18). Essas divisões se alinham com ITU-T (principalmente devido a restrições de *crosshaul*) e com as iniciativas O-RAN e SCF (GAVRILOVSKA et al., 2020; MAVENIR, 2019). Naturalmente, a arquitetura D-RAN tradicional também é suportada para fornecer cenários, em que o *crosshaul* é limitado (ITU-T, 2018) (Figura 10 - DRC19).

Em resumo, o vNG-RAN pode ser implementado como um serviço de rede virtual, ou seja, um grupo de VNFs com um determinado conjunto de características. Primeiro, o serviço consiste em processar uma pilha de protocolos por dispositivo de RF em NG-RAN. Este processamento implica uma ordenação do fluxo de comunicação entre as VNFs, ou seja, o encadeamento a ser respeitado. Os VNFs são instanciados nas unidades da RAN, que também são elementos virtuais que podem ser executados em diferentes dispositivos de computação no NG-RAN. A escolha de onde posicionar as unidades da RAN e suas VNFs afeta os recursos aplicados, incluindo computação e rede. Para cada topologia NG-RAN e conjunto de recursos, pode haver várias opções para o posicionamento de nós VNFs e RAN. Em geral, o objetivo é consumir o mínimo de recursos e agrupar o máximo dos VNFs relacionados ao mesmo protocolo ou camada. Entretanto, cada opção de posicionamento implica diferentes demandas de computação e de rede que não devem exceder os recursos disponíveis. Dessa forma, o posicionamento da função torna-se um problema de otimização complexo. Nesse sentido, a resolução e a solução ótima dependem da escalabilidade das entradas do problema. Sendo assim, a modelagem matemática do mesmo é apresentada formalmente a seguir.

4.3.2 Modelagem do Sistema

Nesta seção, é introduzido o modelo do sistema representado e sintetizado com as notações do problema pela Tabela 5, dividindo a mesma em parâmetros e variáveis do problema. A tabela é detalhada conforme a apresentação do modelo e a formulação do problema. Sendo assim, a topologia de rede é definida como um grafo $G = (\mathcal{V}, \mathcal{E})$, com $\mathcal{V} = \{v_0\} \cup \mathcal{B} \cup \mathcal{C} \cup \mathcal{T}$ para os conjuntos de nós/unidades de rede e $\mathcal{E} = \{e_{ij}; v_i, v_j \in \mathcal{V}\}$, $v_i, v_j \in \mathcal{V}$ representando o conjunto dos *links* (arestas) que conectam os nós/unidades. v_0 representa o CN e é convencionado como a origem (*downlink*) de todo o fluxo do tráfego (CN também poderia ser convencionado como o destino do tráfego, ou seja, *uplink*). Cada *link* $e_{ij} \in \mathcal{E}$ possui uma capacidade de taxa de dados e_{ij}^{Cap} (dada em Gbps) e uma latência e_{ij}^{Lat} (dada em milissegundos). Os nós de rede da topologia são divididos em três conjuntos distintos. O primeiro é o conjunto $\mathcal{B} = \{b_1, b_2, \dots, b_{|\mathcal{B}|}\}$ que representa

entrada das unidades RUs, composta pelos protocolos Low PHY e RF. O segundo é o conjunto $\mathcal{C} = \{c_1, c_2, \dots, c_{|\mathcal{C}|}\}$ de RCs que processam as VNFs conectadas às RUs do conjunto \mathcal{B} . Cada RC c_m possui uma capacidade de processamento c_m^{Proc} com unidade em *cores* de CPUs (m é o identificador/índice de cada RC). Ademais, cada RC possui capacidades de memória e de armazenamento, porém no que se refere às características da desagregação da RAN, os mesmos não possuem profundas avaliações e não são explorados. Por fim, o terceiro conjunto de nós é o $\mathcal{T} = \{t_1, t_2, \dots, t_{|\mathcal{T}|}\}$ referente aos nós de transporte *crosshaul*, no quais são conectados as unidades de rádio virtuais (vCU, vDU e vRU), os RCs e o CN.

Tabela 5: Resumo das notações da formulação do problema

Parâmetro	Definição
$\mathcal{C} = \{c_1, c_2, \dots, c_{ \mathcal{C} }\}$	Conjunto de RCs.
c_m^{Proc}	Representa o recurso computacional do RC $c_m \in \mathcal{C}$. Sendo m o identificador/índice do RC e <i>proc</i> a capacidade de CPU do RC.
v_0	Representa o CN, origem do fluxo do tráfego.
$\mathcal{B} = \{b_1, b_2, \dots, b_{ \mathcal{B} }\}$	Conjunto de BSs, definido pela entrada de unidades RUs.
$\mathcal{F} = \{f_1, f_2, \dots, f_8\}$	Conjunto de funções virtualizadas baseado nos protocolos da RAN.
$\mathcal{T} = \{t_1, t_2, \dots, t_{ \mathcal{T} }\}$	Conjunto de nós de transporte.
$\mathcal{D} = \{D_1, D_2, \dots, D_{ \mathcal{D} }\}$	é o conjunto de DRC's.
$\mathcal{P}_l = \{P_{b_1}, P_{b_2}, \dots, P_{b_{ \mathcal{B} }}\}$	é o conjunto de caminhos que tem como origem o Core e destino $b \in \mathcal{B}$. Sendo l o identificador/índice da BS.
p_{BH}	Sub-caminho da comunicação CN-CU ou das redes <i>backhaul</i> .
p_{MH}	Sub-caminho da comunicação CU-DU ou das redes <i>midhaul</i> .
p_{FH}	Sub-caminho da comunicação DU-RU ou das redes <i>fronthaul</i> .
$y_{e_{ij}}^{p_{BH}}$	Link e_{ij} <i>backhaul</i> em um caminho $p \in \mathcal{P}_l$ de um DRC $D_r \in \mathcal{D}$.
$y_{e_{ij}}^{p_{MH}}$	Link e_{ij} <i>midhaul</i> em um caminho $p \in \mathcal{P}_l$ de um DRC $D_r \in \mathcal{D}$.
$y_{e_{ij}}^{p_{FH}}$	Link e_{ij} <i>fronthaul</i> em um caminho $p \in \mathcal{P}_l$ de um DRC $D_r \in \mathcal{D}$.
α_{BH}^r	Largura de banda entre CN-CU ou <i>backhaul</i> conforme DRC $D \in \mathcal{D}$.
α_{MH}^r	Largura de banda entre CU-DU ou <i>midhaul</i> conforme o DRC $D \in \mathcal{D}$.
α_{FH}^r	Largura de banda entre DU-RU ou <i>fronthaul</i> conforme o DRC $D \in \mathcal{D}$.
β_{BH}^r	Latência entre CN-CU ou <i>backhaul</i> conforme DRC $D \in \mathcal{D}$.
β_{MH}^r	Latência entre CU-DU ou <i>midhaul</i> conforme DRC $D \in \mathcal{D}$.
β_{FH}^r	Latência entre DU-RU ou <i>fronthaul</i> conforme DRC $D \in \mathcal{D}$.
γ_m^s	Processamento necessário para a VNF $f_s \in \mathcal{F}$ ser executada em $c \in \mathcal{C}$.
D_r^ω	Prioridade do DRC D_r . Sendo ω o peso da prioridade e r o requisito do DRC.
Variável	Definição
$u_m^p = \{0, 1\}$	Valor 1 se $c \in \mathcal{C}$ faz parte do caminho $p \in \mathcal{P}_b$.
$x_l^{p,r} = \{0, 1\}$	Valor 1 se o caminho $p \in \mathcal{P}_b$ e DRC $D \in \mathcal{D}$ é escolhido a BS $b \in \mathcal{B}$.
$M(c_m, f_s, b_l, D_r) = \{0, 1\}$	Valor 1 se o RC $c_m \in \mathcal{C}$ executa $f_s \in \mathcal{F}$ da BS $b_l \in \mathcal{B}$ de acordo com $D_r \in \mathcal{D}$.
†Cada caminho $p \in \mathcal{P}_l$ é formado por três sub caminhos p_1, p_2 e p_3 . Onde $p = p_1 + p_2 + p_3$.	

Outras duas definições para a construção do modelo são o encaminhamento/roteamento e as funções de rádio virtualizadas (ou as VNFs de rádio). Neste sentido, o encaminhamento/roteamento é todo o tráfego da rede. Embora o tráfego do CN possa ser de *downlink* e *uplink*, para o modelo, define-se que todo o tráfego tem o sentido do CN para o RU. O encaminhamento/roteamento é definido por \mathcal{P}_l (l é o identificador/índice de cada BS definida pela entrada de RUs) como um conjunto de k -caminhos mais curtos desde o CN até cada RU $b_l \in \mathcal{B}$. Cada caminho $p \in \mathcal{P}_l$ é composto por três sub-caminhos: p_{BH}

(Comunicação CN-CU ou *backhaul*), p_{MH} (Comunicação CU-DU ou *midhaul*) e p_{FH} (Comunicação DU-RU ou *fronthaul*), nos quais pelo menos um desses sub-caminhos não pode ser vazio. Por exemplo, o DRC2 possui os três caminhos p_{BH} , p_{MH} e p_{FH} , pois têm duas desagregações somadas a comunicação do BS com o CN. Por outro lado, o DRC19 (sem desagregações) possui apenas o caminho p_{BH} , referente à comunicação entre a BS e o CN. Vale ressaltar que os caminhos independem da posição das unidades de rádios. Com isso, por exemplo, um caminho p_{BH} DRC19 pode ser um caminho p_{MH} para um DRC2.

Para as funções de rádio virtualizadas (conjunto de protocolos da RAN), a execução das VNFs deve seguir a pilha de protocolos (por exemplo, protocolo RRC comunica-se apenas com o protocolo PDCP, e o PDCP se comunica com os protocolos RRC e o RLC.) da RAN e a fronteira entre as VNFs obedece aos requisitos demandados pelas desagregações entre os protocolos da pilha, conforme a Tabela 1. Sendo assim, as VNFs são compostas em ordem crescente do protocolo PHY Low com f_1 até o protocolo RRC com f_8 (conforme elucidado, o protocolo RF ainda não possui virtualização viável), de acordo com Tabela 5. Sendo assim, cada VNF da RAN desagregada possui um conjunto de protocolos que formam as unidades de rádio, e as VNFs são orientadas pelo conjunto de nove DRCs escolhidos $\mathcal{D} = \{D_1, D_2, \dots, D_{|\mathcal{D}|}\}$, priorizados conforme descrito na Subseção 4.3.1 e ilustrados na Figura 10.

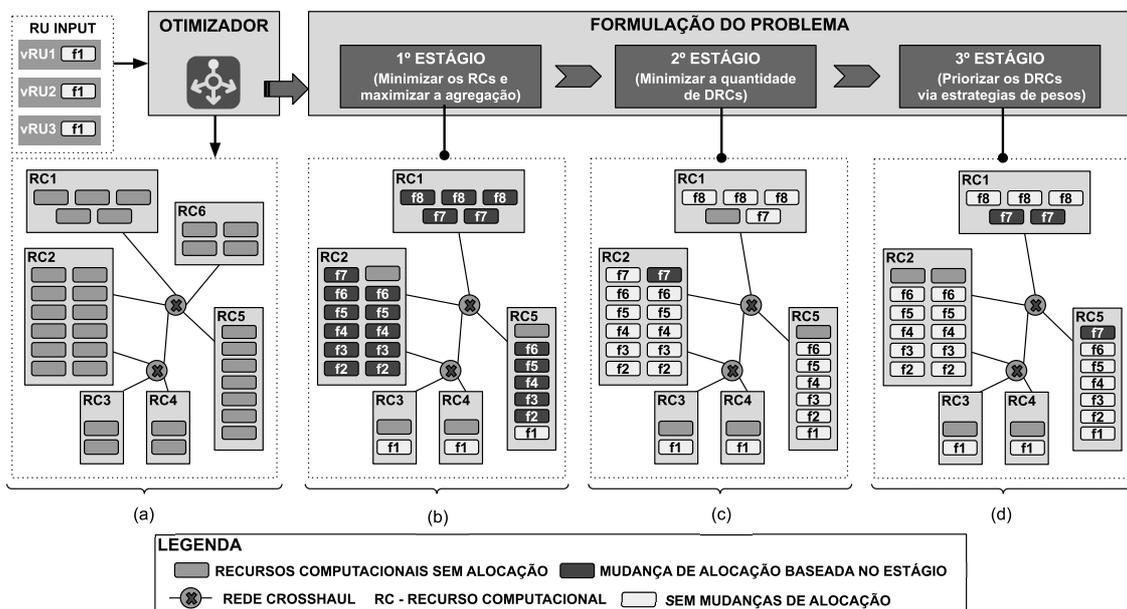
4.3.3 Formulação do Problema

A presente seção desenvolve a formulação do problema com dois objetivos: (i) maximizar o nível de agregação das VNFs da RAN e (ii) minimizar o número de RCs usados para agregação dos protocolos da RAN. Como as capacidades de computação e de rede são limitadas, diminuir o número de recursos de computação pode não implicar em um aumento no nível de agregação, o que cria objetivos conflitantes. Sendo o nível de agregação medido pela quantidade de VNFs $\mathcal{F} = \{f_2, \dots, f_8\}$ agrupadas em um mesmo recurso computacional. A medição é dada entre a relação das 6 VNFs $\mathcal{F} = \{f_2, \dots, f_8\}$ passíveis de agregação versus o número de entrada de RU. Por exemplo, uma entrada de três RUs ocasiona uma agregação máxima de 18. Porém, a agregação máxima é limitada pela capacidade das redes *crosshaul* e RCs. Por conta destas restrições, não há uma relação direta entre o número de RCs e o nível de agregação. Além disso, o nível de agregação das divisões funcionais não é medido apenas pelo número de VNFs e RCs, sendo também afetado por duas outras métricas (além da relação entre a minimização de recursos computacionais e a agregação das VNFs da RAN): o número de DRCs empregados e a prioridade ou preferência de uso de cada DRC. Como três métricas incompatíveis medem o nível de agregação, projetou-se a formulação em três estágios. A solução ótima pode ser eventualmente obtida no primeiro ou no segundo estágio, mas a validação da formulação

é considerada apenas com a conclusão do terceiro estágio, após resolver todos os empates potenciais.

Para elucidar os três estágios, a Figura 11 apresenta um exemplo de tomada de decisão por cada estágio e uma representação do desenvolvido na formulação. A Figura 11 é dividida em quatro visualizações. A Figura 11(a) exibe as entradas do sistema, com o *input* de três vRUs base formadas apenas pela VNF f_1 , a infraestrutura computacional, e a topologia e capacidade das redes *crosshaul*. As demais visualizações são exploradas conforme cada estágio. A legenda apresenta, além da rede *crosshaul* e dos RCs, as mudanças de estado dos recursos computacionais, sendo estes recursos sem alocação e, conseqüentemente, disponíveis, alocados conforme decisão do estágios e sem alteração com a evolução entre os estágios. A seguir, cada estágio é discutido.

Figura 11: Apresenta exemplo do funcionamento dos estágios desenvolvidos para a formulação do problema. A Figura 11a representa a entrada de RUs no Otimizador e a infraestrutura computacional conectada pelas redes *crosshaul*. A Figura 11b apresenta solução do estágio 1 com a resultante de 01 DRC1, 01 DRC2 e 01 DRC13. A Figura 11c mostra a resultante do estágio 2 com a redução na quantidade de DRCs para 2 (02 DRC1 e 01 DRC13). Por fim, a Figura 11d exibe a resultante final, com a priorização por pesos dos DRCs (02 DRC2 e 01 DRC12).



Fonte: Elaborada pelo autor

Estágio 1

No primeiro estágio, o objetivo é unificar a maximização do número de VNFs de unidades da RAN agrupados nos mesmos RCs e a minimização do número de RCs usados para executar essas VNFs (agregação concentrada nas VNFs $\mathcal{F} = \{f_2, \dots, f_8\}$). Desta forma, definiu-se $x_l^{p,r} \in \{0, 1\}$ como a variável de decisão que representa qual caminho $p \in \mathcal{P}_l$ e DRC $D_r \in \mathcal{D}$ (r são os requisitos de latência e taxa de dados de cada DRC) é selecionado para servir o RU $b_l \in \mathcal{B}$. A partir dos dados de entrada, determinou-se $u_m^p \in \{0, 1\}$ para indicar se $c_m \in \mathcal{C}$ faz parte de $p \in \mathcal{P}_l$. Adicionalmente, definiu-se a função de mapeamento $M(c_m, f_s, b_l, D_r) \in \{0, 1\}$ sobre os dados de entrada, o que indica se o RC $c_m \in \mathcal{C}$ executa a VNF $f_s \in \mathcal{F}$ do RU $b_l \in \mathcal{B}$, de acordo com o DRC $D_r \in \mathcal{D}$. Portanto, definiu-se a seguinte função objetivo:

$$\text{minimize } \Phi_1 - \Phi_2, \quad (4.1)$$

onde Φ_1 representa a quantidade de RCs, dada por:

$$\Phi_1 = \sum_{c_m \in \mathcal{C}} \left[\frac{\sum_{b_l \in \mathcal{B}} \sum_{D_r \in \mathcal{D}} \sum_{p \in \mathcal{P}_l} (x_l^{p,r} \cdot u_m^p)}{|\mathcal{C}|} \right], \quad (4.2)$$

e Φ_2 representa a quantidade de VNFs RAN agrupados, dado por:

$$\Phi_2 = \sum_{c_m \in \mathcal{C}} \sum_{f_s \in \mathcal{F}} \left(\sum_{b_l \in \mathcal{B}} \sum_{D_r \in \mathcal{D}} \sum_{p \in \mathcal{P}_l} [x_l^{p,r} \cdot u_m^p \cdot M(c_m, f_s, b_l, D_r)] - \left[\frac{\sum_{D_r \in \mathcal{D}} \sum_{p \in \mathcal{P}_l} \sum_{b_l \in \mathcal{B}} [x_l^{p,r} \cdot u_m^p \cdot M(c_m, f_s, b_l, D_r)]}{|\mathcal{F}|} \right] \right). \quad (4.3)$$

Para cada RU $b_l \in \mathcal{B}$, há exatamente um DRC $D_r \in \mathcal{D}$, usando um único caminho $p \in \mathcal{P}_l$, que deve ser selecionado, conforme representado pela seguinte restrição:

$$\sum_{D_r \in \mathcal{D}} \sum_{p \in \mathcal{P}_l} x_l^{p,r} = 1, \quad \forall b_l \in \mathcal{B}. \quad (4.4)$$

A capacidade de transmissão e_{ij}^{Cap} de cada *link* e_{ij} não deve ser excedida, conforme descrito pela seguinte restrição:

$$\sum_{b_l \in \mathcal{B}} \sum_{D_r \in \mathcal{D}} \sum_{p \in \mathcal{P}_l} \left[x_l^{p,r} \left(y_{e_{ij}}^{p_{BH}} \cdot \alpha_{BH}^r + y_{e_{ij}}^{p_{MH}} \cdot \alpha_{MH}^r + y_{e_{ij}}^{p_{FH}} \cdot \alpha_{FH}^r \right) \right] \leq e_{ij}^{Cap}, \forall e_{ij} \in \mathcal{E}, \quad (4.5)$$

onde $y_{e_{ij}}^{p_{BH}}$, $y_{e_{ij}}^{p_{MH}}$, e $y_{e_{ij}}^{p_{FH}}$ indica se o *link* e_{ij} faz parte do *backhaul*, *midhaul* ou *fronthaul*, em um caminho $p \in \mathcal{P}_l$ que transporta um DRC específico $D_r \in \mathcal{D}$. Cada $D_r \in \mathcal{D}$ tem taxa de dados associadas ao *backhaul*, *midhaul* e *fronthaul*, representado por α_{BH}^r , α_{MH}^r , e α_{FH}^r . Existem divisões funcionais nas quais o caminho $p \in \mathcal{P}_l$ tem menos de três sub-caminhos, como, por exemplo, DRC12, DRC13, DRC17, DRC18 e DRC19 (conforme os DRCs escolhidos para o desenvolvimento do trabalho e ilustrados na Figura 10).

Cada $D_r \in \mathcal{D}$ tolera uma latência máxima em cada sub-caminho (*backhaul*, *midhaul* e *fronthaul*) do caminho $p \in \mathcal{P}_l$, que é descrito pelas seguintes restrições:

$$\sum_{e_{ij} \in \mathcal{E}} x_l^{p,r} \cdot y_{e_{ij}}^{p_{BH}} \cdot e_{ij}^{Lat} \leq \beta_{BH}^r, \forall b_l \in \mathcal{B}, p \in \mathcal{P}_l, D_r \in \mathcal{D}, \quad (4.6)$$

$$\sum_{e_{ij} \in \mathcal{E}} x_l^{p,r} \cdot y_{e_{ij}}^{p_{MH}} \cdot e_{ij}^{Lat} \leq \beta_{MH}^r, \forall b_l \in \mathcal{B}, p \in \mathcal{P}_l, D_r \in \mathcal{D}, \quad (4.7)$$

$$\sum_{e_{ij} \in \mathcal{E}} x_l^{p,r} \cdot y_{e_{ij}}^{p_{FH}} \cdot e_{ij}^{Lat} \leq \beta_{FH}^r, \forall b_l \in \mathcal{B}, p \in \mathcal{P}_l, D_r \in \mathcal{D}, \quad (4.8)$$

onde β_{BH}^r , β_{MH}^r e β_{FH}^r representam a latência máxima tolerada no *backhaul*, *midhaul* e *fronthaul* de um caminho $p \in \mathcal{P}_l$ que transporta um DRC $D_r \in \mathcal{D}$ específico. Existem divisões funcionais nas quais o caminho $p \in \mathcal{P}_l$ tem menos de três sub-caminhos. Da mesma forma, se o sub-caminho estiver ausente, nenhum *link* faz parte dele.

Finalmente, os VNFs selecionados para operar em um RC $c_m \in \mathcal{C}$ não devem exceder sua capacidade de processamento c_m^{Proc} , conforme representado pela seguinte restrição:

$$\sum_{f_s \in \mathcal{F}} \sum_{b_l \in \mathcal{B}} \sum_{D_r \in \mathcal{D}} \sum_{p \in \mathcal{P}_l} x_l^{p,r} \cdot u_m^p \cdot M(c_m, f_s, b_l, D_r) \cdot \gamma_m^s \leq c_m^{Proc}, \quad \forall c_m \in \mathcal{C}, \quad (4.9)$$

onde γ_m^s é a demanda de computação das VNFs $f_s \in \mathcal{F}$. Conforme apresentado na Figura 11(b), o Estágio 1 aloca as VNFs, desde f_2 até f_8 , tendo, como resultante da função objetivo, a redução de seis RCs para cinco RCs e agregação máxima em RC1, que, contudo, acontece com a tomada de decisão pelos DRC1, DRC2 e DRC13.

Estágio 2

Depois de resolver o primeiro estágio, obtém-se o número mínimo de RCs necessários para atingir o nível máximo de agregação das VNFs da RAN. Como esses dois objetivos podem ser conflitantes, o resultado final do primeiro estágio é a melhor análise entre os dois objetivos. No entanto, o nível de agregação alcançado pode não ser o ideal, em virtude da possibilidade de encontra-se soluções distintas e com o mesmo valor final da função objetivo gerada a partir da equação 4.1.

Ao comparar-se diferentes soluções com o mesmo valor para a função objetivo (Equação (4.1)) no primeiro estágio, observa-se que aquelas com um número menor de DRCs escolhidos (a resultante do estágio 1 pode ser desde um DRC até nove DRCs, conforme a escolha adotada) têm um agrupamento maior das mesmas funções virtualizadas \mathcal{F} em um único RC. Sendo assim, tendem a melhorar o desempenho da RAN ao compartilhar o mesmo RC e trazer eficiência no uso dos recursos computacionais. Por exemplo, em um cenário de empate no resultado da função objetivo para duas análises, sendo um resultado com dois DRCs (por exemplo, DRC1 e DRC12) e outro com três DRCs (por exemplo, DRC1, DRC2 e DRC12), sem alterações no consumo dos RCs, a solução que possui apenas dois DRCs é melhor do que a de três DRCs. Logo, a função objetivo do segundo estágio é minimizar o número de DRCs:

$$\text{minimize} \quad \sum_{D_r \in \mathcal{D}} \left[\frac{\sum_{b_l \in \mathcal{B}} \sum_{p \in \mathcal{P}_l} x_l^{p,r}}{|\mathcal{B}|} \right]. \quad (4.10)$$

Este segundo estágio deve considerar apenas soluções com exatamente o mesmo valor da função objetivo alcançada pela solução ótima do primeiro estágio. A seguinte restrição garante esta situação:

$$\Phi_1 - \Phi_2 = f_{1st_stage}(x_l^{p,r*}), \quad (4.11)$$

onde $f_{1st_stage}(x_l^{p,r*})$ representa o valor da função objetivo do primeiro estágio quando a solução ótima é encontrada (* referencia a obrigatoriedade do atendimento dos requisitos do estágio anterior na análise da variável x). Além disso, todas as restrições do primeiro estágio também devem ser satisfeitas, ou seja, o segundo estágio está sujeito às restrições (4.4) - (4.9). A Figura 11(c) apresenta uma análise de um exemplo prático que minimiza para dois DRCs (DRC1 e DRC13) sem reduzir a função objetivo. A solução do estágio 2 difere do estágio 1, uma vez que o estágio 1 resulta em três DRCs distintos (DRC1, DRC2 e DRC13).

Estágio 3

Depois de resolver o segundo estágio, elimina-se as soluções potenciais que não atingem compatibilidade entre as mesmas funções virtualizadas \mathcal{F} que compartilham RCs comuns. Embora isso melhore o nível de agregação, ainda há a possibilidade de se obter soluções diferentes com o mesmo número de DRCs, mas não equivalentes. Sendo assim, cada DRC possui características que o tornam único, ou seja, é possível classificar os DRCs e a qualidade da solução. Ao classificar os DRCs, também é possível diferenciar as soluções novamente. Essa classificação pode ser extraída diretamente dos padrões, que já especificam a ordem de preferência dos DRCs. Naturalmente, diferentes instituições de padronização (por exemplo, O-RAN e ITU) podem atribuir outras prioridades aos DRCs, mas o modelo proposto é genérico e funciona corretamente com qualquer um deles. Na atribuição de prioridade para cada DRC, quanto menor o valor, maior a prioridade. Portanto, a função objetivo do terceiro estágio é minimizar a soma dos valores atribuídos aos DRCs:

$$\text{minimize} \quad \sum_{b_l \in \mathcal{B}} \sum_{D_r \in \mathcal{D}} \sum_{p \in \mathcal{P}_l} (x_l^{p,r} \cdot D_r^\omega), \quad (4.12)$$

onde D_r^ω representa a prioridade do DRC D_r (ω é o peso atribuído a cada DRC).

O terceiro estágio também deve considerar apenas soluções com exatamente o mesmo valor da função objetivo alcançada pela solução ótima do primeiro estágio. Isso significa atender a restrição descrita pela equação 4.11 e o valor exato de função objetivo dada pela solução ótima do segundo estágio. A seguinte restrição garante esta definição:

$$\sum_{D_r \in \mathcal{D}} \left[\frac{\sum_{b_l \in \mathcal{B}} \sum_{p \in \mathcal{P}_l} x_l^{p,r}}{|\mathcal{B}|} \right] = f_{2nd_stage}(x_l^{p,r*}), \quad (4.13)$$

onde $f_{2nd_stage}(x_l^{p,r*})$ representa o valor da função objetivo do segundo estágio quando a solução ótima é encontrada, da mesma forma que, no segundo estágio, todas as restrições do primeiro estágio também devem ser satisfeitas, ou seja, o terceiro estágio está sujeito às restrições (4.4) - (4.9). Com base na formulação do problema e para concluir o exemplo, a Figura 11(d) apresenta a tomada de decisão do posicionamento das VNFs fundamentada no Estágio 3. Sendo assim, é mantido o número total de DRCs, mas eles são alterados para os DRC2 e DRC12, ao contrário do estágio 2 que possuía os DRC1 e DRC13. A escolha realizada é guiada pela resultante da função objetivo do Estágio 3, em que, no exemplo, a soma dos pesos dos DRC2 (peso 1) e DRC12 (peso 10) é menor do que as demais possibilidades (DRC1 tem peso 4 e DRC13 tem peso 9). A solução com 2 DRC2 e 1 DRC12 totaliza peso total 12, enquanto a solução de 2 DRC1 e

1 DRC13 totaliza 17. Por fim, para analisar o orquestrador e o Otimizador, o próximo Capítulo apresenta a metodologia aplicada para avaliação da solução PlaceRAN.

5 METODOLOGIA DE AVALIAÇÃO

Este capítulo detalha a metodologia de avaliação da solução PlaceRAN. Tendo em vista uma solução que tem como objetivo um otimizador de posicionamento e um orquestrador, a avaliação está dividida em duas provas de conceito distintas. A primeira avaliação é realizada de forma conceitual, conforme a seção 5.1, e analisa a solução do bloco Otimizador. A segunda parte da avaliação, seção 5.2, por sua vez, trata-se de uma verificação do orquestrador no formato de um experimento para a avaliação fim-a-fim da solução.

5.1 Metodologia de Avaliação Conceitual

Esta seção fornece uma descrição geral dos cenários avaliados, o detalhamento das topologias reais aplicadas, dos tipos de recursos, das demandas e dos parâmetros utilizados para avaliar os resultados da solução. O modelo de avaliação utilizado não possui padronização, porém segue diretrizes semelhantes e complementares aos trabalhos de Garcia-Saavedra et al. (2018) e Murti et al. (2020). A Tabela 6 resume os três cenários desenvolvidos para a avaliação do Otimizador. Os três tipos de cenários são divididos em Capacidade Baixa (CB), Capacidade Média (CM) e Capacidade Alta (CA). Cada cenário possui quatro nós de interconexão: nó de agregação 1 (AG1), nó de agregação 2 (AG2), nó de acesso 1 (AC1) e nó de acesso 2 (AC2). Embora os nós AG1, AG2, AC1 e AC2 sejam definidos pela topologia das redes de *crosshaul*, os mesmos também representam os nós RCs devido à igual geolocalização, conforme decisão de projeto da Seção 4.1. A nomenclatura dos nós é baseada na proximidade com o CN (PASSION Project, 2020). Sendo assim, para cada nó, existem quatro características/ métricas de redes reais: (i) número de recursos de computação, (ii) taxa de dados, (iii) latência e (iv) número de unidades de RU. Esta avaliação está focada em comparar os cenários e dois tipos diferentes de topologia, um representando as redes atuais e o outro ilustrando uma tendência de futuros projetos de rede. A seguir, apresentam-se essas topologias e as características consideradas na avaliação.

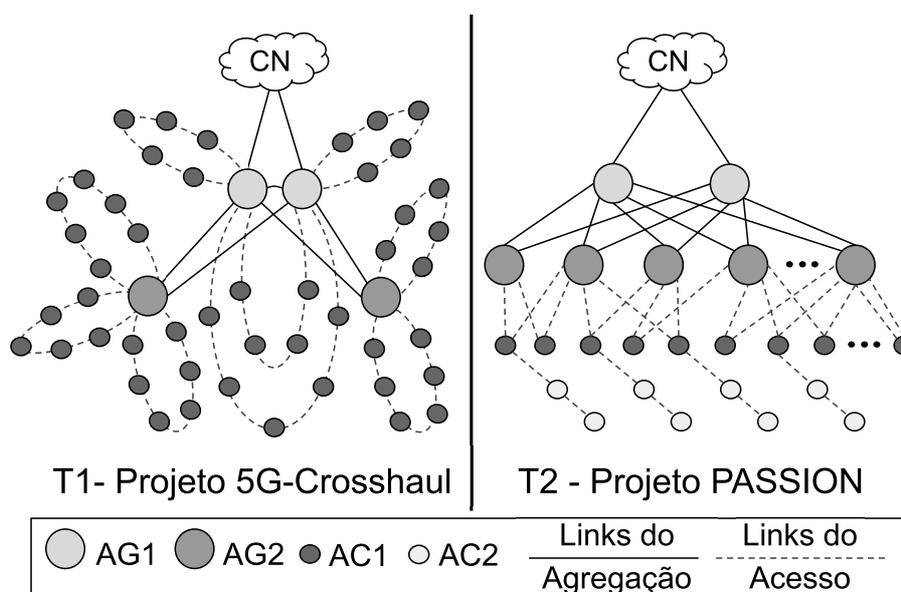
As duas topologias utilizadas para o desenvolvimento do trabalho foram baseadas nos projetos Europeus 5G-crosshaul (5G-crosshaul Project, 2017) (T1 - topologia 1) e PASSION (T2 - topologia 2) (PASSION Project, 2020). O projeto 5G-Crosshaul fornece uma rede operacional atual (foco no *backhaul* do LTE) com 51 nós no formato de anel, sendo formada por um anel de agregação e os demais anéis de acesso. O projeto PASSION define uma tendência de projeto de redes futuras com uma estrutura de árvore hierárquica, apresentando duas etapas de agregação e outras etapas de acesso. Para definir a quantidade

Tabela 6: Cenários para a avaliação da solução de otimização de posicionamento.

Cenários	Capacidade Baixa (CB)				Capacidade Média (CM)				Capacidade Alta (CA)				Topologia
	AG1	AG2	AC1	AC2	AG1	AG2	AC1	AC2	AG1	AG2	AC1	AC2	T1/T2
Nós de Transporte													
Recursos Computacionais (CPU core)	16	16	8	8	16	16	8	8	32	32	16	16	T1
	16	16	8	8	64	64	32	32	64	64	32	32	T2
Capacidade Links (Gbps)	100	40	25	10	100	40	25	10	400	100	40	25	T1
	800	100	50	40	1000	400	100	50	1000	400	100	50	T2
Latência Computacional (ms)	0,002				0,05-0,002				0,05				T1/T2
Latência de Transmissão (ms)	0,000005				0,000005				0,000005				T1/T2
Latência Equip. Óticos (ms)	0,010625				0,010625				0,010625				T2
Latência de Regeneração (ms)	0,0005				0,0005				0,0005				T2
Unidades RU	-				-				-				T1/T2
	F1 e R1				F1 e R1				F1 e R1				

de nós da avaliação conceitual, foi utilizada a premissa de que uma rede (definida como *cluster* pelo projeto) deve ter, no máximo, 20 nós AG2. Com isso realizou-se a razão entre o total de nós AG2 e nós AC1/2 (com base na premissa anterior) e limitou-se a 128 nós (total de todos os tipos de nós) para a avaliação da topologia do projeto PASSION. A Figura 12 ilustra as duas topologias, T1 e T2, empregadas nesta avaliação. Em ambas as topologias, os nós de transporte são classificados em quatro tipos (AG1, AG2, AC1 e AC2, porém AC2 encontra-se apenas na topologia T1), conforme ilustrado pela Figura 12. Esta classificação ajuda a ter flexibilidade na escolha dos parâmetros, ou seja, apresenta os valores das características para cada topologia e cenário.

Figura 12: Topologias para avaliação conceitual do Otimizador de Posicionamento.



Fonte: Elaborada pelo autor

No que se refere a recursos computacionais, este trabalho concentra-se na capacidade

de processamento (CPU), porque esse tem sido o gargalo mais comum para os dispositivos de computação no contexto de planejamento de rede e otimização de vRAN (HU; WANG, 2019; LIN; HUANG, 2018). Os valores de RC na Tabela 6 indicam o número de *cores* de CPUs ou um intervalo, como no caso do cenário CM. Neste cenário, o valor especificado é selecionado aleatoriamente (dentro do intervalo) na geração dos dados de entrada. A capacidade dos RCs foi projetada levando em consideração o perfil de utilização da CPU do *software* emulador da RAN OAI em testes de desempenho (KAZUNARI, 2018). Por outro lado, os valores depurados do consumo de processamento por protocolos da pilha da RAN são conforme a Tabela 7. Os valores exatos podem variar de acordo com os componentes de *software* e dispositivos de computação adotados, mas perfis semelhantes foram relatados em diferentes trabalhos (GARCIA-SAAVEDRA et al., 2018; HU; WANG, 2019; YEOH et al., 2016).

Tabela 7: Utilização de CPU por core da pilha de protocolos da RAN (HU; WANG, 2019; KAZUNARI, 2018; YEOH et al., 2016).

RAN Protocol	CPU Utilization (cores)
RRC	0,49
PDCP	0,49
High RLC	0,0245
Low RLC	0,0245
High MAC	0,343
Low MAC	0,343
High PHY	0,833
Low PHY	2,352
Total	4,9

A capacidade dos *links* considerados para avaliação das topologias T1 e T2 são divididas em taxa de dados e latência. Inicialmente, a taxa de dados empregada na avaliação está alinhada com as estratégias das topologias. Sendo assim, definiu-se a taxa de dados seguindo os padrões fornecidos pela IEEE Alliance (D'AMBROSIA, 2017). Nesse sentido, distinguiu-se as capacidades de interface entre as duas topologias de rede. A topologia T1 (rede atual) usa *links* de capacidade entre 40Gbps e 400Gbps nos nós de agregação (AG1 e AG2) e entre 10Gbps e 40Gbps nos nós de acesso (AC1 e AC2). A topologia T2 (rede futura) opera entre 100Gbps e 1Tbps nos nós de agregação e entre 40Gbps e 100Gbps nos nós de acesso.

Para alcançar as entradas da métrica de latência, empregou-se duas estratégias para gerar os dados de representação das topologias adequadamente e quatro componentes causadores de atraso e, conseqüentemente, latência na rede: (i) Computacional, (ii) Transmissão, (iii) Equipamentos Ópticos e (iv) Regeneração. A primeira estratégia, para a topologia T1, é usar apenas os componentes latência computacional e transmissão, enquanto

a segunda estratégia, para a topologia T2, utilizou-se de, todos os componentes. Ambas as estratégias são alinhadas com as definições dos seus projetos. Nesse sentido, Latência Computacional refere-se ao atraso consumido no processo de encaminhamento gerado por dispositivos comutadores e roteadores. A Latência de Transmissão está relacionada ao atraso de propagação nas fibras ópticas, sendo a mesma diretamente proporcional à distância (LANDON, 2018). As latências dos equipamentos ópticos e de regeneração são parâmetros baseados em sistemas de multiplexação ótica e são utilizados apenas na topologia T2. A latência dos equipamentos ópticos é o atraso causado pelos componentes dos dispositivos óticos, e estes não possuem processamento eletrônico, enquanto a Latência de Regeneração é responsável por transformar os sinais óticos em sinais elétricos (HUAWEI, 2018).

Com base nos componentes aplicados para latência, obteve-se os valores por *link*. Para estes, desenvolveu-se dois métodos diferentes, pois os dados, para cada topologia usada, são expressos de forma distintas. Na topologia T1, todas as Distâncias dos *Links* (entre os 51 nós) estão disponíveis nas informações do projeto (5G-crosshaul Project, 2017). Assim, a latência é expressa pela relação direta entre a distância (em Km) e a latência de transmissão, somadas, as duas, à latência computacional, para a composição da latência do *link*. Já na topologia T2, as informações de latência são derivadas do projeto PASSION com base em informações estatísticas sobre as topologias atuais, conforme é apresentado na Tabela 8. Sendo assim, a mesma fornece as informações da distância e quantidade de nós para a comunicação direta entre os nós de diferentes e de mesmas hierarquias. Com base na comunicação entre os nós, a topologia T2 propõe a utilização de sistemas de multiplexação ópticos para a criação de *links* entre nós de diferentes hierarquias, conforme a aderência de T2 à transformação para atendimento aos requisitos das redes de quinta geração. Por exemplo, um nó AC1 possui uma distância média de 24,04Km e 8,81 outros nós AC1s para conseguir alcançar um nó AG2. Nesse sentido, o uso de sistemas de multiplexação óptica isolam o tráfego entre os AC1, criando uma conexão direta com AG2. Entretanto, inserem as latência dos dispositivos ópticos e regeneradores nos *links*. Por conta disso, as latências de dispositivos ópticos e os regeneradores são adicionadas aos *links* da topologia T2.

Tabela 8: Parâmetros de distância e quantidade de saltos da topologia T2

Links entre Nós	Distância Média (Km)	Qtde de Saltos Médios	Distância Máxima (Km)	Qtde de Saltos Máximo	Distância Mínima (Km)	Qtde de Saltos Mínimos
CN - AG1	13,10	2,68	36,07	7	1,61	2
AG1 - AG2	24,73	7,17	36,80	12,00	11,25	4,00
AG2 - AC1	24,04	8,81	64,70	24,00	2,33	2,00
AC1 - AC2	1,6	1,99	5,88	4	0	1

Outro importante componente da avaliação são as entradas das unidades RU (composta pelos protocolos Low PHY e RF). Para avaliar os cenários (e topologias) com diferentes demandas, considerou-se duas configurações para o número de unidades de RU conectados aos nós de transporte: F1 (Fixo 1) - exatamente um nó RU é conectado a cada nó de transporte e R1 (*Random 1*) - zero ou um (inserido 0 para todos os nós AG2 e para os nós AC1 que concentram nós AC2). O nó RU é conectado a cada nó RC, que está conectado a um nó de transporte. No entanto, nenhum nó RU está conectado a qualquer nó AG1 em qualquer cenário de topologia, conforme recomendado pelos projetos 5G-Crosshaul (5G-crosshaul Project, 2017) e PASSION (PASSION Project, 2020).

Para a avaliação, definiu-se a Tabela 9, que resume os parâmetros das divisões funcionais empregadas na avaliação. Esses parâmetros são introduzidos como restrições na formulação do problema. Sendo assim, a tabela é composta pela priorização de cada DRC (usada na terceira etapa do modelo PlaceRAN) seguindo as especificações da O-RAN Alliance (ALLIANCE, 2020a). De acordo com as opções de divisão, a latência máxima tolerada é definida para cada sub-caminho da RAN: *backhaul* (Core-CU), *midhaul* (CU-DU) e *fronthaul* (DU-RU) (THOR, 2018). Além disso, de acordo com as opções de divisão, a taxa de Dados mínima aceitável é definida para cada sub-caminho da RAN. Entretanto, adotou-se diferentes perfis de RF para as topologias T1 e T2, o que implicou em diferentes larguras de banda para cada rede, conforme mostrado na Tabela 9. Na topologia T1, assumiu-se que as unidades RU possuem as seguintes características: largura de banda do canal de 40 MHz, 32 portas de antena, 8 camadas MIMO, 216 blocos de recursos físicos (PRB - *Physical Resource Block*) e espaçamento de subportadora de 15 kHz por macro BS (THOR, 2018). Na topologia 2, utilizou-se as unidades RU com as seguintes características: largura de banda de 100 MHz, 32 portas de antena, 8 camadas MIMO, 132 PRBs e espaçamento de subportadora de 60 kHz por macro BS (THOR, 2018). Ambos os canais são desejados para a evolução da quinta geração das redes móveis.

Tabela 9: Parâmetros das opções de desagregação avaliadas.

N ^o	DRC Prioridade	Latência - sentido único (ms)			T1 - Taxa de Dados (Gbps)			T2 - Taxa de Dados (Gbps)		
		CN-CU	CU-DU	DU-RU	CN-CU	CU-DU	DU-RU	CN-CU	CU-DU	DU-RU
1	4	1,5~10	1,5~10	250	2,97	5,4	17,4	9,9	13,2	42,6
2	1	1,5~10	1,5~10	250	2,97	5,4	17,4	9,9	13,2	42,6
7	6	1,5~10	1,5~10	250	2,97	5,4	5,6	9,9	13,2	13,6
8	5	1,5~10	1,5~10	250	2,97	5,4	5,6	9,9	13,2	13,6
12	10	1,5~10	1,5~10	-	2,97	5,4	-	9,9	13,2	-
13	9	1,5~10	1,5~10	-	2,97	5,4	-	9,9	13,2	-
17	8	1,5~10	-	250	2,97	-	5,6	9,9	-	13,6
18	7	1,5~10	-	250	2,97	-	17,4	9,9	-	42,6
19	25	1,5~10	-	-	2,97	-	-	9,9	-	-

Finalmente, todos os experimentos são executados em uma VM com Ubuntu 18.04, 16 vCPUs, 1TB de RAM e 40GB de disco virtual. A VM está hospedada em um servidor

DELL PowerEdge M620 com dois Intel Xeon E5-2650@2 GHz. Para essa prototipação, utilizou-se Python 2.7.17 e docplex 2.4.61 para implementar o modelo PlaceRAN. Além disso, o *solver* utilizado foi o IBM CPLEX 12.8.0. O código-fonte e os dados de entrada usados na avaliação estão disponíveis publicamente em Github*. Conforme descrito na seção 4.3, o resultado da solução conceitual é consolidado apenas após a conclusão dos três estágios, porém, é possível extrair resultados de cada estágio.

5.2 Metodologia de Avaliação Experimental

A fim de integrar a solução PlaceRAN, que envolve orquestrador e otimização de posicionamento, permitindo emular a sua aplicação em uma rede, um protótipo foi desenvolvido visando à execução do sistema. Para isso, a presente seção é dividida em duas partes. A subseção 5.2.1 apresenta o desenvolvimento da arquitetura e as ferramentas escolhidas. Por fim, a subseção 5.2.2 detalha o estudo de caso aplicado.

5.2.1 Protótipo

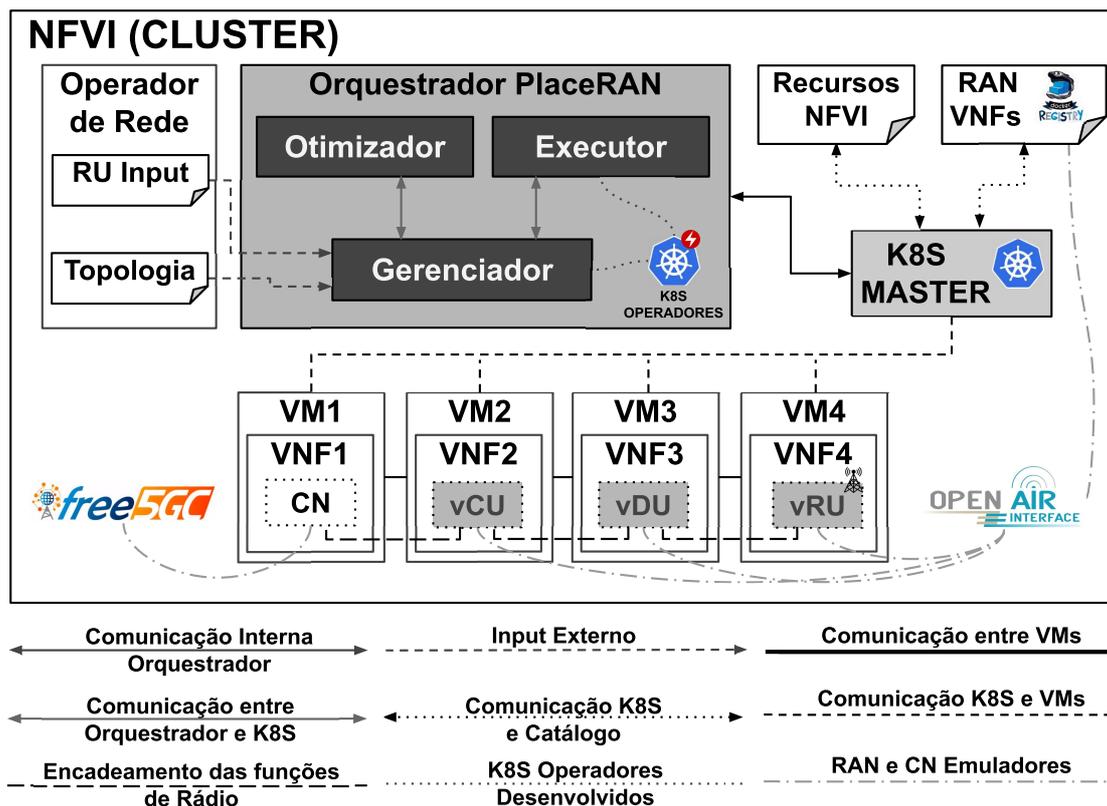
O protótipo da solução PlaceRAN foi construído com a ferramenta de orquestração e gerenciamento de infraestruturas computacionais baseadas em contêineres K8S. A arquitetura NFV e os blocos do MANO VNFM e VIM não possuem customização, sendo proprietários do K8S. Entretanto, o bloco MANO NFVO é customizado via funcionalidade de operadores e com o desenvolvimento dos sub-blocos Gerenciador e Executor. O Catálogo é abordado de diferentes maneiras, tais como:

- O catálogo Topologia *Crosshaul* e a entrada das RUs são definidos como um *ConfigMap* (padrão de armazenamento de configurações estruturado no modelo chave e valor) representado por um arquivo Objeto de notação *JavaScript* (JSON - *JavaScript Object Notation*).
- Os dados do catálogo Recursos NFVI são coletados diretamente do K8S e também são representados por um arquivo JSON.
- O catálogo de VNFs é representado por um repositório de imagens containerizadas (imagem é um binário com todo o conteúdo necessário para execução de um contêiner, como bibliotecas, metadados e requisitos) das funções de rádio virtualizadas (vCU, vDU e vRU).

No protótipo, o catálogo de VNFs é desenvolvido na ferramenta Docker Registry. A Figura 13 apresenta a adequação da solução de referência do capítulo 4 para o protótipo.

*<https://github.com/Unisinos-PPGCA/PlaceRAN-Solution.git>

Figura 13: Protótipo desenvolvido para a solução de Orquestração PlaceRAN. Os sub-blocos Executor e Gerenciador são desenvolvidos sob a funcionalidade do K8S operadores. O Otimizador utiliza as mesmas plataformas e *software*, porém virtualizada em um contêiner. O sub-bloco Otimizador é desenvolvido sob um contêiner virtualizado. Catálogo RAN VNFs é suportado pela plataforma Docker Registry. Os Catálogo Topologia e Recursos NFVI, assim como a entrada RU, são concebidos como um *ConfigMap* sob um arquivo JSON. A RAN foi emulada pela ferramenta OAI e o CN pela ferramenta Free5GC. Toda a infraestrutura computacional NFVI é gerenciada pela plataforma K8S no formato nativo de *cluster*.



Fonte: Elaborada pelo autor

Para o bloco NFVI são utilizados os nós RCs no formato de VMs e não *hardwares* dedicados. A emulação da RAN (no bloco VNFs) e do UE é desenvolvida pela ferramenta OAI LTE conforme detalhado na Figura 13. Para alcançar os objetivos do presente trabalho, uma customização foi realizada, uma vez que, atualmente, o emulador não possui suporte a quebra em três unidades de rede (CU, DU e RU) e containerização nativa. A customização foi gerada com suporte às opções de desagregações O2 e O6, resultando em DRC8, DRC13, DRC17 e DRC19. Além disso, o UE é emulado e está alocado junto à unidade RU. A emulação do CN é realizada pela ferramenta Free5GC terceiro estágio. Nesse estágio, a mesma possui o desenvolvimento parcial do *release 15* do 3GPP e, para o protótipo, é o responsável por prover autenticação dos UEs e conexão com a Internet.

Figura 14: Interface de Posicionamento e Função para o novo posicionamento.

```

Interface Otimizador {
    Posicionar([]*encadeamento) (boolean, erro)
}
func NovoPosionamento(topologia, RUs, RCs,
    requisitos) Posicionamento {}

```

Por fim, a solução de posicionamento, o sub-bloco Otimizador, é integrada como um serviço virtualizado em um contêiner com as mesmas ferramentas do modelo conceitual. O mesmo é baseado na requisição de execução e entrega do posicionamento ao protótipo, conforme a descrição da Interface Otimizador, como pode ser observado na Figura 14. Para comunicação com o sub-bloco Gerenciador, foi desenvolvido um gestor de agendamento, que recebe as entradas do Gerenciador, monitora a finalização da otimização e retorna a solução para o Gerenciador. O Otimizador utiliza as mesmas diretrizes da avaliação executada na seção 5.1.

A gestão dos RCS para a orquestração dos contêineres é realizada pelo K8S no modo *cluster* (nativo da ferramenta). A ferramenta tem o controle total de um conjunto de RCs/VMs (FOUNDATION, 2020b; KUBERNETES.IO, 2020). Cada RC/VM possui todos os componentes necessários para alocar as VNFs da RAN, na menor unidade de implementação computacional do K8S, denominado de *Pod*. Para o protótipo, cada *Pod* gerencia um contêiner com as funções de rádio virtualizadas (vCU, vDU e vRU), embora um *Pod* possa gerenciar mais de um contêiner. Entretanto, para o cenário de mais de um contêiner, existe a limitação de compartilhamento de endereçamento IP e de portas do sistema operacional (FOUNDATION, 2020b; KUBERNETES.IO, 2020). No que se refere às imagens containerizadas da vNG-RAN, o Docker Registry é a ferramenta de armazenamento dos mesmos e cada nó do cluster deve ter uma comunicação individual com o Docker Registry. Sendo assim, o K8S registra Docker Registry e o utiliza para criar os contêineres de forma alinhada com os *Pods* (Docker Inc., 2020).

Para os blocos Gerenciador e Executor, duas extensões do K8S são realizadas via funcionalidade de operadores. Ambos os operadores, o Gerenciador e o Executor, são desenvolvidos sob a linguagem de programação Golang (DOBIES; WOOD, 2020). As funções de agendamento (*scheduling*) e de controle do K8S gerenciam os dois operadores. Sendo assim, a determinação da alocação dos *Pods* é realizada pelos operadores, sendo o primeiro o construtor das informações e o segundo, o executor. Entretanto, o agendador nativo do K8S avalia a demanda de recursos computacionais antes de executar a inicialização dos *Pods* (FOUNDATION, 2020b). Outra ação relevante realizada pelo Executor é a customização dos arquivos de configuração do OAI para a criação do encadeamento das funções de rádio virtualizadas. Esse encadeamento é executado para cada

Figura 15: Exemplo de objeto do Gerenciador.

```

apiVersion: oai.unisinos/v1beta
kind: Gerenciador
metadata:
  name: posicionamento
  namespace: oai
spec:
  coreIP: 10.235.10.4
  topologyConfig: "topology"
  rus:
  - ranDeployerName: "1"
    ruNode: "node6"
  - ranDeployerName: "2"
    ruNode: "node7"
  - ranDeployerName: "3"
    ruNode: "node8"

```

Figura 16: Exemplo de saída do Executor para o K8S

```

apiVersion: oai.unisinos/v1beta1
kind: Executor
metadata:
  name: ngran-1
  namespace: oai
spec:
  coreIP: 10.235.10.4
  cuNode: node3
  duNode: node4
  ruNode: node2

```

VNF da RAN, em que são configurados o endereçamento IP da própria VNF e das VNFs que a mesma necessita encadear. Por exemplo, o encadeamento dos arquivos de configuração do OAI utilizam as interfaces de comunicação F1 (desagregação O2 - CU-DU) e *network Functional Application Platform Interface* (nFAPI) (desagregação O6 - DU-RU). Um exemplo do objeto de saída para o envio dos dados do Gerenciador para o Otimizador pode ser observado na Figura 15. A saída do Executor também gera um objeto de saída. Nesta etapa, o envio de informações é para o agendador do K8S, conforme pode ser verificado no exemplo do objeto de saída da Figura 16.

As escolhas das ferramentas do protótipo são fundamentadas a partir de avaliação de desenvolvimentos de solução para as redes móveis de quinta geração. Primeiramente, devido à granularidade e à alta distribuição previstas para as redes vNG-RAN, a escolha do tipo de virtualização por contêineres é orientada pela capacidade da tecnologia

em definir melhores locais de execução, pelo descobrimento de serviços, pelo balanceamento de cargas e pela melhor utilização de recursos computacionais (GAVRILOVSKA et al., 2020). Nesse contexto, a função de orquestração emerge para operacionalizar e automatizar o tipo de virtualização. Desta forma, a ferramenta K8S foi escolhida, pois a mesma é uma das mais populares (ampla comunidade com contribuições de código aberto), customizáveis e utilizadas em ambientes produtivos de orquestração de contêineres (CASALICCHIO, 2019; Red Hat, 2020). Adicionalmente, o K8S é alinhado com o órgão padronizador Cloud Native Computing Foundation (CNCF) (Cloud Native Computing Foundation, 2020) e com as principais iniciativas de orquestração para a virtualização das redes móveis: OSM e ONAP. Os dois principais concorrentes do K8S são o Docker Swarm e o Apache Mesos, apresentando, o primeiro, aplicações de baixa escala e o segundo, uma proposta para produção de baixa customização e com forte cunho profissional (CASALICCHIO, 2019).

Outra decisão relevante e alinhada com o K8S é a utilização do padrão de operadores para os sub-blocos Gerenciador e Executor, pois o padrão trabalha como uma extensão do plano de controle e Interface de programação de aplicações (API - *Application Programming Interface*) do K8S (DOBIES; WOOD, 2020). Sendo assim, um operador é construído para automatizar todo o ciclo de vida do *software* que ele gerencia, fornecendo automação específica para uma aplicação, sem a necessidade de customizar a base do K8S. Nesse contexto, o padrão de operadores contribui para a implementação das vNG-RANs, visto que os operadores podem ser utilizado para tratar da inicialização e da configuração das funções de rádio virtualizadas, a partir de uma definição declarada pelo operador de rede, e gerenciar questões do ciclo de vida das VNFs (DOBIES; WOOD, 2020). No que se refere às escolhas dos emuladores da RAN e do CN, atualmente, as ferramentas OAI e Free5GC são as principais disponibilizadas por código aberto. Ademais, o OAI é a única ferramenta que opera com desagregação da RAN (ALLIANCE, 2020b) e o Free5GC é a primeira solução de código aberto do *Release 15* (FREE5GC, 2020).

5.2.2 Estudo de Caso

O estudo de caso do protótipo foi realizado seguindo as mesmas diretrizes das entradas da metodologia conceitual, porém com um cenário único, pois o objetivo do estudo de caso é avaliar o comportamento do orquestrador a partir da integração com a solução de otimização de posicionamento. Sendo assim, a metodologia de avaliação do protótipo é embasada na topologia de rede (*crosshaul* e nos nós computacionais), na capacidade dos RCs (métricas de CPU e RAM), na capacidade dos *links* (taxa de dados e latência), na entrada de unidades RUs e nos requisitos de desagregação por DRCs. Além disso, o estudo de caso foi baseado apenas na topologia T2, pois a topologia T1 possui tamanho

fixo e uma análise traz distorções nos resultados devido ao formato em anel da mesma. A topologia T2 foi limitada a 25 nós entre AG1, AG2 e AC1, devido à capacidade de RCs. Seguindo as orientações do projeto (PASSION Project, 2020), T2 tem dois nós AG1 e uma distribuição de quatro AG2s para agregar 19 AC1s, totalizando 25 nós.

No que se refere aos recursos computacionais, os RCs foram disponibilizados conforme os tipos de nós. AG1 possui capacidade de 16 *Cores* de CPU e 16GB de memória RAM, AG2 foi configurado com oito *Cores* de CPU e 8GB de memória e por fim, AC1 possui a capacidade de dois *Cores* de CPU e 4GB de memória RAM. As máquinas virtuais utilizadas para hospedar os nós do *cluster* K8S estão em um ambiente compartilhado composto por 8 servidores *blade* Dell PowerEdge M610, com dois processadores Inter(R) Xeon(R) CPU (Central Processing Unit) X5660 @ 2.80GHz 6 núcleos, 192 GB de memória RAM DDR3 com frequência de 1333 MHz e dois adaptadores de rede Broadcom de 10 Gb. Como plataforma de virtualização, foi utilizado o *hypervisor* VMware ESXI 6.7.

Diferentemente da avaliação conceitual, no protótipo, a capacidade de CPU e de RAM dos RCs é uma entrada do Gerenciador para o Otimizador, com base nos dados do próprio K8S. Sendo assim, o consumo considerado é 0,5 *cores* de CPU e 0,5GB para CU e para DU, 0,8 *cores* de CPU e 1GB e para RU. Os valores de CPU e de RAM foram determinados a partir da avaliação do emulador OAI e com reserva de recursos para o pico das métricas. Quando o Otimizador decidir que as unidades de rádio do mesmo encadeamento devam estar no mesmo RC, os valores de consumo devem ser somados.

Em se tratando da capacidade dos *links*, a latência empregada utilizou a mesma estratégia da topologia T2 da análise conceitual com base nos quatro parâmetros de latência da Tabela 6 e Tabela 8. A taxa de dados dos *links* é definida conforme os tipos dos nós e possui 10 Gbps entre CN-AG1 e AG1-AG2 e 1 Gbps entre AG2-AC1. Estes valores são apenas entradas para o Otimizador, pois não há uma rede *crosshaul* simulada ou emulada. Nesse sentido, a capacidade e o comportamento dos *links* não tiveram desempenho avaliado.

No que diz respeito à entrada de RUs, dois tipos foram avaliados e seguem a mesma estratégia de F1 e R1 da avaliação conceitual, sendo F1 para todos os nós AG1 e AC1 e R1 apenas para os nós AC1, seguindo a lógica de que um nó AC1, sem agregação de outros nós e sem RU, não possui necessidade de existir. No que se refere aos DRCs, foram avaliados os DRC8, DRC13, DRC17 e DRC19 devido às limitações da ferramenta de emulação OAI. No entanto, os DRCs foram utilizados com os requisitos das redes LTE, devido à aplicação do emulador OAI ser de quarta geração. A Tabela 10 detalha as informações, com as especificações da RU: largura de banda de 20 MHz, 1 antena, 100 PRBs e espaçamento de subportadora de 15 kHz por macro BS (FORUM, 2016).

Para o experimento, a avaliação dos resultados foi guiada de forma alinhada com

Tabela 10: Parâmetros das opções de desagregação avaliadas para a tecnologia LTE (FORUM, 2016).

N ^o	DRC		Latência - sentido único (ms)			Largura de Banda (Gbps)		
	Prioridade		CN-CU	CU-DU	DU-RU	CN-CU	CU-DU	DU-RU
8	5		30	30	2	151	151	152
13	9		30	30	-	151	151	-
17	8		30	-	2	151	-	152
19	25		30	-	-	151	-	-

as métricas e análises da arquitetura NFV (ETSI, 2014b,a) e da solução OSM (OSM, 2017). Nesse sentido, duas métricas de qualidade foram analisadas: *(i)* a funcionalidade do experimento e *ii* o comportamento computacional do experimento. Para a avaliação do comportamento dos recursos computacionais, foram realizados dez testes com tempo determinado a partir da estabilização do sistema. Para a funcionalidade do experimento, foram analisadas a veracidade entre a saída do sub-bloco Otimizador, e a implementações dos encadeamentos das funções de rádio enviadas do sub-bloco Executor para a alocação pelo K8s. Todo o detalhamento do estudo de caso está disponível no Github* . O próximo capítulo discute os resultados com base na metodologia apresentada no presente capítulo.

*<https://github.com/Unisinos-PPGCA/PlaceRAN-Solution.git>

6 RESULTADOS

Neste capítulo, são apresentados os resultados da solução PlaceRAN. O capítulo é dividido em duas seções. A seção 6.1 apresenta a avaliação conceitual da modelagem matemática e a seção 6.2 exibe a avaliação experimental baseada em um protótipo e em um estudo de caso.

6.1 Resultados da análise conceitual do Otimizador de posicionamento

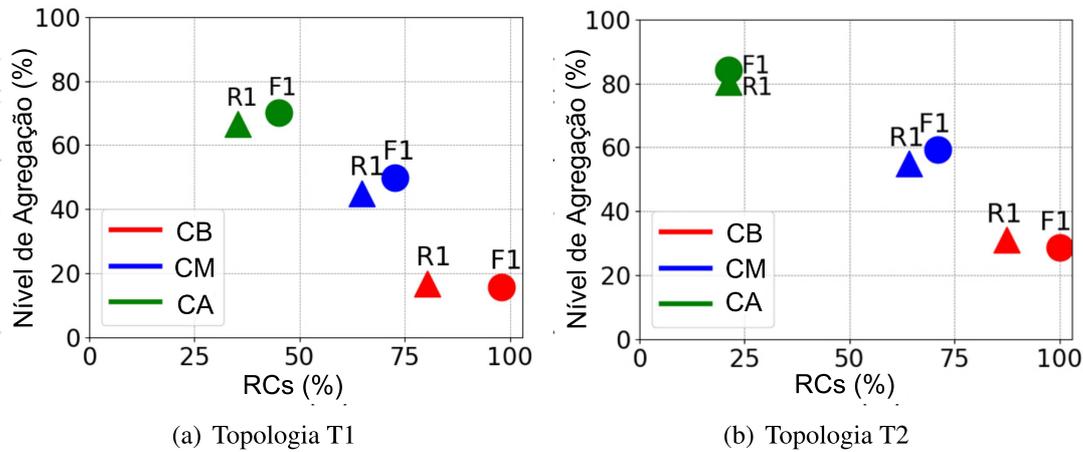
A avaliação de desempenho do modelo PlaceRAN foi organizada em três análises distintas e um fechamento com considerações finais. A primeira avaliação examina a relação entre a concentração das VNFs da RAN para os RCs e o nível de agregação resultante. Os três cenários, ou seja, CB, CM e CH e as duas topologias são avaliadas. A segunda investigação explora as soluções decorrentes da aplicação DRCs com os mesmos cenários. A terceira análise apresenta os impactos das restrições da topologia da rede, especificamente para a topologia T1, devido à complexidade combinatória da topologia. Por fim, são discutidas as associações entre os três resultados com uma visão global do modelo PlaceRAN.

6.1.1 Recursos Computacionais versus Nível de Agregação

Esta análise apresenta os resultados da pesquisa com o relacionamento de dois objetivos principais, a minimização dos RCs e a maximização do nível de agregação dos RAN VNFs. A Figura 17 expõe as duas topologias, T1 e T2, avaliando a relação dos três cenários estipulados (CB, CM e CA) e duas entradas de unidades vRU (F1 e R1). Para as unidades RUs, apenas as funções virtualizadas $\mathcal{F} = \{f_2, f_3, f_4, f_5, f_6, f_7, f_8\}$ são avaliadas uma vez que a função $\mathcal{F} = \{f_1\}$ (protocolo *Low PHY*) é fixa. Portanto, f_1 não foi computado na ocupação dos RCs. Consequentemente, o nível de agregação é produzido a partir da relação entre a quantidade total de vRU versus as seis funções virtualizadas detalhadas anteriormente. Por exemplo, a topologia T1 entrada F1 possui 49 vRUs e um nível de agregação total de 294 (multiplicação de 49 vRUs pelas 6 funções virtualizadas, em que 294 é 100% do nível de agregação). Vale ressaltar que o nível de agregação apenas é atingido caso todas as 49 vRUs sejam concentradas em um mesmo nó.

Primeiramente, a Figura 17 apresenta altos índices de agregação e baixo uso de RCs para o cenário CA. O nível de agregação na topologia T1 é superior a 60%, consumindo entre 30% e 40% da quantidade de nós RCs disponíveis, conforme mostrado na Figura 17(a). Para a topologia T2, o nível de agregação apresentou maior eficiência, próximo a

Figura 17: Análise da relação entre quantidade de RCs utilizados e o nível de agregação das funções de rádio virtualizadas.



80%, consumindo em torno de 20% dos RCs, conforme observado na Figura 17(b). Entretanto, o cenário CB apresentou queda no nível de agregação (para ambas a topologias) e, conseqüentemente, o aumento no uso da quantidade de CRs. O nível de agregação na topologia T1 é inferior a 20%, com 100% de uso de CRs para F1. Na topologia T2, o nível de agregação é inferior a 40% com o uso de CRs semelhantes. Comparando os três cenários (CB, CM e CA) e diferentes entradas de vRUs (R1 e F1), o nível de agregação mostra que a topologia T2 atinge sempre níveis superiores de agregação quando comparados os mesmos cenários entre as topologias. De forma geral, os cenários CB, CM e CA, de cada topologia, exibem resultados semelhantes entre as entradas R1 e F1, com exceção para o uso dos RCs do cenário CB, que demonstraram disparidade de utilização.

Fundamentado nos resultados da Figura 17, que representa os níveis totais de agregação e de usabilidade dos RCs, para um estudo mais detalhado, desenvolveu-se a Tabela 11 e a Tabela 12, considerando os resultados do nível de agregação dividido pelos nós de transporte (AG1, AG2, AC1 e AC2) e a quantidade de RCs. Entretanto, conforme apresentado na seção 5.1 a topologia T1 não possui o nó AC2, e para T2, agrupou-se os nós AC1 e AC2 nos mesmos dados e os denominou de AC. Os níveis de agregação e o uso de RCs totalizam 100% dos percentuais encontrados na figura 17. De forma a representar referência para as duas Tabelas, a resultante ideal para o sistema é a agregação de 100% das funções virtualizadas em AG1 e o uso de apenas um RC para agregação. Com base no detalhamento das tabelas, na Tabela 11, pode-se observar que AG1 concentra agregações das VNFs em 92,3% para o cenário CA com R1 e 83,7% para F1. Além disso, a entrada F1 utiliza apenas dois RCs dos nós AG2 e nenhum nó de AC1, diferentemente da entrada R1 que apresenta um baixo nível de agregação em AC1. Para o cenário CM, ainda em T1, as agregações foram inferiores a 60% em AG1 e o restante distribuído nos nós AG2

e AC1. Por fim, para T1, observou-se um alto nível de agregação em AC1, com 62,8% e 81,6% para ambas as entradas de vRUs no cenário CB. Já para os cenários CB e CM, observou-se um aumento do uso de RCs em nós AC1, o que resulta em baixa agregação de VNFs de RANs.

Tabela 11: Nível de agregação por nó na topologia T1.

		CB		CM		CA	
		R1	F1	R1	F1	R1	F1
AG1	*	26,7	18,4	58,9	53,1	92,3	83,7
	#	2	2	2	2	2	2
AG2	*	10,5	0	25,7	18,4	0	16,3
	#	1	0	1	2	0	2
AC1	*	62,8	81,6	15,4	28,5	7,7	0
	#	12	20	4	7	1	0

* Nível de Agregação (%); # Número de RCs

Para a topologia T2, a Tabela 12 apresenta os resultados encontrados com uma topologia projetada com aderência para as redes de quinta geração. Inicialmente, a análise da planilha traz o expressivo resultado de 0% de agregação para qualquer nó de acesso, seja este, AC1 ou AC2. Com base nesse dado, toda a agregação da topologia T2 é concentrada nos nós AG1 e AG2. Dentro desse contexto, o comportamento dos três cenários é diretamente orientado à capacidade de recursos de rede. Nesse sentido, o cenário CA atinge a máxima agregação no nó AG1 para a entrada R1 com 86,2% e o cenário CB, de forma oposta, alcança a máxima agregação no nó AG2 com a entrada R1 com 80,2%. O cenário CM apresenta equilíbrio na agregação entre os nós AG1 e AG2. A utilização de RCs atinge o melhor desempenho no cenário CA e entrada R1 com apenas dois RCs no nó AG2 e utiliza todos os nós AG2 da topologia no cenário CB com entrada F1. Para conclusão da análise da topologia T2, dois pontos são relevantes. Primeiramente, os resultados evidenciam que existe a necessidade de uma maior capacidade de recursos para alavancar a vNG-RAN, e, em segundo lugar, o comportamento das entradas R1 e F1 na topologia T2 apresenta melhor desempenho para R1, uma vez que existe menor demanda de capacidade de recursos devido a diferença de formato e às consequentes restrições do *crosshaul* para as topologias.

6.1.2 Definição dos DRCs

Esta seção analisa os DRCs escolhidos pelo modelo PlaceRAN. A Figura 18 mostra a quantidade e as opções de DRCs resultantes dos três cenários (CB, CM e CA) com as duas entradas de vRU (F1 e R1) para ambas as topologias investigadas. Para isso, agrupou-se os nove DRCs das cinco arquiteturas mostradas na Figura 10 em quatro grupos para uma

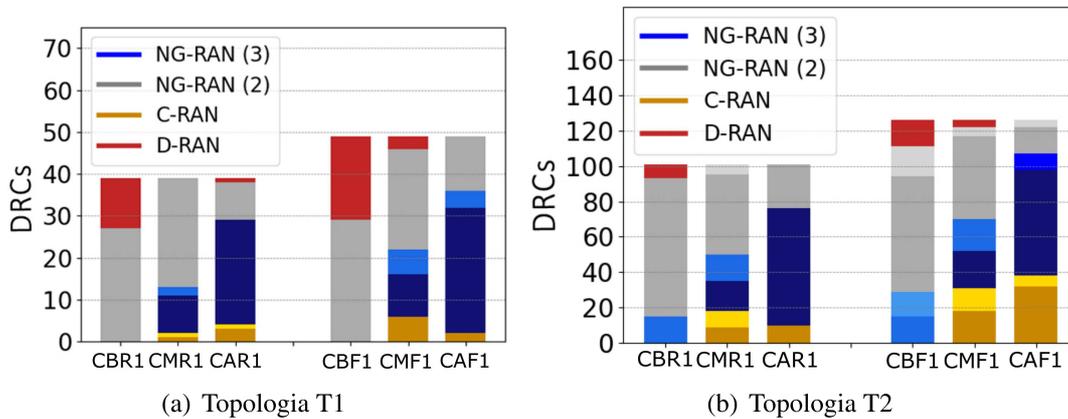
Tabela 12: Nível de agregação por nó na topologia T2.

		CB		CM		CA	
		R1	F1	R1	F1	R1	F1
AG1	*	19,8	27	52,5	46	86,2	77,8
	#	2	2	2	2	2	2
AG2	*	80,2	73	47,5	54	13,8	22,2
	#	18	20	7	11	2	3
AC	*	0	0	0	0	0	0
	#	0	0	0	0	0	0

* Nível de Agregação (%); # Número de RCs

análise precisa. Neste contexto, agrupou-se os quatro DRCs (DRC1, DRC2, DRC7 e DRC8) e os denominados NG-RAN (3) a partir de arquiteturas de 3 unidades independentes (O-RAN Low Split e SCF Low Split). Intitulou-se de NG-RAN (2) os DRC12 e DRC13 da arquitetura de integração RU e DU. Ainda, considerou-se DRC17 e DRC18 como a arquitetura C-RAN, e o DCR19 como a arquitetura D-RAN. Além disso, a tonalidade de suas respectivas cores está de acordo com a prioridade (representada pelos pesos no modelo) determinada para DRCs definidos na terceira etapa do modelo Place-RAN. Nesse caso, a escolha do melhor conjunto de DCRs está associada às restrições de latência e taxa de dados impostas pela rede e aos RCs disponíveis.

Figura 18: Escolhas dos DRCs do Otimizador para as duas topologias e os três cenários.



A primeira observação diz respeito ao número total de DRCs avaliados por cada cenário, dada a diferença de topologias e estratégias entre as entradas R1 e F1 de RU. Nesse sentido, a Figura 18(a) apresenta a topologia T1 com 39 DRCs para R1 e 49 DRCs para F1. Além disso, a Figura 18(b) mostra 101 DRCs para R1 e 126 DRCs para F1 na topologia T2. Ambas as figuras possuem as quantidades de nós e estratégias de entradas de RU descritas na seção 5.1. Observando o comportamento dos resultados dos DRCs, a Figura 18 mostra que a solução de otimização PlaceRAN aplica um número distinto de

tipos de DRCs para cada cenário. Por exemplo, o modelo definiu dois tipos de DRCs para o cenário CB com RUs F1 para a topologia T1 e cinco tipos de DRCs para o mesmo cenário na topologia T2. O aumento dos tipos de DRCs é esperado devido a dois fatores principais: (i) maximização da agregação e (ii) restrições de *crosshaul*. Para o primeiro, a prioridade do modelo é agregar o máximo de VNFs, seguindo o peso dos DRCs da Tabela 6, mesmo que, para isso, aumente o número de tipos de DRCs. Embora o estágio 2 tenha o objetivo de reduzir o número de DRCs, isso apenas ocorre a partir de diferentes resultantes da função objetivo. Sendo assim, o resultado do estágio 2 é determinado pelas opções encontradas no estágio 1. Para o segundo fator, a restrição *crosshaul* limita o número de tipos de DRCs escolhidos, devido à capacidade dos *link crosshaul* (taxa de dados e latência). Vale ressaltar que a capacidade dos RCs aplicada não apresentou restrição para nenhum cenário.

No que se refere à solução dos tipos de DRCs escolhidos, outra análise significativa é a diferença entre os cenários com as arquiteturas definidas. Por exemplo, o modelo PlaceRAN não definiu D-RAN para cenários CA em ambas as topologias, exceto na topologia T1 CAR1. Entretanto, D-RAN tem alta relevância na topologia T1 com o cenário CB, principalmente no cenário CA com entrada de RU F1, atingindo 40,8% (# 20). Este comportamento está diretamente relacionado à restrição da capacidade de *crosshaul* e é mais significativo na topologia T1, devido ao formato em anel. Nesse caso, a capacidade dos *links* AC1 é compartilhada com vários nós. Além disso, observando os dois DRCs da arquitetura C-RAN, pode-se ver que o modelo PlaceRAN determinou um número maior de DRCs para o cenário CA para a topologia T2, tanto para a entrada R1 quanto para a F1. Ainda, o modelo não atingiu nenhum DRC da arquitetura C-RAN no cenário CB, diferentemente do cenário CM e CA. Este comportamento ocorre porque a arquitetura C-RAN maximiza a centralização da função vCU e, neste caso, é necessária uma maior capacidade das redes *crosshaul*.

Os DRCs da arquitetura NG-RAN (2) exigem menos capacidade de rede. Portanto, esses DRCs são amplamente escolhidos pelos cenários CB e CM. Por exemplo, o modelo atingiu, no cenário CB, entrada F1, índices superiores a 70% (#28) na topologia T1 e superiores a 80% (#81) na topologia T2. O modelo PlaceRAN apoiado nos DRCs NG-RAN (2), são a melhor opção para capacidade *crosshaul* estrita, porque a divisão funcional tem uma alta tolerância de latência e baixa capacidade em comparação com uma divisão funcional de DU-RU apresentada na Tabela 6. Finalmente, a arquitetura NG-RAN (3), embora seja o tipo de DRC mais relevante do estágio 3 do modelo e siga as iniciativas O-RAN, não é totalmente viável por causa da capacidade das redes *crosshaul*. Entretanto, o modelo alcançou resultados expressivos no cenário CA, chegando a 64,1% (#25) com a entrada R1 e 67,34% (#33) com a entrada F1 na topologia T1. Além disso, o modelo atingiu 77,2% (#78) para os nós RU R1 e 54,8% (#69) F1 para a topologia T2. A apli-

cação da arquitetura NG-RAN (3) não possui viabilidade para o cenário CB da topologia T1 como em análises dos outros cenários, devido às restrições impostas pelo *crosshaul* estarem alinhadas à característica de alta distribuição em razão das três unidades de rádio.

Análise de Efetividade dos 3 Estágios

Inserido no contexto da tomada de decisão dos DRCs, os estágios 2 e 3 do modelo PlaceRAN precisam lidar com casos empatados e definir qual solução é a melhor. Nesse sentido, a Tabela 13 é um exemplo do comportamento da topologia T1 no cenário CM e entrada de RU R1, em que as soluções encontradas pela segunda e terceira etapas apresentam melhorias significativas para a solução final. Portanto, a Tabela 13 mostra a configuração dos DRCs das soluções encontradas pelos 3 estágios. No estágio 1, a resultante é a definição do nível máximo de agregação e o número mínimo de RCs. No estágio 2, é definido o menor número de DRCs que obtêm o nível de agregação alinhado com o alcançado no estágio 1. Observa-se que o número total de DRCs usados entre o estágio 1 e 2 apresentam uma redução de sete tipos de DRCs para cinco DRCs na Tabela 13. Por fim, no estágio 3, as escolhas dos DRCs são feitas de acordo com os pesos definidos na terceira etapa e representados na Tabela 6. Nesse sentido, observou-se que o montante de DRC19 (D-RAN) diminui significativamente, de 5,2% para 0%, na terceira etapa, visto que DRC19 tem o pior peso neste caso, assim como, o DRC2, de maior peso, obteve um incremento de 5,2% desde o estágio 1 até o 3. Portanto, a Tabela 13 permite evidenciar a importância dos três estágios desenvolvidos no modelo, bem como as melhorias obtidas por cada um.

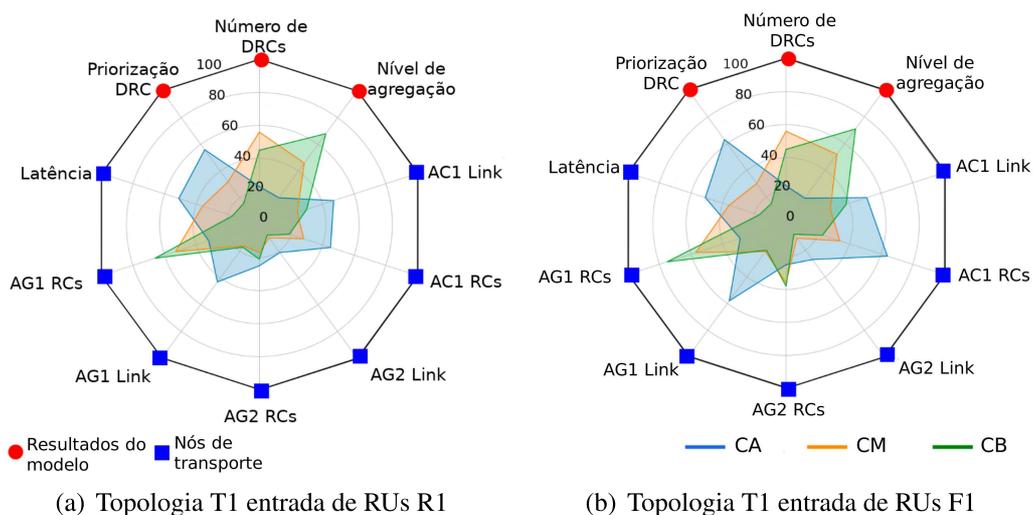
Tabela 13: Análise dos 3 estágios do Modelo do Otimizador da topologia T1 cenário CM.

DRCs	Estágio 1	Estágio 2	Estágio 3
1 - NG-RAN(3)	0%	0%	0%
2 - NG-RAN(3)	17,9%	20,5%	23,1%
7 - NG-RAN(3)	0%	0%	0%
8 - NG-RAN(3)	5,1%	10,2%	5,1%
12 - NG-RAN(2)	0%	0%	0%
13 - NG-RAN(2)	61,4%	64,1%	66,6%
17 - C-RAN	5,2%	0%	2,6%
18 - C-RAN	5,2%	2,6%	2,6%
19 - D-RAN	5,2%	2,6%	0%
Total	100%	100%	100%

6.1.3 Impacto dos Recursos de Rede

A avaliação em questão, investiga a correlação entre os três estágios do modelo do Otimizador e a resultante ocupação dos recursos da rede. Com base nas duas análises anteriores, observou-se que o *crosshaul* foi a principal restrição para o modelo. Nesse sentido, foi escolhida a topologia T1, especialmente para uma avaliação profunda deste comportamento e, assim, destacar-se as restrições desta topologia para o desenvolvimento de projetos de redes de quinta geração. Portanto, selecionou-se dez características para investigar essa correlação para os três cenários (CB, CM e CA), com as duas entradas de RUs, conforme mostrado na Figura 19. As características são divididas em dois tipos: uma característica associada ao modelo e a outra à capacidade da rede. As características baseadas no modelo são indicadas com um marcador vermelho (nível de agregação, uso dos DRCs e peso dos DRCs). A capacidade da rede está associada às características dos nós de transporte (*Link AC1*, *CRs AC1*, *Link AG2*, *CRs AG2*, *Link AG1*, *CRs AG1* e *Latência*), representados pela cor azul na Figura 19.

Figura 19: Avaliação dos estágios do modelo versus recursos de rede.



Observando-se as características do modelo PlaceRAN, os nós RU R1 na Figura 19(a) e F1 na Figura 19(b) mostram resultados semelhantes para os três cenários. O nível de agregação apresentou o mesmo comportamento nos cenários CA e CB, superior a 60% e inferior a 20%, respectivamente. O uso de DRCs mostrou o número de tipos de DRCs em unidade de porcentagem (nove DRCs é igual a 100%). O modelo atingiu a aplicação de quase 60% do total de tipos de DRCs no cenário CM e cerca de 20% no cenário CA. Nesse caso, é possível ver o nível de agregação apropriado para o cenário CA e o alto uso de DRCs para o cenário CM. Além disso, pode-se observar o peso dos DRCs, que é o resultado da soma das prioridades escolhidas dos DRCs detalhadas na Tabela 6. Para

este indicador, quanto menor é o valor, melhor é a solução. Consequentemente, o cenário CA atingiu um percentual inferior a 20% e é a melhor solução entre os cenários. Esse comportamento é o oposto para o cenário CB, com resultados em torno de 60%. Esses resultados estão diretamente relacionados com a capacidade da rede, sendo apresentados a seguir.

As características de consumo dos nós de transporte, RCs, *links* e latência são analisadas em conjunto, pois os comportamentos são complementares. Essas características mostram que o cenário CA teve a maior ocupação de RCs sob os nós AG1, obtendo cerca de 70% para os nós RU R1 e 80% F1. Além disso, a média de latência apresentou uma diferença significativa entre o cenário CM e o cenário CA. Por exemplo, observou-se uma variação de aproximadamente 45% para os nós RU R1 e 38% para F1. Considerando valores absolutos, a latência é superior a 5ms no cenário CB, contribuindo diretamente para o baixo desempenho neste cenário. Por exemplo, o requisito de DU-RU é 0,250ms, conforme apresentado na Tabela 6.

A capacidade dos *links* dos cenários CA não atingiu mais do que 25% da ocupação, resultando em uma alta agregação de VNFs conforme seu respectivo nível de agregação. O cenário CB apresentou um alto uso dos RCs e uma ocupação dos *links* dos nós AC1s, em comparação com o cenário CA, porque essa topologia possui *links* com suporte de baixa capacidade. Este comportamento ocorre na ocupação dos *links* de AG1: por exemplo, quase 60% da ocupação e com latência alta. Isso resulta em um baixo nível de agregação e um uso dos tipos de DRCs com baixa prioridade (alto peso no estágio 3). Entretanto, o *link* AG2 não apresentou variação significativa de consumo para os cenários. Este comportamento ocorre por que a topologia T1 tem uma característica particular, em que os nós AG1 concentram todos os 51 nós e os nós AG2 apenas parte desses nós, conforme é possível ser visualizado na Figura 12

6.1.4 Considerações Finais

Conforme mostrado na pesquisa, para que sejam atingidos os objetivos do modelo, é necessária a superação das restrições das redes *crosshaul*, para que, assim, seja alcançado o melhor posicionamento de VNFs de RAN, a fim de maximizar a agregação, minimizar os CRs, se houver redução nos números de DRCs, e priorizar as escolhas de DRC com base em estratégias. Nesse sentido, a solução sempre visa à melhor decisão de posicionamento, focada em superar as limitações do *crosshaul*. No entanto, a limitação não restringiu a arquitetura D-RAN escolhida em alguns cenários, o que resulta em uma limitação da implantação da rede de quinta geração. Contudo, para os cenários de topologia com baixa restrição no *crosshaul*, o modelo PlaceRAN atinge um alto nível de agregação nos dois RCs AG1, centralizando os VNFs da RAN e potencializando as redes de quinta

geração, com base na priorização das iniciativas O-RAN. Frente ao estado da arte, relativo ao que há de melhor nesta área do conhecimento, o modelo PlaceRAN é um trabalho inovador e, assim, não é possível comparar os seus resultados aos de outros trabalhos.

O comportamento da posição da unidade vCU e sua agregação nos nós de transporte (principalmente nos nós AG1) tem importância significativa para o desenvolvimento da RAN. Para a arquitetura NG-RAN (3), a agregação das VNFs das unidades vCU está toda concentrada nos nós AG1 para ambas as topologias, com exceção para a topologia T1 com o cenário CB, que não tem DRCs NG-RAN (3). Essa agregação total apresenta a melhor opção de implantação do vNG-RAN. Entretanto, a arquitetura C-RAN não tem o mesmo desempenho, uma vez que atinge 100% dos nós CU VNFs para os cenários CA nos nós AG1 e 100% das unidades vCU VNFs para os cenários CB com nós AC1 para a topologia T1. Nesse caso, a agregação nos nós AC1 não contribui para a centralização da vCU. O resultado NG-RAN (2) apresentou uma agregação dispersa de unidades vCU para os cenários, mas atinge 0% da posição da unidade vCU no AC1 para quaisquer cenários da topologia T2 e 0% em T1 apenas nos cenários CA. A arquitetura D-RAN não possui essa análise por conta de ser monolítica, com todas as unidades da RAN concentradas no nó de entrada da vRU.

Embora o trabalho não explore o quesito custos, a escolha dos três cenários está diretamente orientada para o investimento nas redes. Por exemplo, para cenários CA, com melhor desempenho da solução, há maior necessidade de recursos na rede e, consequentemente, alto investimento. Além disso, a influência do formato da topologia também é fator de impacto no investimento. A atual topologia T1 não é propícia à evolução para redes de quinta geração, principalmente por causa da latência total, considerando diversos processamentos computacionais dos dispositivos de roteamento na camada AC1. Além disso, analisou-se a topologia T1 com um canal de largura de banda de 40MHz, diferente do canal almejado para a quinta geração que é de 100MHz. Entretanto, a topologia T2, aderente à quinta geração, também precisa de investimento, principalmente por causa dos dispositivos ópticos para a topologia no formato árvore. Finalmente, os cenários aplicados mostram um desequilíbrio entre a ocupação dos recursos. Por exemplo, no cenário CA, os nós AC1 e AC2 apresentam baixa ocupação nos *links* e RCs, que podem ser otimizados, e, consequentemente, reduzir o investimento.

6.2 Resultados do experimento da Solução PlaceRAN

Esta seção tem como objetivo apresentar os resultados do estudo de caso com base no protótipo desenvolvido. A primeira avaliação averigua a funcionalidade do estudo de caso da solução PlaceRAN, com base na integração do modelo de otimização com o Orquestrador. A segunda avaliação explora o comportamento dos recursos computacionais

Figura 20: Exemplo de alocação de um encadeamento de VNFs na ferramenta K8S.

NAME	READY	STATUS	RST	AGE
cu-sample-1-cb998cbd5-b2m7k	1/1	Running	0	14h
du-sample-1-6b6cd4d4b8-pbbzk	1/1	Running	0	14h
ru-sample-1-6c7bf85f6b-4zw8r	1/1	Running	0	14h

para com a solução PlaceRAN.

6.2.1 Funcionalidade do Estudo de Caso da Solução PlaceRAN

O objetivo da presente análise é atestar o funcionamento do orquestrador desenvolvido. Primeiramente, conforme descrito na seção 5.2, o estudo de caso possui uma entrada de 23 vRUs e uma consequente saída de 23 encadeamentos das funções de rádio virtualizadas (vCU, vDU e vRU). Para alcançar esta verificação, foram executadas 10 amostras do sistema e todas apresentaram resultado funcional e preciso. Ademais, a saída do posicionamento gerado pelo Otimizador e a alocação realizada com base no sub-bloco Executor obtiveram o mesmo resultado. Dessa forma, a integração do Otimizador e do Orquestrador apresentou estabilidade nas operações e comunicações. Nesse sentido, a Figura 20 apresenta um exemplo de saída de um encadeamento do protótipo por cada Pod vCU, vDU e vRU. O encadeamento das VNFs é verificado com base na conectividade do protocolo IP entre o UE, o CN e a saída para a Internet pelos emuladores OAI e CN. Em média o ping para o site *google.com* via o encadeamento avaliado é de 73.9ms.

Para a conexão do experimento com o Otimizador, a Tabela 14 exhibe os 23 encadeamentos criados a partir da solução do Otimizador, assim como os DRCs escolhidos e os nós de rede da topologia (AG1, AG2 e AC1) na qual foram alocados as unidades de rádio vRU, vDU e vCU. No que se refere às tomadas de decisões, o Otimizador indicou as melhores escolhas, uma vez que os DRC8 ((NG-RAN(3)) e DRC13 (NG-RAN(2)) foram selecionados. A variação entre os mesmos, é apenas a posição dos nós RCs, pois o tipo de DRC8 é selecionado quando a entrada vRU é dada em um nó AC1 e o DRC13 é escolhido com nós da posição AG2. Em se tratando dos objetivos da otimização, o estudo de caso atingiu um nível de agregação de 88% e utilizou apenas seis RCs, seguindo a mesma premissa de 6.1.1 para o resultado da agregação, onde apenas as VNFs $\mathcal{F} = \{f_2, f_3, f_4, f_5, f_6, f_7, f_8\}$ são avaliadas. A agregação não atingiu 100%, pois o emulador OAI e a configuração utilizada não possuem suporte a flexibilidade de posicionamento da VNF $\mathcal{F} = \{f_2\}$. Por outro lado, o alto nível de agregação acontece porque o experimento possui abundantes recursos computacionais e capacidade nas redes *crosshaul*. Tal comportamento se dá, ainda, por conta da topologia T2 ter aderência evolutiva e os requisitos das redes LTE serem muito inferiores aos das redes de quinta geração.

Tabela 14: Escolhas dos DRCs pelo Otimizador para a avaliação experimental.

Encadeamento das VNFs	DRC	Alocação VNFs por Nó RC		
		vCU	vDU	vRU
1	13	2-AG1	3-AG2	3-AC1
2	13	2-AG1	4-AG2	4-AC1
3	13	2-AG1	5-AG2	5-AC1
4	13	2-AG1	6-AG2	6-AC1
5	8	1-AG1	3-AG2	7-AC1
6	8	1-AG1	3-AG2	8-AC1
7	8	1-AG1	3-AG2	9-AC1
8	8	1-AG1	3-AG2	10-AC1
9	8	2-AG1	4-AG2	11-AC1
10	8	1-AG1	3-AG2	12-AC1
11	8	1-AG1	3-AG2	13-AC1
12	8	2-AG1	4-AG2	14-AC1
13	8	2-AG1	4-AG2	15-AC1
14	8	2-AG1	3-AG2	16-AC1
15	8	1-AG1	6-AG2	17-AC1
16	8	1-AG1	5-AG2	18-AC1
17	8	2-AG1	6-AG2	19-AC1
18	8	1-AG1	5-AG2	20-AC1
19	8	1-AG1	5-AG2	21-AC1
20	8	1-AG1	5-AG2	22-AC1
21	8	1-AG1	6-AG2	23-AC1
22	8	1-AG1	5-AG2	24-AC1
23	8	1-AG1	5-AG2	25-AC1

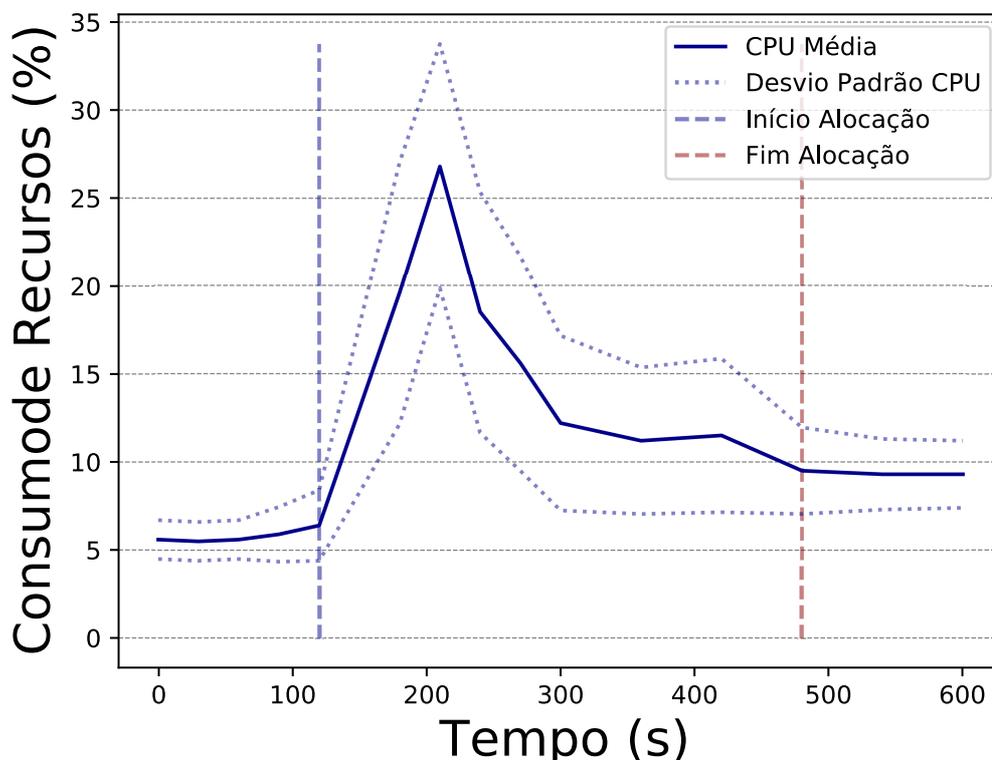
6.2.2 Análise do comportamento dos recursos computacionais

O objetivo dessa análise é investigar o comportamento dos recursos computacionais e a alocação das VNFs através dos Pods. Dessa forma, são avaliados os comportamentos dos consumos do *cluster* (análise total do nós), dos nós divididos em CPU e RAM (de forma particular para cada nó) e dos Pods (a menor unidade virtualizada do K8S, onde estão as VNFs vRU, vDU e vCU). Toda a coleta de dados de consumo é extraída da própria ferramenta de orquestração K8S. Conforme detalhado na seção 6.2, a topologia *crosshaul* não possui análise na avaliação do experimento, pois a mesma é uma entrada no orquestrador e não uma rede simulada ou emulada.

Para análise do *cluster*, as Figuras 21 e 22 apresentam o consumo das métricas CPU e RAM dos RCs ao longo do tempo do experimento. As amostras de tempo foram realizadas em 600 segundos, sendo executadas por dez vezes devido à estabilidade do experimento. Os RCs são medidos pelas médias de CPU e memória RAM. Além disso, as

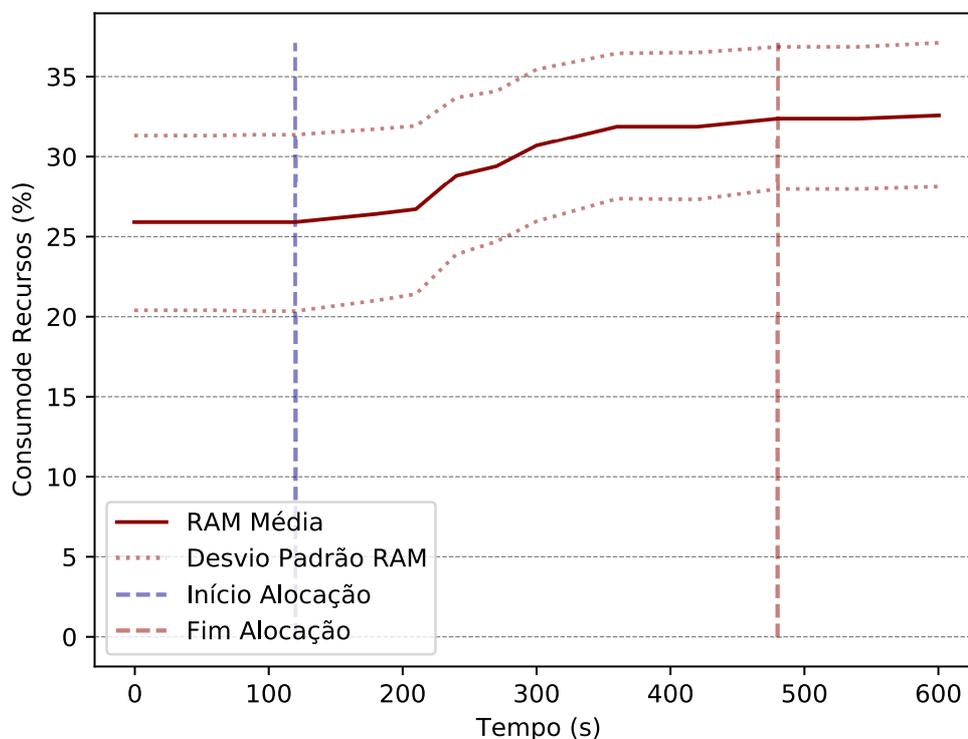
Figuras 21 e 22 mostram o início e o fim do processo de alocação das VNFs. Como é possível verificar o comportamento do *cluster* começa a ter alterações por volta de 100s e um crescimento expressivo do consumo com início em 120s, aproximadamente. Em específico para a métrica CPU, Figura 21, ainda há um decréscimo do consumo, mostrando que a solução é elástica de acordo com a demanda de alocação dos Pods.

Figura 21: Consumo médio e desvio padrão da métrica de processamento CPU dos recursos do cluster (25 RCs) em função do tempo.



Em conformidade com o crescimento do consumo, o comportamento das métricas de CPU e RAM têm respostas distintas. Para a métrica de CPU, é possível verificar um crescimento próximo do linear a partir de 120s, chegando a um consumo superior a 25% do *cluster* por volta de 210s. Entretanto, o consumo de CPU começa a decrescer até atingir a estabilidade, com início próximo a 490s. A RAM possui um crescimento menos abrupto do que a CPU e apresenta estabilidade em torno de 480s. O comportamento de ambas as métricas são esperados, pois o início da alocação das VNFs é próximo de 120s e a conclusão, próxima de 420s. Com base na análise, é possível identificar que a métrica de CPU, quando mal dimensionada ou com restrições de recursos, causa impactos diretos na alocação das VNFs. Embora o tempo não seja uma restrição da solução e não seja

Figura 22: Consumo médio e desvio padrão da métrica de memória RAM dos recursos do cluster (25 RCs) em função do tempo.

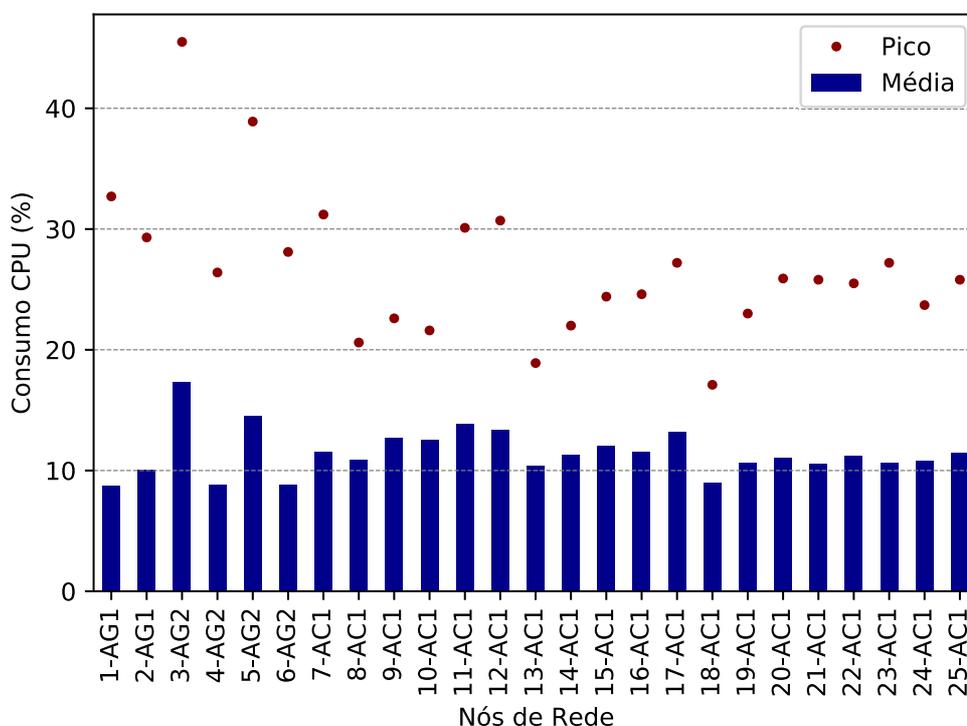


abordado no trabalho, toda a análise do orquestrador é executada em até 120s. Entretanto, vale ressaltar que o tempo do Otimizador apresentado é específico para a abordagem atual, visto que o tempo de análise depende da complexidade da análise (dimensão e possibilidades de encaminhamento do tráfego).

Outra análise referente ao comportamento dos recursos computacionais é relacionada às métricas de CPU e RAM para cada nó RC. A avaliação pode ser visualizada na Figura 23 para a CPU e na Figura 24 para a RAM. A avaliação é realizada em diferentes pontos de análise, pois os comportamentos são distintos. Primeiramente, para a CPU, há uma oscilação média e do pico de consumo. Sendo assim, a Figura 23 apresenta o consumo de CPU médio e de pico em porcentagem para cada nó RC. Entretanto, o comportamento da RAM, visto na Figura 24, mostra um crescimento de consumo e estabilidade, sem oscilação. Considerando esse comportamento, a Figura 24 apresenta um incremento do consumo da RAM de cada nó RC em porcentagem. É importante ressaltar que cada tipo de nó (AG1, AG2 e AC1) deve ser analisado de forma distinta, pois os mesmos possuem diferentes capacidades, conforme exposto na seção 5.2 e, dessa forma, o percentual de

consumo não reflete a comparação entre todos nós.

Figura 23: Apresenta o consumo do recurso de CPU médio e de pico de todos os nós RCs.

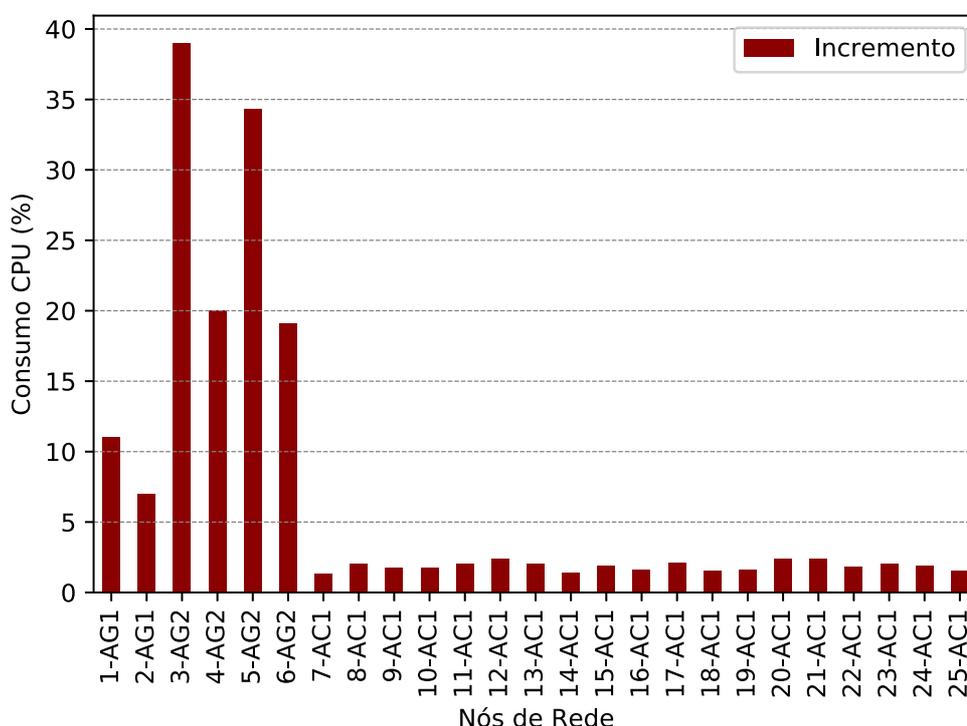


Fonte: Elaborada pelo autor

Inicialmente, de acordo com a Figura 23, é possível constatar que cada nó RC segue o mesmo comportamento verificado no *cluster*, em que a média de consumo de CPU é muito inferior ao consumo de pico, tendo nós RC, por exemplo, como o 3-AG2, no qual o pico é quase três vezes a sua média. De forma geral, os nós RCs AC1 possuem um consumo médio semelhante, em torno de 10% da sua capacidade e com uma flutuação no pico de consumo, mesmo que todos tenham a mesma aplicação vRU. Para os nós AG2, existe uma flutuação de consumo médio de CPU em virtude da quantidade de alocação de Pods e este é explorado na próxima análise. O comportamento da RAM apresenta uma relação direta com a quantidade de Pods alocados. Esse comportamento é também explorado na próxima análise. No entanto, a mesma exibe um consumo semelhante para os nós AC1 com uma variação de alocação muito baixa para a aplicação vRU.

Para o consumo de recursos por Pods, a Tabela 15 apresenta os resultados encontrados no experimento. Foram priorizadas as análises de todos os nós RCs, por exemplo, AG1 e AG2, e de quatro nós RC AC1, pois esses nós constituem-se em uma totalidade de 19 e apresentam comportamentos semelhantes, conforme a Figura 23 e Figura 24. A Tabela 15

Figura 24: Consumo incremental dos recursos de RAM de todos os nós RCs.



Fonte: Elaborada pelo autor

mostra a quantidade de Pods divididos nas unidades vCU, vDU e vRU por nó RCs, assim como a totalidade de Pods por nó RC. Os dados estão alinhados com a Tabela 14. Além disso, a Tabela 15 segue as mesmas premissas da segunda análise, na qual são analisados a CPU média e de pico e o incremento de RAM, todos com a unidade porcentagem.

Para a avaliação do consumo dos RCs, é válido ressaltar que as diferenças de porcentagens por tipos de nós RCs (AC1, AC2 e AC1) refletem a capacidade entre os mesmos. Embora os nós RCs AG1 recebam uma maior carga de Pods, não apresentam grandes variações devido à sua maior capacidade de recursos e, também, por característica da aplicação. Nesse caso, a vCU com o emulador OAI possui os protocolos da RAN RRC e PDCP. Nesse sentido, como a abordagem da solução PlaceRAN não avalia o tráfego de UEs, não há carga de trabalho para as unidades vCU, havendo um consequente baixo uso dos recursos. Contudo, os nós RCs AG2 possuem dois tipos de unidades: (i) vDU e (ii) vRU, sendo que o vDU possui alto volume de unidades. Entretanto, o comportamento das unidades apresenta estabilidade, mesmo com diferença entre a quantidade de vDU. Por fim, os Pods vRUs em nós AC1 possuem o mesmo incremento de consumo da RAM e comportamento não uniforme da métrica CPU, tendo, para alguns, pico de 100% de uso

Tabela 15: Consumo computacional com base na alocação das VNFs da RAN por Pods.

Nós de Rede	Pods vCU	Pods vDU	Pods vRU	Total Pods	Média CPU(%)	Pico CPU (%)	Incremento RAM (%)
1-AG1	14			14	0,68	6,25	0,75
2-AG1	9			9	0,82	3,10	0,74
3-AG2		8	1	9	9,47	12,50	5,20
4-AG2		4	1	5	9,81	12,50	3,98
5-AG2		7	1	8	8,55	12,50	5,20
6-AG2		4	1	5	8,64	12,50	5,20
7-AC1			1	1	35,00	50,00	13,50
12-AC1			1	1	35,01	50,00	13,50
17-AC1			1	1	65,00	100,00	13,50
25-AC1			1	1	60,00	100,00	13,50

de CPU. Tal comportamento, de maior consumo de CPU, é verificado para unidade vRU do emulador OAI, uma vez que as unidades vRU emulam os protocolos PHY, o qual está mais próximo das camadas físicas da RAN e, no estudo de caso, possui menos recursos computacionais.

6.2.3 Considerações Finais

Conforme mostrou-se neste estudo, a solução PlaceRAN é a união entre o Orquestrador e o modelo de otimização de posicionamento. Sendo assim, o principal objetivo do experimento é criar um ambiente de suporte de testes para o modelo de otimização. Nesse sentido, a integração apresentou funcionalidade total entre as três principais tarefas do sistema, tais como: *(i)* a tomada de decisão do Otimizador, *(ii)* a interação entre os sub-blocos Gerenciador, Otimizador e Executor e *(iii)* a alocação das VNFs da RAN pelo agendador do K8S. Embora a métrica tempo não seja avaliada por conta de a solução ser direcionada ao planejamento de redes, para o cenário desenvolvido com 25 nós, o comportamento temporal obteve resposta inferior a 100s.

No que se refere ao comportamento dos recursos computacionais para o desenvolvimento da vNG-RAN, as duas métricas, CPU e RAM, apresentaram comportamentos distintos, porém esperados. A CPU apresentou alto consumo na alocação das VNFs, mas retornou, de forma geral, a um ponto de estabilidade. Entretanto, a métrica RAM apresentou estabilidade de consumo após a implementação e comportamento muito próximo

entre as mesmas VNFs das unidades de rádio (vCU, vDU e vRU). Outro ponto relevante da análise relacionado ao resultado do comportamento das métricas analisadas é a aplicação, que, neste caso, é feita pelo emulador OAI, que no presente trabalho é apenas utilizado para validar o encadeamento das VNFs. Entretanto, é possível ressaltar a necessidade de estabilidade entre as aplicações de emulação ou reais da RAN e as infraestruturas computacionais para que as arquiteturas vNG-RAN possam ser alavancadas. Os resultados não possuem comparações com outros trabalhos, pois a solução PlaceRAN aborda um tópico ainda não desenvolvido na literatura, conforme a máxima exploração de nosso conhecimento.

7 CONCLUSÃO

A evolução das redes móveis de quinta geração impulsiona a transformação digital em diversos segmentos da sociedade. A liderança dessas transformações é conduzida pelas organizações padronizadoras ITU-T, 3GPP e ETSI, que, juntas, introduzem grandes mudanças de arquiteturas, conceitos e tecnologias. Tais mudanças são introduzidas especificamente pelo *Release 15* do 3GPP (3GPP, 2018), uma vez que endereça as evoluções em todas as cadeias das redes móveis, em especial na RAN, que evolui para NG-RAN, passando por transformações significativas na sua arquitetura e imergindo nos conceitos de virtualização, resultando, assim, na vRAN. Tais evoluções são complementares e possibilitam o desenvolvimento da vNG-RAN, o que abre espaço para inúmeras oportunidades de pesquisa.

O principal ponto de investigação, neste trabalho, foi endereçado na seguinte questão de pesquisa: *Como poderia constituir-se uma solução de orquestração para o posicionamento das funções de rádios desagregados e virtualizados da rede de acesso de nova geração?* A solução PlaceRAN responde à pergunta, com base no planejamento de redes desenvolvido com um orquestrador que suporta a otimização do posicionamento das funções de rádio virtualizadas. Pelo fato de a solução PlaceRAN possuir duas investigações complementares, duas questões de pesquisa específicas foram desenvolvidas para colaborar com a questão de pesquisa principal.

A primeira questão específica é direcionada à orquestração: *Como modelar um orquestrador para o posicionamento dos rádios desagregados e virtualizados alinhados com os padrões definidos para vNG-RAN?* Três sub-blocos foram criados baseados na arquitetura NFV para responder essa pergunta: (i) Gerenciador, (ii) Executor e (iii) Otimizador. O Gerenciador é orientado pela orquestração de recursos de rede do NFVO e gerencia todas as entradas internas e externas. Além disso, esse sub-bloco comunica-se com a infraestrutura computacional e envia a demanda para a alocação para o sub-bloco Executor. O sub-bloco Executor é alinhado com a coordenação dos serviços de rede e produz toda a execução do encadeamento das funções virtualizadas que são geradas pelo Otimizador.

O sub-bloco Otimizador foi concebido com foco na segunda questão de pesquisa específica: *Como representar o posicionamento das funções virtuais de uma rede de rádio de acesso desagregada usando o menor número de recursos computacionais?* Para responder essa segunda questão, foi criada uma modelagem de otimização exata para ser utilizada na arquitetura vNG-RAN. A modelagem tem como base a melhor tomada de decisão para a escolha das desagregações da pilha de protocolos da RAN, e supera as limitações dos recursos computacionais e das redes *crosshaul*. Para realizar essa mode-

lagem, primeiramente, foi introduzido o conceito DRC com o mapeamento de todas as opções de combinações da vNG-RAN e alinhamento com as iniciativas O-RAN. Nesse sentido, o Otimizador foi concebido com três estágios: (i) minimizar os recursos computacionais e maximizar as agregações das funções de rádio virtualizadas, (ii) minimizar o número do uso de DRCs e (iii) priorizar, por meio de pesos, as escolhas dos DRCs.

Baseado no desenvolvimento do orquestrador e do modelo de otimização, foram definidas duas avaliações complementares: uma avaliação experimental com estudo de caso para o orquestrador e uma conceitual para o modelo de otimização, a fim de ser utilizado no sub-bloco Otimizador. O Orquestrador foi desenvolvido sob a ferramenta K8S, explorando a funcionalidade dos Operadores dessa ferramenta. Além disso, as desagregações dos elementos vRU, vDU e vCU foram desenvolvidas utilizando a ferramenta OAI. Para analisar o comportamento do orquestrador, são observadas as funcionalidades entre a saída do sub-bloco Otimizador e as implementações dos encadeamentos das funções de rádio enviadas do sub-bloco Executor para a alocação pelo K8S. Por fim, o trabalho apresentou uma avaliação de desempenho considerando as métricas de CPU e RAM.

Para a análise conceitual, foram utilizadas duas topologias de redes reais, dados de recursos computacionais com base em ensaios reais e a determinação dos canais de rádio sob a perspectiva da quinta geração de redes móveis. As cargas de trabalho aplicadas na análise conceitual dizem respeito à entrada de unidades vRU, à capacidade dos RCs e a capacidade das redes *crosshaul*. Além disso, o Otimizador foi avaliado de forma experimental. O estudo de caso foi analisado com dados de uma topologia real, com RCs no formato de VMs, com parâmetros de desagregação da rede LTE (devido ao suporte do OAI) e com entrada de unidades vRU. Além disso, o estudo de caso foi desenvolvido em alinhamento com a arquitetura NFV, sob a gerência do Orquestrador pelo sub-bloco Gerenciador e com envio dos dados de alocação das VNFs pelo sub-bloco Executor.

Os resultados indicaram que a solução PlaceRAN pode alcançar até 80% de agregação total (independente do nó de rede) das funções de rádio virtualizadas em conjunto com menos de 20% dos recursos computacionais das duas topologias avaliadas. Esses resultados são totalmente alinhados com as prerrogativas do desenvolvimento da arquitetura vNG-RAN. Entretanto, tais números dependem diretamente das capacidades disponíveis na rede. Também, a integração do otimizador com o experimento demonstrou total viabilidade do desenvolvimento do vNG-RAN para os ambientes de orquestração de contêineres a partir da emulação da RAN. Os resultados indicam que o orquestrador é funcional desde a saída do sub-bloco Otimizador e até alocação das VNFs nas VMs para todas as dez amostras. Finalmente, o tempo total de execução do processo, desde a entrada dos dados pelo operador de rede até a alocação das VNFs, apresentou uma estabilidade média inferior a 500s e o tempo médio da solução PlaceRAN foi inferior a 120s, embora o fator tempo não seja determinante para o objetivo da solução de Planejamento.

Dessa forma, pode-se concluir que a solução PlaceRAN promove o desenvolvimento da arquitetura vNG-RAN.

7.1 Contribuições

O PlaceRAN é uma solução de orquestração e otimização do posicionamento das funções virtualizadas da vNG-RAN que busca uma solução ótima e que seja aplicada por meio de um orquestrador alinhado com as prerrogativas da principal arquitetura de virtualização das redes móveis, i.e., a arquitetura NFV. Sendo assim, traz contribuições a partir das características que apresenta, pontuadas abaixo:

- O orquestrador especializado na otimização do posicionamento da vNG-RAN, com ciência de topologia *crosshaul* e de recursos computacionais. Também possui alinhamento com a arquitetura de referência NFV, e, além disso, é passível de integração com outras plataformas de desenvolvimento das redes móveis de quinta geração.
- É um modelo de otimização de posicionamento exato para a vNG-RAN flexível e com capacidade de seguir os preceitos da nova geração da RAN, que é baseada em duas funcionalidades principais: (i) alta agregação e (ii) alta distribuição. O modelo visa à melhor tomada de decisão com alto índice de DRCs, aplicado a qualquer topologia de rede e a quaisquer capacidades de recursos computacionais.
- Para a comunidade científica e sociedade, a solução PlaceRAN é um contribuinte para o impulsionamento da tecnologia de quinta geração, que é considerada o principal propulsor da transformação digital. Além disso, a exploração da vRAN colabora para a redução de investimento, o que é determinante para o avanço rápido de novas tecnologias.

Em comparação com os trabalhos relacionados apresentados no capítulo 3, a solução PlaceRAN assume destaque, pois é a única solução de orquestração para a otimização do posicionamento especializada para vNG-RAN. Comparando o PlaceRAN com o trabalho de Matoussi et al. (2020), identifica-se que a pesquisa citada ainda utiliza a arquitetura C-RAN, parâmetros da tecnologia LTE, e é focada no emulador OAI e sem orientação para uma arquitetura de referência. Além disso, no trabalho de Matoussi et al. (2020), apenas a rede *fronthaul* é investigada e não há flexibilidade na desagregação. No âmbito da modelagem de otimização, a solução PlaceRAN é a primeira a ter uma abordagem exata com até nove possibilidades de desagregação (9 DRCs) e três unidades de rádio (CU, DU e RU), a fim de tomar a melhor decisão para superar as limitações das redes *crosshaul* e dos recursos computacionais.

7.2 Limitações

Na presente seção são elencadas possíveis limitações identificadas durante a construção da solução PlaceRAN. A primeira das limitações do modelo é o fator tempo para redes de maior escala (maior número de nós) e de topologias com maiores possibilidades de encaminhamento/caminhos para o tráfego. Nesse sentido, faz-se necessária uma máquina com alta capacidade para realizar os testes. Para o orquestrador, o experimento não pode ter uma escala semelhante à análise conceitual devido à limitação de recursos. Além disso, a entrada dos dados das redes *crosshaul* é estática e não é gerada a partir de uma rede em operação, também por conta da limitação de recursos. Por fim, o emulador OAI, embora seja uma das ferramentas utilizadas no desenvolvimento da RAN, possui instabilidades para realizar testes mais elaborados.

7.3 Trabalhos Futuros

Com base nas limitações apresentadas anteriormente, identifica-se oportunidades para o desenvolvimento de futuros trabalhos para a solução PlaceRAN, conforme detalhado abaixo:

- Redução do tempo de análise: possibilidade de implementar métodos e técnicas de modelagem que possam contribuir para a redução do tempo de execução da análise e que não ocasionem impacto nas funções objetivos dos estágios do modelo.
- Carga de trabalho baseada em UEs: oportunidade de gerar diferentes cargas de trabalho nas UEs, representando serviços variados. Realizar testes com diferentes comportamentos temporais do tráfego, por exemplo, urbano, suburbano e rural.
- Aprimoramento do modelo de otimização dos recursos das redes *crosshaul*: abertura para introduzir o conceito de fluxo de tráfego na modelagem, pois hoje o modelo opera com a alocação de largura de banda total nos caminhos. Com base no conceito de fluxo, integrar ao orquestrador os conceitos de redes programáveis por *software* (SDN - *Software Defined Networking*) para uma entrada real dos dados das redes *crosshaul*.
- Orquestrador de soluções de otimização genérica: possibilidade de atribuir melhorias ao orquestrador para que ele possa tornar-se uma plataforma genérica para o desenvolvimento de otimizações de posicionamento da RAN, tendo em vista a lacuna existente de análises experimentais no tópico sobre posicionamento de funções virtualizadas.

7.4 Publicações

Uma publicação foi obtida durante o período do mestrado. Essa publicação tem relação direta com os desenvolvimentos da RAN e com conceitos de *software*, conforme abaixo:

1. *When SDN meets C-RAN: A survey exploring multi-point coordination, interference, and performance*. Journal of Network and Computer Applications, p. 102655, 2020. <https://doi.org/10.1016/j.jnca.2020.102655>.

Quanto à solução PlaceRAN, estão sendo geradas duas publicações, no que se refere à submissão de artigos: uma para a formalização do problema de otimização do posicionamento e outra para o orquestrador. O primeiro, referente à otimização do posicionamento, está disponível conforme indicado abaixo no formato de *preprint*, estando, contudo, ainda sem aprovação da submissão. O segundo trabalho, relacionado ao orquestrador, encontra-se em elaboração.

1. *PlaceRAN: Optimal Placement of Virtualized Network Functions in the Next-Generation Radio Access Networks*. <https://arxiv.org/abs/2102.13192>.

REFERÊNCIAS

- 3GPP. **Study on New Radio Access Technology; Radio Access Architecture and Interfaces (Release 14)**. [S.l.]: 3rd Generation Partnership Project (3GPP), 2017. Technical Recommendation (TR). (38.801).
- 3GPP. **Service Requirements for the 5G System, document TS 22.261 v16.0.0**. [S.l.]: 3rd Generation Partnership Project (3GPP), 2017. Technical Specification (TS). (22.261).
- 3GPP. **System Architecture for the 5G (Release 15)**. [S.l.]: 3rd Generation Partnership Project (3GPP), 2018. Technical Recommendation (TR). (23.501).
- 5G-crosshaul Project. **5G-crosshaul, D1.2: final 5g-crosshaul system design and economic analysis**. 2017.
- AFOLABI, I. et al. Network slicing and softwarization: a survey on principles, enabling technologies, and solutions. **IEEE Communications Surveys & Tutorials**, [S.l.], v. 20, n. 3, p. 2429–2453, 2018.
- AGRAWAL, R. et al. Cloud RAN challenges and solutions. **Annals of Telecommunications**, [S.l.], v. 72, n. 7-8, p. 387–400, 2017.
- AL-DULAIMI et al. **5G Networks: fundamental requirements, enabling technologies, and operations management**. [S.l.]: John Wiley & Sons, 2018.
- ALLIANCE, O. R. Cloud Architecture and Deployment Scenarios for O-RAN Virtualized RAN. **White Paper**, [S.l.], 2020.
- ALLIANCE, O. S. **OpenAirInterface**. [Último acesso: 26-abril-2020], <https://www.openairinterface.org/>.
- AROUK, O. et al. Multi-objective placement of virtual network function chains in 5G. In: IEEE 6TH INTERNATIONAL CONFERENCE ON CLOUD NETWORKING (CLOUDNET), 2017., 2017. **Anais...** [S.l.: s.n.], 2017. p. 1–6.
- AROUK, O. et al. Cost optimization of cloud-RAN planning and provisioning for 5G networks. In: IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS (ICC), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p. 1–6.
- BERNARDOS, C. J. et al. **Network virtualization research challenges**. [S.l.]: IETF, 2019.
- BERTENYI, B. et al. NG radio access network (NG-RAN). **Journal of ICT Standardization**, [S.l.], v. 6, n. 1, p. 59–76, 2018.
- BHAMARE, D. et al. Efficient virtual network function placement strategies for cloud radio access networks. **Computer Communications**, [S.l.], v. 127, p. 50–60, 2018.

BONATI, L. et al. Open, Programmable, and Virtualized 5G Networks: state-of-the-art and the road ahead. **arXiv preprint arXiv:2005.10027**, [S.l.], 2020.

BOTH, C.; GUIMARAES, R.; SLYNE, F.; WICKBOLDT, J.; MARTINELLO, M.; DOMINICINI, C.; MARTINS, R.; ZHANG, Y.; CARDOSO, D.; VILLACA, R. et al. FUTEBOL Control Framework: enabling experimentation in convergent optical, wireless, and cloud infrastructures. **IEEE Communications Magazine**, [S.l.], v. 57, n. 10, p. 56–62, 2019.

CASALICCHIO, E. Container orchestration: a survey. In: **Systems Modeling: methodologies and tools**. [S.l.]: Springer, 2019. p. 221–235.

CHEN, T. et al. Software defined mobile networks: concept, survey, and research directions. **IEEE Communications Magazine**, [S.l.], v. 53, n. 11, p. 126–133, 2015.

Cloud Native Computing Foundation. **Kubernetes Documentation | Kubernetes**. 2020.

DALLA-COSTA, A. G.; BONDAN, L.; WICKBOLDT, J. A.; BOTH, C. B.; GRANVILLE, L. Z. Orchestra: a customizable split-aware nfv orchestrator for dynamic cloud radio access networks. **IEEE Journal on Selected Areas in Communications**, [S.l.], v. 38, n. 6, p. 1014–1024, 2020.

DOBIES, J.; WOOD, J. **Kubernetes Operators - Automating the container orchestration platform**. [S.l.]: o'reilly, 2020.

Docker Inc. **Docker Documentation**. 2020.

DZOGOVIĆ, B. et al. Connecting Remote eNodeB with Containerized 5G C-RANs in OpenStack Cloud. In: IEEE INTERNATIONAL CONFERENCE ON CYBER SECURITY AND CLOUD COMPUTING (CSCLOUD)/2019 5TH IEEE INTERNATIONAL CONFERENCE ON EDGE COMPUTING AND SCALABLE CLOUD (EDGECON), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 14–19.

D'AMBROSIA, J. **The State of Ethernet**. 2017.

ERIKSSON, M. **Placement of services in distributed clouds**. 2017.

ETSI, G. N.-P. Network Functions Virtualisation (NFV): nfv performance & portability best practises. , [S.l.], 2014.

ETSI, G. N.-T. Network Functions Virtualisation (NFV) Release 2; Testing; NFVI Compute and Network Metrics Specification. , [S.l.], 2018.

ETSI, I. **Network functions virtualisation (nfv)-network operator perspectives on industry progress**. [S.l.: s.n.], 2013.

ETSI, N. **Network functions virtualisation (NFV); terminology for main concepts in NFV**. [S.l.: s.n.], 2014.

ETSI, T. **5G; NG-RAN; Architecture description (ETSI TS 138 401 V15.5.0 (2019-05))**. [S.l.: s.n.], 2019.

ETSI, G. N.-E. **Network Functions Virtualisation (NFV); Virtualisation Technologies; Report on the application of Different Virtualisation Technologies in the NFV Framework**. [S.l.: s.n.], 2016.

FIORANI, M. et al. Transport Abstraction Models for an SDN-Controlled Centralized RAN. **IEEE Communications Letters**, [S.l.], v. 19, n. 8, p. 1406–1409, 2015.

FONSECA, F. F. et al. Optimizing allocation and positioning in a disaggregated radio access network. In: XXXVII SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS, 2019. **Anais...** [S.l.: s.n.], 2019. p. 791–804.

FORUM, S. C. Small cell virtualization: functional splits and use cases. , [S.l.], 2016.

FOUNDATION, L. **Open Network Automation Platform**. 2020.

FOUNDATION, T. L. **Kubernetes Documentation**. 2020.

FREE5GC. **Free5GC**. [Último acesso: 08-fevereiro-2020], <https://free5gc.org/>.

GARCIA-SAAVEDRA, A. et al. Fluidran: optimized vran/mec orchestration. In: IEEE INFOCOM 2018-IEEE CONFERENCE ON COMPUTER COMMUNICATIONS, 2018. **Anais...** [S.l.: s.n.], 2018. p. 2366–2374.

GAVRILOVSKA, L. et al. From Cloud RAN to Open RAN. **Wireless Personal Communications**, [S.l.], p. 1–17, 2020.

GSMA. **Migration from Physical to Virtual Network Functions – Best Practices and Lessons Learned**. 2018.

GUERZONI, R. et al. Network functions virtualisation: an introduction, benefits, enablers, challenges and call for action, introductory white paper. In: SDN AND OPENFLOW WORLD CONGRESS, 2012. **Anais...** [S.l.: s.n.], 2012. v. 1, p. 5–7.

HABIBI, M. et al. A Comprehensive Survey of RAN Architectures Toward 5G Mobile Communication System. **IEEE Access**, [S.l.], v. 7, p. 70371–70421, 2019.

HARUTYUNYAN, D. et al. Flex5G: flexible functional split in 5g networks. **IEEE Transactions on Network and Service Management**, [S.l.], v. 15, n. 3, p. 961–975, 2018.

HARUTYUNYAN, D. et al. CU placement over a reconfigurable wireless fronthaul in 5G networks with functional splits. **International Journal of Network Management**, [S.l.], v. 30, n. 1, p. e2086, 2020.

HOSSAIN, F. et al. Recent research in cloud radio access network (C-RAN) for 5G cellular systems-A survey. **Journal of Network and Computer Applications**, [S.l.], v. 139, p. 31–48, 2019.

HU, Y.; WANG, J. Architectural and Cost Implications of the 5G Edge NFV Systems. In: IEEE 37TH INTERNATIONAL CONFERENCE ON COMPUTER DESIGN (ICCD), 2019. **Anais...** [S.l.: s.n.], 2019. p. 594–603.

HUAWEI. White Paper on Huawei MS-OTN Low-latency Network solution.

Available at

<https://www-file.huawei.com/.../huawei-optical-network-low-Latency-solution-white-paper-en.pdf> (access on November 11, 2020).

ITU-T. **Information technology – Cloud computing – Overview and vocabulary.** Geneva: International Telecommunication Union, 2014. Recommendation. (Y.3500).

ITU-T, G.-T. **Transport network support of IMT-2020/5G.** [S.l.]: ITU-T, 2018.

ITU-T, G.-T. **Application of optical transport network Recommendations to 5G transport - Serir G Supplement 67.** [S.l.]: ITU-T, 2019.

ITU-T, S. S.; OF. **ITU-T Focus Group IMT-2020 Deliverables.** [S.l.: s.n.], 2017.

KAZUNARI, K. **Approach to Commercial Use of OAI.** Available at

<https://www.openairinterface.org/docs/.../20171108.pdf> (access on November 11, 2020).

KONDEPU, K. et al. Orchestrating lightpath adaptation and flexible functional split to recover virtualized RAN connectivity. In: OPTICAL FIBER COMMUNICATIONS CONFERENCE AND EXPOSITION (OFC), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p. 1–3.

KREHER, R.; GAENGER, K. **LTE signaling: troubleshooting and optimization.** [S.l.]: John Wiley & Sons, 2010.

KUBERNETES.IO. **Kubernetes Plataform.** Acessado: 2020-06-30,

<https://kubernetes.io/>.

LAGHRISSI, A.; TALEB, T. A survey on the placement of virtual resources and virtual network functions. **IEEE Communications Surveys & Tutorials**, [S.l.], v. 21, n. 2, p. 1409–1434, 2018.

LANDON, P. **DWDM & Packet Optical Fundamentals: how to troubleshoot the transmission layer.** Available at

<https://archive.nanog.org/sites/.../20150602-Landon-Tutorial-DwDM-v3.pdf> (access on November 11, 2020).

LARSEN, L. M. et al. A survey of the functional splits proposed for 5G mobile crosshaul networks. **IEEE Communications Surveys & Tutorials**, [S.l.], v. 21, n. 1, p. 146–172, 2018.

LIN, P.-C.; HUANG, S.-L. Performance profiling of cloud radio access networks using openairinterface. In: ASIA-PACIFIC SIGNAL AND INFORMATION PROCESSING ASSOCIATION ANNUAL SUMMIT AND CONFERENCE (APSIPA ASC), 2018. **Anais...** [S.l.: s.n.], 2018. p. 454–458.

- LUO, Y.-C. et al. A computation workload characteristic study of c-ran. In: IEEE 38TH INTERNATIONAL CONFERENCE ON DISTRIBUTED COMPUTING SYSTEMS (ICDCS), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p. 1599–1603.
- MAHAPATRA, B. et al. Optimal Placement of Centralized BBU (C-BBU) for Fronthaul and Backhaul Optimization in Cloud-RAN Network. In: INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY (ICIT), 2017., 2017. **Anais...** [S.l.: s.n.], 2017. p. 107–112.
- MAKRIS, N.; ZARAFETAS, C.; CHOUMAS, K.; FLEGKAS, P.; KORAKIS, T. Virtualized Heterogeneous 5G Cloud-RAN deployment over Redundant Wireless Links. In: IEEE INTERNATIONAL SYMPOSIUM ON LOCAL AND METROPOLITAN AREA NETWORKS (LANMAN), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 1–2.
- MAMUSHIANE, L. et al. Overview of 9 Open-Source Resource Orchestrating ETSI MANO Compliant Implementations: a brief survey. In: IEEE 2ND WIRELESS AFRICA CONFERENCE (WAC), 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 1–7.
- MARSCH, P. et al. **5G system design: architectural and functional considerations and long term research.** [S.l.]: John Wiley & Sons, 2018.
- MASDARI, M. et al. An overview of virtual machine placement schemes in cloud computing. **Journal of Network and Computer Applications**, [S.l.], v. 66, p. 106–127, 2016.
- MASOUDI, M. et al. Cost-Effective Migration Toward Virtualized C-RAN With Scalable Fronthaul Design. **IEEE Systems Journal**, [S.l.], 2020.
- MATOUSSI, S. et al. 5G RAN: functional split orchestration optimization. **IEEE Journal on Selected Areas in Communications**, [S.l.], v. 38, n. 7, p. 1448–1463, 2020.
- MAVENIR. **Transforming the Radio Access Network - Mavenir's Virtualized RAN.** [S.l.]: Mavenir, 2019.
- MOLNER, N. et al. Optimization of an integrated fronthaul/backhaul network under path and delay constraints. **Ad Hoc Networks**, [S.l.], v. 83, p. 41–54, 2019.
- MOSAIC5G.IO. **Mosaic5G.** Acessado: 2020-12-30, <https://mosaic5g.io//>.
- MURTI, F. W. et al. Optimal Deployment Framework for Multi-Cloud Virtualized Radio Access Networks. **IEEE Transactions on Wireless Communications**, [S.l.], p. 1–1, 2020.
- MURTI, F. W. et al. On the Optimization of Multi-Cloud Virtualized Radio Access Networks. **arXiv preprint arXiv:2002.10681**, [S.l.], 2020.
- NIKAEIN, N. e. a. Mosaic5G: agile and flexible service platforms for 5g research. **ACM SIGCOMM Computer Communication Review**, [S.l.], v. 48, n. 3, p. 29–34, 2018.

NOVAES, C. et al. Virtualized C-RAN Orchestration with Docker, Kubernetes and OpenAirInterface. **XXXVII SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS - SBrT2019**, [S.l.], 2020.

OSM. **OSM Release Three, A Technical Overview**. 2017.

OSNM, E. **Open Source MANO**. 2020.

PASSION Project. **Photonic technologies for programmable transmission and switching modular systems based on Scalable Spectrum/space aggregation for future agile high capacity metro Networks**. 2020.

PLIATSIOS, D. et al. Realizing 5G vision through Cloud RAN: technologies, challenges, and trends. **EURASIP Journal on Wireless Communications and Networking**, [S.l.], n. 1, p. 136, 2018.

RADUKIĆ, S. et al. Effects of Digital Transformation and Network Externalities in the Telecommunication Markets. **Economics**, [S.l.], v. 7, n. 2, p. 31–42, 2019.

Red Hat. **What is container orchestration?** 2020.

REHMAN, A. et al. Network functions virtualization: the long road to commercial deployments. **IEEE Access**, [S.l.], v. 7, p. 60439–60464, 2019.

RESEARCH, A. **Economic Advantages of Virtualizing the RAN in Mobile Operators' Infrastructures**. 2019.

RUIZ, L. et al. A genetic algorithm for VNF provisioning in NFV-Enabled Cloud/MEC RAN architectures. **Applied Sciences**, [S.l.], v. 8, n. 12, p. 2614, 2018.

SAHA, R. K.; TSUKAMOTO, Y.; NANBA, S.; NISHIMURA, K.; YAMAZAKI, K. Novel M-CORD based multi-functional split enabled virtualized cloud ran testbed with ideal fronthaul. In: **IEEE GLOBECOM WORKSHOPS (GC WKSHPs)**, 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p. 1–7.

SEUNG-QUE, L. et al. Virtualization and Orchestration of Radio Access Network. In: **IEEE INTERNATIONAL CONFERENCE ON COMMUNICATION SYSTEMS (ICCS)**, 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p. 321–325.

SHARMA, S. **ICT Development trends and approaches for Digital Transformation**. 2019.

SONG, S. et al. Clustered virtualized network functions resource allocation based on context-aware grouping in 5G edge networks. **IEEE Transactions on Mobile Computing**, [S.l.], v. 19, n. 5, p. 1072–1083, 2019.

THOR. **ThoR - Deliverable D2.1 - Requirements for B5G backhaul/fronthaul**. Available at <https://thorproject.eu/wp-content/uploads/2018/12/ThoR-D2.1-Requirements-for-B5G-back-fronthaul.pdf> (access on November 11, 2020).

YEOH, C. Y. et al. Performance study of LTE experimental testbed using OpenAirInterface. In: INTERNATIONAL CONFERENCE ON ADVANCED COMMUNICATION TECHNOLOGY (ICACT), 18., 2016. **Anais...** [S.l.: s.n.], 2016. p. 617–622.

YI, B. et al. A comprehensive survey of network function virtualization. **Computer Networks**, [S.l.], v. 133, p. 212–262, 2018.

YUSUPOV, J. et al. Multi-objective function splitting and placement of network slices in 5G mobile networks. In: IEEE CONFERENCE ON STANDARDS FOR COMMUNICATIONS AND NETWORKING (CSCN), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p. 1–6.

ZHANG, Y. **Network Function Virtualization: concepts and applicability in 5G networks**. [S.l.]: John Wiley & Sons, 2018.