

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**ALOCÇÃO DE DOIS NÍVEIS PARA UMA
ARQUITETURA H-CRAN BASEADA EM
*OFFLOADING***

MARIANE DE PAULA DA SILVA GONÇALVES

DM 02/2019

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil

01/2019

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**ALOCÇÃO DE DOIS NÍVEIS PARA UMA
ARQUITETURA H-CRAN BASEADA EM
*OFFLOADING***

MARIANE DE PAULA DA SILVA GONÇALVES

DM 02/2019

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil

01/2019

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

MARIANE DE PAULA DA SILVA GONÇALVES

**ALOCAÇÃO DE DOIS NÍVEIS PARA UMA
ARQUITETURA H-CRAN BASEADA EM
*OFFLOADING***

Dissertação submetida à Banca Examinadora do Programa de Pós-Graduação em Engenharia Elétrica da UFPA para a obtenção do Grau de Mestre em Engenharia Elétrica na área de Concentração em Computação Aplicada.

Orientador: Prof. Dr. Diego Lisboa Cardoso
Coorientador: Prof. Dr. Fabrício José Brito Barros

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém-Pará-Brasil

01/2019

**Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD
Sistema de Bibliotecas da Universidade Federal do Pará
Gerada automaticamente pelo módulo Ficat, mediante os dados fornecidos pelo(a) autor(a)**

D111a da Silva Gonçalves, Mariane de Paula.
ALOCAÇÃO DE DOIS NÍVEIS PARA UMA ARQUITETURA H-CRAN BASEADA EM
OFFLOADING / Mariane de Paula da Silva Gonçalves, . — 2019.
67 f. : il. color.

Orientador(a): Prof. Dr. Diego Lisboa Cardoso
Coorientador(a): Prof. Dr. Fabrício José Brito Barros
Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia Elétrica, Instituto de Tecnologia,
Universidade Federal do Pará, Belém, 2019.

1. Redes Móveis, 5G, H-CRAN, QoS, Offloading, TLA.. I. Título.

CDD 621.31098115

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

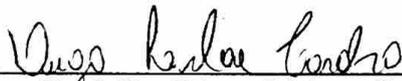
"ALOCAÇÃO DE DOIS NÍVEIS PARA UMA ARQUITETURA H-CRAN BASEADA EM OFFLOADING"

AUTORA: MARIANE DE PAULA DA SILVA GONÇALVES

DISSERTAÇÃO DE MESTRADO SUBMETIDA À BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, SENDO JULGADA ADEQUADA PARA A OBTENÇÃO DO GRAU DE MESTRA EM ENGENHARIA ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

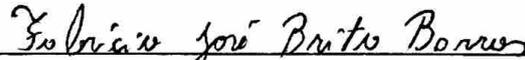
APROVADA EM: 24/01/2019

BANCA EXAMINADORA:



Prof. Dr. Diego Lisboa Cardoso

(Orientador – PPGEE/UFPA)



Prof. Dr. Fabrício José Brito Barros

(Co-Orientador – PPGEE/UFPA)



Prof. Dr. Carlos Renato Lisboa Francês

(Avaliador Interno – PPGEE/UFPA)



Prof. Dr. Edvar da Luz Oliveira

(Avaliador Externo – UFRA)

VISTO:

Prof.ª Dr.ª Maria Emília de Lima Tostes

(Coordenadora do PPGEE/ITEC/UFPA)

Agradecimentos

A Deus, pela dádiva da vida e por me permitir realizar tantos sonhos nesta existência. Obrigado por me permitir errar, aprender e crescer, por sua eterna compreensão e tolerância, e por seu infinito amor.

Aos meus pais, José Renivaldo Lemos Gonçalves e Virgínia da Silva Gonçalves, que não só neste momento, mas em toda minha vida estiveram ao meu lado, apoiando as minhas decisões e torcendo pela minha felicidade.

À minha família, por todo incentivo e apoio durante esses anos, que nunca mediram esforços para me ajudar. Aos meus irmãos: Márcia, Michelly, Silmara, Alvim, Arquimimo, Gleyson e Osair. Em especial, à minha irmã Márcia de Paula Medeiros, que foi uma das peças fundamentais para a minha chegada até aqui.

Ao meu noivo Igor Casemiro, que no decorrer desta jornada sempre me apoiou, incentivou e cuja sua contínua presença foi inestimável para o término desta jornada.

Ao Prof. Dr. Diego Lisboa Cardoso, pela orientação, competência, profissionalismo e dedicação tão importantes. Ao professor Dr. Fabrício José Brito Barros pela coorientação e contribuição ao longo do desenvolvimento desta pesquisa. E a todos os professores que contribuíram para a minha formação.

Agradeço aos amigos que fiz no LPO, por todo companheirismo no decorrer dessa jornada: Daniel, Carlos, Rafael, Rita, Igor, Paulo Henrique, Lucivaldo, Paulo Tássio, Nilze, Ermínio, Áurea, Welton, Ladislav e em especial ao Matheus Leto, que teve contribuição fundamental para a conclusão deste trabalho.

Agradecimentos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro ao longo do programa. Por fim, a todos aqueles que contribuíram direta ou indiretamente, para a realização desta dissertação, os meus sinceros agradecimentos.

“Este trabalho é dedicado às pessoas que sempre estiveram ao meu lado, ao longo de toda a vida: meus pais, José Renivaldo Lemos Gonçalves e Virgínia da Silva Gonçalves.

Se eu vi mais longe, foi por estar sobre ombros de gigantes.
(Isaac Newton)

Sumário

1	Introdução	1
1.1	Motivações	2
1.2	Objetivos	3
1.3	Organização da Dissertação	4
2	Referencial Teórico	5
2.1	Considerações Iniciais	5
2.2	Evolução das Redes Móveis	5
2.3	HetNets	9
2.4	CRAN	10
2.5	H-CRAN	11
2.6	<i>Offloading</i> em Redes Móveis	13
2.7	Inteligência Computacional	15
2.7.1	PSO	16
2.7.1.1	Conceitos Sociais:	16
2.7.1.2	Características Computacionais:	16
2.8	Considerações Finais	19
3	Trabalhos Relacionados	20
3.1	Considerações iniciais	20
3.2	Trabalhos Relacionados	20
3.3	Considerações finais	24
4	Modelo do Sistema, Parâmetros de Rede e Algoritmo TLA	25
4.1	Considerações iniciais	25
4.2	Modelo do Sistema	25
4.3	Parâmetros de Rede	28
4.4	KPIs	30
4.4.1	Indicador de Performance para Usuários Bloqueados (KPI_{BU})	30
4.4.2	Inter-BBU Handovers:	30
4.4.3	Intra-BBU Handovers:	31
4.4.4	Forced Handovers:	31
4.4.5	QoS do Sistema	31
4.5	Ambiente de Simulação	32
4.5.1	Algoritmo TLA	34
4.5.1.1	Alocação	35

4.5.1.2	Balanceamento RRH-BBU	36
4.5.1.3	Decisão de <i>Offloading</i>	37
4.6	Considerações Finais	38
5	Análise dos Resultados	39
5.1	Considerações iniciais	39
5.2	Resultados	39
5.3	Considerações finais	43
6	Conclusões	44
6.1	Trabalhos futuros	45
6.2	Dificuldades Encontradas	46
6.3	Contribuições	46
	Referências	48

Lista de ilustrações

Figura 1	Evolução das Redes Móveis, adaptado de [ZHANG,2014]	6
Figura 2	Arquitetura do Sistema 5G, [RODRIGUEZ 2015]	8
Figura 3	Três principais cenários de serviço de redes celulares 5G, [CHANG et. al., 2018]	9
Figura 4	Arquitetura HetNets, adaptado de [ZHANG, 2014]	10
Figura 5	Arquitetura CRAN , adaptado de [PAIXÃO, 2018]	11
Figura 6	Componentes e Recursos 5G, adaptado de [PENG et. al., 2015]	12
Figura 7	Arquitetura H-CRAN, adaptado de [ZHOU et. al., 2018]	12
Figura 8	Modelo Geral de <i>Offloading</i> , adaptado de [OLTEANU et. al., 2014 <i>apud</i> SILVA,2018]	13
Figura 9	Categorias de <i>offloading</i> , [ZHOU, 2018]	14
Figura 10	Fluxograma Genérico PSO	18
Figura 11	Camadas Redes 5G	26
Figura 12	Arquitetura Proposta	27
Figura 13	Blocos Funcionais da Modelagem do Sistema	33
Figura 14	Demonstração do Cenário	33
Figura 15	Algoritmo TLA	35
Figura 16	Fluxograma do mapeamento de RRHs nas BBUs	36
Figura 17	Fluxograma da decisão de <i>offloading</i>	37
Figura 18	Número de Conexões por tipo de RAN	40
Figura 19	Média de Usuários Conectados	41
Figura 20	Média de Usuários Desconectados	41
Figura 21	Média de Taxa de Dados	42

Lista de tabelas

Tabela 1	Características das principais tecnologias celulares, adaptado de [SILVA, 2018]	7
Tabela 2	Síntese dos Principais Trabalhos Relacionados	23
Tabela 3	Parâmetros de Simulação	34

Lista de abreviaturas e siglas

1G	Primeira Geração Móvel
2G	Segunda Geração Móvel
3G	Terceira Geração Móvel
3GPP	3rd Generation Partnership Project
4G	Quarta Geração Móvel
5G	Quinta Geração Móvel
ANATEL	Agência Nacional de Telecomunicações
BBUs	Baseband Unit
BS	Base Station
CAPEX	Capital Expenditure
CRAN	Cloud Radio Access Network
CDMA	Code Division Multiple Access
CDI	Cell Differentiation and Integration
D2D	Device to Device
DL	Downlink
EE	Energy Efficiency
EDGE	Enhanced Data Rates For GSM Evolution
EIRP	Effective Isotropic Radiated Power
eMBB	Enhanced mobile broadband
FDMA	Frequency Division Multiple Access
FIFO	First in, first out
GPRS	General Packet Radio Services

GSM	Global System for Mobile Communication
HC	Hard Capacity
H-CRAN	Heterogeneous Cloud Radio Access Network
HetNets	Heterogeneous Networks
HPN	High Power Node
H-WDRF	Heterogeneous Weighted Dominant Resource Fairness
IC	Inteligência Computacional
IA	Inteligência Artificial
KPIs	Key Performance Indicator
LPNs	Low Power Nodes
LTE	Long Term Evolution
MATLAB	MATrix LABoratory
mMTC	Massive machine-type communication
MBSs	Macro cell base stations
OPEX	Operational Expenditure
OFDMA	Orthogonal Frequency Division Multiple Access
PSO	Optimization Particle Swarm
QoS	Quality of Service
RAT	Radio Access Technology
RSRP	Reference Signal Received Power
RRH	Remote Radio Head
RSL	Received Signal Level
SC-FDMA	Single Carrier Frequency Division Multiple Access
SE	Spectral Efficiency
SCNs	Small Cell Networks
SON	Self-Optimising Network

SOCRAN	Self-optimized Cloud Radio Access Network
SINR	Signal-to-interference-plus-noise ratio
SNR	signal-to-noise ratio
SDN	Software Defined Network
SUI	Stanford University Interim
TDMA	Time Division Multiple Access
TLA	Two-Level Allocation
UBs	User Bases
UE	User equipment
URLLC	Ultra reliable and low latency communication
V2V	vehicle-to-vehicle
V2I	vehicle-to-infrastructure
Wifi	Wireless Fidelity

Resumo

O crescimento acelerado de dados e aplicativos representa desafios significativos para a próxima geração de redes móveis (5G). Dentre eles, destaca-se a necessidade de uma coexistência de novos e antigos padrões durante a transição das arquiteturas. Assim, na coexistência de arquiteturas, o processo de *offloading* entre diferentes arquiteturas é essencial. Aliado a isso, processos de *offloading* tradicionais, que focam apenas nos recursos de rádio, podem ser ineficazes, uma vez que o processo dinâmico de alocação de RRH-BBU influencia no resultado. Desta forma, este trabalho investigou soluções para descarregamento em uma arquitetura híbrida, também conhecida como H-CRAN (*Heterogeneous Cloud Radio Access Network Architecture*), que centraliza o processamento e busca um melhor uso dos recursos da rede. Através de um algoritmo evolutivo, buscou-se encontrar uma solução subótima para a alocação de dois níveis (TLA) na arquitetura H-CRAN e outra baseada no FIFO (*First In, First Out*), para fins de *benchmarking*. A média SNR (*Noise Interference Signal*), a máxima taxa de bits, o número de usuários com ou sem conexões e o número de conexões em RRHs e macro foram usados como medidas de desempenho. Através dos resultados, percebeu-se uma melhora de aproximadamente 60% na Taxa Máxima de Bits, quando comparada à abordagem tradicional, possibilitando um melhor atendimento aos usuários.

Palavras-chave: Redes Móveis, 5G, H-CRAN, QoS, Offloading, TLA.

Abstract

The accelerated data and apps growth represents significant challenges to the next generation of mobile networks. Amongst them, it is highlighted the necessity for a co-existence of new and old patterns during the transition of architectures. Thus, this paper has investigated solutions for offloading into a hybrid architecture, also known as H-CRAN (Heterogeneous Cloud Radio Access Network Architecture), that centralizes processing and searches a better use of the network resources. The strategy of optimization was analyzed through the evolutive algorithm PSO (Particle Swarm Optimization), in order to find a suboptimal solution to the allocation of two levels (TLA) in the H-CRAN architecture and another one based on FIFO (First In, First Out), for benchmarking purposes. SNR (Noise Interference Signal) average, Maximum Bit Rate, the number of users with or without connections and number of connections in RRHs and macro were used as performance measurements. Through the results, it was noticed an improvement of approximately 60% in the Maximum Bit Rate when compared to the traditional approach, enabling a better service to the users.

Keywords: Mobile Networks, 5G, H-CRAN, QoS, Offloading, TLA.

1 Introdução

As comunicações móveis estão em um constante processo de transformação, que se pode constatar na evolução desde a primeira geração de redes móveis (1G), que transportava para milhões de usuários apenas voz, passando para uma prospecção para as futuras redes móveis (5G), que irá suportar trilhões de dispositivos transmitindo dados e voz. O aumento do volume de serviços de dados e aplicações, juntamente com o crescimento acelerado das demandas de acesso sem fio representam desafios significativos para a redes móveis 5G.

O objetivo principal desta futura rede móvel se concentra nas técnicas e mecanismos de habilitação para garantir que várias aplicações sem fio possam ser realizadas com garantias distintas de QoS (*Quality of Service*), com limite de atraso para uma ampla gama de serviços multimídia, aplicativos e usuários com diferentes requisitos. Espera-se que as redes suportem não apenas aplicativos com uso intenso de dados, como vídeo de alta definição e realidade virtual, mas também outras aplicações diversas que exigem latência rigorosa e desempenho de confiabilidade [ZHANG et. al., 2018a].

De acordo com [CISCO, 2017], o tráfego mensal global de dados móveis até 2021 será de 49 exabytes e deverá fornecer cerca de 1000 vezes maior capacidade, aliada a uma economia de até 90% no consumo de energia. Com isso, as operadoras de telefonia móvel estão investindo cada vez mais em novas tecnologias, a fim de aumentar a capacidade da rede, mitigar a interferência e melhorar a QoS dos usuários finais.

Para melhorar a taxa de dados, capacidade e cobertura, a prática comum é a densificação da rede com pequenas células ou estações base, a fim de maior reutilização de largura de banda. HetNets (*Heterogeneous Networks*) foram implantadas visando aumentar a capacidade das redes celulares em áreas densas e com altas demandas de tráfego, preenchendo lacunas existentes na área de cobertura e descarregando o tráfego de dados das MBSs (*Macro Base Station*) [DUAN et. al.,2015].

No entanto, uma rede densa ¹ traz maiores despesas de Capital - CAPEX (*Capital Expenditure*) e Operacionais - OPEX (*Operational Expenditure*), por conta da implantação e manutenções adicionais na rede. Além disso, torna-se complexo planejar e gerenciar multicamadas, ocasionando um aumento na interferência entre as células e coordenação no envio do sinal [RAZA et. al., 2016]. Em contrapartida a HetNets, surgiu a arquitetura CRAN (*Cloud Radio Access Network*) centralizada e que se propõe a ser escalável e flexível,

¹Densidade é a relação existente entre o total de usuários e a área geográfica da rede.

buscando solucionar os diversos problemas existentes na arquitetura anterior. Assim, visa-se mitigar problemas das arquiteturas descentralizadas, tais como ineficiência na utilização dos recursos, principalmente em cenários de altas cargas de tráfego [WU et. al.,2015].

Nesta nova arquitetura, que move o processamento do sinal primário, funções desempenhadas nas BBUs (*Base Band Unit*) dos eNodesBs típicos para uma localização centralizada, denominada BBU-Cloud. E as RRHs (*Remote Radio Head*), com menos funções, mantém o acesso de rádio nos sites da célula, possibilitando uma implantação densa com custo mínimo [WU et. al.,2015]. Entretanto, a arquitetura CRAN ainda possui muitos desafios, dentre eles destaca-se o rigoroso controle da latência, do *jitter* e o alto custo de infraestrutura, principalmente no que concerne à implantação do *fronthaul*, que devido às limitações de latência, precisa ser implantado com tecnologias de alta capacidade.

Esse processo de migração de arquiteturas tradicionalmente não é realizado de forma brusca. Quando uma rede precisa ser atualizada, quase todo o equipamento de rede deve ser substituído e, por isso, existe a necessidade de uma fase de transição, a fim de satisfazer a coexistência de padrões novos antigos. Dessa forma, surge o H-CRAN, que é uma arquitetura onde ambos HetNets (descentralizada) e CRAN (centralizada) são combinados para permitir a implantação de redes densas e heterogêneas, visando centralizar o processamento de carga de trabalho e realizar as funcionalidades de processamento de sinal cooperativo em larga escala e as funcionalidades de rede e, assim, o desempenho de SE (*Spectral Efficiency*) e EE (*Energy Efficiency*) é substancialmente melhorado [MAROTTA et. al.,2017].

Como resultado, H-CRAN foi reconhecido como uma arquitetura de rede promissora para redes 5G e tem ganhado bastante atenção da indústria e da academia. Apesar disso, há uma série de desafios a serem enfrentados, incluindo a alocação otimizada sobre a fronteira restrita, o consumo de energia, o desenvolvimento de padrões compatíveis com versões anteriores, altos eventos de bloqueio de chamadas e a QoS [ZHANG et. al.,2018].

1.1 Motivações

Um dos maiores desafios presentes em redes móveis é lidar com a alta demanda que está por vir, propondo uma melhor QoS a todos os usuários finais. O *offloading* em redes móveis é amplamente utilizado, no intuito de desafogar pontos críticos na rede, evitando assim possíveis congestionamentos.

Em diversos trabalhos foi estudado o desempenho de *offloading* em redes móveis, o qual, em suma, mostrou-se eficiente em inúmeras arquiteturas. Ainda não há uma arquitetura fechada para 5G, e a comunidade científica vem estudando arquiteturas que possam lidar com esta alta demanda, que ofereçam ao usuário um serviço de conexão e

cobertura de qualidade, com múltiplos serviços, flexível e que tenham baixos custos de CAPEX/OPEX. Nesse viés, a arquitetura H-CRAN tem se mostrando eficiente, por solucionar muitos dos problemas existentes em diversos cenários. No entanto, ainda possui várias lacunas, como a alocação otimizada sobre a fronteira restrita, o consumo de energia, o desenvolvimento de padrões compatíveis com versões anteriores, altos eventos de bloqueio de chamadas e manter altos níveis de QoS [ZHANG et. al., 2018].

Porém, no levantamento realizado pelos autores, não foi encontrado nenhum trabalho que investigue soluções de *offloading*, visando o balanceamento de carga da rede² e melhoria de QoS, em uma arquitetura H-CRAN.

A principal contribuição deste trabalho é a proposição de um novo esquema de *offloading* para arquiteturas H-CRAN, focando no balanceamento de carga nas BBUs, melhor utilização dos recursos da rede, maximização da capacidade, melhoria da cobertura e maximização da QoS dos usuários finais. Dessa forma, promovendo a heterogeneidade e, conseqüentemente, melhorando a eficiência do espectro.

1.2 Objetivos

Nesta dissertação, como objetivo primordial, têm-se o desenvolvimento de uma arquitetura de *offloading* em redes móveis H-CRAN, considerando aspectos de rádio e BBU, aliado ao balanceamento de carga da rede, buscando maximização de carga. O balanceamento de carga da rede e o mapeamento dinâmico BBU-RRH é analisado, buscando oferecer uma melhor QoS a usuários finais. Como objetivos específicos, destacam-se:

- Técnica de *offloading* eficiente, que considera recursos de rádio e BBU em conjunto, para alocação otimizada de usuários com melhor aproveitamento de recursos em uma arquitetura H-CRAN;
- Estudo de algoritmos de *offloading* aplicados a arquiteturas híbridas;
- Implementação de testes para validação da proposta em estudos de caso;
- Validação da proposta e eventuais publicações a cerca do tema proposto;

²Balanceamento de Carga é um método eficaz para equilibrar e aliviar o congestionamento em redes móveis.

1.3 Organização da Dissertação

Este documento está dividido como segue:

- Capítulo 2: Neste capítulo são mostrados os conceitos, as tecnologias e apresentação do estado da arte, necessários para a compreensão do estudo realizado. O capítulo é iniciado com um breve histórico da evolução das redes móveis. Em seguida, apresentam-se as principais arquiteturas que norteiam o entendimento da arquitetura e a técnica de *offloading* usada nesta dissertação;
- Capítulo 3: São mostrados os trabalhos relacionados à problemática desta dissertação, abordando principalmente as soluções para balanceamento de carga em redes centralizadas e as lacunas em aberto quanto à temática.
- Capítulo 4: Neste capítulo são apresentados todos os parâmetros de rede, o modelo do sistema, o algoritmo proposto e as métricas utilizadas para avaliação de desempenho abordado neste trabalho.
- Capítulo 5: Apresenta-se a discussão dos resultados da referida dissertação.
- Capítulo 6: A dissertação é finalizada com as considerações finais, apontando as principais contribuições do estudo e as possíveis incrementações em trabalhos futuros.

2 Referencial Teórico

2.1 Considerações Iniciais

Neste capítulo, é feita uma descrição de todas as tecnologias que fundamentaram este trabalho. É iniciado com a evolução das redes móveis, seguido por conceituação das arquiteturas HetNets, CRAN e H-CRAN, respectivamente. Já no penúltimo subtópico, é abordado o *Offloading* em Redes Móveis. O capítulo é finalizado descrevendo a técnica de Inteligência Computacional utilizada neste trabalho.

2.2 Evolução das Redes Móveis

A ANATEL (Agência Nacional de Telecomunicações) informou que existem cerca de 234,3 milhões de celulares, com densidade de 11,84 cel/habitantes na rede móvel, celulares ativos na operadora. Tal densidade, foi calculada com a projeção de população do IBGE (Rev. 2013) para o mês respectivo [ANATEL,2018]. De acordo com a [CISCO, 2017], até 2021 a nova geração de rede móvel 5G será responsável por 1,5% do tráfego móvel global, e o 2G será responsável por apenas 0,6%. Para dar conta desta demanda, as redes de telefonia móvel estão investindo em tecnologias de ponta para transmitir a melhor qualidade do sinal para usuários finais.

A evolução das redes móveis e as respectivas tecnologias existentes nelas é mostrada na Fig. 1, no qual pode-se observar que a medida que as gerações foram progredindo, novas tecnologias surgiram para suprir diversos problemas ocasionados pela alta demanda e limitações da rede.

A primeira geração de redes móveis (1G) foi introduzida durante os anos 80, com padrões de telecomunicações analógicos. Usava FDMA (*Frequency Division Multiple Access*), que é uma técnica em qual a frequência alocada para comunicação é dividida por vários canais e cada canal é capaz de transportar um sinal de dados. Com capacidade de canal de 30 kHz e faixa de frequência de 824-894 MHz, foi baseado em uma tecnologia conhecida como Advance Mobile Phone Service AMPS (*Advanced Mobile Telephone System*)[BABA et. al., 2018].

A segunda geração de redes móveis (2G) foi dividida em CDMA (*Code Division Multiple Access*) e GSM (*Global System for Mobile Communication*), nos EUA e Europa, respectivamente. CDMA é um método de acesso de canal usado por várias tecnologias

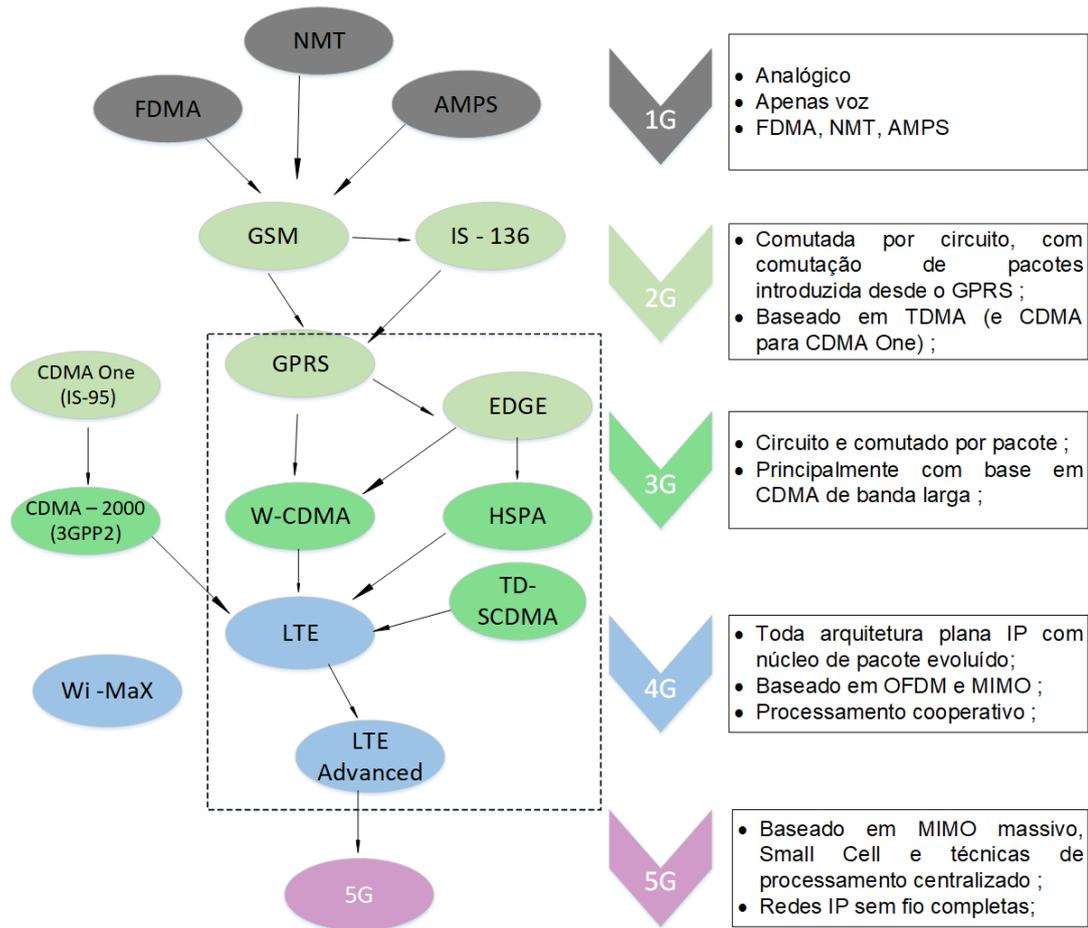


Figura 1 – Evolução das Redes Móveis, adaptado de [ZHANG,2014]

de comunicação de rádio, onde vários transmissores podem enviar informações simultaneamente, através de um único canal de comunicação. Já GSM, utiliza uma variação de acesso múltiplo por divisão de tempo TDMA (*Time Division Multiple Access*), digitaliza e comprime os dados, enviando-os por um canal com dois outros fluxos de dados do UE, cada um em seu próprio intervalo de tempo[VORA et. al.,2015].

Já nas redes móveis 3G, tecnologias RAT (*Radio Access Technology*) foram utilizadas, como o WCDMA (*Wide-Band Code-Division Multiple Access*), buscando uma comunicação de dados em banda larga na telefonia móvel, reaproveitando a mesma rede de suporte GPRS (*General Packet Radio Services*) e EDGE (*Enhanced Data Rates For GSM Evolution*), facilitando a implementação, diferindo apenas pelos protocolos e interface aérea entre os sistemas.

Na arquitetura 4G / LTE (*Long Term Evolution*) subsequente, voz e dados são baseados em pacotes IP, e a interface aérea desta arquitetura possui uma variabilidade de canais de banda larga, variando entre 1,4 MHz - 20 MHz, utilizando técnicas de OFDMA (*Orthogonal Frequency Division Multiple Access*) e SC-FDMA (*Single Carrier Frequency Division Multiple Access*) para downlink e uplink, respectivamente[AGIWAL ET. AL.,2016].

Tabela 1 – Características das principais tecnologias celulares, adaptado de [SILVA, 2018]

Geração	2G (1991-2000)			3G (2001-2008)				4G (2009-IMT)		5G (IMT - 2020)
Tecnologia	GSM	GPRS	EDGE	WCDMA	HSPA	HSPA +	LTE (Rel.8)	LTE-Advanced	LTE-Advanced Pro (4.5 G)	-
Vazão Teórica (downlink)	14.4 kbps	114 kbps	384 kbps	2 Mbps	7.2/14.4 Mbps	21/42 Mbps	100 Mbps	1 Gbps	3 Gbps	20 Gbps
Vazão Teórica (uplink)	-	-	473.6 Kbps	474 Kbps	5.76 Mbps	7.2/11.5 Mbps	50 Mbps	0.5 Gbps	1.5 Gbps	10 Gbps
Tamanho do Canal (MHz)	0.2	0.2	0.2	5	5	5	20	100	610	Até 1000
Latência (ms)	500	500	300	250	~70	~30	~10	<5	<2	<1
Frequência	1.8 GHz (digital)			1.6 - 2.0 GHz				2 - 8 GHz		3 - 300 GHz

Após toda essa evolução e com altas demandas esperadas nos próximos anos, a rede 5G vêm sendo planejada, que vai além do aumento da vazão, pois buscam também reduzir significativamente a latência de comunicação, aumentar o alcance, a capilaridade e o número de usuários da rede. Com a junção dos sistemas 2G (GSM), 3G UMTS (*Universal Mobile Telecommunications System*) e 4G (LTE / WiMax), é esperada uma taxa de dados máxima alcançável de 20 Gbit /s, com 1 Gbit / s a taxa de dados alcançável em toda área de cobertura [SERIES, 2015], visando superar diversos problemas existentes e prevenir outros que possam vir.

Toda essa evolução, bem como as características e demandas de capacidade de cada tecnologia celular, pode ser consultada na Tabela 1.

Há um emaranhado de proposições de tecnologias para 5G, que precisarão coexistir e partilhar recursos da rede. A proposta da nova arquitetura de rede 5G, que irá suportar uma ampla gama de aplicações de voz, vídeo em tempo real, V2V (*vehicle-to-vehicle*), V2I (*vehicle-to-infrastructure*), entre outros. O desempenho da rede em relação a cobertura, capacidade e EE será amplamente reforçado, com o objetivo de aumentar cerca de mil vezes a capacidade da rede, como pode ser visto na Fig. 2.

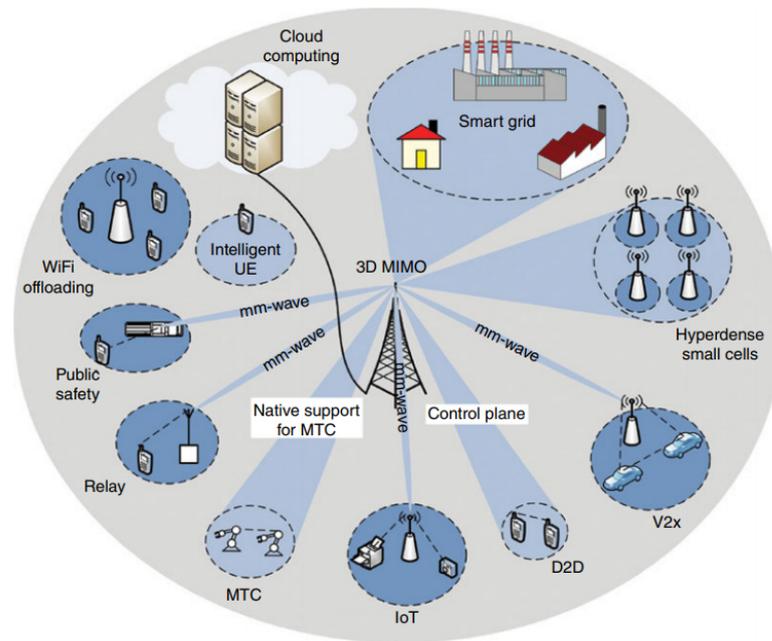


Figura 2 – Arquitetura do Sistema 5G, [RODRIGUEZ 2015]

Em um projeto de pesquisa de pré-padroneização do 5G, o projeto METIS identificou três categorias de cenários de serviço de redes celulares 5G, exemplificada na figura abaixo [CHANG et. al., 2018]:

1. eMBB (*enhanced mobile broadband*) ;
2. URLLC (*ultra reliable and low latency communication*) ;
3. mMTC (*massive machine-type communication*) ;

1 - Abrange os serviços que exigem alta taxa de dados, como serviços de vídeo de alta definição e serviços de realidade aumentada.

2- Inclui principalmente os serviços que exigem alta confiabilidade e latência de comunicação rigorosa, como a comunicação V2V e V2I.

3 - Consiste principalmente nos serviços em que um grande número de dispositivos precisa estar conectado via redes celulares.

Esses diferentes cenários de serviço têm diferentes requisitos de desempenho nas redes móveis celulares. Para suportar esses serviços heterogêneos, as redes precisam não apenas superar os sistemas legados em termos de taxa de dados suportada, mas também obter grande redução da latência de serviço de ponta a ponta e maior confiabilidade. Enquanto isso, as redes celulares 5G precisam fornecer conectividade a dispositivos com densidade.

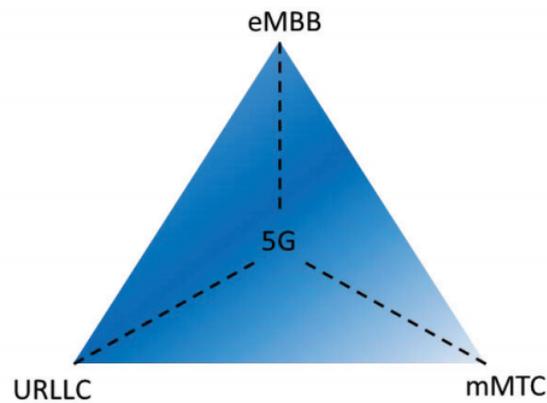


Figura 3 – Três principais cenários de serviço de redes celulares 5G, [CHANG et. al., 2018]

2.3 HetNets

Decorre do uso de vários tipos de nós de acesso em uma rede sem fio (por exemplo, *macrocells*, *microcells*, *smallcells*, etc.) e várias tecnologias de acesso por rádio (GSM, 3G, LTE, Wifi (*Wireless Fidelity*), por exemplo), buscando oferecer uma ampla cobertura de rede sem fio em ambientes que variam desde escritórios, residências, até a áreas subterrâneas [RAMAZANALI et. al., 2016].

Esta arquitetura foi considerada uma das soluções mais interessantes para o iminente problema de tráfego de dados móveis no contexto de LTE-A e 5G existentes. A sobreposição das células pequenas (pico, femto) com as células macro existentes, melhora a capacidade da rede e estende a cobertura, ocasionando uma reutilização eficiente de frequência. A arquitetura da rede HetNets pode ser vista na Fig. 4.

Em HetNets é feita uma expansão de faixa de células através de polarização de *handover* e particionamento de recursos entre diferentes classes de poder de nó. Isso permite o balanceamento de carga, onde, dependendo do valor de polarização, a rede pode controlar o número de unidades UE (*User equipment*) associadas aos nós de baixa potência e, portanto, controlar a demanda de tráfego nesses nós. O particionamento de recursos, que também pode ser adaptativo, permite a configuração e o ajuste de recursos protegidos por interferência, permitindo que o UE em uma área expandida de célula receba dados [CLAUSSEN, 2017].

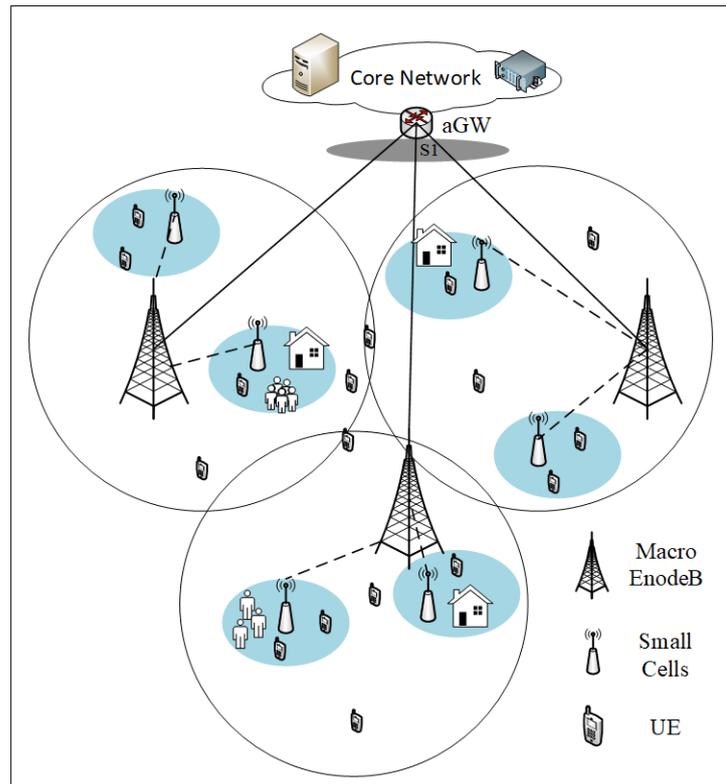


Figura 4 – Arquitetura HetNets, adaptado de [ZHANG, 2014]

2.4 CRAN

A busca incessante por novas tecnologias emergentes vêm sendo feita pela comunidade científica, visando solucionar problemas existentes nas redes móveis. Em 2010, na China Mobile Research, foram feitos testes e protótipos de CRAN que, desde então, tem sido uma das promessas para 5G. Seu nome vem dos quatro "C's" nas principais características do sistema CRAN, (*Clean, Centralized processing, Collaborative radio, and a real-time Cloud Radio Access Network*) [RAZA et. al., 2016].

Esta arquitetura de rede centralizada, segundo [MAROTTA et. al., 2017], é uma evolução da rede celular distribuída, onde os recursos computacionais do BS (*Base Station*) são alocados em um local central, chamado BBU Cloud.

Como pode ser visto na Fig. 5, uma arquitetura típica de CRAN consiste principalmente em três partes diferentes:

- RRHs: são unidades de rádio distribuídas com antenas localizadas em um site remoto;
- BBU: composta por equipamentos programáveis de alto desempenho, tecnologias de virtualização de rede e tecnologias RRM (*Radio resource management*);
- A rede de transmissão que liga os RRHs e a BBU Pool com alta velocidade de banda

e baixa latência (*fronthaul*).

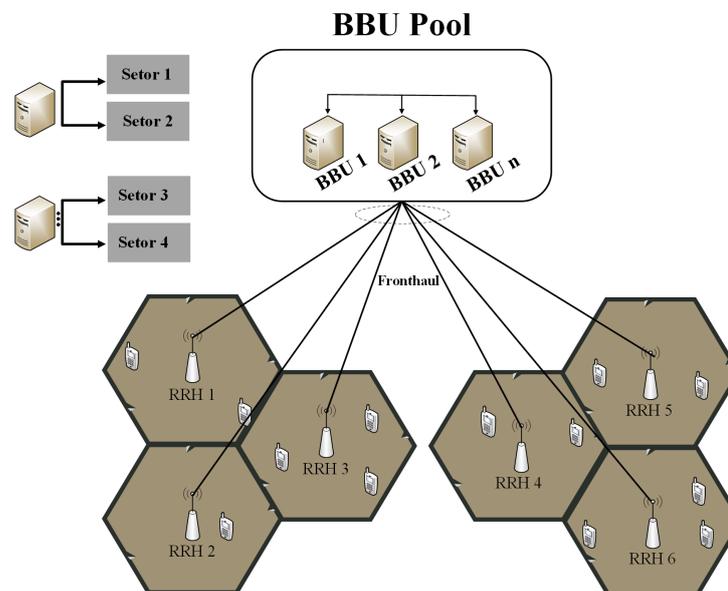


Figura 5 – Arquitetura CRAN , adaptado de [PAIXÃO, 2018]

Possui as seguintes características em sua arquitetura: gerenciamento centralizado de recursos de computação, reconfigurabilidade dos recursos do espectro, comunicações e computação em nuvem em tempo real em plataformas genéricas.

2.5 H-CRAN

Esta arquitetura é considerada promissora nas redes sem fio de quinta geração (5G). Combina atributos de HetNets e CRAN, permitindo a implantação de redes densas e heterogêneas, visando centralizar o processamento de carga de trabalho e realizar as funcionalidades de processamento de sinal cooperativo em larga escala e gerenciar todos os recursos da rede.

O sistema 5G baseado nessa arquitetura possuem três planos: o plano de usuário (U), o plano de controle (C) e o plano de gerenciamento (M), cada um processando diferentes áreas de operações e funções [PENG et. al., 2015], como pode ser visto na Fig. 6.

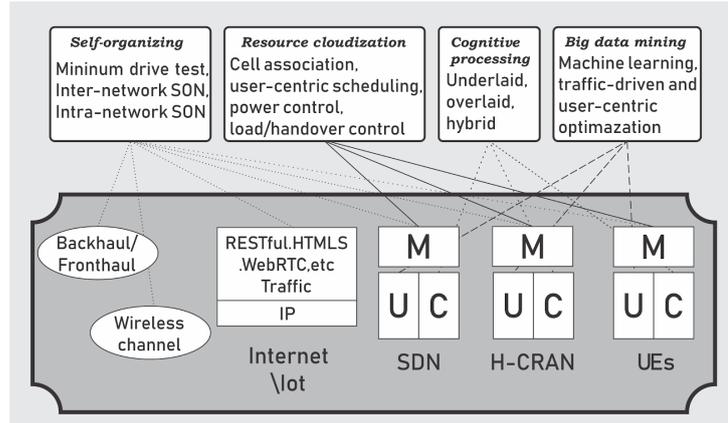


Figura 6 – Componentes e Recursos 5G, adaptado de [PENG et. al., 2015]

De modo geral, o plano (U) transfere o tráfego real do usuário e executa o processamento de tráfego relacionado, para satisfazer os vários requisitos de QoS. O plano (C) transporta a sinalização de controle e é encarregado da alocação de recursos e do processamento do tráfego, visando melhorar SE e EE. Já o plano (M) tem como função executar a administração e operação, e é responsável por adicionar, excluir, atualizar e modificar as interações com os outros planos [PENG et. al., 2015].

Essa arquitetura permite que os usuários disponha de diversos serviços com alta eficiência energética, alta eficiência espectral e operação de baixo custo, que são alcançados usando técnicas de virtualização e computação em nuvem [ZHOU et. al., 2018]. Trata-se de uma solução centralizada, eficiente e escalável, para gerenciamento e distribuição de informações na rede, como pode ser visto na Fig. 7.

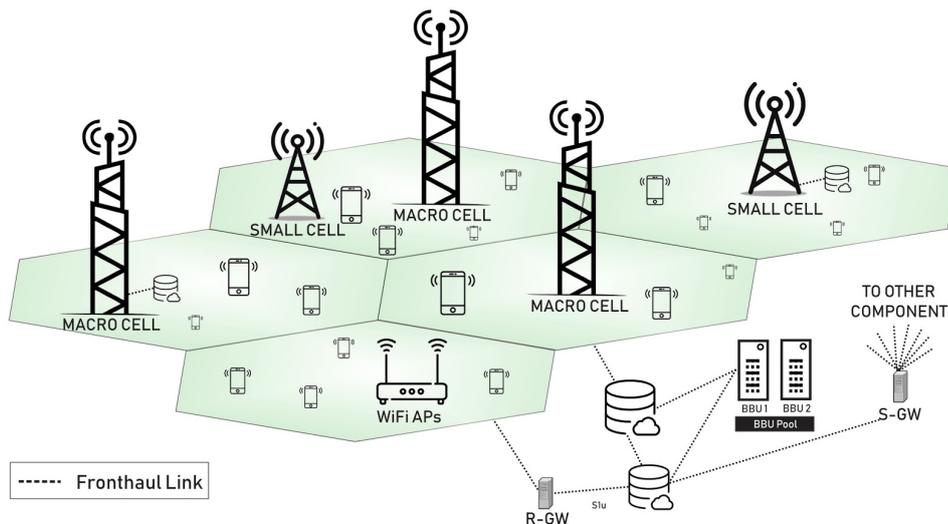


Figura 7 – Arquitetura H-CRAN, adaptado de [ZHOU et. al., 2018]

A nuvem centralizada no núcleo 5G consiste em poderosos nós computacionais de alto desempenho que fornecem acesso de rede onipresente, abrangente, conveniente e sob demanda a um conjunto compartilhado de recursos de computação configuráveis,

como redes, servidores, armazenamento, aplicativos e serviços que são provisionados e liberados rapidamente com esforço mínimo de gerenciamento ou interação com o provedor de serviços [LEE et. al., 2018].

2.6 Offloading em Redes Móveis

O *offloading* ou descarregamento de dados é a técnica de usar tecnologias de redes complementares para desafogar pontos críticos na rede, visando aumentar o desempenho e reduzir o consumo de energia de dispositivos móveis, por meio do descarregamento do processamento de dados para outras infraestruturas, com maior poder computacional e armazenamento. Vem sendo estudado desde o 3G e perdura como uma das propostas para o 5G que buscam reduzir a quantidade de dados transportados nas bandas celulares, liberando largura de banda para outros usuários na rede, evitando possíveis congestionamentos, permitindo assim uma melhor conectividade [ZHANG et. al., 2018b].

Em [SILVA et. al., 2018] é apresentado um levantamento abrangente de técnicas de *offloading* em redes celulares, com o uso em conjunto de nós de baixa potência (*picocell*, *femtocell*, *microcell*). A figura abaixo mostra o modelo geral de *offloading* em redes móveis:

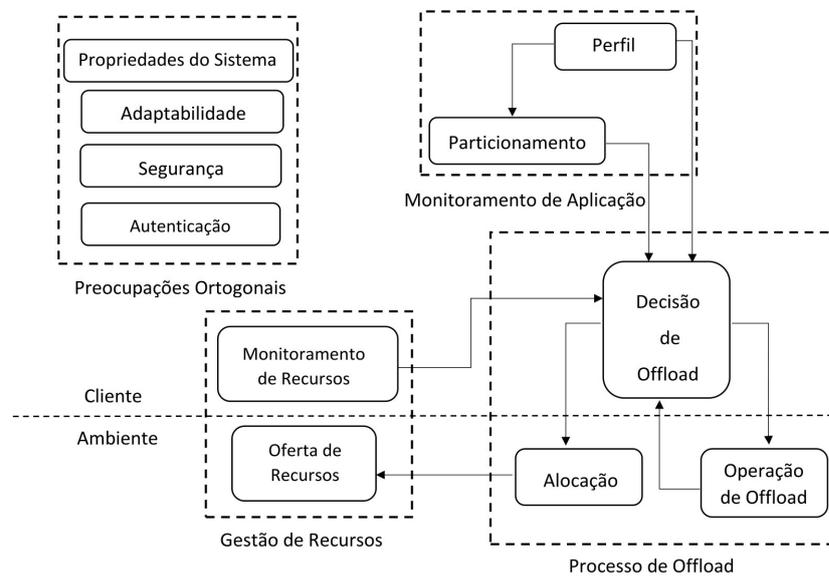


Figura 8 – Modelo Geral de *Offloading*, adaptado de [OLTEANU et. al., 2014 apud SILVA,2018]

De acordo com a Fig. 8, pode-se observar que a tomada de decisão para *offloading* depende de diversos fatores da rede, os quais devem ser respeitados para obter uma alta performance. Em [ZHOU, 2018], é feita uma revisão das principais estratégias de *offloading* e é fornecida uma categorização abrangente das tecnologias de descarregamento de dados móveis existentes. O autor divide em quatro categorias:

- Descarregamento de dados através de SCNs (*Small Cell Networks*);
- Descarregamento de dados através de redes WiFi;
- Descarregamento de dados através de redes móveis oportunistas ;
- Transferência de dados através do HetNets (ou descarregamento de dados híbridos);

Como mostra a Fig. 9 a seguir:

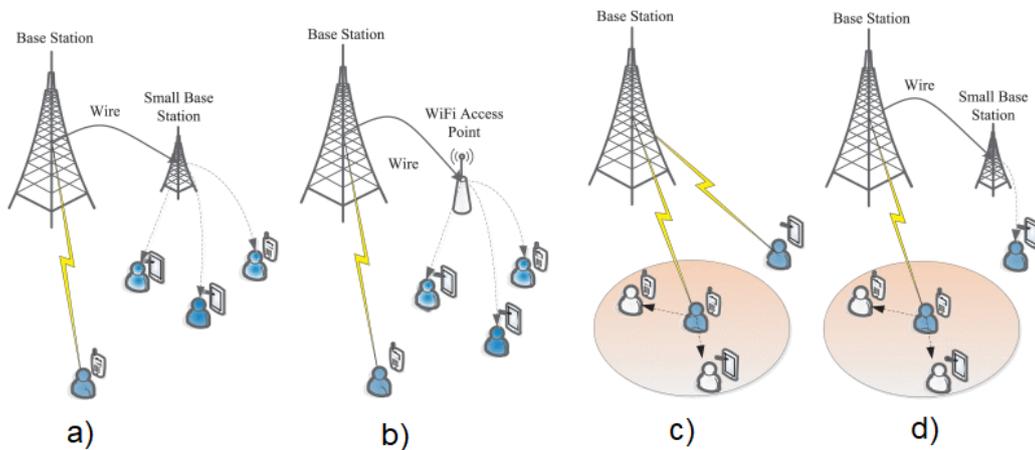


Figura 9 – Categorias de *offloading*, [ZHOU, 2018]

- Nesta categoria, há uma variedade de SBSs (*small-cell base stations*) de baixo consumo de energia para descarregar o tráfego em SCNs. Estes também podem ser realizados com a tecnologia de rádio distribuída, utilizando BBUs e RRHs.
- O tráfego de dados das redes celulares tradicionais é descarregado para redes WiFi, objetivando reduzir custos e carga de tráfego de redes celulares quando os usuários móveis entram em áreas com cobertura WiFi, por oferecer altas taxas de transmissão de dados.
- No descarregamento de dados através de redes móveis oportunistas, os dispositivos móveis vizinhos podem se comunicar sem usar a infraestrutura de rede e a proximidade de dispositivos móveis pode fornecer altas taxas de dados com um baixo consumo de energia. A maioria destes dados é proveniente de provedores de serviços de conteúdo, incluindo condições de tráfego, música, multimídia, entre outros, buscando economizar o custo monetário. Isso, por sua vez, é um dos fatores mais atraentes nestas redes.
- Redes multicamadas heterogêneas com pequenas estações base de baixa potência (SBSs), bem como comunicações oportunistas subjacentes a MBSs (*macro cell base stations*), buscando aumentar consideravelmente a capacidade das redes celulares.

2.7 Inteligência Computacional

A busca por soluções em uma grande variedade de problemas, das mais diversas áreas de conhecimento, fez com que pesquisadores apresentassem estratégias que tentam emular alguns comportamentos observados em sistemas biológicos naturais. A Inteligência Computacional (IC) é um objeto de estudo do design de agentes inteligentes que, por sua vez, é um sistema que age inteligentemente, buscando modelar sistemas complexos do mundo real, através do comportamento de muitos sistemas biológicos, incluindo seres humanos [KENNEDY et. al., 2011].

Em subsequente, e tendo como principal objetivo entender o que torna a inteligência possível, surgiu a Inteligência Artificial (IA), na década de 50. Da mesma forma que podem ser feitas soluções através de comportamentos biológicos (como Redes Neurais e os Algoritmos Genéticos), existem casos em que tentam se aproximar a comportamentos observados em sistemas naturais (por exemplo, o raciocínio aproximado dos sistemas nebulosos), percebendo seu ambiente e realizando ações que maximizam sua chance de atingir seus objetivos com sucesso [RUSSELL et. al., 2016].

A área de IA pode ser considerada como composta por duas abordagens principais (RUSSEL; NORVIG, 2003): A simbólica (ou clássica) e a conexionista (redes neurais artificiais).

A abordagem simbólica é predominante na IA, onde a manipulação de símbolos discretos é muito eficiente, permitindo uma maior facilidade de interação o usuário. É uma abordagem *top-down*, onde o conhecimento é introduzido explicitamente no sistema, que faz com que os sistemas simbólicos encontrem dificuldade para lidar em ambientes reais (contínuos) desconhecidos.

Já a conexionista, adota uma abordagem *bottom-up* baseada no ajuste de conexões entre processadores elementares ligados em rede, os neurônios artificiais. A semântica do domínio não precisa ser introduzida explicitamente no sistema, que pode induzir este conhecimento, através de um processo de aprendizagem. Entretanto, a maioria das técnicas atuais não permite aprender eficientemente em ambientes complexos. Além disso, o conhecimento aprendido não se torna facilmente interpretável pelo usuário.

Neste trabalho, dentre as diversas técnicas disponíveis, utilizou-se a meta-heurística, conhecida como PSO (*Optimization Particle Swarm*), que utiliza estruturas primitivas e operadores matemáticos sem grande custo computacional, a qual será melhor discutida abaixo.

2.7.1 PSO

O PSO foi proposto por James Kennedy e Russell Eberhart em 1995. Trata-se de uma metaheurística inspirada no comportamento social e cooperativo exibido por várias espécies (como por exemplo, o movimento de cada indivíduo dentro de um bando de pássaros ou de um cardume de peixes), por meio do qual os comportamentos coletivos de agentes não sofisticados, que estão interagindo localmente com seu ambiente, criam padrões funcionais. Ele é baseado por duas disciplinas fundamentais: Ciências Sociais e Ciência da computação, e, segundo [DEL VALLE, 2008], pode ser descrito da seguinte maneira:

2.7.1.1 Conceitos Sociais:

A inteligência humana é resultado da interação social, através da avaliação, comparação e imitação dos outros, bem como aprender com a experiência alheia, o que permite a adaptação dos seres humanos ao ambiente e determina padrões de comportamento, atitudes e correlatos.

2.7.1.2 Características Computacionais:

O termo “partículas” se refere a membros da população. São sujeitos a velocidades e acelerações para um melhor modo de comportamento e possuem os seguintes atributos computacionais:

- a) Partículas individuais (células) atualizadas em paralelo.
- b) A cada novo valor, depende do valor anterior da partícula e sua vizinhança.
- c) Todas as atualizações são realizadas de acordo com as mesmas regras.

O PSO inicia o processo em busca das melhores soluções candidatas, as partículas avaliam o custo da solução candidata e lembram-se qual a melhor solução onde estiveram, conhecida como mínimo local ou mínimo da partícula, no qual essa informação também é dada a seus vizinhos. O mínimo global também é conhecido de todas as partículas, no qual os movimentos, através do espaço de pesquisa, são guiados por estas informações [ZHANG et. al., 2015].

A posição de cada partícula é determinada pelo vetor; e seu movimento, pela velocidade da partícula, que pode ser calculada de acordo com a fórmula abaixo:

$$\vec{x}_i(t) = \vec{x}_i(t-1) + \vec{v}_i(t) \quad (2.1)$$

Vale ressaltar que quanto mais longe, maior o valor dessa velocidade. A informação disponível para cada indivíduo é baseada nas experiências até o momento, (e o sucesso em cada decisão) e no conhecimento do desempenho dos indivíduos em sua vizinhança. São

comumente atribuídos pesos aleatórios a cada parte, portanto a velocidade é reformulada e determinada por:

$$\vec{v}_i(t) = \vec{v}_i(t-1) + \varphi_1 \cdot rand_1 \cdot (\vec{p}_i - \vec{x}_i(t-1)) \dots$$

$$\varphi_2 \cdot rand_2 \cdot (\vec{p}_g - \vec{x}_i(t-1)) \quad (2.2)$$

onde, 1 e 2 são dois números positivos e aleatórios com distribuição uniforme, na faixa de $[0,0, 1,0]$.

Em (2.2), há três componentes principais para a atualização da velocidade:

1. Por vezes referido como “inércia” e/ou “Momentum”, modela a tendência da partícula de continuar na direção que tem viajado;
2. O pbest (melhor local) de cada partícula indica o quão próximo que os dados já chegaram ao destino;
3. E o gbest (melhor global), só muda quando o valor do pbest de qualquer partícula se aproxima da meta, que é o gbest. A cada iteração do algoritmo, o gbest se aproxima do alvo, até que uma das partículas atinja o objetivo final.

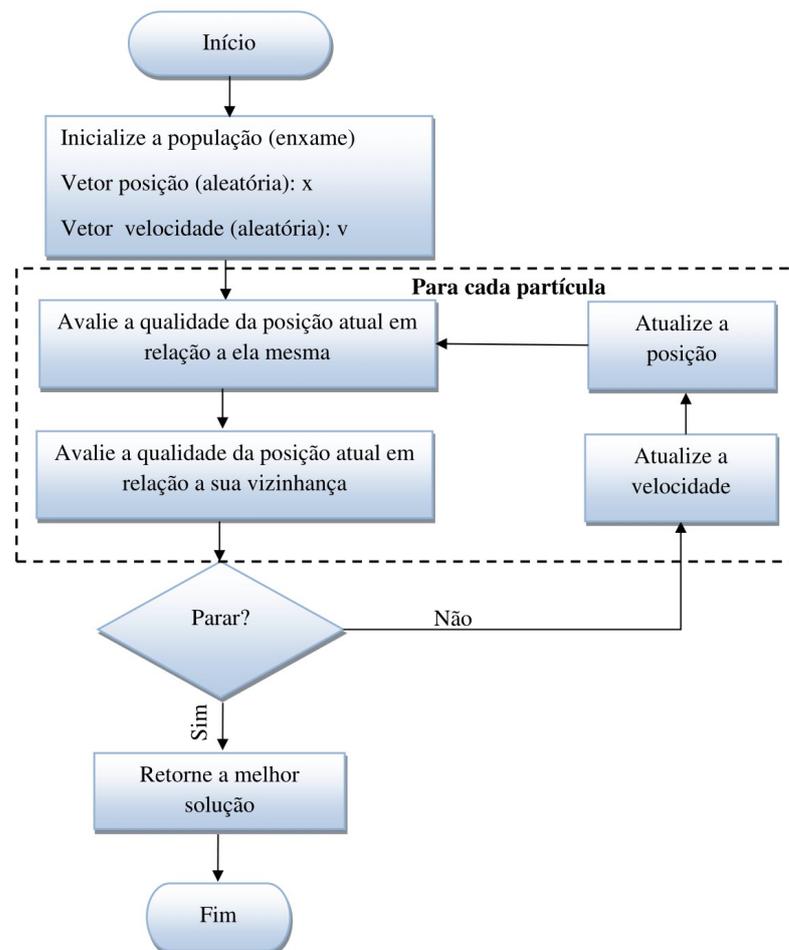


Figura 10 – Fluxograma Genérico PSO

Em [ANANDAKUMAR, et. al. 2018], o autor toma como exemplo para um melhor entendimento um bando de pássaros sobrevoando em uma determinada área, onde eles têm em sua mira uma fonte de comida. O pássaro que estiver mais próximo do alvo, faz um barulho mais alto e os outros pássaros vão em sua direção. Se qualquer um dos outros pássaros se aproximar do alvo primeiro, ele soa outro som mais alto, para que os demais pássaros venham em direção novamente. E esse ciclo continua até que um dos pássaros apareça com a comida.

A Fig. 10 ilustra um algoritmo genérico do funcionamento do PSO, com todas suas etapas. Em suma, ele procura o espaço de uma função objetivo, ajustando as partículas, como os caminhos por partes formadas por vetores de posição, buscando a melhor solução. Esta técnica têm sido amplamente utilizada por ter baixa complexidade computacional e rápida convergência.

2.8 Considerações Finais

Neste capítulo, foi apresentado o levantamento do estado da arte das tecnologias envolvidas nesta pesquisa, que serviram de embasamento para o referido trabalho. No próximo capítulo, é apresentado os trabalhos relacionados a esta pesquisa.

3 Trabalhos Relacionados

3.1 Considerações iniciais

Neste capítulo, é feito um levantamento das soluções propostas que norteiam o tema proposto neste trabalho. Através deste levantamento, objetiva-se ratificar que essa área ainda apresenta lacunas e problemáticas não resolvidas. Na Tabela 2, são apresentados todos os trabalhos, mostrando de maneira concisa as considerações dos mesmos.

3.2 Trabalhos Relacionados

Com o crescimento explosivo dos serviços de dados móveis, a capacidade das redes sem fio de transportar serviços de dados se torna cada vez mais exigente. Isso vem se mostrando preocupante pela comunidade científica, em ter uma arquitetura de rede capaz de suprir esta demanda. Dessa forma, o balanceamento de carga da rede tem se mostrado como uma das alternativas para solucionar esse problema. Neste sentido, a técnica de *offloading* aparece como uma das possíveis soluções para esse problema.

Redes Hetnet é aceito como um caminho promissor para satisfazer a implacável demanda de tráfego nas redes móveis. Oferece a função de decisão de *offloading* em uma gama de serviços de transmissão sem fio com diferentes classes de serviço. Essa decisão é explorada em [LEI et. al.,2013], onde o autor considera tanto o ganho de *offloading*, quanto o custo de usar a HetNet quando um SLA (*Service Level Agreement*) é estabelecido com ele. Porém, a técnica de *offloading* utilizada neste trabalho, gera altos custos de CAPEX/OPEX, e altas interferências na rede.

Em [DING et al, 2016] foi proposto um Algoritmo de Handover em HetNets, para tratar o problema de alocação de recursos de PBS(Pico Base Station), baseado na posição do UEs e no RSRP((Reference Signal Received Power))em tempo real. O modelo de handover eficaz baseia-se no modelo exponencial de previsão de movimento do UE onde inicialmente é avaliado o problema da ocorrência de handovers desnecessários que são causados pelos UEs de alta velocidade.

Como dito no capítulo anterior, CRAN é uma aposta de muitos autores para redes 5G. Esta arquitetura tem a capacidade de ajustar dinamicamente as conexões lógicas entre as BBUs e os RRHs, de acordo com as condições do tráfego. Neste sentido, o balanceamento da rede é investigado em diversos trabalhos, que serão mostrados a seguir.

Propõe-se, em [KHAN et. al., 2015], o balanceamento de carga da rede em C-RAN, no qual é proposto um modelo de mapeamento otimizado para RRH-BBU, com uma abordagem do SON (*Self-Optimising Network*). Os autores propõem um conjunto de KPIs (*Key Performance Indicator*) para obter uma estrutura de rede auto-otimizada, resolvendo o problema de balanceamento de carga na rede, mantendo os níveis de QoS. Resultados aceitáveis foram obtidos, maximizando o processo de mapeamento RRH-BBU e, ao mesmo tempo, causando um desequilíbrio de carga na rede, pois o balanceamento foi realizado sem se preocupar com o balanceamento entre os setores.

Já no trabalho de [PAIXÃO et. al., 2018], o autor buscou aperfeiçoar o balanceamento dos recursos entre os setores de mais de uma BBU, além de possibilitar o balanceamento de cargas nas UBS. Com isso, o número de chamadas bloqueadas teve uma redução considerável, ao mesmo tempo que se notou um equilíbrio de carga entre as BBUs, otimizando o balanceamento da rede.

Em [KHAN et. al., 2017], é investigada a capacidade de remapeamento dinâmico de uma rede CRAN, para configurar as RRHs em setores das BBUs, em um ambiente de tráfego variante no tempo. É redistribuída a capacidade da rede em uma área geográfica, através de uma configuração do setor RRH. A QoS é formulada como um problema de otimização pelo autor, através de uma combinação de indicadores chaves de desempenho (KPIs), para o número de usuários bloqueados e handovers na rede. Então, é proposto o algoritmo SOCRAN (*Self-optimized Cloud Radio Access Network*) para melhorar a QoS da rede, por meio de balanceamento de carga de tráfego, com o mínimo de handovers possíveis. Resultados satisfatórios foram obtidos, no entanto o autor não se preocupa em oferecer uma melhor conexão e melhor aproveitamento dos recursos da rede, através da técnica de *offloading* em redes móveis.

Em [KHAN et. al., 2018], é enfatizada a técnica de auto-otimização para arquiteturas CRAN, visando melhorias de desempenho da rede e focando na modelagem de um problema de otimização multi-objetivo. O algoritmo CDI (*Cell Differentiation and Integration*) é usado para selecionar as BBUs e RRHs apropriadas para ativação / desativação após um ciclo CDI fixo, e a parte dinâmica executa o mapeamento de BBU apropriado para RRH, para balanceamento de rede, visando a máxima QoS com os handovers mínimos possíveis. Porém, a arquitetura CRAN utilizada nos trabalhos citados anteriormente requerem altos custos de CAPEX/OPEX e causam altos níveis de interferência na rede. Neste caso, é necessária uma fase de transição entre as arquiteturas. Nesse viés, a arquitetura H-CRAN vem sendo amplamente utilizada na literatura, sendo uma das propostas para 5G.

Conforme pode ser consultado em [LEE et. al., 2018], um fatiamento de rede para H-CRANs multi-tenant leva em consideração os recursos de banda base, capacidades de *fronthaul* e *backhaul*, QoS e interferência. Porém, não é explorada a eficiência da alocação

balanceada nos setores BBUs, explorando o uso de *offloading* em redes móveis, evitando altos congestionamentos na rede e altos eventos de chamadas bloqueadas.

Buscando explorar os potenciais benefícios da arquitetura H-CRAN, em [LIU et. al., 2018], é estudado o problema de associação de usuários e formulado como um jogo de correspondência bidirecional, onde os nós HPN (*High Power Node*) são a capacidade esgotada, e os UEs precisam recorrer à capacidade de *fronthaul* dos nós LPNs (*Low Power Nodes*). Os UEs devem considerar não apenas as taxas de dados atingidas, mas também o impacto negativo causado por outros UEs. Já os LPNs devem evitar o desempenho dos UEs excessivamente associados e, portanto, o balanceamento de tráfego deve ser levado em consideração. Apesar de ter resultados satisfatórios, esta técnica ocasiona altas interferências e ineficiência na alocação dos recursos da rede.

Em [TRAN et. al., 2018], o autor aborda o problema de backhaul sem fio em uma rede H-CRAN, e uma alocação de energia é feita para múltiplas células de diferentes níveis, levando em conta o consumo de energia de diferentes tipos de células e backhaul sem fio. Um algoritmo de iteração é proposto para determinar o número máximo de células que podem ser suportadas no H-CRAN.

O autor propõe um algoritmo de alocação de recursos, denominado H-WDRF (*Heterogeneous Weighted Dominant Resource Fairness*). Para resolver a transmissão de problema de alocação de energia, é formulado como um problema de beamforming, cuja função objetivo é minimizar a potência total de transmissão com computação individual, capacidade de recursos, SINR (*signal-to-interference-plus-noise ratio*) e restrições máximas de potência de transmissão. Porém, o autor não se preocupa com o balanceamento entre os setores da rede, o que causa instabilidade na mesma.

[ZHANG et. al., 2018] propôs uma arquitetura de virtualização de função de rede SDN para virtualizar de forma escalável os recursos sem fio e infraestruturas físicas, com base nas localizações e solicitações do usuário, em três tipos de redes sem fio virtuais: redes virtuais sem descarregamento, redes virtuais com WiFi *offloading* e redes virtuais com dispositivo D2D (*Device to Device*) de *offloading*.

Em nenhum dos trabalhos citados acima foi investigada diretamente a técnica de *offloading*, em conjunto com balanceamento de carga da rede e o mapeamento dinâmico entre BBU-RRH. Estes, por sua vez, podem ser considerados uma das soluções para estes problemas.

Objetivando um melhor esclarecimento das diferentes vertentes dos trabalhos citados acima, a Tabela 2 mostra de maneira clara e concisa os principais trabalhos relacionados a esta pesquisa, ressaltando o diferencial do referido trabalho proposto nesta dissertação.

Tabela 2 – Síntese dos Principais Trabalhos Relacionados

Referência	Objetivo	Principais Características	Desafios
[KHAN et. al., 2015]	Mapeamento otimizado para RRH-BBU.	Análise através dos KPIs, para balanceamento no C-RAN	Foi realizado um balanceamento sem se preocupar com o balanceamento entre os setores, ocasionando um desequilíbrio na rede
[PAIXÃO et. al., 2018]	Balanceamento otimizado dos recursos entre os setores de mais de uma BBU, em conjunto com o balanceamento de carga da rede.	Reformula os KPIs, otimizando o balanceamento entre os setores.	Utiliza a arquitetura CRAN, que ocasiona altas interferências na rede.
[KHAN et. al., 2017]	Remapeamento dinâmico de uma rede C-RAN. É redistribuída a capacidade da rede em uma área geográfica através de uma configuração do setor RRH.	Análise através de 4 KPIs distintos, para balanceamento no C-RAN em um ambiente de tráfego variante no tempo.	Não faz uso da técnica de <i>offloading</i> para desafogar pontos críticos da rede, liberando recursos para os demais usuários.
[KHAN et. al., 2018]	Enfatiza a técnica de auto-otimização no C-RAN sobre a melhoria do desempenho da rede.	Algoritmo CDI fixo seleciona BBUs e RRHs para ativação e desativação, e a parte dinâmica executa o mapeamento de carga da rede visando o máximo QoS com o mínimo de handovers.	Utiliza arquitetura CRAN que requer altos custos de CAPEX/OPEX e causam altos níveis de interferência na rede.
[LEE et. al., 2018]	Fatiamento de rede para H-CRANs multi-tenant	São explorados recursos de banda base, capacidades de fronthaul e backhaul, QoS e interferência.	Não é explorada a eficiência da alocação balanceada nos setores BBUs, explorando o uso de <i>offloading</i> em redes móveis.
[Proposta da Dissertação]	Arquitetura de <i>offloading</i> em redes móveis H-CRAN, considerando aspectos de rádio e BBU, aliado ao balanceamento de carga da rede.	Técnica de <i>Offloading</i> em conjunto com mapeamento BBU-RRH, buscando a maximização de carga.	Arquitetura de <i>offloading</i> em redes móveis H-CRAN, considerando aspectos de rádio e BBU, aliado ao balanceamento de carga da rede, buscando maximização de carga.

3.3 Considerações finais

Neste capítulo, foram apresentados trabalhos que norteiam o estado da arte do tema proposto nesta dissertação, a fim de mostrar a diversidade de estratégias e lacunas existentes ainda não resolvidas. O modelo do sistema, parâmetros de Rede e o Algoritmo TLA proposto é apresentado no capítulo que segue.

4 Modelo do Sistema, Parâmetros de Rede e Algoritmo TLA

4.1 Considerações iniciais

Neste capítulo, são feitas importantes considerações a respeito da descrição do Modelo do Sistema, juntamente com todos os parâmetros de rede. Por conseguinte, apresentam-se as métricas de avaliação de desempenho e a descrição do algoritmo TLA utilizado neste trabalho.

4.2 Modelo do Sistema

A arquitetura H-CRAN é caracterizada por atributos de SDN (*Software Defined Network*) e computação em nuvem, podendo aumentar consideravelmente a flexibilidade da arquitetura da rede, melhorar o sistema SE e reduzir significativamente o consumo de energia e os gastos operacionais. Além disso, a QoS pode ser notavelmente melhorada devido à distância reduzida entre RRHs e UEs, bem como a flexibilidade para permitir a associação do UE com diferentes níveis [PENG et. al., 2015].

A combinação das arquiteturas de redes Heterogêneas e CRAN, formando H-CRAN é uma grande escolha para redes 5G, pois permite o aproveitamento de recursos de arquiteturas antigas, fazendo a heterogeneidade, e tecnologia de centralização, permitindo maior controle de processamento e funcionalidades. Dessa forma, a Macro eNodeB cobre os usuários em áreas não alcançadas pelas RRHs, bem como pode atender usuários que não puderam ser conectados nelas [LEE et. al.,2018].

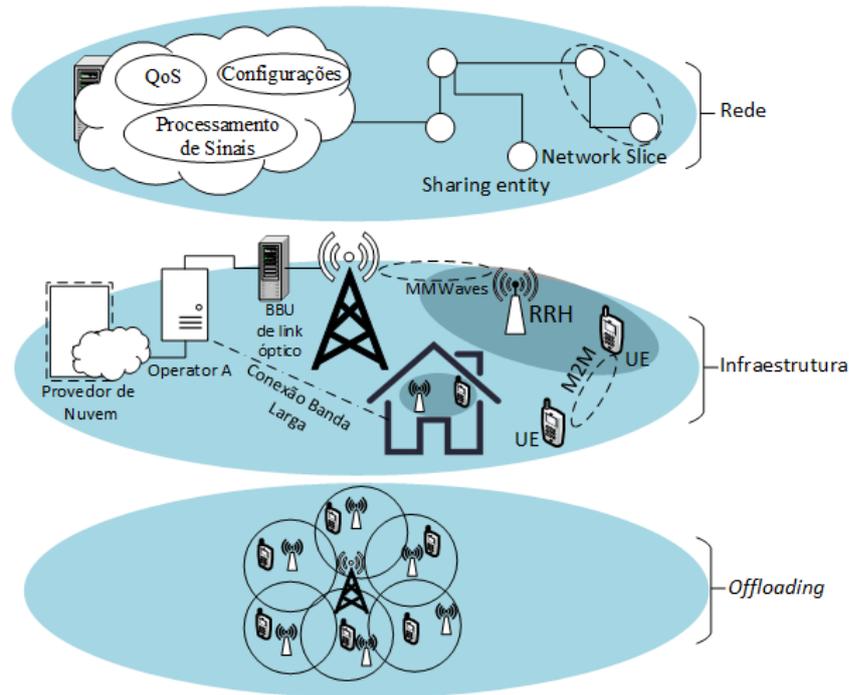


Figura 11 – Camadas Redes 5G

As propostas para 5G são muitas, e as proposições das camadas das mesmas surgiram no intuito de estimar todo o suporte necessário para implantação da rede. Na Fig. 11, é exemplificada a visão dessas camadas, que estão divididas entre: Camada de Rede, de Infraestrutura e de *Offloading*.

- Camada de Rede: tem o suporte necessário para a sinalização e monitoramento da rede.
- Camada de Infraestrutura: visa centralizar o processamento de carga buscando ganhos cooperativos.
- Camada de *Offloading*: busca descongestionar pontos críticos da rede, descarregando para Macro EnodeB usuários que antes estavam bloqueados, buscando a maximização de QoS, melhor utilização dos recursos da rede e, conseqüentemente, a maximização da rede.

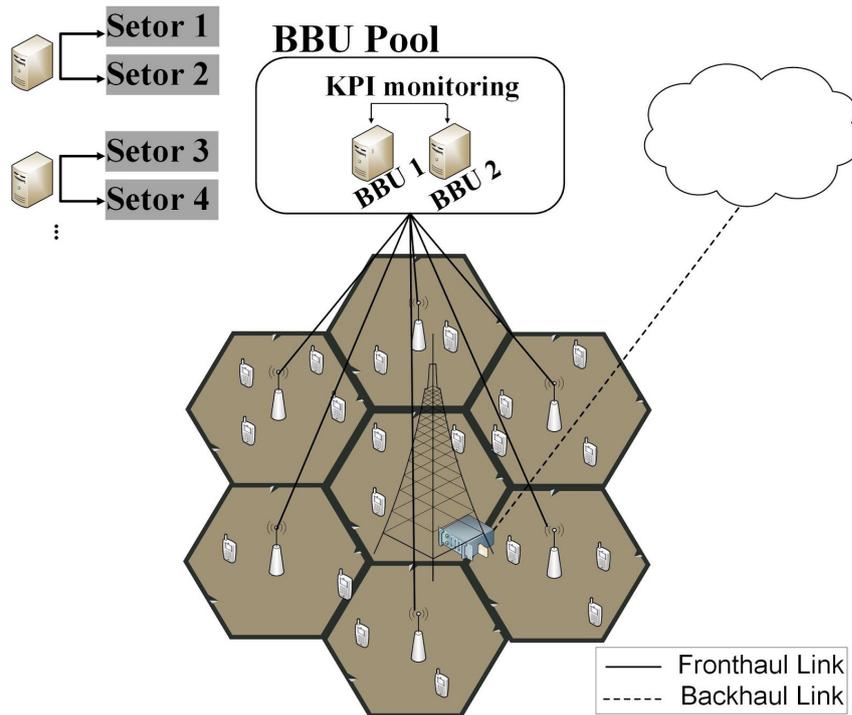


Figura 12 – Arquitetura Proposta

Neste estudo, é considerada uma rede celular de dois níveis, que consiste de um macro enodeB e 6 RRHs (de acordo com a Fig. 12), com baixo consumo de energia, que cooperam um com outro no BBU pool centralizado, visando alcançar altos ganhos cooperativos. Este trabalho, difere de [PAIXÃO et. al., 2018], no que diz respeito a arquitetura adotada, que no referido trabalho utilizava CRAN, e tratava apenas como métrica o número de usuários bloqueados na rede, e o balanceamento de carga do mesmo. Já neste trabalho, além de ser utilizado uma arquitetura H-CRAN, foi usado um conjunto de métricas aliado a técnica de *offloading* para avaliar a performance da rede.

É utilizado 1 BBU Pool, com 2 setores cada, cada um com capacidade máxima definida como ($HC_{max} = 30$). O sistema se inicia com a alocação dos usuários de forma inteligente, na qual é avaliado o SNR dos usuários com as antenas (RRHs e macro eNodeB), nos quais o usuário se encontra na área de cobertura. Para permitir a alocação do usuário, os seguintes critérios devem ser satisfeitos:

- A antena deve ter recursos disponíveis, pois a medida que os usuários vão sendo conectados a ela, os recursos vão se esgotando;
- O usuário deve estar na área de cobertura da antena;
- No caso das RRHs, o número de usuários conectados não deve ultrapassar o limite da capacidade (HC) de um setor da BBU, pois, nesse caso, está única antena obteria usuários bloqueados.

4.3 Parâmetros de Rede

Foi utilizado o modelo de perda de propagação SUI (*Stanford University Interim*), que é amplamente utilizado na literatura para perdas de trajetórias em redes celulares operando acima de 2 GHz. Onde $PL_{SUI}(d)$ em dB é encontrado por [CASTRO et. al., 2011] [SULYMAN et. al., 2016]:

$$PL_{SUI}(d) = PL_{(d0)} + 10n \log_{10} \left(\frac{d}{d_0} \right) + X_{(fc)} + X_{(RX)} + X_{\sigma} \quad (4.1a)$$

$$PL_{(d0)} = 10n \log_{10} \left(\frac{4\pi d_0}{\lambda} \right) \quad (4.1b)$$

$$n = a - b \cdot h_{TX} + \frac{c}{h_{TX}} \quad (4.1c)$$

$$X_{fc} = 6 \log_{10} \left(\frac{f_{MHZ}}{2000} \right), f_c > 2GHz \quad (4.1d)$$

$$X_{RX} = -10.8 \cdot \log_{10} \left(\frac{h_{RX}}{2} \right) \quad (4.1e)$$

- λ denota o comprimento de onda do portador em metros;
- $PL_{(d0)}$ representa a perda de caminho do espaço livre, em dB a uma distância inicial $d_0 = 100\text{metros}$;
- X_{σ} , é o desvanecimento causado pelas áreas de sombreamento, podendo estar entre $8.2 < \sigma < 10.6\text{dB}$;
- Os parâmetros a, b e c são constantes, utilizadas para modelar os tipos de terreno. Utilizamos o terreno A, que é o modelo adequado para vegetação montanhosa e densa, com valores ($a = 4.6$ / $b = 0.0075$ / $c = 12.6$);
- f_{MHZ} é a frequência portadora (fc) em MHz;
- h_{TX} é a altura da estação base, podendo estar entre 10 a 80 metros;
- n é o expoente de perdas;

Os UEs só são conectados, se estiverem no raio de cobertura da RRH ou macro EnodeB. A potência recebida no interior de cada célula é dada pela fórmula (4.2), na qual é realizada a conversão de dBm para Watts:

$$Pot_w = \frac{Pot_i - L_{SUI}}{1000} \quad (4.2)$$

Onde Pot_w é dado em watts, Pot_i é a potência inicial ou irradiada em dBm e L_{SUI} é o valor do desvanecimento do sinal. A partir do valor da potência, é possível calcular o valor de SNR e então a vazão máxima do UE conectado, conforme descrito abaixo:

$$SNR = \frac{Pot_w}{N} \quad (4.3)$$

$$V_{max} = B * \log_2(1 + SNR) \quad (4.4)$$

Então, Pot_w é a potência, enquanto N é o valor do ruído branco. Para obter a taxa de dados que o usuário obtém, é usado o teorema de *Shannon–Hartley* [SHANNON,1949]. Este teorema especifica a taxa máxima, em bits por segundo, que pode ser obtida por um canal, dado uma largura de banda B , que é a quantidade de largura de banda disponível para os usuários.

Para se conectar a rede, cada usuário deve conseguir um QoS mínimo, primeiramente estabelecido pela vazão de 200 kbps. Dado o valor de bit rate que o usuário conseguiu em um dado canal equação (4.4), para atingir o QoS estabelecido, este usuário precisará de N canais, estabelecidos nas equações (4.5) e (4.6).

$$X = \frac{QoS}{bitRate * 1024} \quad (4.5)$$

$$N = \min\{N \in Z | N \geq X\} \quad (4.6)$$

A equação (4.5) mostra quantos canais o usuário vai precisar para atingir o QoS mínimo de 200 kbps [KOUTLIA, et. a.2015], dado que esse usuário tem uma determinada vazão bitRate (multiplicando por 1024 para obter na mesma unidade de kilobits). Como o valor X pode ser qualquer valor real e não é possível fornecer um valor não inteiro de canais, a equação (4.6) fornece o número de canais N necessários pelo usuários que será o menor inteiro maior ou igual a X .

Essa relação é feita para todos os usuários e as antenas, nas quais eles estão no raio de atuação. O algoritmo de alocação prioriza primeiramente as RRHs, pois assim se tem uma melhor distribuição de atendimento dos usuários. Após todos relacionados, a alocação começa de fato pela melhor SNR de todas as relações, assim é garantido que os usuários que vão ocupando ocupem o mínimo, e o maior número possível de usuários seja alocado.

Quando não houver mais usuários para serem alocados nas RRHs, seja por falta de canais dessas para que o usuário consiga o mínimo de sua conexão, por ter atingido o limite de usuários conectados dela ou por todos os usuários nelas conseguirem se conectar, a alocação focará em atender os usuários que ainda estão desconectados através da macro.

Se um usuário não conseguir canais suficientes de uma antena para conseguir seu QoS mínimo, este usuário ficará desconectado. Finalizada a alocação de usuários antenas, o algoritmo passa a se preocupar com a setorização nas BBUs, voltando para uma alocação de RRHs-BBUs.

4.4 KPIs

Como função objetivo do PSO implementado, foi utilizado uma combinação de KPIs propostos em [KHAN et. al., 2017], seguindo os mesmos critérios. Segundo o autor, pode-se avaliar o índice de desempenho da rede, através das seguintes métricas: Indicador de Performance para Usuários Bloqueados, Inter-BBU Handovers, Intra-BBU Handovers e Forced Handovers. As formulações utilizadas são mostradas abaixo:

4.4.1 Indicador de Performance para Usuários Bloqueados (KPI_{BU})

O KPI de chamadas bloqueadas, visa avaliar a capacidade das BBUs e de seus setores (HCs). Quando tal valor é excedido, os usuários são considerados bloqueados. O número de usuários bloqueados BU na rede no tempo $t + 1$, é dado por [KHAN et. al., 2017]:

$$BU = \sum_s \max \left[\left(\left(\sum_i U_i R_{is}^{t+1} \right) - HC_s \right), 0 \right] \quad (4.7)$$

Onde $i = 1, 2, \dots, M$ e $s = 1, 2, \dots, S$. Em seguida, o KPI para usuários bloqueados (KPI_{BU}) pode ser apresentado como:

$$KPI_{BU} = \begin{cases} 1 & \text{if } BU = 0 \\ \frac{1}{BU} & \text{otherwise} \end{cases} \quad (4.8)$$

4.4.2 Inter-BBU Handovers:

São as transferências necessárias de uma BBU para outra. Dado uma configuração atual no tempo t , se na configuração $t+1$ a RRH_i e RRH_j estiverem em BBUs diferentes pode ocorrer uma handover entre BBUs. Isso apresenta um custo para a rede, que é representado nos cálculos de KPIs como custo de handover (Inter BBU) na qual é ponderada da relação da quantidade de usuários sobre a distância das RRHs.

$$\text{Inter-BBU}_{HO} = \sum_i \sum_{j \neq i} \frac{U_i}{D_{ij}} \left(1 - \left(\sum_n \sum_{s \in \text{SOS}_n} \left(R_{is}^{t+1} \cdot R_{js}^{t+1} \right) \right) \right) \quad (4.9)$$

4.4.3 Intra-BBU Handovers:

De forma similar ao KPI Inter-BBU, mede um custo de uma configuração que possa ocorrer o handover em uma única BBU, através da troca de setores. Para isso, a equação gera um valor dos custos, que seria a relação de usuários na RRH e a distância entre as RRHs, que estão em uma mesma BBU, mas em setores diferentes.

$$\text{Intra-BBU}_{\text{HO}} = \sum_i \sum_{j \neq i} \frac{U_i}{D_{ij}} \left[\sum_n \sum_{s \in \text{SOS}} (R_{is}^{t+1} \cdot R_{js}^{t+1}) - \sum_s (R_{is}^{t+1} \cdot R_{js}^{t+1}) \right] \quad (4.10)$$

$$\text{KPI}_{\text{intra}} = \begin{cases} 1 & \text{if Intra-BU}_{\text{HO}} = 0 \\ [\text{Intra-BBU}_{\text{HO}}]^{-1} & \text{otherwise.} \end{cases} \quad (4.11)$$

4.4.4 Forced Handovers:

Usado para quando uma RRH faz uma troca de setor. Ao trocar de configuração, mede o quanto seria custoso, na nova configuração mudar determinada RRH de setor, pois mudar uma RRH de setor é o mesmo que trocar todos os seus usuários conectados para um novo setor.

$$f_{\text{HO}} = \sum_s \sum_i (R_{is}^t + R_{is}^{t+1}) U_i \quad (4.12)$$

$$\text{KPI}_f = \begin{cases} 1 & \text{if } f_{\text{HO}} = 0 \\ [f_{\text{HO}}]^{-1} & \text{otherwise.} \end{cases} \quad (4.13)$$

4.4.5 QoS do Sistema

Dadas as equações dos KPIs, o QoS do sistema proposta em [KHAN et. al., 2017], pode ser definido como:

$$QoS_{\text{max}} = w_1 \cdot \text{KPI}_{\text{BU}} + w_2 \cdot \text{KPI}_{\text{inter}} + w_3 \cdot \text{KPI}_{\text{intra}} + w_4 \cdot \text{KPI}_f \quad (4.14)$$

Os níveis de prioridade ou pesos selecionados para (4.14) são baseados no método do *Rank Rank Centroid* (ROC), que é um método simples de atribuir pesos a algumas funções,

classificando de acordo com sua prioridade ou importância. A prioridade de cada função é tomada como uma entrada e convertida em peso usando a seguinte fórmula [AHN et. al., 2011]:

$$W_i = \left(\frac{1}{F}\right) \sum_{n=i}^F \frac{1}{n} \quad (4.15)$$

onde F é o número de funções (KPIs) e W_i é o peso da função i . KPI_{BU} é ponderado como $(1 + 12 + 13 + 14) / 4 = 0,52$, KPI_{inter} é ponderado como $(12 + 13 + 14) / 4 = 0,27$, KPI_{intra} é ponderado como $(13 + 14) / 4 = 0,15$, KPI_f em quarto lugar é ponderado como $(14) / 4 = 0,06$. Em que os pesos w_1 , w_2 , w_3 e w_4 são: 0.52, 0.27, 0.15 e 0.06, respectivamente.

4.5 Ambiente de Simulação

A simulação foi descrita por *Shannon* como "o processo de concepção de um modelo informatizado de um sistema (ou processo) e realizando experimentos com este modelo para o propósito de compreender o comportamento do sistema ou de avaliar várias estratégias para a operação do sistema"[POOCH, 1977].

A técnica utilizada nesta dissertação é a simulação computacional que, segundo [SILVA, 2018], tem menor custo e complexidade, o que torna mais práticos a modelagem e estudo de um sistema celular móvel. Usando uma simulação por computador, é possível projetar e analisar uma série de cenários de celulares móveis. Após a implantação de um sistema, a análise de desempenho pode ser aplicada para encontrar possíveis gargalos e sugerir alternativas para a sua expansão [FRANCÊS, 2001].

Foi utilizado o MATLAB® (do inglês MATrix LABoratory), a fim de avaliar a performance da proposta. Ele é um software interativo de alta performance, voltado para o cálculo numérico e gráficos científicos.

A Fig. 13 descreve os principais blocos funcionais da estrutura desenvolvida para implementação e modelagem do trabalho desenvolvido nesta dissertação.

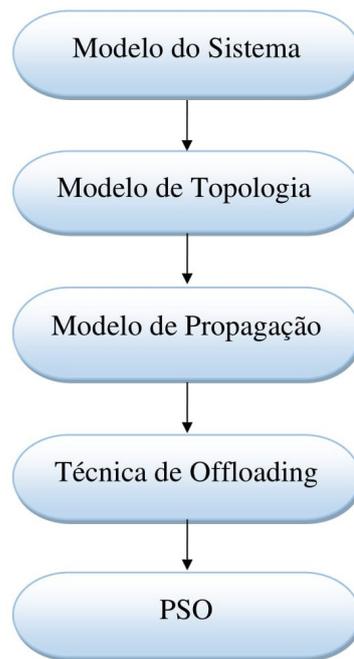


Figura 13 – Blocos Funcionais da Modelagem do Sistema

Assumiu-se uma área geográfica de 1000 m x 1000 m, onde as antenas foram dispostas, 6 RRHs e 1 macro EnodeB no centro. Os UEs foram dispostos aleatoriamente no cenário, como pode ser visto na Figura 14 abaixo:

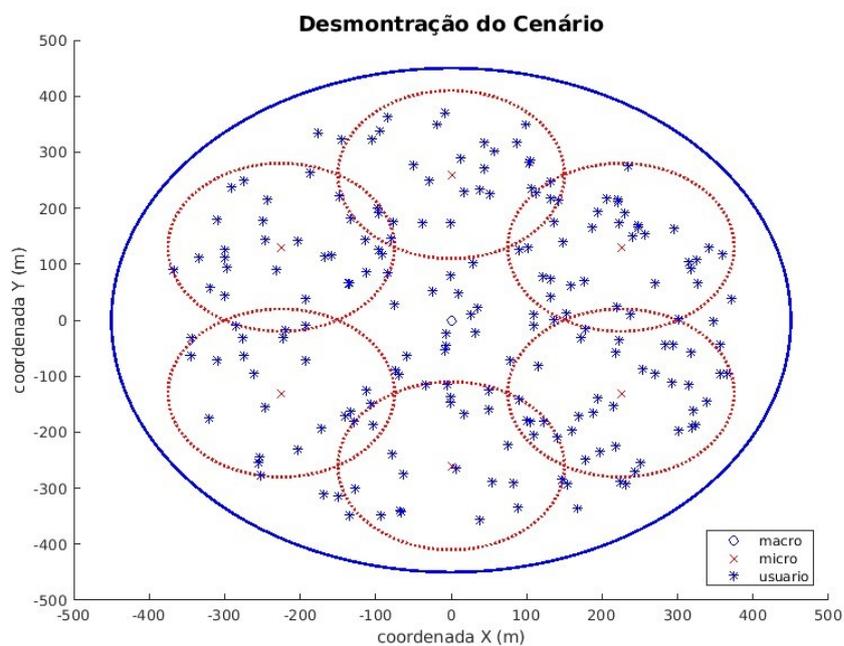


Figura 14 – Demonstração do Cenário

Os principais parâmetros utilizados na modelagem, são mostrados na tabela abaixo [PHAIWITTHAYAPHORN et. al., 2017][ZHANG et. al., 2018]:

Tabela 3 – Parâmetros de Simulação

	Macro EnodeB	RRH
Frequência	3.5 GHZ	2.5 GHZ
Potência de Transmissão	43 dbm	33 dbm
Raio de Cobertura	1000 m	150 m
Largura de Banda	20 MHZ	

4.5.1 Algoritmo TLA

O algoritmo busca maximizar os fatores de QoS do sistema, enquanto busca um balanceamento de carga. Para a quantificação do QoS, são levados em consideração os KPIs mostrados anteriormente. Através da avaliação, há uma tomada de decisão para a reorganização da rede. É aplicado o PSO para a reorganização e a busca pelo melhor resultado, que tem melhores KPIs e atende melhor o QoS.

O PSO começa com a geração aleatória de partículas, com uma distribuição linear, onde cada partícula é uma solução possível para o problema e representa o conjunto de RRHs, e seus respectivos setores, nos quais eles estão conectados. Na sequência, cada partícula é avaliada pela função objetivo, na qual retorna a qualidade da solução. Assim, os valores das partículas são atualizados, buscando o melhor G_{best} , depois disso, ele está em um loop até atingir os critérios de parada, definidos como 1000 iterações, onde a cada iteração atualiza suas partículas e calcula sua adequação.

Quando a solução atual é melhor que a anterior, os valores de partículas são atualizados (P_{Best}), se não, a solução é descartada e a solução anterior é mantida. Para cada iteração, verifica-se se a melhor solução encontrada permanece a mesma, caso contrário, ela é atualizada pelos melhores encontrados (G_{Best}). Este, é mostrado no subtópico 4.4.5 deste capítulo, bem como todos seus respectivos níveis de prioridade e/ou pesos selecionados.

A tomada de decisão do *offloading* é explicada na figura 15, onde as análises de KPIs e QoS são pré-determinadas. Uma das regras de restrição na alocação é que uma RRH não pode ter mais usuários conectados nela do que a capacidade máxima de um setor de BBU (HC). Ou seja, ela não pode estar ocupando sozinha um setor e ter usuários bloqueados. Nesse caso, todas as RRHs são selecionadas para escolher o usuário que sofrerá o *offloading*, onde o UE que possui o melhor SNR com a macro EnodeB é alocado para ele.

Porém, quando o PSO é executado e a organização da rede resultou ainda com usuários bloqueados, o sistema então seleciona os setores das BBUs que tiveram seu HC superado, escolhendo, dentre as antenas presentes nesses setores, qual usuário tem o melhor SNR com a macro. Se for detectado que o usuário consiga, com o SNR que tem com a macro, conecta nela, atendendo seu QoS mínimo, e então é realizado o *offloading*

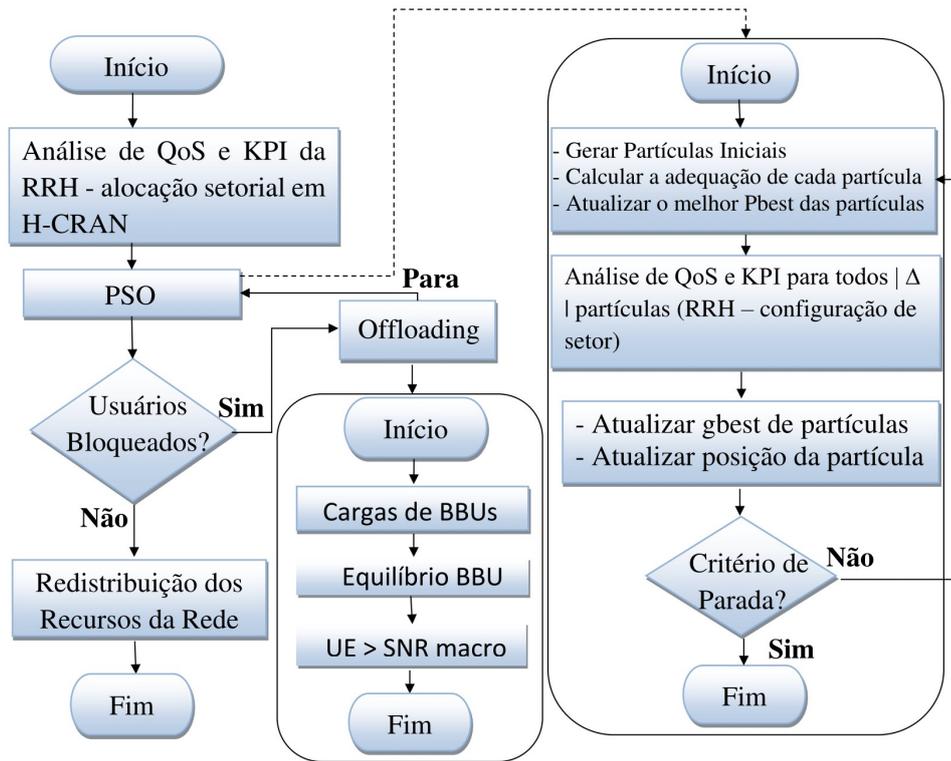


Figura 15 – Algoritmo TLA

com a mesma.

Quando o *offloading* não é mais possível, devido à macro não ter mais como atender o usuário, o processo se torna para escolher qual usuário ficará desconectado. É escolhido, dentre a RRH com maior número de usuários, o que apresenta pior conexão com a antena que está conectada. Com isso, há um melhor aproveitamento dos recursos da rede e manutenção dos níveis de QoS.

4.5.1.1 Alocação

O algoritmo inicia procurando a melhor forma de alocação, aquela na qual consiga maior número de usuários conectados e com melhor qualidade. Para isso, gera as relações SNR dos usuários com as antenas (RRHs ou *macro*), nas quais o usuário se encontra na área de cobertura.

Essa relação é feita para todos os usuários e as antenas. O algoritmo de alocação prioriza a conexão nas RRHs e, posteriormente, para a *macro* enodeB. Dessa forma, se mantém a proposta de rede centralizada, em que são melhor atendidos os locais com maior densidade, fornecendo maiores taxas de transmissão através das RRHs que possuem menor potência e raio de cobertura.

Após todos os clientes estarem devidamente relacionados, a alocação começa pela

melhor SNR de todas as relações, ou seja, o usuário que exige menos canais, assim, é garantido que os usuários que vão sendo alocados nas antenas ocupem o mínimo de canais possíveis e o maior número de usuários possam ser conectado.

Quando não houver mais alocações de usuários nas RRHs, seja por falta de canais dessas para que o usuário consiga o QoS mínimo, devido ao limite de usuários conectados na antena ser atingido ou por todos os usuários nelas conseguirem se conectar, a alocação passa a focar em atender os usuários que ainda estão desconectados através da *macro*, a qual passa pelo mesmo processo das RRHs com os usuários restantes.

Se um usuário não conseguir canais suficientes de uma antena para obter seu valor de QoS mínimo, este usuário ficará desconectado. Finalizada a alocação de usuários nas antenas, o algoritmo passa a tratar a setorização das BBUs.

4.5.1.2 Balanceamento RRH-BBU

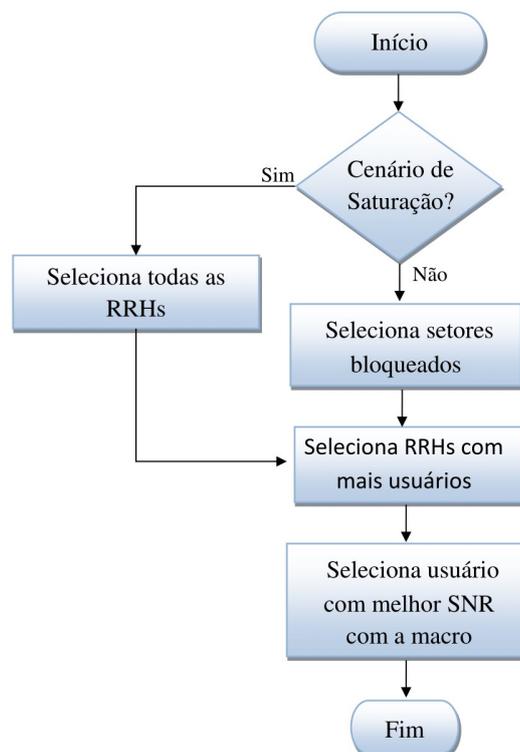


Figura 16 – Fluxograma do mapeamento de RRHs nas BBUs

Quando for finalizada a fase de alocação, as RRHs com seus usuários conectados terão que ser alocadas nos setores das BBUs, obedecendo a capacidade máxima dos setores (HC). Para que não ocorram casos de usuários bloqueados, inicialmente não pode haver mais usuários conectados nas RRHs do que a capacidade máxima (soma das capacidades de todos os setores) da BBU pool.

Quando o sistema se encontra nesse estado, em que é impossível alcançar uma alocação sem obter usuários bloqueados, o algoritmo tenta realizar *offloading*, até que o

número de usuários conectados nas RRHs se torne igual à capacidade máxima da BBU *pool*, como mostra o fluxograma da figura 16.

Quando o cenário não está saturado, o PSO é executado para tentar mapear as RRHs nos setores. Cada partícula é um vetor em que o índice representa a antena, e o valor representa o setor da BBU.

Os resultados são gerados pela função objetivo de acordo com os KPIs definidos na seção 4.4. O critério de parada é definido como 1000 iterações, então o Gbest é obtido como resultado, sendo este a configuração que a rede passa a ter. Essa configuração é testada, para saber se há usuários bloqueados. Caso haja, o algoritmo tenta realizar *offloading*, se não, a configuração é mantida para a fase de redistribuição de recursos.

4.5.1.3 Decisão de *Offloading*

Quando há usuários bloqueados nos setores das BBUs, pelo menos um usuário não consegue obter comunicação, então um usuário é escolhido para ser atendido pela *macro*, ocorrendo o *offloading* para liberar espaço nos setores de BBU.

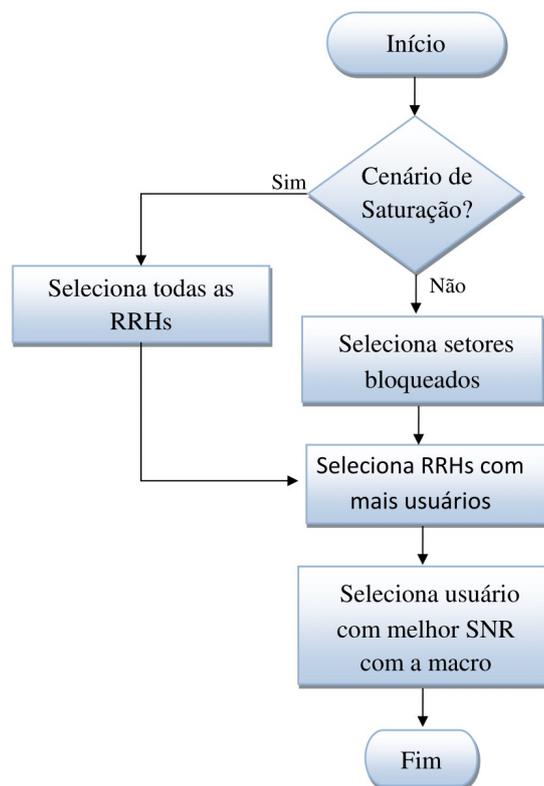


Figura 17 – Fluxograma da decisão de *offloading*

Como mostra o fluxograma da figura 17, para escolher o usuário em cenários não saturados, são selecionadas as RRHs dos setores das BBUs que possuem usuários bloqueados. Dentre essas RRHs, são selecionadas as que possuem maior número de usuários

conectados. Por fim, desse seleto grupo de RRHs, escolhem-se os usuários com melhor SNR com a macro.

Entretanto, quando se está partindo de um cenário saturado, ou seja, o número de usuários nas RRHs é maior que o HC máximo da BBU pool, são consideradas todas as RRHs, ao invés daquelas que estão em um setor da BBU com usuários bloqueados, para procurar um usuário que será alvo do offloading. Para que, ao fim, seja testado se é possível concluir o offloading do usuário, ou seja, se a macro dispõe de canais suficientes para atendê-lo.

4.6 Considerações Finais

Neste capítulo, foi mostrado o modelo do sistema, e todos os parâmetros de rede utilizados nesta dissertação. Em subsequente, foi detalhado o ambiente de simulação e o algoritmo TLA proposto. No capítulo posterior, é feita a discussão dos resultados obtidos no referido trabalho.

5 Análise dos Resultados

5.1 Considerações iniciais

O *offloading* em redes móveis, em conjunto com o balanceamento de carga e mapeamento BBU-RRH, foi proposto neste trabalho. Neste capítulo, discutem-se os resultados obtidos e suas respectivas análises, objetivando explicar que a proposta desta dissertação se mostrou eficaz.

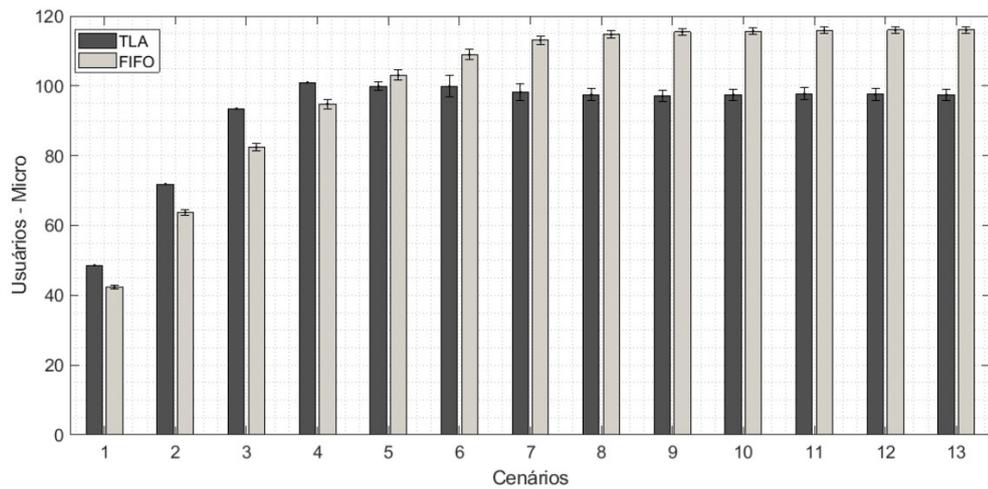
5.2 Resultados

Como *benchmarking* foi utilizado uma abordagem baseada no FIFO, que é amplamente utilizado na literatura. Neste método, os recursos são alocados sem nenhum tipo de prioridade, o primeiro usuário a solicitar irá ser atendido por uma RRH disponível com o melhor SNR. Já na proposta TLA, é priorizada a conexão nas RRHs, até o esgotamento dos recursos da rede, visando que os usuários sejam atendidos com a maximização da capacidade da rede e manutenção dos níveis de QoS.

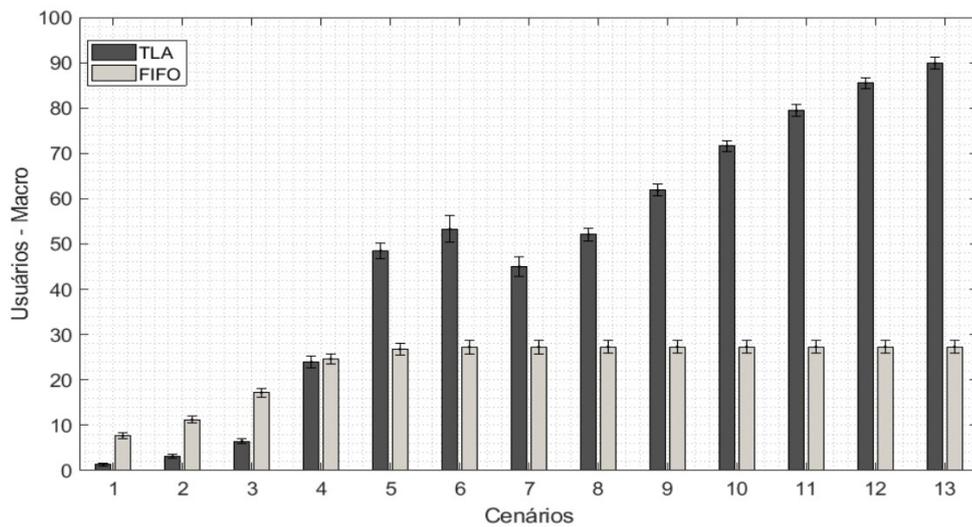
Uma característica importante para a análise estatística correta dos resultados da simulação são os intervalos de confiança. Neste trabalho, foi adotado o intervalo de confiança de 95%, no qual todos os cenários foram simulados 50 vezes, e os resultados apresentados são referentes à média destas.

As medidas de desempenho adotadas foram o número de conexões por tipo de RAN (nas RRHs e na Macro ENodeB), a média de número de usuários conectados, desconectados, e a Taxa de Bits Máxima na rede. Foram gerados 13 cenários na coleta de resultados, contendo 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350 usuários, respectivamente.

O número de conexões por tipo de RAN é mostrado na Figura 18 (a) e (b), a qual mostra as limitações do equipamento durante a modelagem. Pode-se verificar que o número de conexões em RRHs é maior em FIFO, já em relação ao número de conexões na macro, TLA obteve um número de conexões maior, visto que oferece uma melhor conexão. O número de usuários desconectados e conectados apresenta uma tendência linear, devido à adição de usuários nos cenários (que é linear).



a) Número de conexões em RRHs



b) Número de conexões na macro EnodeB

Figura 18 – Número de Conexões por tipo de RAN

A alocação dos recursos de rede é redistribuída de forma eficiente, garantindo uma melhor conexão dos usuários na rede TLA.

Na Figura 18 (b), no cenário 6, o TLA perde a linearidade e o FIFO tende à saturação, devido aos recursos da rede, que são subutilizados no FIFO. A técnica de offloading presente no TLA garante altos níveis de QoS, para os melhores usos dos recursos da rede.

Quando comparados o número de usuários conectados, percebe-se que a proposta TLA se mantém superior, como pode ser visto na Figura 19. A média de usuários conectados é cerca de 25% maior no TLA. Isso acontece devido ao esquema de descarregamento, que abre o tráfego da rede e descarrega na macro EnodeB os usuários que foram bloqueados na BBU.

Já o número de usuários desconectados é 30% maior no FIFO, de acordo com a Figura 20, constatando que TLA conseguiu uma melhor conexão e manteve o número de usuários desconectados menor que FIFO.

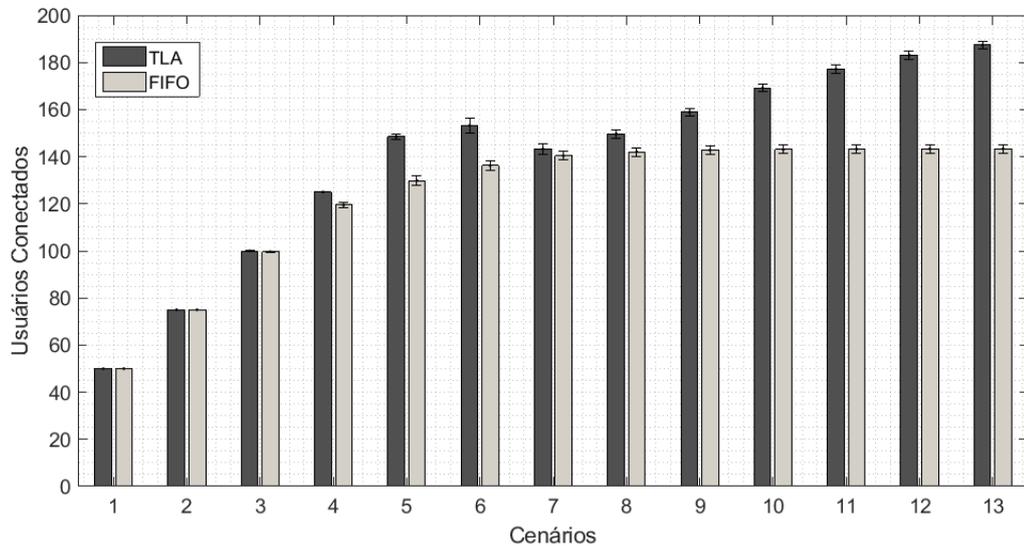


Figura 19 – Média de Usuários Conectados

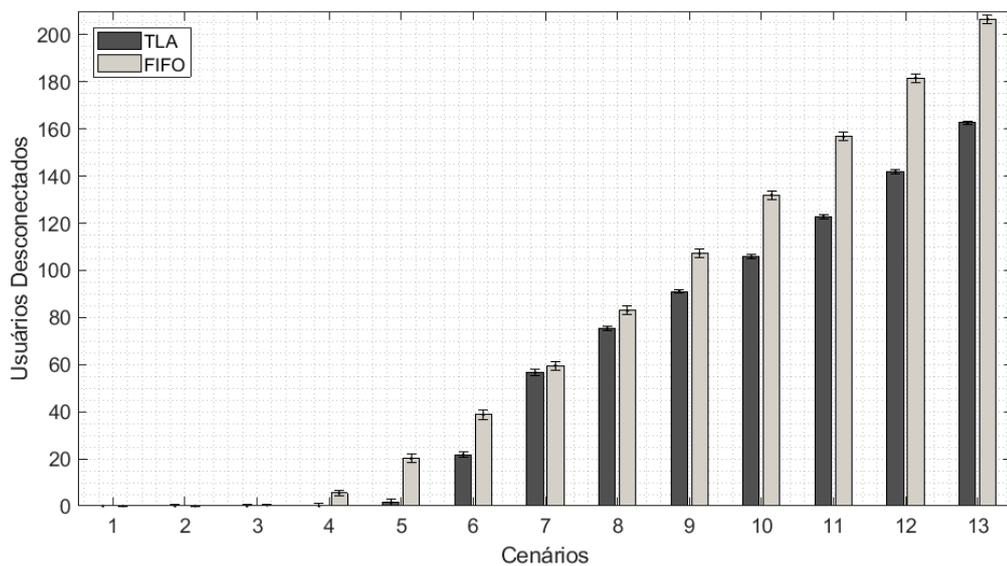


Figura 20 – Média de Usuários Desconectados

O método de alocação de TLA faz com que os usuários com melhores SNR conectem-se nas antenas e, se não houver canais suficientes, os usuários de cenários anteriores que possuíam pior sinal passam a não conseguir conexão. Devido a possuírem melhor sinal, possuem também melhores taxas de dados, o que faz com que sejam necessários menos canais para atingir os valores desejados, o que implica na rede conseguir atender mais usuários.

A Figura 21 apresenta a média de taxa de dados, onde se observa a manutenção dos níveis mínimos de QoS em todos os cenários. Nota-se que TLA e FIFO se mantiveram na média, o primeiro garantiu uma melhor conexão quando comparado ao segundo, onde ambos se mantiveram constantes, pois, no início, os requisitos mínimos de QoS dos UEs são cumpridos e, posteriormente, os recursos são redistribuídos, obtendo-se maior número de usuários ao longo da modelagem.

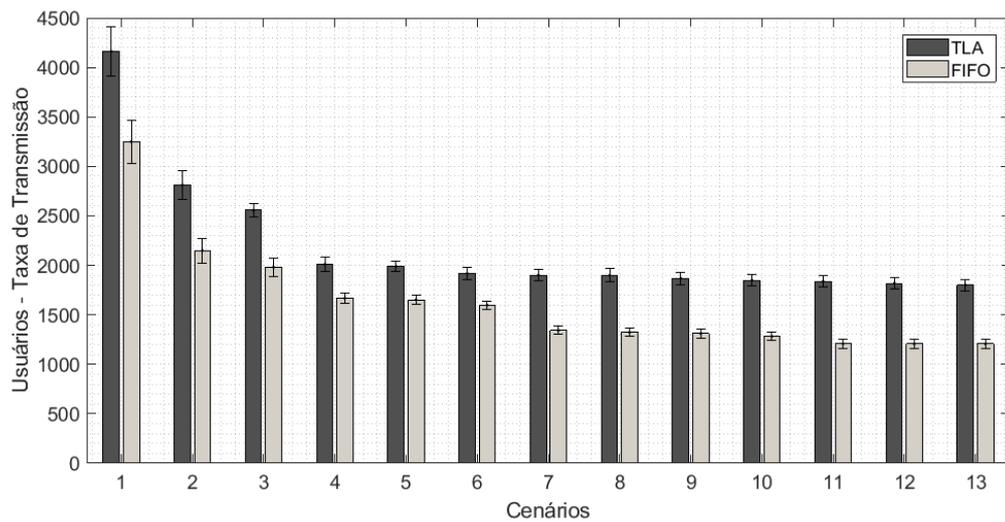


Figura 21 – Média de Taxa de Dados

Assim, é fato que a proposta da TLA obteve um desempenho satisfatório de cerca de 60% em relação ao FIFO, respeitando todos os requisitos estabelecidos, maximizando sua capacidade, mantendo os níveis mínimos de QoS e cobertura de rede.

5.3 Considerações finais

Neste capítulo, foram demonstrados os resultados obtidos neste trabalho, com a técnica de *offloading* proposta. Pode-se constatar que o algoritmo TLA teve um desempenho satisfatório e consegue oferecer altos níveis de QoS, mostrando-se superior ao *benchmarking* FIFO, oferecendo melhor conexão, minimizando o número de usuários desconectados e tendo elevado número de usuários conectados na rede.

6 Conclusões

Com a evolução dos dados móveis para os próximos anos e de novas exigências de aplicações, no que diz respeito à largura de banda, torna-se indispensável um estudo acerca dos parâmetros de QoS, ao aumento da cobertura da rede, a busca por técnicas eficientes para alocação de canais, assim como um melhor aproveitamento dos recursos da rede para prover serviços de qualidade a usuários finais.

As operadoras de telefonia móveis fazem altos investimentos para dar conta destas premissas, buscando prover um aumento considerável na capacidade das redes móveis. Com isso, a proposta do 5G busca atender as exigências dos usuários, aliada com eficiência energética e espectral.

A busca da comunidade científica por soluções centralizadas, em conjunto com técnica de *offloading*, visando descongestionar pontos críticos da rede, vêm sendo uma das propostas para as redes móveis 5G. Neste trabalho, foi apresentada uma proposta de Alocação de Dois Níveis, baseado em *offloading* em uma Arquitetura H-CRAN, visando centralizar o processamento de carga e obter altos desempenhos cooperativos. Foi apresentado um levantamento do estado da arte de todas as áreas envolvidas na pesquisa, a partir de eventos renomados (periódicos, conferências e grupos de pesquisa de referência), arquiteturas centralizadas e seus principais benefícios. Por conseguinte, foram apresentados todos os trabalhos que nortearam o tema proposto, ressaltando pontos em aberto na literatura.

A arquitetura de rede H-CRAN foi adotada para o desenvolvimento do trabalho, na qual se mostrou eficiente em relação as outras arquiteturas existentes. Uma simulação foi feita com vários cenários e diferentes disposições de usuários para gerar o intervalo de confiança dos resultados, que comprovaram a eficácia do algoritmo TLA, em correlação com o balanceamento de carga da rede, desafogando pontos nos quais a rede está sobrecarregada, realizando o *offloading* para a macro enodeB e, com isso, liberando largura de banda aos demais usuários, proporcionando uma melhor conexão, e gerou mais usuários atendidos e menos desperdícios da rede.

A rede foi organizada em um cenário onde se encontram várias RRHs e uma macro EnodeB. Uma BBU pool contendo BBUs com maiores capacidades de cargas, organizada através de um PSO. A rede conta com a decisão de *offloading* e com o mapeamento das RRHs nos setores das BBUs sendo feita a avaliação do QoS do sistema levando em conta os KPIs de indicadores de performance.

Todas as métricas para avaliação de desempenho da rede foram apresentadas, assim como os parâmetros de rede, ambiente de simulação e modelo de propagação adotados. O método de alocação, balanceamento de carga e técnica de *offloading* foram detalhados, afim de mostrar de maneira concisa o objetivo do trabalho.

Este estudo teve como objetivo principal o desenvolvimento de um novo esquema de *offloading* para arquiteturas H-CRAN, considerando aspectos de rádio e BBU, aliado no balanceamento de carga nas BBUs, tendo uma melhor utilização dos recursos da rede, maximização da capacidade, melhoria da cobertura e maximização do QoS dos usuários finais.

As simulações permitiram constatar a eficácia da proposta, onde TLA obteve desempenho satisfatório quando comparada ao FIFO, obtendo melhorias de aproximadamente 60% na Taxa Máxima de Bits. Isso, por sua vez, proporciona uma melhor conexão a todos os usuários da rede, distribuindo eficientemente todos os recursos da rede.

6.1 Trabalhos futuros

Como possíveis desdobramentos deste trabalho, podem-se destacar:

- Um estudo em novos cenários e diferentes topologias;
- Variar a capacidade da BBU, o número de antenas e os arranjos das mesmas;
- Analisar outras medidas de desempenho;
- Estudo de sensibilidade, para avaliar o comportamento da proposta em diferentes configurações de rede, como com frequências diferentes para RRHs, diferentes limites para os setores da BBUs, variação dos valores de exigência da taxa de transmissão dos usuários.

6.2 Dificuldades Encontradas

Durante o desenvolvimento deste trabalho foram encontrados diversos obstáculos, referentes à pesquisa das tecnologias envolvidas, compreensão dos trabalhos bases, desenvolvimento dos algoritmos e testes. Porém estes não impediram a realização, apenas mediram os esforços e revelaram a magnitude do trabalho. Pode-se destacar:

- A realização de um estudo aprofundado da tecnologia LTE, compreendendo as principais características para a realização do estudo;
- Na fase de implementação e modelagem no MATLAB®, no que diz respeito à parametrização da simulação das antenas;
- Escolha do algoritmo evolutivo com um tempo de convergência satisfatória para validação da proposta;
- Os testes com as simulações, que foram demasiadamente longos e necessários para cada uma das várias etapas: distribuição dos usuários e antenas, alocação, balanceamento através PSO, *offloading*, decisão de desconexão e redistribuição de recursos.

Contudo, os esforços foram válidos para sanar qualquer dificuldade e resolver os problemas para a obtenção dos resultados.

6.3 Contribuições

As principais contribuições do trabalho podem ser resumidas da seguinte forma:

- Técnica de descarregamento eficiente, com melhor aproveitamento de recursos na arquitetura H-CRAN;
- Estudo de algoritmos de *offloading* aplicados a arquiteturas híbridas;
- Algoritmo de alocação de usuários em arquiteturas de dois níveis H-CRAN;
- Distribuição Multiobjetivo de Recursos;
- Estudo detalhado da camada física do Modelo de Propagação da SUI;
- Implementação de testes para validar a proposta em estudos de caso;

Além disto, a divulgação do trabalho por meio de publicação de artigos em conferências Nacionais e Internacionais, onde são apresentadas a proposta e os resultados do estudo de caso desta dissertação. Os artigos são listados abaixo:

- GONCALVES, M. P. S., LETO, M. B., FOGAROLLI, R., BARROS, F. J. B., CARDOSO, D. L. Offloading approach for a Two-Level Allocation in H-CRAN Architecture. In: Cloud Technologies and Applications (CloudTech), 2018 International Conference on. IEEE, 2018.
- Gonçalves, M. D. P. D. S. et al. Abordagem de Offloading para a Alocação de Dois Níveis em uma Arquitetura H-CRAN. In: VIII CONFERÊNCIA NACIONAL EM COMUNICAÇÕES, REDES E SEGURANÇA DA INFORMAÇÃO, 8. 2018. Salvador. Anais. p. 79-80.
- LETO, M. B., GONCALVES, M. P. S., JUNIOR, L. G. L., PEREIRA, P. H. A., CARDOSO, D. L. Análise de Realocação de Recursos baseado em Offloading usando o Algoritmo Two Level Allocation. II Congresso de Tecnologias e Desenvolvimento da Amazônia (II CTDA);
- Goncalves, M. P. S., Leto, M. B., Fogarolli, R., Barros, F. J. B., Cardoso, D. L. Two-Level Allocation for H-CRAN Architecture Based in *Offloading*. (Submetido a periódico, aguardando avaliação).
- Orientação de Trabalho de Conclusão de Curso do Aluno Matheus Barros Leto, cuja a coorientação era do Professor Drº Diego Lisboa Cardoso, sobre o Tema de “Redistribuição de Recursos Em Rede H-CRAN Baseado em Offloading Usando o Algoritmo Two-Level Allocation”.

Referências

- AGIWAL, Mamta; ROY, Abhishek; SAXENA, Navrati. Next generation 5G wireless networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, v. 18, n. 3, p. 1617-1655, 2016.
- AHN, Byeong Seok. Compatible weighting method with rank order centroid: Maximum entropy ordered weighted averaging approach. *European Journal of Operational Research*, v. 212, n. 3, p. 552-559, 2011.
- ANANDAKUMAR, H.; UMAMAHESWARI, K. A bio-inspired swarm intelligence technique for social aware cognitive radio handovers. *Computers & Electrical Engineering*, v. 71, p. 925-937, 2018.
- ANATEL. Estatísticas de Celulares no Brasil. Disponível em: <<http://www.teleco.com.br/ncel.asp>>. Acesso em: Outubro de 2018.
- BABA, Majid Irfan et al. Evolution of Mobile Wireless Communication Systems from 1G to 5G: A Comparative Analysis. 2018.
- CASTRO, Bruno Souza Lyra et al. Modelo de propagação para redes ~ sem fio fixas na banda de 5, 8 GHz em cidades típicas da região ~ amazônica. 2010.
- CHANG, Peiliang. Cross-Layer Energy-Efficient Mobile Network Design. 2018. Tese de Doutorado. KTH Royal Institute of Technology.
- Cisco. 2017. "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014– 2021."
- CLAUSSEN, Holger et al. *Small Cell Networks: Deployment, Management, and Optimization*. John Wiley & Sons, 2017.
- DA PAIXÃO, Ermínio Augusto Ramos et al. Optimized load balancing by dynamic BBU-RRH mapping in C-RAN architecture. In: *Fog and Mobile Edge Computing (FMEC)*, 2018 Third International Conference on. IEEE, 2018. p. 100-104.
- DEL VALLE, Yamille et al. *Particle swarm optimization: basic concepts, variants and applications in power systems*. 2008.
- DING, Z.; WANG, X.; WANG, X.; XIAO, L. "An effective handover scheme in heterogeneous networks," 2016 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2016.

FRANCÊS, C. R. L. Statecharts Estocásticos e Queuing Statecharts – Novas Abordagens para Avaliação de Desempenho Baseadas em Especificação Statecharts. 2001. Tese (Doutorado em Ciência da Computação e Matemática Computacional). Universidade de São Paulo, São Carlos, 2001.

KENNEDY, James. Particle swarm optimization. In: Encyclopedia of machine learning. Springer, Boston, MA, 2011. p. 760-766.

KHAN, M.; ALHUMAIMA, R. S.; AL-RAWESHIDY, H. S. Quality of service aware dynamic BBU-RRH mapping in cloud radio access network. In: Emerging Technologies (ICET), 2015 International Conference on. IEEE, 2015. p. 1-5.

KHAN, Muhammad; ALHUMAIMA, Raad S.; AL-RAWESHIDY, Hamed S. QoS-Aware Dynamic RRH Allocation in a Self-Optimized Cloud Radio Access Network With RRH Proximity Constraint. IEEE Transactions on Network and Service Management, v. 14, n. 3, p. 730- 744, 2017.

KHAN, M.; FAKHRI, Zainab H.; AL-RAWESHIDY, Hamed S. Semistatic Cell Differentiation and Integration With Dynamic BBU-RRH Mapping in Cloud Radio Access Network. IEEE Transactions on Network and Service Management, v. 15, n. 1, p. 289-303, 2018.

KOUTLIA, Katerina; PÉREZ-ROMERO, Jordi; AGUSTI, Ramón. On enhancing almost blank subframes management for efficient eICIC in HetNets. In: Vehicular Technology Conference (VTC Spring), 2015 IEEE 81st. IEEE, 2015. p. 1-5.

LEE, Ying Loong et al. Dynamic Network Slicing for Multitenant Heterogeneous Cloud Radio Access Networks. IEEE Transactions on Wireless Communications, v. 17, n. 4, p. 2146-2161, 2018.

LEI, Lei et al. Challenges on wireless heterogeneous networks for mobile cloud computing. IEEE Wireless Communications, v. 20, n. 3, p. 34-44, 2013.

LIU, Shaobo et al. A Delay and Traffic Balancing Sensitive Scheme for User Association in Limited Fronthaul H-CRAN s. In: 2018 15th International Symposium on Wireless Communication Systems (ISWCS). IEEE, 2018. p. 1-5.

MAROTTA, Marcelo A. et al. Design considerations for software defined wireless networking in heterogeneous cloud radio access networks. Journal of Internet Services and Applications, v. 8, n. 1, p. 18, 2017.

OLTEANU, Alexandru-Corneliu; ȚĂPUȘ, Nicolae. Offloading for mobile devices: A survey. UPB Scientific Bulletin, v. 76, n. 1, p. 3-16, 2014.

- PENG, Mugen et al. Energy-efficient resource assignment and power allocation in heterogeneous cloud radio access networks. *IEEE Transactions on Vehicular Technology*, v. 64, n. 11, p. 5275-5287, 2015.
- PENG, M. et al. Heterogeneous cloud radio access networks: a new perspective for enhancing spectral and energy efficiencies. *IEEE Wireless Communications*, v. 21, n. 6, p. 126-135, 2014.
- PHAIWITTHAYAPHORN, Prapassorn et al. Cell throughput based sleep control scheme for heterogeneous cellular networks. In: *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2017 14th International Conference on. IEEE, 2017. p. 584-587.*
- RAMAZANALI, H, et al. Survey of user association in 5G HetNets. In: *Communications (LATINCOM), 2016 8th IEEE Latin-American Conference on. IEEE, 2016. p. 1-6.*
- RAZA, Muhammad Rehan et al. Demonstration of dynamic resource sharing benefits in an optical C-RAN. *Journal of Optical Communications and Networking*, v. 8, n. 8, p. 621-632, 2016.
- RODRIGUEZ, Jonathan (Ed.). *Fundamentals of 5G mobile networks*. John Wiley Sons, 2015.
- RUSSELL, Stuart J.; NORVIG, Peter. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,, 2016.
- RUSSEL, S., NORVIG, P., *Artificial Intelligence*, Prentice Hall, 2003.
- SERIES, M. IMT Vision–Framework and overall objectives of the future development of IMT for 2020 and beyond. *Recommendation ITU*, p. 2083-0, 2015.
- Shannon C E, “Communication in the Presence of Noise”. *Proceedings of the IRE*, vol. 37 pp. 10-21, 1949.
- SILVA, Ketyllen da Costa. *Estratégia de planejamento e otimização do handover em redes móveis densificadas*. 88 f. Tese (Doutorado) - Universidade Federal do Pará, Instituto de Tecnologia, Belém, 2018. Programa de Pós-Graduação em Engenharia Elétrica. Disponível em:<<http://repositorio.ufpa.br/jspui/handle/2011/10249>>. Acesso em: Setembro de 2018.
- SULYMAN, Ahmed Iyanda et al. Directional radio propagation path loss models for millimeter-wave wireless networks in the 28-, 60-, and 73-GHz bands. *IEEE Transactions on Wireless Communications*, v. 15, n. 10, p. 6939-6947, 2016.

- TRAN, Huu Q. et al. On the energy efficiency of NOMA for wireless backhaul in multi-tier heterogeneous CRAN. 2017. WANG, Ke; ZHAO, ZhongYuan; CEN, Yi. Allocating Multi-type Resources in Heterogeneous Cloud Radio Access Networks. *Mobile Networks and Applications*, v. 23, n. 3, p. 611-623, 2018.
- VORA, Lopa J. Evolution of mobile generation technology: 1G to 5G and review of upcoming wireless technology 5G. *International Journal of Modern Trends in Engineering and Research*, v. 2, n. 10, p. 281-290, 2015.
- ZHOU, Fuhui et al. Energy-efficient NOMA enabled heterogeneous cloud radio access networks. *IEEE Network*, v. 32, n. 2, p. 152-160, 2018.
- ZHANG, B. et al. Resource Allocation for 5G Heterogeneous Cloud Radio Access Networks with D2D Communication: A Matching and Coalition Approach. *IEEE Transactions on Vehicular Technology*, 2018a.
- ZHANG, X.; ZHU, Q. Scalable Virtualization and Offloading Based Software-Defined Architecture for Heterogeneous Statistical QoS Provisioning Over 5G Multimedia Mobile Wireless Networks. *IEEE Journal on Selected Areas in Communications*, 2018b.