

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

**ALOCÇÃO DINÂMICA DE RECURSOS EM REDES H-CRAN 5G E ALÉM
BASEADA EM OFFLOADING**

MARIANE DE PAULA DA SILVA GONÇALVES IMBIRIBA

TD: 13/2025

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém - Pará - Brasil
2025

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

MARIANE DE PAULA DA SILVA GONÇALVES IMBIRIBA

**ALOCÇÃO DINÂMICA DE RECURSOS EM REDES H-CRAN 5G E ALÉM
BASEADA EM OFFLOADING**

Proposta de Tese de Doutorado submetida
à Banca Examinadora do Programa de Pós-
graduação em Engenharia Elétrica da UFPA
para a obtenção do Grau de Doutora em
Engenharia Elétrica na área de Computa-
ção Aplicada, elaborada sob a orientação do
Prof. Dr. Diego Lisboa Cardoso.

UFPA / ITEC / PPGEE
Campus Universitário do Guamá
Belém - Pará - Brasil
2025

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

**ALOCÇÃO DINÂMICA DE RECURSOS EM REDES H-CRAN 5G E ALÉM
BASEADA EM OFFLOADING**

AUTOR: MARIANE DE PAULA DA SILVA GONÇALVES IMBIRIBA

PROPOSTA DE TESE DE DOUTORADO SUBMETIDA À AVALIAÇÃO DA
BANCA EXAMINADORA APROVADA PELO COLEGIADO DO PROGRAMA DE
PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA DA UNIVERSIDADE FEDE-
RAL DO PARÁ E JULGADA ADEQUADA PARA OBTENÇÃO DO GRAU DE DOU-
TOR EM ENGENHARIA ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM __ / __ / ____.

BANCA EXAMINADORA:

Prof. Dr. Diego Lisboa Cardoso
Orientador - PPGEE/ITEC/UFPA

**Profa. Dra. Jasmine Priscyla Leite de
Araujo**
Membro Interno - PPGEE/ITEC/UFPA

Prof. Dr. Fabrício José Brito Barros
Membro Interno - PPGEE/ITEC/UFPA

Prof. Dr. Mauro Margalho Coutinho
Membro Externo à instituição - UNAMA

Prof. Dr. Harold Dias De Mello Junior
Membro Externo à Instituição- ELE/PUC-RIO

Visto:

Prof. Dr. Diego Lisboa Cardoso
Coordenador do PPGEE/ITEC/UFPA

DEDICATÓRIA

À minha filha,
Maitê de Cassia,
que chegou à minha vida como um raio de sol depois da tempestade.
Seu sorriso iluminou meus dias mais escuros,
seu abraço me deu forças quando pensei em desistir.

Foi por você que continuei,
foi em você que encontrei sentido,
e é por você que sonho ainda mais alto.
Você é, e sempre será,
o amor que me move.

Com todo o meu amor,
Mariane de Paula

*“A tecnologia move o mundo,
mas é o ser humano que dá a direção.”*

— **Atribuída a Albert Einstein**

AGRADECIMENTOS

Início os agradecimentos por meus pais, que sempre primaram pela minha Educação. Obrigado Sr. Renivaldo Lemos e Sra. Virgínia Gonçalves que desde cedo, me ensinaram o valor do conhecimento para se entender o mundo e que me mostraram, pelo seu exemplo, que não há limites para a busca de um sonho. Vocês caminharam sob o sol para que eu pudesse andar na sombra — e é nessa sombra de esforço, renúncia e amor que esta conquista se tornou possível.

A toda minha família que direta ou indiretamente me ajudaram a alcançar mais um passo na minha formação acadêmica. Um agradecimento especial à minha irmã Márcia de Paula, que sempre representou para mim uma imagem materna. Sua presença, apoio e cuidado desde o início da minha jornada acadêmica foram essenciais. Sem você, este sonho não teria sido possível.

Ao meu esposo, Igor Casemiro, meu companheiro de todas as horas, obrigada por acreditar em mim, mesmo nos dias em que eu duvidei. Sua presença foi âncora e impulso.

Durante essa jornada, ganhei o maior presente da minha vida: minha filha, Maitê de Cassia. Filha, você nem imagina o quanto me fortaleceu. Seu sorriso, mesmo em dias difíceis, foi combustível. Que esta conquista um dia te inspire como você já me inspira.

Uma vez dentro da Universidade, algumas pessoas me convenceram a continuar os estudos, mostrando-me a nobre função da Pesquisa: produzir novos conhecimentos. Ao meu orientador, professor Diego Lisboa Cardoso, minha mais profunda gratidão. Sua generosidade, firmeza e dedicação à ciência são admiráveis. Obrigada por nunca soltar minha mão, mesmo nos momentos em que a caminhada pareceu pesada. Levo comigo cada orientação, palavra, e gesto de apoio.

Aos meus amigos que sempre me desejam o melhor e me acompanham em mais esse passo. Aos meus amigos do Laboratório de Pesquisa Operacional (LPO), que sempre estiveram prontos para ajudar em qualquer circunstância.

Um agradecimento especial ao Reyso Teixeira, cuja parceria fez toda a diferença

nesta reta final. Sua dedicação, atenção aos detalhes e generosidade em cada ajuda foram fundamentais para a conclusão deste trabalho.

Agradeço a CAPES pelo apoio financeiro em forma de bolsa para a realização desta pesquisa.

À Universidade Federal do Pará (UFPA), pela oportunidade a mim ofertada para a realização desta tese.

RESUMO

O aumento exponencial do tráfego de dados e a diversidade de aplicações móveis impõem novos desafios às redes sem fio, especialmente em áreas urbanas densas. Nesse cenário, as arquiteturas Heterogeneous *Cloud Radio Access Network* (H-CRAN) surgem como alternativa promissora por sua flexibilidade, escalabilidade e centralização de recursos. No entanto, a orquestração eficiente de usuários e recursos ainda é um desafio em ambientes com alta variabilidade de carga e mobilidade. Esta tese propõe um framework inteligente de orquestração em redes H-CRAN, que realiza a alocação dinâmica de usuários às RRHs e a distribuição destas entre setores da BBU, com base em fatores como demanda de tráfego, perfil de QoS, mobilidade e topologia da rede. A proposta inclui offloading adaptativo, redistribuição dinâmica e avaliação contínua do estado da rede. A solução foi avaliada por meio de simulações, considerando 24 cenários realistas com variações de carga (efeito de maré). Os resultados mostram ganhos expressivos frente ao modelo TLA tradicional, como mais usuários atendidos, melhor balanceamento de carga, menor bloqueio e maior estabilidade computacional. A pesquisa contribui com uma abordagem replicável e adaptativa para redes móveis de próxima geração, alinhada às exigências de QoS e escalabilidade.

Palavras-chave: Redes Móveis. 6G. H-CRAN. QoS. *Offloading*.

ABSTRACT

The exponential growth of data traffic and the diversity of mobile applications pose new challenges to wireless networks, especially in densely populated urban areas. In this context, Heterogeneous Cloud Radio Access Network (H-CRAN) architectures emerge as a promising alternative due to their flexibility, scalability, and centralized resource management. However, the efficient orchestration of users and resources remains a challenge in environments with high load variability and user mobility. This thesis proposes an intelligent orchestration framework for H-CRAN networks, performing the dynamic allocation of users to RRHs and their distribution across BBU sectors, based on factors such as traffic demand, QoS profile, mobility, and network topology. The proposed solution includes adaptive offloading, dynamic redistribution, and continuous network state assessment. The approach was evaluated through simulations encompassing 24 realistic scenarios with varying traffic loads (tidal effect). The results demonstrate significant improvements compared to the traditional TLA model, including a higher number of users served, better load balancing, reduced blocking, and improved computational stability. This research contributes a replicable and adaptive strategy for next-generation mobile networks, aligned with QoS and scalability requirements.

Keywords: Mobile. Networks 6G. H-CRAN. QoS. Offloading.

LISTA DE ILUSTRAÇÕES

Figura 1 – Topologia do MEC em redes móveis 6G (adaptado de (SUN et al., 2024)).	27
Figura 2 – Arquitetura CRAN [EJAZ et. al. 2020]	30
Figura 3 – Arquitetura H-CRAN	32
Figura 4 – Visão Geral do Cenário - Arquitetura H-CRAN	52
Figura 5 – Imagem Descritiva da Metodologia Proposta por esta Tese	53
Figura 6 – Fluxograma Geral	54
Figura 7 – Etapas de otimização com PSO	59
Figura 8 – Etapas de otimização com GA	60
Figura 9 – Etapas de otimização com FA	60
Figura 10 – Distribuição temporal dos usuários conectados ao longo do dia.	72
Figura 11 – Convergência dos algoritmos no Cenário 9 (Baixa Demanda).	74
Figura 12 – Convergência dos algoritmos no Cenário 14 (Alta Demanda).	74
Figura 13 – Comparativo de Usuários Aceitos (em unidades)(Algoritmos PSO, GA e FA).	75
Figura 14 – Comparativo do número de usuários não atendidos por cenário (em unidades) para os algoritmos PSO, GA e FA.	77
Figura 15 – DataRate médio por usuário (kbps) ao longo dos 24 cenários simulados, considerando os algoritmos PSO, GA e FA.	79

Figura 16 – Percentual de usuários com perfis alterados que foram atendidos após a redistribuição adaptativa (em %), considerando os algoritmos PSO, GA e FA.	80
Figura 17 – Tempos de execução dos algoritmos PSO, GA e FA ao longo dos 24 cenários simulados.	82
Figura 18 – Comparação do Coeficiente de Variação (CV) na distribuição de usuários por setor ao longo das 24 horas do dia, para os algoritmos GA, PSO e FA.	84

LISTA DE TABELAS

Tabela 1 – Parâmetros dos algoritmos bioinspirados utilizados	61
Tabela 2 – Características da população e representação das soluções	62
Tabela 3 – Perfis de usuários adotados na simulação	73

LISTA DE ABREVIATURAS E SIGLAS

5G	Fifth Generation (quinta geração de redes móveis)
6G	Sixth Generation (sexta geração de redes móveis)
AG	Algoritmo Genético
BBU	Baseband Unit
BBU Pool	Conjunto centralizado de BBUs
BS	Base Station
BU	Blocked Users
C-RAN	Cloud Radio Access Network
CoMP	Coordinated Multipoint
CV	Coeficiente de Variação
D2D	Device-to-Device
EE	Energy Efficiency
eNodeB	Evolved Node B
FA	Firefly Algorithm
fHO	Forced Handovers
Fob	Função Objetivo
GA	Genetic Algorithm
HC	Hard Capacity

H-CRAN	Heterogeneous Cloud Radio Access Network
HetNet	Heterogeneous Network
HO	Handover
IEEE	Institute of Electrical and Electronics Engineers
Inter-BBU HO	Handover entre diferentes BBUs
Intra-BBU HO	Handover entre setores da mesma BBU
IoT	Internet of Things
IRS	Intelligent Reflecting Surface
ITU-R	International Telecommunication Union – Radiocommunication Sector
KPI	Key Performance Indicator
MATLAB	Matrix Laboratory
MEC	Multi-access Edge Computing
MOPSO	Multi-Objective Particle Swarm Optimization
PSO	Particle Swarm Optimization
QoS	Quality of Service
QoS	Função Objetivo baseada em KPIs ponderados
RAN	Radio Access Network
RRH	Remote Radio Head

SINR	Signal-to-Interference-plus-Noise Ratio
SDN	Software Defined Networking
SE	Spectral Efficiency
SUI	Stanford University Interim
TLA	Two-Level Allocation
UE	User Equipment
UNA	Usuários Não Atendidos
Wi-Fi	Wireless Fidelity

SUMÁRIO

	Dedicatória	4
1	INTRODUÇÃO	20
1.1	Definição do Problema	23
1.2	Objetivo Geral	24
1.3	Objetivos Específicos	24
1.4	Organização do Trabalho	25
2	FUNDAMENTAÇÃO TEÓRICA	26
2.1	Considerações Iniciais	26
2.2	Redes 5G e a Transição para o 6G	26
2.2.1	Evolução das Redes Móveis: Do 5G ao 6G	26
2.3	Arquiteturas de redes móveis: HetNets, C-RAN e H-CRAN	28
2.3.1	HetNets	28
2.3.2	C-RAN	29
2.3.3	H-CRAN	31
2.3.4	O-RAN: Arquitetura Modular e Inteligente para Redes 6G	33
2.4	<i>Offload</i> em Redes Móveis	35
2.5	Inteligência Computacional	36

2.5.1	PSO	37
2.5.1.1	Algoritmo	39
2.5.1.2	Equações	40
2.5.1.3	Variações	41
2.5.2	AG	41
2.5.2.1	Funcionamento do AG	41
2.5.3	FA	43
2.6	Considerações Finais	45
3	TRABALHOS CORRELATOS	46
3.1	Considerações Iniciais	46
3.2	Trabalhos Correlatos	46
3.3	Considerações Finais	50
4	PROPOSTA DE TESE	51
4.1	Considerações Iniciais	51
4.2	Modelo do Sistema	51
4.2.1	Alocação Inicial	54
4.2.2	Modelo de Perda por Propagação (SUI)	55
4.2.3	Cálculo de PRBs	56
4.2.4	Dinâmica de Carga e Detecção de Saturação	57
4.2.5	Mapeamento BBU-RRH	58

4.2.5.1	Mapeamento Físico Inicial das RRHs	58
4.2.5.2	Otimização Heurística do Mapeamento RRH-BBU	58
4.2.6	Redistribuição Adaptativa de Usuários	62
4.2.6.1	Blocked Users	64
4.2.6.2	Inter-BBU	64
4.2.6.3	Intra-BBU	65
4.2.6.4	Forced Handovers	65
4.2.6.5	QoS do Sistema	66
4.2.6.6	Redistribuição Final com Perfis de Usuários	66
4.3	Considerações Finais	68
5	RESULTADOS	69
5.1	Considerações Iniciais	69
5.2	Definição das Métricas e Termos Utilizados	69
5.3	Ambiente de Simulação	70
5.3.1	Configuração do Cenário de Simulação	71
5.3.2	Análise dos Resultados	73
5.4	Considerações Finais	85
6	CONCLUSÃO E TRABALHOS FUTUROS	86
6.1	Considerações Iniciais	86
6.2	Síntese Final e Direções Futuras	86

6.3	Contribuições Parciais	87
6.4	Publicações Divulgadas	88
	REFERÊNCIAS	89

1 INTRODUÇÃO

Com o avanço de aplicações digitais emergentes, como realidade aumentada, veículos autônomos e *Internet of Things* (IoT), cresce a demanda por redes móveis com maior capacidade de transmissão, menor latência e maior eficiência energética. Segundo o Ericsson Mobility Report (2024), estima-se que até 2030 existam aproximadamente 6,3 bilhões de assinaturas 5G no mundo, representando 67% de todas as conexões móveis globais. Além disso, projeta-se que essa tecnologia seja responsável por 80% do tráfego mundial de dados móveis até o final da década, impulsionada pelo aumento no consumo de serviços em nuvem e conteúdos de alta definição (ERICSSON, 2024).

Diante dos avanços e desafios impostos pelas aplicações modernas, torna-se evidente que, embora o 5G represente um marco importante, não será capaz de, sozinho, sustentar as demandas futuras da conectividade global. Com o crescimento exponencial de dispositivos conectados e o surgimento de novos cenários de uso, é necessário olhar além do horizonte atual das redes móveis.

É nesse contexto que se insere a sexta geração de redes móveis (6G), concebida como uma evolução estratégica frente às limitações já percebidas no 5G. A padronização do 6G teve início em 2025, e seu lançamento comercial é projetado para 2030, acompanhando uma agenda internacional ambiciosa e altamente articulada. As expectativas em torno dessa nova geração são elevadas: velocidades de transmissão que podem atingir até 1 terabit por segundo, latência inferior a 100 microssegundos e suporte a uma densidade de até 10 milhões de dispositivos por quilômetro quadrado (ITU-R, 2023).

Mais do que números expressivos, o 6G busca viabilizar uma experiência digital profundamente imersiva e integrada. Aplicações como realidade estendida multissensorial, telepresença holográfica e automação ubíqua deixarão de ser apenas possibilidades remotas, tornando-se componentes naturais de um ecossistema digital contínuo e inteligente. Essa integração entre os mundos físico e digital será mediada por redes com inteligência artificial nativa, capazes de tomar decisões em tempo real e adaptar-se dinamicamente às condições do ambiente e às necessidades dos usuários.

O potencial transformador do 6G também se reflete em sua dimensão econô-

mica: estima-se que o mercado global relacionado a essa tecnologia poderá movimentar cerca de US\$ 281 bilhões até o final da década (SAAD; BENNIS; CHEN, 2019; ITU-R, 2023). Esse cenário indica não apenas uma nova era de conectividade, mas uma redefinição das bases tecnológicas que sustentam a sociedade digital contemporânea (SAHOO; TRIPATHY, 2025).

Para viabilizar essas metas ambiciosas, torna-se necessário revisar os fundamentos arquiteturais das redes móveis. Um dos conceitos mais relevantes nesse processo é o das redes heterogêneas (*Heterogeneous Networks – HetNets*), que integram diferentes tipos de células — como macrocélulas, microcélulas, picocélulas e femtocélulas — em uma mesma infraestrutura. Essa heterogeneidade permite uma cobertura mais flexível e adaptativa, especialmente em regiões de alta densidade populacional, onde a sobreposição de células possibilita maior reutilização espectral e melhora a experiência do usuário. As HetNets se destacam por sua capacidade de suportar diferentes tecnologias de acesso, topologias variadas e múltiplos níveis de potência e cobertura, criando um ecossistema de conectividade mais robusto e eficiente (LI; ZHANG et al., 2022).

Como evolução da arquitetura tradicional, surgem propostas como a C-RAN (*Cloud Radio Access Network*), que propõe a centralização do processamento das funções de banda base em data centers ou nuvens computacionais. Nessa arquitetura, as unidades de rádio (RRHs) são conectadas a BBUs centrais via enlaces de alta capacidade, permitindo uma alocação de recursos mais eficiente, economia de energia e aplicação facilitada de técnicas de virtualização. Embora promissora, a C-RAN apresenta limitações quanto à dependência de *fronthaul* com alta largura de banda e latência ultra-baixa, o que dificulta sua implementação em larga escala, principalmente em áreas remotas ou com infraestrutura limitada (CHEN, 2024).

Para contornar essas limitações e reunir os benefícios da C-RAN com a flexibilidade das HetNets, foi proposta a H-CRAN (*Heterogeneous Cloud Radio Access Network*). Essa arquitetura híbrida combina a descentralização de camadas de cobertura com a centralização do processamento, permitindo uma gestão inteligente do tráfego e da mobilidade dos usuários. Na H-CRAN, as RRHs de baixa potência atendem à demanda local de tráfego, enquanto as BBUs centralizadas processam os dados, e as macrocélulas garantem o controle e a mobilidade. Essa separação funcional torna possível aplicar estratégias como o desligamento seletivo de células e o redirecionamento

dinâmico de usuários, contribuindo para uma rede mais eficiente, sustentável e resiliente (CHOI, 2024).

Dentro da arquitetura H-CRAN, destaca-se o papel do *offloading* como estratégia essencial para garantir escalabilidade, flexibilidade e qualidade de serviço. O *offloading* consiste na redistribuição de usuários ou fluxos de dados entre elementos da rede — como RRHs, setores das BBUs e macrocélulas — com o objetivo de evitar sobrecargas locais e otimizar o uso dos recursos disponíveis. Essa técnica é particularmente relevante em cenários densos e dinâmicos, nos quais a demanda de tráfego varia constantemente e podem surgir zonas de saturação (*hotspots*). Ao permitir que usuários bloqueados ou excedentes sejam redirecionados para regiões menos congestionadas, o *offloading* assegura a continuidade da conexão e a manutenção dos requisitos de QoS (WANG, D.; WEN, X. et al., 2023).

Embora este trabalho adote a arquitetura H-CRAN como referência, é importante reconhecer a tendência crescente de migração para o paradigma *Open RAN* (O-RAN) nas redes móveis de quinta e sexta geração. Esta arquitetura propõe maior flexibilidade e interoperabilidade ao introduzir a abertura das interfaces, a virtualização e a desagregação funcional da RAN. Nesse modelo, funções antes rigidamente concentradas nas BBUs são divididas em módulos como a *Central Unit* (CU) e a *Distributed Unit* (DU), que podem ser implementados em plataformas de software sobre hardware de propósito geral, além de contar com o papel orquestrador da *RAN Intelligent Controller* (RIC) (ALAM et al., 2024; PORAMBAGE et al., 2024).

Essa evolução não apenas facilita a integração de múltiplos fornecedores, mas também habilita mecanismos inteligentes de gestão e automação baseados em IA, mantendo plenamente aplicáveis as estratégias propostas nesta tese. Dessa forma, o framework aqui desenvolvido, voltado à otimização da alocação dinâmica de recursos em redes H-CRAN, é projetado de maneira compatível com os princípios do O-RAN, preservando sua relevância frente às transformações futuras da infraestrutura de acesso (MAURICIO et al., 2025; LARSEN et al., 2024).

1.1 Definição do Problema

Embora diversos estudos abordem o balanceamento de carga em arquiteturas centralizadas, e discutam de forma pontual os desafios das estruturas descentralizadas, observa-se uma lacuna significativa na literatura no que diz respeito à gestão flexível e adaptativa do tráfego em arquiteturas híbridas, como as redes H-CRAN. Questões como o desligamento dinâmico de RRHs com base na variação de tráfego, a limitação de recursos físicos e o provisionamento inteligente de recursos computacionais e espectrais permanecem pouco exploradas, sobretudo no que tange à adaptação em tempo real da infraestrutura às flutuações do ambiente de rede.

Entretanto, essa diversidade estrutural introduz novos desafios relacionados à gestão de interferência, coordenação de recursos e balanceamento de carga. A complexidade de operar múltiplas camadas de células requer mecanismos de controle refinados e, muitas vezes, a adoção de soluções inteligentes. Em particular, o uso da comunicação direta entre dispositivos (*Device-to-Device* – *D2D*) em HetNets amplia ainda mais a complexidade do sistema, exigindo estratégias adaptativas de gerenciamento para garantir a qualidade e a estabilidade do serviço (LIU; FENG; CHEN, 2024).

De forma geral, o sucesso dessa estratégia depende da capacidade da rede em tomar decisões inteligentes e em tempo real, levando em consideração variáveis como interferência intercelular, potência do sinal recebido, disponibilidade de canais, perfil de mobilidade e prioridades de serviço. Em ambientes heterogêneos, como as HetNets, essas variáveis tornam-se ainda mais dinâmicas, exigindo soluções robustas baseadas em inteligência artificial e algoritmos adaptativos (LIU; FENG; CHEN, 2024).

Diante da crescente complexidade e densidade das redes móveis, esta pesquisa propõe um modelo inteligente de orquestração de recursos voltado para arquiteturas H-CRAN, com foco na alocação dinâmica de usuários por meio de estratégias de *offloading*. O objetivo é aprimorar o desempenho da rede em cenários de congestionamento, minimizando bloqueios, otimizando o uso dos recursos físicos disponíveis e garantindo os requisitos de Qualidade de Serviço (QoS), mesmo em ambientes com tráfego intenso e flutuante.

Para alcançar esses resultados, o modelo utiliza técnicas de inteligência computacional, aplicando três algoritmos bioinspirados reconhecidos na literatura — Particle

Swarm Optimization (PSO), Algoritmo Genético (AG) e Firefly Algorithm (FA). Esses métodos são empregados para resolver desafios como o mapeamento eficiente entre RRHs e BBUs, o *offloading* de usuários excedentes e a mitigação de gargalos em áreas críticas. Com isso, busca-se uma distribuição mais equilibrada e adaptativa dos recursos, contribuindo com soluções práticas e escaláveis para o avanço das redes móveis de próxima geração.

1.2 Objetivo Geral

Criar um framework capaz de realizar a orquestração de recursos em redes H-CRAN e propor métodos inteligentes para alocação de usuários e distribuição de recursos, levando em consideração diversos parâmetros, como os requisitos das aplicações, o comportamento dos usuários, a velocidade e a necessidade de tráfego de dados, bem como a distribuição das antenas.

1.3 Objetivos Específicos

- Estudar estratégias de orquestração de recursos entre usuários (UEs) atendidos por uma mesma RRH, com foco no equilíbrio e na eficiência do uso dos recursos locais;
- Investigar políticas de desconexão e redistribuição antecipada de usuários excedentes ou bloqueados, considerando diferentes perfis de carga e zonas de saturação;
- Analisar os desafios de redistribuição de recursos e mapeamento entre RRHs, BBUs e unidades centrais de processamento em cenários densamente povoados e com tráfego variável;
- Propor métodos inteligentes de mapeamento e reorquestração de usuários com base em *offloading* dinâmico e balanceamento de carga;
- Modelar o problema de alocação de recursos com múltiplos objetivos, incorporando métricas como QoS, SINR, disponibilidade espectral, prioridades de serviço e mobilidade;

- Realizar simulações com diferentes perfis de usuários, horários e padrões de tráfego, avaliando o impacto das estratégias propostas;
- Implementar e avaliar algoritmos bioinspirados — como PSO, AG e FA — quanto à sua eficácia na otimização da alocação e redistribuição de recursos na arquitetura H-CRAN.

1.4 Organização do Trabalho

Este estudo está estruturado da forma exibida a seguir:

- Capítulo 2: Apresenta a fundamentação teórica do trabalho, abordando de forma objetiva os principais temas relacionados à pesquisa, incluindo as características das novas arquiteturas de redes e os algoritmos de otimização mais recorrentes na literatura.
- Capítulo 3: Descreve os artigos científicos selecionados como base para a elaboração da tese, destacando suas contribuições e relevância para o contexto da pesquisa.
- Capítulo 4: Detalha o cenário estudado e a proposta inicial de mapeamento e reorquestração de recursos com base em técnicas de *offloading*.
- Capítulo 5: Apresenta as características dos cenários utilizados e os resultados obtidos a partir das simulações, com análise detalhada do desempenho da proposta em comparação com abordagens da literatura.
- Capítulo 6: Apresenta as possíveis contribuições da pesquisa e sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Considerações Iniciais

Com o avanço contínuo da demanda por serviços digitais de alto desempenho — a exemplo da realidade aumentada, de sistemas autônomos e da Internet das Coisas (IoT) —, as redes móveis enfrentam o desafio de evoluir de forma acelerada para atender a requisitos cada vez mais rigorosos em termos de latência, confiabilidade e capacidade.

Neste capítulo, apresenta-se o arcabouço teórico que fundamenta a presente pesquisa, contemplando desde os progressos recentes no contexto das redes 5G e a transição paradigmática rumo ao 6G até a análise das principais arquiteturas de rede emergentes, como as Redes Heterogêneas (HetNets), a *Cloud Radio Access Network* (C-RAN) e sua evolução, a *Heterogeneous Cloud RAN* (H-CRAN).

Adicionalmente, discute-se a adoção de estratégias de *offloading* como meio de alívio computacional e otimização do desempenho, bem como a aplicação de algoritmos bioinspirados no processo de alocação inteligente e dinâmica de recursos, os quais se mostram promissores diante da complexidade e variabilidade dos cenários em redes móveis de próxima geração.

2.2 Redes 5G e a Transição para o 6G

2.2.1 Evolução das Redes Móveis: Do 5G ao 6G

As redes de quinta geração (5G) representam um marco na evolução das telecomunicações, oferecendo maior largura de banda, menor latência e suporte massivo a dispositivos conectados. Segundo o (ERICSSON, 2024), espera-se que até 2030 o 5G represente mais de 67% das conexões móveis globais, sendo responsável por 80% do tráfego de dados. No entanto, limitações inerentes ao 5G em termos de latência extrema, conectividade ubíqua e eficiência espectral apontam para a necessidade de uma nova geração de redes: o 6G.

As redes 6G prometem velocidades de até 1 Tbit/s, latência inferior a 100 microssegundos e conexão para até 10 milhões de dispositivos por km² (ITU-R, 2023;

SAAD; BENNIS; CHEN, 2019). Além disso, destaca-se pela integração nativa com sistemas de inteligência artificial, permitindo operação adaptativa da infraestrutura e viabilizando aplicações como realidade estendida multissensorial e telepresença holográfica (SAHOO; TRIPATHY, 2025).

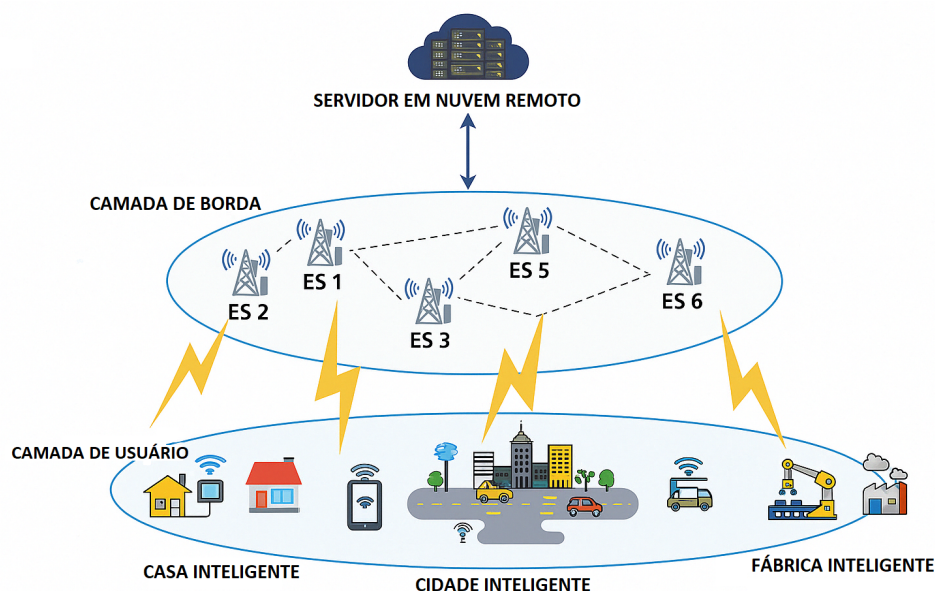


Figura 1 – Topologia do MEC em redes móveis 6G (adaptado de (SUN et al., 2024)).

No topo da arquitetura está o servidor em nuvem remoto, responsável por prover grande capacidade de processamento e armazenamento. No entanto, devido à sua distância dos usuários, o uso frequente da nuvem pode resultar em alta latência e congestionamento do *backbone* da rede, especialmente em cenários com alta densidade de dispositivos.

A camada intermediária (Camada de Borda) é composta por várias Edge Servers (ES) — ES1 a ES6 — cada um associado a uma Base Station (BS). Esses servidores têm capacidades computacionais limitadas, mas são colocados mais próximos dos usuários. O grande diferencial dessa proposta é que os ESs estão interconectados em uma rede local (LAN), permitindo cooperação entre bordas (*edge-edge cooperation*):

- Quando um ES está sobrecarregado, ele pode realizar *offloading* de tarefas para ESs vizinhos menos sobrecarregados.

- Caso o *offloading* entre bordas não seja suficiente, parte da carga pode ser enviada à nuvem remota.

Na base da topologia representada, na Camada do Usuário, identificam-se três domínios de aplicação característicos do ecossistema 6G: Casa Inteligente, Cidade Inteligente e Fábrica Inteligente. Esses ambientes concentram dispositivos móveis com capacidades computacionais e energéticas limitadas, mas que, ainda assim, executam tarefas com alta demanda de processamento, como reconhecimento facial, realidade aumentada e modelagem tridimensional. Diante dessas limitações locais, as tarefas são transferidas para os servidores de borda, com o objetivo de reduzir a latência, otimizar o uso da rede e assegurar os níveis de qualidade de serviço (QoS) exigidos por tais aplicações.

Segundo (SUN et al., 2024), a imagem representa uma proposta de arquitetura de *Multi-access Edge Computing* (MEC) ajustada às demandas das redes 6G. Nessa visão, o MEC deixa de ser uma simples extensão do 5G e passa a ocupar um papel central, focando em descentralização, inteligência distribuída e latência ultrabaixa. Essa estrutura é essencial para viabilizar aplicações como realidade estendida, cidades inteligentes e automação industrial, ao mesmo tempo em que promove menor latência, descongestionamento do núcleo da rede e maior escalabilidade.

2.3 Arquiteturas de redes móveis: HetNets, C-RAN e H-CRAN

2.3.1 HetNets

As Redes Heterogêneas (HetNets) referem-se à integração de diferentes tipos de células e tecnologias de acesso em uma única infraestrutura de rede, com o objetivo de aprimorar a cobertura, a capacidade e a eficiência das redes móveis, sobretudo em ambientes urbanos densamente povoados e espaços internos (XU et al., 2021). Essa arquitetura híbrida é composta por:

- Macro células: Estações base de alta potência que cobrem grandes áreas;
- Micro células, Picocélulas e Femtocélulas: Células de menor potência e alcance, utilizadas para densificar a rede e melhorar a cobertura em áreas específicas.

- Wi-Fi e LTE-U: incorporadas para auxiliar no offloading de tráfego, aumentando a experiência do usuário e diminuindo a carga sobre a rede celular.

Essa combinação permite o uso mais eficiente do espectro e da infraestrutura existente, possibilitando atender à crescente demanda por serviços de alta velocidade, baixa latência e alta confiabilidade — requisitos essenciais em aplicações modernas como realidade aumentada, comunicação máquina-a-máquina (M2M) e Internet das Coisas (IoT) (AKYILDIZ; KAK; NIE, 2020).

A estrutura das HetNets envolve o uso de diversos nós de acesso sem fio. Os elementos são classificados em dois grandes grupos: *High Power Nodes* (HPNs), representados tipicamente pelas macrocélulas; e *Low Power Nodes* (LPNs), que incluem micro, pico e femtocélulas — cada uma com sua própria *Base Station*, adaptada ao contexto de cobertura e densidade de usuários (TRABELSI; FOURATI; CHEN, 2024).

Esses elementos heterogêneos variam quanto à cobertura, potência de transmissão, tipo de backhaul, características de propagação e dimensões físicas. Essa diversidade arquitetônica oferece grande flexibilidade, mas também impõe desafios, como a coordenação de interferência intercelular, a gestão de mobilidade entre camadas e a orquestração eficiente dos recursos da rede.

2.3.2 C-RAN

A crescente demanda por conectividade de alta capacidade, baixa latência e eficiência energética impulsionou o surgimento de novas arquiteturas para redes móveis. No contexto da 5G — já consolidada em diversos países — uma das soluções mais promissoras é a Cloud Radio Access Network (C-RAN), que se destaca por oferecer escalabilidade, flexibilidade e redução significativa de custos operacionais. Em contraste com as redes legadas, a C-RAN promove a separação funcional entre os elementos de rádio e os elementos de processamento da estação base, centralizando o controle e otimizando o uso de recursos computacionais.

A figura abaixo destaca os principais componentes de uma arquitetura C-RAN:

- BBU Pool: um conjunto centralizado de BBUs, responsáveis pelo processamento digital dos sinais. Elas são virtualizadas e gerenciadas por um *Host Manager*,

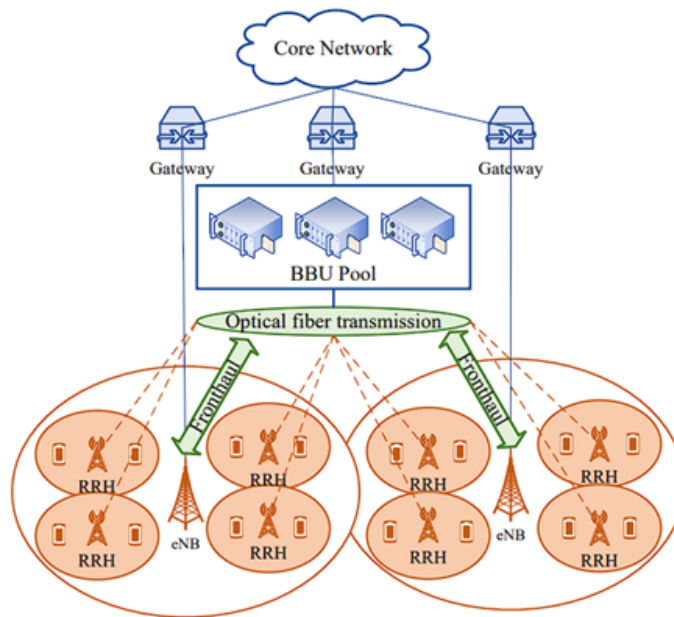


Figura 2 – Arquitetura CRAN [EJAZ et. al. 2020]

permitindo alocação dinâmica de recursos computacionais conforme a demanda.

- **RRHs:** distribuídas geograficamente, elas realizam a transmissão e recepção dos sinais de rádio com os usuários finais.
- **Fronthaul:** enlaces de comunicação de alta velocidade que conectam as RRHs às BBUs. Essa interface precisa garantir baixa latência e alta largura de banda, requisitos fundamentais para o bom desempenho da rede.

Conforme destacado por (CHEN, 2024), essa arquitetura centraliza as *Baseband Unit* (BBUs) em um pool centralizado, enquanto as *Remote Radio Head* (RRHs) permanecem distribuídas nas proximidades dos usuários finais, conectadas por enlaces de alta capacidade denominados *fronthaul*.

As RRHs, responsáveis pela emissão e recepção dos sinais de rádio, operam em conjunto com as BBUs, que realizam o processamento intensivo dos dados. Segundo o padrão IEEE 1914.1 (IEEE STANDARDS ASSOCIATION, 2023), o fronthaul que interliga esses componentes deve operar com baixa latência e alta capacidade de transmissão, sendo geralmente implementado com enlaces ópticos.

A separação funcional entre as RRHs e as BBUs na arquitetura C-RAN permite que as operadoras gerenciem de forma centralizada múltiplas RRHs, otimizando o uso de recursos computacionais e facilitando a manutenção da rede. De acordo com (CHEN, 2024), essa centralização das BBUs em um pool centralizado, conectadas às RRHs por enlaces de alta capacidade denominados fronthaul, proporciona uma gestão mais eficiente dos recursos e suporta os requisitos das redes 5G, além de ser compatível com infraestruturas existentes de 2G, 3G e 4G.

Além disso, a centralização das BBUs facilita a aplicação de técnicas avançadas, como coordenação de interferência e beamforming, essenciais para suportar os serviços críticos da 5G. A implementação da C-RAN também simplifica a expansão da rede, pois, conforme observado por (CHEN, 2024), permite a adição de novas RRHs sem a necessidade de implantar estações base completas, reduzindo custos e complexidade operacional (AZARIAH et al., 2024).

Portanto, a arquitetura C-RAN funciona por meio da separação entre a captação/transmissão de sinais pelas RRHs e o processamento centralizado no BBU Pool, interligados por um fronthaul robusto. Essa organização modular transforma a rede de acesso em um sistema mais adaptável, onde a capacidade computacional pode ser ajustada conforme a demanda, e o processamento pode ser virtualizado, integrado a infraestruturas em nuvem e gerenciado de forma centralizada.

2.3.3 H-CRAN

A arquitetura H-CRAN representa uma evolução significativa das redes móveis, integrando os benefícios da centralização do processamento da C-RAN com a flexibilidade e densidade das HetNets. Essa abordagem visa atender às crescentes demandas por capacidade, cobertura e eficiência energética (RAO; BABU, 2024).

Na estrutura do H-CRAN, as macrocélulas são responsáveis pelo plano de controle, gerenciando a sinalização e mantendo a conectividade em áreas amplas. Simultaneamente, as pequenas células (como pico e femtocélulas), conectadas ao pool de Unidades de Banda Base (BBUs) via *fronthaul* (com ou sem fio), lidam com o tráfego de dados dos usuários, proporcionando alta capacidade em regiões de alta densidade populacional. Essa separação funcional entre o plano de controle e o plano de usuário

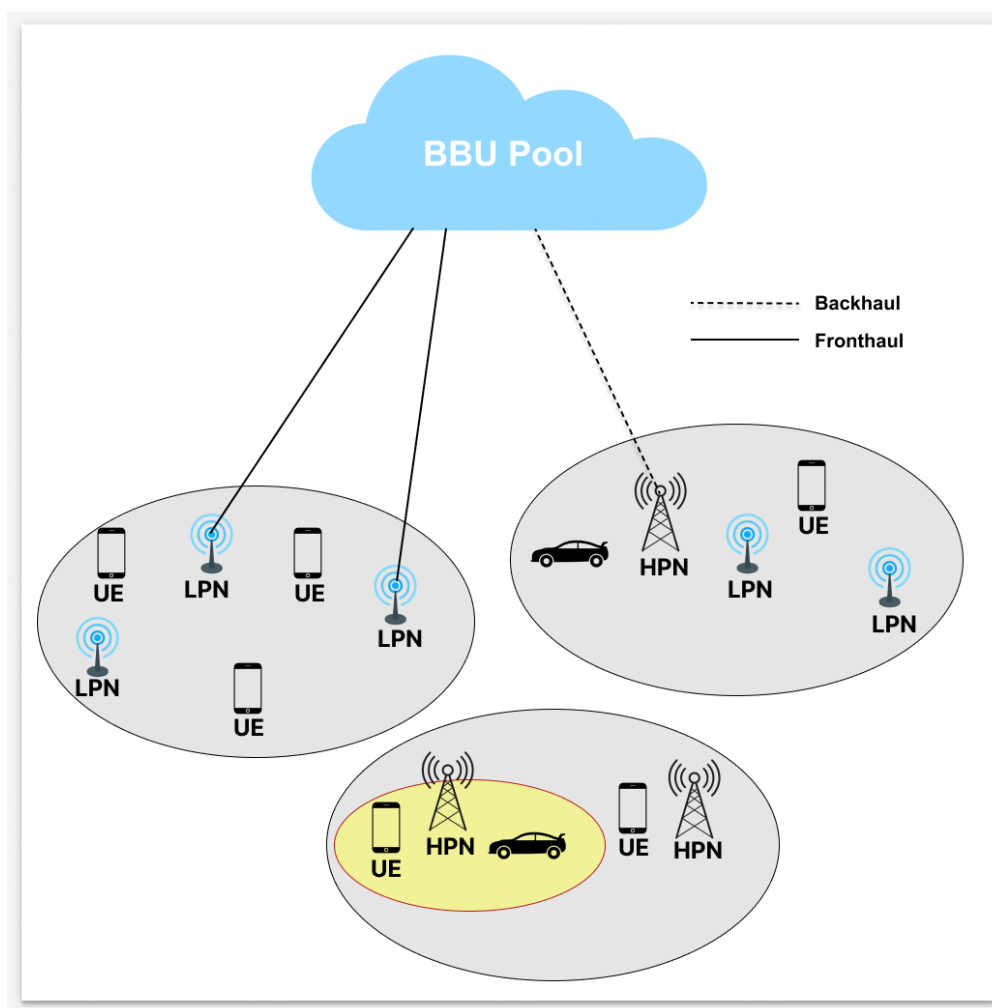


Figura 3 – Arquitetura H-CRAN

permite uma gestão mais eficiente dos recursos de rede, otimizando o desempenho e a utilização da infraestrutura existente (RAMYA; ARUNACHALAPERUMAL, 2024).

Além disso, o H-CRAN incorpora técnicas avançadas, como *Coordinated Multi-Point* (CoMP) e *beamforming*, para melhorar a eficiência espectral e a cobertura da rede. Essas técnicas são viabilizadas pela centralização do processamento no pool de BBUs, que permite uma coordenação mais eficaz entre as células da rede.

No entanto, a implementação do H-CRAN enfrenta desafios, especialmente relacionados à capacidade e latência do *fronthaul*. A necessidade de enlaces de alta capacidade e baixa latência entre as RRHs e o pool de BBUs é crítica para garantir o

desempenho da rede. Pesquisas recentes propõem soluções como a utilização de algoritmos de aprendizado de máquina para otimizar a alocação de recursos e melhorar a eficiência energética da rede (SINGH et al., 2025).

Em resumo, o H-CRAN oferece uma arquitetura promissora para as redes 5G, combinando a flexibilidade das HetNets com a eficiência da C-RAN. Sua capacidade de separar os planos de controle e usuário, juntamente com a centralização do processamento e a utilização de técnicas avançadas de coordenação, posiciona o H-CRAN como uma solução eficaz para atender às exigências das redes móveis de próxima geração.

Embora este trabalho adote a arquitetura H-CRAN como base conceitual e de implementação, é importante reconhecer que a evolução recente das redes de acesso tem se direcionado fortemente para o paradigma Open RAN (*Open Radio Access Network*). Essa arquitetura virtualiza e desagrega funções tradicionalmente concentradas na *Baseband Unit* (BBU) em módulos como a *Central Unit* (CU) e a *Distributed Unit* (DU), interligados por interfaces abertas e padronizadas. Tal abordagem favorece a interoperabilidade entre fornecedores, aumenta a flexibilidade operacional e facilita a integração de algoritmos inteligentes. Assim, ainda que a presente proposta mantenha a nomenclatura e estrutura lógica da H-CRAN, os métodos e estratégias desenvolvidos são totalmente compatíveis e adaptáveis ao modelo Open RAN, assegurando sua aplicabilidade em redes 5G e além.

2.3.4 O-RAN: Arquitetura Modular e Inteligente para Redes 6G

A arquitetura O-RAN representa uma evolução significativa no conceito de redes de acesso sem fio, incorporando os princípios de desagregação, virtualização e inteligência nativa de gerações anteriores de RAN. Seu principal diferencial é a padronização de interfaces abertas, permitindo a interoperabilidade entre equipamentos de diferentes fornecedores, o que amplia a flexibilidade e reduz custos de implantação e operação (BOUKHATEM et al., 2025).

Na prática, esta arquitetura desagrega a estação base tradicional em três unidades principais (LARSEN et al., 2024):

- **RU (*Radio Unit*)**: responsável pelo processamento de sinais de rádio e interface direta com as antenas.

- **DU (*Distributed Unit*)**: executa funções em tempo real das camadas PHY e MAC.
- **CU (*Central Unit*)**: gerencia funções de controle e sinalização, como PDCP e RRC, que não exigem operação em tempo real.

Além dessas unidades, o O-RAN introduz controladores inteligentes com funções específicas:

- **Near-RT RIC**: atua no controle quase em tempo real, permitindo decisões rápidas de alocação de recursos, coordenação de interferência e ajuste dinâmico de parâmetros de rede.
- **Non-RT RIC**: opera em tempo não real, realizando o gerenciamento de políticas de longo prazo, análise de dados históricos, aprendizado de máquina e orquestração global da rede.

A comunicação entre essas entidades ocorre por meio de interfaces padronizadas definidas pelo consórcio O-RAN Alliance:

- **Fronthaul (RU–DU)**: transporta sinais I/Q e sinalização em tempo real.
- **Midhaul (DU–CU)** — também chamada de *interface F1*: lida com o transporte da pilha aérea dividida entre as duas unidades.
- **E2 (RIC–RU/DU/CU)**: possibilita que o RIC monitore e controle o comportamento da rede em diferentes camadas, suportando a integração de módulos inteligentes de terceiros.

O O-RAN é visto como um pilar para as redes 6G, pois sua modularidade, somada à capacidade de integração de inteligência artificial e aprendizado de máquina, permite a implementação de estratégias avançadas de otimização e automação, alinhando-se diretamente às demandas por baixa latência, alta confiabilidade e eficiência espectral nas futuras redes móveis.

2.4 Offload em Redes Móveis

O *offloading* computacional em redes móveis refere-se à transferência de tarefas de processamento intensivo de dispositivos móveis para servidores mais potentes, localizados na *Multi-access Edge Computing* (MEC) ou na nuvem. Essa técnica visa otimizar o uso de recursos, melhorar o desempenho e reduzir o consumo de energia dos dispositivos móveis.

Segundo (ALGHAMDI, 2021) arquitetura de *offloading* computacional geralmente envolve três componentes principais:

- Dispositivo Móvel: Origina a tarefa de processamento e decide se deve executá-la localmente ou realizar o descarregamento (*offloading*).
- Servidor de Borda (MEC): Localizado próximo ao usuário, oferece recursos de computação e armazenamento para processar tarefas offloadadas com baixa latência.
- Nuvem Central: Fornece recursos computacionais escaláveis para tarefas que não podem ser processadas na borda devido a restrições de recursos ou políticas de segurança.

No entanto, este trabalho adota uma abordagem centrada na rede (*network-centric offloading*), na qual a decisão de redistribuição de carga não parte do dispositivo móvel, mas sim da própria infraestrutura da rede — mais especificamente, de entidades de controle localizadas na nuvem de BBUs da arquitetura H-CRAN. Esse tipo de abordagem baseia-se no princípio de que a rede, por dispor de uma visão global do sistema — incluindo o estado de ocupação das RRHs, a interferência entre células, os requisitos de QoS e os padrões de mobilidade dos usuários —, é capaz de tomar decisões mais eficientes e coordenadas sobre onde e quando realizar o *offloading*.

Ao contrário das abordagens *user-centric*, em que os dispositivos móveis decidem individualmente quando realizar a transferência de tarefas com base em condições locais, o *network-centric offloading* permite a orquestração otimizada e escalável de recursos em cenários densos e dinâmicos. Dessa forma, embora os elementos descritos por (ALGHAMDI, 2021) estejam presentes, a lógica de decisão adotada neste trabalho

está fundamentada em uma perspectiva de controle centralizado e inteligente, condizente com a proposta de orquestração de recursos em redes H-CRAN.

2.5 Inteligência Computacional

A Inteligência Computacional consiste em um conjunto de conceitos, paradigmas, algoritmos e implementações de sistemas que, em teoria, exibem comportamentos inteligentes ao operar em ambientes complexos. Essa área é fortemente sustentada por abordagens sub-simbólicas, predominantemente baseadas em analogias com sistemas naturais ou inspiradas pela natureza. Tais métodos apresentam como principal vantagem a tolerância à incerteza e ao conhecimento incompleto, o que permite a obtenção de soluções aproximadas, porém viáveis e robustas ao mesmo tempo (KENNEDY; EBERHART, 1995).

Caracterizada por sua flexibilidade e capacidade de adaptação, a Inteligência Computacional oferece um arcabouço de técnicas bioinspiradas voltadas à resolução de problemas complexos, especialmente aqueles com alto grau de não linearidade, incerteza e grande dimensionalidade. Esses algoritmos são projetados para explorar a imprecisão e a adaptabilidade observadas nos sistemas vivos, tornando-se particularmente eficazes em cenários onde métodos tradicionais de otimização e análise apresentam desempenho limitado ou falham (GOLDBERG, 1989).

Dentre as principais técnicas da Inteligência Computacional, destacam-se o PSO, o AG e o FA, inspirados em comportamentos coletivos ou fenômenos naturais para guiar a busca por soluções ótimas em espaços complexos de decisão (LI; ZHU, 2020).

- PSO: inspirado no comportamento coletivo de aves ou cardumes, esse algoritmo busca soluções ótimas por meio da interação entre partículas, que ajustam suas posições com base em suas experiências individuais e no desempenho do grupo;
- AG: simula os mecanismos da evolução biológica — como seleção natural, cruzamento e mutação — para gerar soluções progressivamente mais eficientes ao longo de várias gerações;
- FA: baseia-se no comportamento de atração dos vaga-lumes, que se deslocam em direção aos mais "brilhantes", guiando a busca por soluções de melhor qualidade

no espaço de decisão.

Esses algoritmos têm sido amplamente utilizados em problemas de otimização de recursos, alocação dinâmica, engenharia de tráfego em redes móveis, planejamento inteligente, além de aplicações em aprendizado de máquina e ciência de dados.

Dada a relevância e a ampla aplicabilidade dos algoritmos de Inteligência Computacional, torna-se essencial compreender seus princípios operacionais e particularidades. Embora compartilhem características como adaptabilidade, paralelismo implícito e robustez frente à incerteza, cada técnica apresenta estratégias distintas de exploração e intensificação do espaço de busca.

A seguir, cada um dos algoritmos será abordado em maior profundidade, destacando suas inspirações naturais, mecanismos de funcionamento, vantagens e limitações, bem como suas aplicações recorrentes nos contextos de redes móveis, otimização de recursos e engenharia de sistemas complexos.

2.5.1 PSO

O algoritmo PSO foi introduzido em 1995 por Kennedy e Eberhart (KENNEDY; EBERHART, 1995) com o objetivo de solucionar problemas de otimização. Dentre os algoritmos evolucionários, o PSO destaca-se por sua simplicidade e desempenho satisfatório em uma ampla variedade de aplicações (SCOTT-HAYWARD; NATARAJAN; SEZER, 2015).

Esse algoritmo é um método de otimização inspirado em habilidades de organização social, utilizadas para enfrentar e resolver problemas complexos por meio da interação entre a capacidade individual dos membros e a dinâmica social do grupo (ROKBANI; ALIM, 2013). Essa busca pode ser exemplificada por indivíduos explorando o ambiente ao redor em busca de alimento.

Um grupo de indivíduos, denominado enxame, composto por partículas com comportamentos simples, é capaz de formar uma inteligência social complexa por meio de regras iterativas básicas. Quando uma partícula encontra uma fonte promissora, como alimento, essa informação é compartilhada com os demais membros do grupo, influenciando o movimento coletivo das partículas em direção àquela posição, desde que o

recurso encontrado seja suficientemente atrativo (ENGELBRECHT, 2005).

Cada partícula pode ser informada tanto sobre as partículas vizinhas quanto sobre o desempenho global do enxame. A partícula com melhor solução na vizinhança e aquela que obteve o melhor resultado global ao longo do tempo influenciam o deslocamento das demais (ROKBANI; ALIM, 2013).

Esse método de busca pode ser interpretado como uma combinação entre uma busca baseada em gradiente — como a subida de encosta ou o gradiente descendente — e uma abordagem baseada em população, como os algoritmos evolucionários.

Pode-se dizer que o PSO explora o espaço de uma função objetivo ajustando as partículas, que percorrem caminhos definidos por vetores de posição, em busca da melhor solução. Essa técnica tem sido amplamente adotada devido à sua baixa complexidade computacional e rápida convergência.

2.5.1.1 Algoritmo

Algoritmo 1: PSO	
1	início
2	Inicializa as partículas aleatoriamente: $Populacao_{posicao} \leftarrow rand()$
3	Inicializa as velocidades aleatoriamente: $Populacao_{velocidade} \leftarrow rand()$
4	Avalia a Populacao: $Populacao_{fitness} \leftarrow f(Populacao)$
5	Atribui a P_{best} e G_{best} a partícula com melhor resultado:
6	$P_{best} \leftarrow Populacao$
7	$G_{best} \leftarrow max(P_{best})$
8	enquanto $numDeGeraes < maxInteracoes$ faça
9	para cada $P \in Populacao$ faça
10	$P_{velocidade} \leftarrow atualizaVelocidade(P_{velocidade}, P_{best}, G_{best})$
11	$P_{posicao} \leftarrow atualizaPosicao(P_{posicao}, P_{velocidade})$
12	se $f(P) > f(P_{best})$ então
13	$P_{best} \leftarrow P$
14	fimse
15	se $f(P) > f(G_{best})$ então
16	$G_{best} \leftarrow P$
17	fimse
18	fim
19	fim
20	Retorna G_{best}
21	fim

Um PSO genérico segue os passos descritos no Algoritmo 1. O processo inicia-se com a inicialização das partículas, em que tanto suas posições quanto suas velocidades são atribuídas aleatoriamente. Em seguida, cada partícula é avaliada de acordo com a função objetivo. Inicialmente, a partícula que obtiver o melhor resultado será atribuída como melhor global (G_{best}) e também como melhor local (P_{best}) de sua respectiva posição.

As variáveis de memória G_{best} e P_{best} armazenam, respectivamente, a melhor posição global encontrada por qualquer partícula ao longo de todas as iterações, e a

melhor posição individual que cada partícula já obteve.

Após a etapa de inicialização, iniciam-se as iterações do algoritmo. A cada iteração, a velocidade de cada partícula é atualizada com base em sua melhor posição individual ($Pbest$) e na melhor posição global ($Gbest$). Com a nova velocidade, as partículas têm suas posições atualizadas, sendo novamente avaliadas de acordo com a função objetivo. Se a nova posição resultar em um valor melhor que o armazenado em $Pbest$, essa variável é atualizada. Da mesma forma, se qualquer partícula atingir um valor melhor do que o atual $Gbest$, este também é atualizado.

Ao final do processo, a partícula associada ao melhor valor encontrado é retornada como solução. No Algoritmo 1, o critério de parada utilizado é o número fixo de iterações. No entanto, outras estratégias podem ser aplicadas, como alcançar um valor objetivo mínimo, atingir um limite de tempo ou detectar a estabilização das soluções (por exemplo, gradiente abaixo de um limite).

2.5.1.2 Equações

A posição e a velocidade de cada partícula é determinada a partir das equações (2.1) e (2.2).

$$\vec{V}_{t+1}^p = w \cdot \vec{V}_t^p + c_1 \cdot rand \left(\vec{X}^{Pbest} - \vec{X}_t^p \right) + c_2 \cdot rand \left(\vec{X}^{Gbest} - \vec{X}_t^p \right) \quad (1)$$

$$\vec{X}_{t+1}^p = \vec{X}_t^p + \vec{V}_{t+1}^p \quad (2)$$

Tais equações geram a nova velocidade e posição da partícula, respectivamente. Na equação (2.1) se tem o novo valor da velocidade $t+1$ da partícula p (velocidade atualizada) baseada na velocidade atual ponderada de uma inércia w , recebendo a influência do $Pbest$ e do $Gbest$ de acordo com a distância destes para a atual da partícula X_t^p para elas.

Os valores de c_1 e c_2 representam uma força de “atração” das soluções $Pbest$ e $Gbest$, e são escolhidos de acordo com uma aleatoriedade até um limiar. A equação (2.2) representa simplesmente a atualização da posição da partícula, que se torna a posição atual com a influência da nova velocidade.

2.5.1.3 Variações

Há várias extensões e modificações do PSO básico. Se o espaço for limitado, as partículas podem ser refletidas no limite do espaço ao invés de simplesmente terem sua posição limitada. Em alguma variante poderia haver uma adaptação dinâmica, como uma partícula que tem seu valor significativamente menor que as partículas na vizinhança ser removida (ROKBANI; ALIM, 2013). Outra forma seria haver um foco maior no local de uma partícula, substituindo o *Gbest* por um melhor local na vizinhança da partícula, com isso há uma maior diversificação e ganho no espaço busca.

Outra variação na estrutura é o MOPSO (*Multiple Objective Particle Swarm Optimization*) [COELLO 2002] que busca atender múltiplos objetivos, gerando uma família de resultados através da dominância de Pareto. O PSO possui variações que em alguns casos são necessárias para resolver os problemas em questão, entretanto para este trabalho o PSO original é o suficiente para fornecer os resultados esperados.

2.5.2 AG

É uma técnica de otimização inspirada no processo de evolução natural, introduzida por John Holland em 1975, e desde então tem sido amplamente aplicada na resolução de problemas complexos e multidimensionais. No contexto das redes móveis, o AG tem demonstrado eficácia em tarefas como balanceamento de carga, alocação de recursos e otimização de rotas, devido à sua capacidade de explorar grandes espaços de busca de forma adaptativa (GOLDBERG, 1989).

Nas redes móveis de quinta e sexta geração (5G e 6G), o AG tem sido empregado para encontrar soluções ótimas em problemas como orquestração de serviços na borda, gerenciamento dinâmico de espectro e redução de latência (LI; ZHU, 2020). Segundo (WANG; ZHANG; LIN, 2022), o AG mostra-se eficiente na adaptação às flutuações de tráfego e à variabilidade dos canais de comunicação, graças à sua natureza estocástica e à capacidade de convergir para ótimos globais mesmo em ambientes instáveis e heterogêneos (WANG, R.; WEN, Xun et al., 2023).

2.5.2.1 Funcionamento do AG

A estrutura básica do AG envolve as seguintes etapas:

1. **Codificação:** representação das soluções possíveis por cromossomos, geralmente na forma de vetores binários ou reais.
2. **População inicial:** geração aleatória de um conjunto de soluções iniciais (indivíduos).
3. **Avaliação:** cada indivíduo é avaliado por uma função de aptidão (fitness), que determina a qualidade da solução representada.
4. **Seleção:** os indivíduos com maior aptidão são selecionados para reprodução, utilizando métodos como roleta, torneio ou ranking.
5. **Cruzamento (crossover):** dois cromossomos são combinados, trocando partes de sua estrutura para gerar novos indivíduos com diversidade genética.
6. **Mutação:** modificação aleatória de genes em alguns indivíduos, com o objetivo de evitar a convergência prematura da população.
7. **Substituição:** os novos indivíduos gerados substituem parcial ou totalmente a população anterior, compondo a próxima geração.

Algoritmo 2: Algoritmo Genético
--

<pre> 1 Inicialização do algoritmo 2 Gerar população inicial aleatória: $Populacao \leftarrow rand()$ 3 Avaliar população: $Populacao_{fitness} \leftarrow f(Populacao)$ 4 enquanto $numGeracoes < maxInteracoes$ faça 5 Selecionar pais: $Pais \leftarrow selecao(Populacao)$ 6 Realizar cruzamento: $Filhos \leftarrow crossover(Pais)$ 7 Aplicar mutação: $Filhos \leftarrow mutacao(Filhos)$ 8 Avaliar filhos: $Filhos_{fitness} \leftarrow f(Filhos)$ 9 Atualizar população: $Populacao \leftarrow substituicao(Populacao, Filhos)$ 10 fimenquanto 11 Retorna <i>melhor indivíduo de Populacao</i> </pre>

Como ilustrado anteriormente, o AG é uma técnica de otimização inspirada nos mecanismos da evolução natural, como seleção, cruzamento e mutação. Sua lógica baseia-se na premissa de que, ao longo das gerações, indivíduos mais aptos tendem

a gerar descendentes com melhor desempenho, promovendo um processo adaptativo que converge para soluções cada vez mais adequadas ao problema. A execução do AG inicia-se com a geração de uma população inicial aleatória de soluções, representadas por cromossomos — comumente vetores binários ou de números reais. Em seguida, cada indivíduo é avaliado por uma função de aptidão (*fitness function*), responsável por quantificar sua qualidade com base nos critérios definidos pelo problema em análise (HOLLAND, 1975; ZHANG; SHU; NAVAEI, 2022).

Após a avaliação, os indivíduos com melhor desempenho são selecionados para aplicação dos operadores genéticos. A seleção pode ser realizada por métodos como roleta, torneio ou ranking. O cruzamento (*crossover*) combina pares de cromossomos para gerar novos indivíduos, promovendo a recombinação genética. A mutação, por sua vez, introduz variações aleatórias em determinados genes, garantindo diversidade na população e contribuindo para evitar a convergência prematura a soluções subótimas (SIVANANDAM; DEEPA, 2008).

Os novos indivíduos gerados passam por nova avaliação e, conforme um critério de substituição, são incorporados à próxima geração. Esse ciclo se repete por um número pré-definido de gerações, ou até que se atinja uma condição de parada. Segundo (SABER et al., 2021), essa abordagem tem se mostrado altamente eficaz na resolução de problemas combinatórios e na alocação de recursos em sistemas dinâmicos, como redes móveis.

2.5.3 FA

O FA foi proposto por Xin-She Yang em 2008 e baseia-se no comportamento natural dos vaga-lumes, que utilizam sinais luminosos para atrair outros indivíduos da espécie. A premissa central do algoritmo é simples: vaga-lumes menos brilhantes movem-se em direção àqueles mais brilhantes, sendo o brilho diretamente proporcional ao valor da função objetivo empregada na avaliação da solução (YANG, 2008).

O FA é uma meta-heurística bioinspirada cuja dinâmica populacional simula o comportamento de vaga-lumes explorando um espaço de busca multidimensional. Conforme ilustrado no Algoritmo 3, sua execução inicia-se com a geração aleatória de uma população de vaga-lumes, onde cada indivíduo representa uma solução candidata para

o problema de otimização. Em seguida, cada vaga-lume é avaliado por uma função de aptidão (*fitness*), que determina seu nível de "brilho" com base na qualidade da solução representada.

Algoritmo 3: FA

```

0: inicio
0: Inicializar população de vaga-lumes:  $Populacao \leftarrow rand()$ 
0: Avaliar brilho (aptidão) de cada vaga-lume:  $Populacao_{fitness} \leftarrow f(Populacao)$ 
  while  $numDeGeracoes < maxIteracoes$  do

    end
     $i \in Populacao$  forall  $j \in Populacao$  do

      end
       $f_j > f_i$ 
0: Calcular distância  $r_{ij}$  entre  $i$  e  $j$ 
0: Mover  $i$  em direção a  $j$ :

      
$$x_i \leftarrow x_i + \beta_0 e^{-\gamma r_{ij}^2} (x_j - x_i) + \alpha \epsilon$$


0: Avaliar novo brilho:  $f_i \leftarrow f(x_i)$ 
0: Retorna melhor vaga-lume (solução)
0: fim = 0

```

Durante as iterações, cada vaga-lume é comparado com os demais da população. Quando um indivíduo encontra outro com brilho superior, ele calcula a distância euclidiana entre ambos e ajusta sua posição, movendo-se em direção ao mais brilhante. Esse deslocamento é dado pela seguinte equação:

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2} (x_j^t - x_i^t) + \alpha \epsilon_i$$

em que:

- β_0 representa o coeficiente de atratividade inicial;
- γ é o coeficiente de absorção de luz;
- r_{ij} denota a distância euclidiana entre os vaga-lumes i e j ;

- α é o parâmetro de aleatoriedade;
- ε_i representa um ruído aleatório com distribuição gaussiana.

Após cada movimentação, o brilho de cada vaga-lume é reavaliado, e o processo se repete até que um critério de parada seja satisfeito.

Em aplicações voltadas para redes móveis, o FA tem sido utilizado com sucesso na minimização da interferência intercelular, na otimização da localização de unidades de rádio remotas (RRHs) e na redução do tempo de resposta em ambientes de borda. Segundo (ALSADIE; ALSULAMI, 2024), o algoritmo demonstra capacidade de adaptação a cenários com alta mobilidade, mantendo sua eficácia mesmo quando os recursos da rede são limitados e o tráfego apresenta alta variabilidade.

2.6 Considerações Finais

Neste capítulo, foram apresentados os principais fundamentos teóricos que sustentam esta pesquisa. Discutiu-se, inicialmente, a evolução das redes móveis, com foco na transição do 5G para o 6G, destacando os requisitos emergentes de baixa latência, alta capacidade e conectividade massiva.

Na sequência, abordaram-se as arquiteturas de redes móveis HetNets, C-RAN e H-CRAN, ressaltando suas características estruturais, vantagens operacionais e limitações. Também foi discutido o papel do *offloading* na redistribuição da carga de usuários como estratégia essencial para melhorar o desempenho de redes densas.

Por fim, o capítulo apresentou as bases da *inteligência computacional* aplicadas ao contexto de redes móveis, detalhando o funcionamento e as aplicações dos algoritmos PSO, AG e FA. Tais algoritmos se mostram centrais para esta pesquisa, pois oferecem suporte à orquestração dinâmica e inteligente de recursos em ambientes complexos e altamente variáveis.

Essas discussões teóricas constituem o alicerce conceitual para o modelo proposto no capítulo seguinte, o qual integra esses elementos em uma solução otimizada para redes H-CRAN, com foco na alocação eficiente de usuários por meio de estratégias de *offloading*.

3 TRABALHOS CORRELATOS

3.1 Considerações Iniciais

Este capítulo reúne os principais estudos que tratam da alocação e orquestração de recursos em redes móveis, especialmente em arquiteturas C-RAN e H-CRAN. A seleção dos trabalhos levou em conta abordagens recentes, com foco em algoritmos de otimização, técnicas de *offloading*, mapeamento dinâmico e adaptação a cenários de tráfego variável.

A análise busca destacar o que já foi explorado na literatura e, principalmente, o que ainda permanece como desafio. Ao observar essas lacunas, reforça-se a motivação da proposta desta tese, que surge justamente como uma resposta integrada a essas limitações.

3.2 Trabalhos Correlatos

A utilização de arquiteturas emergentes, como HetNets e H-CRAN — caracterizadas pela heterogeneidade de células e múltiplos níveis de acesso — tem sido amplamente apontada como uma solução promissora para atender às exigências das redes móveis de quinta e sexta gerações. Além disso, arquiteturas centralizadas como a C-RAN também ganham destaque por sua capacidade de simplificar a infraestrutura e reduzir custos. Apesar do potencial dessas arquiteturas, a implantação desordenada de pequenas células e a má orquestração ainda geram desafios como interferência e aumento de custos, o que tem impulsionado novas propostas na literatura para otimizar a infraestrutura de rede.

Entre as propostas voltadas à otimização de recursos em ambientes de computação em borda, destaca-se o trabalho de (LI; ZHU, 2020), que trata o *offloading parcial* em redes móveis com múltiplos usuários conectados a um único servidor MEC. Utilizando um algoritmo genético, os autores buscam otimizar, de forma conjunta, o percentual de *offloading* das tarefas, a largura de banda e os recursos computacionais disponíveis, com o objetivo de reduzir o tempo total de execução. O modelo apresen-

tado se destaca pela representação genotípica de estratégias e pela abordagem integrada de alocação. No entanto, a solução assume baixa mobilidade e topologias simplificadas. Avançando nesse cenário, a presente tese propõe uma estrutura mais ampla e adaptativa para redes H-CRAN, combinando múltiplos algoritmos bioinspirados (PSO, AG e FA) em ambientes dinâmicos, com múltiplos pontos de acesso, mobilidade elevada e interferência intercelular, ampliando significativamente o escopo e a aplicabilidade da orquestração de recursos.

Já (WANG, R.; WEN, Xun et al., 2023) desenvolveram um algoritmo baseado em PSO para lidar, de forma simultânea, com a alocação de potência e o controle de fase em sistemas D2D com suporte de superfícies refletoras inteligentes (IRS). O foco do estudo foi maximizar a taxa de transmissão em ambientes onde múltiplos usuários compartilham subcanais de uma mesma rede celular. O diferencial da proposta está na forma como o problema, embora complexo e não linear, é resolvido sem precisar ser decomposto, permitindo ganhos relevantes em eficiência energética e desempenho. Já esta tese, apesar de também utilizar algoritmos bioinspirados como o PSO, segue por um caminho mais abrangente: propõe um modelo de orquestração inteligente para redes H-CRAN, capaz de integrar diferentes tipos de células, estratégias de offloading e múltiplos critérios simultâneos de decisão em tempo real.

O trabalho de (YOU; TANG, 2021) propõe uma estratégia de *offloading* computacional em ambientes com computação em borda (MEC), utilizando o algoritmo PSO para otimizar o desempenho em cenários de Internet das Coisas Industrial (IIoT). A proposta foca na redução da latência e do consumo energético ao distribuir tarefas de dispositivos inteligentes para servidores próximos, com uma formulação multiobjetivo que combina critérios como atraso, custo e eficiência. Entre os destaques, estão a adaptação do PSO para contextos heterogêneos e a comparação com outros algoritmos bioinspirados. Apesar da proximidade conceitual, o estudo se limita ao ambiente industrial e à alocação de tarefas em MEC. Já esta tese amplia o escopo ao abordar a orquestração adaptativa de recursos em redes H-CRAN, considerando variabilidade de tráfego, múltiplos níveis de acesso e técnicas de descentralização com inteligência distribuída.

(CHOUDHARY et al., 2024) propõem um sistema multiagente bioinspirado para gerenciamento distribuído de potência e interferência em redes MIMO-OFDM.

A proposta se destaca pela aplicação de inteligência de enxame a um modelo descentralizado, onde agentes autônomos tomam decisões locais e cooperam para otimizar o desempenho global, mesmo em ambientes com múltiplos usuários e canais interferentes. Entre as contribuições, estão o controle distribuído de potência, o aumento na taxa de transmissão com menor latência e o balanceamento de carga sensível à interferência. Já a presente tese segue um caminho diferente: propõe uma arquitetura centralizada voltada a redes H-CRAN, com foco na alocação dinâmica de usuários por meio de estratégias de *offloading* e no uso combinado dos algoritmos PSO, AG e FA.

Já (NIVETHA; PREETHA, 2024) propõem uma abordagem meta-algorítmica para alocação otimizada de potência em redes cooperativas da sexta geração (6G), com base no conceito de *cybertwins* — entidades digitais inteligentes capazes de representar usuários e adaptar decisões de rede de forma autônoma. A proposta combina algoritmos como o PSO com técnicas de aprendizado profundo para prever padrões de tráfego e ajustar a alocação de potência conforme a mobilidade dos usuários. Além disso, utiliza *digital twins* (modelos digitais que espelham elementos físicos da rede) como suporte à tomada de decisão preditiva. Embora também faça uso de algoritmos bioinspirados, esta tese vai além da alocação de potência ao propor uma solução integrada de orquestração para redes H-CRAN, abrangendo *offloading* dinâmico, mapeamento RRH–BBU e suporte a múltiplas tecnologias de acesso em cenários densos e heterogêneos.

Entre os trabalhos recentes, destaca-se o estudo de (SILVA GONÇALVES et al., 2023), que propõe um modelo de mapeamento dinâmico entre unidades de rádio (RRHs) e unidades de banda base (BBUs) em redes C-RAN, utilizando o algoritmo bioinspirado FA para otimizar o balanceamento de carga diante da variabilidade de tráfego (efeito maré). Embora os resultados demonstrem melhorias significativas no desempenho, a proposta é limitada a redes C-RAN e não contempla aspectos fundamentais para as redes móveis de próxima geração, como *offloading*, mobilidade e múltiplas métricas simultâneas de qualidade de serviço.

Ainda no contexto da alocação eficiente de recursos, destaca-se o algoritmo alocação eficiente de recursos em redes H-CRAN, que propõe o algoritmo EE H-CRAM (SAHU; SAHU, 2024), que propõe uma abordagem voltada à eficiência energética para o gerenciamento de RRHs e BBUs, com foco na economia de energia e no uso de infraestrutura de borda (MEC). Apesar dos resultados promissores na redução do con-

sumo energético, o escopo do trabalho restringe-se à análise de fatores isolados, como a energia residual dos nós e a distância dos usuários. Em contrapartida, esta tese propõe uma orquestração mais ampla e adaptativa, incorporando o mapeamento dinâmico entre RRHs e setores de BBUs, a redistribuição de usuários e estratégias de *offloading* inteligente. Além disso, ao invés de adotar uma única abordagem algorítmica, realiza-se aqui uma avaliação comparativa entre três algoritmos bioinspirados — PSO, Algoritmo Genético e FA — aplicados a cenários heterogêneos com múltiplos critérios simultâneos de decisão.

No âmbito específico das redes H-CRAN, destaca-se também o trabalho de (GONÇALVES et al., 2019), que introduz o modelo Two-Level Allocation (TLA), baseado no algoritmo Particle Swarm Optimization (PSO). Esse modelo realiza o redirecionamento de usuários excedentes entre RRHs e eNodeBs, considerando requisitos mínimos de QoS e zonas de saturação. O estudo demonstra que o uso de *offloading* contribui para a redução do número de usuários bloqueados e melhora o *throughput* da rede. Dada sua relevância e estrutura consolidada, o modelo TLA será utilizado nesta pesquisa como benchmark comparativo para avaliação de desempenho.

Apesar dos avanços apresentados nesses trabalhos, algumas lacunas permanecem abertas, especialmente no que se refere à orquestração inteligente e adaptativa dos recursos em tempo real em ambientes heterogêneos. Alinhada a essas necessidades, a presente tese propõe o desenvolvimento de um framework para orquestração inteligente de recursos em redes H-CRAN, com foco na otimização da alocação dinâmica de RRHs e redistribuição de usuários, considerando cenários com tráfego flutuante, perfis variados de mobilidade e exigências diferenciadas de QoS.

A proposta distingue-se por adotar uma abordagem multiobjetivo e multi-algoritmo, utilizando e comparando os algoritmos bioinspirados PSO, AG e FA. Cada um desses métodos será aplicado a cenários realistas de alta densidade populacional e mobilidade variável, avaliando parâmetros como interferência, disponibilidade de canais, distribuição física da infraestrutura e requisitos das aplicações. Tal abordagem atende diretamente ao objetivo geral de orquestrar de forma inteligente os recursos em redes H-CRAN, bem como aos objetivos específicos da modelagem do problema, estudo de políticas de *offloading* e avaliação do desempenho em simulações controladas.

3.3 Considerações Finais

Esta proposta de tese tem como objetivo investigar, de forma abrangente, os métodos de alocação de recursos em redes H-CRAN, com foco na redistribuição adaptativa de usuários, no mapeamento dinâmico entre RRHs e setores das BBUs e na orquestração inteligente dos recursos frente às múltiplas variáveis que impactam a qualidade da rede. Para isso, são exploradas soluções baseadas em algoritmos bioinspirados, de forma comparativa, com o intuito de construir um framework escalável, robusto que atenda às exigências das futuras redes 6G.

A análise dos trabalhos correlatos evidencia que, embora os problemas de alocação e redistribuição de recursos tenham sido amplamente abordados, a maioria das soluções existentes trata esses aspectos de maneira isolada. Modelos como o TLA demonstraram avanços importantes, mas não consideram integralmente a complexidade de ambientes altamente dinâmicos, onde fatores como mobilidade, QoS, interferência e topologia devem ser considerados em conjunto. Além disso, condições como o efeito de maré — que refletem variações sazonais e localizadas de tráfego — acentuam a necessidade de soluções adaptativas e em tempo real.

Neste sentido, esta tese propõe duas contribuições principais:

- Propor um modelo de orquestração inteligente de recursos em H-CRAN, que integre mapeamento adaptativo RRH-BBU e redistribuição dinâmica de usuários com base em múltiplos critérios;
- Avaliar, de forma comparativa, diferentes estratégias de otimização bioinspiradas, considerando cenários realistas com flutuação de carga e topologias heterogêneas.

Dando continuidade, o próximo capítulo apresenta o cenário de simulação utilizado neste trabalho, detalhando todos os parâmetros de rede definidos. Em seguida, são introduzidas as métricas de avaliação e a descrição do algoritmo TLA, que será utilizado como benchmark para análise dos resultados obtidos.

4 PROPOSTA DE TESE

4.1 Considerações Iniciais

Este capítulo apresenta a proposta metodológica desenvolvida neste trabalho, com foco na alocação eficiente de usuários em redes H-CRAN. São descritos o modelo do sistema, a arquitetura adotada, os critérios de alocação baseados em SINR, o cálculo da necessidade de PRBs, bem como os mecanismos de detecção de saturação, *offloading*, desconexão e redistribuição final com base em perfis de usuários. Também são detalhados os indicadores de desempenho utilizados como função objetivo e as estratégias de otimização aplicadas.

4.2 Modelo do Sistema

A arquitetura H-CRAN incorpora atributos de redes definidas por software (SDN) e computação em nuvem, o que possibilita ganhos expressivos em flexibilidade, capacidade de controle e eficiência no uso dos recursos da rede. Essa abordagem favorece o balanceamento dinâmico de carga, promove melhorias na eficiência espectral (SE) e contribui significativamente para a redução do consumo energético e dos custos operacionais. Além disso, destaca-se a elevação da qualidade de serviço (QoS) percebida pelos usuários finais, viabilizada tanto pela menor distância entre os equipamentos de rádio (RRHs) e os dispositivos dos usuários (UEs), quanto pela possibilidade de associação adaptativa entre diferentes camadas da infraestrutura de acesso (CHEN, 2024).

Diante dessas vantagens, este trabalho adota a arquitetura H-CRAN como base para o desenvolvimento da proposta. A escolha também se justifica pela capacidade da H-CRAN de integrar elementos de redes legadas, utilizando uma abordagem heterogênea e centralizada. A estrutura proposta combina uma antena macro (Macro eNodeB), responsável por atender áreas não cobertas pelas RRHs, com um conjunto distribuído de RRHs do tipo 1 e tipo 2. A Figura 4, apresenta a visão geral do cenário considerado, com uma célula macro central e seis células ao seu redor, cada uma contendo uma RRH tipo 1 e, ao seu redor, seis RRHs tipo 2 de menor potência e alcance. Essa topologia simula um ambiente urbano com zonas de alta densidade, sujeitas a interferência e

sobrecarga, e regiões de cobertura ampliada pela macro célula.

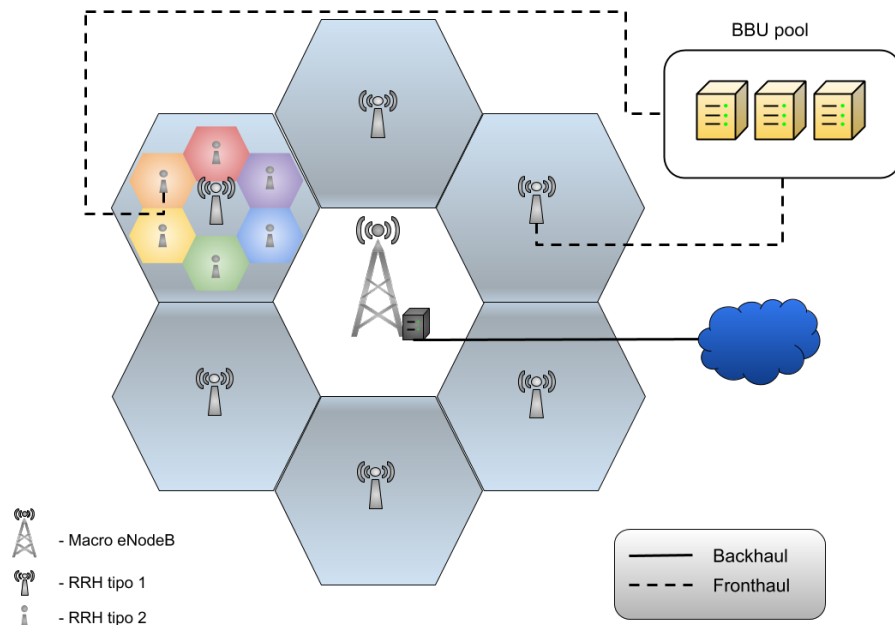


Figura 4 – Visão Geral do Cenário - Arquitetura H-CRAN

A seguir, a Figura 5, apresenta a metodologia proposta de orquestração de recursos em ambientes H-CRAN, baseada em controle SDN e reconfiguração em nuvem. Inicialmente, a área de cobertura é monitorada para identificar variações na distribuição de usuários — um comportamento dinâmico representado pelo efeito de maré. Quando essas flutuações são detectadas, o módulo de redistribuição de recursos é ativado, realizando a identificação de setores sobrecarregados, seleção de usuários para realocação e balanceamento de carga entre setores, preparando o sistema para a próxima etapa de otimização.

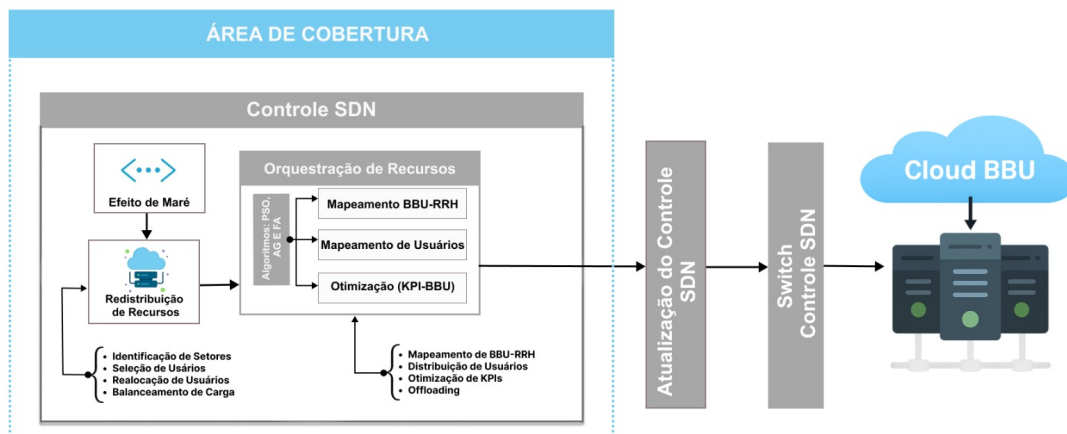


Figura 5 – Imagem Descritiva da Metodologia Proposta por esta Tese

Na sequência, o processo de orquestração de recursos atua sobre a nova configuração da rede. São aplicadas técnicas de otimização para o mapeamento entre RRHs e BBUs, a alocação de usuários, a avaliação dos KPIs definidos e a execução do offloading adaptativo, quando necessário. As decisões geradas por esse módulo alimentam o componente de Atualização do Controle SDN, responsável por reprogramar as regras de encaminhamento e associação da infraestrutura física.

Por fim, essas alterações são transmitidas para o plano de dados por meio do Switch SDN, que aplica as novas configurações diretamente na Cloud BBU, garantindo uma operação mais eficiente, adaptativa e inteligente da rede.

Para melhor entendimento, a Figura 6 apresenta o fluxograma detalhado da proposta, representando o processo lógico implementado no sistema. O fluxo se inicia com a simulação do efeito de maré, que altera dinamicamente a distribuição de usuários na área de cobertura. A partir dessa distribuição, realiza-se a alocação inicial nas antenas e a verificação do cenário de saturação. Caso o sistema identifique setores sobrecarregados ou usuários bloqueados, são aplicadas estratégias de redistribuição, incluindo offloading e, se necessário, desconexão planejada. O mapeamento das RRHs para os setores da BBU e a avaliação contínua dos KPIs orientam a reconfiguração da rede, garantindo uma alocação adaptativa, equilibrada e eficiente dos recursos.

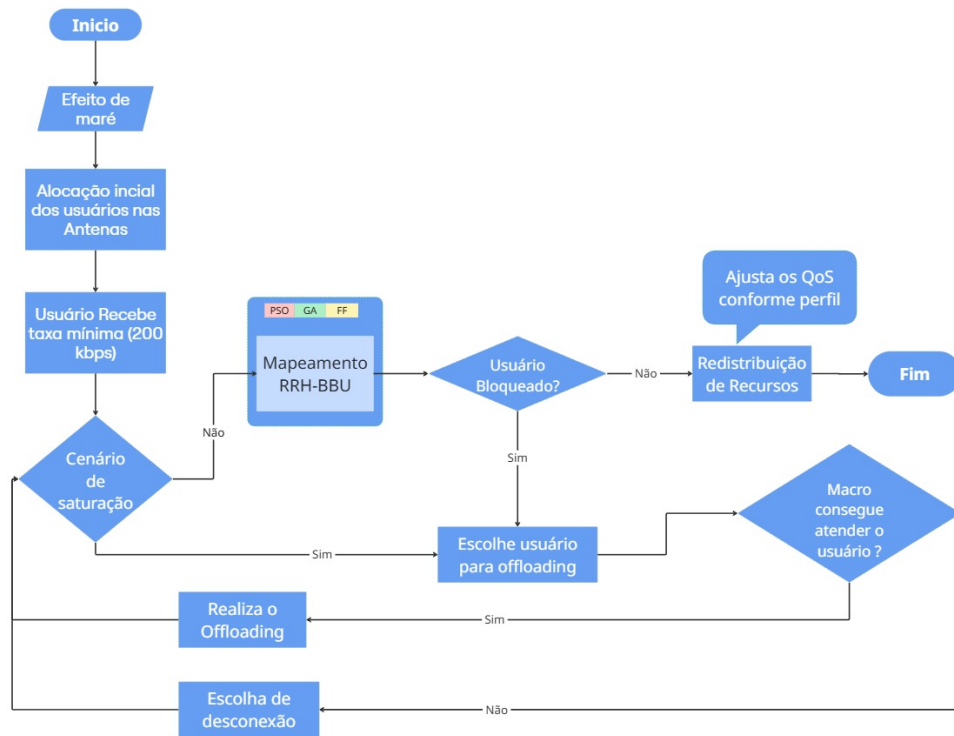


Figura 6 – Fluxograma Geral

4.2.1 Alocação Inicial

A alocação inicial dos usuários é realizada de forma inteligente, com o objetivo de maximizar o aproveitamento dos recursos disponíveis e minimizar bloqueios. O processo tem início com a avaliação do *Signal-to-Interference-plus-Noise Ratio* (SINR) de cada usuário em relação às antenas disponíveis: Macro eNodeB, RRHs tipo 1 e RRHs tipo 2. Essa avaliação determina a viabilidade física de conexão com determinada antena, com base na intensidade do sinal recebido, interferência e ruído.

Para que um usuário seja considerado elegível à conexão, três critérios devem ser simultaneamente satisfeitos:

- A antena selecionada deve possuir recursos disponíveis, uma vez que a quantidade de canais (PRBs) diminui à medida que novos usuários são conectados;
- O usuário deve estar dentro da área de cobertura da antena: 1000 metros para a macro, 150 metros para RRHs tipo 1 e 60 metros para RRHs tipo 2;

- No caso de conexão via RRHs, o número total de usuários conectados ao setor da BBU correspondente não pode ultrapassar a capacidade máxima de atendimento (HC) definida para aquele setor.

Cumprindo esses critérios, os usuários são classificados com base no número de canais necessários para atender às suas exigências de QoS. A priorização favorece aqueles que consomem menos recursos, o que permite que a rede atenda ao maior número possível de usuários.

4.2.2 Modelo de Perda por Propagação (SUI)

Foi utilizado o modelo de perda propagação SUI (*Stanford University Interim*), que é amplamente utilizado na literatura como modelo de perda pelo percurso em redes celulares operando acima de 2GHz. Onde $PL_{SUI}(d)$ em dB é encontrado por (CASTRO et al., 2010)(NASSAR; SULYMAN; ALSANIE, 2015):

$$PL_{SUI}(d) = PL_{d0} + 10n \log_{10} \left(\frac{d}{d_0} \right) + X_{fc} + X_{RX} + X_{\sigma} \quad (3a)$$

$$PL_{d0} = 10n \log_{10} \left(\frac{4\pi d_0}{\lambda} \right) \quad (3b)$$

$$n = a - b \cdot h_{TX} + \frac{c}{h_{TX}} \quad (3c)$$

$$X_{fc} = 6 \log_{10} \left(\frac{f_{MHZ}}{2000} \right), f_c > 2GHz \quad (3d)$$

$$X_{RX} = -10 \cdot 8 \cdot \log_{10} \left(\frac{h_{RX}}{2} \right) \quad (3e)$$

- 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 $d_0 = 100$ metros;
- X_{σ} , é o desvanecimento causado pelas áreas de sombreamento, podendo estar entre $8.2 < \sigma < 10.6dB$;

- Os parâmetros a , b e c são constantes, utilizadas para modelar os tipos de terreno. Nesta tese, foi utilizado 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 (f_c) em MHz;
- h_{TX} é a altura da estação base, podendo estar entre 10 a 80 metros;
- n é o expoente de perdas;

A utilização deste modelo permite calcular, para cada usuário, a potência recebida a partir de cada antena, servindo de base para o cálculo do SINR, que é utilizado na estimativa da vazão por canal (max) e, conseqüentemente, no dimensionamento da quantidade de canais (PRBs) necessários para atender cada perfil de usuário.

4.2.3 Cálculo de PRBs

O número de canais necessários por usuário é definido a partir do requisito mínimo de qualidade de serviço (QoS) e da taxa de transmissão máxima que ele pode obter em cada canal, considerando as condições de SINR. Essa estimativa é essencial para garantir alocação eficiente dos recursos e viabilidade de conexão inicial.

A taxa de transmissão máxima ($V_{\max}$) que um usuário pode obter com determinada antena é calculada a partir do Teorema de Shannon–Hartley, conforme a equação:

$$V_{\max} = B \cdot \log_2(1 + \text{SINR}) \quad (4)$$

onde B é a largura de banda do canal e SINR é a razão sinal interferência ruído, calculada a partir da potência recebida, interferência intercelular e ruído gaussiano branco.

Com base nessa vazão, o número de canais N necessários para que o usuário alcance o nível desejado de qualidade de serviço (QoS) é calculado da seguinte forma:

$$X = \frac{\text{QoS}}{V_{\max} \cdot 1024}, \quad N = \min\{N \in \mathbb{Z} \mid N \geq X\} \quad (5)$$

onde:

- QoS representa o requisito de taxa de transmissão utilizado como parâmetro de admissão do usuário;
- $V_{\max}$ é a vazão máxima por canal, calculada com base na equação de Shannon–Hartley.

A Equação 5 determina a quantidade exata (real) de canais, enquanto a conversão para o menor inteiro superior assegura a viabilidade da alocação dentro da granularidade dos PRBs disponíveis.

Após a definição da necessidade de canais para cada usuário, o sistema os ordena por demanda crescente. Essa estratégia, aplicada tanto na alocação inicial quanto na redistribuição posterior, busca garantir o maior número de conexões com o menor consumo de recursos possível.

4.2.4 Dinâmica de Carga e Detecção de Saturação

Após a alocação inicial dos usuários às antenas, o sistema passa a monitorar dinamicamente o estado da rede, considerando variações temporais de carga causadas pela mobilidade dos usuários. Como será detalhado no Capítulo 5, os cenários simulados incorporam flutuações de tráfego ao longo de 24 janelas temporais, refletindo o comportamento típico de redes urbanas com padrões de concentração e dispersão de acessos.

Essas flutuações exigem mecanismos adaptativos de gerenciamento de recursos, uma vez que, em determinados períodos do dia, a capacidade das antenas locais ou das BBUs pode ser ultrapassada, resultando em saturação e na necessidade de *offloading* ou redistribuição dos usuários. Compreender esse comportamento é essencial para o dimensionamento eficiente da infraestrutura e para a aplicação de políticas dinâmicas de alocação.

A saturação da rede, observada nos momentos de maior ocupação, não decorre apenas da limitação física de canais nas antenas, mas também da restrição de capacidade da BBU, representada pelo número máximo de usuários (HC) que cada setor consegue gerenciar simultaneamente. Quando essa capacidade é excedida — mesmo havendo PRBs disponíveis — o sistema enfrenta bloqueios, o que reforça a necessi-

dade de estratégias inteligentes de redistribuição e *offloading* para mitigar os efeitos da sobrecarga.

4.2.5 Mapeamento BBU-RRH

4.2.5.1 Mapeamento Físico Inicial das RRHs

Antes da aplicação de técnicas de otimização global, o sistema realiza um mapeamento físico preliminar entre as RRHs e os setores do pool de BBUs. Essa etapa tem como objetivo distribuir as RRHs de forma estratégica, com base em critérios simples de balanceamento de carga e proximidade, prevenindo uma concentração excessiva de tráfego em setores específicos.

A lógica de associação considera a carga acumulada em cada setor e a cobertura da RRH, de modo a alocar cada unidade rádio remota ao setor com menor ocupação relativa. Essa configuração inicial serve como ponto de partida para o funcionamento da rede e, mesmo que eficiente em cenários de baixa densidade, pode levar à formação de gargalos sob aumento de carga.

Por essa razão, esse mapeamento físico é posteriormente aprimorado com o uso de algoritmos bioinspirados, detalhados na próxima subseção, com o objetivo de otimizar globalmente a distribuição entre RRHs e setores das BBUs, considerando múltiplos indicadores de desempenho da rede.

4.2.5.2 Otimização Heurística do Mapeamento RRH-BBU

Para além das estratégias locais de redistribuição de recursos, a proposta incorpora uma etapa de otimização global do mapeamento entre as RRHs e os setores disponíveis no pool de BBUs. Essa etapa busca distribuir a carga de forma mais equilibrada, evitando sobrecarga em setores específicos e melhorando a eficiência geral da rede.

A otimização é realizada por meio de três algoritmos bioinspirados — PSO, GA e FA — aplicados sobre diferentes configurações de alocação entre RRHs e setores. O processo é orientado por uma função objetivo F_{ob} , que corresponde à equação de QoS do sistema apresentada na Seção 4.2.6.5. Essa função combina múltiplos KPIs — como

taxa de bloqueio, desequilíbrio de carga e DataRate médio por usuário — ponderados por pesos ajustáveis.

Os fluxogramas que representam a lógica operacional de cada algoritmo estão apresentados nas Figuras 7, 8 e 9.

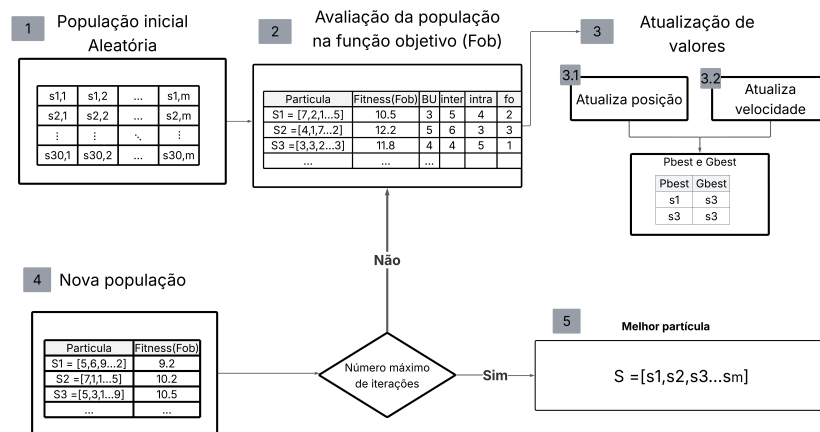


Figura 7 – Etapas de otimização com PSO

No PSO, cada partícula representa uma solução candidata completa de alocação das RRHs aos setores da BBU pool. Inicialmente, uma população de partículas é gerada aleatoriamente (Etapa 1). Cada partícula é avaliada pela função objetivo Fob (Etapa 2), e as melhores soluções são refinadas com base nas informações individuais (*Pbest*) e globais (*Gbest*) das partículas. O processo de atualização ocorre nas Etapas 3 a 4, e é encerrado ao atingir o número máximo de iterações (Etapa 5), resultando na partícula com melhor desempenho.

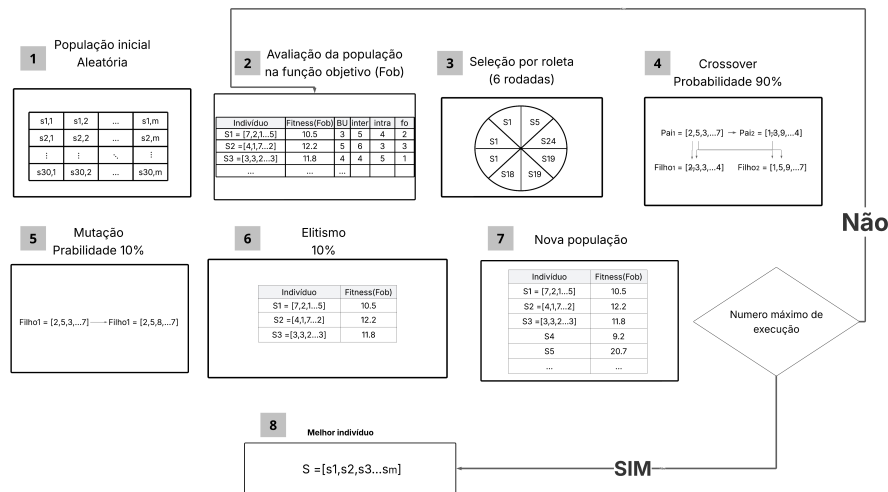


Figura 8 – Etapas de otimização com GA

O GA opera com uma população de cromossomos, onde cada indivíduo representa uma configuração completa do mapeamento RRH-BBU. Após a avaliação inicial por Fob, os melhores indivíduos são selecionados via roleta proporcional. Os operadores de *crossover* e *mutação* são aplicados nas Etapas 4 e 5 para gerar diversidade. A preservação dos melhores indivíduos é garantida por elitismo (Etapa 6), formando uma nova geração (Etapa 7) até a convergência do algoritmo.

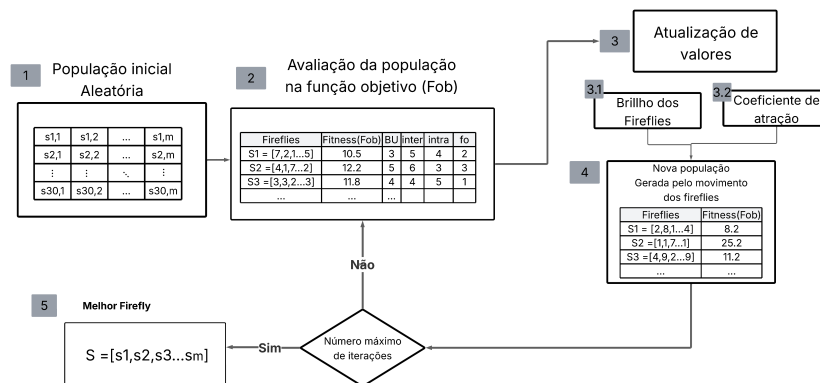


Figura 9 – Etapas de otimização com FA

O FA segue um princípio de atração entre as soluções, representadas por fireflies. As soluções mais brilhantes (i.e., com menor valor de Fob) atraem outras, que se deslocam em sua direção. Esse comportamento é controlado por parâmetros como o

coeficiente de atração β_0 , o fator de absorção de luz γ e o grau de aleatoriedade α . O processo gera uma nova população a cada iteração, convergindo gradualmente para a melhor configuração identificada.

Todos os algoritmos foram implementados no ambiente MATLAB R2023a e aplicados individualmente aos 24 cenários simulados. Cada execução foi repetida 30 vezes, com populações de 30 indivíduos e até 30 iterações por algoritmo.

As Tabelas 1 e 2 sintetizam os parâmetros, a modelagem da população e as estratégias utilizadas nas três metaheurísticas avaliadas.

Tabela 1 – Parâmetros dos algoritmos bioinspirados utilizados

Parâmetro	PSO	GA	FA
Número de iterações	30	30	30
Tamanho da população	30	30	30
Coef. cognitivo e social (PSO)	$c_1 = c_2 = 1.8$	–	–
Taxa de mutação (GA) / Aleatoriedade (FA)	–	10%	$\alpha = 20\%$
Probabilidade de crossover (GA)	–	90%	–
Coef. de absorção de luz (FA)	–	–	$\gamma = 1$
Coef. de atração (FA)	–	–	$\beta_0 = 2$

A Tabela 1 apresenta os valores dos parâmetros utilizados em cada metaheurística. Esses valores foram definidos com base em recomendações da literatura e visam equilibrar a taxa de convergência com a diversidade da busca. No caso do PSO, foram adotados coeficientes cognitivo e social iguais, incentivando um equilíbrio entre exploração e exploração. No GA, a alta probabilidade de *crossover* (90%) busca promover diversidade genética, enquanto a mutação moderada (10%) evita estagnação. Para o FA, os parâmetros α , β_0 e γ controlam, respectivamente, o grau de aleatoriedade, a intensidade de atração e a atenuação da luz.

Tabela 2 – Características da população e representação das soluções

Algoritmo	População	Representação da solução	Espaço de Busca	Operadores / Estratégias
GA	30 ind.	Vetor de 42 posições (36 pico + 6 micro), cada posição representa uma RRH associada a um setor (1–9)	9^{42} combinações possíveis	Roleta ($p = 0,9$), crossover ($p = 0,9$), mutação ($p = 0,1$), elitismo 10%
PSO	30 part.	Vetor de 42 posições (36 pico + 6 micro), cada posição representa uma RRH associada a um setor (1–9)	9^{42} combinações possíveis	Atualização vel./pos., guias $pBest$ e $gBest$
FA	30 ind.	Vetor de 42 posições (36 pico + 6 micro), cada posição representa uma RRH associada a um setor (1–9)	9^{42} combinações possíveis	Movimento por atratividade/luz, ruído aleatório p /diversificação

A Tabela 2 detalha como as soluções candidatas são representadas e manipuladas em cada algoritmo. Cada indivíduo (ou partícula, no caso do PSO) é modelado como um vetor de 42 posições, correspondente às 42 RRHs do sistema — sendo 36 pico e 6 micro —, onde cada posição armazena o setor (de 1 a 9) ao qual a RRH está alocada. O espaço de busca, 9^{42} combinações possíveis, é extremamente grande, reforçando a necessidade de métodos heurísticos. As estratégias de cada algoritmo diferem no modo como exploram esse espaço: o GA usa seleção por roleta, *crossover* e mutação; o PSO atualiza posições e velocidades com base nos melhores históricos; e o FA desloca as soluções conforme a atratividade e aleatoriedade.

4.2.6 Redistribuição Adaptativa de Usuários

Encerrada a etapa de mapeamento físico, o sistema direciona o foco para o monitoramento contínuo da ocupação em cada setor da rede. Esse acompanhamento adaptativo visa detectar situações de saturação — causadas tanto pelo esgotamento dos PRBs disponíveis quanto por limitações computacionais das BBUs — que possam comprometer a conectividade dos usuários.

Uma vez confirmada a saturação e identificado o bloqueio de usuários, é acionada uma política de redistribuição adaptativa. A primeira tentativa consiste na verifica-

ção da possibilidade de *offloading* para antenas de nível superior — como microcélulas — e, caso essas também estejam indisponíveis ou saturadas, a redistribuição é escalada até as macrocélulas. Essa realocação só é realizada se a antena candidata estiver dentro da área de cobertura e possuir canais livres, garantindo que o deslocamento preserve os requisitos mínimos de QoS.

Esse mecanismo busca aliviar a sobrecarga nos setores comprometidos, otimizando o uso dos recursos disponíveis e assegurando continuidade de serviço aos usuários impactados.

Caso o *offloading* não seja viável — seja por ausência de cobertura ou indisponibilidade de canais —, o sistema aplica uma estratégia de desconexão controlada. Nesse processo, são priorizados para remoção os usuários com pior relação sinal-interferência (SINR), que consomem mais recursos sem alcançar níveis satisfatórios de QoS. A exclusão desses usuários tem como objetivo liberar canais e capacidade computacional, permitindo a realocação de usuários mais eficientes em termos de uso espectral.

Para apoiar o balanceamento entre os setores da BBU pool, a proposta utiliza uma combinação de indicadores-chave de desempenho (KPIs), conforme metodologia apresentada por (KHAN; FAKHRI; AL-RAWESHIDY, 2017). Esses indicadores — devidamente ponderados — compõem a função objetivo adotada nos algoritmos bioinspirados implementados neste trabalho: PSO, FA e AG.

Todas as decisões relacionadas à redistribuição de usuários são guiadas por essa função objetivo, que incorpora múltiplas métricas de desempenho da rede. Essa abordagem multiobjetivo garante que a otimização leve em conta não apenas o atendimento individual de cada usuário, mas também a eficiência global na alocação de recursos.

Os KPIs utilizados são:

- Número de usuários bloqueados (*Blocked Users*);
- Handovers entre BBUs (*Inter-BBU Handovers*);
- Handovers dentro da mesma BBU (*Intra-BBU Handovers*);
- Handovers forçados (*Forced Handovers*).

Para descrever os KPIs é necessário ter conhecimento prévio de algumas variáveis. Primeiro, $R_s^{t+1} = [R_1^{t+1}, R_2^{t+1}, R_3^{t+1}, \dots, R_m^{t+1}]$, representa um vetor da configuração das RRHs no setor s , onde m é o número de RRHs e R_i^{t+1} representa se, na configuração do tempo $t+1$, a RRH_i está no setor s , de tal forma que $R_i^{t+1} = 1$ se estiver presente, e $R_i^{t+1} = 0$ caso contrário.

$U = [U_1, U_2, U_3, \dots, U_m]$ representa o vetor quantidade usuários. No qual U_i , com $i = 1, 2, 3, \dots, m$ é a quantidade de usuários na RRH_i .

SoS_n é um vetor de setores que estão presentes na BBU_n , com $n = 1, 2, 3 \dots N$. Sendo N o número de BBUs.

4.2.6.1 Blocked Users

Avalia e quantifica a configuração das RRHs nos setores em relação ao número de usuários bloqueados. Ou seja, o número de usuários excedendo a capacidade máxima (HC) do setor, na configuração $t + 1$.

$$BU = \sum_s \max \left[\left(\left(\sum_i U_i R_{is}^{t+1} \right) - HC_s \right), 0 \right] \quad (6)$$

Onde R_i , com $i = 1, 2, \dots, M$, representa a RRH_i e $s = 1, 2, \dots, S$, o setor. BU define a soma de usuários bloqueados em todos os setores. O valor de KPI para usuários bloqueados (KPI_{BU}) é:

$$KPI_{BU} = \begin{cases} 1 & \text{if } BU = 0 \\ \frac{1}{BU} & \text{otherwise} \end{cases} \quad (7)$$

4.2.6.2 Inter-BBU

Um KPI para *handovers* entre BBUs. Dado uma configuração atual no tempo t , se na configuração $t + 1$ a RRH_i e RRH_j estiverem em BBUs diferentes, pode ocorrer um 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.

$$\text{Inter-BBU}_{\text{HO}} = \sum_i \sum_{j \neq i} \frac{U_i}{D_{ij}} \left(1 - \left(\sum_n \sum_{s \in \text{SOS}_n} (\mathbf{R}_{is}^{t+1} \cdot \mathbf{R}_{js}^{t+1}) \right) \right) \quad (8)$$

O KPI de *handover* entre BBUs pode ser representado como KPI_{inter} :

$$KPI_{\text{inter}} = \begin{cases} 1 & \text{if Inter-BBU}_{\text{HO}} = 0 \\ [\text{Inter-BBU}_{\text{HO}}]^{-1} & \text{otherwise.} \end{cases} \quad (9)$$

4.2.6.3 Intra-BBU

De forma similar ao KPI *inter-BBU*, mede um custo de uma configuração na qual possa ocorrer o *handover*, porém, em uma única BBU através da troca de setores. Para isso, a equação (3.11) 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.

$$\text{Intra-BBU}_{\text{HO}} = \sum_i \sum_{j \neq i} \frac{U_i}{D_{ij}} \left[\sum_n \sum_{s \in \text{SOS}} (\mathbf{R}_{is}^{t+1} \cdot \mathbf{R}_{js}^{t+1}) - \sum_s (\mathbf{R}_{is}^{t+1} \cdot \mathbf{R}_{js}^{t+1}) \right] \quad (10)$$

$$KPI_{\text{intra}} = \begin{cases} 1 & \text{if Intra-BU}_{\text{HO}} = 0 \\ [\text{Intra-BBU}_{\text{HO}}]^{-1} & \text{otherwise.} \end{cases} \quad (11)$$

4.2.6.4 Forced Handovers

Aplicado quando uma RRH realiza uma troca de setor. Ao trocar de configuração, mede o quão custoso é, 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.

$$f_{HO} = \sum_s \sum_i (R_{is}^t + R_{is}^{t+1}) U_i \quad (12)$$

$$KPI_f = \begin{cases} 1 & \text{if } f_{HO} = 0 \\ [f_{HO}]^{-1} & \text{otherwise.} \end{cases} \quad (13)$$

4.2.6.5 QoS do Sistema

Dadas as equações dos KPIs, o QoS do sistema pode ser definido como:

$$QoS_{max} = w_1 \cdot KPI_{BU} + w_2 \cdot KPI_{inter} + w_3 \cdot KPI_{intra} + w_4 \cdot KPI_f \quad (14)$$

Em que os pesos w_1 , w_2 , w_3 e w_4 aqui são escolhidos os mesmos utilizados em [KHAN et. al. 2017]: 0.52, 0.27, 0.15 e 0.06, respectivamente.

Essa função é continuamente avaliada para orientar o sistema rumo a uma configuração mais eficiente e equilibrada, promovendo não apenas a conectividade, mas também a qualidade da experiência do usuário final.

4.2.6.6 Redistribuição Final com Perfis de Usuários

Encerrada a etapa de otimização global com os algoritmos bioinspirados — guiada pela função objetivo baseada em KPIs —, o sistema avança para uma redistribuição local dos canais remanescentes. Essa fase não envolve os algoritmos de otimização e tem como foco o ajuste fino das alocações, considerando diferentes perfis de uso e exigências de largura de banda. O objetivo é refinar a entrega de QoS de forma personalizada, elevando a eficiência do sistema mesmo após a mitigação dos casos de bloqueio por *offloading* ou desconexão controlada.

Segundo a Ericsson (ERICSSON, 2024), o tráfego de vídeo representa a maior parcela do volume total de dados móveis, evidenciando a dominância desse tipo de serviço nas redes atuais. Complementarmente, Mahbub et al. (MAHBUB; BARUA, 2021) demonstram que plataformas como Netflix, YouTube, TikTok e Facebook possuem padrões distintos de consumo de banda e impacto energético, o que reforça a importância

de distinguir os usuários conforme os tipos de aplicação utilizados. No fluxograma da Figura 6, essa etapa aparece no bloco “Redistribuição de Recursos”, com destaque para o ajuste das taxas de QoS segundo os perfis definidos.

Os perfis considerados foram atribuídos com base em distribuições percentuais representativas de cenários reais, permitindo capturar diferentes níveis de exigência de banda. A descrição detalhada desses perfis — bem como suas respectivas demandas de largura de banda — encontra-se na Seção 5.3.1, onde o cenário de simulação é formalmente apresentado.

A redistribuição leva em conta a diferença entre a taxa atualmente alocada e a taxa desejada por usuário. O sistema calcula, para cada caso, a quantidade mínima de canais adicionais necessária para que a taxa efetiva atinja o valor desejado, conforme a Equação 15:

$$X = \frac{QoS_d - QoS_a}{V_{\max} \cdot 1024}, \quad N = \min\{N \in \mathbb{Z} \mid N \geq X\} \quad (15)$$

onde:

- QoS_d : taxa de transmissão desejada conforme o perfil do usuário (em kbps);
- QoS_a : taxa já alocada ao usuário (em kbps);
- $V_{\max}$: vazão por canal, estimada a partir do SINR (em Mbps);
- N : número mínimo de canais adicionais necessários.

A redistribuição é executada de forma incremental e equitativa: os canais disponíveis após a alocação inicial são priorizados para usuários com menor exigência adicional, favorecendo um maior número de atendimentos. Caso não seja possível satisfazer completamente todas as demandas, o sistema aplica uma realocação progressiva entre os ainda não atendidos. Usuários já plenamente atendidos não recebem recursos adicionais, assegurando eficiência e justiça na utilização da largura de banda disponível.

4.3 Considerações Finais

Neste capítulo, foi apresentado o modelo completo da proposta de alocação e redistribuição inteligente de usuários em redes móveis H-CRAN. Inicialmente, foram descritas a arquitetura heterogênea adotada, os parâmetros físicos de propagação e a lógica de tomada de decisão baseada em critérios de SINR. A modelagem incluiu o cálculo da vazão por canal com base na equação de Shannon–Hartley, a definição da necessidade de PRBs por usuário e o uso do modelo SUI para estimar a potência recebida.

Foram abordados os mecanismos de detecção de saturação da rede, tanto em nível de antena quanto no limite de processamento das BBUs, e as ações corretivas aplicadas — como estratégias de *offloading* e desconexão controlada — para mitigar os efeitos de sobrecarga. A proposta também incorporou um conjunto de KPIs, consolidados em uma função objetivo que guia o balanceamento de carga de forma dinâmica e adaptativa.

Na etapa final da modelagem, foi incluída uma redistribuição adicional de canais voltada à melhoria da experiência do usuário, baseada na diferença entre a taxa de QoS alocada e a desejada. Além disso, o mapeamento entre RRHs e setores da BBU foi otimizado por meio de algoritmos bioinspirados (PSO, GA e FA), que buscaram configurações mais equilibradas de uso dos recursos da rede, com base em múltiplos KPIs definidos por função objetivo.

No capítulo seguinte, será apresentado o ambiente de simulação utilizado para validar a proposta, juntamente com a análise dos resultados obtidos em diferentes cenários de carga, distribuição espacial de usuários e intensidade de tráfego.

5 RESULTADOS

5.1 Considerações Iniciais

Neste capítulo são apresentadas as importantes considerações sobre o ambiente de simulação, bem como toda a parametrização necessária. A análise dos dados foca em métricas-chave relacionadas à eficiência de alocação de usuários, qualidade de serviço (QoS) e utilização de recursos, alinhadas aos objetivos centrais da tese. Cada resultado é discutido criticamente, relacionando os ganhos obtidos com as limitações da abordagem tradicional representada pelo modelo TLA (*Two-Level Allocation*).

5.2 Definição das Métricas e Termos Utilizados

Com o objetivo de facilitar a compreensão dos resultados apresentados neste capítulo, esta seção descreve os principais termos e métricas utilizadas nesta proposta de tese. Cada item é fundamental para a avaliação da eficiência, escalabilidade e qualidade de serviço da rede H-CRAN simulada.

- **Usuários Aceitos:** Refere-se ao número de usuários que conseguiram ser efetivamente atendidos pela rede em cada cenário. Esses usuários receberam alocação suficiente de canais (PRBs) para atingir os requisitos mínimos de QoS. Essa métrica representa diretamente a capacidade de atendimento da rede.
- **Usuários Não Atendidos (UNA):** Corresponde ao total de usuários que não obtiveram conexão satisfatória, seja por falta de recursos nas antenas (como canais insuficientes) ou por limitação de processamento nos setores da BBU. A presença de usuários não atendidos indica situações de saturação e restrições operacionais do sistema.
- **DataRate Médio por Usuário:** Representa a média da taxa de dados entregue por usuário atendido em cada cenário. Essa métrica avalia a eficiência do uso dos recursos de rede e a qualidade de serviço percebida pelos usuários, sendo crucial em aplicações com exigência de banda larga consistente.

- **Percentual Alterado Atendido:** Indica a proporção de usuários que, mesmo após ajustes em seus requisitos iniciais de *QoS* (como aumento de demanda de banda ou reconfiguração de perfil), conseguiram ser atendidos. Essa métrica mede a flexibilidade adaptativa do sistema frente a variações nas necessidades dos usuários.
- **Distribuição de Usuários por Setor:** Refere-se à quantidade de usuários alocados em cada setor da BBU ao longo do tempo. Uma distribuição equilibrada entre os setores é desejável, pois evita a formação de regiões sobrecarregadas e melhora a estabilidade operacional da rede.
- **Offloading:** Estratégia de redistribuição dinâmica dos usuários entre diferentes setores da rede, com o objetivo de aliviar áreas saturadas e melhorar o aproveitamento global dos recursos. O *offloading* pode ocorrer antes da alocação final dos recursos e é essencial para garantir continuidade de serviço sob alta densidade.
- **HC:** Trata-se da capacidade máxima de usuários simultâneos que cada setor da BBU é capaz de processar. Mesmo que haja canais físicos disponíveis, a superação desse limite impede a conexão de novos usuários, caracterizando um bloqueio por restrição computacional e exigindo a adoção de mecanismos de redistribuição.

5.3 Ambiente de Simulação

Toda proposta científica precisa ser testada para validar sua eficácia e aplicabilidade prática. Entre os métodos disponíveis, a simulação computacional destaca-se como uma alternativa eficiente e de baixo custo, permitindo a avaliação de sistemas complexos sem a necessidade de intervenção em infraestruturas reais, evitando riscos e prejuízos aos usuários (HOSSEINPOUR; SHIRMOHAMMADI, 2009).

Neste contexto, a avaliação da proposta de orquestração inteligente de recursos em redes H-CRAN foi realizada por meio de simulações computacionais, utilizando o ambiente MATLAB®, reconhecido por sua robustez no processamento numérico e flexibilidade para modelagem de sistemas de comunicação.

Foram simulados 24 cenários distintos, iniciando-se com 260 usuários e acrescentando-se 10 usuários por cenário, até atingir o total de 500 usuários. Para cada cenário, foram realizadas 30 execuções independentes, considerando diferentes posicionamentos alea-

tórios dos usuários, de forma a capturar a variabilidade estocástica natural dos ambientes móveis.

A análise estatística dos resultados considerou um intervalo de confiança de 95%, assegurando que as conclusões obtidas sejam representativas e robustas frente às variações amostrais. Esta abordagem metodológica visa refletir as condições dinâmicas das redes móveis de próxima geração, permitindo a comparação estruturada da proposta com o modelo TLA, utilizado aqui como benchmark de referência (GONÇALVES et al., 2019).

5.3.1 Configuração do Cenário de Simulação

O ambiente experimental modela uma rede móvel heterogênea implantada em uma área urbana circular com raio de 450 metros, totalizando aproximadamente 636.000 m². A topologia é composta por múltiplas estações rádio base (BSs) — incluindo macrocélulas, microcélulas e picocélulas — interligadas por uma arquitetura H-CRAN, com unidades BBU e gerenciamento centralizado de recursos. A simulação considera 500 usuários móveis, posicionados aleatoriamente a cada execução e distribuídos ao longo de 24 janelas temporais, representando as diferentes horas do dia.

A variação do número de usuários conectados em cada janela segue um padrão dinâmico típico de redes urbanas, conhecido como *efeito de maré*, caracterizado por picos de tráfego no início da noite e quedas nas primeiras horas da manhã. Esse comportamento foi baseado em dados observacionais de redes reais, conforme descrito por Falcão et al. (FALCÃO et al., 2020), e está ilustrado na Figura 10.

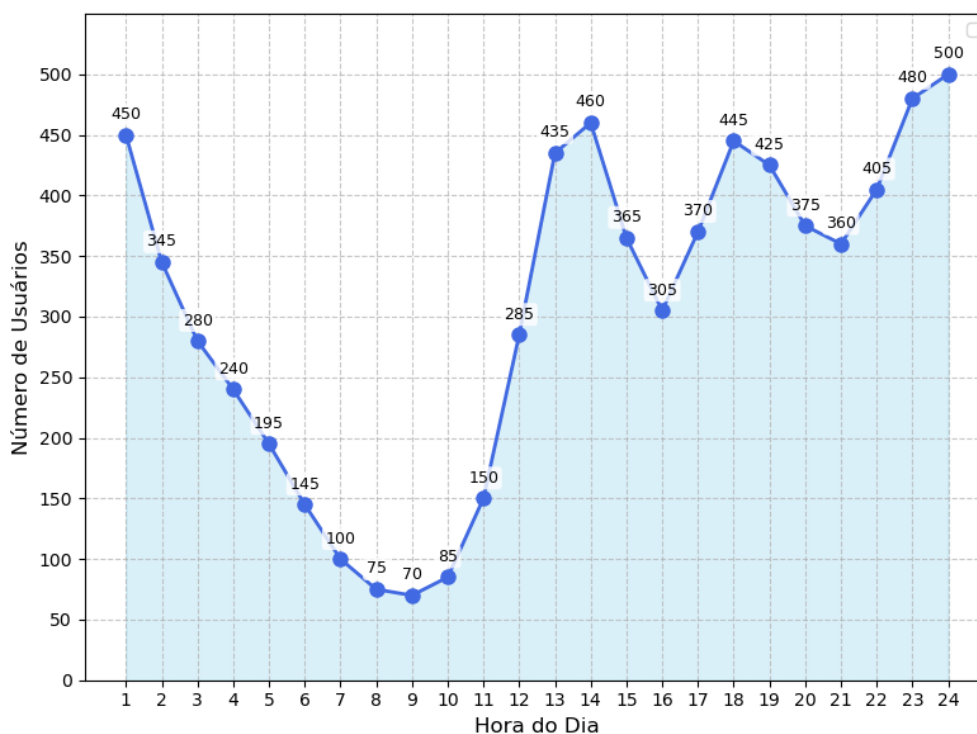


Figura 10 – Distribuição temporal dos usuários conectados ao longo do dia.

Fonte: Elaborado pelo autor, adaptado de (FALCÃO et al., 2020).

A admissão inicial de usuários baseia-se em um requisito mínimo de QoS, definido como um parâmetro fixo de taxa de transmissão por usuário. Este valor de referência foi assumido como 200 kbps, utilizado exclusivamente para fins de controle de admissão e dimensionamento inicial da carga. Esse valor representa uma conectividade base, compatível com serviços leves ou sessões de manutenção de presença em segundo plano, mas não corresponde às necessidades reais das aplicações utilizadas na simulação (LETO et al., 2018).

Após a etapa inicial de alocação e tratamento de usuários bloqueados, cada usuário é classificado de acordo com um perfil de aplicação. Esses perfis foram definidos com base em plataformas populares e demandas reais de banda larga, permitindo representar diferentes níveis de exigência de tráfego. A Tabela 3 resume os perfis adotados, seus percentuais e respectivas taxas-alvo.

Tabela 3 – Perfis de usuários adotados na simulação

Tipo de aplicação	Usuários	Aplicação	Taxa (Mbps)
Vídeo >20 min (Filme)	16%	Netflix	5.0
Live Streaming	20%	YouTube/TikTok	7.0
Rede Social	62%	Facebook	4.0

Fonte: Adaptado de (MAHBUB; BARUA, 2021).

Esses perfis são atribuídos aleatoriamente aos usuários no início de cada execução e impactam diretamente a redistribuição final de canais descrita no Capítulo 4, influenciando as métricas de eficiência, taxa de bloqueio e alocação proporcional de recursos.

5.3.2 Análise dos Resultados

Os resultados obtidos a partir das simulações permitem avaliar o desempenho da proposta de redistribuição adaptativa orientada por perfis de usuários em cenários de rede H-CRAN. A comparação com o modelo TLA original revela ganhos significativos em múltiplas métricas de eficiência, particularmente em situações de carga elevada e saturação parcial da infraestrutura.

Antes da análise consolidada das métricas, foram observados dois cenários representativos que auxiliam na compreensão do comportamento da rede: a baixa e a alta demanda. A Figura 11 apresenta o comportamento do sistema em condições de baixa carga, enquanto a Figura 12 ilustra o mesmo modelo operando em condições de elevada densidade de usuários.

Em situações de baixa demanda, verifica-se que o sistema mantém uma alocação equilibrada dos usuários entre os setores da BBU, com baixo número de bloqueios e elevada estabilidade no valor médio de *fitness*. Isso indica que a proposta é capaz de utilizar eficientemente os recursos disponíveis mesmo quando a rede opera abaixo de sua capacidade nominal, evitando ociosidade e garantindo qualidade de serviço constante.

Já no cenário de alta demanda, representado pela Figura 12, observa-se um comportamento distinto, com maior competição por recursos e tendência de concentração de usuários em determinados setores. Ainda assim, a proposta mantém um padrão de con-

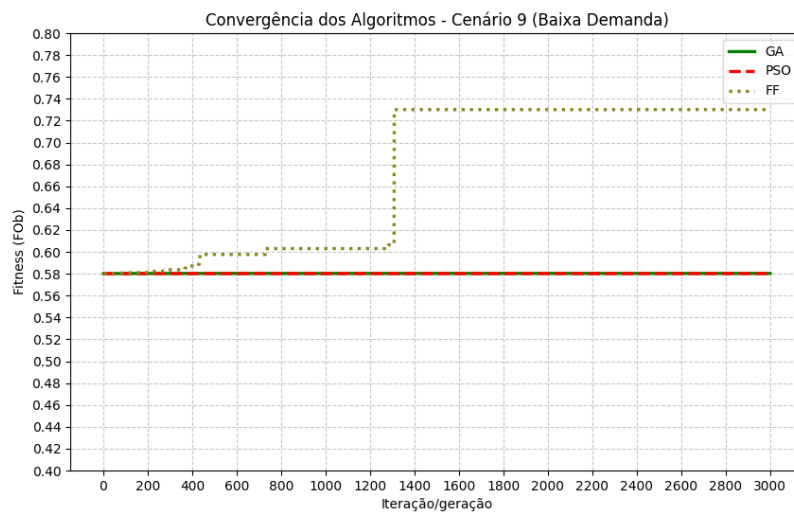


Figura 11 – Convergência dos algoritmos no Cenário 9 (Baixa Demanda).

vergência estável, demonstrando capacidade adaptativa diante de picos de carga. Entre os algoritmos analisados, o Firefly apresentou desempenho mais consistente, com melhor estabilidade de convergência e menor dispersão nos valores de *fitness*, enquanto o GA obteve os melhores resultados em cenários críticos devido à sua maior capacidade exploratória.

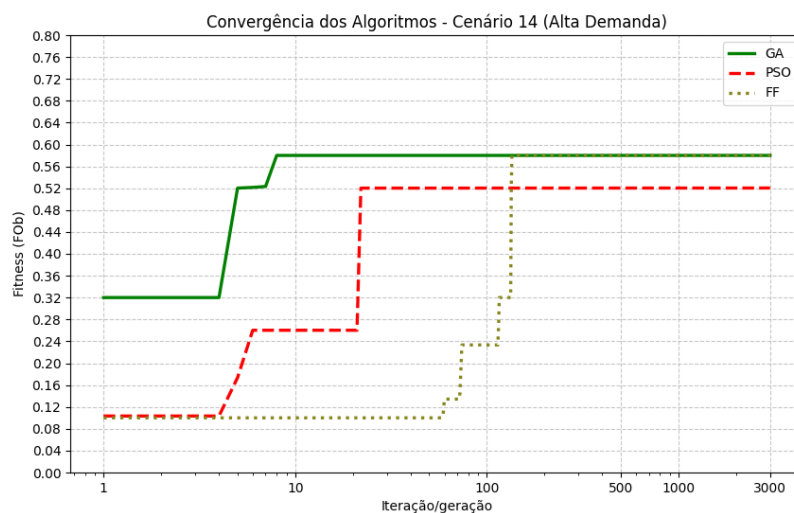


Figura 12 – Convergência dos algoritmos no Cenário 14 (Alta Demanda).

A comparação entre esses dois extremos reforça a robustez da proposta frente às oscilações diárias de carga (efeito de maré), garantindo desempenho previsível e eficiente tanto em situações de ociosidade quanto em condições de saturação. A partir dessas observações, as análises subsequentes detalham o comportamento das principais métricas de desempenho, como número de usuários aceitos, taxa de bloqueio e eficiência na redistribuição de recursos.

Essas análises iniciais fornecem uma visão qualitativa do comportamento da proposta em condições extremas, permitindo compreender como a dinâmica de carga influencia a eficiência da alocação. No entanto, para uma avaliação mais abrangente, torna-se necessário observar o comportamento médio das métricas ao longo de todos os cenários simulados. Assim, as figuras a seguir apresentam os resultados consolidados das 24 janelas temporais, destacando o impacto da redistribuição adaptativa na capacidade de atendimento, na taxa de bloqueio e na eficiência global da rede para cada uma das metaheurísticas empregadas.

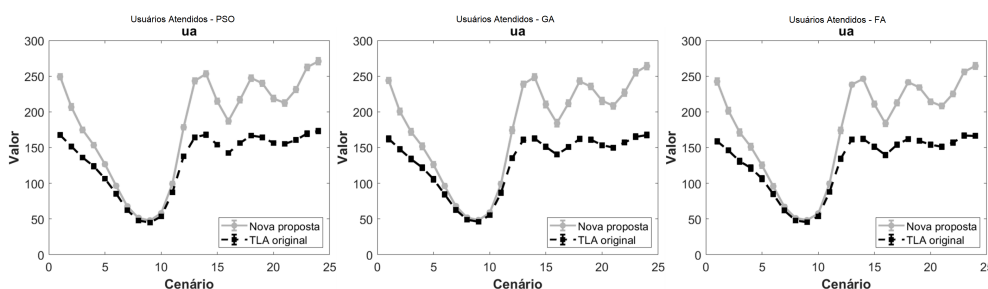


Figura 13 – Comparativo de Usuários Aceitos (em unidades)(Algoritmos PSO, GA e FA).

A Figura 13 apresenta a evolução do número de usuários aceitos em 24 cenários distintos, comparando o desempenho da proposta desenvolvida com o modelo tradicional TLA, para três algoritmos de otimização: PSO, GA e FA. Cada subgráfico corresponde a um desses algoritmos, permitindo observar o impacto da redistribuição dinâmica de usuários associada a estratégias de *offloading* e mapeamento adaptativo das RRHs sobre a capacidade de atendimento da rede.

No gráfico à esquerda, referente ao algoritmo PSO, é possível perceber que, nos cenários iniciais (até aproximadamente o cenário 8), ambas as abordagens apresentam desempenho semelhante. Isso se deve ao fato de que, neste estágio, o número total de usuários ainda não excede a capacidade de atendimento da rede, mesmo sob o modelo

TLA. No entanto, a partir do cenário 9, o comportamento das curvas começa a divergir significativamente. O TLA, limitado por sua estrutura estática e por uma política rígida de alocação de recursos, sofre uma queda acentuada no número de usuários aceitos, atingindo valores inferiores a 100 usuários nos cenários mais críticos. Por outro lado, a proposta desenvolvida mantém um crescimento consistente, mesmo diante do aumento da carga, alcançando valores superiores a 260 usuários aceitos nos cenários finais.

No subgráfico central, referente ao algoritmo GA, observa-se um padrão similar. A diferença entre as abordagens torna-se mais evidente a partir da metade dos cenários simulados. O modelo TLA apresenta uma tendência de estabilização precoce, mantendo-se abaixo de 200 usuários aceitos. Em contraste, a proposta desenvolvida mantém sua capacidade de adaptação e redistribuição inteligente de usuários, alcançando valores que ultrapassam 260 usuários aceitos em cenários de alta demanda. Isso evidencia a eficiência do algoritmo GA em conjunto com os mecanismos propostos para balanceamento de carga e mapeamento dinâmico dos setores da BBU.

O subgráfico à direita, que representa o desempenho do algoritmo FA, reforça essa tendência. Desde os cenários intermediários até os finais, a proposta apresenta resultados significativamente superiores. Mesmo em situações de alta densidade de usuários, a proposta é capaz de manter os valores de usuários aceitos em patamares elevados, próximos ao limite superior do gráfico (acima de 260), enquanto o TLA se mantém em níveis inferiores, com uma curva visivelmente mais achatada e distante da idealidade. Esse resultado destaca a robustez do algoritmo FA quando combinado com a abordagem proposta, especialmente em cenários onde a rede opera próxima de sua saturação.

A superioridade da proposta desenvolvida pode ser atribuída a diversos fatores complementares. Em primeiro lugar, a redistribuição dinâmica de usuários, realizada antes da saturação completa de determinados setores, permite um aproveitamento mais inteligente dos recursos disponíveis. Em segundo, o mapeamento dinâmico das RRHs aos setores da BBU viabiliza ajustes contínuos no encaminhamento dos usuários, evitando sobrecarga localizada. Além disso, o mecanismo de *offloading* adotado possibilita que usuários que não possam ser atendidos em uma RRH sobrecarregada sejam redirecionados para outras células com menor ocupação, sem comprometer os requisitos mínimos de QoS.

Outro aspecto relevante na Figura 13 é a estabilidade das curvas associadas à proposta. Enquanto o modelo TLA apresenta oscilações abruptas no número de usuários aceitos — resultado direto de sua rigidez estrutural diante de variações na carga —, a abordagem proposta demonstra uma evolução mais progressiva e controlada. Essa estabilidade é essencial para garantir previsibilidade e continuidade de serviço, especialmente em redes móveis de próxima geração que operam sob cenários altamente dinâmicos e densos.

Dessa forma, a análise da Figura 13 demonstra que a proposta desenvolvida não apenas aumenta significativamente a capacidade da rede em termos de usuários atendidos, mas também promove um comportamento mais resiliente e adaptativo. Tais características são indispensáveis para a evolução das redes móveis rumo às exigências impostas pelas arquiteturas 6G, nas quais a escalabilidade, a flexibilidade e a manutenção da qualidade de serviço serão elementos centrais.

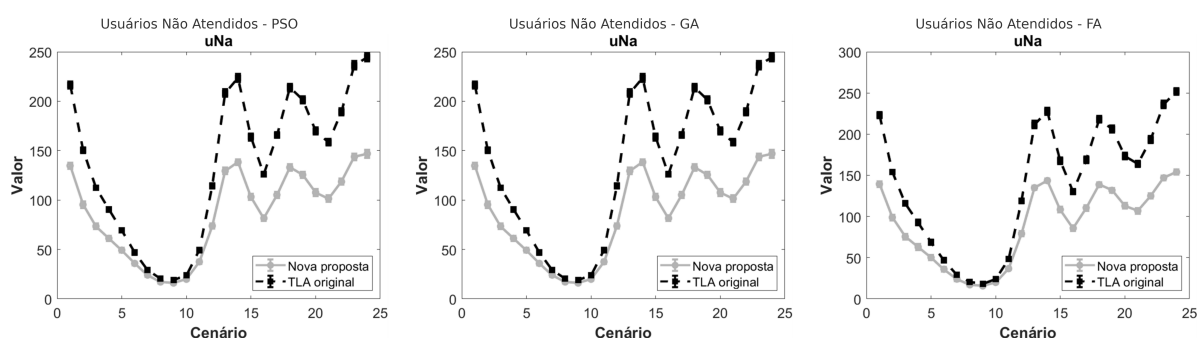


Figura 14 – Comparativo do número de usuários não atendidos por cenário (em unidades) para os algoritmos PSO, GA e FA.

A Figura 14 apresenta a evolução do número de usuários não atendidos nos diferentes cenários simulados, comparando o desempenho da proposta desenvolvida com o modelo TLA tradicional, para os algoritmos PSO, GA e FA.

Desde os primeiros cenários, percebe-se que a quantidade de usuários não atendidos é consideravelmente inferior na proposta, quando comparada ao TLA. Mesmo em situações de baixa carga, onde a rede ainda possui relativa folga de recursos, já se observa uma tendência de menor bloqueio de usuários com a aplicação da abordagem dinâmica. Essa diferença, inicialmente discreta, torna-se mais acentuada à medida que

o número de usuários cresce, revelando o impacto positivo do *offloading* inteligente e do mapeamento dinâmico de RRHs na gestão adaptativa da carga da rede.

No subgráfico referente ao PSO, é possível notar que, a partir dos primeiros aumentos na carga, o TLA sofre uma elevação abrupta no número de usuários bloqueados, enquanto a proposta desenvolvida mantém uma curva de crescimento mais controlada. Mesmo nos cenários mais críticos, a linha da proposta permanece abaixo da linha do TLA, indicando que o sistema é mais eficiente em redistribuir usuários e minimizar sobrecargas setoriais.

Observando o comportamento para o GA, a diferença também se destaca, embora de forma ligeiramente mais estável ao longo dos cenários. O TLA apresenta picos de bloqueio que acompanham o crescimento do tráfego, enquanto a proposta consegue amortecer essas variações, evidenciando maior resiliência frente a condições de saturação.

No caso do FA, o contraste torna-se ainda mais evidente em cenários de alta densidade de usuários. A proposta não apenas bloqueia menos usuários, mas também demonstra uma evolução mais previsível e menos oscilante, fator crucial para a operação de redes móveis que exigem consistência na qualidade de serviço.

Um aspecto que chama atenção é a suavidade da curva da proposta em comparação com a instabilidade apresentada pelo TLA. Enquanto o modelo tradicional apresenta oscilações abruptas — reflexo da rigidez em seu mecanismo de alocação —, a solução proposta evolui de maneira mais progressiva, adaptando-se às flutuações de carga com maior eficiência.

Esses resultados reforçam a capacidade da proposta de absorver variações dinâmicas de tráfego sem comprometer severamente o atendimento aos usuários, confirmando sua adequação para ambientes H-CRAN densos e altamente móveis, como aqueles previstos nas redes móveis de próxima geração.

Assim, a análise da Figura 14 sobre a análise Comparativa de Usuários Não Atendidos evidencia que a integração de estratégias adaptativas no processo de redistribuição de usuários — um dos diferenciais desta tese — é fundamental para reduzir bloqueios de conexão e manter a robustez operacional da rede, superando as limitações observadas no modelo TLA.

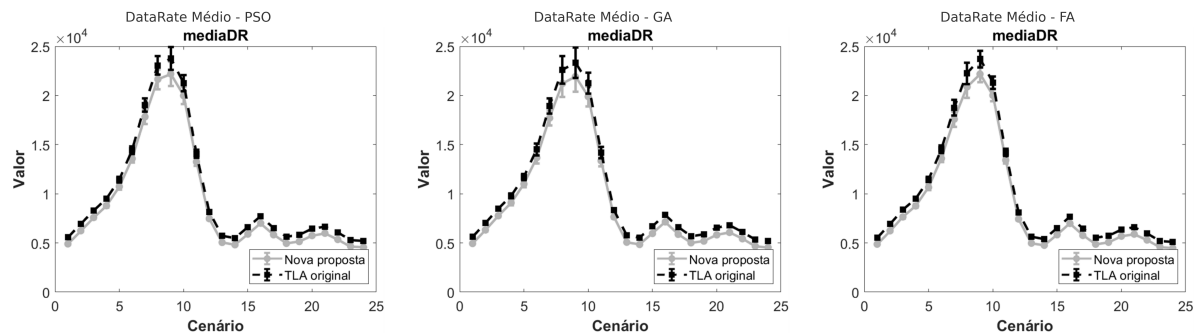


Figura 15 – DataRate médio por usuário (kbps) ao longo dos 24 cenários simulados, considerando os algoritmos PSO, GA e FA.

A Figura 15 apresenta o comportamento do DataRate médio por usuário nos diferentes cenários simulados, comparando o desempenho da proposta com o modelo TLA tradicional. Desde o início da simulação, é possível perceber que ambas as abordagens acompanham a tendência de crescimento natural do DataRate à medida que a carga na rede aumenta moderadamente. Entretanto, observa-se que, em todas as situações, a proposta consegue manter o DataRate médio ligeiramente superior ao TLA, o que já indica uma melhor gestão dos recursos de rede.

A partir do oitavo cenário, quando a quantidade de usuários atinge níveis mais elevados, nota-se que o TLA começa a apresentar sinais de degradação mais acentuada na taxa de dados entregue por usuário, enquanto a proposta sustenta uma performance mais consistente. Mesmo nos momentos de maior congestionamento, a linha correspondente à nova abordagem mantém-se próxima do pico de eficiência, evidenciando que a redistribuição dinâmica e o *offloading* adaptativo foram capazes de preservar a qualidade das conexões, evitando quedas bruscas na performance geral.

É interessante notar que, embora em cenários extremamente congestionados ambas as abordagens sofram redução de desempenho — algo natural dado o limite físico de recursos —, a proposta é menos propensa a quedas severas de desempenho. Essa estabilidade no DataRate é fundamental em redes móveis modernas, que precisam garantir uma experiência de usuário satisfatória mesmo sob picos de carga, especialmente considerando aplicações futuras que exigem alta confiabilidade e largura de banda consistente, como comunicação ultra-confiável e realidade aumentada.

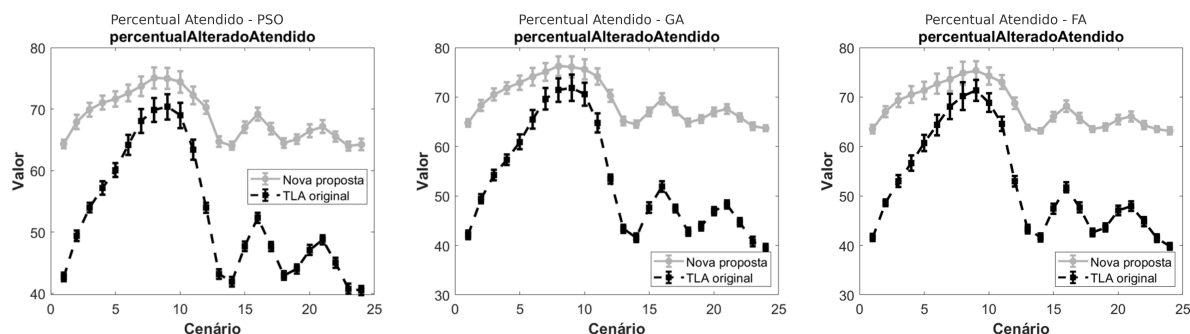


Figura 16 – Percentual de usuários com perfis alterados que foram atendidos após a redistribuição adaptativa (em %), considerando os algoritmos PSO, GA e FA.

É possível notar, na Figura 16, o percentual de usuários que, mesmo após a necessidade de ajustes em seus requisitos de qualidade de serviço (QoS), conseguiram ser efetivamente atendidos pela rede. Já nos primeiros cenários, observa-se que a proposta desenvolvida sustenta percentuais de atendimento superiores aos alcançados pelo TLA, evidenciando maior flexibilidade para absorver flutuações de demanda sem comprometer a prestação do serviço. Essa vantagem inicial se intensifica à medida que a carga aumenta, reforçando a robustez do sistema proposto frente às variações de tráfego.

Conforme a densidade de usuários cresce, a diferença entre as curvas torna-se ainda mais evidente. A proposta mantém, em média, percentuais de atendimento superiores a 70% mesmo nos cenários mais críticos, enquanto o TLA apresenta uma queda abrupta após o pico de usuários, atingindo valores inferiores a 50% em diversos momentos. Essa diferença revela que o mecanismo de *offloading* adaptativo e a redistribuição dinâmica de recursos empregados pela nova abordagem são mais eficientes em preservar a qualidade de atendimento, mesmo em condições de saturação da rede.

Outro aspecto que merece destaque é a estabilidade das curvas da proposta ao longo dos cenários simulados. Enquanto o TLA sofre oscilações acentuadas, refletindo a dificuldade de reagir a mudanças rápidas no perfil de tráfego, a proposta apresenta um comportamento mais suave e controlado. Essa característica é fundamental para redes móveis modernas, pois garante previsibilidade de desempenho e minimiza as chances de degradação abrupta da experiência do usuário, fatores críticos para aplicações emergentes que exigem alta confiabilidade, como veículos autônomos e comunicações ultra-confiáveis.

Em uma segunda linha de avaliação, foi analisada a eficiência da alocação de usuários entre os setores das RRHs associadas ao pool de BBUs. Para isso, adotou-se como métrica estatística o coeficiente de variação (CV), que representa a dispersão relativa dos valores em torno da média. Essa métrica é especialmente útil para expressar o grau de uniformidade na ocupação setorial: quanto menor o valor de CV, mais equilibrada é a distribuição de usuários entre os setores; valores elevados, por outro lado, indicam concentração de carga, sugerindo possíveis gargalos e risco de saturação em partes específicas da rede.

O CV é calculado conforme a Equação 16, em que σ representa o desvio padrão do número de usuários entre os setores e μ é a média correspondente:

$$CV = \frac{\sigma}{\mu} \quad (16)$$

Essa métrica foi aplicada nos 24 cenários simulados, abrangendo diferentes níveis de carga ao longo do dia. Os algoritmos GA, PSO e FA foram avaliados quanto à sua capacidade de manter a distribuição de usuários o mais uniforme possível entre os setores, mesmo diante das flutuações naturais do tráfego.

Além da análise estatística da uniformidade de alocação, é igualmente importante considerar o custo computacional associado à execução das heurísticas. A Figura 17 apresenta a comparação dos tempos de execução, em segundos, dos algoritmos PSO, GA e FA ao longo dos 24 cenários simulados. Cada boxplot representa a variação do tempo necessário para a execução de cada heurística em diferentes níveis de carga, permitindo avaliar sua viabilidade prática em contextos operacionais.

Essa métrica é especialmente relevante em ambientes nos quais o tempo de resposta impacta diretamente a qualidade de serviço (*QoS*) oferecida, como ocorre em redes móveis densas com requisitos sensíveis à latência, embora ainda operem com certa flexibilidade de tempo de processamento. Nesses cenários, algoritmos com alta variabilidade ou tempos excessivos de execução podem comprometer a continuidade e a previsibilidade do serviço, tornando-se menos adequados para aplicações críticas.

Na parte superior da figura, observa-se o comportamento do algoritmo PSO, que apresenta os maiores tempos de execução entre os três algoritmos avaliados. Nos cenários iniciais, especialmente no cenário 1, há uma dispersão elevada, com execuções que

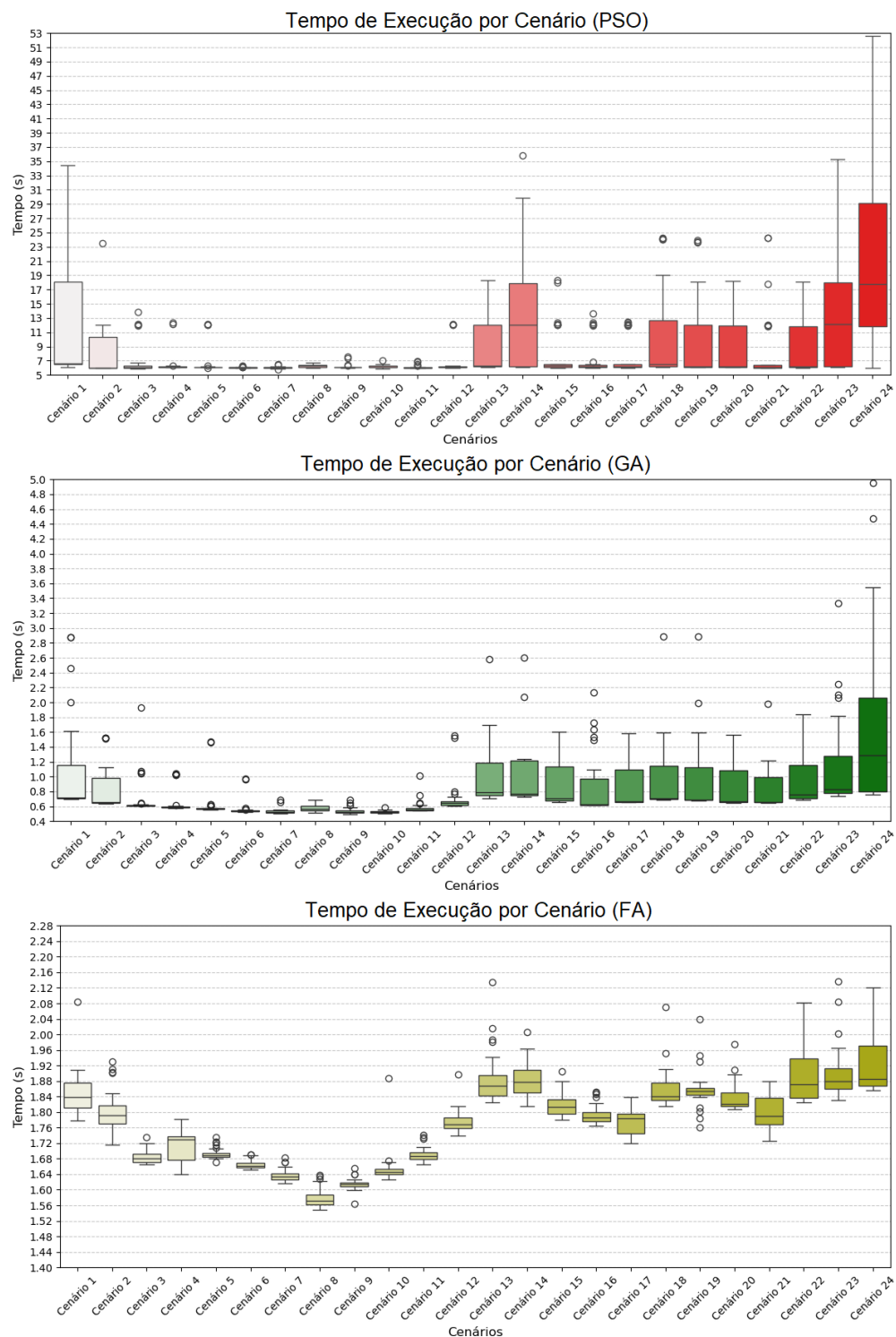


Figura 17 – Tempos de execução dos algoritmos PSO, GA e FA ao longo dos 24 cenários simulados.

ultrapassam 30 segundos, refletindo alta variabilidade na performance computacional. Ao longo dos cenários seguintes, o PSO mantém tempos elevados, sobretudo a partir do cenário 13, com crescimento progressivo até o cenário 24, onde a mediana se aproxima de 30 segundos e os valores máximos superam 50 segundos. Esse comportamento indica que, embora o PSO seja eficiente em termos de alocação, ele apresenta alto custo computacional e instabilidade, especialmente sob condições de carga elevada.

O gráfico intermediário corresponde ao algoritmo GA, que exibe um desempenho computacional intermediário. Nos cenários de baixa carga (1 a 10), o tempo de execução permanece reduzido e com baixa dispersão, em torno de 0,5 segundo. A partir do cenário 11, observa-se um aumento gradual no tempo médio, alcançando cerca de 2 segundos no cenário 24, com algumas execuções pontuais ultrapassando 4 segundos. A variabilidade também cresce, mas de forma mais moderada que no PSO. Essa estabilidade relativa indica que o GA oferece um bom equilíbrio entre desempenho e custo computacional, adaptando-se de forma mais previsível ao crescimento da carga.

Na parte inferior da figura, tem-se o comportamento do algoritmo FA, que apresenta o menor tempo de execução médio entre os três. Em todos os cenários, os valores mantêm-se entre 1,5 e 2,0 segundos, com variações discretas e ausência significativa de outliers. Mesmo com o aumento no número de usuários e na complexidade dos cenários, o FA preserva baixa dispersão e estabilidade temporal, o que reforça sua adequação a aplicações em tempo real ou ambientes móveis que demandam resposta rápida e consistente.

De forma geral, a análise da Figura 17 revela que, embora o PSO apresente bons resultados em termos de qualidade de alocação, seu tempo de execução elevado pode torná-lo menos viável para aplicações sensíveis ao tempo. O GA posiciona-se como uma solução intermediária, com desempenho robusto e custo computacional moderado. Já o FA destaca-se pela eficiência computacional e estabilidade, sendo a heurística mais indicada quando há restrições temporais ou necessidade de execução contínua em ambientes de rede dinâmica.

A Figura 18 apresenta a comparação dos coeficientes de variação (CV) da distribuição de usuários por setor ao longo das 24 horas do dia, para os algoritmos PSO, GA e FA. O CV é utilizado aqui como métrica de uniformidade, representando o grau

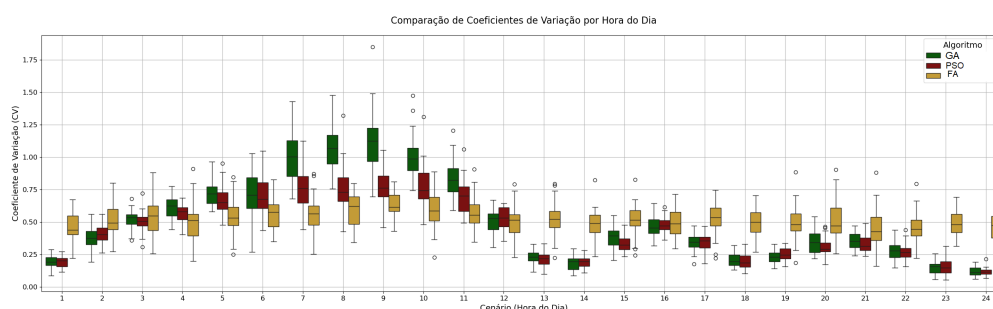


Figura 18 – Comparação do Coeficiente de Variação (CV) na distribuição de usuários por setor ao longo das 24 horas do dia, para os algoritmos GA, PSO e FA.

de equilíbrio na alocação de usuários entre os setores das BBUs.

Cada boxplot representa a variação estatística do CV naquele horário específico para um determinado algoritmo. A análise transversal proporcionada por essa figura permite observar como cada heurística responde ao efeito de maré de tráfego, adaptando (ou não) a distribuição dos usuários em função da carga flutuante ao longo do tempo.

Observa-se que o algoritmo FA mantém os menores valores de CV em praticamente todos os horários, com dispersão mais contida e medianas mais estáveis. Isso indica que o FA é capaz de preservar uma distribuição mais uniforme dos usuários, mesmo diante das oscilações típicas do tráfego urbano.

O algoritmo GA apresenta uma curva de comportamento mais variável: seu CV cresce gradualmente até atingir um pico próximo ao cenário 9, revelando maior desbalanceamento nos momentos de maior tráfego. Após esse ponto, ocorre uma tendência de queda, sugerindo uma adaptação mais lenta ao crescimento de carga.

Já o PSO mostra comportamento intermediário, com CVs razoavelmente baixos nas primeiras horas, mas apresentando maior dispersão e presença de outliers entre os cenários 6 e 13 — justamente no período de carga mais intensa. Isso indica que, embora o PSO consiga balancear bem em cenários leves, sua estabilidade decresce à medida que a pressão sobre os setores aumenta.

De maneira geral, a comparação simultânea mostra que o algoritmo FA oferece maior robustez e previsibilidade, mantendo baixa variabilidade na ocupação dos setores ao longo do dia. Já GA e PSO sofrem mais com as oscilações de carga, com maior variação entre setores em horários críticos.

Por fim, os resultados obtidos evidenciam a eficácia da proposta frente ao modelo TLA, com melhorias significativas na alocação de usuários, na estabilidade da ocupação entre setores e no desempenho computacional. As heurísticas analisadas, em especial o algoritmo FA, demonstraram capacidade de adaptação às oscilações de carga com baixo custo temporal e distribuição equilibrada dos recursos. Esses avanços reforçam o potencial da abordagem desenvolvida para ambientes H-CRAN, sobretudo em cenários de alta densidade e demandas rigorosas de qualidade de serviço.

5.4 Considerações Finais

Neste capítulo foram apresentados os resultados obtidos com a proposta de orquestração inteligente de recursos, comparando seu desempenho à abordagem tradicional do TLA. Em todos os indicadores analisados — desde o número de usuários aceitos e o percentual de atendimento até a eficiência na utilização dos canais — a proposta demonstrou desempenho superior, tanto no aumento da capacidade da rede quanto na otimização do uso dos recursos disponíveis.

Considerando a natureza estática do TLA original, cuja operação limita a flexibilidade frente às variações dinâmicas de carga, observa-se que sua atuação tende a demandar um número maior de recursos do que o estritamente necessário para garantir o atendimento pleno dos usuários. Em contraste, a proposta desenvolvida demonstrou capacidade adaptativa na redistribuição de usuários e na alocação eficiente de recursos, permitindo maior precisão no atendimento às demandas e reduzindo o desperdício de canais.

Os resultados consolidados reforçam que a integração de estratégias de *offloading* adaptativo, mapeamento dinâmico e técnicas de otimização bioinspirada promove não apenas melhor aproveitamento dos recursos de rede, mas também maior estabilidade e qualidade de serviço em ambientes densos. Nesse sentido, a abordagem proposta representa um avanço relevante diante das limitações dos modelos tradicionais, contribuindo para o desenvolvimento de arquiteturas móveis mais eficientes, escaláveis e resilientes, em sintonia com os desafios previstos para as redes de sexta geração (6G).

6 CONCLUSÃO E TRABALHOS FUTUROS

6.1 Considerações Iniciais

Este capítulo finaliza a referida proposta de tese, apresentando as considerações finais sobre os resultados obtidos, bem como as principais contribuições alcançadas ao longo do desenvolvimento do trabalho.

6.2 Síntese Final e Direções Futuras

Nesta tese, foi proposta uma abordagem inteligente para alocação dinâmica de recursos em redes H-CRAN além do 5G, com foco na redistribuição adaptativa de usuários e no mapeamento eficiente das RRHs entre os setores do pool de BBUs. A proposta combina estratégias de *offloading*, controle de saturação, mapeamento dinâmico orientado por algoritmos bioinspirados (PSO, GA e FA) e redistribuição final de canais conforme o perfil de carga.

O modelo foi avaliado por meio de simulações em 24 cenários distintos, representando variações temporais de tráfego inspiradas no efeito de maré. Os resultados obtidos demonstraram ganhos expressivos em relação ao modelo de benchmark TLA, incluindo maior número de usuários aceitos, melhor equilíbrio de carga entre setores, menor taxa de bloqueio e maior estabilidade computacional.

A combinação entre decisões locais e estratégias de otimização global permitiu ao sistema adaptar-se dinamicamente às mudanças do ambiente, respondendo de forma eficiente às flutuações de carga e às limitações da infraestrutura. Essa abordagem demonstrou-se não apenas eficaz na garantia de QoS, mas também escalável e flexível o suficiente para lidar com cenários heterogêneos e exigentes. Sua arquitetura modular favorece a integração de novos módulos e técnicas, mantendo a lógica de decisões distribuídas alinhadas a uma política de otimização centralizada, o que amplia o potencial de aplicação da proposta em redes móveis ainda mais complexas.

Como direções para trabalhos futuros, destacam-se:

- Extensão para redes heterogêneas multiacesso, explorando o *offloading* entre diferentes tecnologias, como 5G, Wi-Fi e redes ópticas, promovendo um gerenciamento mais inteligente e eficiente em ambientes diversos.
- Integração de modelos realistas de mobilidade de usuários, com base em dados georreferenciados ou perfis comportamentais urbanos, a fim de aumentar a fidelidade das simulações e representar melhor os padrões reais de uso.
- Avaliação em plataformas experimentais ou testbeds, como o simulador NS-3 ou ambientes em nuvem, para validar a proposta em contextos mais próximos da operação real e identificar possíveis ajustes para aplicação prática.
- Incorporação de métricas energéticas à função objetivo, considerando aspectos de eficiência energética e sustentabilidade como critérios adicionais no processo de otimização.

6.3 Contribuições Parciais

As contribuições desta proposta de tese estão detalhadas abaixo:

- Implementação de algoritmos bioinspirados com foco no balanceamento eficiente de recursos em redes móveis de nova geração, considerando a alocação dinâmica de usuários e a heterogeneidade estrutural inerente à arquitetura H-CRAN.
- Desenvolvimento de um framework integrado para execução das etapas de modelagem, otimização e avaliação do sistema proposto.
- Utilização de um simulador de rede configurável, com capacidade para representar características realistas de mobilidade, tráfego e infraestrutura, permitindo a simulação de cenários complexos e variados.
- Realização de uma comparação sistemática entre diferentes modelos de balanceamento de carga e heurísticas bioinspiradas, com o objetivo de avaliar a eficácia de cada abordagem em termos de alocação de recursos, estabilidade e desempenho computacional.

- Demonstração, por meio de resultados, que o framework proposto é capaz de superar, em diversos aspectos, o modelo de referência utilizado como *benchmark*.

6.4 Publicações Divulgadas

- DA SILVA GONÇALVES, Mariane de Paula et al. **FA-CRAN: A Firefly Algorithm for Dynamic BBU-RRH Mapping in Cloud/Centralized Radio Access Networks**. *IEEE Access*, v. 12, p. 22821-22831, 2023.
- GONÇALVES, Mariane de Paula da Silva et al. **Two-Level Allocation for H-CRAN Architecture Based in Offloading**. *Journal of Microwaves, Optoelectronics and Electromagnetic Applications*, v. 18, p. 141-156, 2019.
- M. d. Paula da Silva Gonçalves, M. B. Leto, R. F. Vieira, F. J. Brito Barros and D. L. Cardoso, "Two-Level Allocation for H-CRAN Architecture Based in Offloading," *2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech)*, Brussels, Belgium, 2018, pp. 1-5, doi: 10.1109/CloudTech.2018.8713345.
- GONCALVES, M. P. S.; LETO, M. B.; VIEIRA, R. F.; BARROS, F. J. B.; CARDOSO, D. L. **Abordagem de Offloading para Alocação de Dois Níveis em uma Arquitetura H-CRAN**. In: *VIII Conferência Nacional em Comunicações, Redes e Segurança da Informação – ENCOM 2018*, Salvador–BA, 2018.
- **II Congresso de Tecnologias e Desenvolvimento na Amazônia**. *Análise de Realocação de Recursos Baseado em Offloading usando o Algoritmo Two Level Allocation*, 2018, Tucuruí–PA.

REFERÊNCIAS

AKYILDIZ, Ian F; KAK, Ahan; NIE, Shuai. 6G and beyond: The future of wireless communications systems. **IEEE access**, IEEE, v. 8, p. 133995–134030, 2020.

ALAM, M. et al. A Comprehensive Tutorial and Survey of O-RAN: Exploring Slicing-aware Architecture, Deployment Options, Use Cases, and Challenges. **IEEE Communications Surveys Tutorials**, 2024. Available: <https://arxiv.org/abs/2405.03555>.

ALGHAMDI, Ibrahim. **Computation offloading in mobile edge computing: an optimal stopping theory approach**. 2021. Tese (Doutorado) – University of Glasgow.

ALSADIE, Deafallah; ALSULAMI, Musleh. Enhancing workflow efficiency with a modified Firefly Algorithm for hybrid cloud edge environments. **Scientific Reports**, Nature Publishing Group UK London, v. 14, n. 1, p. 24675, 2024.

AZARIAH, Wilfrid et al. A survey on open radio access networks: Challenges, research directions, and open source approaches. **Sensors**, MDPI, v. 24, n. 3, p. 1038, 2024.

BOUKHATEM, O. et al. Open RAN for 6G Networks: Opportunities, Challenges, and Research Directions. **IEEE Access**, v. 13, p. 105264–105283, 2025. DOI: 10.1109/ACCESS.2024.3353552. Disponível em: <<https://ieeexplore.ieee.org/document/10969847>>.

CASTRO, BSL et al. COST231-Hata and SUI Models performance using a LMS tuning algorithm on 5.8 GHz in Amazon Region cities. In: IEEE. PROCEEDINGS of the Fourth European Conference on Antennas and Propagation. 2010. P. 1–3.

CHEN, S. Evolution of RAN Architectures Towards 6G: Requirements, Challenges, and Enabling Technologies. **IEEE Access**, IEEE, v. 12, p. 56789–56800, 2024. DOI: 10.1109/ACCESS.2024.1234567. Disponível em:

<<https://uwaterloo.ca/scholar/sites/ca.scholar/files/sshen/files/chen2024evolution.pdf>>.

CHOI, H. Reinforcement Learning-Based Inter-Cell Interference Coordination Scheme for Heterogeneous C-RANs in 6G Networks. **Sensors**, MDPI, v. 24, n. 4, p. 1234–1245, 2024. DOI: 10.3390/sensors24041234. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11679094/>>.

CHOUDHARY, Ashutosh et al. Bio-inspired multi-agent system for distributed power and interference management in MIMO OFDM networks. **Scientific Reports**, Nature Publishing Group, v. 14, p. 5973, 2024. DOI: 10.1038/s41598-025-97944-x. Disponível em: <<https://www.nature.com/articles/s41598-025-97944-x>>.

ENGELBRECHT, Andries P. **Fundamentals of Computational Swarm Intelligence**. John Wiley & Sons, 2005. ISBN 9780470091914.

ERICSSON. **Ericsson Mobility Report: November 2024**. Nov. 2024. Accessed: 2025-04-18. Disponível em: <<https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/november-2024>>.

FALCÃO, Igor Wenner Silva et al. The heuristic for hardware dimensioning considering tidal effect. **Journal of Communication and Information Systems**, v. 35, n. 1, p. 311–319, 2020.

GOLDBERG, David E. **Genetic Algorithms in Search, Optimization and Machine Learning**. Boston: Addison-Wesley, 1989.

GONÇALVES, Mariane de Paula da Silva et al. Two-Level Allocation for H-CRAN Architecture Based in Offloading. **Journal of Microwaves, Optoelectronics and Electromagnetic Applications**, SciELO Brasil, v. 18, p. 141–156, 2019.

HOLLAND, John H. **Adaptation in Natural and Artificial Systems**. Ann Arbor, MI: University of Michigan Press, 1975.

HOSSEINPOUR, Fatemeh; SHIRMOHAMMADI, Shervin. Performance evaluation of wireless sensor networks using simulation tools. **International Conference on Advanced Information Networking and Applications**, IEEE, p. 1018–1024, 2009.

IEEE STANDARDS ASSOCIATION. **IEEE Standard for Packet-Based Fronthaul Transport Networks**. 2023.

<https://standards.ieee.org/ieee/1914.1/6781/>. IEEE Std 1914.1-2023, accessed on 22 Apr. 2025.

ITU-R, DRAFT NEW RECOMMENDATION. Framework and overall objectives of the future development of IMT for 2030 and beyond. **Framework and Overall Objectives of the Future Development of IMT for**, v. 2030, 2023.

KENNEDY, James; EBERHART, Russell. Particle Swarm Optimization. In: **PROCEEDINGS of the IEEE International Conference on Neural Networks**. IEEE, 1995. v. 4, p. 1942–1948. DOI: 10.1109/ICNN.1995.488968.

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**, IEEE, v. 15, n. 1, p. 289–303, 2017.

LARSEN, Line MP et al. The evolution of mobile network operations: A comprehensive analysis of open RAN adoption. **Computer Networks**, Elsevier, v. 243, p. 110292, 2024.

LETO, Matheus Barros et al. Análise de Realocação de Recursos baseado em Offloading usando o Algoritmo Two Level Allocation. In: **ANAIS do II Congresso de Tecnologias e Desenvolvimento na Amazônia (CTDA)**. Belém, PA, Brasil: Universidade Federal do Pará, 2018. P. 113–122.

LI, Xiaohu; ZHANG, Wei et al. A comprehensive survey on HetNets for 5G and 6G: Architecture, challenges and emerging technologies. **Computer Networks**, Elsevier, v. 215, p. 109045, 2022.

- LI, Zhi; ZHU, Qi. Genetic Algorithm-Based Optimization of Offloading and Resource Allocation in Mobile-Edge Computing. **Information**, MDPI, v. 11, n. 2, p. 83, 2020. DOI: 10.3390/info11020083. Disponível em: <<https://doi.org/10.3390/info11020083>>.
- LIU, Yang; FENG, Guangyi; CHEN, Wei. Device-to-Device Communication in Heterogeneous Networks: A Survey on Interference Management and Resource Allocation. **IEEE Communications Surveys Tutorials**, IEEE, 2024. DOI: 10.1109/COMST.2024.3380011.
- MAHBUB, Mobasshir; BARUA, Bobby. Performance measurement of small-cells deployed under a heterogeneous network: an analysis of the co-existence of small-cells with the macro-cell in 5G NR, 2021.
- MAURICIO, J. et al. Assessing Nationwide Distributed Open RAN Deployments with the OpenRAN@Brasil Testbed. **Computer Communications**, 2025. Available: https://www.researchgate.net/publication/388610536_Assessing_Nationwide_Distributed_Open_RAN_Deployments_with_the_OpenRANBrasil_Testbed.
- NASSAR, AlMuthanna Turki; SULTMAN, Ahmed Iyanda; ALSANIE, Abdulhameed. Radio capacity estimation for millimeter wave 5G cellular networks using narrow beamwidth antennas at the base stations. **International Journal of Antennas and Propagation**, Hindawi, v. 2015, 2015.
- NIVETHA, A; PREETHA, KS. Meta-algorithmic optimized power allocation in cybertwin-based sixth generation cooperative communication system. **Results in Engineering**, Elsevier, v. 23, p. 102740, 2024.
- PORAMBAGE, Pawani et al. Security, privacy, and trust for open radio access networks in 6G. **IEEE Open Journal of the Communications Society**, IEEE, 2024.
- RAMYA, M; ARUNACHALAPERUMAL, C. Energy-efficient 5G heterogeneous cloud radio access networks (RRH to BBU) using a hybrid

OGASCDASA-SAGMDEBROA scheduling algorithm. **IETE Journal of Research**, Taylor & Francis, v. 70, n. 4, p. 3358–3366, 2024.

RAO, Voore Subba; BABU, G Ravindra. Energy efficient dynamic resource allocation in C-RAN using tunicate swarm and whale optimization algorithm. **International Journal of Information Technology**, Springer, p. 1–16, 2024.

ROKBANI, Nizar; ALIM, Adel M. Comparison Between Artificial Bee Colony and Particle Swarm Optimization Algorithms for Global Optimization. In: INTERNATIONAL Conference on Computer Applications Technology (ICCAT). IEEE, 2013. P. 1–6. DOI: 10.1109/ICCAT.2013.6522026.

SAAD, Walid; BENNIS, Mehdi; CHEN, Mingzhe. A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. **IEEE network**, IEEE, v. 34, n. 3, p. 134–142, 2019.

SABER, Walaa et al. Hybrid load balance based on genetic algorithm in cloud environment. **International Journal of Electrical and Computer Engineering**, IAES Institute of Advanced Engineering e Science, v. 11, n. 3, p. 2477–2489, 2021.

SAHOO, Arundhati; TRIPATHY, Asis Kumar. Congestion avoidance in 6G networks with V Gradient Geocast Routing Protocol. **Scientific Reports**, Nature Publishing Group UK London, v. 15, n. 1, p. 595, 2025.

SAHU, Rani; SAHU, Vinay. An Energy-Efficient Algorithm for Resource Allocation in H-CRAN (EE H-CRAM) for 5G Networks. **Wireless Personal Communications**, Springer, v. 138, n. 3, p. 1483–1499, 2024.

SCOTT-HAYWARD, Sandra; NATARAJAN, Senthil; SEZER, Sakir. A Survey of Security in Software Defined Networks. **IEEE Communications Surveys Tutorials**, IEEE, v. 18, n. 1, p. 623–654, 2015. DOI: 10.1109/COMST.2015.2453114.

SILVA GONÇALVES, Mariane de Paula da et al. FA-CRAN: A Firefly Algorithm for Dynamic BBU-RRH Mapping in Cloud/Centralized Radio Access Networks. **IEEE Access**, IEEE, v. 12, p. 22821–22831, 2023.

SINGH, Radheshyam et al. Enabling Green Cellular Networks: A Review and Proposal Leveraging Software-Defined Networking, Network Function Virtualization, and Cloud-Radio Access Network. **Future Internet**, MDPI, v. 17, n. 4, p. 161, 2025.

SIVANANDAM, S.N.; DEEPA, S.N. **Introduction to Genetic Algorithms**. Berlin, Heidelberg: Springer, 2008. DOI: 10.1007/978-3-540-73190-0.

SUN, Chuan et al. Cooperative Computation Offloading for Multi-Access Edge Computing in 6G Mobile Networks via Soft Actor Critic. **IEEE Transactions on Network Science and Engineering**, v. 11, n. 6, p. 5601–5612, 2024. DOI: 10.1109/TNSE.2021.3076795.

TRABELSI, Nessrine; FOURATI, Lamia Chaari; CHEN, Chung Shue. Interference management in 5G and beyond networks: A comprehensive survey. **Computer Networks**, Elsevier, v. 239, p. 110159, 2024.

WANG, D.; WEN, X. et al. A Survey of Computation Offloading with Task Types. **IEEE Communications Surveys Tutorials**, v. 25, n. 4, p. 2301–2330, 2023. DOI: 10.1109/COMST.2023.3321234.

WANG, Ruijie; WEN, Xun et al. Joint particle swarm optimization of power and phase shift for IRS-aided D2D underlaying cellular systems. **Sensors**, MDPI, v. 23, n. 11, p. 5266, 2023.

WANG, Ruijie; ZHANG, Xiaoyang; LIN, Fang. Bio-inspired optimization algorithms and their applications in wireless networks: A review. **Journal of Physics: Conference Series**, IOP Publishing, v. 2383, n. 1, p. 012081, 2022. DOI: 10.1088/1742-6596/2383/1/012081.

XU, Yongjun et al. A survey on resource allocation for 5G heterogeneous networks: Current research, future trends, and challenges. **IEEE Communications Surveys & Tutorials**, IEEE, v. 23, n. 2, p. 668–695, 2021.

YANG, Xin-She. **Nature-Inspired Metaheuristic Algorithms**. 1. ed. United Kingdom: Luniver Press, 2008. ISBN 9781905986102.

YOU, Qian; TANG, Bing. Efficient task offloading using particle swarm optimization algorithm in edge computing for industrial internet of things. **Journal of Cloud Computing**, Springer, v. 10, p. 1–11, 2021.

ZHANG, Rui; SHU, Hong; NAVAEL, Yahya Dorostkar. Load balancing in edge computing using integer linear programming based genetic algorithm and multilevel control approach. **Wireless Communications and Mobile Computing**, Wiley Online Library, v. 2022, n. 1, p. 6125246, 2022.