

# SEQUOIA - Solução para Energia e Qualidade de serviço Otimizadas em Infraestruturas Abertas

Maria Barbosa<sup>1</sup>, Matheus Pinheiro<sup>1</sup>, Marcos Alves<sup>1</sup>, Anderson Queiroz<sup>1</sup>, Kelvin Dias<sup>1</sup>

<sup>1</sup>Centro de Informática – Universidade Federal de Pernambuco (CIn - UFPE)  
Recife – PE – Brasil

{mksb, mfp2, mgma, aalq, kld}@cin.ufpe.br

**Abstract.** *The radio access network (RAN) is the most energy-consuming component in the 5G system. Overall, 5G RAN energy saving relies on switching on/off base station or radio components leveraging proprietary solutions. Furthermore, service categories of 5G such as URLCC and eMBB have stringent quality-of-service (QoS) requirements, which demand flexible mechanisms to achieve energy-efficient RAN operation while keeping adequate levels of QoS for those applications. Open RAN enables real-time monitoring and network adaptability through open interfaces and RAN intelligent controllers. Thus, this paper proposes SEQUOIA, a solution based on Open RAN and a multi-objective optimization to guarantee QoS and reduce energy consumption.*

**Resumo.** *A rede de acesso rádio (RAN) é o componente que mais consome energia no sistema 5G. No geral, a economia de energia da RAN 5G depende da ativação/desativação de estação base ou componentes de rádio, utilizando soluções proprietárias. Além disso, as categorias de serviço do 5G, como URLCC e eMBB, possuem requisitos rigorosos de qualidade de serviço (QoS), o que exige mecanismos flexíveis para alcançar uma operação eficiente em termos de energia na RAN, mantendo níveis adequados de QoS para essas aplicações. Open RAN permite monitoramento em tempo real e adaptabilidade da rede por meio de interfaces abertas e controladores inteligentes da RAN. Assim, este artigo propõe o SEQUOIA, uma solução baseada em Open RAN e otimização multiobjetivo para garantir QoS e reduzir o consumo de energia.*

## 1. Introdução

O setor de telecomunicações está entre os maiores consumidores de energia elétrica no Brasil, ocupando a quarta posição entre os setores comerciais, com um consumo total de 3.591 gigawatt-hora (GWh), equivalente a 5,2% da participação nacional <sup>1</sup>. Com a implementação do 5G, projeta-se um aumento no consumo energético das redes de telecomunicações, estimado entre 150% e 170% até 2026. Este crescimento é impulsionado pela necessidade de suportar mais dispositivos conectados e atender aos rigorosos requisitos de qualidade de serviço (QoS) de aplicações, como realidade virtual e aumentada, *cloud gaming* e transmissão de vídeo, que demandam maior capacidade de transmissão e, consequentemente, mais energia [Gao et al. 2023].

Esses cenários exigem o uso de recursos específicos na rede, como o *Multiple Input Multiple Output* (MIMO), fazendo com que a Rede de Acesso de Rádio (RAN)

---

<sup>1</sup><https://revistapotencia.com.br/portal-potencia/artigos/como-o-5g-ira-impactar-o-consumo-de-energia>

5G (*Next Generation Node B* - gNodeB) seja a principal responsável pelo consumo energético, representando cerca de 70% do total na infraestrutura 5G [Kolts et al. 2021]. No entanto, estima-se que 50% das gNodeBs recebem apenas 5% do tráfego dos usuários [Perner 2024], o que, em casos de superdimensionamento, pode agravar esse consumo, mesmo quando os recursos não estejam sendo plenamente utilizados.

Tais desafios no ecossistema de telecomunicações intensifica a crescente demanda por soluções que permitam uma maior flexibilidade na rede. Nesse contexto, a arquitetura de Redes de Acesso de Rádio Abertas (*Open Radio Access Networks* - *Open RAN*) surge como uma tecnologia habilitadora, caracterizando-se pela desagregação dos componentes da estação rádio-base e por sua capacidade de promover interoperabilidade entre fornecedores através de interfaces abertas, permitir programabilidade e inteligência na RAN. Os Controladores Inteligentes da Rede de Acesso (*RAN Intelligent Controllers* - RICs) definidos por *software* existentes na arquitetura Open RAN possibilitam o controle e otimização do uso dos recursos disponíveis na gNodeB, que podem colaborar para a redução do consumo de energia [Li et al. 2024].

A O-RAN Alliance<sup>2</sup>, responsável pelas especificações da Open RAN, na fase 1 dos casos de uso, inclui o Projeto de *Hardware* de Caixa Branca, que promove a eficiência energética. Essa integração foi abordada em vários artigos recentes. Alguns estudos utilizam Open RAN para ativar/desativar células [Liang et al. 2024][Bordin et al. 2024], ajustar configurações e otimizar os recursos disponíveis [Urumkar et al. 2023][Hoffmann 2024] [Demir et al. 2024] visando eficiência energética. Contudo, observa-se uma limitação na literatura em relação a soluções que busquem otimizar as configurações da rede para aplicações *Enhanced Mobile Broadband* (eMBB) e *Ultra-Reliable Low Latency Communications* (URLLC), mantendo os requisitos mínimos de desempenho da rede enquanto aumentam os recursos disponíveis para atender a esses requisitos e garantir um menor consumo de energia.

Sendo assim, este trabalho propõe uma solução para otimizar a energia e a qualidade de serviço em infraestruturas abertas (SEQUOIA). A abordagem busca mitigar o superdimensionamento da rede por meio de uma otimização multiobjetivo, utilizando as principais métricas de desempenho (*Key Performance Metrics* – KPM) para configurar a RAN de modo a atender aos requisitos mínimos da aplicação, considerando os compromissos inerentes. Como resultado, a solução contribui para a redução da pegada de carbono da gNodeB e o aumento da eficiência energética. Com isso, as principais contribuições deste estudo são:

- **Otimização de recursos visando redução no consumo de energia:** Desenvolvimento do SEQUOIA, uma solução que busca eficiência energética ciente de QoS para apoiar a tomada de decisões e realizar a orquestração de nós E2 (gNodeBs). A proposta também contempla dois xApps (eXtended application): o primeiro, utilizado para o monitoramento das KPMs e armazenamento dessas informações em banco de dados para alimentar o SEQUOIA; o segundo, para ajustar os parâmetros de rádio e redirecionar o tráfego conforme necessário. Por fim, a tomada de decisão da otimização é implementada como um rApp.
- **Concepção de um *testbed* e avaliação de desempenho:** Avaliamos a solução proposta usando um *testbed* baseado em *Software Defined Radio* (SDR) para co-

---

<sup>2</sup><https://www.o-ran.org/>

letar dados, projetar e testar a solução em diferentes casos de uso: *streaming* de vídeo baseado em Computação de Borda Multi-acesso (*Multi-access Edge Computing* - MEC) e *Cloud Gaming*. As métricas de avaliação de desempenho incluem atraso, vazão, perda, tamanho de fila e potência média utilizada sob tráfego dos dois casos de uso.

As próximas seções foram estruturadas da seguinte forma: as Seções 2 e 3 apresentam os conceitos fundamentais utilizados neste artigo. A Seção 4 descreve os trabalhos relacionados à utilização de Open RAN para eficiência energética. A proposta SEQUOIA e seus componentes arquiteturais são detalhados na Seção 5. A Seção 6 apresenta a concepção de um protótipo de baixo custo para rede 5G. A Seção 7 detalha a avaliação de desempenho bem como os resultados obtidos. Por fim, as conclusões e trabalhos futuros são apresentados na Seção 8.

## 2. Relação entre Vazão e Consumo de Energia

Esta seção aborda os impactos das tecnologias habilitadoras do 5G, com foco no aumento da vazão e nas implicações do maior consumo de energia da RAN. De acordo com a especificação TS 38.306 do 3GPP, a vazão ( $R$ ) em uma rede 5G NR, considerando uma única portadora, é expressa pela equação a seguir [3GPP 2022]:

$$R = \nu_{Layers} Q_m f R_{max} \frac{12 N_{PRB}^{\mu}}{T_s^{\mu}} (1 - OH) \quad (1)$$

Na equação acima,  $\nu_{Layers}$  representa o número de camadas, referente ao uso de uma ou múltiplas antenas como transmissores e receptores, configurando o *Multiple Input Multiple Output* (MIMO), como nas configurações 2x2 ou 4x4. O termo  $Q_m$  refere-se à ordem de modulação máxima suportada. O fator  $f$  é um fator de escala cujo valor adotado para uma única portadora é 1. O parâmetro  $R_{max}$  é a taxa máxima de código alvo, conforme as tabelas de índice *Modulation Coding Scheme* (MCS) da TS 13.214, que variam entre 948 e 1024, o que resulta em uma eficiência espectral de 5,5547 bps/Hz com 64 *Quadrature Amplitude Modulation* (QAM) e 7,4042 bps/Hz com 256-QAM. Já o termo  $N_{PRB}^{\mu}$  se refere ao número máximo de blocos de recursos utilizados, sendo que  $\mu$  corresponde à numerologia (*Subcarrier Spacing* - SCS). Neste estudo, utiliza-se a SCS de 30 kHz, totalizando  $N_{PRB}$  igual a 51 em uma banda de 20 MHz. O parâmetro  $T_s^{\mu}$  é a duração média do símbolo *Orthogonal Frequency Division Multiplexing* (OFDM) em um subquadro para a numerologia correspondente. Para uma SCS de 30 kHz, essa duração é de 35.714  $\mu s$ . Por fim,  $OH$  refere-se ao *overhead* que, para o intervalo de frequências 1 (FR1) e frequências abaixo de 6 GHz (Sub-6GHz), assume os valores de 0,14 para o *downlink* e 0,08 para o *uplink*. Como ilustrado na equação 1, a vazão, em configurações sem agregação de portadora, depende diretamente da modulação e do número de camadas.

### 2.1. Modulação

A modulação é uma técnica que manipula as características de uma onda portadora por meio de uma onda modulante, permitindo otimizar diversos aspectos da transmissão de dados. No contexto do 5G NR, são empregadas diferentes técnicas de modulação, cada uma com um propósito específico. Dentre essas técnicas, destacam-se a QPSK, 16-QAM, 64-QAM e 256-QAM [Info 2024]. A modulação QAM utiliza duas ondas portadoras em

quadratura, possibilitando a transmissão de mais bits por símbolo. Com a utilização dessas duas ondas, a modulação se torna bidimensional, permitindo a ampliação dos níveis de amplitude, conhecidos como estados. Cada estado é associado a uma combinação única de bits. Por exemplo, na 64-QAM, há sessenta e quatro estados, correspondendo a seis bits por símbolo. Assim, a quantidade de bits transmitidos por símbolo impacta diretamente na vazão de dados [Thangamayan et al. 2022]. Contudo, o aumento da modulação resulta em um incremento na potência de transmissão [Chauhan et al. 2019], o que, por sua vez, pode levar a um maior consumo de energia.

## 2.2. MIMO

O *Multiple Input Multiple Output* (MIMO) utiliza múltiplas antenas para transmitir e receber sinais, aproveitando o fenômeno natural de *multipath* (múltiplos caminhos), onde as ondas de rádio percorrem diferentes trajetórias até o destino. Com isso, as antenas combinam as informações recebidas, melhorando a transmissão. O MIMO aumenta a capacidade de dados e reduz erros, usando múltiplas antenas para enviar e receber dados simultaneamente, o que é chamado de *Spatial Multiplexing* (SM). Essa técnica maximiza a utilização do sistema sem precisar aumentar a potência ou a largura de banda, resultando em maior capacidade, menos congestionamento e mais estabilidade na rede [Pereira de Figueiredo 2022]. Entretanto, existem desvantagens a se considerar antes de optar pelo uso do MIMO, uma delas sendo o consumo de energia. Isto ocorre pela necessidade do uso de múltiplas antenas e correntes de transmissão, o que demanda um maior poder de processamento. Devido ao aumento de componentes de *hardware* e a complexidade envolvida em gerenciar os diferentes fluxos de dados, e posteriormente decodificá-los nos receptores, o consumo energético pode aumentar consideravelmente, especialmente em sistemas maiores como no *Massive MIMO* [Mahmood et al. 2018].

## 3. Visão Geral Open RAN

A *Open RAN* representa uma evolução significativa nas estruturas das redes móveis, promovendo a desagregação dos componentes de hardware e software e a adoção de interfaces abertas e interoperáveis. Seu principal objetivo é proporcionar maior flexibilidade, inovação e redução de custos, permitindo que operadoras integrem soluções de múltiplos fornecedores conforme suas necessidades e requisitos de cada ambiente [Azariah et al. 2024].

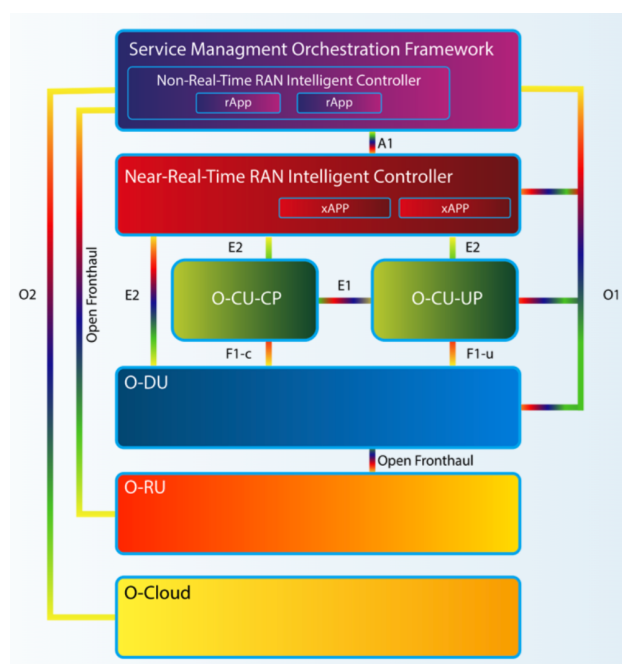
A iniciativa da *Open RAN* ganhou destaque com a formação da O-RAN Alliance em 2018, uma colaboração global que reúne operadoras de telecomunicações, fabricantes e instituições de pesquisa. A aliança visa definir especificações técnicas e promover a padronização das interfaces abertas no RAN.

O financiamento e o gerenciamento da evolução da *Open RAN* são impulsionados por um ecossistema diversificado, incluindo investimentos de operadoras de telecomunicações, fornecedores de tecnologia e apoio de entidades governamentais interessadas em fomentar a inovação e a competitividade no setor. Empresas líderes, têm desempenhado um papel significativo na industrialização e no desenvolvimento de soluções Open RAN, contribuindo para a maturidade e a adoção dessa arquitetura no mercado global [Constantine and Akhtar 2024].

A *Open RAN* surge como uma solução para à necessidade de redes mais abertas e flexíveis, idealizado por um consórcio de atores do setor de telecomunicações e financiado

por investimentos conjuntos das corporações e instituições de padronização comprometidas com esta transformação e inovação tecnológica, possibilitando o crescimento de todo ecossistema beneficiando toda sociedade.

A Figura 1 apresenta a arquitetura Open RAN proposta pela O-RAN Alliance. O Controlador Inteligente da Rede de Acesso (*RAN Intelligent Controller* - RICs) é dividido em dois tipos: o Near-RT RIC, que opera em intervalos de 10 ms a 1 s, e o Non-RT RIC, que funciona em escalas maiores que 1 segundo. O Non-RT RIC, parte da estrutura *Service Management and Orchestration* (SMO), gerencia o consumo de hardware, o ciclo de vida das funções de rede e a orquestração das gNodeBs (nós E2).



**Figura 1. Arquitetura Open RAN - O-RAN Alliance.**

O Near-RT RIC gerencia os nós E2, xApps e suas interações. A interface E2 conecta o Near-RT RIC aos nós E2, permitindo ciclos de controle e alguns procedimentos, como o RIC *Subscription* que gerencia as assinaturas de xApps, RIC *Indication* que envia notificações da gNodeB e o RIC *Control* que envia ações de controle baseadas nos xApps. A plataforma Open RAN utilizada neste trabalho é a *O-RAN Software Community* (OSC), uma iniciativa de código aberto da O-RAN Alliance e da *Linux Foundation*, que desenvolve *software* conforme as especificações da *O-RAN Alliance*, incluindo componentes para Near-RT e Non-RT RIC, xApps como *Traffic Steering* e *Anomaly Detection* (AD).

#### 4. Trabalhos Relacionados

Esta seção apresenta o estado da arte em termos de soluções para eficiência energética em arquiteturas Open RAN. A Tabela 1 destaca os principais aspectos dos estudos que relacionam a este trabalho de pesquisa. O estudo [Urumkar et al. 2023] inova ao tratar da economia de energia em nível de processamento, propondo um escalonamento dinâmico de CPUs no O-Cloud, oferecendo benefícios computacionais, mas com impacto restrito à camada de infraestrutura. A pesquisa [Liang et al. 2024] explora o potencial de xApps inteligentes para reduzir o consumo energético em múltiplos elementos da RAN, destacando

a modularidade da arquitetura, mas carecendo de validações em redes reais. O estudo [Hoffmann 2024] emprega reconfigurações de canais de radiofrequência (RF) utilizando modelos baseados em heurísticas de carga, apresentando implementações simples, porém com flexibilidade limitada frente a cenários de alta variabilidade no tráfego.

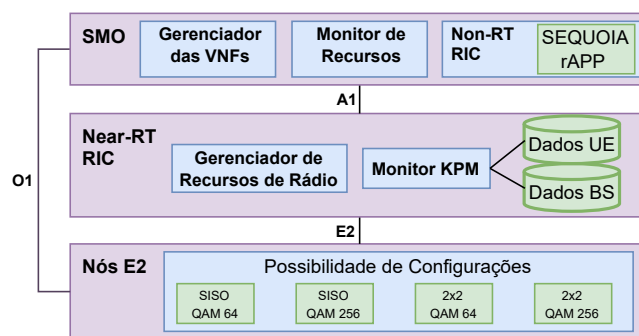
Tabela 1. Sumário dos Trabalhos Relacionados				
Artigo	Solução Proposta	Ambiente de Avaliação	Tomada de Decisão	Métricas e Cenários de Avaliação
[Urumkar et al. 2023]	Algoritmo parametriza os núcleos de CPU dinamicamente com base na carga de trabalho	Simulação	Algoritmo de Monitoramento	Teste de Impacto do Algoritmo de agendamento para Energia, CPU e Banda de Rede
[Liang et al. 2024]	xApps para gerenciar o estado operacional dos rádios	Simulação (TeraVM)	Medições orientadas por KPMs	Consumo de energia em diferentes cargas de trabalho e número de <i>radio cards</i> desligados
[Hoffmann 2024]	rApp DRL para reconfiguração de canais de RF em Open RAN	Simulação em Python	Aprendizado por Reforço baseado no ES-rApp projetado com um agente	Comparação com algoritmos de referência, para carga e ML.
[Demir et al. 2024]	Orquestração de Recursos com Consciência Energética na RAN	Simulação	Uso de RICs para otimização e coordenação de recursos	Aproximações para resolver restrições de potência e capacidade de processamento
[Bordin et al. 2024]	Economia energética utilizando DRL através de um xApp que ativa ou desativa a RF	Simulação (ns-O-RAN)	Um xApp gerencia células 5G em diferentes modos de economia de energia	Eficiência energética, vazão, falhas de enlace e custo de ativação
SEQUOIA	Solução baseada em Open RAN para a flexibilização no uso de recursos da RAN visando reduzir o consumo de energia	Ambiente Real	Algoritmo de otimização multi-objetivo	eMBB: vazão e tamanho de fila / URLLC: perda e latência. Consumo de energia

Já em [Demir et al. 2024], é proposta uma solução integrada que orquestra recursos de rádio, transporte óptico e processamento em nuvem para redes *Cell-Free Massive MIMO*, fornecendo ganhos significativos em eficiência energética em sistemas distribuídos, embora a complexidade de coordenação limite sua escalabilidade em redes densas. Por fim, o trabalho de [Bordin et al. 2024], o foco recai sobre algoritmos de *Deep Reinforcement Learning* (DRL) para controle dinâmico da ativação e desativação de *frontends* RF, destacando ganhos adaptativos em cenários dinâmicos, mas com a limitação de depender de treinamento off-line extensivo. Esses trabalhos buscam evidenciar os possíveis avanços com relação à eficiência energética que podem ser obtidos quando da implementação dos sistemas de comunicação Open RAN, com diferentes compromissos entre complexidade, adaptabilidade e aplicabilidade.

A partir do levantamento do estado da arte apresentado, identificou-se a inexistência de artigos que levem em consideração KPMs em conjunto com técnicas de otimização visando fornecer os requisitos mínimos de QoS de aplicações eMBB e URLLC evitando superdimensionamento, garantindo uma redução do consumo de energia, em ambientes reais. Sendo assim, este trabalho, localizado na última linha na Tabela 1, propõe SEQUOIA, uma solução para mitigar a utilização excessiva de recursos da rede por meio de uma otimização multiobjetivo, baseado nas KPMs capturadas por um xApp para determinar a melhor configuração de recursos a ser utilizada diante do contexto da rede. Além disso, para a implementação e avaliação de desempenho da proposta foi empregado um protótipo de rede 5G utilizando placas SDR e plataformas de código aberto.

## 5. SEQUOIA: Eficiência Energética Baseada em O-RAN

Esta seção apresenta os componentes arquiteturais e a solução de otimização energética proposta neste artigo. O SEQUOIA visa otimizar a Eficiência Energética, utilizando otimização multiobjetivo e incorporando as métricas de QoS do usuário, permitindo uma tomada de decisão para a orquestração dos nós E2. A arquitetura ilustrada na Figura 2 é composta por três blocos principais: o SMO, o Near-RT RIC e os nós E2. O SMO dispõe de funcionalidades para o gerenciamento das funções de rede virtuais (*Virtual Network Functions* - VNFs), o monitoramento do uso dos recursos de *hardware* e o SEQUOIA rApp, localizado no Non-RT RIC. O Near-RT RIC contém os xApps responsáveis pelo monitoramento das métricas de desempenho (*Key Performance Metrics* - KPM) e pela gestão dos recursos de rádio. Por fim, os nós E2 apresentam as diferentes configurações utilizadas neste estudo.



**Figura 2. Visão Geral da Arquitetura Proposta.**

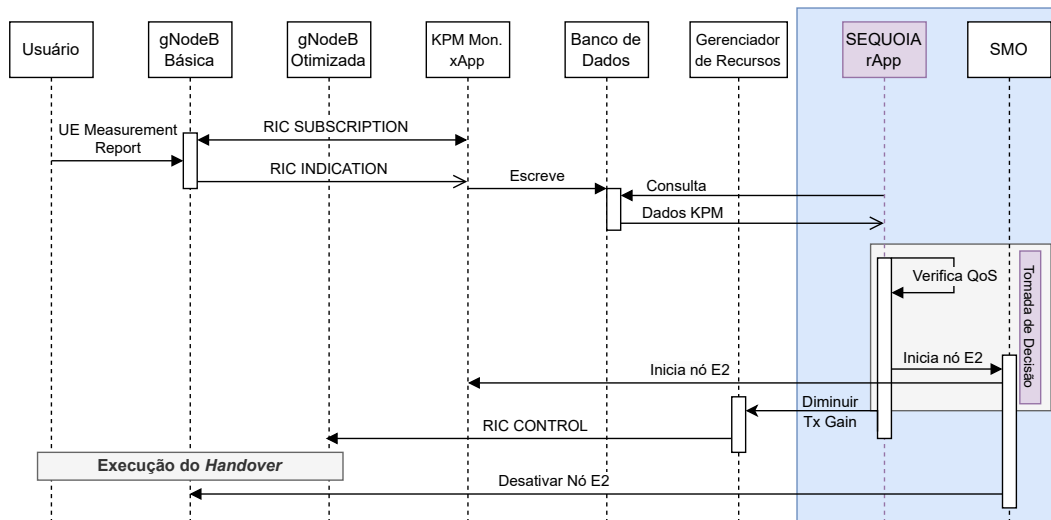
Os nós E2 podem ser configurados de quatro maneiras diferentes neste trabalho. A configuração básica utiliza uma única camada, caracterizando um sistema *Single Input Single Output* (SISO) com modulação 64-QAM, alcançando até 64 Mbps, conforme a Equação 1. A segunda configuração usa a modulação 256-QAM, mantendo o sistema SISO. A terceira configuração amplia para um sistema MIMO 2x2 com duas camadas, mantendo a modulação 64-QAM. Por fim, a última versão combina as duas melhorias (MIMO 2x2 e 256-QAM), com uma vazão teórica de 436 Mbps, conforme a Equação 1.

O xApp Monitor KPM é responsável pelo recebimento e armazenamento das informações dos usuários e da estação rádio base (BS), sendo essas informações de QoS dinâmicas, dependendo da aplicação utilizada na rede, como, por exemplo, no caso de

URLLC, onde são monitorados a latência e a taxa de perdas. Além disso, o xApp de gerenciamento de recursos é responsável pelo aumento e diminuição da amplificação do sinal transmitido pela estação base, influenciando o alcance e a qualidade do sinal. O SMO é responsável pela gestão do ciclo de vida dos nós E2, ou seja, pela orquestração e automatização da configuração desses nós, além de realizar o monitoramento do uso de recursos de hardware. No SMO, encontra-se o Non-RT RIC, que contém os rApps. O rApp desenvolvido neste artigo é denominado SEQUOIA, e é responsável pela tomada de decisão sobre qual a melhor configuração entre as possíveis para os nós E2 que será utilizada.

### 5.1. Sinalização da proposta SEQUOIA

A Figura 3 apresenta a sinalização da solução proposta. Inicialmente, o Monitor KPM envia uma subscrição para o serviço de monitoramento de um nó E2 (*RIC SUBSCRIPTION*), especificando as métricas desejadas. O nó E2 associa essa solicitação a um *callback*, atendendo à solicitação de subscrição do xApp. Em seguida, o usuário envia *Measurement reports* para a gNodeB, que converte essas informações em uma *RIC INDICATION* a ser enviada para o Monitor KPM. Em conjunto as informações do usuário, a gNodeB adiciona suas métricas de observação, tais como o *Physical Resource Block (PRB) Usage*. Esses dados são transmitidos para o xApp a cada segundo e ao recebê-los, os armazena em um banco de dados, tornando-as acessíveis para o processo de tomada de decisão.



**Figura 3. Fluxo de Sinalização da Solução Proposta.**

Após a disponibilização dos dados, o SEQUOIA realiza consultas periódicas para analisar o estado do usuário e da gNodeB. Essa verificação consiste na aplicação dos dados na otimização multiobjetivo, avaliando possíveis degradações das KPMs. Se nenhuma degradação for detectada, o sistema mantém o ciclo de monitoramento. Caso contrário, o SEQUOIA notifica o gerenciador das VNFs no SMO, solicitando a inicialização de um nó E2 otimizado. Uma vez ativado o novo nó E2, o SEQUOIA envia uma mensagem ao gerenciador de recursos no Near-RT RIC para reduzir o ganho de transmissão da gNodeB básica. Com a diminuição da amplificação do sinal, o handover é executado, transferindo a UE para a gNodeB otimizada. Por fim, a gNodeB básica é desativada.



## 5.2. Formulação da Política

Esta seção descreve a formulação da política do SEQUOIA, incluindo seleção das métricas e do algoritmo para tomada de decisão. A metodologia selecionada para a realização de verificação de QoS, que disparará a inicialização do nós E2, foi baseada em otimização multiobjetivo. O Algoritmo 1 apresenta o funcionamento do SEQUOIA para o gerenciamento de políticas levando em consideração tanto informações do usuário quanto da gNodeB para realizar a tomada de decisão.

---

**Algorithm 1:** Otimização Multiobjetivo para Seleção da Configuração do nó E2

---

**Input:** KPMs,  $\alpha_1, \alpha_2, \gamma_{i1}, \gamma_{i2}$

**Function** CalcularCusto ( $S_t, \alpha_1, \alpha_2, \gamma_{i1}, \gamma_{i2}$ ):

- └ **return**  $J_i = \alpha_1 \cdot \gamma_{i1} \cdot S_t[0] + \alpha_2 \cdot \gamma_{i2} \cdot S_t[1]$ ;

**Function** AtualizarHistorico ( $Hist, C_{ótima}, N$ ):

- └ Adicionar  $C_{ótima}$  ao histórico  $Hist$ ;
- └ **if**  $|Hist| > N$  **then**
- └ Remover a configuração mais antiga de  $Hist$ ;
- └ **return**  $Hist$ ;

**Function** EscolherConfiguração ( $Hist$ ):

- └ **return** Configuração mais frequente em  $Hist$ ;

**Begin**

- └ Inicializar histórico  $Hist = []$ ;
- └ **foreach** estado  $S_t$  **do**
- └ **foreach** configuração  $C_i$  **do**
- └ Calcular  $J_i$  usando CalcularCusto ( $S_t, \alpha_1, \alpha_2, \gamma_{i1}, \gamma_{i2}$ );
- └ Selecionar  $C_{ótima} = \arg \min J_i$ ;
- └ Atualizar o histórico  $Hist$  com
- └ AtualizarHistorico ( $Hist, C_{ótima}, N$ );
- └ Estabilizar a recomendação com
- └ EscolherConfiguração ( $Hist$ );
- └ **return** Retornar o configuração sugerida;

---

Conforme apresentado no Algoritmo 1, a tomada de decisão é dividida em três etapas principais: Inicialmente, o modelo recebe o estado atual da rede ( $S_t$ ), que corresponde às métricas de perda de pacotes ( $S_t[0]$ ) e atraso ( $S_t[1]$ ), para as aplicações URLLC e tamanho de fila do buffer ( $S_t[0]$ ) e vazão ( $S_t[1]$ ) para o eMBB. A partir desse  $S_t$  é calculado o custo referente ao uso de cada uma das configurações. Ou seja, ao utilizar uma configuração baseada em SISO e 64-QAM para apenas um usuário, o custo  $J_{i1}$  seria o mesmo para todas as configurações, independente da aplicação. Com a chegada de  $K$  usuários gera-se um novo custo  $J_{i2}$ , onde  $J_{i1} < J_{i2}$ .

A função de custo adotada segue um modelo multiobjetivo, no qual os pesos ( $\alpha_1$ ) e ( $\alpha_2$ ) determinam a prioridade entre diferentes métricas. Para aplicações do tipo URLLC,  $\alpha_1$  corresponde à perda de pacotes, enquanto  $\alpha_2$  representa o atraso. Já no caso do eMBB,

$\alpha_1$  está associado ao tamanho da fila do buffer e  $\alpha_2$  à vazão. Essas ponderações seguem a fronteira de Pareto, uma vez que o problema busca otimizar múltiplas métricas de QoS minimizando o custo. Assim, estabelece-se um compromisso, onde os pesos são ajustados para encontrar um ponto de equilíbrio entre as métricas, garantindo que ambas sejam atendidas simultaneamente. Após a realização de experimentos prévios, em ambos os casos foram adotados os valores de  $\alpha_1 = 0.7$  e  $\alpha_2 = 0.3$ . Além disso, são utilizados os fatores  $\gamma_{i1}$  e  $\gamma_{i2}$ , que representam o fator de probabilidade de redução da métrica para cada tipo de configuração, dado por  $\gamma = \frac{1}{\rho}$ , onde  $\rho$  é a probabilidade do dado estado, ou seja, para apenas um usuário,  $\gamma = 1$ , independente da configuração, com a chegada de  $K$  usuários, no SISO e 64-QAM,  $\gamma$  tende a aumentar drasticamente, pois a probabilidade de conseguir atingir valores menores diminui. Os valores de  $\gamma$  iniciais foram definidos utilizando um histórico de testes prévios, e foram atualizados com a captura de novos valores.

Após o cálculo do custo, os valores obtidos são comparados e a configuração que apresenta o menor custo é considerada ótima ( $C_{ótima}$ ). Contudo, visando evitar uma sobrecarga de trocas, foi implantado no SEQUOIA um sistema de estabilização, visando garantir que a troca seja realmente necessária. Para isso o  $C_{ótima}$  é armazenado em um histórico de recomendações recentes para estabilizar a escolha do modelo mais adequado. Vale ressaltar que o algoritmo visa sempre selecionar a configuração mínima para manter o QoS e, conseqüentemente, garantindo menores consumo de energia.

## 6. Protótipo de rede 5G SEQUOIA

Esta seção apresenta os componentes de *hardware* e *software* utilizados na montagem do protótipo de rede 5G *Standalone* (SA), desenvolvido para viabilizar a implementação e avaliação de desempenho do SEQUOIA. A Tabela 2 apresenta os componentes que compõem esse ambiente de testes. O núcleo 5G selecionado foi o Open5GS<sup>3</sup>, implantado em uma infraestrutura de virtualização baseada em contêineres *Docker*.

Docker Host	Intel NUC 11 - Core i7 - 16GB
Núcleo 5G	Open5Gs
gNodeB ( <i>software</i> )	srsRAN
gNodeB ( <i>hardware</i> )	Placa RDS Ettus B210
Número de usuários	3
Dispositivo móvel	Motorola Moto G53
RIC	OSC

**Tabela 2. Configuração do ambiente.**

A implantação da gNodeB adotou o split 8 [3GPP 2017], utilizando o srsRAN<sup>4</sup> para executar as funcionalidades das unidades central (CU) e distribuída (DU), com o processamento direto no *Docker Host*. Essa abordagem permitiu monitorar o uso de recursos computacionais sem os impactos da camada de virtualização. A unidade de rádio (RU) utilizou a placa de rádio definida por *software* (RDS) *Ettus* B210 para gerar o sinal de rádio, que, em conjunto com as antenas *Ettus Log Periodic*, estabelecem a comunicação

<sup>3</sup><https://open5gs.org/>

<sup>4</sup><https://www.srsran.com/>

5G com dispositivos móveis reais para a realização dos testes. Por fim, a plataforma utilizada para o RIC foi a fornecida pela OSC, com implantação baseada em Docker<sup>5</sup>.

## 7. Avaliação de desempenho de casos de uso

A avaliação visa analisar o desempenho da solução proposta para otimização de energia da infraestrutura da rede com uma arquitetura 5G Open RAN. A estrutura se concentra em aprimorar o QoS por meio da melhoria da eficiência de orquestração dos nós E2. Busca-se determinar se a solução melhora os indicadores de eficiência energética e mantém o desempenho da rede diante dos diferentes casos de uso. Esta seção descreve os cenários de avaliação, as métricas avaliadas e a discussão dos resultados das avaliações conduzidas em diferentes casos de uso. Os experimentos foram divididos em duas aplicações: uma de *Cloud Gaming* e a outra de transferência de vídeo com suporte da computação de borda.

### 7.1. Medição do Consumo de Energia

O monitor de recursos no SMO captura o consumo de CPU e memória da CU/DU de cada gNodeB. Para estimar a potência utilizada em Watts a partir dessas medições, empregamos a seguinte equação para calcular o consumo total de energia ( $P_{total}$ ):

$$P_{total} = P_{CPU} + P_{memória} \quad (2)$$

Onde:

$$P_{CPU} = \frac{\text{Uso da CPU (\%)}}{100} \times P_{TDP} \quad (3)$$

$$P_{memória} = \text{Memória (GB)} \times P_{memória \text{ por GB}} \quad (4)$$

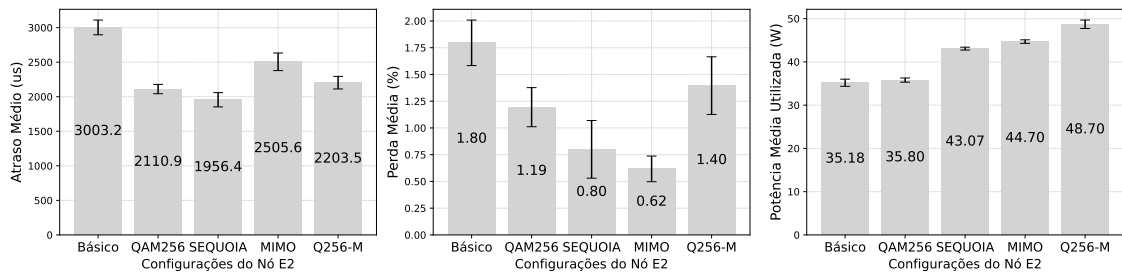
A potência máxima da CPU ( $P_{TDP}$ ) do Docker Host utilizado no protótipo da rede é de 28 W, enquanto o consumo de memória ( $P_{memória \text{ por GB}}$ ) é de 2,5 W por GB utilizado. Esse método permite estimar a potência consumida pelo sistema com base no uso de CPU e memória, fornecendo uma aproximação do consumo energético em Watts.

### 7.2. Caso de uso 1: *Cloud Gaming*

O primeiro caso de uso selecionado foi o *Cloud Gaming*, com o objetivo de analisar os impactos da solução proposta nas métricas de perda de pacotes e atraso na camada *Radio Link Control* (RLC). Para a realização dos experimentos foi estabelecida uma conexão entre o cliente e o servidor *Edge* (Notebook Dell i7) por meio do aplicativo *Moonlight*, um cliente de *streaming* baseado no protocolo *NVIDIA GameStream*. No lado do servidor, utilizou-se a plataforma *Sunshine*, que permite o acesso remoto do cliente. Além disso, o *Steam Remote Play* foi empregado para acessar o jogo *EFootball 2024*, executado com resolução de 1080p a 60 FPS. Para os testes, foram utilizados três clientes e três nós *edge*. Inicialmente, conectou-se apenas um par cliente-servidor e, gradualmente, os demais pares foram conectados a fim de observar os impactos nas métricas de QoS.

A Figura 4 ilustra que o SEQUOIA obteve o menor atraso e a segunda menor taxa de perda, em comparação com as configurações estáticas. Isso evidencia os impactos da

<sup>5</sup><https://github.com/srsran/oran-sc-ric>

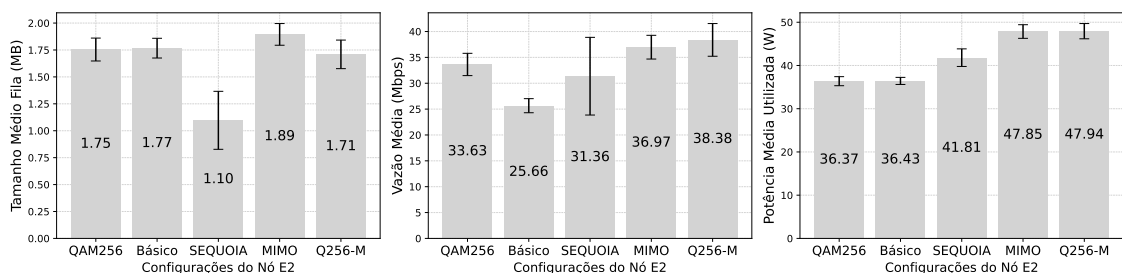


**Figura 4. Resultados obtidos para os experimentos do Cloud Gaming.**

otimização multiobjetivo na solução, pois, ao atribuir maior peso à ponderação da perda em relação ao atraso, prioriza-se a garantia do atraso em detrimento da perda. Em relação à perda média, o sistema MIMO 2x2 apresentou os melhores resultados. Para avaliar a similaridade entre os resultados do MIMO e do SEQUOIA, ambos foram submetidos ao teste estatístico ANOVA. O valor de  $F = 1,99$  representa a razão entre a variabilidade entre os grupos e a variabilidade dentro dos grupos. O p-value obtido foi de 0,158, superior ao limiar de 0,05, indicando a ausência de diferenças estatisticamente significativas entre os dois grupos. Assim, o teste confirma que os resultados do SEQUOIA e do MIMO são semelhantes para a métrica analisada. Por fim, em relação ao consumo médio de potência, o SEQUOIA apresentou uma redução de aproximadamente 11,5% em comparação com a configuração de maior consumo, 256-QAM+MIMO (Q256-M). Portanto, o SEQUOIA foi eficaz para esse caso de uso, pois conseguiu equilibrar o compromisso entre atraso e perda, garantindo métricas de QoS aceitáveis enquanto minimizava o consumo de energia.

### 7.3. Caso de uso 2: Transferência de vídeo

O segundo caso de uso analisa a vazão e o tamanho da fila do *buffer* usando um aplicativo de transferência de vídeo. Para esses experimentos, um aplicativo *Python* baseado em *Flask*<sup>6</sup>, que fornece serviços de vídeo e grava métricas de QoS, foi criado e implantado em um contêiner Ubuntu. Esse contêiner foi conectado apenas ao *User Plane Function* (UPF) por meio da interface N6, localizando-se na borda da rede. No lado do usuário, um aplicativo *Android* baseado em Java solicita um vídeo (*Big Buck Bunny 30fps/1080p*) do servidor, faz o *download* e assiste no telefone. O aplicativo calcula o tempo de carregamento e o rendimento da rede e, em seguida, envia essas métricas ao servidor. A Figura 5 apresenta os resultados obtidos para o cenário de transferência de vídeo.



**Figura 5. Resultados obtidos para os experimentos de Transferência de vídeo.**

<sup>6</sup><https://flask.palletsprojects.com/en/3.0.x/>

A primeira métrica analisada foi o tamanho médio da fila do *buffer* de *download*, um indicador do nível de congestionamento da rede. A Figura 5 mostra que a solução proposta apresentou o menor valor para essa métrica. No entanto, em relação à vazão, observa-se que a configuração 256-QAM+MIMO (Q256-M) obteve o maior valor, conforme esperado, uma vez que, de acordo com a Equação 1, essa configuração apresenta a maior vazão teórica. No caso do SEQUOIA, a configuração inicial adotada foi SISO+64-QAM, com chaveamento subsequente para outras opções a fim de selecionar a melhor alternativa, equilibrando o compromisso entre tamanho de fila e vazão. Como resultado, o SEQUOIA apresentou um aumento na vazão média e desempenho semelhante às configurações 256-QAM e MIMO, com valores de p-value de 0,43 e 0,6, respectivamente. Por fim, no que se refere ao consumo médio de potência, o SEQUOIA demonstrou uma redução de aproximadamente 12,7% em comparação com a configuração de maior consumo, 256-QAM+MIMO, evidenciando a eficácia da solução.

## 8. Conclusão

Este trabalho propôs o SEQUOIA, uma solução baseada em otimização multiobjetivo que considera o estado da rede, métricas de QoS para diferentes casos de uso (*cloud gaming* e transferência de vídeo) e redução do consumo de energia na rede de acesso 5G com o suporte da arquitetura Open RAN. A proposta ativa/desativa dinamicamente nós E2 e reconfigura parâmetros/componentes de rádio (modulação e número de antenas), ajustando a RAN com configurações que melhor atendam aos requisitos de aplicações URLLC e eMBB. Além disso, também foi proposto um ambiente experimental baseado em plataformas de código aberto, com os seguintes componentes: Open5GS para o núcleo da rede, srsRAN para a rede de acesso de rádio e O-RAN Software Community para o Near-RT RIC.

Os resultados indicam a eficiência da solução proposta SEQUOIA diante do uso de configurações estáticas tanto para as métricas de QoS quanto para a potência utilizada. Em ambos os casos de uso, *cloud gaming* e transferência de vídeo, o SEQUOIA reduziu o consumo em cerca de 11,5% e 12,7%, respectivamente, em comparação com a configuração QAM256+MIMO. Além de diminuir o atraso e o percentual de perdas para o caso do *cloud gaming* e o tamanho da fila do *buffer* para o caso de transferência de vídeo. Para trabalhos futuros, buscar-se-á a integração do SEQUOIA ao ambiente do *Open Source Management and Orchestration* (OSM MANO) do *European Telecommunications Standards Institute* (ETSI), além do aumento do número de usuário e utilização das gNodeB em ambientes Kubernetes.

## Referências

- 3GPP (2017). Radio Access Architecture and Interfaces (Release 14). Technical Report TR 38.801 V2.0.0, 3rd Generation Partnership Project.
- 3GPP (2022). User equipment (ue) radio access capabilities (release 17). Technical Report TS 38.306, 3rd Generation Partnership Project.
- Azariah, W., Bimo, F. A., Lin, C.-W., Cheng, R.-G., Nikaein, N., and Jana, R. (2024). A survey on open radio access networks: Challenges, research directions, and open source approaches. *Sensors*, 24(3).

- Bordin, M., Lacava, A., Polese, M., Satish, S., Nittoor, M. A., Sivaraj, R., Cuomo, F., and Melodia, T. (2024). Design and evaluation of deep reinforcement learning for energy saving in open ran. *arXiv preprint arXiv:2410.14021*.
- Chauhan, A., Bansal, K., and Chaturvedi, A. K. (2019). Weighted sum of spectral efficiency and energy efficiency in spatial modulation-mimo systems. In *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, pages 1–5.
- Constantine, J. and Akhtar, H. (2024). Open ran progress report - ericsson. *Major Milestones and What's Next in the Transformation Journey*, 2.
- Demir, T., Masoudi, M., Björnson, E., and Cavdar, C. (2024). Cell-free massive mimo in o-ran: Energy-aware joint orchestration of cloud, fronthaul, and radio resources. *IEEE Journal on Selected Areas in Communications*, 42(2):356–372.
- Gao, Y., Wang, L., Xie, Z., Qi, Z., and Zhou, J. (2023). Energy- and quality of experience-aware dynamic resource allocation for massively multiplayer online games in heterogeneous cloud computing systems. *IEEE Transactions on Services Computing*, 16(3):1793–1806.
- Hoffmann, M. e. (2024). Energy efficiency in open ran: Rf channel reconfiguration use case. *IEEE Access*, 12:118493–118501.
- Info, H. (2024). Understanding 32 qam, 64 qam, 128 qam, and 256 qam. Online resource. Accessed: 2024-12-27.
- Kolta, E., Hatt, T., and Moore, S. (2021). Going green: Benchmarking the energy efficiency of mobile. Technical report, GSMA Intelligence.
- Li, H., Emami, A., Assis, K. D. R., Vafeas, A., Yang, R., Nejabati, R., Yan, S., and Simeonidou, D. (2024). Drl-based energy-efficient baseband function deployments for service-oriented open ran. *IEEE Transactions on Green Communications and Networking*, 8(1):224–237.
- Liang, X., Al-Tahmeesschi, A., Wang, Q., Chetty, S., Sun, C., and Ahmadi, H. (2024). Enhancing energy efficiency in o-ran through intelligent xapps deployment. *arXiv preprint arXiv:2405.10116*.
- Mahmood, F. E., Perrins, E. S., and Liu, L. (2018). Energy consumption vs. bit rate analysis toward massive mimo systems. In *2018 IEEE International Smart Cities Conference (ISC2)*, pages 1–7.
- Pereira de Figueiredo, F. A. (2022). An overview of massive mimo for 5g and 6g. *IEEE Latin America Transactions*, 20(6):931–940.
- Perner, J. v. e. (2024). Beyond 5g white paper supplementary volume “sustainability and energy efficiency”. Project: Green Future Networks.
- Thangamayan, S., Walunjkar, M. D., Kumar Ray, D., Venkatesan, M., Banik, A., and Amrutkar, K. P. (2022). 5g modulation technique comparisons using simulation approach. In *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, pages 848–856.
- Urumkar, S., Ramamurthy, B., and Sharma, S. (2023). Improving energy efficiency in open ran through dynamic cpu scheduling. In *2023 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, pages 288–293.