

A Parameterized Model for Multimedia-Content Placement of the Streaming Service on 5G Cellular Networks

Um Modelo Parametrizado para Posicionamento de Conteúdo Multimídia do Serviço de Streaming em Redes Celulares 5G

Rodrigo Augusto Cardoso da Silva¹, Carlo Kleber da Silva Rodrigues^{1*}, Vladimir Rocha¹

Abstract: In the context of streaming services on 5G cellular networks of CRAN architecture with MEC servers, this article proposes a parametrized linear-programming model for efficient multimedia-content placement. The model resolution is obtained using one of these two alternatives: (i) transmission-time prioritization and (ii) data-traffic prioritization. To demonstrate the model's applicability, several streaming scenarios are evaluated, each involving different object-request patterns, server-storage capacities, and access concentrations (i.e., object's popularity). Three performance metrics are assessed in the experiments: average delay, network traffic, and hit rate. The overall results show that: (i) the first alternative should be considered when the object's popularity follows a non-uniform distribution; and (ii) the second alternative is recommended when the object's popularity tends to be evenly distributed. As its main contribution, this research thus offers relevant subsidies for streaming projects in 5G cellular networks.

Keywords: Multimedia Streaming — 5G Cellular Networks — Parametrized Model — CRAN Architecture

Resumo: No contexto de serviços de *streaming* em redes celulares 5G de arquitetura CRAN com servidores MEC, este artigo propõe um modelo parametrizado de programação linear para o posicionamento eficiente de conteúdos multimídia. A resolução do modelo é obtida por meio de uma das seguintes alternativas: (i) priorização do tempo de transmissão e (ii) priorização do tráfego de dados. Para demonstrar a aplicabilidade do modelo, vários cenários de *streaming* são avaliados, cada um deles envolvendo diferentes padrões de requisição de objetos, capacidades de armazenamento dos servidores e concentrações de acessos (i.e., popularidade dos objetos). Três métricas de desempenho são avaliadas nos experimentos: atraso médio, tráfego na rede e taxa de acertos. Os resultados globais mostram que: (i) a primeira alternativa deve ser empregada quando a popularidade dos objetos segue uma distribuição não uniforme; e (ii) a segunda alternativa é recomendada quando a popularidade dos objetos tende a ser uniformemente distribuída. Como principal contribuição, esta pesquisa oferece, portanto, subsídios relevantes para projetos de *streaming* em redes celulares 5G.

Palavras-Chave: Streaming Multimídia — Redes Celulares 5G — Modelo Parametrizado — Arquitetura CRAN

¹ Centro de Matemática, Computação e Cognição (CMCC), Universidade Federal do ABC (UFABC), Brasil

*Corresponding author: carlo.kleber@ufabc.edu.br

DOI: <http://dx.doi.org/10.22456/2175-2745.140750> • Received: 12/06/2024 • Accepted: 25/09/2024

CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

1. Introdução

O tráfego de dados móvel global tem crescido vertiginosamente e a tecnologia 5G se mostra progressivamente mais predominante. As estimativas apontam que o número de assinantes móveis já atinge 5,7 bilhões de pessoas (i.e., 71% da população global), e os dispositivos e conexões 5G correspondem a mais de 10% do total de dispositivos e conexões móveis existentes. Além disso, os dispositivos móveis somam mais de 13,1 bilhões, sendo 1,4 bilhão compatível com a tecnologia 5G [1, 2, 3, 4].

Uma das principais razões para o crescimento do tráfego

de dados móvel global é a popularização do serviço de *streaming* sob demanda de objetos multimídia, disponibilizado a partir de provedores de conteúdos na Internet [3, 5, 6]. Esse cenário traz grandes desafios para as arquiteturas das redes celulares em geral, tendo em vista que um significativo volume de dados precisa ser transmitido desde os provedores de conteúdos até os terminais móveis dos assinantes (do inglês, *User Equipments* - UEs) [2, 7].

Sob arquiteturas de redes celulares 5G, o caminho do tráfego de dados é constituído pelos seguintes dois subcaminhos em sequência: 1) desde os provedores de conteúdo na Internet

até a *Core Network* (CN); e 2) desde a CN até os UEs por meio da *Radio Access Network* (RAN). A RAN é formada pelas infraestruturas de *Backhaul* (BH) e *Fronthaul* (FH), além das denominadas estações bases (do inglês, *Base Stations* - BSs) [8, 2, 9].

Grande parte do volume de dados do serviço de *streaming* sob demanda corresponde a conteúdos repetidos. Isso ocorre devido à popularidade dos objetos multimídia: quanto mais popular é um objeto, mais requisições de usuários (i.e., assinantes) ocorrem para esse objeto. Se esse volume cresce demasiadamente, a probabilidade de congestionamento na RAN aumenta. Consequentemente, o retardo percebido pelos assinantes para visualizarem esses objetos sob *streaming* também aumenta, impactando diretamente os níveis de qualidade de serviço (do inglês, *Quality of Service* - QoS) e de qualidade de experiência (do inglês, *Quality of Experience* - QoE), respectivamente.

A conjuntura explicada acima suscita o seguinte problema de pesquisa: Como otimizar o volume de dados transmitido por meio da RAN devido a conteúdos multimídia repetidos? O paradigma *Multi-access Edge Computing* (MEC) é uma possível solução tecnológica para o supracitado problema [6, 10, 11, 12]. A ideia geral consiste em levar as capacidades de armazenamento e de processamento para mais próximo dos assinantes. Para o caso do serviço de *streaming* sob demanda, tem-se o detalhamento dado a seguir.

A Figura 1 mostra a organização hierárquica em camadas de uma rede 5G genérica [2, 8, 13]. No cenário de análise deste presente trabalho de pesquisa, servidores MEC são posicionados na infraestrutura de FH na camada RAN. Esses servidores são equipados com memória secundária (i.e., discos) para armazenar os objetos multimídia mais populares de provedores de conteúdo localizados na camada *Outer Layer*. Ademais, esses servidores têm interface de rede com fio para se conectarem à infraestrutura de FH na camada RAN.

Sob essa configuração, o serviço de *streaming* dos objetos multimídia é realizado pelos servidores MEC diretamente para os assinantes, localizados na camada *User Layer*, sem a necessidade de qualquer acesso aos provedores de conteúdos na Internet. Isso evita a transmissão de conteúdos repetidos por meio da camada CN e da infraestrutura de BH. Caso o objeto requisitado ainda não esteja disponível no servidor MEC, este servidor deve então primeiramente baixá-lo (do inglês, *download*) a partir dos provedores de conteúdos, armazená-lo localmente e, somente então, enviá-lo para o assinante via *streaming*. Ressalta-se que esse conteúdo é baixado uma única vez por meio da CN e da infraestrutura de BH, ficando disponível localmente no servidor MEC para atender a futuras requisições de assinantes, utilizando exclusivamente a infraestrutura de FH.

Ante o exposto e considerando um catálogo de objetos multimídia armazenados em provedores de conteúdos na Internet, este artigo de pesquisa propõe um modelo parametrizado de programação linear inteira multiobjetivo para abordar o problema de pesquisa acima identificado. Para tanto, esse

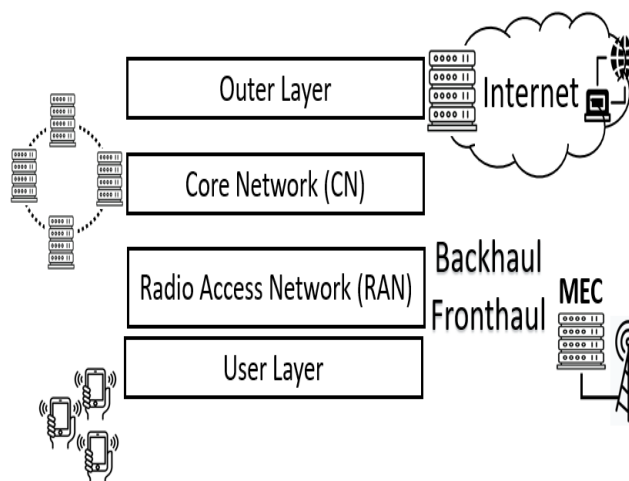


Figura 1. Organização em camadas da rede celular 5G.

modelo contempla a decisão do posicionamento eficiente dos objetos mais populares em um conjunto de servidores sob o paradigma MEC (i.e., servidores MEC). Como consequência, os dois objetivos a seguir são aqui definidos para serem alcançados.

- 1) Otimização do tempo médio de transmissão dos objetos multimídia na rede celular, o que nesta pesquisa se traduz na minimização do tempo médio para que os objetos multimídia armazenados nos servidores MEC sejam transmitidos para os assinantes;
- 2) Otimização do volume total de dados transmitido pela rede celular, o que nesta pesquisa se traduz na minimização do volume total de dados na infraestrutura de FH, que é originado devido à transmissão dos objetos multimídia pelos servidores MEC.

Para efeito de resolução do modelo, analisamos duas alternativas: i) Priorizar pelo primeiro objetivo e, em seguida, selecionar a solução que melhor atende ao segundo objetivo; e ii) Priorizar pelo segundo objetivo e, em seguida, selecionar a solução que melhor atende ao primeiro objetivo. Em complemento, um estudo de caso é aqui examinado para demonstrar a aplicabilidade do modelo. De forma geral, a principal contribuição deste artigo é prover subsídios teóricos para o desenvolvimento de projetos visando à implantação do serviço de *streaming* multimídia sob demanda em redes celulares 5G, com base no emprego do paradigma MEC.

O restante deste artigo é organizado como segue. Fundamentos sobre arquiteturas de redes celulares 5G constam na Seção 2. A Seção 3 discorre sobre trabalhos relacionados. Na Seção 4, tem-se a explicação do modelo parametrizado aqui proposto. A avaliação e a demonstração da aplicabilidade do modelo aparecem na Seção 5. Finalmente, na Seção 6 encontram-se as conclusões finais e as sugestões de trabalhos futuros.

2. Fundamentos

Para fins de organização de exposição, esta seção se encontra dividida em três subseções, como explicado a seguir. A primeira subseção fornece uma visão geral sobre as categorias das arquiteturas da RAN. A segunda subseção detalha a categoria Cloud RAN. Por último, a terceira discorre sobre vantagens e preocupações relacionadas à arquitetura CRAN.

2.1 Categorias de Arquiteturas

As arquiteturas da RAN de tecnologia 5G podem estar em uma das seguintes três categorias gerais [14, 15, 16]: Cloud RAN (CRAN), Heterogeneous CRAN (HCRAN) e Fog RAN (FRAN), as quais são brevemente explicadas a seguir.

A CRAN traz precipuamente a ideia de centralização tanto do processamento de dados quanto do gerenciamento da rede, usando conceitos de virtualização e computação na nuvem [2, 15, 17]. A HCRAN traz a densificação da rede (i.e., aumento de taxa de dados e cobertura do sinal), por meio do uso de células de diferentes tamanhos, combinada com a centralização tanto do processamento de dados quanto do gerenciamento da rede, que são também características da CRAN [2, 18, 19]. Finalmente, a FRAN vislumbra principalmente prover serviço de baixa latência com base na ideia de levar recursos computacionais, residentes na nuvem, para a borda da rede, criando então múltiplos núcleos de processamento e gerenciamento centralizados mais próximos dos usuários [14, 18, 20].

Neste artigo de pesquisa é considerada exclusivamente a categoria CRAN, combinada com o uso do paradigma MEC. A razão para tanto se deve à ampla disseminação dessa categoria na academia e indústria [14, 15, 16].

2.2 Arquitetura CRAN

A BS tradicional é constituída de dois componentes integrados: *Base-Band Unit* (BBU) e *Radio Head* (RH). Sob a CRAN, separam-se fisicamente esses dois componentes. Isso é feito para se conseguir: (i) processamento de alta capacidade a partir de um conjunto centralizado de BBUs, fazendo uso de virtualização e computação em nuvem; e (ii) serviço de rádio colaborativo a partir de RHs distintos remotamente localizados, nomeados como *Remote Radio Heads* (RRHs) [2, 15, 17].

A parte da rede que conecta o conjunto de BBUs (na nuvem) à CN constitui a infraestrutura de BH, enquanto que a parte da rede que conecta o conjunto de BBUs aos RRHs forma a infraestrutura de FH. Esse entendimento é ilustrado na Figura 2. Os recursos de processamento de sinal do conjunto de BBUs são alocados dinamicamente sob demanda, incluindo várias funções como, por exemplo, modulação, codificação, e seleção de frequência ou de canal [2, 15, 17].

Por sua vez, os RRHs, que são equipados com antenas, ficam encarregados de transmitir sinais de rádio do conjunto de BBUs para os usuários que se conectam por meio de UEs, bem como encaminhar os sinais em banda base desses usuários para o conjunto de BBUs para fins de processamento [2, 15, 17].

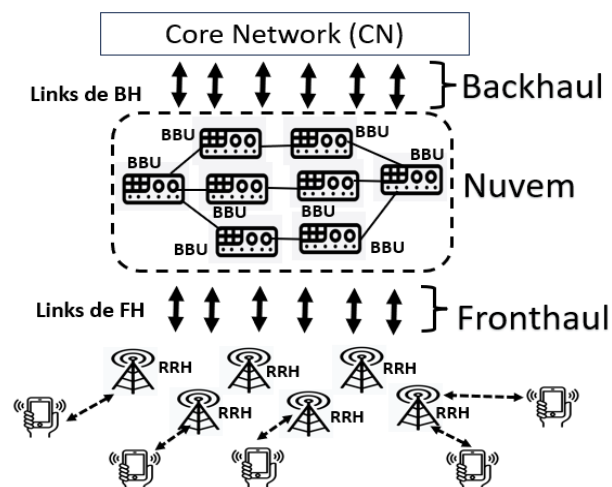


Figura 2. Arquitetura simplificada da CRAN.

Neste contexto, as principais funções dos RRHs incluem amplificação de sinais de radiofrequência (RF), conversão dos sinais de descida (i.e., em direção aos usuários) e subida (i.e., em direção à nuvem), filtragem, processamento digital, conversão analógico-digital (e digital-analógico), e adaptação de interface. Como a maioria das funções de processamento de sinal é realizada na nuvem, os RRHs podem ser equipamentos relativamente simples e, assim, serem espalhados sobre uma ampla área geográfica de forma efetiva e a baixo custo [2, 15, 17].

Sobre as infraestruturas de FH e BH, tem-se o seguinte entendimento. A infraestrutura de FH provê os *links* de comunicação entre BBUs e RRHs. Várias tecnologias como, por exemplo, comunicação por fibra óptica, comunicação sem fio padrão ou, até mesmo, comunicação por onda milimétrica, podem ser usadas para implementar esses *links*. A escolha da tecnologia leva em conta a prioridade do projeto. Por exemplo, no caso da comunicação por fibra óptica, tem-se uma alta capacidade de transmissão de dados, mas com um alto custo econômico e uma implantação inflexível. Já no caso da comunicação sem fio com frequências portadoras de 5 GHz a 40 GHz, tem-se um projeto mais econômico e mais flexível à custa de uma capacidade mais reduzida. Para a infraestrutura de BH, tem-se um projeto voltado para a adoção de *links* em fibra óptica na maioria dos casos [2, 15, 17].

2.3 CRAN: Vantagens e Preocupações

A Tabela 1 apresenta as principais vantagens relacionados à arquitetura CRAN. A partir dessa tabela, é possível observar a notória evolução tecnológica com respeito às arquiteturas legadas. Todavia, há algumas preocupações que devem estar em mente quando da implantação da arquitetura CRAN, conforme brevemente discorrido a seguir. A primeira se refere ao problema de segurança, pois ataques cibernéticos estão constantemente no horizonte. Nesse sentido, além das ameaças comuns de segurança das redes sem fio tradicionais, como ataque de emulação de usuário primário (do inglês, *Primary User Emulation*), há também o ataque de emulação de usuário secundário (do inglês, *Secondary User Emulation*), que visa a obter acesso não autorizado aos recursos da rede.

mary User Emulation Attack - PUEA), a arquitetura CRAN enfrenta outras ameaças de segurança, e.g., *eavesdropping*, *spoofing* e *flooding*. As pesquisas apresentadas nos trabalhos de [9, 21, 22] são indicadas como exemplos de fontes para maior detalhamento sobre essa preocupação, bem como sobre possíveis soluções já propostas na literatura.

A segunda preocupação advém do fato de as BBUs serem agrupadas na nuvem, estabelecendo um paradigma lógico de processamento centralizado. Em que pese as vantagens ressaltadas na Tabela 1, isso implica que a CRAN tem o evidente risco de falha de ponto único, ou seja, se a nuvem falhar, toda a rede fica fora de operação. A terceira preocupação se refere à eventual sobrecarga nos *links* de FH (entre RRHs e a nuvem), cujos requisitos podem ser 50 vezes maiores que aqueles dos *links* de BH. Por fim, a latência e o *jitter* para transmissão/recepção de dados entre a nuvem e os RRHs, devido à distância geográfica, também se constituem em uma preocupação quando da implementação de aplicações com estritos requisitos de tempo. As pesquisas apresentadas nos trabalhos de [21, 23, 24] se constituem em exemplos de fontes de consulta para aprofundamento dessas últimas preocupações, bem como conhecimento de algumas soluções da literatura.

Por fim, para encerrar esta subseção, convém reconhecer que o assunto redes celulares 5G envolve uma quantidade significativa de siglas e acrônimos. Assim sendo, para facilitar a leitura deste trabalho de pesquisa, tem-se então a Tabela 2, a qual fornece um breve glossário de siglas e acrônimos visando a consultas que eventualmente se fizerem necessárias.

3. Revisão da Literatura

Esta seção traz uma discussão sobre algumas obras da literatura que contribuem direta ou indiretamente para o objetivo deste trabalho, especialmente oportunizando uma percepção geral do estado da arte sobre o assunto redes celulares 5G.

Em [33], antevendo o futuro da implantação das redes celulares 5G e além, os autores descrevem no ano de 2015 as opções para a construção das infraestruturas de BH e FH, as quais podem suportar os serviços de banda larga móvel para as tecnologias sucessoras do 4G LTE Advanced. Para tanto, primeiro os autores revisam a arquitetura da RAN vigente na época, incluindo o acesso sem fio para alcançar os UEs. Em seguida, são descritas as tecnologias para as futuras infraestruturas de BH e FH, as quais são capazes de aumentar significativamente a largura de banda sem fio e permitem a implantação massiva de células de menores coberturas, promovendo a densificação da rede. O trabalho desses autores pode ser considerado como seminal e, sobretudo, contribui com valiosos subsídios teóricos para melhor compreender e analisar as arquiteturas que estão disponíveis na conjuntura corrente (i.e., CRAN, HCRAN e FRAN) e aquelas que eventualmente estejam por vir ou simples variantes.

Os trabalhos em [14, 15, 34] se constituem em *surveys* que discutem arquiteturas para implantação das redes celulares 5G, com destaque para CRAN, HCRAN e FRAN. Esses trabalhos possuem conteúdos que contemplam, sob uma vi-

são cronológica evolutiva, aspectos teóricos e práticos do assunto tecnologia 5G, incluindo um detalhamento especial sobre as vantagens, os desafios e as soluções já apresentadas na literatura corrente, bem como sobre a técnica de *caching*. Esses trabalhos podem ser vistos como fontes de consulta para projetos sobre arquiteturas para RANs de tecnologia 5G.

Em [23], os autores discutem as funcionalidades pertinentes ao conjunto centralizado de BBUs da arquitetura CRAN. São propostos esquemas flexíveis de divisão funcional baseada em RRHs virtuais (V-RRHs), visando otimizar a utilização de recursos e, ao mesmo tempo, reduzir o custo geral do conjunto de BBUs. Mais especificamente, três tipos de esquemas são propostos, os quais diferem quanto ao número de funções de processamento, nativamente pertencentes ao conjunto de BBUs, que podem ser executadas por V-RRHs. O objetivo é encontrar a alternativa de divisão funcional ideal entre os V-RRHs e o conjunto centralizado de BBUs. Os experimentos realizados mostram que os esquemas de divisão funcional propostos superam em grande medida os esquemas mais tradicionais em termos de CAPEX/OPEX.

Em [20], os autores evidenciam a possível alta latência de um RRH para o conjunto de BBUs da arquitetura CRAN. É destacado que essa latência pode ser até ser inaceitável para algumas aplicações com estritos requisitos de tempo de resposta. Nesse artigo, os autores então tratam o problema de alocação (divisão) ideal de recursos do tradicional conjunto de BBUs entre o conjunto de RRHs e os subconjuntos de BBUs da arquitetura FRAN, de modo que apenas um número mínimo de subconjuntos de BBUs esteja ligado (energizado) para atender todos os RRHs. Isso leva à redução do consumo de energia elétrica. Além disso, também discute-se o balanceamento de carga entre os subconjuntos de BBUs da FRAN visando à estabilidade e à tolerância a falhas. Os resultados de simulação mostram que os esquemas propostos são efetivos.

Em [35, 36], os respectivos autores abordam as características operacionais do projeto de implantação de redes celulares 5G, desde a camada física até a camada de aplicação, bem como discutem detalhadamente resultados reais alcançados em experimentos. Mais especificamente, em [36] são realizadas medições em uma cidade de alta densidade demográfica, considerando uma área de 0,5 km x 0,92 km, rodeada por edifícios, árvores, com diversas atividades urbanas (e.g., cinema, escolas, escritórios, etc.) e atendida por seis BSs. Em [35], tem-se um trabalho mais amplo, pois as medições ocorrem em três cidades distintas nos EUA, a saber, Mineápolis, Chicago e Atlanta, além de considerar três operadoras comerciais de redes celulares, a saber, VZ, T-Mobile e Sprint. As contribuições desses dois trabalhos se dão pelos aportes de cunho prático e realista, viabilizados a partir das inúmeras medições executadas. Esses aportes se constituem em valiosos guias para o desenvolvimento de projetos de implantação da tecnologia 5G. Ademais, os desafios identificados nesses trabalhos podem ser enfrentados para redução dos CAPEX e OPEX, respectivamente, relativos à implantação de redes 5G.

Mais diretamente relacionado ao problema de pesquisa do

Tabela 1. Vantagens da arquitetura CRAN.

Vantagem	Entendimento
Reduzidos CAPEX/OPEX	Baixo custo, menos espaço e menor tempo para instalar e operar RRHs. Há também uma maior facilidade para compartilhamento de equipamentos, levando a reduzidos investimentos de capital (do inglês, <i>Capital Expenditure</i> - CAPEX). Os recursos computacionais são agregados na nuvem, deixando as funções mais simples para os RRHs, o que reduz a redundância em despesas operacionais (do inglês, <i>Operational Expenditure</i> - OPEX) [25, 26, 27].
Uso eficiente de energia elétrica	O número de unidades de processamento (i.e., BBUs) é menor que em uma RAN convencional (i.e., BSs), levando a uma redução de consumo de energia elétrica. Ademais, o sistema de arrefecimento das unidades de rádio (i.e., RRHs) é de potência reduzida, pois há uma arrefecimento natural em virtude da elevação das torres de sustentação ou dos prédios em que estão instaladas [25, 28, 29].
Uso eficiente de espectro de frequências	A implementação de estratégias de transmissão/recepção coordenadas e cooperativas como, por exemplo, a coordenação aprimorada de interferência entre células (do inglês, <i>enhanced inter-cell interference coordination</i> - eICIC) e a transmissão multiponto coordenada (do inglês, <i>coordinated multipoint transmission</i> - CoMP) entre os RRHs podem ser muito mais simples e eficazes, redundando em uso eficiente de frequências [25, 30, 31].
Latência reduzida para execução de operações	Várias operações do sistema podem ser executadas dentro da nuvem, sem depender de entidades externas. Por exemplo, o tempo necessário para realizar <i>handovers</i> é reduzido, pois é realizado dentro da nuvem em vez de entre BSs tradicionais. Além disso, o volume de informações das operações, que antes era enviado para a CN, é também reduzido [14, 25].
Uso eficiente de BBUs	Não há necessidade de que as BSs permaneçam sempre ativas (i.e., energizadas). Como as ações de processamento são realizadas na nuvem, o consumo de energia elétrica e o balanceamento de carga podem ser otimizados pela alocação dinâmica de recursos de processamento e migração de tarefas entre BBUs. Dessa maneira, BBUs podem ser colocadas em modo econômico de operação (do inglês, <i>sleep mode</i>) ou mesmo desligadas para economizar energia [26].
Gerenciamento eficiente de interferência	Existe facilidade de compartilhamento de informação do estado do canal (do inglês, <i>channel state information</i> - CSI), tráfego de dados e informação de controle de serviços móveis entre as unidades de processamento (i.e., BBUs). Isso permite a cooperação multiponto mais efetiva e habilita ter a multiplexação de mais fluxos em um mesmo canal com substancial redução de interferência mútua. Devido à redução da interferência, a qualidade do <i>link</i> e a vazão são significativamente otimizadas [15, 26].
Manutenção e expansão facilitadas	A centralização na nuvem promove a escalabilidade sistêmica, o que simplifica os processos de atualização/expansão e de manutenção da rede. Por exemplo, uma operadora de rede precisa apenas adicionar novos RRHs conectados à nuvem para fins de atender a mais áreas de serviço. Também há facilidade para a instalação sob demanda de recursos virtuais na nuvem. Além disso, a arquitetura fornece uma maneira fácil de introduzir novos padrões, pois o <i>hardware</i> eventualmente necessário precisa ser colocado apenas na nuvem. Além disso, essa centralização permite atualizações na capacidade de processamento de CPUs de maneira mais fácil e mais frequente do que em sistemas tradicionais. Por fim, o suporte centralizado nativo para operação multipadrão torna a operação inter-RAN relativamente simplificada [15, 30, 32].
Adaptação ao tráfego não uniforme	As redes celulares modernas exibem uma quantidade substancial de diversidade tempo-espacial no tráfego. No entanto, as BSs tradicionais são dimensionadas para horários de pico, implicando um desperdício de poder de processamento fora dos horários de pico. Como o processamento em banda base na CRAN é realizado na nuvem (i.e., no conjunto de BBUs), então os recursos de processamento podem ser alocados de maneira otimizada considerando a demanda instantânea de tráfego, melhorando a taxa geral de utilização de recursos [15].

presente trabalho, estão as obras que tratam sobre soluções para reduzir a latência experimentada pelas aplicações computacionais que acessam a camada mais externa da arquitetura de rede 5G, a saber, a camada *Outer Layer* (vide Figura 1) ou, equivalentemente, as soluções para reduzir o tráfego de

dados escoado pela rede celular por meio das camadas CN e RAN, respectivamente (vide Figura 1), incluindo a otimização do consumo de energia elétrica. Neste contexto, é ponto pacífico que existe uma quantidade significativa de trabalhos dentre desse escopo. Daí, para efeito de maior objetividade de

Tabela 2. Acrônimos e siglas.

Acrônimo/Sigla	Significado
UE	<i>User Equipment</i>
CN	<i>Core Network</i>
RAN	<i>Radio Access Network</i>
BH	<i>Backhaul</i>
FH	<i>Fronthaul</i>
BS	<i>Base station</i>
QoS	<i>Quality of Service</i>
QoE	<i>Quality of Experience</i>
MEC	<i>Multiaccess Edge Computing</i>
CRAN	<i>Cloud RAN</i>
HCRAN	<i>Heterogeneous CRAN</i>
FRAN	<i>Fog RAN</i>
BBU	<i>Base-Band Unit</i>
RH	<i>Radio Head</i>
RRH	<i>Remote Radio Head</i>
OPEX	<i>Operational Expenditure</i>
CAPEX	<i>Capital Expenditure</i>

apresentação e facilidade de entendimento, a seguir são então discutidas conjuntamente apenas alguns dos mais recentes obras.

Em [37, 38, 39, 40], os autores abordam soluções e/ou realizam *surveys* com foco na utilização da comunicação *device-to-device* (D2D), em que admite-se o armazenamento de conteúdos nos próprios UEs e pode-se ainda objetivar a otimização do consumo de energia elétrica. Por sua vez, em [2, 41, 42, 43, 44], os respectivos autores propõem um problema de otimização envolvendo a combinação de diferentes tamanhos de células, com base no conceito de HCRAN, conjuntamente com a estratégia de armazenamento de conteúdos nas células que possuem mais usuários associados. A diferenciação entre esses trabalhos reside na formulação e/ou método de solução matemática utilizada, bem como em outras métricas de desempenho associadas como, por exemplo, a energia elétrica consumida. Os resultados numéricos obtidos a partir de inúmeros experimentos se mostram bem satisfatórios, embora não haja avaliação contundente da complexidade de uma eventual implementação real.

Ante o exposto, este presente trabalho complementa a literatura disponível, pois propõe uma solução que otimiza dois objetivos conjuntamente: (i) a redução do tráfego de dados na infraestrutura de comunicação móvel e (ii) a redução do tempo de transmissão de objetos multimídia para os usuários. Como já mencionado ao início deste trabalho, isso é conseguido por meio do posicionamento estratégico de conteúdos populares em servidores MEC. Mais especificamente em comparação a trabalhos de semelhante propósito (e.g., [2, 41, 42, 43, 44]), este presente trabalho se diferencia por derivar um modelo matemático de programação linear inteira multiobjetivo, cujas avaliação e aplicabilidade são feitas por meio da realização de um estudo de caso. Por fim, a Tabela 3 traz um síntese dos

trabalhos percorridos nesta seção.

4. Modelo Parametrizado

Esta seção explica o modelo de programação linear. Para fins de organização e facilidade de entendimento, tem-se a subdivisão a seguir. A Subseção 4.1 traz a modelagem e a notação adotadas na formulação do problema, e a Subseção 4.2 trata sobre as equações que descrevem o modelo.

4.1 Modelagem

Considere os seguintes entendimentos preliminares. Localidade é uma área geográfica com usuários móveis que é atendida por um único e exclusivo RRH. Todo servidor MEC está associado a uma localidade. Embora cada servidor se conecte fisicamente a um único e exclusivo RRH, os UEs podem acessar um objeto multimídia em qualquer servidor MEC, mesmo que esse servidor não esteja na mesma localidade em que o assinante móvel se encontra.

As características do problema são modeladas de forma individual (e.g., localidades, objetos multimídia e servidores MEC). Ademais, a infraestrutura de FH que interconecta os RRHs é representada pelos enlaces individuais de comunicação. O conjunto de todos esses elementos constitui a entrada do problema, que é detalhada na notação apresentada a seguir.

- $\mathcal{L} = \{1, \dots, L\}$: conjunto de localidades;
- $\mathcal{M} = \{1, \dots, M\}$: conjunto de objetos multimídia;
- S_m , onde $m \in \mathcal{M}$: tamanho do objeto multimídia m , medido em bytes;
- $\mathcal{N} = \{1, \dots, N\}$: conjunto de servidores MEC;
- L_n , onde $n \in \mathcal{N}$, $L_n \in \mathcal{L}$: localidade do servidor n ;
- C_n , onde $n \in \mathcal{N}$: capacidade de disco do servidor n , medido em bytes;
- P_{ml} , onde $m \in \mathcal{M}$, $l \in \mathcal{L}$: popularidade do objeto multimídia m na localidade l . Conceitualmente, P_{ml} é definida como o número de requisições na localidade l direcionadas ao conteúdo m no momento corrente;
- A_{ml} , onde $m \in \mathcal{M}$, $l \in \mathcal{L}$: variável binária que indica se objeto multimídia m é acessado por algum usuário na localidade l . Conceitualmente, $A_{ml} = 1$ se $P_{ml} > 0$, e $A_{ml} = 0$ em caso contrário;
- D_{mlk} , onde $m \in \mathcal{M}$, $l \in \mathcal{L}$, $k \in \mathcal{L}$: tempo para transmitir o objeto multimídia m para um usuário na localidade l a partir de um servidor na localidade k ;
- $\mathcal{E} = \{1, \dots, E\}$: conjunto dos enlaces da infraestrutura de FH;
- $\mathcal{L}_e = \{(l_{e1}, k_{e1}), (l_{e2}, k_{e2}), \dots\}$, onde $e \in \mathcal{E}$: conjunto de todos caminhos (l, k) de uma origem $l \in \mathcal{L}$ até um destino $k \in \mathcal{L}$ que passam pelo enlace e ;

Tabela 3. Síntese dos trabalhos relacionados.

Referência	Propósito
[33]	Descrever as opções para a construção das infraestruturas de BH e FH que podem suportar os serviços de banda larga móvel para as tecnologias sucessoras do 4G LTE Advanced. É um trabalho seminal e, sobretudo, contribui com valiosos subsídios teóricos para melhor entender as arquiteturas que estão disponíveis na conjuntura corrente.
[14, 15, 16, 34]	Apresentar <i>surveys</i> que discutem alternativas de arquiteturas para implantação das redes celulares 5G, e.g., CRAN, HCRAN, e FRAN. São quatro trabalhos que se constituem em importantes fontes de consulta para projetos de pesquisa.
[23]	Propor esquemas flexíveis de divisão funcional baseada em RRHs virtuais (V-RRHs), visando otimizar a utilização de recursos e, ao mesmo tempo, reduzir o custo geral (CAPEX/OPEX) do conjunto de BBUs.
[20]	Tratar o problema de alocação (divisão) ideal de recursos do tradicional conjunto centralizado de BBUs entre o conjunto de RRHs e os subconjuntos de BBUs da arquitetura FRAN, de modo que apenas um número mínimo de subconjuntos de BBUs esteja ligado (energizado) para atender todos os RRHs, reduzindo o consumo de energia elétrica.
[35, 36]	Abordar de forma prática as características operacionais do projeto de implantação de redes celulares 5G, desde a camada física até a camada de aplicação, bem como discutir detalhadamente os resultados reais alcançados. Ademais, a identificação feita nesses trabalhos dos desafios ainda presentes, considerando aspectos tecnológicos e econômicos, pode ser utilizada para redução do CAPEX/OPEX referente à implantação de redes 5G.
[37, 38, 39, 40]	Propor soluções (ou realizar <i>surveys</i>) que se baseiam na comunicação <i>device-to-device</i> (D2D), considerando o armazenamento de conteúdos em UEs e/ou visando à otimização do consumo sistêmico de energia elétrica.
[2, 6, 41, 42, 43, 44]	Propor problema de otimização envolvendo a combinação de diferentes tamanhos de células, com base no conceito de HCRAN, conjuntamente com a estratégia de armazenamento de conteúdo nas células que possuem mais usuários associados. A diferenciação entre esses trabalhos reside na formulação e/ou método de solução matemática utilizada.
Este trabalho de pesquisa	Semelhantemente aos trabalhos de [2, 6, 41, 42, 43, 44], também propor um problema de otimização sistêmica, diferindo por utilizar o conceito de MEC e derivar um modelo matemático de programação linear inteira multiobjetivo, cujas avaliação e aplicabilidade são feitas por meio da realização de um estudo de caso, asseverando a exequibilidade da abordagem considerada.

- B_e , com $e \in \mathcal{E}$: capacidade do enlace e , medida em bytes.

A resolução de otimização é feita sobre um intervalo de tempo Δt . Mais especificamente, a popularidade P_{ml} de um objeto multimídia representa o número de requisições feitas a ele dentro do intervalo de tempo Δt em uma determinada localidade. Isso porque o tempo durante o qual um objeto multimídia permanece popular tende a ser limitado. Por consequência, a resolução de otimização deve ser periodicamente executada a fim de permitir o melhor posicionamento dos objetos multimídia no transcorrer de todo o tempo. Ademais, a rede que interconecta todos RRHs é modelada por meio de seus enlaces $e \in \mathcal{E}$. Cada enlace tem definida uma capacidade B_e , que é a quantidade máxima de bytes que ele poderá transmitir em cada uma das direções dentro do intervalo de tempo Δt . Ainda, note que um par de RRHs em localidades distintas pode naturalmente ser interconectado por um ou mais enlaces.

Admita a topologia da infraestrutura de FH exemplificada na Figura 3. Há dez RRHs que se conectam a um anel constituído por cinco roteadores. Esses roteadores são responsáveis

pelo encaminhamento dos seguintes tráfegos: (1) da nuvem em direção aos assinantes; (2) dos assinantes em direção à nuvem; e (3) entre servidores MEC. Os enlaces entre RRHs e roteadores e entre roteadores são modelados no conjunto $\mathcal{E}$, juntamente com suas capacidades B_e .

Os tráfegos (1) e (2) ocorrem sempre que um objeto multimídia requisitado não está presente em nenhum dos servidores MEC. Esses tráfegos passam pelos enlaces da CN, BH e FH antes de chegar aos assinantes (vide Figura 2). Por sua vez, o tráfego (3) ocorre quando o objeto multimídia requisitado já está disponível em algum dos servidores MEC. Relembramos que o objetivo 2 da presente pesquisa é otimizar o volume total de dados na infraestrutura de FH, considerando a transmissão dos objetos multimídia armazenados nos servidores MEC (i.e., tráfego (3)), e não aquele entre provedores de conteúdos e os UEs (i.e., tráfegos (1) e (2)). A otimização conjunta dos tráfegos (1), (2) e (3) é deixada como trabalho futuro.

Ainda nesse exemplo da Figura 3, suponha que se queira calcular o valor de D_{116} , i.e., tempo de transmissão do objeto 1 desde a localidade 1 até a localidade 6. Para tanto, assumimos

S_1 previamente conhecido e calculamos o tempo de transmissão entre RRH 1 e RRH 6, considerando todos enlaces no caminho então existente, i.e., enlaces entre RRH1 e r1, entre r1 e r2, entre r2 e r3, e entre r3 e RRH6, respectivamente.

Por fim, além dos valores de entrada definidos até aqui, há ainda as variáveis de decisão. Todas variáveis são inteiras e binárias, caracterizando assim o modelo proposto como um modelo de programação linear inteira. A descrição das variáveis está apresentada a seguir.

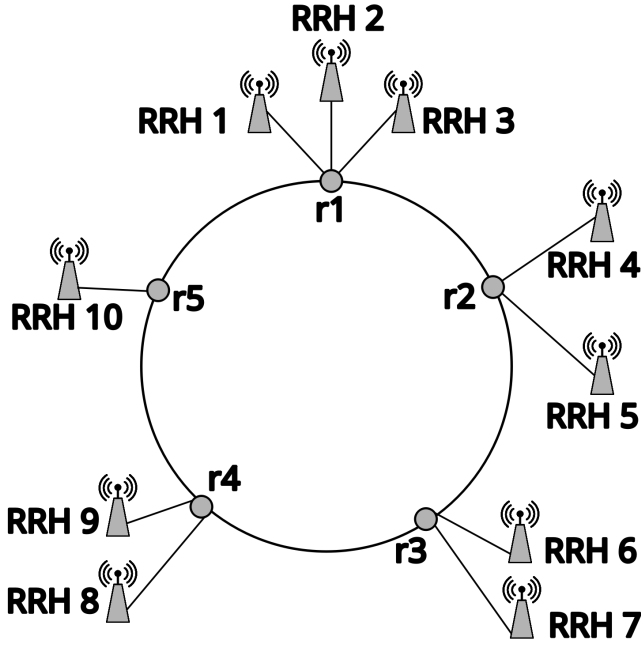


Figura 3. Exemplo de rede de FH.

- x_{mlk} , onde $m \in \mathcal{M}, l \in \mathcal{L}, k \in \mathcal{L}$: variável binária. $x_{mlk} = 1$ se as requisições originadas na localidade l para o objeto multimídia m acessam o conteúdo por meio de um servidor na localidade k , e $x_{mlk} = 0$ em caso contrário;
- y_{nm} , onde $n \in \mathcal{N}, m \in \mathcal{M}$: variável binária. $y_{nm} = 1$ se o servidor n é escolhido para hospedar o objeto multimídia m , e $y_{nm} = 0$ em caso contrário.

4.2 Equações

Relembramos que a decisão do posicionamento eficiente dos objetos mais populares em um conjunto de servidores MEC tem, como consequência, a necessidade do atendimento dos dois objetivos a seguir definidos nesta pesquisa.

- 1) Otimização do tempo médio de transmissão dos objetos multimídia na rede celular, o que nesta pesquisa se traduz na minimização do tempo médio para que os objetos multimídia armazenados nos servidores MEC sejam transmitidos para os assinantes;

- 2) Otimização do volume total de dados transmitido pela rede celular, o que nesta pesquisa se traduz na minimização do volume total de dados na infraestrutura de FH, que é originado devido à transmissão dos objetos multimídia pelos servidores MEC.

O primeiro objetivo é garantido com base na variável P_{ml} , que define a popularidade do objeto multimídia m na localidade l . Essa variável é usada como um peso na equação da função objetivo definida na sequência. Dessa forma, se um objeto tem uma popularidade muito alta, a otimização do seu tempo de transmissão tem prioridade sobre outros objetos, fazendo com que ele seja colocado em um servidor mais próximo da localidade de sua requisição.

Para esse primeiro objetivo, tem-se a Equação (1), a qual descreve a primeira função objetivo. O termo entre parênteses indica o tempo de transmissão de um objeto multimídia. Somente um termo da somatória tem $x_{mlk} = 1$, sendo este o termo correspondente ao servidor usado para acessar o objeto. As somatórias dessa equação garantem que todos os objetos multimídia e localidades são considerados.

$$\text{minimizar} \sum_{m \in \mathcal{M}} \sum_{l \in \mathcal{L}} P_{ml} \left(\sum_{k \in \mathcal{L}} x_{mlk} D_{mlk} \right) \quad (1)$$

Para o segundo objetivo, tem-se a Equação (2), a qual descreve a segunda função objetivo. Essa equação calcula a quantidade de bytes transmitida através de todos os enlaces de comunicação.

$$\text{minimizar} \sum_{e \in \mathcal{E}} \sum_{(l,k) \in \mathcal{L}_e} \sum_{m \in \mathcal{M}} x_{mlk} S_m \quad (2)$$

Finalmente, tem-se as restrições do problema nas Equações (3)–(6), como detalhado a seguir. A Equação (3) garante que todo objeto multimídia requisitado por qualquer usuário pode ser acessado. A Equação (4) garante que a capacidade de armazenamento dos servidores é respeitada. A Equação (5) garante que servidores em uma certa localidade só são acessados se possuírem o objeto multimídia desejado. Por último, a Equação (6) garante que nenhum enlace é sobrecarregado.

$$\sum_{k \in \mathcal{L}} x_{mlk} = A_{ml}, \forall m \in \mathcal{M}, \forall l \in \mathcal{L} \quad (3)$$

$$\sum_{m \in \mathcal{M}} y_{nm} \cdot S_m \leq C_n, \forall n \in \mathcal{N} \quad (4)$$

$$x_{mlk} \leq \sum_{n \in \mathcal{N} | L_n = k} y_{nm}, \forall m \in \mathcal{M}, \forall l \in \mathcal{L}, \forall k \in \mathcal{L} \quad (5)$$

$$\sum_{(l,k) \in \mathcal{L}_e} \sum_{m \in \mathcal{M}} x_{mlk} S_m \leq B_e, \forall e \in \mathcal{E} \quad (6)$$

5. Avaliação e Aplicabilidade

Esta seção avalia e demonstra a aplicabilidade do modelo proposto, bem como sua adaptabilidade a diferentes cenários de análise de *streaming* sob demanda. Isso é feito por meio da sua aplicação em um estudo de caso considerando diferentes padrões de requisição de objetos, capacidades de armazenamento dos servidores MEC e concentrações de acessos (i.e., popularidade de objetos). Para tanto, inicialmente são apresentadas respectivamente as caracterizações da rede e dos objetos multimídia e, na sequência, tem-se os experimentos e a análise dos resultados obtidos.

5.1 Caracterização da rede

A topologia da rede do estudo de caso é detalhada a seguir sob o entendimento introdutório dado para a rede da Figura 3. Ressalta-se que a mesma é inspirada nos trabalhos de [45, 46], cujo dimensionamento geográfico é parcialmente baseado no bairro Morumbi da cidade de São Paulo, SP, no Brasil [47].

A topologia é constituída por um anel de sete nós conectados entre si por enlaces bidirecionais de capacidade individual de transmissão 10 Gbps (em cada direção). Cada nó da rede é um roteador de tráfego que se conecta a três RRHs via enlaces bidirecionais (um para cada RRH) de capacidade individual de transmissão 1 Gbps (em cada direção). Cada RRH tem associado um servidor MEC com uma certa capacidade de armazenamento em disco.

A Tabela 4 recapitula os valores numéricos acima indicados, bem como informa outros parâmetros complementares da topologia da rede, provendo sua completa caracterização operacional. Por fim, a visualização simplificada dessa topologia é mostrada na Figura 4. Por simplicidade, os enlaces entre RRHs e roteadores/usuários/MECs não aparecem em sua totalidade, sendo apenas indicados para um RRH do anel.

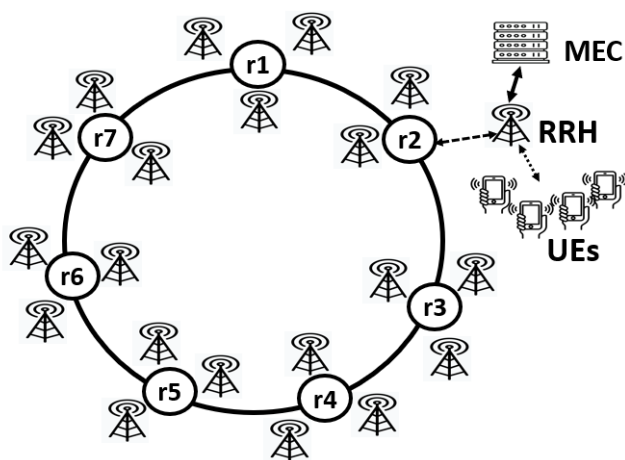


Figura 4. Topologia simplificada da rede.

5.2 Caracterização dos objetos multimídia

A Tabela 5 traz a caracterização dos objetos multimídia. Os valores explicitados se baseiam em padrões adotados nas plataformas da empresa americana de serviço *online* de *streaming* Netflix [48, 49]. Note que essa caracterização independe da topologia da rede considerada, pois os provedores de conteúdo estão localizados na camada *Outer Layer*.

5.3 Experimentos

5.3.1 Configuração dos Experimentos

No cenário analisado nesta pesquisa, os servidores MEC são usados para armazenar apenas uma parte dos objetos multimídia do catálogo completo. Isso ocorre devido à capacidade de armazenamento mais limitada dos servidores MEC em comparação com aquela dos provedores de conteúdos.

Considerando a limitação de armazenamento supracitada, nos experimentos é definida a *proporção* de objetos multimídia que devem ser armazenados na infraestrutura MEC, cujos valores são: 1%, 2% e 3%. Daí, os objetos multimídia são selecionados em ordem decrescente de popularidade. Por exemplo, para o primeiro valor estimado de *proporção*, deve-se garantir que 1% dos objetos multimídia mais requisitados sejam armazenados nos servidores MEC. Valores maiores que 3% inviabilizam a resolução do problema, pois a soma das capacidades de todos os discos dos servidores MEC é insuficiente para armazenar todos esses objetos multimídia. Sendo assim, apenas valores até 3% são considerados.

Para a avaliação dos experimentos, é necessário definir o número de requisições feitas em cada localidade correspondente a um RRH. Esse número é modelado a partir de uma distribuição normal, com valores entre 100 e 1000, em incrementos de 100. Para determinar qual objeto multimídia é solicitado em cada requisição, é utilizada uma distribuição de Zipf. Essa abordagem permite avaliar o impacto da popularidade dos objetos no desempenho do sistema [50, 51, 52, 53]. Isso porque, quanto maior é o valor do parâmetro da Zipf, mais requisições são direcionadas aos mesmos objetos multimídia, estabelecendo uma concentração de acessos em um subconjunto de objetos. Esse parâmetro é denominado de *skew factor*. São utilizadas duas distribuições Zipf de mesmo valor de *skew factor*: uma para a categoria *shows/séries* e outra para categoria *filmes*. A categoria a que cada uma das requisições se refere é determinada aleatoriamente. Os valores do *skew factor* considerados nos experimentos são indicados mais adiante.

5.3.2 Métricas de Desempenho

São consideradas três métricas de desempenho: *atraso médio*, *tráfego na rede* e *taxa de acertos*. O *atraso médio* é a razão entre o valor obtido para a primeira função objetivo e o número total de requisições. O *tráfego na rede* é o valor obtido pela segunda função objetivo, que contabiliza o número total de bits transferidos nos enlaces cabeados da infraestrutura de FH. Por fim, a *taxa de acertos* é a razão entre o número de requisições atendidas por servidores MEC conectados ao RRH mais próximo do usuário e o número total de requisições.

Tabela 4. Caracterização operacional da topologia de rede [45, 46, 47].

Definição	Valor
Área de abrangência geográfica da topologia	4 km ²
Alcance (raio) de cobertura rádio de cada RHH	500 m
Capacidade de armazenamento de cada servidor MEC	10 GB
Distância física entre RHH e servidor MEC associado	20 m
Distância física entre nós (roteadores) adjacentes no anel	200 m
Distância física entre pares de RHHs associados a um mesmo servidor MEC	30 m
Capacidade de transmissão do enlace (em cada direção) entre roteadores adjacentes no anel	10 Gbps
Capacidade de transmissão do enlace (em cada direção) entre RHH e roteador associado	1 Gbps
Capacidade de transmissão do enlace (em cada direção) entre RHH e <i>User Equipment</i>	0,5 Gbps

Tabela 5. Caracterização dos objetos multimídia baseada na empresa Netflix [48, 49].

Definição	Valor
Número total de títulos da plataforma	8400
Capacidade individual dos provedores de conteúdo (Netflix)	100 TB
Tipos de objetos	Shows/Séries; Filmes.
Tamanho de Show/Série	250 MB
Tamanho de filme	1 GB
Número total de títulos do tipo Show/Série ($\approx 82\%$)	6888
Número total de títulos do tipo Filme ($\approx 18\%$)	1512

5.3.3 Metodologia de Resolução

As equações do modelo proposto na Seção 4 são codificadas no resolutor Gurobi Optimizer [54], derivando um modelo matemático. Esse resolutor é um programa computacional que implementa diversos algoritmos (e.g., Simplex) para solução de modelos de programação matemática, como modelos de programação linear inteira, linear misto-inteira e quadrática.

Os dados do problema de otimização são então mapeados para as entradas do modelo matemático derivado e, na sequência, o resolutor é executado. Ao fim da execução, os valores obtidos para as variáveis de decisão representam a solução para o problema modelado. Ante o fato de que temos duas funções objetivos a serem consideradas, adotamos as seguintes duas alternativas de resolução nos experimentos.

- Alternativa do *primeiro objetivo*: Priorizar pelo primeiro objetivo e, em seguida, selecionar a solução que melhor atende ao segundo objetivo. Esta alternativa visa priorizar os assinantes móveis, reduzindo o *atraso médio*.
- Alternativa do *segundo objetivo*: Priorizar pelo segundo objetivo e, em seguida, selecionar a solução que melhor atende ao primeiro objetivo. Esta alternativa visa priorizar os provedores de conteúdo, minimizando o *tráfego na rede* e, conseqüentemente, otimizando o uso da infraestrutura de FH.

Ademais, para efeito de modelagem dos cenários dos experimentos, assumimos e/ou relembramos o que se segue.

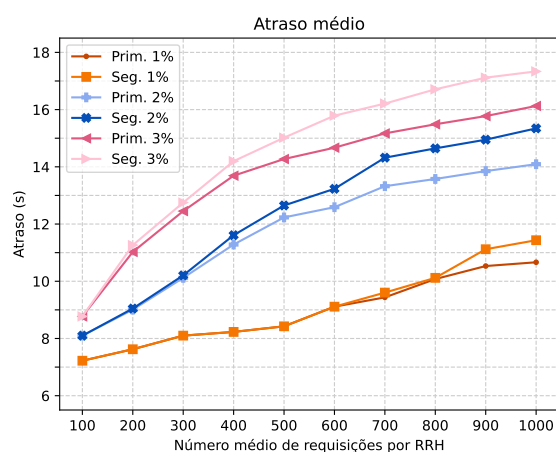
- Os tempos de transmissão de todos objetos multimídia para todos pares de RHHs da infraestrutura de FH são pré-calculados e usados como entrada para o problema de otimização;
- Os assinantes não são modelados individualmente. Caso um assinante faça mais de uma requisição, isso é representado no modelo matemático como múltiplas requisições distintas. Essa abordagem não compromete a representação do problema real nem a sua solução;
- Caso a requisição seja feita para um objeto multimídia já armazenado em algum servidor MEC, esse objeto é transmitido via *streaming* desse servidor até o assinante que fez a requisição. Caso contrário, o servidor MEC da localidade do assinante que fez a requisição deve baixar (do inglês, *download*) o objeto a partir do provedor de conteúdo, armazená-lo localmente e, somente então, enviá-lo via *streaming* para esse assinante;
- Para o caso de múltiplas requisições não sincronizadas para um mesmo objeto devido a múltiplos assinantes em uma mesma localidade, técnicas de comunicações dos tipos *multicast* e *device-to-device* (D2D) são convenientemente empregadas (e.g., [55, 56]). Isso para que apenas uma cópia do objeto multimídia seja transmitida via *streaming* para atender aos assinantes da localidade.

5.3.4 Resultados e Análises

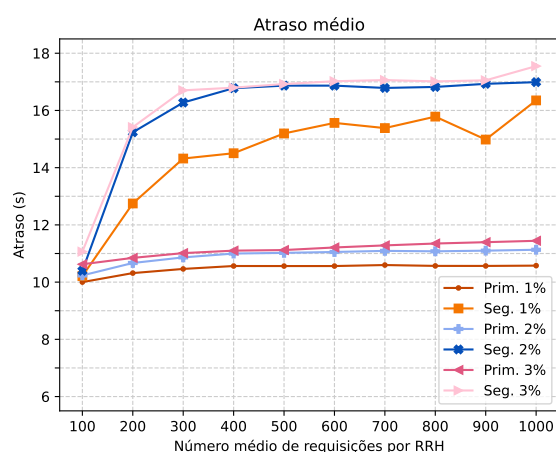
Os resultados das métricas estão nos gráficos das figuras desta seção. Esses resultados são analisados com base nas duas alternativas de resolução, bem como nos valores dos parâmetros

proporção e skew factor, respectivamente. Para facilidade de interpretação dos gráficos, tem-se as observações a seguir.

- O rótulo *Prim.* identifica a alternativa do primeiro objetivo, enquanto o rótulo *Seg.* identifica a alternativa do segundo objetivo.
- Como anteriormente informado, os valores do parâmetro *proporção* são 1%, 2% e 3%;
- Para o parâmetro *skew factor*, são inicialmente considerados dois valores extremos: 0,2 (i.e., os objetos tendem a possuir a mesma popularidade, sem concentração de acessos) e 2,0 (i.e., há um subconjunto de objetos mais populares que concentram mais acessos). Ao final, tem-se também uma discussão sobre a variação desse parâmetro.



(a) Skew factor de 0,2



(b) Skew factor de 2,0

Figura 5. Atraso médio em função do número médio de requisições em cada RRH para diferentes valores de *proporção*.

1) Métrica Atraso médio

A Figura 5a traz o *atraso médio* nas duas alternativas de resolução, com *skew factor* de 0,2 e os diferentes valores de *proporção*. A partir dessa figura, tem-se a análise que segue.

O aumento no número de requisições por RRH aumenta o *atraso médio* percebido. Note que um maior número de requisições implica mais acessos de uma mesma localidade para diferentes objetos multimídia, dificultando manter todos os títulos requisitados no servidor MEC mais próximo. Daí, exige-se o acesso a servidores remotamente localizados. Essa tendência também ocorre para um *skew factor* de 2,0 (vide Figura 5b). Porém, esse aumento é mais modesto. Isso ocorre porque a maioria das requisições se concentra em poucos objetos multimídia, facilitando que estes sejam replicados em vários servidores e mantidos mais próximos dos usuários.

Para um mesmo padrão de requisições por RRH, quanto maior é o valor de *proporção*, maior é o *atraso médio*. Note que, ao aumentar o valor de *proporção*, aumenta-se também o número de objetos multimídia diferentes a armazenar nos servidores MEC e, dessa forma, reduzem-se as chances de que um objeto multimídia esteja armazenado em mais de um servidor. Isso decorre do fato de que a quantidade de objetos multimídia a serem armazenados é maior, mas a capacidade de disco de todos servidores MEC se mantém inalterada. Logo, com menos replicações, aumentam-se as chances de que uma requisição seja atendida por um servidor MEC mais distante, aumentando assim o *atraso médio*. Essa tendência também ocorre para um *skew factor* de 2,0 (vide Figura 5b). Todavia, os valores do *atraso médio* da alternativa do primeiro objetivo são menores que todos aqueles da alternativa do segundo objetivo. Isso ocorre porque a maioria das requisições se concentra em poucos objetos multimídia, o que termina sendo melhor explorado pela alternativa do primeiro objetivo.

2) Métrica tráfego na rede

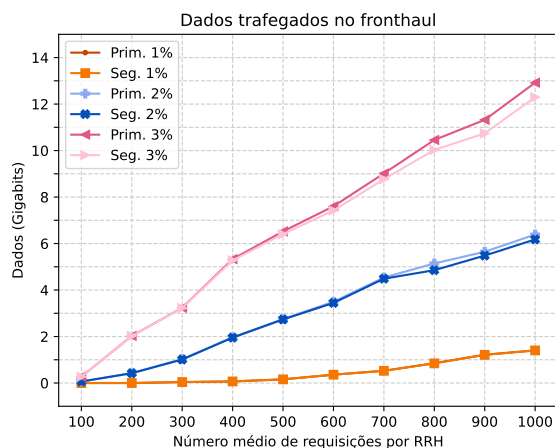
A Figura 6 traz a métrica *tráfego na rede* nas duas alternativas de resolução, com os diferentes valores de *proporção*. A partir dessa figura, tem-se a análise que segue.

Para um mesmo valor de *proporção*, a alternativa do primeiro objetivo apresenta um valor de *tráfego na rede* maior que aquele obtido pela alternativa do segundo objetivo. Em específico, no caso de pouca concentração de acessos, i.e., todos objetos com semelhante popularidade (vide Figura 6a), essa diferença é modesta. Por outro lado, no caso de maior concentração de acessos, i.e., há um subconjunto de objetos mais populares (vide Figura 6b), essa diferença é mais expressiva. Isso ocorre porque, nesse último caso, há menos objetos multimídia que precisam ser armazenados, i.e., apenas aqueles mais frequentemente acessados. Assim, a replicação desses objetos em vários servidores é mais facilmente realizada.

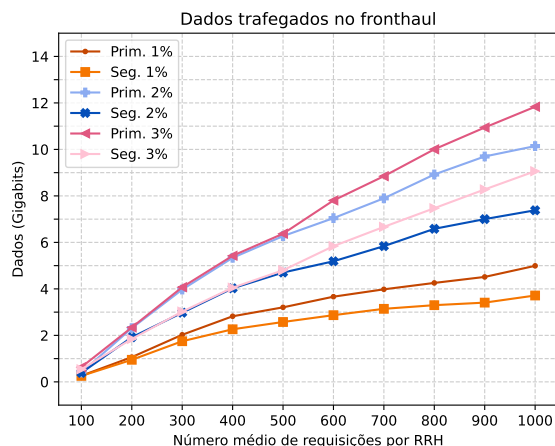
A análise conjunta dos resultados das métricas *tráfego na rede* e *atraso médio* sugere, portanto, que a escolha de uma das duas alternativas de resolução depende do cenário considerado. Quando não há um subconjunto de objetos mais populares (i.e., baixo valor de *skew factor*), o ganho em utilizar uma ou outra alternativa de resolução não é quantitativa-

mente evidente. No entanto, para esse caso deve-se optar pela alternativa do *segundo objetivo*, já que essa alternativa é a que mais deixa os enlaces da rede de FH livres para transmissão de dados provenientes de outras fontes de tráfego concorrentes (e.g., voz e imagem).

Por outro lado, para cenários em que há uma concentração de acessos (i.e., alto valor de *skew factor*), a escolha de uma ou outra alternativa de resolução tem diferenças mais evidentes. Neste contexto, a escolha da alternativa deve então depender do que é mais importante de ser priorizado. Se a prioridade for a QoE (e.g., atrasos percebidos pelo usuário), então se deve utilizar a alternativa do *primeiro objetivo*. Porém, se a prioridade for a QoS (e.g., tráfego na rede de FH), então se deve utilizar a alternativa do *segundo objetivo*. Uma discussão específica do impacto da variação do *skew factor* é feita mais adiante neste texto.



(a) *Skew factor* de 0,2



(b) *Skew factor* de 2,0

Figura 6. Tráfego na rede em função do número médio de requisições em cada RRH para diferentes valores de *proporção*.

3) Métrica *taxa de acertos*

A Figura 7 traz a métrica *taxa de acertos* nas duas alter-

nativas de resolução, com os diferentes valores de *proporção*. A partir dessa figura, tem-se a análise que segue.

Para ambas alternativas de resolução, a *taxa de acertos* diminui com o aumento do número de requisições. Isso ocorre porque o número de servidores MEC e suas respectivas capacidades de armazenamento se mantêm inalteradas. Daí, a chance de os usuários estarem próximos a um servidor com os conteúdos solicitados é reduzida. Em particular, note que a alternativa do *primeiro objetivo* tem resultados melhores porque, como já mencionado, essa alternativa busca alocar objetos multimídia nos servidores próximos aos usuários requisitantes. Isso reduz o tempo de acesso e, ainda, aumenta a *taxa de acertos*.

Também é possível notar que a *taxa de acertos* tem comportamentos distintos em função valor do *skew factor*. Para o *skew factor* de 2,0 (vide Figura 7b), a alternativa do *primeiro objetivo* produz um *taxa de acertos* próxima de 100%. Isso não ocorre na alternativa do *segundo objetivo*, pois essa estratégia prioriza reduzir o tráfego na rede de forma geral, sem necessariamente priorizar o posicionamento dos conteúdos nos servidores próximos dos usuários requisitantes, como o faz a alternativa do *primeiro objetivo*. Para o *skew factor* de 0,2 (vide Figura 7a), a diferença entre as alternativas de resolução não é expressiva. Isso ocorre porque, como não há um subconjunto de títulos mais populares, as duas alternativas terminam pouco diferindo em desempenho, com pequena vantagem para a alternativa do *primeiro objetivo*.

4) Variação do parâmetro *skew factor*

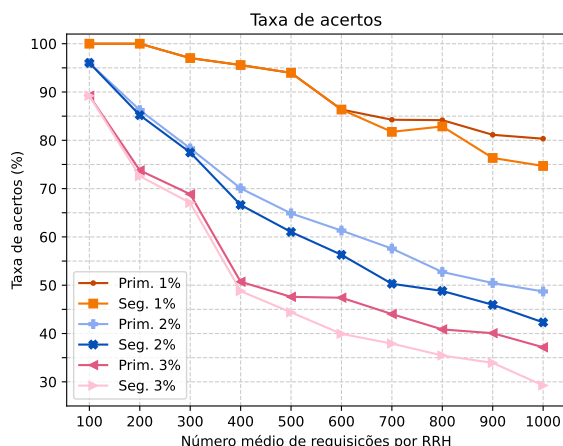
Para a análise do impacto do parâmetro *skew factor*, adota-se o seguinte procedimento. Fixa-se a *proporção* de objetos multimídia armazenados nos servidores MEC em 1%, e observa-se o *atraso médio* obtido para diferentes valores do *skew factor* usando a alternativa do *primeiro objetivo*.

Aumentar o valor *skew factor* implica fazer com que requisições passem a estar muito mais concentradas em menos títulos. Isso permite armazenar menos títulos nos servidores, priorizando aqueles que estão mais próximos dos usuários requisitantes. A tendência esperada é então a redução do *atraso médio* de todas as requisições.

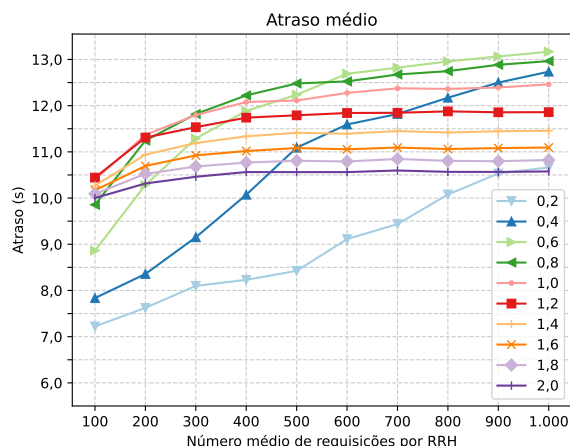
Essa tendência é confirmada para valores de *skew factor* entre 1,2 e 2,0 (vide Figura 8b). No entanto, essa tendência não é evidente para valores de *skew factor* mais baixos (até 1,0) (vide Figura 8a). Daí, conjecturamos que a modelagem de cenários de alta concentração de acessos deve ser feita apenas com valores de *skew factor* acima de 1,0. A discussão mais detalhada dessa tendência é deixada como um possível trabalho futuro.

6. Conclusões e Trabalhos Futuros

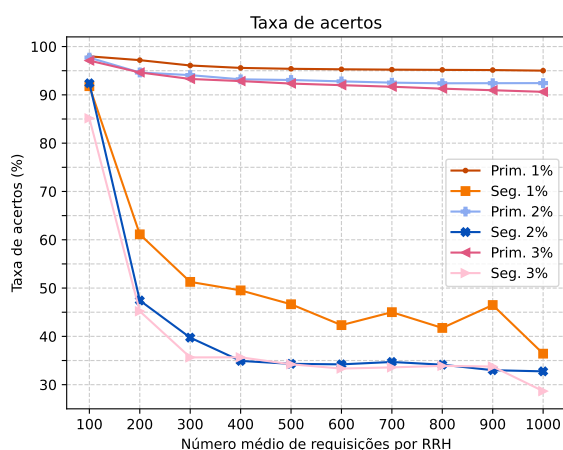
Este artigo propôs um modelo parametrizado de programação linear para o posicionamento eficiente de conteúdos multimídia, considerando o contexto de serviços de *streaming* em redes celulares 5G de arquitetura CRAN com servidores MEC. Como consequência, os dois objetivos a seguir precisavam ser alcançados: (i) otimização do tempo médio de transmissão



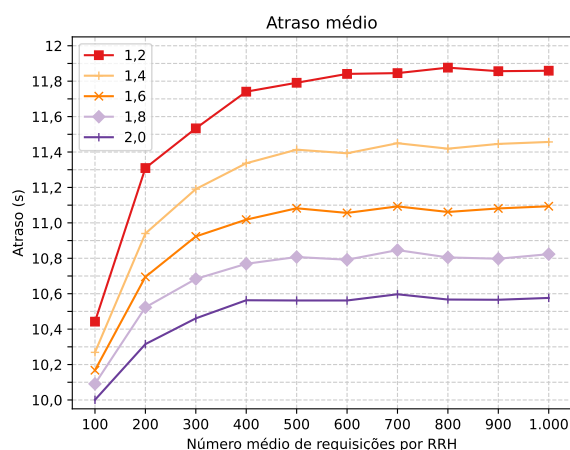
(a) Skew factor de 0,2



(a) Skew factor entre 0,2 e 2,0



(b) Skew factor de 2,0



(b) Skew factor entre 1,2 e 2,0

Figura 7. Taxa de acertos em função do número médio de requisições em cada RRH para diferentes valores de proporção.

Figura 8. Atraso médio com diferentes valores de skew factor, considerando a alternativa do primeiro objetivo com valor de proporção de 1%.

dos objetos multimídia na rede celular; e (ii) otimização do volume do tráfego de dados transmitido pela rede celular.

A resolução do modelo foi obtida utilizando uma das seguintes duas alternativas de resolução: (i) priorização do tempo de transmissão dos objetos multimídia; e (ii) priorização do tráfego de dados na rede. Para demonstrar a aplicabilidade do modelo, cenários de *streaming* foram analisados, cada um deles envolvendo diferentes padrões de requisição de objetos, capacidades de armazenamento dos servidores MEC e concentrações de acessos (i.e., popularidade dos objetos). Três métricas de desempenho foram avaliadas nos experimentos: *atraso médio*, *tráfego na rede* e *taxa de acertos*.

Para se ter a maior eficiência no posicionamento dos conteúdos multimídia, os resultados globais dos experimentos mostraram que: (i) a primeira alternativa de resolução deve ser considerada quando a popularidade dos objetos segue uma distribuição não uniforme, resultando em um determinado subconjunto de objetos com muito mais requisições; e

(ii) a segunda alternativa de resolução é mais recomendada quando a popularidade do objeto tende a ser uniformemente distribuída, pois possibilita deixar os enlaces de comunicação mais livres para transmissão de outras fontes concorrentes de tráfego. Como principal contribuição, esta pesquisa ofereceu subsídios teóricos relevantes para o desenvolvimento de projetos de *streaming* em redes celulares 5G.

Como trabalhos futuros e cientes das limitações desta pesquisa, sugerimos o seguinte: (i) avaliação do modelo proposto em áreas de maior abrangência geográfica, considerando maior diversidade de padrões de requisições, mais níveis de popularidade de objetos e outras capacidades de armazenamento dos servidores MEC; (ii) análise do impacto da implementação do serviço de *streaming* nos níveis de QoS e QoE; (iii) estudo do problema de otimização do volume de dados na infraestrutura de FH juntamente com aquele originado ou com destino aos provedores de conteúdo, envolvendo os enlaces na

infraestrutura de BH e na CN. Por fim, (iv) análise de diferentes categorias de arquiteturas e suas particularidades, como as arquiteturas HCRAN e FRAN, respectivamente. Para essas sugestões, conjecturamos que nem sempre será possível derivar uma solução matemática exata (ótima) para o problema do posicionamento de conteúdos, que poderá ser resolvido alternativamente por meio de heurísticas.

Contribuições dos autores

Os autores contribuíram igualmente para: concepção e elaboração da pesquisa; realização dos experimentos; interpretação e análise dos resultados; e redação do manuscrito.

Referências

- [1] CISCO. Cisco Annual Internet Report (2018–2023). 2020. White Paper. Disponível em: <<https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>>.
- [2] LI, Y. et al. Optimized Content Caching and User Association for Edge Computing in Densely Deployed Heterogeneous Networks. IEEE Transactions on Mobile Computing, v. 21, n. 6, p. 2130–2142, 2022.
- [3] SANDVINE. 2023 Global Internet Phenomena Report. 2023. White Paper. Disponível em: <<https://www.sandvine.com/global-internet-phenomena-report-2023>>.
- [4] RACHAKONDA, L. P.; SIDDULA, M.; SATHYA, V. A comprehensive study on IoT privacy and security challenges with focus on spectrum sharing in Next-Generation networks (5G/6G/beyond). High-Confidence Computing, v. 4, n. 2, p. 100220, 2024. ISSN 2667-2952. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2667295224000230>>.
- [5] SHRAMA, L.; JAVALI, A.; ROUTRAY, S. K. An Overview of High Speed Streaming in 5G. In: 2020 International Conference on Inventive Computation Technologies (ICICT). [S.l.: s.n.], 2020. p. 557–562.
- [6] LIN, P. et al. Joint Optimization of Preference-Aware Caching and Content Migration in Cost-Efficient Mobile Edge Networks. IEEE Transactions on Wireless Communications, p. 1–1, 2023.
- [7] LI, K. et al. Coordination of Macro Base Stations for 5G Network with User Clustering. Sensors, v. 21, n. 16, 2021. Disponível em: <<https://www.mdpi.com/1424-8220/21/16/5501>>.
- [8] PATTARANANTAKUL, M.; VORAKULPIPAT, C.; TAKAHASHI, T. Service Function Chaining security survey: Addressing security challenges and threats. Computer Networks, v. 221, p. 109484, 2023. ISSN 1389-1286. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1389128622005187>>.
- [9] BHANDARI, A. et al. Latency optimized C-RAN in optical backhaul and RF fronthaul architecture for beyond 5G network: A comprehensive survey. Computer Networks, v. 247, p. 110459, 2024. ISSN 1389-1286. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1389128624002913>>.
- [10] AGUILAR-ARMIJO, J.; TIMMERER, C.; HELLWAGNER, H. SPACE: Segment Prefetching and Caching at the Edge for Adaptive Video Streaming. IEEE Access, v. 11, p. 21783–21798, 2023.
- [11] SARAH, A.; NENCIONI, G.; KHAN, M. M. I. Resource Allocation in Multi-access Edge Computing for 5G-and-beyond networks. Computer Networks, v. 227, p. 109720, 2023. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1389128623001652>>.
- [12] LIANG, B.; GREGORY, M. A.; LI, S. Multi-access edge computing fundamentals, services, enablers and challenges: A complete survey. Journal of Network and Computer Applications, v. 199, p. 103308, 2022. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1084804521002976>>.
- [13] RODOSHI, R. T.; KIM, T.; CHOI, W. Fuzzy Logic and Accelerated Reinforcement Learning-Based User Association for Dense C-RANs. IEEE Access, v. 9, p. 117910–117924, 2021.
- [14] HOSSAIN, M. F. et al. Recent research in cloud radio access network (C-RAN) for 5G cellular systems - A survey. Journal of Network and Computer Applications, v. 139, p. 31–48, 2019. ISSN 1084-8045. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1084804519301432>>.
- [15] PANA, V. S.; BABALOLA, O. P.; BALYAN, V. 5G radio access networks: A survey. Array, v. 14, p. 100170, 2022. ISSN 2590-0056. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2590005622000315>>.
- [16] ALAM, M. J. et al. An overview of LTE/LTE-A heterogeneous networks for 5G and beyond. Transactions on Emerging Telecommunications Technologies, Wiley Online Library, p. e4806, 2023.
- [17] SILVEIRA, L. B. et al. Tutorial on communication between access networks and the 5G core. Computer Networks, v. 216, p. 109301, 2022. ISSN 1389-1286. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1389128622003528>>.
- [18] ULLAH, Y. et al. A Survey on Handover and Mobility Management in 5G HetNets: Current State, Challenges, and Future Directions. Sensors, MDPI AG, v. 23, n. 11, p. 5081, May 2023. ISSN 1424-8220. Disponível em: <<http://dx.doi.org/10.3390/s23115081>>.
- [19] REHMAN, A. U.; ROSLEE, M. B.; JIAT, T. J. A Survey of Handover Management in Mobile HetNets: Current Challenges and Future Directions. Applied Sciences, MDPI AG, v. 13, n. 5, p. 3367, Mar 2023. ISSN 2076-3417. Disponível em: <<http://dx.doi.org/10.3390/app13053367>>.

- [20] LIN, C.-H. et al. Energy Efficient Fog RAN (F-RAN) with Flexible BBU Resource Assignment for Latency Aware Mobile Edge Computing (MEC) Services. In: 2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall). [S.l.: s.n.], 2019. p. 1–6.
- [21] TIAN, F.; ZHANG, P.; YAN, Z. A Survey on C-RAN Security. IEEE Access, v. 5, p. 13372–13386, 2017.
- [22] NIU, B. et al. A Dynamic Resource Sharing Mechanism for Cloud Radio Access Networks. IEEE Transactions on Wireless Communications, v. 15, n. 12, p. 8325–8338, 2016.
- [23] MEI, H.; PENG, L. Flexible functional split for cost-efficient C-RAN. Computer Communications, v. 161, p. 368–374, 2020. ISSN 0140-3664. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0140366420318417>>.
- [24] MANGLA, C. et al. Mitigating 5G security challenges for next-gen industry using quantum computing. Journal of King Saud University - Computer and Information Sciences, v. 35, n. 6, p. 101334, 2023. ISSN 1319-1578. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1319157822002373>>.
- [25] DAO, N.-N. et al. A review on new technologies in 3GPP standards for 5G access and beyond. Computer Networks, v. 245, p. 110370, 2024. ISSN 1389-1286. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1389128624002020>>.
- [26] WU, J. et al. Cloud radio access network (C-RAN): a primer. IEEE Network, v. 29, n. 1, p. 35–41, 2015.
- [27] SURYAPRAKASH, V.; ROST, P.; FETTWEIS, G. Are heterogeneous cloud-based radio access networks cost effective? IEEE Journal on Selected Areas in Communications, v. 33, n. 10, p. 2239–2251, 2015.
- [28] ALHUMAIMA, R. S.; KHAN, M.; AL-RAWESHIDY, H. S. Component and parameterised power model for cloud radio access network. IET Communications, v. 10, n. 7, p. 745–752, 2016.
- [29] BASSOLI, R.; RENZO, M. D.; GRANELLI, F. Analytical energy-efficient planning of 5G cloud radio access network. In: 2017 IEEE International Conference on Communications (ICC). [S.l.: s.n.], 2017. p. 1–4.
- [30] CHECKO, A. et al. Cloud RAN for Mobile Networks—A Technology Overview. IEEE Communications Surveys & Tutorials, v. 17, n. 1, p. 405–426, 2015.
- [31] SIMEONE, O. et al. Cloud radio access network: Virtualizing wireless access for dense heterogeneous systems. Journal of Communications and Networks, v. 18, n. 2, p. 135–149, 2016.
- [32] PENG, M. et al. Recent Advances in Cloud Radio Access Networks: System Architectures, Key Techniques, and Open Issues. IEEE Communications Surveys & Tutorials, v. 18, n. 3, p. 2282–2308, 2016.
- [33] KANI, J. ichi; KUWANO, S.; TERADA, J. Options for future mobile backhaul and fronthaul. Optical Fiber Technology, v. 26, p. 42–49, 2015. Next Generation Access Networks. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1068520015000954>>.
- [34] SAFAVAT, S.; SAPAVATH, N. N.; RAWAT, D. B. Recent advances in mobile edge computing and content caching. Digital Communications and Networks, v. 6, n. 2, p. 189–194, 2020. ISSN 2352-8648. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352864819300227>>.
- [35] NARAYANAN, A. et al. A First Look at Commercial 5G Performance on Smartphones. In: Proceedings of The Web Conference 2020. New York, NY, USA: Association for Computing Machinery, 2020. (WWW '20), p. 894–905. Disponível em: <<https://doi.org/10.1145/3366423.3380169>>.
- [36] XU, D. et al. Understanding Operational 5G: A First Measurement Study on Its Coverage, Performance and Energy Consumption. In: Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication. New York, NY, USA: [s.n.], 2020. p. 479–494.
- [37] PRERNA, D.; TEKCHANDANI, R.; KUMAR, N. Device-to-device content caching techniques in 5G: A taxonomy, solutions, and challenges. Computer Communications, v. 153, p. 48–84, 2020. ISSN 0140-3664. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0140366419318225>>.
- [38] AYTAC, K. N. et al. Device-to-device caching for video content delivery. In: 2017 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom). [S.l.: s.n.], 2017. p. 1–2.
- [39] BANIATA, M. et al. Energy-Balancing Unequal Concentric Chain Clustering (MIMO-UCC) Protocol for IoT System in 5G Environment. In: Proceedings of the 2018 Conference on Research in Adaptive and Convergent Systems. New York, NY, USA: Association for Computing Machinery, 2018. (RACS '18), p. 68–74. ISBN 9781450358859. Disponível em: <<https://doi.org/10.1145/3264746.3264747>>.
- [40] GONG, Y.; WANG, J.; LAI, G. Energy-efficient Query-Driven Clustering protocol for WSNs on 5G infrastructure. Energy Reports, v. 8, p. 11446–11455, 2022. ISSN 2352-4847. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S235248472201719X>>.
- [41] TENG, W. et al. Content Placement and User Association for Delay Minimization in Small Cell Networks. IEEE Transactions on Vehicular Technology, v. 68, n. 10, p. 10201–10215, 2019.
- [42] WANG, Y. et al. Joint Caching Placement and User Association for Minimizing User Download Delay. IEEE Access, v. 4, p. 8625–8633, 2016.

- [43] WU, H. et al. Energy and Delay Optimization for Cache-Enabled Dense Small Cell Networks. IEEE Transactions on Vehicular Technology, v. 69, n. 7, p. 7663–7678, 2020.
- [44] LIU, Y. et al. Joint User Association and Caching in Wireless Heterogeneous Networks with Backhaul. In: ICC 2021 - IEEE International Conference on Communications. [S.l.: s.n.], 2021. p. 1–6.
- [45] MOLNER, N. et al. Optimization of an integrated fronthaul/backhaul network under path and delay constraints. Ad Hoc Networks, v. 83, p. 41–54, 2019. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1570870518306206>>.
- [46] NAUDTS, B. et al. How can a mobile service provider reduce costs with software-defined networking? International Journal of Network Management, Wiley Online Library, v. 26, n. 1, p. 56–72, 2016.
- [47] WIKIPÉDIA. Morumbi (distrito de São Paulo). 2023. ONLINE. Disponível em: <[https://pt.wikipedia.org/wiki/Morumbi_\(distrito_de_S%C3%A3o_Paulo\)](https://pt.wikipedia.org/wiki/Morumbi_(distrito_de_S%C3%A3o_Paulo))>.
- [48] COOK, S. Netflix statistics & facts that define the company's dominance in 2023. 2023. ONLINE. Disponível em: <<https://www.comparitech.com/blog/vpn-privacy/netflix-statistics-facts-figures/>>.
- [49] STOLL, J. Countries with most content available on Netflix worldwide as of March 2023. 2023. ONLINE. Disponível em: <<https://www.statista.com/statistics/1013571/netflix-library-size-worldwide/>>.
- [50] WANG, T. et al. Estimating Video Popularity From Past Request Arrival Times in a VoD System. IEEE Access, v. 8, p. 19934–19947, 2020. Disponível em: <<https://ieeexplore.ieee.org/document/8959235>>.
- [51] CHERKASOVA, L.; GUPTA, M. Analysis of enterprise media server workloads: access patterns, locality, content evolution, and rates of change. IEEE/ACM Transactions on Networking, v. 12, n. 5, p. 781–794, 2004. Disponível em: <<https://ieeexplore.ieee.org/document/1344003>>.
- [52] MITRA, S. et al. Characterizing Web-Based Video Sharing Workloads. ACM Trans. Web, Association for Computing Machinery, New York, NY, USA, v. 5, n. 2, may 2011. ISSN 1559-1131. Disponível em: <<https://doi.org/10.1145/1961659.1961662>>.
- [53] ALMEIDA, J. M. et al. Analysis of educational media server workloads. In: Proceedings of the 11th International Workshop on Network and Operating Systems Support for Digital Audio and Video. New York, NY, USA: Association for Computing Machinery, 2001. (NOSSDAV '01), p. 21–30. ISBN 1581133707. Disponível em: <<https://doi.org/10.1145/378344.378348>>.
- [54] Gurobi Optimization. Gurobi Optimizer Reference Manual - Version 11.0. 2024. ONLINE. Disponível em: <<https://www.gurobi.com/documentation/current/refman/index.html>>.
- [55] YIN, J. et al. Nefis: A network coding based flexible device-to-device video streaming scheme. Journal of Network and Computer Applications, v. 227, p. 103892, 2024. ISSN 1084-8045. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1084804524000699>>.
- [56] JIMÉNEZ-SORIA, D.; MARTÍN-VEGA, F. J.; AGUAYO-TORRES, M. C. Coordinated Multicast/Unicast Transmission on 5G: A Novel Approach for Linear Broadcasting. Wireless Personal Communications, v. 121, n. 2, p. 1273–1287, Nov 2021. ISSN 1572-834X. Disponível em: <<https://doi.org/10.1007/s11277-021-09057-z>>.