

# Dimensionamento ótimo de fronthaul óptico com divisão flexível de funções de processamento em CF-RAN

Matias R. P. dos Santos<sup>1</sup>, Rodrigo I. Tinini<sup>2</sup>, Tiago Januario<sup>1</sup>, Gustavo B. Figueiredo<sup>1</sup>

<sup>1</sup>Instituto de Computação – IC - Universidade Federal do Bahia (UFBA)  
Av. Adhemar de Barros, s/n - Ondina, Salvador - BA, 40170-110

<sup>2</sup>Universidade de São Paulo – São Paulo (USP)

matiasrps@ufba.br, rtinini@ime.usp.br, {januario,gustavo}@dcc.ufba.br

**Abstract.** *The baseband processing functions centralization in a cloud pool emerged as an alternative to reduce CAPEX and OPEX for the 5G network. However, this same centralization presented challenges for the fronthaul access networks in terms of bandwidth and delay. To mitigate such problems, several alternatives proposed by academia and industry have shown to be promising, including baseband functional split and the use of hybrid architectures. In this work, we propose an optimal solution to flexibly decide the baseband functional split considering the state of the network for energy efficiency in a CF-RAN hybrid architecture. The main objective is to make the optimal choice of the split option, centralizing processing functions in the cloud as much as possible.*

**Resumo.** *A centralização do processamento das funções de banda base em um pool de nuvem emergiu como alternativa para redução de CAPEX e OPEX para as redes 5G. No entanto, esta centralização apresenta sérios desafios para as redes de acesso fronthaul em termos de largura de banda e de atraso. Para mitigar tais problemas, várias alternativas propostas mostraram-se promissoras, dentre elas a divisão funcional e o uso de arquiteturas híbridas. Neste trabalho, nós propomos uma solução ótima para decidir de forma flexível a divisão funcional considerando o estado da rede e a eficiência energética como decisão de projeto em uma arquitetura híbrida CF-RAN. O objetivo principal é fazer a escolha ideal da opção de divisão centralizando ao máximo as funções de processamento na nuvem.*

## 1. Introdução

As redes 5G são projetadas para suportar tráfego 1000 vezes maior e latência 10 vezes menor que a geração anterior. Esses requisitos exigiram a proposição de novas arquiteturas de Redes de Acesso a Rádio (do inglês *Radio Access Network* – RANs) a fim de proporcionar maior capacidade de processamento e uso eficiente dos recursos disponíveis. Assim, a arquitetura *Cloud-RAN* foi proposta para reduzir custo operacional (do inglês *Operational Expenditure* – OPEX) e de capital (do inglês *Capital Expenditure* – CAPEX) através da centralização dos elementos de processamento da banda base na nuvem. Essa centralização ocorre devido à separação das funções de recepção de rádio, que são realizadas pelas *Remote Radio Head* (RRHs), e das funções de processamento de banda base, executadas pelas *BaseBand Units* (BBUs).

No entanto, o processamento centralizado impõe alta demanda por largura de banda e requisitos estritos de atraso no *fronthaul* (rede de interconexão entre RRHs e BBUs) [Wu et al. 2015, Konstantinou et al. 2020]. Esses desafios exigiram, além da remodelagem da infraestrutura de RAN, a proposição de técnicas de compressão de modulação adaptativa e separação das funcionalidades da cadeia de processamento de banda base [Figueiredo et al. 2016, Alimi et al. 2018, Ranaweera et al. 2018, Larsen et al. 2019] para contornar os limites estabelecidos de capacidade do *fronthaul*.

No que tange à remodelagem, arquiteturas híbridas foram propostas para mitigar problemas oriundos das arquiteturas centralizadas [Guizani and Hamdi 2017, Riva et al. 2018]. Dentre tais arquiteturas, está a Cloud-Fog RAN (CF-RAN) [Tinini et al. 2019]. A CF-RAN beneficia-se do emprego de *fog computing* e permite o processamento de banda base em nós intermediários alocados próximos aos usuários. O uso de nós intermediários promove, porém, aumento no consumo de energia devido à ativação de nós de processamento adicionais, o que é uma desvantagem quando comparada à C-RAN.

Isso fomentou o desenvolvimento de alternativas como a divisão funcional de processamento. A divisão funcional de processamento, proposta pela *3rd Generation Partnership Project (3GPP)*, permite relaxar, de forma eficiente, a carga no *fronthaul* e os requisitos de atraso associados. O alívio de carga no *fronthaul* ocorre dividindo-se as funções de processamento da pilha de protocolos entre RRH e BBU e colocando-as de forma distribuída próxima aos usuários. Isso porque, quanto mais próximo do RRH está a função de processamento, menor o uso de banda e menores as restrições de atraso, uma vez que cada opção de divisão incorre em requisitos específicos de largura de banda e latência que são ajustáveis ao estado da rede. Isso implica em decisões de divisão de funções baseadas nos requisitos da *User Equipment (UE)*, restrições de *link* e *QoS* [Mei and Peng 2020, 3GPP 2017a, 3GPP 2017b].

Claramente, tal abordagem reduz as demandas de transmissão no *fronthaul*, mas por outro lado, minora os ganhos da centralização. Desta forma, a adoção de divisão funcional cria um compromisso entre a operação energeticamente eficiente da rede (promovida pela centralização das BBUs) e o alívio na carga do *fronthaul* e possível aumento de cobertura da rede, uma vez que mais RRHs podem ser ativados. Assim, idealmente, quanto maior a disponibilidade do *fronthaul* em acomodar demandas e maior a capacidade de processamento do *BBU pool*, maior deverá ser o nível de centralização das funções. Em outras palavras, se há 100% de capacidade de processamento em nuvem, sem restrições de capacidade no *fronthaul*, todas as funções de processamento serão alocadas para a nuvem devido os ganhos da centralização. Em havendo limites (como é o caso prático) desses recursos, deve-se usar a divisão e colocação de funções como forma de aliviar as demandas do *fronthaul*. Dessa forma, a decisão sobre cada divisão funcional e como a rede deve se comportar com a sua utilização é um problema de grande relevância [Wang et al. 2017, Larsen et al. 2019]. Além disso, a escolha de uma única opção de divisão funcional para implantação de rede não dá flexibilidade para fornecer *QoS* específico para cada serviço oferecido, que pode requerer latência e requisitos de desempenho específicos. Assim, a divisão funcional flexível surge para melhor captar os benefícios das várias opções de divisão funcional com adaptação e ajuste ao consumo de recursos e requisitos nas diferentes áreas geográficas.

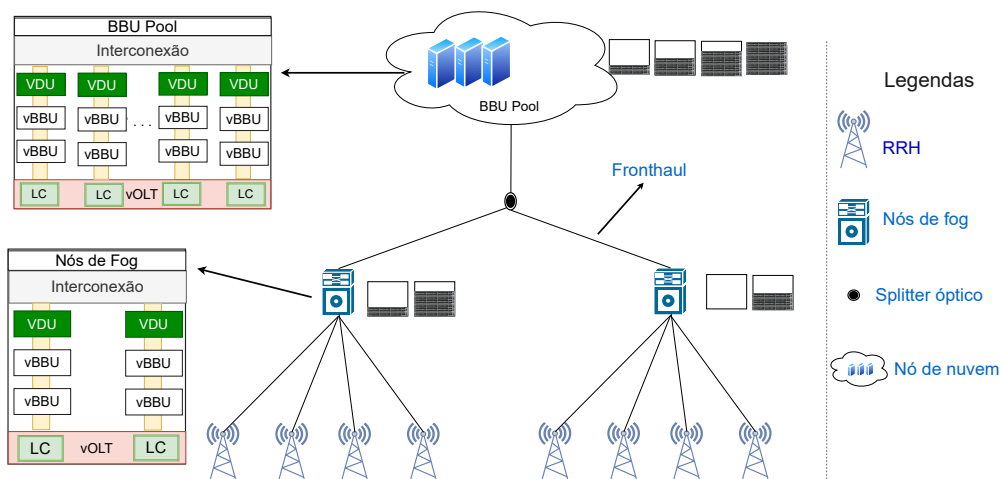
Nesse artigo, propõe-se uma solução ótima para o problema de divisão flexível de funções de processamento mediante o uso de Programação Linear Inteira (PLI) em uma arquitetura híbrida CF-RAN. A solução considera o estado da rede e os altos requisitos de largura de banda para decidir sobre a divisão das funções e a sua colocação na rede. O objetivo é aliviar a carga do fronthaul propondo uma solução ciente do consumo de energia e para isso, uma solução de divisão funcional de processamento flexível é aplicada para cada RRH ativo considerando seus requisitos de *QoS* e o estado da rede. As contribuições desse trabalho são: **1)** proposta de uma solução flexível ótima para particionamento de processamento de banda base. **2)** demonstração que o particionamento e o posicionamento dinâmico das funções com ciência de energia e do *status* da rede permite melhor eficiência no balanceamento entre o consumo de energia, largura de banda e cobertura de rede. **3)** comparação dos resultados entre duas arquiteturas considerando divisão funcional flexível e fixa.

O restante deste trabalho está organizado da seguinte forma: a Seção 2 apresenta trabalhos relacionados focando em divisão funcional e arquiteturas híbridas; a Seção 3 descreve a arquitetura CF-RAN e motivação e a problemática desse trabalho; a Seção 4 apresenta a formulação do PLI e suas restrições; a Seção 5 apresenta os resultados dos experimentos computacionais; a Seção 6 apresenta alguns comentários finais sobre o artigo e fornece orientações futuras de pesquisa.

## 2. Trabalhos relacionados

Vários trabalhos investigam soluções para a melhor aplicação de divisão funcional e posicionamento de BBUs virtuais combinadas com virtualização de funções de rede (*Virtual Network Functions* – VNFs) para hospedar funções de processamento de banda base nos nós de processamento da rede [Tinini et al. 2019, Moreira Zorello et al. 2020]. Alguns estudos aplicam formulações de PLI e heurísticas para mitigar problemas relacionados, como o atraso, a largura de banda e o uso de recursos de processamento [Harutyunyan and Riggio 2018, Gao et al. 2019, Tohidi et al. 2020, Mei and Peng 2020].

Os autores em [Harutyunyan and Riggio 2018] aplicam PLI com o objetivo de minimizar a interferência de acesso conjunto intercelular e a utilização da largura de banda do *fronthaul* selecionando a divisão funcional apropriada. Autores em [Gao et al. 2019] utilizaram um algoritmo de aprendizado por reforço para decidir sobre a alocação e posicionamento de BBUs virtuais em C-RAN. Autores em [Tohidi et al. 2020] estudaram o uso de heurística para o problema de seleção de divisão flexível ideal para aplicações sensíveis ao atraso. Os autores formularam uma heurística com fila para gerar o atraso *end-to-end* e decidir a melhor divisão funcional para o momento. Por fim, os autores em [Mei and Peng 2020] propuseram esquemas para divisão funcional flexível utilizando o conceito de *Virtualized RRHs* (V-RRHs) com o intuito de balancear os recursos disponíveis reduzindo custos operacionais. Para esse fim, os autores modelaram um programa não linear inteiro puro (do inglês *pure integer non-linear programming* (PINLP)) para obter a otimalidade em simulações. Em relação ao dimensionamento do *fronthaul* baseado em consumo de energia e largura de banda, os autores em [Wang et al. 2017] propuseram um modelo de programação por restrição para decidir sobre a divisão funcional ideal.



**Figura 1. Arquitetura do CF-RAN**

Diferentemente dos trabalhos citados acima, apresentamos uma solução baseada em formulação matemática PLI em uma arquitetura híbrida com decisão flexível de divisão funcional para cada RRH na rede. Nossa solução decide a divisão funcional para cada RRH conectada na rede conforme as restrições de capacidade são identificadas.

### 3. A Arquitetura CF-RAN e a Divisão de funções

A arquitetura CF-RAN e as opções de divisão funcional são detalhadas nesta seção. A arquitetura descreve os componentes e a estrutura operacional da rede, enquanto as divisões funcionais detalham a disposição das funções de banda base e seu impacto na latência e largura de banda.

#### 3.1. A Arquitetura CF-RAN

A Figura 1 apresenta a arquitetura CF-RAN com representação de indução de divisão funcional. A arquitetura corresponde a duas camadas principais de processamento, compostas por um nó de nuvem e vários nós de fog. A camada de nuvem centraliza inicialmente todo o processamento de banda base para economizar o consumo de energia e possui uma grande capacidade de processamento. Os nós de fog estão dispostos próximos aos usuários e são ativados para receber a carga de trabalho excedida mediante alguma restrição no *fronthaul* ou nuvem.

A CF-RAN propõe o uso do paradigma *Network Functions Virtualization* (NFV) para que BBUs virtualizadas (vBBUs) sejam ativadas dinamicamente apenas sob demanda de forma que não haja gasto energético desnecessário. Conjuntos de Unidades Digitais Virtuais (do inglês *Virtual Digital Unit* – VDUs) estão presente em ambos os nós de processamento, isto é, na nuvem e na *fog* e realizam o processamento de banda base por intermédio de funções de processamento virtualizadas (do inglês *Virtualized Processing Functions* – VPFs). Uma dessas VPFs é o vBBU, que recebe e processa os sinais de banda base de um RRH específico. São usadas ONUs (do inglês *Optical Network Unit* – ONU) para conectar cada RRH ao *fronthaul*. Cada nó de processamento, por sua vez, implementa uma OLT virtual (*Virtualized Optical Line Terminal* – vOLT), que possui um conjunto de Line Cards, também chamados *transceivers* ópticos, que encaminham

os dados oriundos da fibra de alimentação para sua respectiva VDU. Para a arquitetura CF-RAN, é utilizado *fronthaul* óptico com adição de links ópticos que para conectar os RRHs aos nós de *fog*.

O *fronthaul* da CF-RAN é implementado sobre uma rede TWDM-PON (do inglês *Time Wavelength Division Multiplexing Passive Optical Network*) com dois níveis de *splitters* ópticos [Figueiredo et al. 2016]. Com tal arquitetura, é possível a atribuição dos comprimentos de onda para grupos diferentes de ONUs (que por sua vez estão conectados às RRHs), criando, assim, as chamadas VPONs (do inglês *Virtual Passive Optical Networks*). Como dito, tais VPONs podem ser compartilhadas por vários RRHs no domínio do tempo para transmitir para um nó de processamento comum, o que aumenta os ganhos estatísticos de multiplexação [Riva et al. 2018, Tinini et al. 2020].

É importante destacar que a implementação do *fronthaul* sobre TWDM-PONs é projetada para atingir alcances entre 20 e 40 km, devido às restrições de atraso do protocolo CPRI (do inglês *Common Public Radio Interface*) usado no *fronthaul* para a comunicação *RRH – BBU* e, claro, ao próprio atraso de propagação em tais distâncias. Além disso, os limites tecnológicos do *fronthaul* decorrem também da alta demanda por largura de banda em configurações normais de antenas. A título ilustrativo, tomando-se a arquitetura centralizada C-RAN, mesmo com aplicação de compressão de taxa de 3:1 e utilizando uma configuração típica de RRH com 64X64 *Massive MIMO*, esquema de modulação em 256 *QAM* e portadora de 200 *MHz*, a demanda de largura de banda no *fronthaul* será de  $\approx 200$  *Gbps* por RRH conectado [Consortium et al. 2017].

O compartilhamento e a divisão de processamento de banda base permite diminuir a demanda por largura de banda e aumentar a distância do *link fronthaul* ao relaxar a latência. Isso permite a redução de CAPEX e OPEX na medida em que uma mesma infraestrutura de rede pode cobrir uma área geográfica maior. No entanto, esses benefícios vêm com um incremento no custo de energia. Decorrente disso, a análise do equilíbrio entre largura de banda, consumo de energia e do gastos com infraestrutura é um trabalho de interesse geral [Consortium et al. 2017, Pelekanou et al. 2018, Larsen et al. 2019]. Decorrente disso, a motivação deste trabalho foca em utilizar uma formulação matemática PLI no problema de decisão flexível de divisão funcional para cada RRH na rede em uma arquitetura híbrida CF-RAN. A solução decide a divisão funcional de funções de processamento para cada RRH conectada na rede conforme as restrições de capacidade são identificadas.

### 3.2. Divisões funcionais

A descrição das divisões funcionais segue a pilha de protocolos LTE (do inglês *Long Term Evolution*), na qual a parte inferior da pilha inclui as três camadas suscetíveis ao particionamento: camada física; camada de enlace de dados; camada de rede. A camada física lida com a conversão de bits digitais em ondas de rádio. As divisões funcionais nesta camada implicam em coordenação estrita entre camadas superiores e, portanto, os altos requisitos de latência permanecem, porém ocorre já redução significativa da largura de banda exigida. Já a camada de enlaces compreende três subcamadas: 1) Protocolo de Controle de link de rádio (*Radio Link Control – RLC*); 2) convergência de dados por pacotes (*Packet Data Convergence Protocol – PDCP*); 3) Controle de acesso à mídia (*Medium Access Control – MAC*). Algumas opções de divisões de processamento na camada de

enlace têm requisitos de latência relaxados no *link* de *fronthaul*, o que permite ajustes às demandas de redes, e a largura de banda é amplamente reduzida. Por fim, a divisão na camada de rede lida com planos de usuário e de controle associado ao *Radio Resource Control* (RRC) [Larsen et al. 2019].

Consideramos apenas cinco opções de divisão de processamento para os testes computacionais que serão realizados na seção 5. Listamos as opções escolhidas e seu detalhamento a seguir:

1. **Divisão E:** esta é a opção de divisão de processamento mais centralizada. Nela, a maioria das funções de processamento é executada na nuvem, enquanto algumas funções de RF (do inglês *Radio Frequency*) permanecem no RRH.
2. **Divisão I:** consiste em colocar a Transformação Rápida de *Fourier* (*fast Fourier transform* – FFT), mapeador de camada, modulação e funções de pré-codificação para processamento local (borda ou *fog*) para reduzir a transmissão da taxa de bits no *fronthaul*, removendo o prefixo cíclico e transformando o sinal recebido no domínio da frequência. Em suma, esta opção mantém o ARQ (do inglês *Automatic repeat request*) centralizado na nuvem e nos elementos de mapeamento das subportadoras locais, o que mantém as restrições de latência altas.
3. **Divisão D:** promove a divisão de funções de processamento relacionadas à camada física que devem ser processadas localmente enquanto centraliza as funções MAC e de rede na nuvem. Esta opção promove uma redução elevada no uso da largura de banda do *fronthaul*, mesmo com a restrição de latência ainda muito rigorosa.
4. **Divisão B:** esta opção coloca o *Radio Link Control* (RLC) e o *buffer* de transferência para processamento local. Como os principais recursos de processamento sensíveis ao tempo são processados localmente, as restrições de latência são reduzidas significativamente. No entanto, funções como CoMP (do inglês *Coordinated Multipoint*) e programação coordenada são limitadas nesta opção. Os ganhos mais significativos nesta opção estão associados à multiplexação, redução elevada na sobrecarga do BBU *pool* e centralização da criptografia.
5. **Divisão A:** esta opção seria a mais próxima da *Distributed-RAN*, em que as funções de banda base são todas processadas localmente. Esta opção exige uma BBU dedicada a cada RRH.

### 3.3. Largura de banda por divisão funcional

Como os desafios básicos da implantação da rede 5G está associada ao atraso e à largura de banda, a estratégia de dividir o processamento implica em uma abordagem promissora que reduz a taxa de bits do *fronthaul* e o atraso. Ademais, o uso flexível de divisão funcional eleva a capacidade dos operadores mediante escolha dinâmica entre pontos de processamento mediante o *trade-off* de energia por largura de banda exigida. Além disso, *links* dedicados de alta capacidade são custosos e podem ser fatores limitantes. O requisito de largura de banda utilizado para os testes computacionais presentes na seção 5 para cada opção de divisão funcional pode ser calculado de acordo com [Consortium et al. 2017, Marotta et al. 2019] como descritos a seguir:

$$\left\{ \begin{array}{l} \text{Divisão E} = A \times S_r \times S_{f,iq} \times 2 \times N_p \times \left(\frac{16}{15}\right) \\ \text{Divisão I} = N_{sc} \times N_{sym} \times N_{mm} \times S_{iq} \times 2 + O_{phy,D} \\ \text{Divisão D} = PR_u + O_{MAC,u} \\ \text{Divisão B} = PR_u + OPR_u \end{array} \right.$$

**Tabela 1. Notação utilizada para o modelo.**

Símbolos	Definições
<b>Conjuntos</b>	
$i \in R$	conjuntos de demandas de tráfego da RRH
$n \in N$	conjunto de nós de processamento
$w \in W$	conjuntos de VPONs disponíveis
$s \in S$	conjunto de divisões funcionais disponibilizados
<b>Parâmetros</b>	
$B_i^s$	demanda de largura de banda da RRH $i$ de acordo com a divisão funcional $s$
$B_w$	capacidade do VPON $w$
$P_i^s$	demanda de processamento de demandas do RRH $i$ de acordo com a divisão funcional $s$
$P_n$	capacidade de processamento do nó $n$
$M$	número muito grande
$C_n$	custo energético do nó $n$
$C_{lc}$	custo energético da <i>LineCard LC</i>
$C_{split}$	valor atribuído a cada divisão funcional $s$
<b>Variáveis</b>	
$y_{wns}^i$	1 se a demanda $i$ é processada no nó $n$ delimitada por $s$ sendo transmitida no VPON $w$ , 0 caso contrário.
$z_{wn}$	1 se o VPON $w$ está alocado no nó $n$ , 0 caso contrário.
$x_n$	1 se funções de processamento e elementos do nó $n$ estão ativados, 0 caso contrário.
$d_{in}$	1 se as demandas de $i$ estão alocadas no nó $n$ , 0 caso contrário.
$\delta_{is}$	1 se as demandas de $i$ estão associadas com a opção de divisão $s$ , 0 caso contrário.
$t_{isn}$	1 se as demandas de $i$ tem sua opção de divisão $s$ sendo processada no nó $n$ , 0 caso contrário.

Onde  $A$  representa o número de antenas por setor,  $S_r$  é o *sampling rate* (30 MS/s por 20 MHz),  $S_{f,iq}$  é *bit-width* do rádio sobre fibra (*Radio over fiber* (Rof)) por amostras (16 para LTE),  $N_p$  é o número de portas de antena (consideramos duas),  $N_{sc}$  é o número de sub-portadores (1200 para LTE),  $N_{sym}$  é o número de símbolos (14 por subframe),  $N_{mm}$  é o número de camadas MIMO,  $S_{iq}$  é a largura de bits da amostra (sete para DL),  $O_{phy,D}$  é o *overhead* da camada PHY (9 Mbps),  $PR_u$  é a taxa de pico de dados DL da *Long Term Evolution* (LTE) (150 Mbps) e  $O_{MAC,u}$  é o DL *overhead* MAC-PHY (5 Mbps) e o  $OPR_u$  corresponde ao PDCP *overhead* no DL (16 Mbps).

## 4. Divisão ótima de funções de processamento

Nesta seção, será apresentada a formulação do PLI para resolver o problema de seleção ideal da divisão flexível de processamento de banda base para dimensionamento do fronthaul.

### 4.1. Formulação do PLI

A formulação do PLI proposto realiza o dimensionamento do fronthaul objetivando a minimização do consumo de energia. As notações utilizadas na formulação do problema são apresentadas na Tabela 1.

#### Função Objetivo

$$\text{Min} \sum_{n=1}^N x_n \cdot C_n + C_{lc} \sum_{w=1}^W \sum_{n=1}^N z_{wn} + C_{split} \sum_{i=1}^I \sum_{s=1}^S \delta_{is} \quad (1)$$

A função objetivo (1) minimiza a ativação geral dos elementos de processamento do CF-RAN, como nós de processamento, VDUs, vBBUs, Line Cards, colocando ao máximo as demandas de RRHs em um único VPONs para processamento ao máximo na nuvem. Nossa formulação utiliza valores decrescentes nos custos da opção de divisão funcional para promover a escolha da máxima centralização na nuvem para melhor eficiência

energética. A descentralização das funções de banda base ocorrerá apenas quando a nuvem saturar a sua capacidade de processamento.

### Restrições

$$\sum_{s=1}^S \delta_{is} = 1, \forall i \in R \quad (2)$$

A restrição (2) garante que cada RRH terá ativa apenas uma das opções das divisões funcionais disponíveis. Esta restrição está associada à escolha flexível para cada RRH.

$$\sum_{n=1}^N d_{in} = 2, \forall i \in R \quad (3)$$

$$\sum_{w=1}^W \sum_{n=1}^N \sum_{s=1}^S y_{wns}^i = 2, \forall i \in R \quad (4)$$

Restrições (3) e (4) garantem que dois VPONs serão criados para transmitir a demanda  $i$  para diferentes nós de processamento  $n$  dada a opção de divisão decidida pela formulação, sendo uma direcionada para o nó de fog e outra para a nuvem.

$$\sum_{n=1}^N z_{wn} \leq 1, \forall w \in W \quad (5)$$

A restrição (5) garante que um VPON seja alocado para no máximo um nó de processamento.

$$\sum_{w=1}^W \sum_{n=1}^N y_{wns}^i \leq \delta_{is}, \forall i \in R, \forall s \in S \quad (6)$$

$$\sum_{n=1}^N t_{isn} \leq \delta_{is}, \forall i \in R, \forall s \in S \quad (7)$$

$$\sum_{w=1}^W y_{wns}^i \leq t_{isn}, \forall i \in R, \forall s \in S, \forall n \in N \quad (8)$$

Restrições (6), (7), (8) garantem que a demanda  $i$  seja mapeada usando a restrição (7) e, assim, ocorre a divisão correta do processamento entre nuvem e fog e a correta alocação da demanda para seu respectivo nó de processamento.

$$\sum_{i=1}^R \sum_{n=1}^N \sum_{s=1}^S (y_{wns}^i \times B_i^s) \leq B_w, \forall w \in W \quad (9)$$

$$\sum_{i=1}^R \sum_{w=1}^W \sum_{s=1}^S (y_{wns}^i \times P_i^s) \leq P_n, \forall n \in N \quad (10)$$

Restrições (9), (10) garantem que a capacidade total dos VPONs e dos nós de processamento sejam respeitados para o total de demanda vigente, o que torna possível o envio de demandas apenas se houver capacidade de processamento para isso.



$$M \times x_n \geq \sum_{i=1}^R \sum_{w=1}^W \sum_{s=1}^S y_{wns}^i, \forall n \in N \quad (11)$$

$$x_n \leq \sum_{i=1}^R \sum_{w=1}^W \sum_{s=1}^S y_{wns}^i, \forall n \in N \quad (12)$$

$$M \times z_{wn} \geq \sum_{i=1}^R \sum_{s=1}^S y_{wns}^i, \forall w \in W, \forall n \in N \quad (13)$$

$$z_{wn} \leq \sum_{i=1}^R \sum_{s=1}^S y_{wns}^i, \forall n \in N, \forall w \in W \quad (14)$$

Restrições (11), (12), (13), (14) executam a ativação dos nós de processamento e VPON para reagir corretamente às demandas vigentes e, assim, executar a correta atribuição destes.

$$M \times d_{in} \geq \sum_{w=1}^W \sum_{s=1}^S y_{wns}^i, \forall n \in N, \forall i \in R \quad (15)$$

$$d_{in} \leq \sum_{w=1}^W \sum_{s=1}^S y_{wns}^i, \forall n \in N, \forall i \in R \quad (16)$$

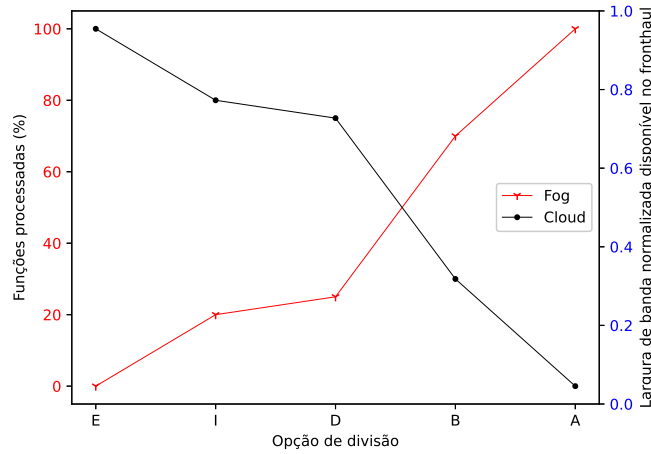
Restrições (15), (16) garantem a ativação de nós de processamento e outros elementos de rede para decisões de redirecionamento de tráfego excedente dado escolhas realizadas na opção de divisão funcional e na largura de banda disponível. Esta restrição garante a ativação de novos elementos de processamento quando não houver recursos suficientes disponíveis relacionados às restrições (9) e (10).

## 5. Exemplos Numéricos

Foram realizadas simulações para avaliar o desempenho da formulação proposta utilizado um simulador de eventos discretos para gerar tráfego de rede semelhante a redes de acesso comerciais [Tinini et al. 2020]. As demandas de tráfego geradas seguem um processo de *Poisson* e possuem um tempo de serviço seguindo uma distribuição exponencial negativa. Este comportamento de tráfego segue o padrão apresentado em [Peng et al. 2011].

Como referência, foram considerados dois casos para comparar a divisão e o posicionamento funcional da rede: A) a base usada para comparação é *Fix Split*, em que o particionamento das funções de banda base e seu posicionamento entre a nuvem e os *fog nodes* são previamente decididos e fixos para todas as execuções subsequentes; B) *Flexible Split* com a decisão de posicionamento e divisão de funções feita dinamicamente. Os parâmetros de simulação são apresentados na Tabela 2 indicando a topologia, a configuração das antenas e o tráfego gerado em cada uma das opções de divisão funcional.

Para os testes dinâmicos em simulação, é considerado uma rede com arquitetura CF-RAN com dez comprimentos de onda ( $w = \{1, 2, \dots, 10\}$ ) de 10 *Gbps* e cinco nós



**Figura 2. Distribuição de funções por opção de divisão funcional**

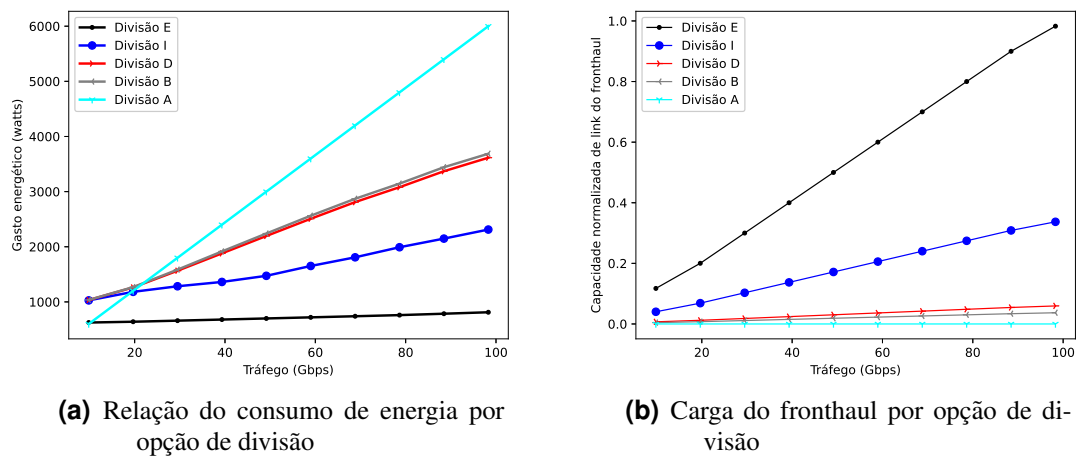
**Tabela 2. Parâmetros de simulações**

Parâmetro	Valor
Topologia	1 cloud, 4 fogs com $\frac{1}{5}$ da capacidade da nuvem
Configuração do RRH	20 MHz, 2x2 MIMO
Custo da Cloud e do Fog	600,300 Watts
Custo de Line Card	20 Watts
Custo da BBU dedicada	$\approx 100$ watts
Custo de OLT	100 Watts
Propagação Óptica	$\approx 4,9 \mu s$ (per Km)
Bandwidth da Divisão E, I D e B no UL	1966, 674.4, 119, 74 Mbps

de processamento ( $n = \{1, 2, 3, 5\}$ ), sendo uma nuvem e quatro nós de fog. A topologia simulada é a mesma da Figura 1.

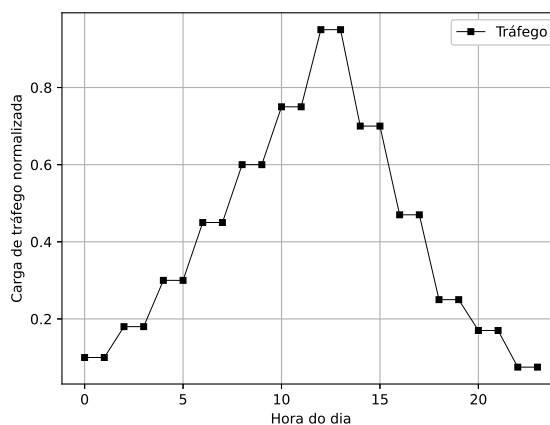
A Figura 2 mostra como a divisão das funções de processamento são distribuídas, dada a disponibilidade de processamento de largura de banda. Além disso, mostra a relação direta entre o maior consumo de energia com a menor disponibilidade de largura de banda disponível na nuvem ao começar a colocar as funções de processamento na *fog*. A figura também mostra a distribuição de funções conforme a seleção da divisão funcional é alterada para contornar restrições no *fronthaul*. Observa-se que a divisão A apresenta um comportamento semelhante ao da *Distributed*-RAN, em que todas as funções de processamento são colocadas de forma distribuída. Enquanto isso, a divisão E representa a função mais centralizada, correspondendo à proposta de arquitetura C-RAN. Observa-se que quando há o correto dimensionamento do *fronthaul* para as demandas, visto pela disponibilidade para processar banda base vigente, todas as funções são processadas em nuvem. Caso não haja disponibilidade de largura de banda para a nuvem, todas as funções são alocadas para processamento em nós de *fog*.

A Figura 3 (a) (b) mostra a relação da largura de banda com a divisão funcional, além da razão de crescimento da largura de banda exigida pelo *fronthaul* conforme a carga de rede gerada aumenta. Note que a Figura 3 (a) apresenta o comportamento esperado do menor consumo de energia ao centralizar as funções na nuvem, enquanto que o consumo de energia aumenta à medida que opções mais distribuídas são utilizadas. Porém, a Figura 3 (b) deixa clara a relação de menor consumo de energia com a largura de



**Figura 3. Avaliação do consumo de energia e da largura de banda por opção de divisão funcional**

banda necessária. A centralização reduz a energia, mas a largura de banda exigida é muito superior às demais divisões de processamento, o que explicita um claro *trade-off*. Esse *trade-off* está relacionado com a disponibilidade de largura de banda e a centralização do processamento. Isso implica que uma maior disponibilidade de largura de banda viabiliza a centralização completa de funções, promovendo uma redução considerável do gasto com energia. No entanto, esta mesma redução emerge com a necessidade de uma rede de acesso mais onerosa para atingir grandes capacidades de *fronthaul*

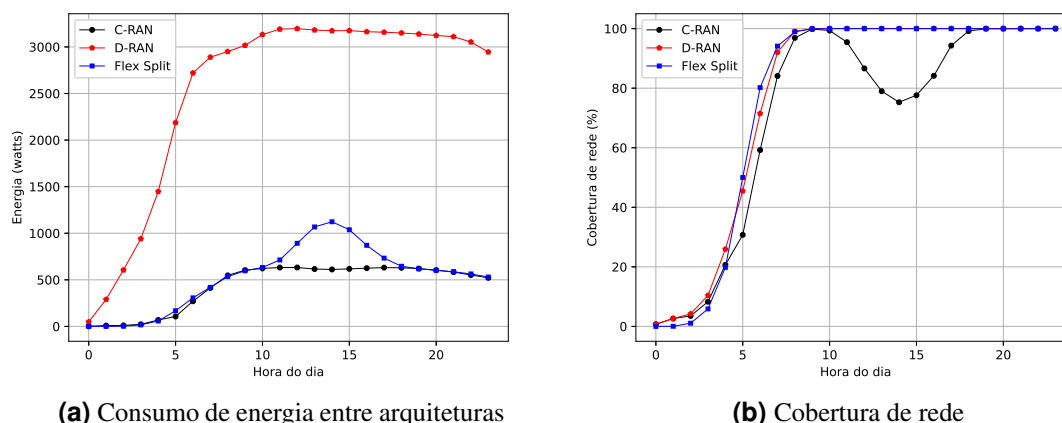


**Figura 4. Carga da rede**

A Figura 4 mostra o comportamento de um dia operacional típico em redes de acesso em uma região de negócios. Nela, as requisições geradas chegam seguindo um processo de *Poisson*, cuja média é igual ao *erlang* da hora atual de operação, e têm um tempo de serviço retirado de uma distribuição exponencial negativa. Este comportamento de tráfego de rede de acesso comercial segue o padrão detalhado em [Peng et al. 2011, Tinini et al. 2020]. Os resultados mostram valores médios adquiridos em 40 execuções

com um nível de confiança de 95%.

Os resultados apresentados na Figura 5 (a) (b) correspondem ao *trade-off* entre a redução do consumo de energia e a cobertura da rede em três arquiteturas. Em relação ao consumo de energia, C-RAN e *Distributed*-RAN se comportam de forma diferente. Neles, o consumo de energia apresenta diferentes destaques. O C-RAN tem um valor consideravelmente menor de consumo de energia em relação aos demais, enquanto o *Distributed*-RAN apresenta um valor superior. Por outro lado, a arquitetura com particionamento flexível apresenta um melhor equilíbrio se comparada as outras duas. Observa-se que a cobertura da rede, que está diretamente relacionada ao bloqueio de requisições vigentes neste cenário, apresenta-se onerosa mesmo com o baixo consumo de energia do C-RAN pois apresentou perda de mais de 27% da cobertura da rede. Enquanto isso, o *Distributed*-RAN e a arquitetura de particionamento flexível permitem 100% de cobertura de rede. O *trade-off* entre as arquiteturas e as métricas testadas no cenário mostra que a arquitetura com particionamento flexível de funções é vantajosa ao balancear a cobertura de rede com um melhor consumo de energia e largura de banda.



(a) Consumo de energia entre arquiteturas

(b) Cobertura de rede

Figura 5. Resultados ilustrativos

Os resultados mostram que à medida que a capacidade do *fronthaul* é expandida para processamento de tráfego, as funções são alocadas em nuvem para reduzir o consumo de energia, que está diretamente relacionado à função objetivo e mantém os resultados do cenário anterior. Em caso de baixa disponibilidade, as funções são distribuídas para os nós de *fog* por indisponibilidade de recursos. Em um cenário com capacidades de processamento limitadas, é mais vantajoso usar arquiteturas híbridas. O uso de particionamento flexível de funções de processamento pode otimizar ainda mais a relação entre largura de banda do *fronthaul* e a eficiência energética.

## 6. Conclusão e trabalhos futuros

Neste trabalho, é apresentado uma solução matemática utilizando formulação PLI para auxiliar na decisão de qual opção de divisão flexível de funções para cada RRH na rede com objetivo de minimizar o custo energético ao mesmo tempo que centraliza ao máximo as funções no BBU *pool*. Resultados apresentados mostram que há uma forte relação entre consumo de energia, necessidade de alocação de largura de banda e centralização de

funções em uma arquitetura híbrida. Dessa forma, os ganhos de energia devem ser balanceados com os requisitos de largura de banda e atraso. Para trabalhos futuros, consideramos o problema de acesso conjunto, o uso de fatiamento de redes e o dimensionamento utilizando rede packetizada ciente de latência. Ademais, utilizaremos soluções baseadas em meta-heurísticas e aprendizado de máquina.

## Agradecimentos

Esta pesquisa é parte do INCT da Internet do Futuro para Cidades Inteligentes, financiado por CNPq (proc. 465446/2014-0), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e FAPESP (procs. 14/50937-1 e 15/24485-9). Também é parte do projeto FAPESP proc. 18/22979-2.

## Referências

- 3GPP (2017a). Study on cu-du lower layer split for nr. Technical report (TR) TS 36.421, 3rd Generation Partnership Project (3GPP). Version 15.0.0.
- 3GPP (2017b). Study on new radio access technology: Radio access architecture and interfaces. Technical report 38.801, 3rd Generation Partnership Project (3GPP). Version 14.0.0.
- Alimi, I. A., Teixeira, A. L., and Monteiro, P. P. (2018). Toward an efficient c-ran optical fronthaul for the future networks: A tutorial on technologies, requirements, challenges, and solutions. *IEEE Communications Surveys Tutorials*, 20(1):708–769.
- Consortium, G. et al. (2017). Study on new radio access technology: Radio access architecture and interfaces. Technical report, Technical Report TR-38.801 Release.
- Figueiredo, G. B., Wang, X., Meixner, C. C., Tornatore, M., and Mukherjee, B. (2016). Load balancing and latency reduction in multi-user comp over twdm-vpns. In *IEEE Intl. Conf. on Communications (ICC)*.
- Gao, Z., Zhang, J., Yan, S., Xiao, Y., Simeonidou, D., and Ji, Y. (2019). Deep reinforcement learning for bbu placement and routing in c-ran. In *Optical Fiber Communications*.
- Guizani, Z. and Hamdi, N. (2017). Cran, h-cran, and f-ran for 5g systems: Key capabilities and recent advances. *International Journal of Network Management*, 27(5):e1973. e1973 nem.1973.
- Harutyunyan, D. and Riggio, R. (2018). Flex5g: Flexible functional split in 5g networks. *IEEE Trans. on Network and Service Management*, pages 961–975.
- Konstantinou, D., Bressner, T. A., Rommel, S., Johannsen, U., Johansson, M. N., Ivashina, M. V., Smolders, A. B., and Tafur Monroy, I. (2020). 5g ran architecture based on analog radio-over-fiber fronthaul over udwdm-pon and phased array fed reflector antennas. *Optics Communications*, 454:124464.
- Larsen, L. M. P., Checko, A., and Christiansen, H. L. (2019). A survey of the functional splits proposed for 5g mobile crosshaul networks. *IEEE Communications Surveys Tutorials*, 21(1):146–172.

- Marotta, A., Cassioli, D., Kondepu, K., Antonelli, C., and Valcarenghi, L. (2019). Exploiting flexible functional split in converged software defined access networks. *IEEE/OSA Journal of Optical Communications and Networking*, 11(11):536–546.
- Mei, H. and Peng, L. (2020). Flexible functional split for cost-efficient c-ran. *Computer Communications*, 161:368–374.
- Moreira Zorello, L. M., Troia, S., Quagliotti, M., and Maier, G. (2020). Power-aware optimization of baseband-function placement in cloud radio access networks. In *2020 International Conference on Optical Network Design and Modeling (ONDM)*, pages 1–6.
- Pelekanou, A., Anastasopoulos, M., Tzanakaki, A., and Simeonidou, D. (2018). Provisioning of 5g services employing machine learning techniques. In *International Conference on Optical Network Design and Modeling (ONDM)*, pages 200–205.
- Peng, C., Lee, S.-B., Lu, S., Luo, H., and Li, H. (2011). Traffic-driven power saving in operational 3g cellular networks. In *Proceedings of the 17th annual international conference on Mobile computing and networking*, pages 121–132.
- Ranaweera, C., Wong, E., Nirmalathas, A., Jayasundara, C., and Lim, C. (2018). 5g c-ran with optical fronthaul: An analysis from a deployment perspective. *Journal of Lightwave Technology*, 36(11):2059–2068.
- Riva, M., Donâncio, H., Almeida, F. R., Figueiredo, G. B., Tinini, R. I., Cesar Jr, R. M., and Batista, D. M. (2018). An elastic optical network-based architecture for the 5g fronthaul. In *Anais do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*. SBC.
- Tinini, R. I., Batista, D. M., Figueiredo, G. B., Tornatore, M., and Mukherjee, B. (2019). Energy-efficient baseband processing via vbbu migration in virtualized cloud-fog ran. In *IEEE GLOBECOM*.
- Tinini, R. I., Batista, D. M., Figueiredo, G. B., Tornatore, M., and Mukherjee, B. (2019). Low-latency and energy-efficient bbu placement and vpon formation in virtualized cloud-fog ran. *IEEE/OSA Journal of Optical Communications and Networking*, 11(4):B37–B48.
- Tinini, R. I., dos Santos, M. R. P., Figueiredo, G. B., and Batista, D. M. (2020). 5GPy: A SimPy-based simulator for performance evaluations in 5G hybrid Cloud-Fog RAN architectures. *Simulation Modelling Practice and Theory*, 101:102030.
- Tohidi, M., Bakhshi, H., and Parsaeefard, S. (2020). Flexible function splitting and resource allocation in c-ran for delay critical applications. *IEEE Access*, 8:26150–26161.
- Wang, X., Alabbasi, A., and Cavdar, C. (2017). Interplay of energy and bandwidth consumption in cran with optimal function split. In *IEEE Intl. Conf. on Communications*.
- Wu, J., Zhang, Z., Hong, Y., and Wen, Y. (2015). Cloud radio access network (c-ran): a primer. *IEEE Network*, 29(1):35–41.