

Posicionamento quase ideal de vBBU e atribuição de comprimento de onda em Cloud Fog RAN

Matias R. P. dos Santos¹, Rodrigo I. Tinini², Tiago Januario¹, Gustavo B. Figueiredo¹

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

²Universidade de São Paulo – São Paulo (USP)

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

Abstract. *In Cloud fog Radio Access Networks (CF-RAN), optimal dynamic activation of fog nodes and virtual BaseBand Units (vBBU) placement is typically performed by Integer Linear Programming (ILP) formulations. However, ILP presents scalability problems and compromises its application in larger network scenarios. So, in this paper we propose a linear relaxation approach to work around this scalability problem and provide a near optimal solution for dynamic vBBU placement in a timely manner in CF-RAN. The results show that a heuristic approach based on linear relaxation significantly reduces the use of computational resources and the execution time compared to ILP while providing sub-optimal solutions that approximated ILP solutions.*

Resumo. *Na arquitetura Cloud fog Radio Access Network (CF-RAN), a ativação ótima e dinâmica de fog nodes e o posicionamento de unidades de banda base virtuais (virtual BaseBand Unit - vBBU) são realizados por meio de propostas utilizando programação linear inteira (PLI). Porém, a PLI apresenta problemas de escalabilidade em cenários de redes maiores. Assim, nesse artigo é proposto uma solução baseada em relaxação linear para evitar problemas de escalabilidade e para prover uma solução aproximada da PLI para o posicionamento dinâmico de vBBUs na CF-RAN. Os resultados mostram que a relaxação linear reduz significativamente o tempo de execução e é capaz de prover soluções sub-ótimas muito próximas das providas pela PLI.*

1. Introdução

A arquitetura de Rede de Acesso por Rádio (do inglês *Cloud Radio Access Network* – C-RAN) reduz os custos operacionais e de capital de redes móveis ao centralizar os elementos de processamento de banda base (do inglês *BaseBand Unit* – BBU) em uma instalação de nuvem chamada *BBU pool* [Wu et al. 2015] e implantar nas células da rede antenas de baixo custo, chamadas de *Remote Radio Heads* (RRHs) responsáveis por agregar os sinais de rádio dos usuários. No entanto, mesmo que forneça economia de energia devido à centralização proposta, algumas desvantagens emergem devido às altas demandas de tráfego, restrições de atraso fim-a-fim e cobertura de rede limitada [Figueiredo et al. 2016].

Para lidar com as restrições da rede de transporte que conecta os BBUs aos RRHs, isto é, o *fronthaul*, a arquitetura *Cloud-Fog RAN* (CF-RAN) foi proposta [Tinini et al. 2019]. CF-RAN faz uso dos paradigmas de Virtualização de Função de

Rede (do inglês *Network Functions Virtualization* – NFV) e de computação em névoa (*fog computing*) para disponibilizar processamento de banda base local caso o *pool* de BBU ou *fronthaul* esteja com altas cargas de trabalho [Tinini et al. 2020]. Na CF-RAN, os nós de processamento de *fog* são dispostos e implantados próximos aos RRHs, e os BBUs virtualizados (vBBUs), que recebem e processam os sinais de banda base, são ativados dinamicamente em função da demanda decorrente da rede. Porém, a ativação dos nós de *fog* aumenta o consumo de energia, reduzindo os ganhos provenientes da C-RAN. Isso torna fundamental que os nós de *fog* sejam ativados apenas quando realmente necessário. Caso todo o tráfego proveniente das demandas de RRHs seja adequadamente processado no *pool* de BBU, nós de *fog* devem permanecer desativados. Porém, o padrão de tráfego nas redes móveis é altamente dinâmico e, assim, a ativação coerente dos nós de *fog* e o posicionamento de vBBUs torna-se crucial para evitar desperdício de recursos.

O *fronthaul* do CF-RAN é implementado em uma Rede óptica Passiva Multiplexada por Divisão de Comprimento de Onda e Tempo (*Time Wavelength Division Multiplexing PON* – TWDM-PON), sendo assim necessário alocar um canal óptico virtual dedicado, conhecido como PON Virtual (VPON), para que os grupos de vBBUs ativos possam permitir a transmissão dos RRHs para seus vBBUs correspondentes. O problema relacionado ao posicionamento de funções de processamento e de atribuição de comprimento de onda emerge da necessidade de melhor dimensionar a rede e evitar falhas ou desperdícios de recursos.

A literatura apresenta como solução ótima para esse problema o uso de formulações matemáticas utilizando Programação Linear Inteira (PLI) [Schweissguth et al. 2020, Tinini et al. 2020, Wang et al. 2017]. Porém, não há algoritmos conhecidos que resolvam formulações de PLI em tempo polinomial [Wolsey 1998]. Relacionado a esse problema de escalabilidade, o uso de relaxação linear pode ajudar a resolver modelos de PLI em tempo polinomial com possível controle na perda da otimalidade [Noor-E-Alam and Doucette 2012]. Assim, neste trabalho, é apresentado uma formulação de relaxação linear para resolver o problema de colocação de vBBU e atribuição de VPONs no CF-RAN. O objetivo é apresentar uma solução válida e escalável que apresente resultados próximos ao PLI com redução no tempo de execução. A abordagem proposta visa identificar o melhor posicionamento de vBBU com quantidades mínimas de VPONs atribuídas com base no estado da rede, recursos disponíveis e restrições de capacidade do *fronthaul*. Em suma, este artigo atinge os seguintes objetivos e apresenta as seguintes contribuições: i) apresenta a formulação de uma relaxação linear para o problema de colocação de vBBUs e atribuição de comprimento de ondas; ii) aloca e ativa recursos de rede para suportar o tráfego objetivando a redução do consumo de energia; iii) apresenta uma solução escalável com resultados muito próximos ao do PLI. Para avaliação do desempenho da solução proposta e comparação com o PLI, simulações estáticas e dinâmicas de tráfego são utilizadas considerando métricas como a cobertura da rede alcançada, o consumo de energia, a intensidade utilizada de CPU, a média do tempo do *solver* e a ativação de VPONs e nós de processamento para cada hora do dia.

Os resultados mostram ganhos nos tempos de execução e mantêm os resultados ótimos gerados pela PLI em 21 das 24 horas testadas nos experimentos computacionais. Nossa formulação proposta reduz o tempo de execução em cerca de ($\approx 98.03\%$), reduz a

intensidade da CPU em até (45.4%), enquanto mantém o *gap* percentual entre relaxação linear e PLI em (0%), e o *gap* de integralidade (IG) em (1) em mais de 90% nos experimentos computacionais.

O restante deste trabalho da seguinte forma: a Seção 2 apresenta alguns trabalhos relacionados concentrando-se na migração de processamento, ativação ou desativação de elementos de rede e posicionamento de BBU virtual; a Seção 3 apresenta a arquitetura CF-RAN e suas funcionalidades; a Seção 4 apresenta o problema e a política para ativação e desativação de vBBUs; a Seção 5 apresenta os resultados obtidos pelos experimentos; a Seção 6 conclui o artigo e fornece orientações futuras de pesquisa.

2. Trabalhos Relacionados

Alguns trabalhos propuseram o uso de formulação PLI para solucionar o gerenciamento de recursos no C-RAN em condições de tráfego dinâmico [Figueiredo et al. 2016, Mohammed Mikaeil et al. 2019]. Ambos os trabalhos buscam solucionar o posicionamento ótimo de vBBUs no C-RAN, objetivando alcançar o melhor provisionamento de recursos no *pool* de BBU para maximizar a cobertura enquanto minimiza as restrições ópticas do *fronthaul*. No entanto, esses trabalhos não consideraram arquiteturas híbridas nem investigaram os benefícios do uso de relaxação linear como alternativa para permitir a escalabilidade da rede. No trabalho de [Tinini et al. 2020] foi utilizada uma arquitetura baseada em *fog* para reduzir a latência, largura de banda e alocação de recursos. Além disso, o trabalho apresenta a PLI para resolver o problema de forma otimizada, mas há uma séria limitação de aplicabilidade da formulação para rede de pequeno porte decorrente do problema da PLI em si.

No que tange o uso da arquitetura híbrida para solucionar as desvantagens do C-RAN, autores em [Nassar and Yilmaz 2019, Katti and Prince 2019] apresentam conceitos e diversas soluções para contornar limitações do C-RAN baseadas em heurísticas e em formulação matemática PLI. Autores em [Gkatzios et al. 2019] apresentaram conceitos de desagregação de recursos de rede em uma RAN virtualizada centralizada. Eles propuseram um esquema de alocação para funções de processamento em vários servidores de acordo com a carga de tráfego da rede. Este trabalho apresenta PLI e abordagens heurísticas para avaliar a proposta.

A literatura aborda diversos trabalhos que tratam do tempo de execução e da escalabilidade da PLI utilizando a relaxação linear. As soluções propostas aplicam diversos tipos de abordagens de relaxação para obter valores próximos ao ótimo em uma infinidade de aplicação de redes e de logística [Noor-E-Alam and Doucette 2012, Zaky Kasem et al. 2012, Baruah et al. 2019]. Essas soluções mostram a busca por alternativas ou outras abordagens para resolver o problema de escalabilidade do PLI sem grande disparidade em comparação ao resultado ótimo.

Diferente dos trabalhos acima mencionados, neste artigo, é utilizado uma relaxação linear que alcança resultados próximos de uma solução PLI ótima, utilizando o modelo PLI proposto em [Tinini et al. 2019], para resolver o problema de posicionamento de vBBU e atribuição de vPONS em CF-RAN. Também realizamos uma comparação com outra heurística de relaxação proposta em [Tinini et al. 2020].

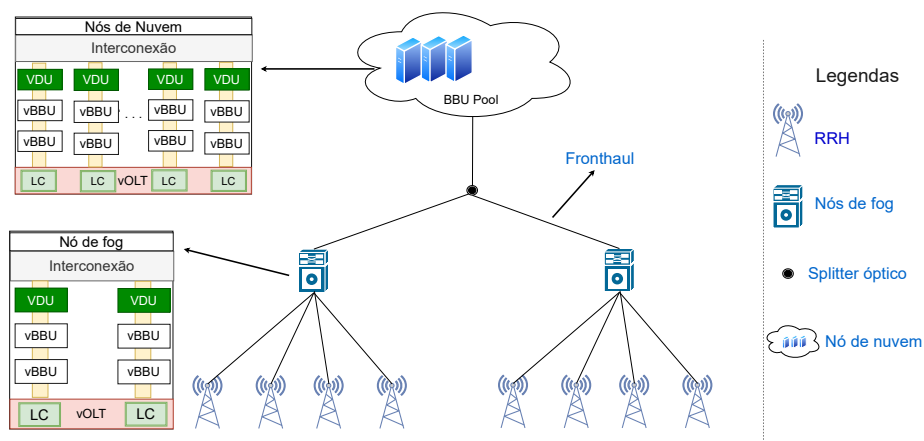


Figura 1. Arquitetura do CF-RAN

3. Arquitetura e funcionamento do CF-RAN

A Figura 1 apresenta a arquitetura CF-RAN. Ela é composta por nós de computação em nuvem e de *fog*. Elas formam a camada de processamento de banda base. A camada de nuvem é composta por um único nó de processamento que centraliza o processamento de banda base e possui grande capacidade de processamento. A outra camada de processamento é composta por nós de *fog* que são posicionados próximos aos usuários.

Tanto na nuvem quanto nos nós de *fog*, um conjunto de unidades digitais virtuais (do inglês *Virtual Digital Unit* – VDUs), ou seja, contêineres de processamento de banda base, implementa um conjunto de funções de processamento virtualizado (do inglês *Virtualized processing functions* – VPFs). Um desses VPFs é o vBBU, que recebe e processa os sinais de banda base de um RRH específico. Assim, para cada RRH que transmite sinais de banda base usando o protocolo CPRI (do inglês *Common Public Radio Interface*), um VDU e um vBBU precisam ser ativados em algum nó para realizar o processamento de banda base para o RRH.

O *fronthaul* CF-RAN é implementado sobre um TWDM-PON que fornece baixo consumo de energia e baixa latência. Uma Unidade de Rede Óptica (*Optical network unit* – ONU) conecta cada RRH, e cada nó de processamento implementa um Terminal de Linha óptica Virtual (do inglês *Virtualized Optical line transport* – vOLT). Lasers sintonizáveis equipam as ONUs e são responsáveis por realizar transmissões em qualquer comprimento de onda concedido pela vOLT. A vOLT implementa um conjunto de *Line Cards* (LC), que são transceptores responsáveis por escalonar comprimento de onda específico para as ONUs. Os VDUs conectam cada LC e o tráfego encaminhado de um LC para seu VDU relacionado. Ele usa um *switch* interno para alternar o tráfego entre VDUs quando necessário. Assumimos o uso de canais VPONs em que cada ONU pode sintonizar qualquer um dos comprimentos de onda disponíveis, de modo que várias ONUs podem compartilhar o mesmo canal óptico em uma maneira de Multiplexação por Divisão de Tempo (do inglês *Time Division Multiplexing* – TDM) para transmitir para um nó de processamento comum. Os VPONs também podem ser ativados/criados dinamicamente para oferecer suporte a demandas de rede específicas nesse sentido.

Com relação à operação de CF-RAN, para um determinado conjunto de RRHs

Tabela 1. Parâmetros de entrada e variáveis de decisão do ILP.

Símbolo	Definições
Conjuntos	
$i \in R$	conjunto de demandas de tráfego RRH
$n \in N$	conjuntos de nós de processamento
$w \in W$	conjuntos de canais de VPONs disponíveis
Parâmetros de Entrada	
B_i	demanda de largura de banda de RRH i
B_w	capacidade da VPON w
B_{en}	largura de banda do backplane switch e no nó n
P_i	demanda de processamento do RRH i
P_n	capacidade de processamento do nó n
I_w	capacidade de processamento do VDU w
M	um número muito grande
C_n	custo energético do nó n
C_{lc}	custo energético do LC
C_{vdu}	custo energético do VDU em cada nó n
C_e	custo energético do backplane switch e
Variáveis de decisão	
x_{wn}^i	1 se a demanda do RRH i é processada no nó n transmitido pela VPON w , 0 caso contrário.
u_{wn}^i	1 se a demanda do RRH i é processado pelo VDU w no nó n , 0 caso contrário.
z_{wn}	1 se o VPON w está alocado no nó n , 0 caso contrário.
t_n	1 se funções de processamento e elementos do nó n estão ativados, 0 caso contrário.
y_{in}	1 se a demanda do RRH i foi alocado para o nó n , 0 caso contrário.
k_{in}	1 se a demanda do RRH i foi direcionado para o VDU w no nó n , 0 caso contrário.
s_{wn}	1 se o VDU w está ativo no nó n , 0 caso contrário.
e_n	1 se o backplane switch e está ativo no nó n , 0 caso contrário.

ativos, um número adequado de vBBUs deve ser colocado na rede e os VPONs precisam ser alocados para promover as transmissões de RRHs para os vBBUs colocados. Uma decisão importante é onde colocar vBBUs, ou seja, na nuvem ou em nós de *fog*, para promover a eficiência energética. Na próxima seção, nós apresentamos o problema e a relaxação linear e a formulação PLI utilizada para resolver o problema de colocação de vBBU e atribuição de VPON no CF-RAN.

Apresentamos na próxima seção a formulação do problema e a proposta de solução do posicionamento de vBBU e atribuição de VPONs no CF-RAN.

4. Definição do Problema

Dado um conjunto de RRHs R gerando tráfego CPRI, um conjunto de nós de processamento N , um conjunto de canais PON W e um conjunto de VDUs com a mesma cardinalidade do conjunto W em cada nó n , nós colocamos cada vBBU de cada demanda em R na menor quantidade de nós de processamento dentro da menor quantidade de VDUs usando a menor quantidade de canais PON para suportar suas transmissões.

4.1. Formulação do PLI

A seguir, é apresentada a formulação PLI utilizada com o objetivo de fornecer a solução ótima para o problema estudado. Foi utilizado a seguinte formulação matemática como base para nosso algoritmo para propor a relaxação linear. Os parâmetros de entrada, assim como as variáveis de decisão, são apresentados na Tabela 1.

Função Objetivo

$$\begin{aligned}
\text{Minimize } & \sum_{n=1}^N x_n \times C_n + C_{lc} \times \sum_{w=1}^W \sum_{n=1}^N z_{wn} + \\
& C_{vdu} \times \sum_{w=1}^W \sum_{n=1}^N s_{wn} + C_e \times \sum_{n=1}^N e_n
\end{aligned} \tag{1}$$

Restrições

$$\sum_{w=1}^W \sum_{n=1}^N x_{wn}^i = 1, \forall i \in R \tag{2}$$

$$\sum_{w=1}^W \sum_{n=1}^N u_{wn}^i = 1, \forall i \in R \tag{3}$$

$$\sum_{i=1}^R u_{wn}^i \geq 0, \forall w \in W, \forall n \in N \tag{4}$$

$$\sum_{n=1}^N y_{in} = 1, \forall i \in R \tag{5}$$

$$y_{in} \leq k_{in}, \forall i \in R, \forall n \in N \tag{6}$$

$$\sum_{n=1}^N z_{wn} \leq 1, \forall w \in W \tag{7}$$

$$z_{wn} \leq s_{wn}, \forall w \in W, \forall n \in N \tag{8}$$

$$\sum_{i=1}^R \sum_{n=1}^N (x_{wn}^i \times B_i) \leq B_w, \forall w \in W \tag{9}$$

$$\sum_{i=1}^R \sum_{w=1}^W (x_{wn}^i \times P_i) \leq P_n, \forall n \in N \tag{10}$$

$$M \times t_n \geq \sum_{i=1}^R \sum_{w=1}^W x_{wn}^i, \forall n \in N \tag{11}$$

$$t_n \leq \sum_{i=1}^R \sum_{w=1}^W x_{wn}^i, \forall n \in N \tag{12}$$

$$M \times z_{wn} \geq \sum_{i=1}^R x_{wn}^i, \forall w \in W, \forall n \in N \tag{13}$$

$$z_{wn} \leq \sum_{i=1}^R x_{wn}^i, \forall n \in N, \forall w \in W \tag{14}$$

$$M \times e_n \geq \sum_{i=1}^R k_{in}, \forall n \in N \quad (15)$$

$$\sum_{i=1}^R \sum_{n=1}^N u_{wn}^i \geq I_w, \forall w \in W \quad (16)$$

$$M \times y_{in} \geq \sum_{w=1}^W x_{wn}^i, \forall n \in N, \forall i \in R \quad (17)$$

$$y_{in} \leq \sum_{w=1}^W x_{wn}^i, \forall n \in N, \forall i \in R \quad (18)$$

$$M \times y_{in} \geq \sum_{w=1}^W u_{wn}^i, \forall n \in N, \forall i \in R \quad (19)$$

$$y_{in} \leq \sum_{w=1}^W u_{wn}^i, \forall n \in N, \forall i \in R \quad (20)$$

$$M \times s_{wn} \geq \sum_{i=1}^R u_{wn}^i, \forall w \in W, \forall n \in N \quad (21)$$

$$s_{wn} \leq \sum_{i=1}^R u_{wn}^i, \forall w \in W, \forall n \in N \quad (22)$$

$$\sum_{i=1}^R k_{in} \times B_i \geq B_{en}, \forall n \in N \quad (23)$$

$$e_n \leq \sum_{i=1}^R k_{in}, \forall n \in N \quad (24)$$

A função objetivo (1) visa reduzir o consumo de energia ativando a menor quantidade de recursos de rede colocando vBBUs ao máximo na *Cloud*. As restrições (2)-(5) garantem que o vBBU para cada demanda i seja colocado em apenas um VDU, em apenas um nó de processamento e use apenas um VPON. A restrição (6) aponta para o conjunto de nós que a VPON pode ser atribuído. A restrição (7) garante que um VPON seja alocado para no máximo um nó de processamento (*fog* ou nuvem). A restrição (8) garante que uma demanda i seja alocada apenas para um nó de processamento fisicamente conectado a ela. Restrições (9)-(15) garante que a capacidade dos nós de processamento, VPONs e switch backplane serão respeitados. As restrições restantes impõem a ativação da chave do painel traseiro em cada nó e a ativação de VDUs adicionais.

4.2. Relaxação Linear

A relaxação linear é uma alternativa viável para resolver problemas de escalabilidade do PLI. A relaxação lagrangiana e a relaxação linear são algumas das abordagens comumente usadas para este propósito, mesmo utilizando abordagens distintas [Noor-E-Alam and Doucette 2012]. Além disso, alguns dos *solvers* mais populares fazem uso de diferentes algoritmos para resolver problemas de PLI usando relaxação para gerar limites inferiores para identificar o tamanho da árvore de busca e alguns outros parâmetros para representar os resultados ideais [Morrison et al. 2016]. Os benefícios da aplicação da relaxação linear está associado ao uso de variáveis contínuas que restringe a exigência de integralidade das variáveis de decisão binárias e apresenta soluções com limite inferior dependendo do tipo de problema de otimização. Portanto, resolver o problema de colocar vBBU e atribuir VPONs para CF-RAN usando relaxação linear se mostra praticável, pois permite dimensionar a rede para instâncias maiores. No entanto, a relaxação linear por si só pode não garantir soluções viáveis e ótimas uma vez que valores não inteiros representando nós de processamento, por exemplo, não são praticáveis. Dessa forma, é necessário o arredondamento de valores não inteiros e a validação dos resultados gerados.

Pode-se obter uma solução integral ótima ou quase ótima simplesmente arredondando os componentes fracionados de uma solução do LP para inteiros, para cima ou para baixo, ou para o número inteiro mais próximo [Matoušek and Gärtner 2007]. Neste artigo, foi utilizado o modelo PLI e o “relaxamos” para uma relaxação linear removendo as restrições de integralidade, resolvemos o programa linear em tempo aproximado polinomial e, em seguida, “arredondamos” a solução relaxada para uma solução inteira. Após extensa análise para identificar a melhor configuração das variáveis, as seguintes foram escolhidas para relaxamento:

$$0 \leq u_{wn}^i \leq 1, \forall i, n, w \in R, N, W \quad (25)$$

$$0 \leq y_{in} \leq 1, \forall i, n \in R, N \quad (26)$$

$$0 \leq k_{in} \leq 1, \forall i, n \in R, N \quad (27)$$

$$t_n \in \{0\} \cup [1, \infty], \forall n \in N \quad (28)$$

$$s_{wn} \in \{0\} \cup [1, \infty], \forall n, w \in N, W \quad (29)$$

$$z_{wn} \in \{0\} \cup [1, \infty], \forall n, w \in N, W \quad (30)$$

Foi retirada a restrição de integralidade das variáveis t_n , s_{wn} e z_{wn} de binárias para semicontínuas, e y_{in} , k_{in} e u_{wn}^i de binárias para contínuas. Ressalta-se que a solução para relaxação linear não é necessariamente integral. No entanto, uma vez que a região viável

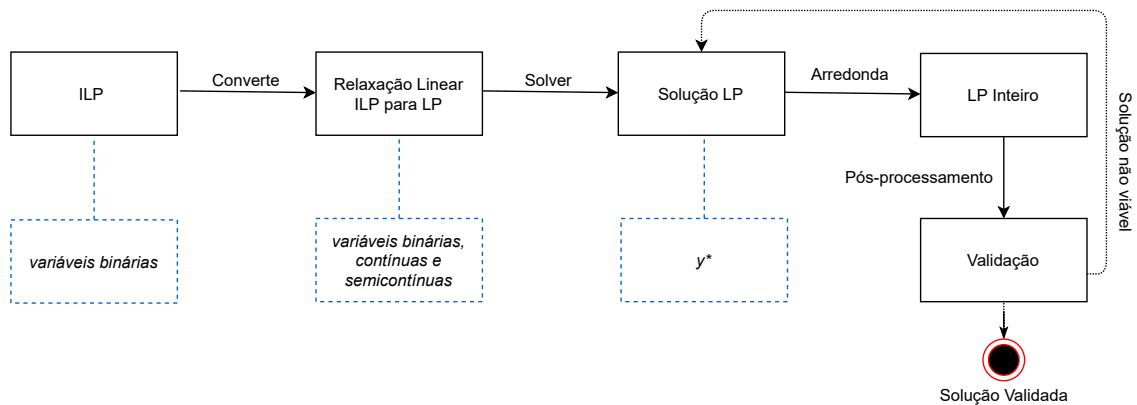


Figura 2. Etapas para resolver e validar a relaxação linear

da relaxação linear é maior do que a região viável do problema original, o valor ótimo do primeiro não é pior do que o valor ótimo do último. Cabe destacar que as variáveis semicontínuas possuem limite inferior em 1, isso significa que elas podem assumir valores igual a 0 ou entre o limite inferior e superior $(1, \infty)$.

O processo para geração de uma solução heurística aproximada segue os passos descritos na Figura 2 e pode ser detalhado da seguinte forma: (1) relaxa-se a restrição de integralidade das variáveis selecionadas; (2) resolve o problema linear resultante da etapa anterior, que em algumas rodadas foi fracionado; (3) a solução fracionária passa por uma função de pós-processamento para arredondamento das variáveis fracionadas utilizando arredondamento randomizado e validação da solução gerada. A otimalidade após este processo não é garantida, mas uma solução com aproximação de tempo polinomial é garantida.

Nas Figura 3 e 4, as soluções obtidas com variáveis inteiras são denotadas por PLI, e as soluções obtidas com o método “*relax-and-round*” são denotados por relaxação LP. Também consideramos como *baseline* outro algoritmo de pós-processamento associado ao método “*relax-and-round*” [Tinini et al. 2020] (LP RR) para comparação em teste de cenários dinâmicos. A partir de nossos resultados, descobrimos que a lacuna de integralidade entre a relaxação linear e a solução PLI é em mais de 90% da vezes igual 1, portanto, o limite inferior fornecido pela relaxação linear é igual ao limite ideal fornecido pelo PLI, enquanto os tempos de execução do relaxamento linear são menores.

5. Resultados Numéricos

Nesta seção, apresentamos os resultados na avaliação de dois tipos diferentes de cenários de tráfego: estático e dinâmico. O cenário de tráfego estático permite avaliar se o LP se aproxima da solução ótima comparando o consumo de energia e a ativação de elementos de rede em um comportamento de carga de tráfego conhecido e igual a todos as formulações testadas. O cenário de tráfego dinâmico permite a verificação da cobertura da rede, probabilidade de bloqueio, média de consumo de energia e disponibilidade de serviços de rede. Os parâmetros de simulação usados na simulação dinâmica estão apresentados na Tabela 2, apresentando a topologia, a configuração das RRHs consideradas, o protocolo utilizado e os custos considerados de cada componente na arquitetura CF-RAN.

5.1. Cenário de testes estático

Realizamos execuções comparando os resultados da relaxação linear e da PLI para demonstrar a eficiência das soluções adquiridas. Consideramos uma arquitetura CF-RAN com vinte e quatro comprimentos de onda ($w = \{1, 2, \dots, 24\}$) de 10 Gbps, e composta por 1 nuvem e onze nós de fog ($n = \{1, 2, \dots, 12\}$). Para avaliar as formulações, usamos a API DOCPLEX Python com CPLEX 12.10 com configurações padrão. A quantidade de RRHs varia de 10 até 345 com um incremento de 5 RRHs a cada rodada para verificar os resultados adquiridos e o comportamento quanto a escalabilidade.

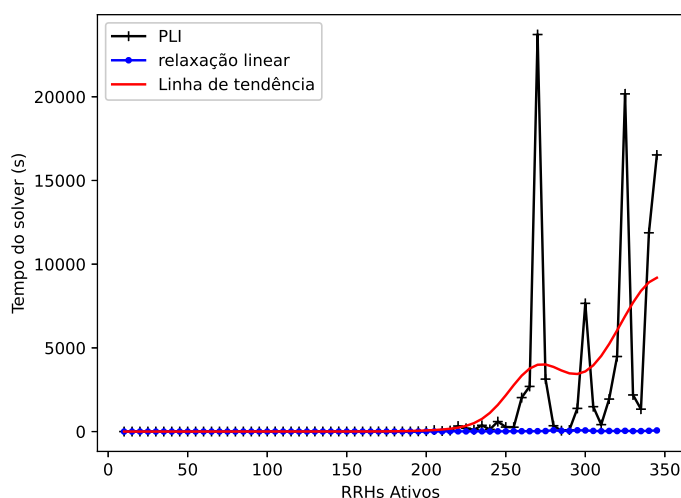


Figura 3. Tempo do *solver* para encontrar uma solução

A Figura 3 mostra a comparação do PLI e da relaxação linear em relação aos tempos de execução do *solver*. Os resultados mostram reduções de cerca de 98.03% no momento que o PLI levou mais tempo para encontrar a solução. Se compararmos o maior tempo de *solver* em ambas as formulações, obtemos uma redução de cerca de 96.32% pela relaxação linear. É traçado uma linha de tendência que mostra o crescimento esperado no tempo do *solver* à medida que as instâncias aumentam. O resultado desta linha de tendência implica em um comportamento de crescimento no tempo de execução que prejudica a escalabilidade da rede ao utilizar a formulação matemática PLI.

Tabela 2. Parâmetros de simulação

Parâmetros	Valores
Topologia	1 cloud, 3 fog
Configuração da RRH	10 MHz, 1x1 MIMO
Comunicação RRH-vBBU	CPRI line 1
Custo de Cloud	600 watts
Custo por nó de Fog	300 watts
Custo por Line Card	20 watts
Custo da vOLT	100 watts
Custo de Switch	15 watts
Custo da VDU da Cloud	100 watts
Custo da VDU da Fog	50 watts

Tabela 3. Comparação da função objetivo, intensidade da CPU e Gap percentual

RRHs	Função Objetivo		Intensidade CPU(%)		Tempo do Solver (s)		Gap PLI - LP(%)
	PLI	LP	PLI	LP	PLI	LP	
10	660	660	23.2	15.1	0.3164	0.3817	0.0
50	720	720	59.5	32.2	1.8134	0.6426	0.0
100	1020	1020	77.8	40.1	6.1563	1.3522	0.0
150	1180	1180	58.6	32.6	4.4714	1.9619	0.0
200	1680	1680	84.9	40.0	17.152	5.3693	0.0
250	3220	3220	90.4	59.4	252.322	12.209	0.0
300	4080	4120	98.7	86.9	8092.916	69.916	0.97
325	4860	4880	99.6	78.7	23079.956	47.956	0.41
345	5180	5300	97.9	82.3	16900.685	69.685	2.26

Em relação aos resultados presentes na Tabela 3, podemos apontar que os ganhos ocorrem não apenas no tempo do *solver*, mas no *gap* percentual, dado como $\frac{\max(PLI,LP) - \min(PLI,LP)}{\max(PLI,LP)} \cdot 100$, da solução final que permaneceu na maior parte do tempo em zero, o que implica que a relaxação linear pode ser usado como uma alternativa viável para PLI uma vez apresentado valores de solução ótima ou muito próxima da ótima. Além disso, a relaxação linear atinge uma redução considerável na intensidade da CPU quando comparado ao PLI. Além disso, a lacuna de integralidade, dada por ($IG = \frac{PLI}{LP}$), é igual a 1 em mais de 96% da execução, o que reforça que a relaxação linear fornece soluções ótimas inteiras. Validamos os resultados alcançados em ambos os programas usando algoritmos de pós-processamento.

Calculamos a intensidade da CPU considerando a quantidade correspondente de consumo de CPU por segundo. Calculamos isso subtraindo o total de CPU disponível e *idle* CPU no segundo t e, em seguida, dividindo a soma da CPU pelo tempo total de execução (T)

$$CPU = \sum_{t=1}^T (CPU_{(t)} - idle_{(t)}) \implies CPU/T.$$

5.2. Cenário Com Tráfego Dinâmico

Usamos um simulador de eventos discretos, 5GPy [Tinini et al. 2020], para avaliar o desempenho da relaxação linear contra o LP RR e a formulação PLI em um cenário de tráfego dinâmico usando a mesma topologia do tráfego do cenário estático, com carga máxima de tráfego ($\epsilon/160$), onde ϵ é o *erlang* de cada hora do dia [Tinini et al. 2020]. As requisições de CPRI chegam seguindo um processo de *Poisson*, cuja média é igual ao *erlang* da hora atual de operação, e tem um tempo de serviço retirado de uma distribuição exponencial negativa. Este comportamento de tráfego segue uma área de rede de acesso comercial seguindo o padrão detalhado em [Peng et al. 2011]. Os resultados mostram valores médios adquiridos em 50 execuções com um nível de confiança de 95%.

A Figura 4 mostra os resultados do desempenho de relaxação linear e LP RR comparadas a formulação PLI. A Figura 4 (a) mostra que a relaxação linear fornece as mesmas soluções ótimas que o PLI para todas as cargas de tráfego durante a operação diária. Enquanto isso, o LP RR apresentou maior consumo de energia do que as outras duas abordagens. Além disso, observe que o tempo de execução, presente na Figura 4

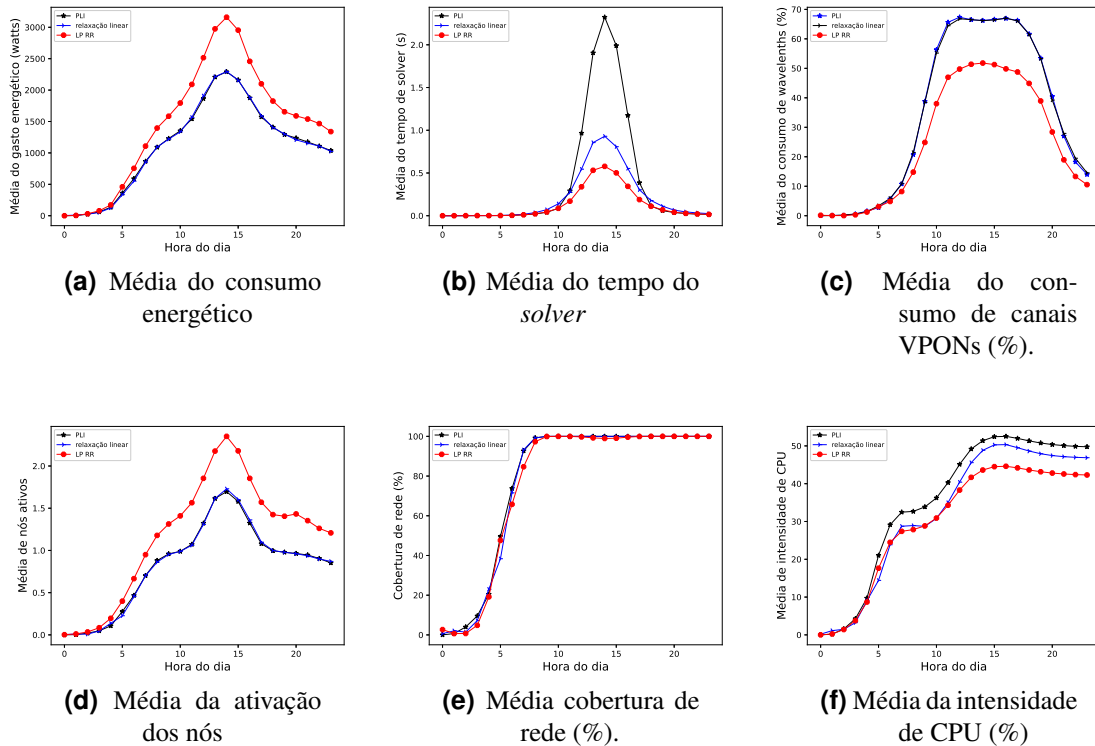


Figura 4. Resultados

(b), é significativamente reduzido pela relaxação linear e LP RR. O resultado mostra que a relaxação linear tem o melhor desempenho, sendo uma alternativa viável à formulação matemática PLI, em que a diferença para LP RR foi estatisticamente significativa, como redução de gasto energético da solução em $\approx 34\%$. Figura 4 (c) (d) mostra que a relaxação linear fornece uma grande aproximação ao PLI em relação ao número de nós de processamento e comprimentos de onda usados para suportar solicitações de CPRI, e isso significa que a relaxação linear e a PLI são estatisticamente iguais. Além disso, os valores adquiridos por LP RR são maiores na ativação do nó e menores no uso de canais VPONs do que os apresentados nas outras duas abordagens. Este resultado está associado à quantidade média de nós e canais VPONs ativados nas execuções. Além disso, esses são maiores do que os outros dois algoritmos. A Figura 4 (e) mostra que a relaxação linear, LP RR e o PLI fornecem resultados ótimos em relação à cobertura da rede. Finalmente, a Figura 4 (f) mostra o uso da intensidade da CPU. Esses resultados reforçam a qualidade adquirida pelo relaxamento proposto, uma vez que apresenta menor consumo que o PLI.

Se compararmos a relaxação linear e LP RR, notamos que a cobertura da rede se aproxima de 100%, e a intensidade da CPU é menor do que as apresentadas em nossa proposta. No entanto, fazendo uma compensação das abordagens, podemos ver que no LP RR esse menor consumo de CPU e tempo de execução veio com maior desperdício de recursos de rede.

Em comparação geral, o LP RR mostrou resultados mais rápidos e menos consumo de CPU. Porém, os resultados promoveram um maior desperdício de recursos, mesmo mantendo a rede ativa e sem bloqueio de requisições durante a simulação. Por

outro lado, o a relaxação linear manteve uma aproximação nos resultados de PLI com menos intensidade de CPU e consumo de tempo, mas não menos que o LP RR. Porém, ao contrário do LP RR, os resultados foram estatisticamente iguais aos do PLI e permitem dimensionar a rede com o mesmo desempenho em um melhor momento.

6. Conclusões e trabalhos futuros

Este estudo apresentou uma formulação matemática de uma PLI e de uma relaxação linear para atribuição de VPONs e posicionamento de vBBU em CF-RAN. A solução proposta foca em resolver o problema mencionado com eficiência energética, e comparamos uma formulações matemática de PLI e uma relaxação linear para verificar a assertividade. Em comparação com a formulação matemática PLI, nossa relaxação linear foi capaz de manter o consumo de energia da solução da arquitetura próximo do ótimo apresentado pelo PLI com menor tempo de execução. Apesar de não garantir a otimalidade, o uso de relaxação permite encontrar configurações viáveis de rede em um sistema de reserva imediata de recursos, algo inviável para a PLI em um cenário de rede grande. Estudos futuros focarão em explorar outras heurísticas, como a baseada em aprendizado de máquina. Além disso, consideramos a aplicação de divisão funcional, diferentes tipos de arquiteturas de rede e o uso de conjuntos de dados para o modelo de rede de 5G.

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

- Baruah, S. K., Bonifaci, V., Bruni, R., and Marchetti-Spaccamela, A. (2019). Ilp models for the allocation of recurrent workloads upon heterogeneous multiprocessors. Journal of Scheduling, 22(2):195–209.
- 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-vpons. In IEEE Intl. Conf. on Communications (ICC).
- Gkatzios, N., Anastasopoulos, M., Tzanakaki, A., and Simeonidou, D. (2019). Efficiency gains in 5g softwarised radio access networks. EURASIP Journal on Wireless Communications and Networking, 2019(1):183.
- Katti, R. and Prince, S. (2019). A survey on role of photonic technologies in 5g communication systems. Photonic Network Communications, 38(2):185–205.
- Matoušek, J. and Gärtner, B. (2007). Integer Programming and LP Relaxation, pages 29–40. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Mohammed Mikaeil, A., Hu, W., and Li, L. (2019). Joint allocation of radio and fronthaul resources in multi-wavelength-enabled c-ran based on reinforcement learning. Journal of Lightwave Technology.

- Morrison, D. R., Jacobson, S. H., Sauppe, J. J., and Sewell, E. C. (2016). Branch-and-bound algorithms: A survey of recent advances in searching, branching, and pruning. Discrete Optimization, 19:79–102.
- Nassar, A. and Yilmaz, Y. (2019). Reinforcement learning for adaptive resource allocation in fog ran for iot with heterogeneous latency requirements. IEEE Access, 7:128014–128025.
- Noor-E-Alam, M. and Doucette, J. (2012). Relax-and-fix decomposition technique for solving large scale grid-based location problems. Computers & Industrial Engineering, 63(4):1062–1073.
- 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.
- Schweissguth, E., Timmermann, D., Parzyjegl, H., Danielis, P., and Mühl, G. (2020). Ilp-based routing and scheduling of multicast realtime traffic in time-sensitive networks. In 2020 IEEE 26th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), pages 1–11.
- 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. (2020). Energy-efficient vbbu migration and wavelength reassignment in cloud-fog ran. IEEE Transactions on Green Communications and Networking, pages 1–1.
- 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.
- 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.
- Wolsey, L. A. (1998). Integer programming, volume 52. John Wiley & Sons.
- Wu, J., Zhang, Z., Hong, Y., and Wen, Y. (2015). Cloud radio access network (c-ran): a primer. IEEE Network, 29(1):35–41.
- Zaky Kasem, A., Doucette, J., et al. (2012). Ilp model and relaxation-based decomposition approach for incremental topology optimization in p-cycle networks. Journal of Computer Networks and Communications, 2012.