

Alocação de Recursos da Interface Aérea 5G a partir de um Critério de Utilidade

Christian Fragoas F. Rodrigues, Lisandro Lovisoló e Luiz da Silva Mello

Resumo—Este artigo discute o desenvolvimento de uma metodologia baseada em utilidade para alocação dos recursos de *downlink* para UEs (*User Equipment*) em uma rede celular. Pontuações de utilidade dos recursos para cada UE são atribuídas de acordo com suas demandas. As utilidades dos UEs são usadas para obter as utilidades dos feixes (*beams*) e das estações base (BSs – *Base Station*), considerando os UEs que eles atendem. Os elementos da RAN são classificados com base nas utilidades para alocar os recursos em ordem de menor esforço. Avaliamos essa proposta e uma versão simplificada de menor complexidade estocasticamente, visando fornecer a mesma capacidade de tráfego para todos os UE conectados à rede. Os resultados mostram a validade da estratégia proposta e que a simplificação não acarreta perda de desempenho.

Palavras-Chave—5G, utilidade dos recursos da RAN, alocação de recursos da RAN.

Abstract—This paper discusses the development of a utility-based methodology for allocating the downlink resources to UEs (User Equipments) in a cellular network. Utility scores for each UE are computed from the resources they demand. The UE's utilities are used to obtain the utilities of the beams and the Base Station (BSs), considering the UEs they serve. These elements are ranked based on the utilities to allocate the resources less effort order. We evaluate the proposal and a simplified version entailing lesser complexity stochastically, aiming at providing the same traffic capacity for every UE connected to the network. The results show the validity of the proposed strategy and that the simplification does not entail performance loss.

Keywords—5G, RAN resource utility, RAN resource allocation.

I. INTRODUÇÃO

Com o advento das redes 5G, são propostos diferentes casos de uso que apresentam demandas de serviços distintas por um número massivo de UEs [1], [2]. O projeto destas se torna cada vez mais complexo, em função do seu tamanho e complexidade, o que requer simulações até o detalhe do enlace de UE e da capacidade resultante [3].

A alocação dos recursos disponíveis em tais redes deve contemplar diversos fatores como ponto de conexão, qualidade do enlace, largura de banda de frequência, sua divisão no tempo e potência. Por exemplo, ela depende da quantidade de UEs admitidos pela rede e de suas associações aos *beams* das diferentes BSs, que resultam em diferentes condições de canal [3], [4]. Consideramos o

caso em que todos os UEs devem ser servidos com uma mesma taxa de dados no *downlink* independentemente de seu ponto de contato e das condições de canal resultantes. Tal cenário se aplica a casos de usos como VoD (*Video on Demand*) e jogos MMO (*Massive Multiplayer Online*).

Para garantir a mesma capacidade (b/s) por UE propomos um critério de menor esforço inspirado em [5]. O critério é computado a partir da demanda por banda e extrapolado para os *beams* e as BSs. A partir do ranqueamento dos *scores* obtidos, faz-se a alocação de banda para os UEs e dos intervalos de tempo ativos dos *beams*.

Usamos $R_{\text{need}}(u)$ para indicar a quantidade de recursos necessários para fornecer a capacidade alvo para o u -ésimo UE e $R_{\text{rsv}}(u)$ para indicar a quantidade de recursos reservados para o mesmo. Seja $\text{CSI}(u)$ o *Channel State Information* do u -ésimo UE, a qualidade de seu enlace [6], ranqueamos os UEs de acordo com a utilidade de atendê-lo

$$Y(u) = \frac{R_{\text{rsv}}(u)}{R_{\text{need}}(u)} f(\text{CSI}(u)). \quad (1)$$

A Eq. 1 é uma utilidade pois avalia individualmente o atendimento da demanda, conforme o critério C_{target} , do UE u a partir de um valor real. Na Eq. (1), $f(\cdot)$ é não monotonicamente crescente em função do CSI, já que quanto maior a qualidade do canal, mais fácil o atendimento. Faz-se ainda que a derivada de $f(\cdot)$ seja decrescente, então quanto melhor a qualidade do enlace, menor será o aumento da utilidade. Observa-se ainda que, quanto menor a quantidade de recursos necessários para atender u , maior é $Y(u)$, isto é a maior é utilidade para a rede em servi-lo, por isso, pode-se dizer que temos um critério de menor esforço. A partir do ranqueamento dos UEs, ranqueiam-se os *beams* e, subsequentemente, as BSs.

Para a avaliação do CSI é necessária uma ferramenta capaz de simular a *Radio Access Network* (RAN) 5G com múltiplos usuários espacialmente distribuídos. Logo, a ferramenta deve considerar a variabilidade intrínseca do canal sem-fio entre UEs e BSs. Desenvolveu-se um simulador capaz de estimar o CSI e avaliou-se o desempenho da proposta de alocação dos recursos de uma RAN 5G.

Após esta introdução, a Seção II apresenta a formulação do problema. A Seção III traz nossa proposta para a alocação dos recursos em uma RAN 5G baseada em utilidade. A Seção IV avalia o desempenho da proposta. Finalmente, na Seção VI as conclusões são apresentadas.

Christian Rodrigues, Centro de Estudos em Telecomunicações (CETUC), Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro-RJ, e-mail: christianfragoas@gmail.com; Lisandro Lovisoló, Laboratório de Processamento de Sinais, Aplicações Inteligentes e Comunicações (PROSAICO), Universidade do Estado do Rio de Janeiro, Rio de Janeiro-RJ, e-mail: lisandro@uerj.br; Luiz da Silva Mello, Centro de Estudos em Telecomunicações (CETUC), Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro-RJ, e-mail: smello@cetuc.puc-rio.br.

II. FORMULAÇÃO DO PROBLEMA

Considera-se que a RAN é composta por N BSs e que cada BS pode ter diferentes setores, e que cada setor pode ter até B beams. A RAN deve prover a mesma capacidade alvo C_{target} para cada UE. Para isso, os recursos em frequência (as subportadoras da RAN) são alocadas para os diferentes UEs ao longo do tempo, em *timeslots*.

A. Déficit de Capacidade

O déficit de capacidade do u -ésimo UE no *timeslot* t é

$$CD(t, u) = C_{\text{target}} - \prod_{\tau=1}^t C(\tau, n_{:u}, b_{:u}, u), \quad (2)$$

onde $C(\tau, n_{:u}, b_{:u}, u)$ é a capacidade provida para o u -ésimo UE no *timeslot* τ pelo *beam* $b_{:u}$ da BS $n_{:u}$. Se o *timeslot* $t=1 \dots T$ e a rede serve U UEs, o déficit de capacidade médio é

$$\overline{CD} = \prod_{u=1}^U CD(T; u) = U C_{\text{target}} \prod_{u=1}^U \prod_{\tau=1}^T C(\tau, n_{:u}, b_{:u}, u); \quad (3)$$

A alocação de recursos deve prover potência e recursos espectrais ao longo do tempo que minimize (3).

B. Modelagem do Problema

Para simplificar o processo de alocação de potência, assume-se que a potência máxima de transmissão é P_{max} . Assumindo que ela é igualmente distribuída por toda a largura de banda. Isto é, $\text{PSD} = P_{\text{max}}/\text{BW}$. Ademais, assume-se que somente um *beam* é ativado por setor durante um *timeslot* (pretende-se modelar o problema com múltiplos beams no futuro). Usa-se $\zeta(t, n, b)$ para representar a matriz de ativação que indica se o b -ésimo *beam* na n -ésima BS está ligado ou desligado durante o *timeslot* t . Com tais suposições, chegamos ao problema

$$\begin{aligned} \min \quad & \prod_{n=1}^N \prod_{b \in \mathcal{B}_n} \prod_{u=1}^U \prod_{t=1}^T C_{\text{target}} - (t; n; b) C(\tau; n; b; u) \quad (4) \\ \text{s.t.} \quad & C_1: \prod_{n=1}^N \prod_{b \in \mathcal{B}_n} \prod_{u=1}^U \prod_{t=1}^T (t; n; b) \leq 1; \quad \forall t; n; \\ & C_2: \prod_{u=1}^U \prod_{t=1}^T \sum_{n=1}^N \sum_{b \in \mathcal{B}_n} (t; n; b) P_{\text{Tx}}(t; n; b; u) \leq P_{\text{max}}; \quad \forall t; b; n; \end{aligned}$$

Acima, \mathcal{B}_n lista os *beams* na n -ésima BS. A primeira condição (C_1) implica que somente um *beam* está ativo por BS em t . Na condição C_2 , $P_{\text{Tx}}(t, n, b, u)$ é a potência de transmissão para o UE u durante o *timeslot* t pelo *beam* b da BS n , assim C_2 condiciona a potência total por *timeslot* em uma BS a ser menor do que a potência máxima.

III. SOLUÇÃO PROPOSTA

Visando prover a mesma capacidade para todos os UEs servidos pela RAN, propõem-se alocar os recursos em ordem de menor esforço.

A. Ranqueamento por Utilidade

1) *Ranqueamento de UEs*: Como função utilidade para os UEs, define-se (ver a Eq. (1))

$$Y_{\text{BW}}(t; u) = (t; u) \log(\text{SNIR}(t; u)); \quad (5)$$

$$(t; u) = \frac{\text{BW}_{\min}(t; n_u; b_u)}{\text{BW}_{\text{need}}(t; u)}. \quad (6)$$

O *Signal to Noise plus Interference Ratio* (SNIR) depende do CSI, como observado em (17). Consequentemente, quanto melhor a qualidade do enlace, maior é a utilidade de alocar recursos para o UE. A Eq. (5) compara a menor largura de banda $\text{BW}_{\min}(t, n_u, b_u)$ que pode ser reservada no *beam* b_u da BS n_u que serve o u -ésimo UE durante o *timeslot* t e a largura de banda $\text{BW}_{\text{need}}(t, u)$ necessária para atender o UE u no *timeslot* t com C_{target} – o logaritmo comprime a faixa dinâmica do SNIR.

Para computar $\text{BW}_{\text{need}}(t, u)$, considera-se a qualidade do enlace usando a Lei de Shannon, i.e.,

$$\text{BW}_{\text{need}}(t, u) = \frac{C_{\text{target}} - C(t-1, u)}{\log_2(1 + \text{SNIR}(t, u))}, \quad (7)$$

$$C(t-1, u) = \prod_{\tau=1}^{t-1} C(\tau, n_{:u}, b_{:u}, u) \quad (8)$$

onde $C(t-1, u)$ é a capacidade provida ao UE u em *timeslots* anteriores. Seja $U(t, n_u, b_u)$ o conjunto de UEs servidos pelo *beam* b_u da n_u -ésima BS que servem o u -ésimo UE durante o *timeslot* t . A granularidade de alocação dos recursos de banda nesse *beam* é

$$\text{BW}_{\min}(t, n_u, b_u) = \nu \frac{\text{BW}}{|U(t, n_u, b_u)|}. \quad (9)$$

Admite-se que a fração ν de recursos é igual para todos os UEs servidos pelo *beam* (assim todos os UEs estão conectados à RAN) durante um *timeslot*.

A Eq. (5) analisa a utilidade de alocar a largura de banda $\text{BW}_{\text{need}}(t, u)$ para o UE u durante o *timeslot* t . Ela é maior, quanto maior for o SNIR do UE ou menor for a quantidade de UEs servidas pelo *beam*, ou menor for a banda necessária para que o UE atinja C_{target} .

2) *Ranqueamento dos Beams*: Após computar $Y_{\text{BW}}(t, u)$, computa-se a utilidade de um *beam* somando as utilidades por ele servidas. Isto é, a utilidade do b -ésimo *beam* da n -ésima BS ($b \in \mathcal{B}_n$) no *timeslot* t é

$$W_{\text{beam}}(t, n, b) = \log(W(t, n, b)), \quad (10)$$

$$W(t, n, b) = \prod_{u \in U(t; n; b)} Y_{\text{BW}}(t, u). \quad (11)$$

3) *Ranqueamento das BSs*: A partir das utilidades dos *beams* de uma BS, Eq. (10), define-se a utilidade da n -ésima BS no *timeslot* t via

$$W_{\text{BS}}(t, n) = \prod_{b \in \mathcal{B}_n} W_{\text{beam}}(t, n, b). \quad (12)$$

B. Alocação de Recursos

Prioriza-se as UEs que demandam menos recursos a fim de garantir C_{target} para a maior quantidade de UEs. Aloca-se os recursos de largura de banda \times tempo em função da qualidade do enlace, atendendo os UEs em ordem decrescente de utilidades. Entretanto, reserva-se uma quantidade mínima de recursos $t_{\min} \times \text{BW}_{\min}$ para todos os UEs. A Eq. (9) define BW_{\min} que é dependente do *beam*. Por sua vez, t_{\min} é o menor intervalo de tempo

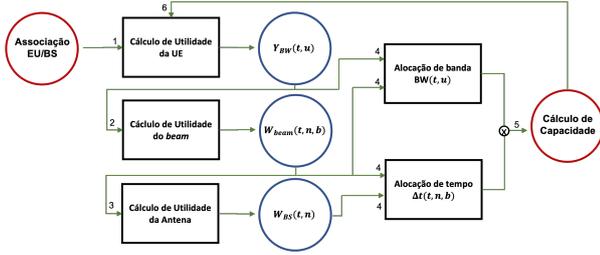


Fig. 1. Alocação de recursos baseada em utilidade. Obtenção das utilidades 1 – das UE, via (5); 2 – dos *beams*, via (11); 3 – dos setores, via (12); Na etapa 4, essas utilidades ditam a alocação dos recursos, via (13) (tempo) e 14 (espectro). Após, em 5, verifica-se as capacidades alcançadas, e atualiza-se as utilidades, em 6.

de ativação de um *beam*, que diminui com t , fazendo $\Delta t(t, n, b) \geq 0 \forall t$, e é igual para todos os *beams* e setores.

A Fig. 1 apresenta o processo. Em t , ativa-se o *beam* b da BS n , i.e. $b \in \mathcal{B}_n$, por

$$\Delta t(t, n, b) = \delta \cdot t_{\min} + \frac{W_{\text{beam}}(t, n, b)}{W_{\text{BS}}(t, n)} (T - |\mathcal{B}_n| \times t_{\min}) \quad (13)$$

timeslots, na qual $\delta = 1 - t/T$. Obviamente, um *beam* só será ativado se $\exists u | CD(t, u) \leq 0, \forall u \in \mathcal{U}_b$. Se o *beam* atendendo a UE u está ativo, a largura de banda reservado para a UE é

$$BW(t; u) = BW_{\min} + \frac{Y_{\text{BW}}(t; u)}{W_{\text{beam}}(t; n_u; b_u)} \quad (14)$$

$$(BW - jU(t; n_u; b_u)j - BW_{\min}(t; n_u; b_u)) :$$

Os recursos de banda são alocados em um *timeslot* até a exaustão. Tal procedimento é iterado para $t = 1 \dots T$, computando-se $Y_{\text{BW}}(t, u)$ para todos os UEs. Observa-se que quando um UE atinge a capacidade alvo, ele passa a ser desconsiderado nos *timeslots* subsequentes.

C. Simplificação da Formulação

Acima, SNIR(t, u) e $CD(t, u)$ são atualizados para UEs que não atingiram a capacidade alvo a cada t . A fim de reduzir a complexidade do algoritmo, SNIR(t, u) e $CD(t, u)$, não são atualizados $\forall t, u$. Ranqueia-se os UEs, *beams* e BSs a partir de SNR(u) em $t = 1$,

$$Y_{\text{BW}}(1, u) = \frac{BW_{\min}(n_u, b_u, s_u)}{BW_{\text{need}}(u)} \log(\text{SNR}(u)) \quad (15)$$

$$BW_{\text{need}}(1, u) = \frac{C_{\text{target}}}{\log_2(1 + \text{SNIR}(u))} \quad (16)$$

e mantém-se o ranqueamento para todo t .

IV. FERRAMENTA DE SIMULAÇÃO DA RAN 5G

Implementa-se uma abstração robusta da RAN, discutimos nesta seção os detalhes.

A. Beamforming

Beamforming usa um *array* para produzir diagramas de radiação diretivos. Para produzir tal diagrama, usa-se a recomendação 3GPP TR 37.840 [7], [8] e gera-se os *beams* a partir de um *array* de elementos de antena.

B. Grid of Beams

O Grid of Beams (GOB) lista as combinações de ângulos de azimute e elevação que definem o apontamento do *beam* para a correspondente antena setorial [9] ¹ Os apontamentos dos *beams* são escolhidos de forma que, suas áreas são complementares no entorno de uma BS [10].

C. Channel State Information - CSI

Para cada UE u e um *beam* b da BS n , o CSI é

$$\text{CSI}(n, b, u) = P_{\text{Tx}}(n) + G(n, b, u) + \text{PL}(n, u), \quad (17)$$

onde, $P_{\text{Tx}}(n)$ é a potência de transmissão, $G(n, b, u)$ é o ganho do b -ésimo *beam* da n -ésima BS para do u -ésimo UE, e $\text{PL}(n, u)$ é o *path loss* (PL) desta e a n -ésima BS. Emprega-se o modelo de propagação ABG (alpha-beta-gama) [11] para modelar o PL com $\sigma^2 = 6$ dB.

D. Associação entre UEs, beams e BSs

Associa-se o u -ésimo UE ao *beam* b_u da BS n_u se

$$(n_u, b_u) = \arg \max_{(n;b) \in \mathcal{N}_g} \max_{f1:Bng} \text{CSI}(n, b, u). \quad (18)$$

Isto é, um UE é associado ao *beam* e BS com o maior CSI.

E. SNIR

A SNIR é a razão entre a potência recebida do *beam* que o atende e a soma das contribuições dos demais *beams* ativos (interferência) e do ruído durante um *timeslot*. Para o UE u associado ao *beam* b_u da BS n_u , temos

$$\text{SNIR}(t; u) = \frac{P_{\text{Rx}}(u; n_u; b_u)}{\sum_{(n;b) \in (n_u; b_u)} (t; n; b) P_{\text{Rx}}(u; n; b) + P_{\text{noise}}(BW(t; u))} \quad (19)$$

onde $P_{\text{Rx}}(u, n, b)$ é a potência recebida em $p_u = (x_u, y_u)$, (posição do u -ésimo UE) a partir do *beam* b da BS n , P_{noise} é a densidade espectral de potência do ruído e $BW(t, u)$ é a banda que serve o UE u durante t , ou seja, $BW(t, u)$ depende da alocação dos recursos. A interferência origina-se de outros *beams* e BSs, já que não é considerada nenhuma coordenação entre as BSs (e seus *beams*).

F. Capacidade Fornecida a um UE

Devemos estimar a taxa de *download* de todas as UEs servidas pela RAN. A Lei de Shannon limita a capacidade [bits/Hz/s] em um *timeslot* t a

$$C_{\text{timeslot}}(t, u) = BW(t, u) \log_2(1 + \text{SNIR}(t, u)). \quad (20)$$

Obtemos a capacidade de um UE via

$$C(u) = \sum_{\tau \in \mathcal{T}_u} C_{\text{timeslot}}(\tau, u) \times \text{slot duration}, \quad (21)$$

onde \mathcal{T}_u indica os *timeslots* em que há recursos alocados para u .

¹ Abstrai-se os setores da discussão já que eles são incorporados nos índices das antena e *beams*.

V. ANÁLISE DA PROPOSTA

A. Avaliação de Desempenho

Usamos o simulador da Seção IV para avaliar a capacidade média e a fração de UEs que atinge C_{target} . Empregam-se múltiplas rodadas para avaliar estatisticamente o desempenho em usar N BSs para atender U UEs.

1) *Configuração*: Avalia-se o desempenho da rede quando composta por $N = \{1, 2, 4, 6, \dots, 14, 16\}$ BSs. Executam-se 2000 rodadas para cada N , em que posiciona-se 800 UEs numa grade de 1000×1000 células de 30m^2 aleatoriamente. Em cada rodada assume-se a rede e as UEs estáticas. O tempo de simulação é $T = 1\text{s}$ e $t_{\text{slot}} = 1\text{ms}$.

2) *Distribuição Estocástica de UEs*: Modela-se a distribuição de UEs na ROI usando a soma de densidades de quatro Gaussianas 2D com seus centros uniformemente espaçados. São empregados dois valores de C_{target} , 30 Mbps e 50 Mbps. Valores típicos para *streaming* de vídeo e jogos em nuvem na resolução 4K.

3) *Arranjo de BSs*: As distâncias entre um UE e a BS mais próxima e, portanto, a CSI e a capacidade dependem do arranjo de BSs. Para mitigar essa influência, as posições das BSs são selecionadas utilizando a clusterização de UEs em função da distância euclidiana [12]. Usam-se N clusters com uma BS no seu centroide. Desta forma, reduz-se a distância média entre UEs e suas BSs, conseqüentemente, minimiza-se a perda de percurso média.

4) *Configuração dos Equipamentos*: Consideram-se todos as BSs iguais, transmitindo a 100 W, largura de banda de 100 MHz e 3 setores com apontamentos $(\phi, \theta) = (60^\circ, 0^\circ)$, $(180^\circ, 0^\circ)$ e $(300^\circ, 0^\circ)$. Cada setor é composto de um *array* de 8×8 elementos com ganho de 5 dBi que geram um GOB de 20 *beams* [8]. As UEs tem 0 dBi de ganho, diagramas de radiação omnidirecionais e altura de 1,5 m.

B. Comparações

Compara-se o desempenho da proposta de alocação de recursos baseada em utilidade, sua versão simplificada, uma estratégia simples igualitária e um limite superior otimista de capacidade. Para distribuir igualmente os recursos de tempo entre os *beams* de um setor e os recursos de frequência entre os UEs de um *beam*, faz-se

$$\Delta t(n, b) = \frac{T}{|\mathcal{B}_n|}, \quad \text{e} \quad \text{BW}(u) = \frac{\text{BW}}{|\mathcal{U}_{b(u)}|}, \quad (22)$$

que são fixas para $t=1 \dots T$. Acima, $|\mathcal{B}_n|$ é o número de *beams* ativos associados com UEs e $\Delta t(n, b)$ é a quantidade de *timeslots* em que o *beam* b da BS n está ativo, $\text{BW}(u)$ é a largura de banda alocada para o UE u e $|\mathcal{U}_{b(u)}|$ é o número de UEs atendidos pelo *beam* b que atende o UE u .

Um limite superior otimista para a capacidade atingível é obtido através da eficiência espectral média para as BSs, seus *beams* e os UEs associados. Considera-se que a mesma largura de banda $\text{BW}(n)$ está disponível para todos os setores da BS n . Somam-se as capacidades médias das BS considerando seus *beams* $|\mathcal{B}_n|$ e divide-se pelo número de usuários por ela atendidos $|\mathcal{U}_n|$). Este cenário hipotético

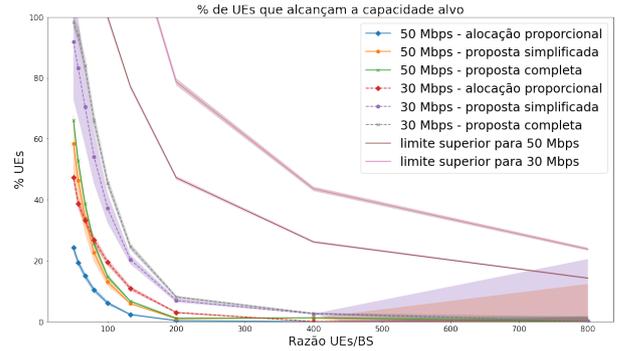


Fig. 2. Porcentagem de UEs que alcançam C_{target} em função de UEs/BS para o algoritmo de alocação de recursos proposto, a alocação uniforme e o limite de capacidade da RAN.

fornece o limite

$$\bar{C}_{\text{top}} = \sum_{n=1}^{\mathcal{N}} \frac{|\mathcal{B}_n| \text{BW}(n)}{|\mathcal{U}_n|} E_{U_n} [\log_2(1 + \text{SNR}(u, n))] \times T. \quad (23)$$

A razão entre \bar{C}_{top} e C_{target} é uma expectativa otimista da fração de UEs que atingem a C_{target} .

C. Resultados

A Fig. 2 apresenta os desempenhos (média \pm dispersão) das diferentes estratégias mencionadas e o limite superior. O eixo vertical é a porcentagem de UEs que atingem a capacidade alvo, e o horizontal é a razão UEs/BS.

Vê-se que com a metodologia proposta mais UEs atingem C_{target} do que com a estratégia uniforme, e está um pouco mais próxima ao limite superior. A simplificação da metodologia proposta acarreta em resultados ligeiramente inferiores e com uma dispersão um pouco maior. A ligeira perda de desempenho da simplificação advém da não atualização do ranqueamento a cada alocação que, na versão completa da metodologia proposta, evita desperdícios de recursos. Também se observa uma grande dispersão nos resultados para as propostas completa e simplificada quando a razão UE/BS aumenta (a quantidade de BSs reduz), isso provavelmente decorre da dificuldade em obter posições para BSs com menor variabilidade da qualidade dos enlaces entre UEs e suas BSs.

A Fig. 3 apresenta os histogramas de déficit de capacidade obtidos com a alocação proposta quando $C_{\text{target}}=50$ Mbps. O eixo vertical apresenta as quantidades de UEs com o déficit indicado no eixo frontal para diferentes valores de BSs no eixo restante. Comparamo-los como os histogramas obtidos pela alocação uniforme na Fig. 4. Vemos que à medida que a quantidade de BSs aumenta, a concentração de déficits usando a alocação proposta se move de C_{target} para a região em torno de zero (mais UEs atingem C_{target}), o que não ocorre com a alocação uniforme, mostrando que a proposta funciona.

Cumpramos observar que visualmente, os histogramas de déficits de capacidade das versões completa e simplificada da proposta são semelhantes. Usou-se o

