

Solução para Gerenciamento de Largura de Banda usando Aprendizado por Reforço

Eriberto Silva, Iago Rego, Dario Vieira, Augusto Venâncio e Vicente Sousa

Abstract—Este trabalho propõe uma solução de aprendizado por reforço para automação de técnicas de coordenação de interferência entre células (ICIC) voltadas à mitigação de um dos principais desafios do projeto de redes móveis: a interferência co-canal (ICI). A solução é testada em um cenário dinâmico com a presença de *hotspots*, um aglomerado de usuários que pode agravar a ICI. As investigações foram efetuadas utilizando o simulador de redes *ns-3*. O impacto do *hotspot* é discutido à luz da distribuição de banda entre borda e centro da célula. Pelos resultados de caracterização do problema, é possível verificar a necessidade de, além do uso de ICIC, alocar de forma dinâmica sua distribuição de banda. Levando em consideração essa dinamicidade do ambiente, propõe-se como solução a utilização da técnica *Multi Armed Bandit* (MAB) para otimização do gerenciamento da largura de banda das técnicas de ICIC. Os experimentos de prova de conceito demonstram que a solução proposta supera, em termos de capacidade de transmissão, a configuração estática de distribuição de largura de banda, comprovando ganhos significativos.

I. INTRODUÇÃO

A maioria dos sistemas modernos de comunicação utiliza fator de reuso de frequência igual a 1, ou seja, as células reutilizam todo o espectro disponível. No entanto, essa maneira de reutilização de frequência não traz somente benefícios e deve ser utilizada com cuidado, especialmente em redes com muitas estações bases próximas. Nessa situação, a interferência co-canal, ou *Inter-Cell Interference* (ICI), pode aumentar de maneira descontrolada, degradando a *SINR* (*Signal-to-Interference-plus-Noise Ratio*) e, consequentemente, a qualidade de serviço (QoS) em termos de taxa útil e taxa de erro. Por outro lado, um bom planejamento do fator de reuso reduz a ICI e aumenta a SINR, levando a melhor QoS, mesmo considerando a diminuição de largura de banda útil da célula, devido a um reuso de frequência menor que 1 [1].

Os sistemas 3GPP têm flexibilidade no gerenciamento de banda e de interferência utilizando, por exemplo, estratégias como MIMO *beamforming*, controle de potência, *bandwidth parts*, uso de *small cells* e o Reuso Fracionário de Frequência (FFR). Especificamente relacionadas a gerência de interferência, técnicas de coordenação de ICI foram desenvolvidas como forma de mitigar os efeitos da interferência, atuando na alocação de banda, na potência de transmissão e na distribuição de usuários por regiões da célula. Essas técnicas são chamadas de *Inter-Cell Interference Coordination* (ICIC).

Dario é do Efrei Research Lab, Paris, France. E-mail: dario.vieira@efrei.fr. Os demais autores são do Leading Advanced Technologies Center of Excellence (LANCE), UFRN, Brasil. E-mails: {eriberto.souto.095, iago.diogenes.072}@ufrn.edu.br, augusto@dimap.ufrn.br, vicente.sousa@ufrn.br. Esse estudo foi financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. As simulações fornecidas por esse trabalho foram realizadas com o apoio do Núcleo de Processamento de Alto Desempenho da UFRN (NPAD/UFRN).

Esse gerenciamento de interferência torna-se ainda mais complexo em cenários dinâmicos, nos quais, frequentemente, surgem picos de tráfego inesperados devido às concentrações de usuários decorrentes de eventos aleatórios, por exemplo, eventos artísticos, feiras, jogos, dentre outros. Esses aglomerados de usuários são denominados como *hotspot* e, geralmente, aparecem de forma aleatória, disputando os recursos disponibilizados aos usuários da região [2].

A aleatoriedade do *hotspot* ocorre tanto em posição quanto em quantidade de dispositivos, sendo o seu surgimento capaz de alterar a capacidade projetada do sistema e aumentar a disputa dos usuários pelos recursos de rádio. Além disso, é possível experimentar um aumento de ICI, principalmente para os usuários próximos à borda da célula, que também experimentam maior degradação de sinal [3].

A técnica de ICIC alvo do estudo proposto neste artigo faz uso do FFR, permitindo o gerenciamento da alocação de banda em diferentes áreas da célula, visando uma maior eficiência espectral. Os autores de [4] propõem um método de ICIC que gerencia dinamicamente a largura de banda da borda da célula. O método se baseia no número de usuários conectados e o uso da interface X2 para troca de informações entre células. Em [5], é proposto um novo esquema de reuso e alocação de frequência chamado *First Tier Interference Free* (FTIF). Ele setoriza as células sem reduzir a potência de transmissão, eliminando a região de interferência entre a célula adjacente e a borda da célula vizinha. Em [1], os autores apresentam um *survey* acompanhado de uma análise de desempenho de técnicas de ICIC para diferentes condições de rádio e cargas da rede. O artigo apresenta uma análise comparativa entre o FFR e o *Strict Frequency Reuse* (SFR), discutindo a natureza estática de cada algoritmo e seus desempenhos perante os reusos 1 e 3. Por fim, os autores de [3] abordam a interferência entre células em cenários dinâmicos com e sem a presença de *hotspot*, avaliando os danos que ele causa no desempenho do sistema LTE. Uma solução utilizando aprendizado de máquina é dedicada ao ajuste da distribuição de usuários entre as diferentes zonas de reuso da célula. A proposta é baseada em *Q-Learning* e ajusta os parâmetros de distribuição de usuários entre o centro e a borda da célula de forma dinâmica, conforme o surgimento dos *hotspots*. O trabalho [3] não explora a gerência de largura de banda do ICIC.

Apesar das publicações identificadas, a literatura é carente de trabalhos sobre técnicas de ICIC que façam alocação dinâmica de banda entre centro e borda da célula em cenários com presença de *hotspot*. Este trabalho propõe uma solução de aprendizado por reforço para definir autonomamente a distribuição de largura de banda do ICIC, alocando diferentes distribuições de banda entre centro e borda das células. Essa proposta contribui com a gerência espacial de largura de banda para mitigar os efeitos da ICI agravados pela presença dos *hotspots*.

II. SOLUÇÃO DE ICIC E METODOLOGIA DE PROTOTIPAGEM

Para este trabalho, optou-se pelo algoritmo SFR, pois apresenta equilíbrio de desempenho em diferentes cenários e é efetivo na mitigação dos efeitos de ICI, como comprovado em [6]. O esquema SFR consiste em dividir a banda total do sistema em uma sub-banda central compartilhada e em uma sub-banda privada de borda, com larguras de banda e potências distintas. Os usuários são alocados para cada sub-banda dependendo de sua RSRQ, obedecendo a um limite pré-estabelecido.

O sistema multicelular utilizado para prova de conceito deste trabalho foi prototipado no Network Simulator (*ns-3*). Ele é um simulador de uso gratuito e de código aberto, bastante utilizado no meio acadêmico e em prototipagem na indústria de telecomunicações. Além disso, possui uma comunidade ativa e colaborativa na solução de problemas em experimentos utilizando o simulador.

O *ns-3* suporta algoritmos de ICIC por intermédio da classe abstrata *LteFfrAlgorithm*. Além do SFR, existem mais seis algoritmos de FFR implementados, descritos com detalhes em [7]. O comportamento do SFR é ilustrado ao comparar as três partes da Figura 1, que apresentam 10 células divididas em *clusters* com 3 células. Enquanto as duas partes de cima da figura apresentam os reúsos 1 e 3, a parte de baixo apresenta o SFR. Note que com os reúsos 1 e 3, as células transmitem com a mesma potência, mas a largura de banda é usada em todas as células com reuso 1 e dividida igualmente pelas três células no reuso 3. No SFR, potência e largura de banda das duas regiões da célula podem ser diferentes.

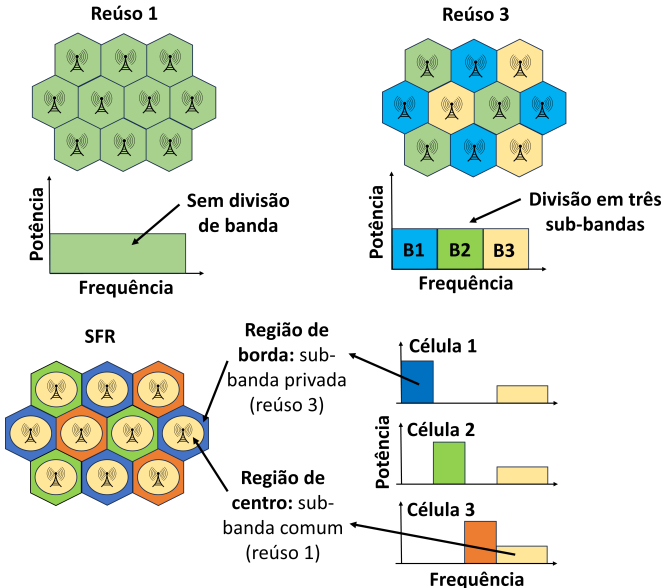


Fig. 1: Comparação entre Reuso 1, Reuso 3 e SFR.

No *ns-3*, é possível definir a largura de banda e a potência de cada região (centro e borda) do SFR. Outro parâmetro importante é o limiar *RsrqThreshold*. Ele determina se o usuário é alocado para a região do centro ou da borda da célula, dependendo de seu valor de *Reference Signal Recieved Quality* (RSRQ) [7].

III. SOLUÇÃO PROPOSTA

Uma das características mais importantes do aprendizado por reforço, que se diferencia dos outros tipos de aprendizado, é a habilidade de usar as informações de treinamento, *avaliando-as* ao invés de puramente *instruir* ao informar as ações corretas. Por exemplo, considere o seguinte problema: um agente se depara repetidamente com um mesmo conjunto de ações. Para cada ação escolhida, o agente recebe uma recompensa a partir de uma distribuição de probabilidade. O objetivo do agente é maximizar a recompensa total esperada durante um período de tempo [8].

O problema descrito no parágrafo anterior é a forma básica do *k-armed Bandit* e recebe esse nome devido a uma analogia com as máquinas caça-níquel, que em inglês eram chamadas alternativamente *one-armed bandits*. Dessa maneira, considerando o problema em questão, cada ação pode ser representada por uma jogada de uma das *k* máquinas, e as recompensas seriam o pagamento por acertar o prêmio, sendo este bom ou ruim. Ao repetir as melhores ações selecionadas, é possível maximizar os ganhos, concentrando as ações nas melhores *k* máquinas. Portanto, para cada ação *k*, há uma recompensa média associada, uma vez que cada uma segue uma distribuição de probabilidade, e essa recompensa é chamada valor da ação [8].

Considerando que cada ação selecionada no tempo *t* é chamada de *A_t* e sua correspondente recompensa de *R_t*, dada uma ação *a* qualquer, o valor da ação ótima *q_{*}(a)* pode ser escrito como na Equação (1).

$$q_*(a) \doteq \mathbb{E}[R_t | A_t = a], \quad (1)$$

sendo $\mathbb{E}$ o valor esperado [8]. Define-se como *Q_t(a)* o valor estimado da ação *a* no tempo *t* e espera-se que *Q_t(a)* aproxime-se de *q_{*}(a)*, a medida que as ações forem sendo escolhidas durante o período de tempo que o agente estiver aprendendo. Com o valor aproximado *Q_{t→∞}(a)*, o agente pode, enfim, escolher sempre a ação que retorna a maior recompensa. Um método para aproximar o valor de *q_{*}(a)* por *Q_t(a)* e usar o valor aproximado para escolher as ações futuras é o *sample-average* [8]. Nele, os valores das ações são estimados pela média das recompensas recebidas:

$$Q_t(a) \doteq \frac{\sum_{i=1}^{t-1} R_i \cdot 1_{A_i=a}}{\sum_{i=1}^{t-1} 1_{A_i=a}}, \quad (2)$$

sendo $1_{A_i=a}$ uma função que é 1 quando *A_i* = *a* for verdadeiro, e 0 para quando for falso. Pela lei dos grandes números, na medida em que o denominador da Equação (2) tende ao infinito, o valor de *Q_t(a)* aproxima-se de *q_{*}(a)*. Assim, garante-se que, se o algoritmo rodar por tempo suficiente, a convergência é alcançada. Por fim, a Equação (2) também pode ser representada pela sua forma incremental como:

$$\begin{aligned} Q_{n+1}(a) &= Q_n(a) + \alpha_n(a) [R_n - Q_n(a)] \\ &= Q_n(a) + \frac{1}{n} [R_n - Q_n(a)], \end{aligned} \quad (3)$$

em que *n* é o número de vezes que a ação *a* foi escolhida até o atual momento. O termo $\alpha_n(a) = \frac{1}{n}$ é o que garante a convergência do algoritmo, pois é a representação do método *sample-average*. Para evitar soluções locais ou subótimas,

pode ser utilizada a estratégia ϵ -greedy, em que um valor ϵ é definido, e a cada passo t , o agente tem probabilidade ϵ de escolher uma ação aleatória, ou probabilidade $(1 - \epsilon)$ de escolher a ação com maior valor $Q_t(a)$ (greedy).

O algoritmo SFR (MAB) proposto neste trabalho tem como ações as seguintes opções de distribuição de banda entre centro e borda: 16-28, 28-24, 40-20, 52-16 e 64-12. A ação 16-28, por exemplo, significa que 16 PRBs são alocados para o centro da célula e 28 para cada região de borda. A recompensa é o *throughput* médio do usuário para o conjunto de três células, calculado entre duas execuções do MAB. O valor de ϵ começa alto (20%) e decresce exponencialmente. Nos experimentos deste artigo, a ação do MAB é realizada a cada 1 s.

IV. CENÁRIOS DE AVALIAÇÃO

Para os resultados de caracterização do problema, foram adotados dois cenários com diferentes posicionamentos dos *hotspots*, mas mantendo a dimensão de célula, bem como a distância entre elas, como ilustrado nas Figs. 2a e 2b. O cenário celular é composto por 3 estações bases distanciadas entre si de 500 m. Os usuários são distribuídos em dois grupos: (i) distribuição uniforme, com 30 usuários uniformemente distribuídos por toda região da célula; e (ii) *hotspots*, que delimitam usuários distribuídos em círculos com raio de 50 m. Existem 3 *hotspots* com 20 usuários cada. No Cenário 1 (Fig. 2a), os *hotspots* são posicionados a 50 m da estação base, enquanto que no Cenário 2 (Fig. 2b), os *hotspots* estão a 450 m.

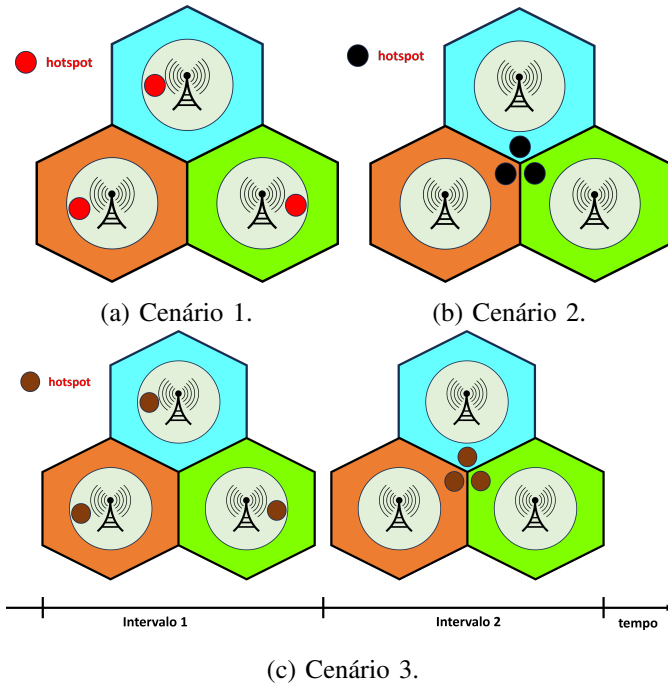


Fig. 2: Posicionamento dos *hotspots* nos cenários avaliados.

Para avaliar o desempenho do MAB, um cenário dinâmico foi configurado. Ele consiste na modelagem da mudança de posicionamento de *hotspot*, como ilustrado na Fig. 2c. Para esse cenário, foram realizados experimentos de 100 s, em que os *hotspots* próximos à estação base se desfazem em 50 s, enquanto os *hotspots* distantes se formam e duram até o final do experimento.

A largura de banda do sistema é de 100 PRBs (20 MHz) e é distribuída na seguinte proporção: 1 fração direcionada ao

centro e 3 frações da banda iguais para cada borda das três células vizinhas (reuso 3). As proporções de banda utilizadas nos experimentos de caracterização do problema foram as mesmas usadas na configuração do MAB.

Os valores 30 e 31 foram utilizados como limiar de *RsrqThreshold*. Segundo [9], utilizar o limiar 30 significa que os usuários com valores de *RSRQ* menores que -4,5 dB são alocados na borda. Os usuários com valores de *RSRQ* maiores usam os recursos de rádio da região central. Esse valor é -4 dB para *RsrqThreshold* igual a 31. O tráfego oferecido por usuário é constante durante todo o experimento e foram realizadas simulações com tráfego oferecido de 2, 1 e 0,5 Mbps.

V. RESULTADOS E DISCUSSÕES

A. Resultados de caracterização do problema

Gráficos de *throughput* de célula são apresentados nas Figs. 3 e nas 4, para *RsrqThreshold* igual a 30; e Figs. 5 e 6, para *RsrqThreshold* igual a 31. Os pares de figuras apresentam resultados para os Cenários 1 e 2. Cada conjunto de três barras representa uma distribuição de banda, sendo cada barra desse conjunto indicando o *throughput* para os valores das taxas oferecidas 2 Mbps, 1 Mbps e 500 kbps, da esquerda para a direita, respectivamente.

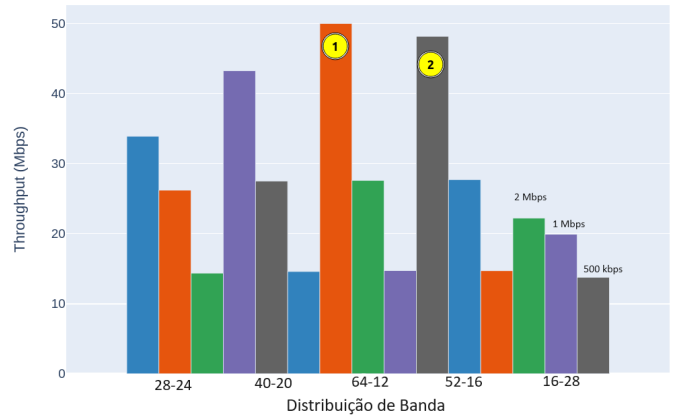


Fig. 3: Cenário 1: *Throughput* de célula - *RSRQThreshold* 30.

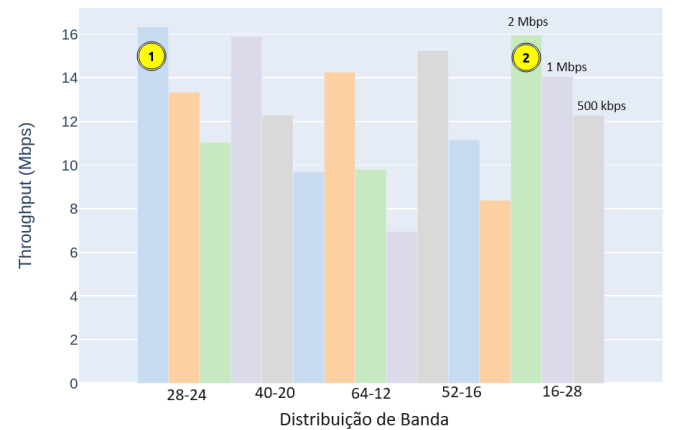


Fig. 4: Cenário 2: *Throughput* de célula - *RSRQThreshold* 30.

Os experimentos com *RSRQThreshold* igual a 30 revelam que ao comparar o *throughput* dos Cenários 1 e 2, o melhor desempenho é alcançado com diferentes configurações de

distribuição de banda, como identificado nas Figs. 3 e 4. Se os *hotspots* estão próximos à estação base (Cenário 1), alocar mais banda para o centro da célula (distribuição 64-12) resulta em maior *throughput*. Contudo, quando os *hotspots* estão afastados, o maior valor de *throughput* é alcançado com a distribuição de banda 28-24, uma alocação equilibrada entre borda e centro da célula.

Os resultados do Cenário 1 para *RSRQThreshold* igual a 31 (Figs. 5 e 6) são semelhantes ao anterior, com maior *throughput* para distribuição de banda 64-12 (mais banda para o centro da célula). Contudo, diferente do primeiro conjunto de resultados do Cenário 2, a melhor distribuição de banda é a 16-28, uma configuração com mais banda para a borda da célula.

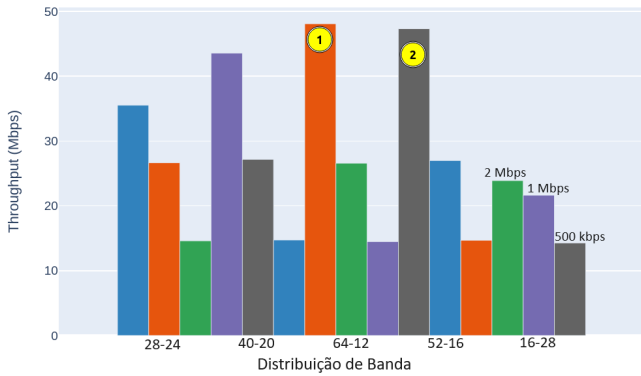


Fig. 5: Cenário 1: *Throughput* de célula - *RSRQThreshold* 31.

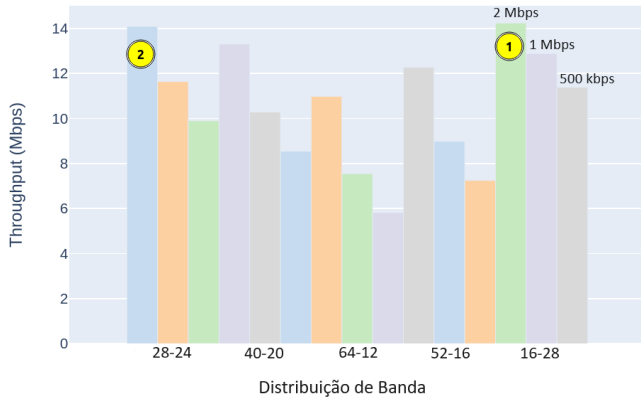


Fig. 6: Cenário 2: *Throughput* de célula - *RSRQThreshold* 31.

Esses resultados indicam que a variação da posição dos *hotspots* afeta significativamente o *throughput* do sistema, e que se agrava para maiores taxas oferecidas, chegando ao ponto de reduzir mais de 50% de desempenho ao comparar os dois cenários testados.

A Tabela I resume os resultados da caracterização do problema ao mostrar a *porcentagem* de redução do *throughput* do Cenário 1 (*hotspots* perto) em relação ao Cenário 2 (*hotspots* longe), i.e., quanto se perde ao mover o *hotspot* de perto para longe da estação base. A coluna *Melhor Escolha* mostra a redução de *throughput* ao se escolher a melhor distribuição de banda para ambos os cenários. A coluna *Escolha Fixa* mostra a redução de *throughput* ao se escolher a melhor distribuição de banda do Cenário 1 e aplicá-la ao Cenário 2. É possível concluir que manter a mesma distribuição de banda nos dois cenários pode aumentar as

perdas de *throughput* em pelo menos 4% (para o caso de *RSRQThreshold* 30 e 2 Mbps) e até 19% (para o caso de *RSRQThreshold* 30 e 500 kbps). Assim, é possível afirmar que não existe uma configuração de distribuição de banda fixa que seja melhor para todos os casos testados e que existe prejuízo de desempenho em escolha uma distribuição de banda errada.

TABELA I: Redução do *Throughput* entre os Cenários 1 e 2.

<i>RSRQThreshold</i> 30		
Taxa Oferecida	Melhor Escolha	Escolha Fixa
2 Mbps	67,34%	71,48%
1 Mbps	48,55%	59,77%
500 kbps	16,57%	52,78%
<i>RSRQThreshold</i> 31		
2 Mbps	70,37%	77,15%
1 Mbps	52,55%	62,13%
500 kbps	22,69%	42,06%

B. Resultados da solução proposta

O solução proposta neste trabalho usa o MAB para definir a distribuição de banda mais apropriada para o sistema, independente da taxa oferecida e da disposição do *hotspot*. Para avaliar essa habilidade, o Cenário 3 foi configurado (Fig. 2c). O desempenho do algoritmo MAB proposto é apresentado nas Figs. 7, 8 e 9, para *RSRQThreshold* igual a 30 e taxas oferecidas de 2, 1 e 0,5 Mbps, respectivamente.

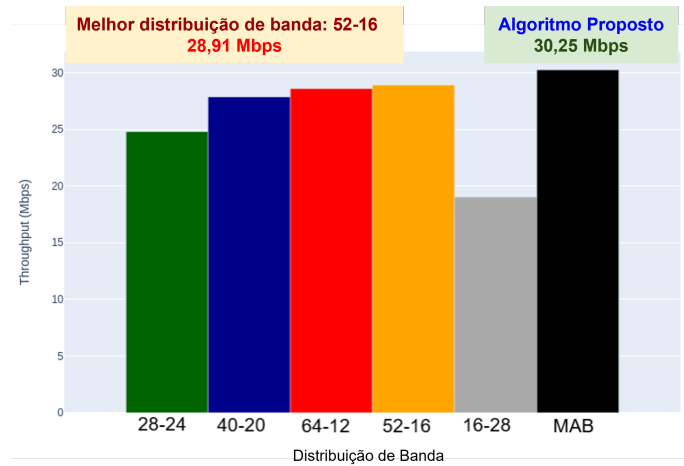


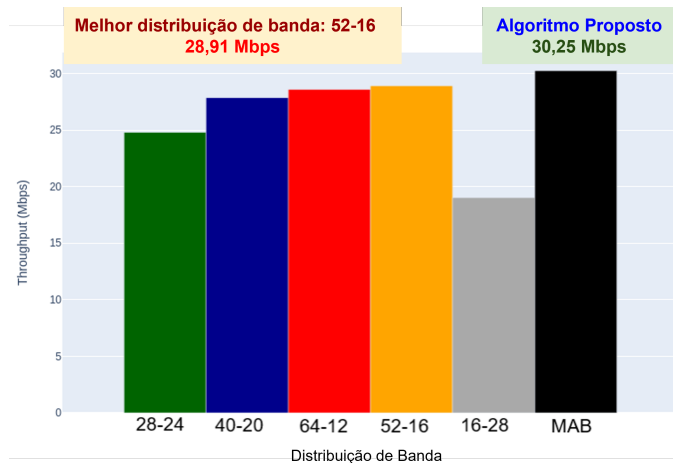
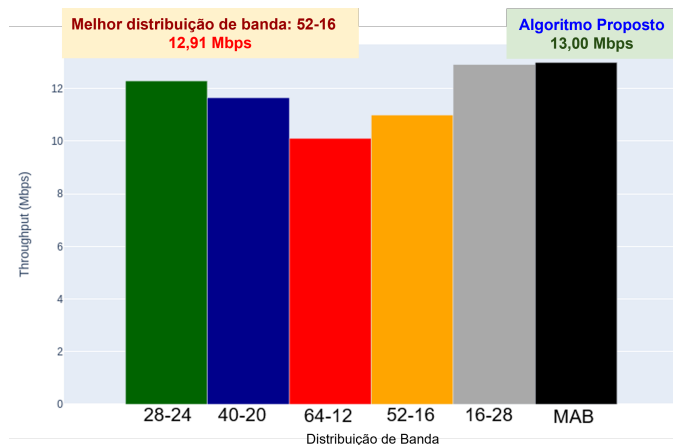
Fig. 7: Cenário 3: *Throughput* de célula - *RSRQThreshold* 30 e taxa oferecida de 2 Mbps.

Os resultados evidenciam ganho do algoritmo proposto em todas as configurações testadas. A ação do MAB em aprender a melhor distribuição de banda para os intervalos do cenário dinâmico se reflete em melhoria de *throughput* para uma parcela significativa de usuários. A Fig. 10 mostra a ECDF (*Empirical Cumulative Distribution Function*) do *throughput* para o caso de 2 Mbps e *RsrqThreshold* igual a 30. É possível visualizar que existe ganho substancial entre o vigésimo e o nonagésimo percentis do *throughput* ao aplicar o MAB.

A Tabela II sumariza os ganhos do algoritmo proposto para dois casos específicos: (i) *Pior escolha fixa*: indica qual é a distribuição de banda fixa que resultou no menor *throughput* e qual o ganho ao aplicar o MAB no Cenário 3; e (ii) *Melhor escolha fixa*: indica qual é a distribuição de banda fixa que resultou no maior *throughput* e qual o ganho ao aplicar o MAB. Existem ganhos significativos de 58,95% (comparando com a

TABELA II: Ganho de *throughput* com a aplicação do algoritmo proposto.

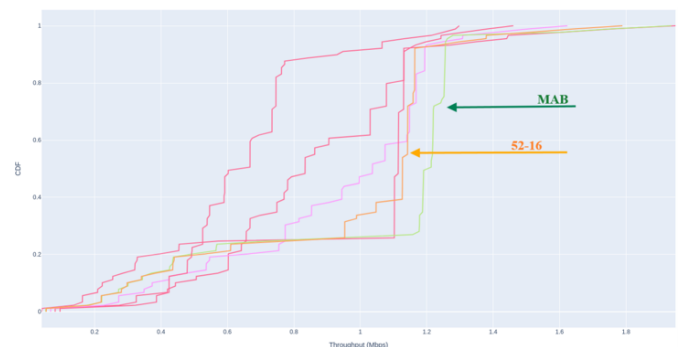
Pior Escolha Fixa					Melhor Escolha Fixa			
Taxa Oferecida	Dist. Banda	Throughput (Mbps)	Ganho do MAB (%)		Dist. Banda	Throughput (Mbps)	Ganho do MAB (%)	
RSRQThreshold 30								
2 Mbps	16-28	19,03	30,25	58,95	52-16	28,92	30,25	4,59
1 Mbps	64-12	16,62	19,78	19,01	28-24	19,15	19,78	3,28
500 kbps	64-12	10,1	13	28,71	16-28	12,9	13	0,77
RSRQThreshold 31								
2 Mbps	16-28	19,57	30,43	55,49	52-16	28,53	30,43	6,65
1 Mbps	64-12	16,25	19,65	20,92	28-24	18,98	19,65	3,53
500 kbps	64-12	9,91	12,92	30,37	16-28	12,91	12,92	0,08


 Fig. 8: Cenário 3: *Throughput* de célula - *RSRQThreshold* 30 e taxa oferecida de 1 Mbps.

 Fig. 9: Cenário 3: *Throughput* de célula - *RSRQThreshold* 30 e taxa oferecida de 500 kbps.

pior escolha fixa) e 6,65% (comparando com a melhor escolha fixa), confirmando que o algoritmo proposto tem a habilidade de fazer uma escolha de distribuição de banda autônoma e sempre vantajosa.

VI. CONCLUSÕES

Este trabalho propõe a otimização do algoritmo ICIC utilizando aprendizado por reforço. A técnica MAB ajusta autonomamente a distribuição de banda do algoritmo SFR baseado no *throughput* médio dos usuários. Uma análise sistêmica utilizando o ns-3 caracterizou o problema e demonstrou que o SFR clássico não é capaz de gerenciar a


 Fig. 10: ECDF do *throughput* - *RSRQThreshold* 30 e taxa oferecida de 2.

interferência em cenários dinâmicos com *hotspots*, chegando a uma redução de *throughput* de mais de 50% de desempenho, dependendo da localização dos *hotspots*.

A aplicação do algoritmo proposto neste trabalho resultou em ganhos de até 58,95% nos cenários avaliados.

Trabalhos futuros envolvem a avaliação de desempenho do algoritmo proposto em cenários mais desafiadores e um estudo aprofundado sobre a convergência do algoritmo, bem como a proposta de outros agentes de aprendizado (e.g., o *Q-learning*).

REFERÊNCIAS

- [1] M. Yassin, M. A. AboulHassan *et al.*, "Survey of ICIC Techniques in LTE Networks under Various Mobile Environment Parameters," *Wireless Networks*, Springer Verlag, vol. 23, no. 2, pp. 403–418, 2017.
- [2] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*. Second edition, 2nd ed. Wiley, 2011.
- [3] I. D. do Rego and J. Vicente A. de Sousa, "Solution for Interference in Hotspot Scenarios Applying Q-Learning on FFR-Based ICIC Techniques," *Sensors*, 2021.
- [4] D. Kimura, Y. Harada, and H. Seki, "De-Centralized Dynamic ICIC using X2 Interfaces for Downlink LTE Systems," *IEEE, 73rd Vehicular Technology Conference (VTC Spring)*, 2011.
- [5] H. Wu, H. Gao, and X. Xiong, "ICIC Based on Adjacent Cells Interference Elimination for Cell Edge," *IEEE, 17th International Conference on Communication Technology (ICCT)*, 2017.
- [6] I. D. do Rego, L. F. de O. Nobrega, and V. A. de Sousa Jr., "Análise de desempenho de técnicas ICIC usando o ns-3," *ECOP*, 2018.
- [7] The University of Washington NS-3 Consortium, "Network simulator(ns-3)," ns-3, Tech. Rep., 2024. [Online]. Available: <https://www.nsnam.org/docs/models/html/lte-user.html>
- [8] R. S. Sutton and A. G. Barto, *Reinforcement Learning*, 2018.
- [9] The University of Washington NS-3 Consortium. (2024) 4g/lte - phy measurement. [Online]. Available: https://sharetechnote.com/html/Handbook_LTE_RSRQ.html