

Estratégia Conjunta para Associação de Usuários e Alocação de Recursos em Redes Móveis de Próxima Geração

Matheus Alves¹, Gustavo Broechl¹, Luna Loyolla¹,
Warley Junior¹, Marcela Alves^{1,2}, Hugo Kuribayashi¹

¹Universidade Federal do Sul e Sudeste do Pará (Unifesspa)
CEP – 68.505-080 – Marabá – PA – Brasil

²Universidade do Estado do Pará (UEPA), Parauapebas – PA, Brasil.

{mathsalves, gustavo.broechl, luna.loyolla, wmvj, hugo}@unifesspa.edu.br
marcela.ad.souza@uepa.br

Abstract. *This study presents an approach based on Reinforcement Learning (RL) to optimize the orchestration of User Association and Resource Allocation (UARA) mechanisms in next-generation heterogeneous networks, focusing on maximizing user satisfaction. The proposed strategy aims to enhance the efficiency of these networks by overcoming operational challenges through adaptive algorithms centered on the user. The results suggest that the strategic application of RL algorithms can lead to significant improvements compared to traditional methods, such as Max-SINR and Cell Range Expansion (CRE), reaching over 90% user satisfaction, highlighting the relevance of this research for the future communications network context.*

Resumo. *Este estudo apresenta uma abordagem baseada em Aprendizagem por Reforço (AR) para otimizar a orquestração dos mecanismos User Association and Resource Allocation (UARA) em redes móveis heterogêneas de próxima geração, com foco na maximização da satisfação dos usuários. A estratégia proposta visa aprimorar a eficiência destas redes, superando desafios operacionais por meio de algoritmos adaptativos e centrados no usuário. Os resultados indicam que a aplicação estratégica de algoritmos de AR pode proporcionar melhorias significativas em relação a métodos tradicionais, como Max-SINR e Cell Range Expansion (CRE), atingindo mais de 90% de satisfação de User Equipments (UEs), destacando a relevância dessa pesquisa para o cenário das redes de comunicações futuras.*

1. Introdução

Nos últimos anos, tem se observado uma crescente demanda por serviços de dados de alta velocidade, impulsionada pelo aumento exponencial no consumo de dados móveis. A necessidade concomitante de garantir *Quality of Service* (QoS) confiável torna-se um desafio significativo para as operadoras de redes móveis. Existe a necessidade inevitável de lidar de forma eficiente com esse massivo volume de dados, ao mesmo tempo em que se mantêm altas taxas de transferência de dados. Diante desse cenário, a busca por inovações na gestão de recursos torna-se imperativa para atender às expectativas dos usuários e garantir uma experiência de conectividade superior [Jayaraman et al. 2023].

A gestão de recursos em redes móveis enfrenta a complexidade decorrente da demanda crescente e das expectativas dos consumidores. As operadoras estão explorando novas abordagens para otimizar a eficiência operacional, concentrando-se em processos de gerenciamento que vão além das práticas convencionais. Esses esforços têm como objetivo não apenas atender à demanda massiva por dados de alta velocidade, mas também assegurar a entrega consistente de QoS em todas as condições de operação. A otimização, nesse contexto, torna-se uma prioridade para lidar com os desafios impostos pelo aumento exponencial do tráfego de dados. Estratégias inovadoras de alocação de recursos, balanceamento de carga e adaptação dinâmica tornam-se cruciais para garantir a eficiência da rede, especialmente em momentos de pico de demanda [Zhang et al. 2019].

A literatura correlata [Alhashimi et al. 2023, Paixão et al. 2023, Zhao et al. 2019] indica que esquemas eficientes de orquestração dos mecanismos de gerenciamento e operação da rede, tornam-se determinantes ao mesmo tempo que representam desafios dada a natureza complexa de tais processos. Além disso, heterogeneidade e densificação destas redes traduzem-se na necessidade de infraestrutura interoperável e (ultra) flexível, de forma a ser capaz de orquestrar os diversos processos envolvidos.

Nesse cenário desafiador, métodos baseados em Aprendizado de Máquina (AM) emergem como uma estratégia promissora para facilitar os processos de alocação de recursos, balanceamento de carga e associação de usuários. Estas técnicas oferecem a possibilidade de aprendizado de padrões complexos e dinâmicos, permitindo uma resposta adaptativa às mudanças nas condições de rede. Porém, apesar do crescente interesse em técnicas baseadas em AM nos últimos anos, é crucial reconhecer que essas abordagens apresentam limitações em cenários específicos do mundo real.

O presente trabalho propõe uma abordagem baseada em Aprendizagem por Reforço (AR) para o ajuste conjunto dos processos de associação de usuários e alocação de recursos. Nesta perspectiva, com suas características de aprendizado contínuo e adaptabilidade, técnicas de AR oferecem uma perspectiva para enfrentar os desafios específicos dessa complexa infraestrutura de rede. Esse mecanismo pode ser projetado a partir da utilização de técnicas automatizadas para fins de elaboração de estratégias que conduzem a soluções eficientes em tempo real, em detrimento às propostas convencionais de gerenciamento, baseadas em abordagens manuais de operação ou na resolução de problemas clássicos de otimização.

Por questões de conveniência, o restante deste trabalho denota a problemática relacionada ao funcionamento conjunto dos mecanismos de associação de usuários e alocação de recursos através do acrônimo *User Association and Resource Allocation* (UARA).

1.1. Contribuições do Trabalho

O esquema proposto neste trabalho baseia-se no desenvolvimento de uma abordagem de AR, como um mecanismo geral para a orquestração dos processos UARA, de tal forma a:

- Desenvolver uma estratégia de AR baseada na maximização dos níveis de QoS dos equipamentos de usuário - *User Equipments* (UEs), apresentando promissores níveis de melhoria na satisfação destes UEs; e
- Desenvolver esquemas analíticos e técnicos voltados para a modelagem do sistema, e parcial automatização do funcionamento dos mecanismos UARA, em observância ao contexto da problemática apresentada.

1.2. Organização do Trabalho

O restante do trabalho está organizado da seguinte forma. A seção 2 apresenta os trabalhos relacionados. Na seção 3 são apresentados os fundamentos teóricos envolvidos no estudo. A seção 4 apresenta a metodologia proposta, enquanto a seção 5 descreve os experimentos aplicados e os resultados obtidos. Por fim, a seção 6 apresenta as conclusões do estudo.

2. Trabalhos Relacionados

Em [Zhao et al. 2019], os autores propõem um método de aprendizado por reforço distribuído para resolver o problema UARA em *Heterogeneous Networks* (HetNets). Diante da natureza não convexa e combinatória do problema, o método utiliza uma abordagem de aprendizado de reforço multiagente para maximizar a utilidade de downlink a longo prazo, ao mesmo tempo em que busca garantir os requisitos de QoS dos usuários. Os resultados das simulações demonstram que o método tende a superar outras abordagens de otimização em termos de capacidade do sistema e eficiência computacional, de tal forma que o método proposto oferece benefícios práticos em termos de desempenho, escalabilidade e eficiência computacional.

Os autores de [Labana and Hamouda 2020] abordam o problema de maximizar a eficiência de redes de comunicação sem fio por meio de uma abordagem de transmissão *Coordinated Multi-Point* (CoMP). O problema é resolvido por meio de uma abordagem de otimização convexa, dividida em três sub-problemas: associação de usuários, alocação de recursos e alocação de energia. Os resultados da simulação mostram melhorias significativas na eficiência da rede quando a abordagem CoMP proposta é utilizada, especialmente em cenários com excesso de recursos e links de *fronthaul* de alta capacidade.

O estudo de [Kim et al. 2023] aborda os desafios UARA em redes sem fio com alta mobilidade dos usuários. Devido à complexidade desses problemas, especialmente em aplicações como a Internet das Coisas, as soluções tradicionais enfrentam dificuldades. O estudo propõe uma abordagem de aprendizado por reforço profundo para lidar com essas questões. A solução proposta envolve a formulação do problema de otimização, onde o objetivo principal é maximizar as taxas de dados dos usuários, enquanto minimiza o número de trocas de célula (*handovers*). Para resolver este problema, os autores adotam o algoritmo *Proximal Policy Optimization* (PPO). Os resultados da simulação mostram que a abordagem proposta oferece uma taxa de convergência mais rápida e um desempenho superior na redução de trocas de célula em comparação com métodos convencionais (redução de 58%), demonstrando a aplicabilidade de métodos de AR.

Por fim, [Mahbub et al. 2021] aborda a maximização da probabilidade de associação do usuário e do número médio de equipamentos de usuário associados em HetNets. A solução proposta aborda a alocação ótima de recursos, como potência, densidade de camadas e densidade de UE, sob restrições de rede. O problema é formulado como um problema de otimização e os resultados revelam várias tendências significativas. Com o aumento da potência de transmissão da camada específica, a probabilidade de associação do UE aumenta, especialmente para um limiar de *Signal-to-Interference-plus-Noise Ratio* (SINR) mais baixo. No entanto, SINR alvos mais altos tendem a reduzir a probabilidade de associação do usuário. O estudo conclui que a abordagem proposta pode ajudar a melhorar a compreensão e otimização de HetNets.

A partir do exposto é importante destacar que parte destes trabalhos utilizam técnicas de aproximação para deduzir uma versão convexa (e simplificada) da função objetivo real do problema, para garantir a aplicação de técnicas de otimização combinatorio. Porém, nem sempre o problema real é de fato convexo, e a formulação de uma simplificação tende a gerar um problema que se distancia do problema real. Além disso, há diversos algoritmos de AR a serem explorados, seja em conjunto com CoMP, ou com o uso de multi-agentes. Desta forma, este trabalho busca endereçar algumas das lacunas existentes nestes trabalhos, como a simplificação do problema de otimização.

3. Associação de Usuários e Alocação de Recursos

HetNets possibilitam a implantação de *Small Base Stations* (SBSs), que possuem limitada potência de transmissão e cobertura. Tais SBSs podem ser organizadas em camadas, com diferentes níveis de densidade de implantação, potência de transmissão e capacidade de serviço, com vantagens relacionadas à facilidade de implantação e reduzido custo de operação e manutenção [Kuribayashi et al. 2020].

Além disso, HetNets têm o potencial de reduzir o desbalanceamento de carga entre diferentes *Base Stations* (BSs), dadas as diferentes opções de capacidade de serviço e potência de transmissão. Por outro lado, a estratégia de seleção de célula, baseada na relação SINR desbalanceia a concentração de carga ao longo da rede. UEs mesmo sob cobertura de SBSs ainda a percebem o maior sinal de *downlink* como sendo de *Macro Base Stations* (MBSs). Eventualmente mesmo que um UE esteja sob a área de cobertura de uma SBS, o SINR da MBS experimentado por este UE pode ser maior do que aquele percebido pela SBS próxima.

Em escala, isso faz com que MBSs permanecem sobrecarregadas, enquanto que as camadas compostas por SBSs permanecem ociosas, com baixo nível de UEs em atendimento. Para fazer melhor uso da infraestrutura de HetNets, UEs devem ser reassociados com SBSs menos sobrecarregadas, de modo a ter seus requisitos de QoS melhor atendidos, dada a potencial maior disponibilidade de recursos de rádio.

Por outro lado, a definição de carga em uma rede celular pode variar de acordo com o cenário ou tecnologia envolvida. No contexto de redes de acesso de rádio, a carga de uma BS pode ser definida como o número de UEs conectados a esta BS. Em consequência, ao se considerar o volume de recursos de rádio disponibilizados por uma BS como constante e o número de UEs servidos por esta BS como variável, sugere-se que o tráfego de dados experimentado por estes UEs passa a ser inversamente proporcional ao volume de UEs. Isto é, se uma BS estiver associada com muitos UEs, considerando uma alocação de recursos igualitária, os recursos de rádio são divididos igualmente, levando a reduzidas taxas de dados ao UEs [Gomez et al. 2018].

Consequentemente, mecanismos diversos para orquestração conjunta dos processos UARA para HetNets têm sido propostos, considerando métricas de desempenho como probabilidade de bloqueio/cobertura, eficiência espectral, eficiência energética, assimetria entre *downlink* e *uplink*, entre outros. Atingir um grau satisfatório de coordenação não representa uma tarefa fácil, se fazendo necessária a utilização de mecanismos inteligentes que considerem a carga de tráfego da rede, assim como condições de rede relacionadas a canal, sinal e interferência. Tais questões representam uma tarefa de complexidade considerável, dada a necessidade de operação em tempo real e a densidade de BSs.

4. Metodologia Proposta

A formulação do problema está baseada na maximização do grau de satisfação dos UEs, mais precisamente quando um UE atinge uma taxa de dados r_{ij} que contemple seus requisitos mínimos de tráfego r_i^{min} ($r_i^{min} \leq r_{ij}$). O conjunto de BSs é definido por \mathcal{B} , enquanto que o conjunto de UEs é denotado por \mathcal{U} . Além disso, considera-se que BSs são agentes orquestráveis do sistema, e podem ser utilizadas para execução de ações que façam com que os requisitos de tráfego dos UEs sejam satisfatoriamente cumpridos. Desta forma denota-se por $\Gamma_j(t)$, o grau de satisfação de todos os UEs associados à j -ésima BS, como:

$$\Gamma_j(t) = \frac{\sum_{\forall i} x_{ij}(t) \varphi_i \psi_i}{\sum_{\forall i} \psi_i}, \forall j \in \mathcal{B}, \quad (1)$$

onde o termo $x_{ij}(t)$ representa o estado de associação binário entre o i -ésimo UE e a j -ésima BS no instante t . Caso $x_{ij}(t) = 1$, há associação entre UE e BS, enquanto que o caso contrário é denotado por $x_{ij}(t) = 0$. De modo análogo, φ_i representa uma variável binária que ilustra o eventual atendimento dos requisitos de tráfego do i -ésimo UE. Quando $r_i^{min} \leq r_{ij}$, considera-se que $\varphi_i = 1$, e para o caso contrário $\varphi_i(t) = 0$.

Adicionalmente, o termo ψ_i representa um fator de ponderação no cálculo do grau de satisfação, ao se considerar UEs prioritários e comuns. Isto é, considera-se que há determinados UEs que possuem prioridade de atendimento. Tal prioridade tem efeito direto nos mecanismos de alocação de recursos, bem como no cômputo do grau de satisfação dos UEs. Em particular, considera-se a condição $\psi_i^{priority} > \psi_i^{ordinary}$, onde $\psi_i^{priority}$ denota a prioridade de UEs prioritários e $\psi_i^{ordinary}$ representa a prioridade de UEs comuns. Assim, o balanço entre UEs prioritários e comuns tem implicações no cômputo de $\Gamma_j(t)$, dado que $\Gamma_j(t)$ é diretamente proporcional ao termo $\varphi_i \psi_i$.

Desta forma, controladores de rede locais através de BSs, podem tomar ações que alterem os mecanismos UARA. E neste contexto, apesar do objetivo geral da orquestração ser o de melhorar as condições de tráfego, eventualmente um conjunto de ações podem levar à degradação do grau de satisfação dos UEs envolvidos. Assim, considera-se uma função de custo $\Phi(t)$, que busca atribuir uma penalidade em tais situações, conforme:

$$\Phi_j(t) = \begin{cases} -\Upsilon, & \text{Se } \Gamma_j(t) < \Gamma_j(t-1), \\ 0, & \text{caso contrário,} \end{cases} \quad (2)$$

onde Υ representa uma constante que atua como uma recompensa negativa, quando há degradação da satisfação dos UEs. A partir do exposto, considera-se uma função utilidade $R_j(t)$ que representa a recompensa associada a j -ésima BS no instante t , conforme:

$$R_j(t) = \sum_j \rho \Gamma_j(t) + \Phi_j(t), \quad (3)$$

onde ρ representa um fator de recompensa associada ao grau de satisfação dos UEs. Neste contexto, a recompensa obtida por um controlador de rede está baseada nas recompensas obtidas por todas as BSs associadas a este controlador. Assim, a partir da Eq. (3), um fator de ponderação das recompensas obtidas a cada instante t passa a ser definido, e finalmente

formula-se a função objetivo apresentada pela Eq. (4). Assim, o mecanismo proposto busca a maximização das recompensas \mathcal{R}_t , obtidas a longo prazo pelo controlador de rede, por meio da soma ponderada das recompensas obtidas a cada instante t , conforme:

$$\text{Maximizar } \mathcal{R}_t = \sum_t \sum_j \gamma_t R_j(t), \quad (4)$$

onde γ_t representa uma taxa de desconto que determina o peso das futuras recompensas. Assim, controladores de rede devem escolher ações que maximizem a longo prazo a função (4). Esta definição de retorno com desconto faz com que controladores valorizem mais recompensas imediatas sobre recompensas futuras, dado que γ_t está definido no intervalo $[0, 1]$. Embora o controlador considere as recompensas que espera receber no futuro, as recompensas mais imediatas têm mais influência quando se trata de tomar uma decisão sobre qual ação realizar. Por fim, a Eq. (4) está sujeita às seguintes restrições:

$$\sum_{j \in \mathcal{B}} x_{ij} = M_{BS}, \forall i \in \mathcal{U}, \quad (5)$$

$$\sum_{i \in \mathcal{U}} x_{ij} \leq M_{UE}, \forall j \in \mathcal{B}, \quad (6)$$

$$n_i^{RRB} \geq T_B, \forall i \in \mathcal{U}. \quad (7)$$

A restrição (5) garante que cada UE está associado a no máximo M_{BS} BSs simultaneamente. Por outro lado, a restrição (6) busca assegurar que cada BS atenda no máximo M_{UE} UEs ao mesmo tempo. Por fim, a restrição (7) garante que o número de *Resource Blocks* (RBs) recebidas por um único UE deve ser maior que um *threshold* mínimo T_B , de forma a emular o conceito de *transport block*.

4.1. Modelagem de Sistema

O modelo de sistema adota uma HetNet, formada por K camadas independentes de BSs. O conjunto de todas as BSs é definido por \mathcal{B} , enquanto que o conjunto de todos os UEs é denotado por \mathcal{U} . Pelo critério de associação baseado na maior potência de sinal recebida (Max-SINR), o i -ésimo UE se associa à j -ésima BS, a partir do maior valor de SINR ζ_{ij} recebido, tal que $j = \arg \max(\zeta_{ij}), \forall j \in \mathcal{B}$, conforme:

$$\zeta_{ij} = \frac{P_j^k h_{ij}}{\sum_{q \in \mathcal{B}, q \neq j} P_q^k h_{iq} + P_N}, \forall j \in \mathcal{B}, \quad (8)$$

onde P_j^k denota a potência de transmissão da BS j na camada k , h_{ij} representa o ganho de canal efetivo entre UE e BS, enquanto que P_N representa a potência de ruído térmico. A partir da Eq. (8), o cálculo da taxa de dados alcançável por canal é expresso por $e_\ell \frac{n_{sc} n_{sym}}{T_{subframe}}$, onde e_ℓ representa uma função de eficiência por subportadora em termos de bits por símbolo *Orthogonal Frequency-Division Multiplexing* (OFDM) para um determinado *threshold* SINR $e_\ell(\zeta_{ij})$. Os termos n_{sc} , n_{sym} e $T_{subframe}$ representam o número

de sub-portadoras por canal, número de símbolos OFDM e a duração de um subframe, respectivamente. Assim, a taxa de dados obtida pelo i -ésimo UE a partir da j -ésima BS pode ser computada conforme:

$$r_{i,j} = e_\ell \cdot n_{i,j}^{RB} \cdot x_{i,j} \cdot \frac{n_{sc} n_{sym}}{T_{subframe}}, \forall j \in \mathcal{B}, \quad (9)$$

onde $n_{i,j}^{RB}$ representa o quantitativo de RBs disponibilizados pela j -ésima BS ao i -ésimo UE. Adicionalmente, adota-se um esquema de alocação baseado em *Proportional Fairness*, e assume-se que os UEs são classificados em comuns e prioritários. Neste caso, UEs prioritário têm um maior peso para alocação de RBs em uma BS, do que UEs comuns. Adicionalmente, um UE comum pode ter seus requisitos mínimos de QoS relaxados, enquanto que UE prioritários devem sempre ter seus requisitos de QoS atendidos. Ainda assim, as condições de tráfego de um UE prioritário dependem diretamente da carga da BS, bem como do quantitativo de outros UE prioritários associados à mesma BS.

5. Experimentos

Para este estudo de caso, considera-se um modelo de HetNet de duas camadas, com BSs uniformemente distribuídas em um cenário de 1.0 km^2 de área, de forma a emular uma estratégia de implantação conduzida pela operadora de rede móvel. Em particular, o cenário de simulação conta com apenas 1 MBS e um quantitativo de 10 SBSs. Neste cenário de simulação são considerados até 300 UE/km^2 , cujos posicionamentos são gerados aleatoriamente por amostras independentes de uma função de Processo de Ponto de Poisson Não-Homogêneo, para emular o posicionamento real de UEs. Além disso, MBS, UEs, e SBSs possuem alturas fixas em 30.0, 1.5 e 10.0m, respectivamente. As posições de UEs são fixas e assim não há mobilidade destes UEs no cenário de simulação.

Cada UE possui um requisito de tráfego r_{min} de 2,0 Mbps, conforme metodologia utilizada em [Kuribayashi et al. 2020]. Adicionalmente, assume-se que uma proporção de 20% UEs possuem requisitos de tráfego prioritário, enquanto que os demais assumem um perfil de tráfego comum. Os pesos para o cálculo de satisfação de UEs são $\psi_i^{priority} = 1.5$ e $\psi_i^{ordinary} = 1.0$, para UEs prioritários e comuns respectivamente. Estes UEs prioritários são escolhidos de forma aleatória, entre os UEs que compõem a simulação. Além disso, considera-se um cenário de simulação com o modelo de *pathloss* e parâmetros de simulação recomendados pelo 3GPP [3GPP TR 36.814 V9.0.0 2010].

A potência de transmissão configurada no sistema é de 46.0 dBm para a MBS, 33.01 dBm para SBSs e -174.0 dBm/Hz de ruído térmico. Em adição, considera-se o esquema de codificação e modulação disponível no *Long Term Evolution* (LTE), conforme dados apresentados na Tabela 1, que apresenta os valores de eficiência e_ℓ em função dos valores de SINR ζ_{ij} , em conformidade com [Kuribayashi et al. 2020]. Os demais parâmetros de camada física são resumidos na Tabela 2.

A partir do cenário proposto, UEs realizam associação com no máximo $M_{BS} = 2$ BSs simultaneamente a partir do critério Max-SINR, enquanto que uma BS deve servir até $M_{UE} = 20$ UEs simultaneamente. Uma vez este processo concluído, computa-se a carga de cada BS e a taxa de dados obtida por cada UE (conforme Eq. 9), bem como o nível de satisfação de cada UE. Finalmente, cabe destacar que no contexto desta implementação de referência, a restrição definida pela Eq. 7 foi relaxada, de forma que $T_B = 0$.

Tabela 1. Relação de eficiência e_ℓ e SINR.

SINR \leq	-6.5	-4.0	-2.6	-1.0	1.0	3.0	6.6	10.0	11.4	11.8	13	13.8	15.6	16.8	17.6
e_ℓ	0.15	0.23	0.38	0.6	0.88	1.18	1.48	1.91	2.41	2.73	3.32	3.9	4.52	5.12	5.55

Tabela 2. Parâmetros de camada física.

Parâmetro	Valor
Frequência de Portadora	2.0 GHz
$T_{subframe}$	1 ms
n_{sc}	12
n_{sym}	14
Ganho de Antena MBS	15.0 dBi
Largura de banda de sub-canal	180 kHz
Ganho de antena de UE	0.0 dBi
Ganho de antena de SBS	5.0 dBi
<i>Path loss</i> entre MBS-UE	$128.0 + 37.6 \log_{10}(\max(d, 35)/1000)$
<i>Path loss</i> entre SBS-UE	$140.7 + 36.7 \log_{10}(\max(d, 10)/1000)$

A cada instante t , a recompensa \mathcal{R}_t de cada BS baseia-se no estado atual \mathcal{S}_m (quantidade de UEs associados as SBSs). Assim, no contexto de AR, o estado seguinte da rede assume um comportamento estocástico, uma vez que no instante $t + 1$, o novo estado da rede é influenciado pelo estado anterior e pelas ações \mathcal{A}_j . Estas ações são baseadas na aplicação de *bias Cell Range Expansion* (CRE) no intervalo contínuos de $[20.0, 80.0]$ db. Assim, a j -ésima BS determina sua próxima ação $a_{n,t}$ ($a_{n,t} \in \mathcal{A}_j$), com base no estado atual da rede $s_{s',t}$, e através de uma política π , de tal forma que $a_{n,t} = \pi(s_{s',t})$. A partir do efeito conjunto das ações de todas as BSs \vec{A} , recompensas são atribuídas às BSs, como uma função estado-ação $Q^\pi(s_{s',t}, a_{n,t})$, expressa como: $Q^\pi(s_{s',t}, a_{n,t}) = E^\pi[\sum_t^T \gamma^t \mathcal{R}_j(t|s_{s',t}, a_{n,t})]$, onde $E^\pi[\cdot]$ representa o operador de esperança. Em todas as ações selecionadas por $\pi(s_{s',t}) = \arg \max Q(s_{s',t+1}, a_{n,t+1})$, $a_{n,t} = \pi(s_{s',t})$ tende a maximizar a função $Q^\pi(s_{s',t}, a_{n,t})$.

Para aproximar uma solução para $\pi(s_{s',t})$, este trabalho adota os algoritmos de *Advantage Actor Critic* (A2C), PPO, *Deep Deterministic Policy Gradient* (DDPG), *Twin Delayed DDPG* (TD3) e *Soft Actor-Critic* (SAC), por meio da biblioteca [Raffin et al. 2021], com os parâmetros: $t = 100000$ (total de iterações), $\gamma = 0.9$ (*discount factor*), $\lambda = 0.001$ (*learning rate*). Estes algoritmos foram escolhidos devida a natureza contínua do conjunto de ações \mathcal{A}_j , excluindo portanto algoritmos que trabalham apenas com conjuntos discretos de ações como *Q-learning* e *Deep Q-Networks* (DQN), por exemplo. Além disso, cada algoritmo foi executado em um *setup* de 100 repetições, para permitir uma análise abrangente e robusta dos resultados obtidos, bem como aproximar uma distribuição estacionária dos resultados.

5.1. Resultados Numéricos

A Figura 1 complementa a descrição dos experimentos realizados neste trabalho, de modo a ilustrar o cenário de simulação. Nesta representação, UEs são identificados por círculos pretos ou vermelhos (quando prioritários), enquanto que SBSs são identificadas por círculos verdes e a MBS é representada por um círculo azul (conforme Fig. 1a).

Adicionalmente, a Figura 1b proporciona uma representação visual do estado de associação e balanceamento de carga após a implementação da métrica 3GPP baseada exclusivamente no Max-SINR. Neste contexto, UEs podem ser marcados com bordas verdes ou vermelhas. Quando um UE alcança uma taxa de dados superior ao mínimo necessário ($r_i^{min} \leq r_{ij}$), há bordas verdes sinalizando o ocorrido. Em contrapartida, círculos com bordas vermelhas representam UEs que não conseguiram atender seus requisitos mínimos de tráfego ($r_i^{min} > r_{ij}$). Assim, os círculos com bordas vermelhas são indesejáveis e evidenciam situações em que a infraestrutura do sistema não conseguiu cumprir os requisitos de tráfego específicos dos UEs.

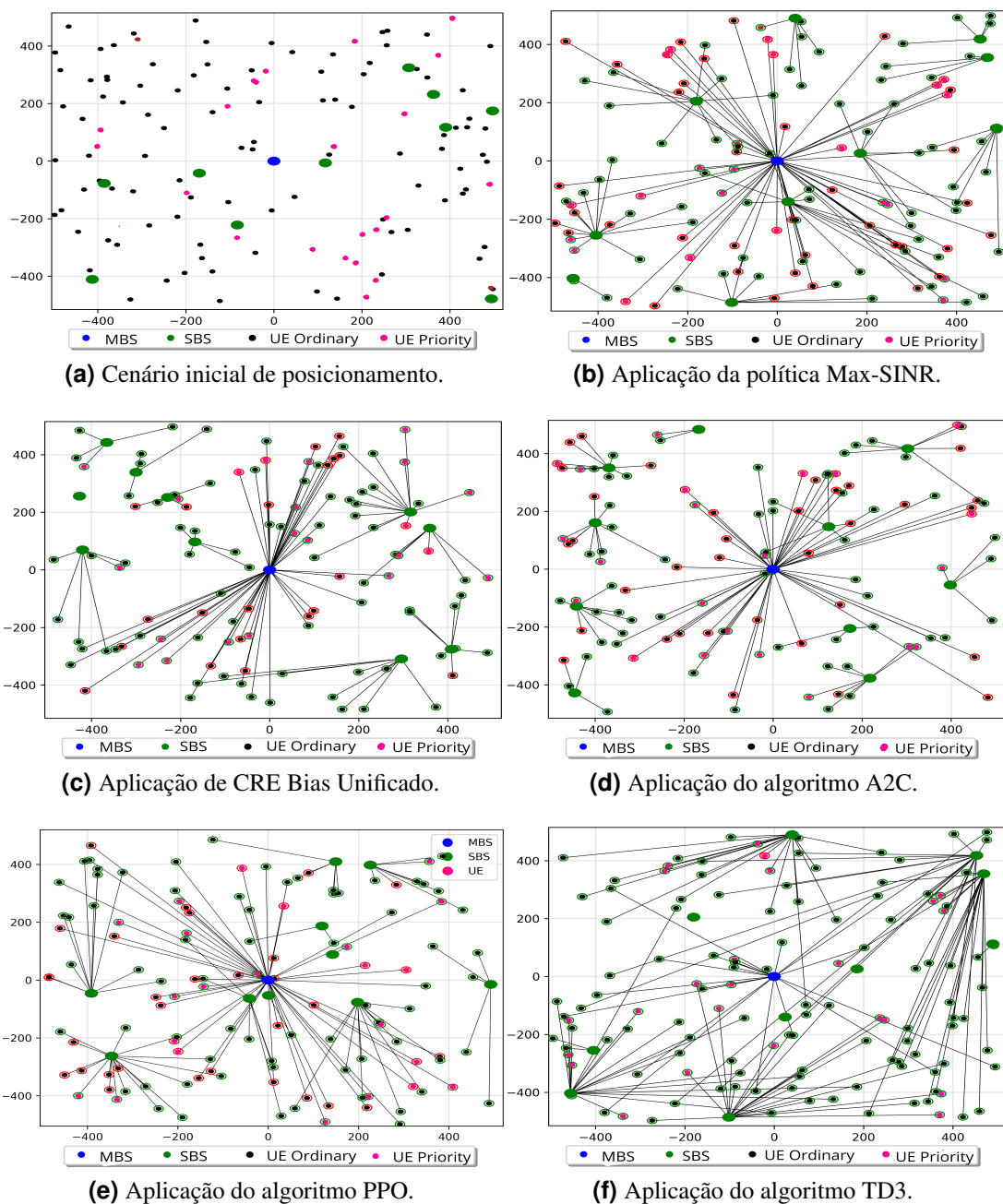


Figura 1. Cenários de distribuição e associação de usuários.

A partir dos resultados apresentados pela Figura 1b torna-se evidente a expressiva quantidade de UEs sem seus requisitos de QoS cumpridos, de forma a totalizar um valor médio de 56,10%, ao longo das 100 realizações de simulações independentes. Além disso, mesmo com a presença de SBSs próximas, é possível visualizar que alguns UEs acabam se associando com a MBS, sobrecarregando-a, e assim diminuindo a média de RBs por UE. Conseqüentemente, neste cenário Max-SINR, registra-se um percentual médio de 48,5% de UEs com requisitos de tráfego atendidos.

Por outro lado, a Figura 1c demonstra o resultado da aplicação de um Bias CRE unificado de 40.0 db a todas as SBSs. Neste tipo de cenário, as SBSs se comportam como uma camada homogênea, de tal forma a forçar a associação de UEs a esta camada, e assim, reduzindo a carga na camada MBS. Neste cenário é possível visualizar uma expressiva redução de UEs associadas com a MBS, fazendo com que na média, 79,91% tenham seus requisitos QoS atendidos, conforme dados da Figura 2a.

De modo complementar, as Figuras 1d e 1e apresentam as aplicações dos algoritmos A2C e PPO, respectivamente. É importante destacar que tais algoritmos apresentam resultados inferiores aqueles observados na estratégia de Bias CRE unificado. Mesmo com a avaliar de variados parâmetros de configuração ou ajustes nos *timesteps* de treinamento, estes algoritmos não consegue aproximar uma solução ótima para a Eq. 4, sendo inclusive as mais inferiores entre todos os algoritmos de AR. Ambos são métodos que visam melhorar a eficiência e estabilidade do treinamento ao balancear a exploração e exploração, utilizando políticas de atores e críticos para atualizar as políticas de ação de forma incremental. Esta tendência pode indicar a inviabilidade de uso destes algoritmos baseados em política no tipo de cenário proposto neste trabalho.

Por questões de conveniência, a Figura 1 não apresenta os resultados dos algoritmos SAC e DDPG, limitando-se a apresentação apenas dos resultados do TD3. Estes algoritmos compartilham várias semelhanças fundamentais, uma vez que estes são baseados em métodos de gradiente de política determinística, que visam aprender políticas determinísticas para mapear diretamente estados para ações, o que é especialmente eficaz em espaços de ação contínua (conforme a definição do conjunto \mathcal{A}_j).

Desta forma, a Figura 1f apresenta os resultados de associação para o algoritmo TD3. Neste cenário é possível verificar o expressivo quantitativo de UEs com requisitos de tráfego atendidos, e o intenso *offload* da MBS em direção às SBSs. É possível verificar que a maioria dos UEs está sendo atendida por SBSs e não pela MBS, onde, observa-se um valor médio de 99,20% de UEs com requisitos de tráfego atendidos. Apenas 15,87% (média) dos UEs estão associados com a MBS.

Em adição ao exposto, a Figura 2 apresenta os resultados de aplicação dos algoritmos de AR, em termos de satisfação e média de UEs por SBS, e um gráfico de boxplot com o percentual de satisfação dos UEs em cada um dos cenários/modelo avaliados neste trabalho. Com base na Figura 2a, é possível verificar que os esquemas que possuem níveis promissores de satisfação de UEs são aqueles que possuem as maiores médias percentuais de UEs por SBS, de tal modo a aliviar a carga da MBS.

De acordo com a Figura 2b é possível observar que a utilização de um *bias* unificado, estabelecido em 40,00 dB, resultou em taxas de satisfação situadas entre 67% e 94%. enquanto que o A2C obteve taxas de satisfação entre 71% e 87%, e PPO, taxas va-

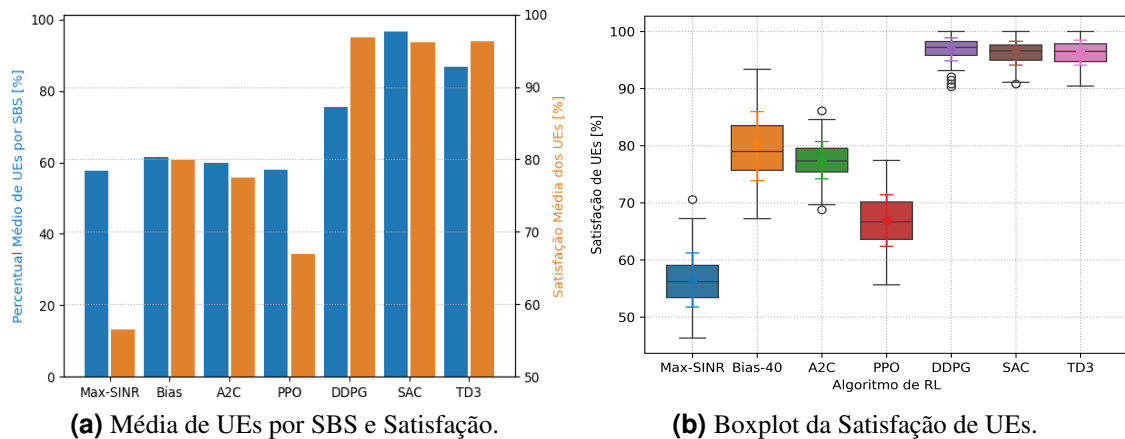


Figura 2. Análise da satisfação de UEs por algoritmo de AR.

riando entre 55% e 75%. Por outro lado, os algoritmos DDPG, SAC e TD3 destacaram-se significativamente em relação aos demais, apresentando os melhores desempenhos. Esses algoritmos demonstraram taxas de satisfação dos usuários consistentemente acima dos 90%, e, em alguns casos, atingindo até 100%. Esses resultados enfatizam a promissora eficácia de abordagens baseadas em AR, indicando seu potencial para a orquestração dos mecanismo UARA em redes de próxima geração.

Por outro lado, ao analisar-se a Figura 2b, torna-se evidente que a metodologia proposta é promissora na orquestração dos processos UARA, uma vez que todas as abordagens baseadas em AR apresentaram valores superiores de satisfação do que aqueles observados na abordagem Max-SINR. Esse aprimoramento é alcançado por meio da redução do total de associações com a MBS, possibilitando que as SBSs assumam a responsabilidade de atender um maior contingente de UEs. Ainda assim, estudos futuros devem esclarecer o motivo da baixa eficiência de algoritmos baseadas em atores e críticos.

A abordagem proposta diferencia-se substancialmente das estratégias convencionais de CRE, que geralmente adotam um *bias* unificado para camada de SBSs. Neste trabalho, a metodologia proposta realiza um ajuste de associação individualizado para cada SBSs, permitindo que UEs sejam reassociados a SBSs específicas (e menos sobrecarregadas), levando em consideração a alocação de recursos de cada BS. Consequentemente, UEs são associados a SBSs específicas e não à camadas específicas.

6. Conclusões

Este estudo empreende uma investigação sobre a aplicação de estratégias de algoritmos baseados em AR com o objetivo primordial de aprimorar significativamente a qualidade de tráfego dos usuários finais. Em particular este estudo concentra-se na capacidade adaptativa desses algoritmos, especificamente no ajuste refinado dos mecanismos UARA. Este enfoque se torna particularmente crucial ao considerar o panorama das redes de próxima geração, onde a presença cada vez mais proeminente de técnicas de AM destaca a necessidade de estratégias operacionais inovadoras.

A abordagem metodológica empregada neste trabalho visa compreender a aplicabilidade prática dos algoritmos de AR no contexto de HetNets de próxima geração.

Os resultados preliminares colhidos neste trabalho sugerem de maneira promissora que a aplicação estratégica de algoritmos de AR representam um caminho viável para superar desafios operacionais e alavancar a eficiência nas redes de próxima geração, com resultados melhores aqueles obtidos por métodos tradicionais (Max-SINR e CRE). A relevância dessa pesquisa transcende o escopo puramente teórico, destacando a necessidade premente de estratégias adaptativas e centradas no usuário no cenário dinâmico das redes de comunicações do futuro.

Agradecimentos

Os autores desejam expressar agradecimentos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Fundação Amazônia de Amparo a Estudos e Pesquisas (FAPESPA) pelo suporte financeiro parcial concedido para a realização desta pesquisa.

Referências

- 3GPP TR 36.814 V9.0.0 (2010). Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9). *3GPP*.
- Alhashimi, H. F. et al. (2023). A Survey on Resource Management for 6G Heterogeneous Networks: Current Research, Future Trends, and Challenges. *Electronics*, 12(3).
- Gomez, C. A., Shami, A., and Wang, X. (2018). Machine Learning Aided Scheme for Load Balancing in Dense IoT Networks. *Sensors*, 18(11).
- Jayaraman, R. et al. (2023). Effective Resource Allocation Technique to Improve QoS in 5G Wireless Network. *Electronics*, 12(2).
- Kim, D. U. et al. (2023). Resource Allocation and User Association Using Reinforcement Learning via Curriculum in a Wireless Network with High User Mobility. In *2023 International Conference on Information Networking (ICOIN)*, pages 382–386.
- Kuribayashi, H. P. et al. (2020). Particle Swarm-Based Cell Range Expansion for Heterogeneous Mobile Networks. *IEEE Access*, 8:37021–37034.
- Labana, M. and Hamouda, W. (2020). Joint User Association and Resource Allocation in CoMP-Enabled Heterogeneous CRAN. In *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pages 1–6.
- Mahbub, M. et al. (2021). Maximizing the Probability of User Association of a Tier of a Multi-Tier Heterogeneous Network by Optimal Resource Allocation. In *2021 Emerging Technology in Computing, Commun. and Electronics (ETCCE)*, pages 1–6.
- Paixão, E. R. et al. (2023). Multilayer Framework for Resource Orchestration in Next Generation Networks. *Journal of Communication and Information Systems*, 38:1–8.
- Raffin, A. et al. (2021). Reliable Reinforcement Learning Implementations. *Journal of Mach. Learning Research*, 22(268):1–8.
- Zhang, L. et al. (2019). 6G Visions: Mobile Ultra-broadband, Super Internet-of-Things and Artificial Intelligence. *China Communications*, 16(8):1–14.
- Zhao, N. et al. (2019). Deep Reinforcement Learning for User Association and Resource Allocation in Heterogeneous Cellular Networks. *IEEE Transactions on Wireless Communications*, 18(11):5141–5152.