Fair Max Rate: Um Escalonador de Recursos Baseado em Aprendizado por Reforço para Redes 5G

Diego Canizio Lopes[1,4], André Nasserala[2], Ian Vilar Bastos[3], Igor M. Moraes[4]

[1]Instituto Federal do Acre (IFAC)
[2]Universidade Federal do Acre (UFAC)
[3]Universidade do Estado do Rio de Janeiro (UERJ)
[4]Laboratório MídiaCom, TCC-PGC-IC, Universidade Federal Fluminense (UFF)

diego.lopes@ifac.edu.br, andre.nasserala@ufac.br,
ian.bastos@eng.uerj.br, igor@ic.uff.br

Abstract. *This paper proposes a resource scheduler for 5G networks based on reinforcement learning that maximizes network throughput by allocating resources according to the spectral efficiency of each user device. The proposed scheduler ensures a minimum throughput for each device, balancing efficiency and fairness. The evaluation is conducted through simulations based on throughput and Jain's Index metrics. The performance of the proposed scheduler is compared to the classic algorithms Round Robin (RR), Proportional Fair (PF), and Max Rate (MR), achieving an increase of up to 27.27% in throughput compared to RR and better fairness indices compared to MR.*

Resumo. *Este artigo propõe um escalonador de recursos para redes 5G baseado em aprendizado por reforço, visando maximizar a vazão da rede ao distribuir recursos conforme a eficiência espectral de cada dispositivo de usuário. O escalonador proposto assegura uma vazão mínima para cada dispositivo, equilibrando eficiência e justiça. A avaliação se dá através de simulação com base nas métricas de vazão e Índice de Jain. O desempenho do escalonador proposto é comparado aos algoritmos clássicos* Round Robin *(RR),* Proportional Fair *(PF) e* Max Rate *(MR), e, com ele, se obtém um aumento de até 27,27% na vazão quando comparado ao RR e com melhores índices de justiça quando comparado ao MR.*

1. Introdução

A quinta geração de redes móveis (5G) provê suporte para aplicações que exigem baixa latência, alta confiabilidade e eficiência energética, tais como Internet das Coisas (IoT), veículos autônomos e realidade estendida (XR) [Agiwal et al. 2016, Holma et al. 2015]. No entanto, o aumento na densidade de dispositivos conectados e a diversidade de requisitos de qualidade de serviço (*Quality of Service* - QoS) impõem desafios à alocação eficiente e justa de recursos de rádio. O escalonamento é essencial para equilibrar a eficiência espectral e a justiça na distribuição de recursos entre dispositivos de usuário (*user devices* - UEs) com diferentes demandas e características, evitando situações de inanição (*starvation*), em que um ou mais UEs fiquem sem receber nenhum recurso [Vihriälä et al. 2016]. Escalonadores clássicos, como *Round Robin* (RR), *Proportional Fair* (PF) e *Max Rate* (MR), demonstram limitações em equilibrar eficiência e

justiça, especialmente em cenários heterogêneos e de alta carga. Esses escalonadores geralmente falham em garantir uma vazão mínima para todos os UEs ou em maximizar o uso eficiente dos recursos de rádio disponíveis [Saraiva et al. 2020]. Assim, é preciso desenvolver soluções avançadas para lidar com o escalonamento de recursos em redes 5G [Rodrigues et al. 2022].

A aplicação de aprendizado por reforço (*reinforcement learning* - RL), é uma abordagem promissora para resolver o problema do escalonamento de recursos de rádio em redes 5G. Em redes móveis, técnicas de aprendizado por reforço permitem que os escalonadores aprendam padrões de uso e demanda em tempo real, ajustando dinamicamente a divisão de recursos e otimizando múltiplos objetivos, como vazão e justiça [Zhang et al. 2020, You et al. 2019]. Este artigo propõe um escalonador de recursos baseado em aprendizado por reforço chamado *Fair Max Rate* (FMR). O mecanismo proposto é projetado para maximizar a vazão e, simultaneamente, aumentar a justiça na divisão de recursos de rádio. O FMR aprende dinamicamente a alocar recursos a cada *slot* de tempo, assegurando uma vazão mínima para todos os dispositivos e evitando situações de inanição. A avaliação do escalonador é conduzida através de simulações, considerando métricas de vazão individual e agregada, bem como o Índice de Jain para medir a equidade na distribuição de recursos entre os dispositivos de usuários. Os resultados mostram que o FMR alcança melhorias de até 27,27% na vazão em comparação ao algoritmo clássico RR, mantendo melhores índices de justiça em relação ao MR.

O restante deste artigo está organizado da seguinte forma. A Seção 2 aborda os princípios de divisão de recursos de rádio em redes 5G. A Seção 3 detalha a proposta do escalonador baseado em aprendizado por reforço e a Seção 4 apresenta os resultados obtidos com base nas métricas definidas. Na Seção 5 são discutidos trabalhos relacionados à utilização de inteligência artificial em escalonadores de recursos. Por fim, a Seção 6 traz as conclusões do trabalho.

2. Alocação de Recursos de Rádio em Redes 5G

A alocação de recursos de rádio em redes 5G é mais complexa do que em redes de gerações anteriores, devido à necessidade de atender simultaneamente a cenários heterogêneos de aplicação. O conceito de numerologia desempenha um papel central nas redes 5G, pois define os parâmetros operacionais que regulam o uso do espectro e os recursos disponíveis. A numerologia está diretamente associada ao espaçamento entre subportadoras (*Subcarrier Spacing* - SCS), que varia de 15 kHz a 240 kHz no padrão 5G *New Radio* - NR. Cada configuração de SCS afeta diretamente a duração dos símbolos OFDM (*Orthogonal Frequency-Division Multiplexing*) e, consequentemente, a largura de banda utilizável por canal [3GPP 2020a].

Os diferentes valores de SCS permitem a coexistência de múltiplos serviços na mesma banda, otimizando o uso do espectro para cenários que vão desde comunicações com latência ultra-baixa (URLLC) até serviços de banda larga aprimorada (eMBB). Um SCS mais largo reduz a duração dos símbolos, aumentando a capacidade de suportar comunicações em cenários de alta mobilidade, enquanto SCS menores favorecem a eficiência espectral em cenários de baixa mobilidade. Cada combinação de SCS e largura de banda resulta em uma determinada quantidade de blocos de recursos (*Resource Blocks* - RBs) disponíveis para alocação [Dahlman et al. 2020]. Os RBs são definidos

como a menor unidade alocável em termos de frequência e tempo. Eles consistem em um conjunto fixo de subportadoras por uma duração definida, geralmente de um *slot*. A quantidade total de RBs em uma largura de banda específica depende da numerologia adotada. Por exemplo, em uma largura de banda de 10 MHz com SCS de 15 kHz, existem 53 RBs disponíveis, enquanto com SCS de 30 kHz, o número de RBs disponíveis é reduzido para aproximadamente 25 RBs [Andrews et al. 2014]. A alocação de recursos também é gerida por parâmetros estabelecidos no *Downlink Control Information* (DCI), que informa aos dispositivos como acessar os recursos disponíveis. O DCI especifica os RBs alocados, os valores de modulação e codificação (*Modulation and Coding Scheme* - MCS) e outros aspectos técnicos essenciais. O MCS, em particular, é fundamental para determinar a eficiência espectral, pois define a taxa de bits transmitidos por símbolo OFDM, variando conforme as condições do canal.

No contexto dos algoritmos de escalonamento, o escalonador determina dinamicamente o número de RBs (*Resource Blocks*) a serem atribuídos a cada UE dentro de um *slot* de tempo, visando otimizar o uso dos recursos de rádio, em termos de eficiência espectral, justiça e qualidade de serviço. Essa divisão de recursos é tipicamente realizada por meio de abordagens tradicionais como *Round Robin* (RR), *Proportional Fair* (PF) e *Max Rate* (MR). O RR aloca recursos de forma cíclica. Um RB é atribuído a cada UE em uma rodada do algoritmo de escalonamento, até que os RBs disponíveis em um *slot* se esgotem. Assim, garante-se que todos os UEs recebam uma parcela justa dos recursos. Porém, sem considerar as condições do canal, o resultado com o RR é uma baixa eficiência em termos da vazão agregada da rede. O PF busca balancear eficiência e equidade, alocando RBs com base na proporção entre a taxa de transmissão atual de um UE e sua taxa média de transmissão desde que está ativo, levando em consideração tanto a demanda individual quanto as condições de canal. Com o PF, um UE que tem maior razão entre a taxa de transmissão e a taxa média de transmissão desde que está ativo, receberá mais RBs em um *slot* de tempo do que um UE que tenha menor razão. Por fim, o MR prioriza UEs com maiores MCSes para maximizar a vazão agregada, embora isso possa levar à inanição de UEs com MCS mais baixo, caso não haja mecanismos de garantia de justiça. Esse comportamento ocorre porque UEs com MCS mais alto podem modular mais bits por símbolo e, portanto, utilizam os RBs com maior eficiência espectral. Assim, em determinados *slots*, o escalonador pode alocar todos os RBs ao UE com o maior MCS para maximizar a vazão agregada da rede, deixando outros UEs sem recursos.

3. O Escalonador Proposto *Fair Max Rate* (FMR)

O escalonador FMR utiliza aprendizado por reforço para otimizar a alocação de recursos em redes 5G. O objetivo do FMR é equilibrar a vazão e a justiça na distribuição de recursos entre os UEs. Diferente dos escalonadores tradicionais, que seguem políticas de alocação predefinidas e fixas ao longo do tempo, o FMR aprende uma política de alocação dinâmica a partir da interação com o ambiente, ajustando sua estratégia de divisão de RBs conforme as condições da rede. Para isso, o FMR implementa um agente de aprendizado por reforço, que é o responsável por tomar as decisões de alocação de recursos no escalonador proposto. O agente do FMR observa o estado atual da rede, escolhe uma ação que define a alocação de recursos e recebe um retorno na forma de recompensa. Ao longo do tempo, esse processo permite que o agente do FMR refine sua política de alocação, aprendendo a distribuir os recursos de maneira eficiente e adaptativa.

O funcionamento do FMR se baseia na priorização da eficiência espectral e na distribuição equilibrada dos RBs entre os UEs. Inicialmente, o FMR identifica o UE com maior MCS, pois este pode modular mais bits por símbolo e utilizar os RBs com maior eficiência espectral. Em seguida, os recursos são alocados prioritariamente para este dispositivo, mas sem ignorar completamente os demais. Diferente do MR, que pode levar à inanição de dispositivos com MCSes mais baixos, o FMR incorpora restrições à sua política de alocação para garantir que todos os UEs recebam um mínimo de recursos. Assim, ao mesmo tempo em que prioriza a eficiência espectral, o agente do FMR evita que UEs com MCSes mais baixos fiquem sem recursos, promovendo um balanceamento entre vazão agregada e justiça. O FMR, portanto, atua de maneira diferente dos algoritmos clássicos. Comparado ao MR, o FMR evita a inanição de UEs com MCSes mais baixos, mantendo um nível mínimo de serviço para todos os UEs. Em relação ao PF, o FMR aumenta a vazão agregada, pois não se baseia apenas no histórico de taxas de transmissão dos UEs, mas aprende a ajustar a alocação de RBs conforme as condições do canal.

O agente do FMR é implementado da seguinte forma. O espaço de estados do agente do FMR é representado como um vetor $\mathbf{s} \in \mathbb{Z}^n$, no qual n é o número total de UEs na rede. Cada componente do vetor $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_n(t)]$ corresponde à quantidade de recursos (RBs) alocados a um dispositivo específico no instante t. Cada valor s_i no vetor varia entre 0 e o número total de RBs disponíveis na rede, denotado como S_{total}, ou seja, $s_i \in [0, S_{\text{total}}]$. A representação do estado permite que o agente do FMR tenha visibilidade completa sobre a alocação de recursos para todos os UES, garantindo que ele possa tomar decisões a respeito da redistribuição de recursos.

O espaço de ação é um vetor de valores contínuos que representam a distribuição de recursos entre os UEs. Cada valor dentro do vetor é um número real no intervalo $[-1, 1]$ e reflete um peso associado ao UE. Esses pesos são utilizados para determinar a alocação relativa de recursos, que pode ser determinada utilizando a função softmax. Trata-se de uma função não-linear que transforma os pesos em uma distribuição de probabilidade, assegurando que a soma dos valores resultantes seja igual a 1. Essa operação é representada pela seguinte expressão:

$$\hat{p}_i(t) = \frac{e^{p_i(t)}}{\sum_{j=1}^{N} e^{p_j(t)}}, \tag{1}$$

na qual $p_i(t)$ é o peso do UE i no instante t, $\hat{p}_i(t)$ é o peso normalizado no instante t e N é o número total de UEs. Após essa transformação, o vetor resultante de probabilidades $\hat{p}(t) = [\hat{p}_1(t), \hat{p}_2(t), \ldots, \hat{p}_N(t)]$ é multiplicado pelo número total de recursos disponíveis RB, gerando a alocação final de recursos A, representada na Equação 2.

$$A_i(t) = \hat{p}_i(t) \cdot RB \quad \text{para todo} \quad i \in \{1, 2, \ldots, N\}. \tag{2}$$

A escolha de utilizar valores contínuos e a normalização permite ao agente do FMR aprender a priorizar UEs de forma proporcional, dependendo das condições de rede observadas e das recompensas acumuladas, ajustando a alocação para maximizar a vazão agregada e garantir que todos os UEs recebam alguma parte de recursos em um instante de tempo.

Para interagir com o agente do FMR, a função de recompensa é baseada na vazão agregada da rede. O objetivo da função de recompensa é maximizar a vazão agregada enquanto assegura que os UEs com menor eficiência espectral não sejam negligenciados. Para isso, a recompensa é calculada da seguinte forma:

Recompensa por vazão: A vazão agregada da rede é comparada com uma vazão agregada máxima, calculada com base no número total de RBs e na eficiência espectral do UE com maior MCS. A recompensa (R) é dada pela razão entre a vazão agregada obtida, denotada por V_{obt}, e a vazão agregada máxima, V_{max}, conforme a Equação 3.

$$R(t) = \max\left(\frac{V_{obt}(t)}{V_{max}(t)}, 0\right), \tag{3}$$

na qual V_{obt} é a vazão agregada obtida no tempo t através da divisão de recursos pelo agente, calculada pela soma da vazão de todos os UEs, e V_{max} é a vazão máxima para o tempo t, que pode ser calculada com base no número total de recursos (RB) e na eficiência espectral máxima (η_{max}) alcançável para o UE com maior eficiência espectral, isto é:

$$V_{max}(t) = RB \cdot \eta_{max}(t). \tag{4}$$

Penalidade para dispositivos com baixa vazão: Para incentivar uma distribuição justa de recursos, é aplicada uma penalidade sempre que algum dispositivo (UE) obtiver uma vazão abaixo de um valor mínimo predefinido V_{min}. Matematicamente, essa penalidade é calculada pela soma dos dispositivos cuja vazão não alcança esse limiar mínimo:

$$P(t) = \sum_{i=1}^{n} \mathbb{I}\left(V_i(t) < V_{min}\right), \tag{5}$$

na qual $V_i(t)$ é a vazão do UE i no instante de tempo t, e $\mathbb{I}$ é a função indicadora definida como:

$$\mathbb{I}(x) = \begin{cases} 1, & \text{se } x \text{ for verdadeiro} \\ 0, & \text{caso contrário,} \end{cases} \tag{6}$$

Dessa forma, para cada dispositivo que tiver vazão abaixo de V_{min} a soma de $P(t)$ aumentará em uma unidade. Caso exista pelo menos um dispositivo com vazão insuficiente (ou seja, $P(t) > 0$), a penalidade final aplicada será dez vezes o valor obtido em $P(t)$. O valor dessa penalidade é, então, subtraído diretamente do valor total acumulado da recompensa R, resultando na recompensa final ajustada que orientará o treinamento do agente.

A política de alocação do FMR visa não apenas maximizar a vazão agregada da rede, mas também tornar a divisão de recursos entre os UEs mais justa, de forma que todos os UEs se beneficiem de maneira proporcional, considerando cada eficiência espectral, evitando que alguns UEs fiquem com uma alocação excessivamente baixa de RBs.

O agente do FMR prevê a alocação de recursos com base nas observações dos MCSes e na função de recompensa, garantindo eficiência na distribuição e evitando a inanição dos UEs. Para isso, utiliza-se o algoritmo *Proximal Policy Optimization* (PPO),

método de aprendizado por reforço adequado para espaços contínuos, diferente de abordagens tradicionais como Q-Learning, restritas a tabelas estado-ação discretas. A escolha do PPO ocorre por sua estabilidade em cenários dinâmicos, como redes móveis, e pela capacidade de adaptação contínua às variações das condições da rede.

O aprendizado ocorre por meio da interação do agente FMR com o ambiente, coletando trajetórias compostas por estados, ações, recompensas e transições. O PPO utiliza essas trajetórias para calcular a função de vantagem A_t,que estima o impacto de uma ação em relação à média das ações possíveis naquele estado. Se a vantagem for positiva, a política aumenta a probabilidade daquela ação; caso contrário, reduz essa probabilidade. Para evitar mudanças bruscas e garantir estabilidade, o PPO aplica um mecanismo de *clipping* na função objetivo, restringindo a razão entre a nova e a antiga política dentro de um intervalo específico. Esse mecanismo limita atualizações agressivas e impede grandes desvios da política atual, permitindo ao agente explorar diferentes estratégias de alocação e otimizar a vazão agregada sem comprometer dispositivos com menor eficiência espectral. Além disso, a incorporação da entropia na função objetivo incentiva a exploração contínua do espaço de políticas, evitando que o agente FMR convirja prematuramente para estratégias subótimas. Esse comportamento é definido na função de custo combinada:

$$L_t^{CLIP+VF+S}(\theta) = \hat{\mathbb{E}}_t \left[L_t^{CLIP}(\theta) - c_1 L_t^{VF}(\theta) + c_2 S[\pi_\theta](s_t) \right]. \tag{7}$$

A função de custo combinada, expressa na Equação 7, desempenha um papel crucial no PPO, combinando três componentes essenciais para otimizar a política de aprendizado. O termo $L_t^{CLIP}(\theta)$, central para a função de custo, garante a otimização da política, evitando mudanças drásticas que possam prejudicar a estabilidade do aprendizado. Esse termo utiliza a razão de probabilidade entre a política atual e a anterior, mantendo-a dentro de limites para controlar as atualizações da política. Quando a distribuição de probabilidade se torna altamente concentrada em uma única ação, a entropia diminui, indicando que a política se tornou mais determinística. O termo $L_t^{VF}(\theta)$ aprimora a precisão da função de valor, crucial para estimar a qualidade das ações tomadas pela política, calculando o erro quadrático entre o valor estimado e o valor alvo. Já o termo de entropia, $S[\pi_\theta](s_t)$, incentiva a exploração, um aspecto fundamental do aprendizado por reforço. A entropia, neste contexto, mede a aleatoriedade das ações escolhidas pela política, e uma política com alta entropia é mais aleatória, explorando mais o ambiente e evitando soluções subótimas. Em outras palavras, o bônus de entropia encoraja a política a experimentar diversas ações, em vez de se fixar em uma única estratégia. Os coeficientes c_1 e c_2 ajustam a importância dos termos $L_t^{VF}(\theta)$ e $S[\pi_\theta](s_t)$ na função de custo total, respectivamente. Em conjunto, esses três termos otimizam a política, aprimoram a precisão da função de valor e incentivam a exploração, resultando em um aprendizado mais eficiente e robusto para o agente do FMR.

4. Resultados e Discussões

Nesta seção, são apresentados os resultados das simulações realizadas para avaliar o desempenho do escalonador FMR, em cenários típicos de redes 5G voltados para serviços eMBB. O desempenho do FMR é comparado com algoritmos clássicos RR, PF e MR. As métricas usadas são a vazão por UE, a vazão agregada e o Índice de Jain [Jain et al. 1984]

para medir justiça. Tal índice é uma métrica utilizada para analisar a equidade na alocação de recursos ou taxas de utilização entre múltiplos usuários em sistemas de comunicação e é definido como:

$$I = \frac{\left(\sum_{m=1}^{M} R_m\right)^2}{M \cdot \sum_{m=1}^{M} R_m^2}, \tag{8}$$

em que M é o número total de usuários e R_m é a taxa de utilização ou alocação de recursos, com base na quantidade de RBs atribuídos, do m-ésimo usuário. O índice de Jain varia entre 0 e 1. Assim, valores próximos a 1 indicam uma alta justiça e valores próximos a 0 indicam uma baixa justiça, com disparidades significativas entre as taxas dos usuários.

O cenário de simulação é composto por uma estação base e três UEs. A topologia da rede é simplificada para uma célula única, com *beamforming* único, e alocação de recursos realizada por *slot* para o canal de *downlink*, considerando uma comunicação do tipo OFDM. Considera-se uma largura de banda (*bandwidth part* - BWP) de 10 MHz e a numerologia μ=1, definindo o espaçamento entre subportadoras como 30 kHz. Com essa configuração, existem 25 RBs para alocação entre os três UEs em um *slot* de tempo.

São usados os dados de MCS do conjunto público de dados da plataforma de testes O-RAN Colosseum [Bonati et al. 2021]. Esse conjunto de dados representa um cenário urbano denso, no qual os UEs apresentam mobilidade aleatória, proporcionando condições de canal dinâmicas e realistas. Dessa forma, os valores de MCS usados na simulação deste artigo são representativos de um ambiente 5G prático, no qual os UEs competem por recursos de rádio sob variações estocásticas do canal. Para as simulações, é necessária uma etapa de pré-processamento do conjunto de dados de MCS com a remoção das entradas contendo valores nulos ou indefinidos. Após isso, foram escolhidos de forma aleatória 3 UEs do conjunto de dados do Colosseum, cujos MCSes são apresentados na Figura 1. Observa-se que o UE 3 possui o maior valor de MCS, indicando uma maior eficiência espectral, conforme definido pela Tabela 5.1.3.1-2 da especificação 3GPP 38.214 [3GPP 2020b]. Além disso, a Figura 1 mostra que os UEs 1 e 2 apresentam valores de MCS semelhantes entre si e inferiores ao do UE 3, caracterizando diferentes condições de canal experimentadas por cada dispositivo ao longo do período analisado.

As simulações são realizadas em um ambiente Python, versão 3.11, combinado com as bibliotecas JAX 0.4.6 e Sympy 1.13. O ambiente recebe como entrada a BWP, a numerologia e os MCSes de cada UE em cada instante da simulação. A implementação dos algoritmos RR, PF e MR é baseada nas implementações desses algoritmos para o simulador NS-3, versão 3.41, combinado com o módulo 5G Lena [Patriciello et al. 2019]. O escalonador proposto usa como base a implementação de um dos escalonadores tradicionais, porém a política de alocação de RBs é feita pelo agente de aprendizado por reforço. Para implementar o agente, é utilizada a biblioteca Stable-Baselines3 com suporte adicional do pacote sb3-contrib.

4.1. Treinamento do Agente do FMR

Os primeiros resultados apresentados referem-se ao treinamento do agente. O treinamento é realizado em um cenário simulado com um total de 204.800 passos de tempo (*timesteps*),

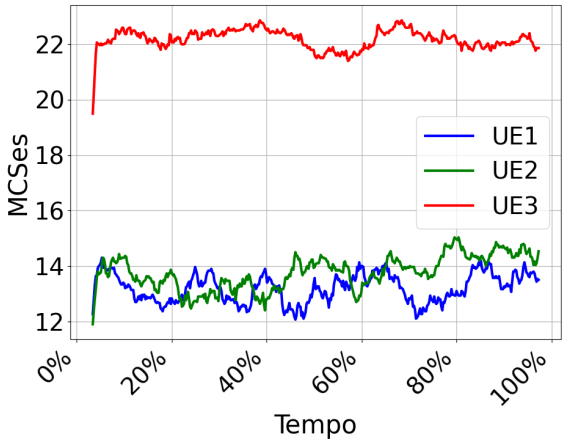


Figura 1. MCSes dos três UEs usados na simulação e obtidos no conjunto de dados do Colosseum [Bonati et al. 2021].

distribuídos em 100 episódios. Nesse contexto, um passo representa uma unidade de tempo na simulação, em que o agente toma uma decisão com base no estado atual do ambiente, e um episódio corresponde a uma sequência completa que termina quando o treinamento alcança a quantidade de 2048 passos. Durante essa etapa, a base de dados contendo os valores de MCSes é dividida em 80% para treinamento e 20% para teste, garantindo que o agente seja testado como dados não vistos por ele previamente.

A Figura 2 apresenta o resultado da recompensa acumulada durante os episódios do treinamento. O gráfico da função de recompensa acumulada demonstra iniciar uma convergência a partir de, aproximadamente, 40% do treinamento. Esse comportamento evidencia que o agente foi capaz de aprender uma política de alocação estável, maximizando a vazão e mantendo um serviço mínimo para todos os UEs.

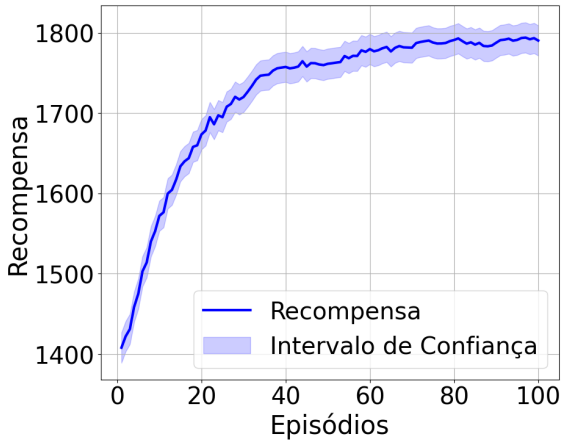
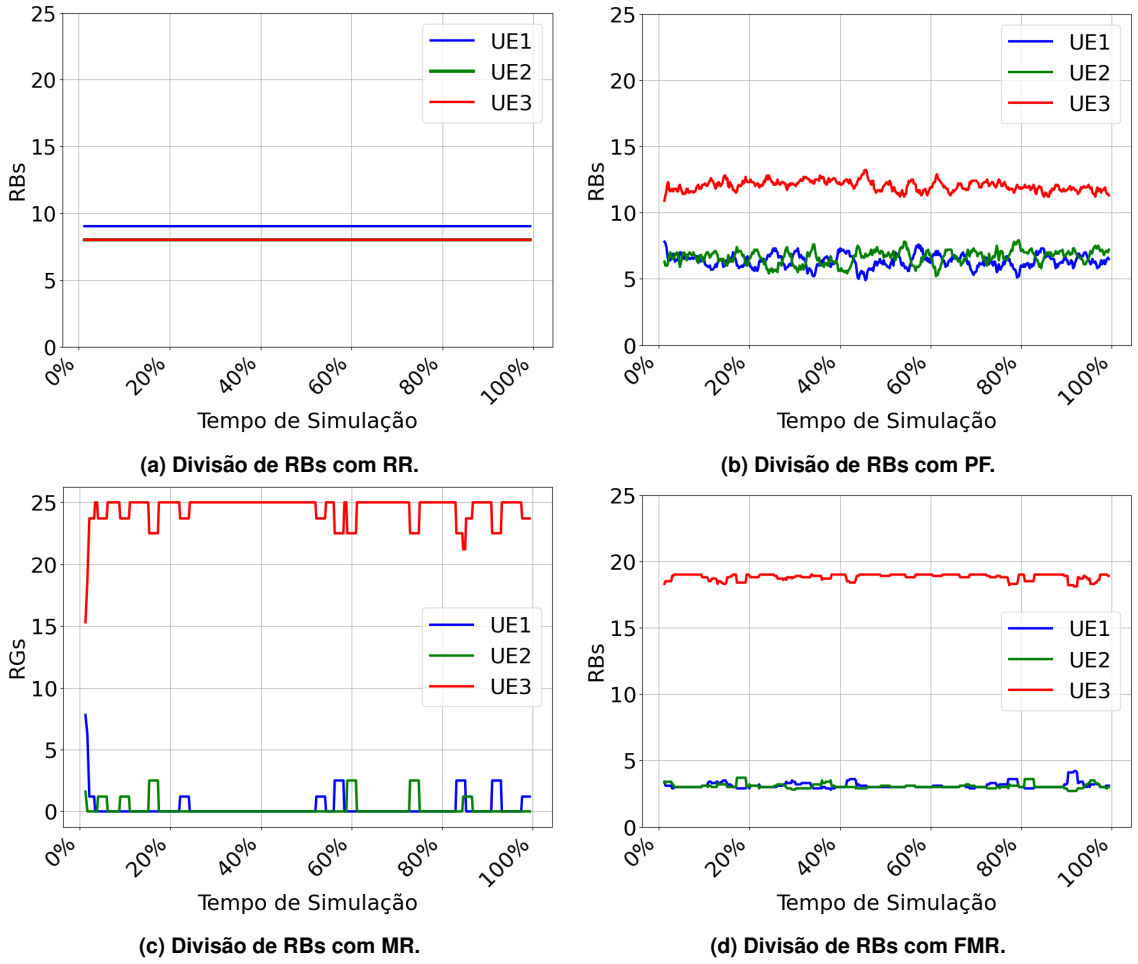


Figura 2. Função de recompensa acumulada do agente do FMR.

4.2. Avaliação do Agente do FMR

A Figura 3 apresenta os resultados da divisão de RBs realizada pelos escalonadores avaliados para uma rodada de simulação. O eixo Y indica a quantidade de RBs alocadas para cada um dos três UEs e o eixo X a porcentagem do tempo de simulação decorrido.



(a) Divisão de RBs com RR.

(b) Divisão de RBs com PF.

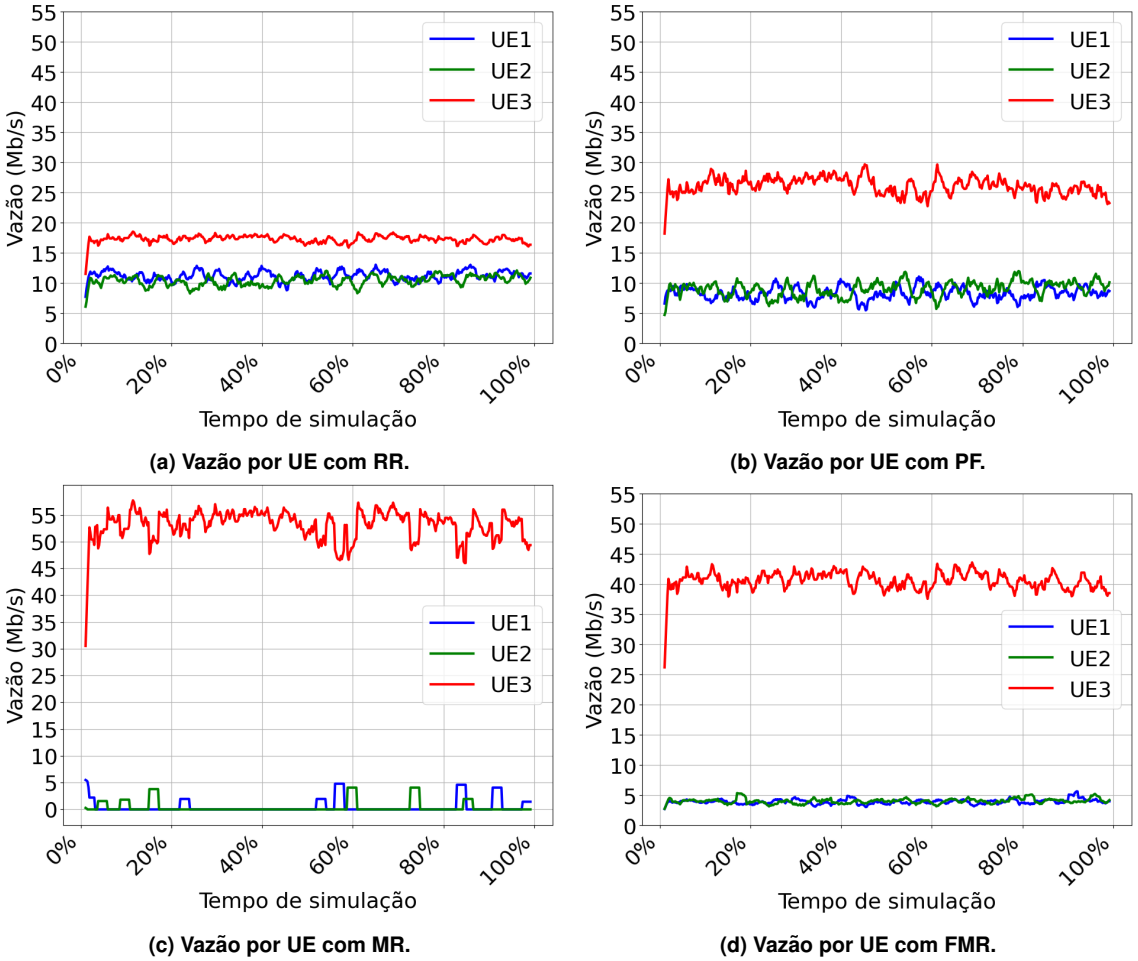
(c) Divisão de RBs com MR.

(d) Divisão de RBs com FMR.

Figura 3. Comparação da divisão de RBs entre escalonadores.

O algoritmo RR, Figura 3a, distribui os recursos uniformemente, resultando em 9 R Bs alocados ao UE 1 e 8 RBs a cada um dos UEs 2 e 3 durante toda a simulação. O Índice de Jain do RR é de 0,9247, evidenciando alta justiça na divisão de recursos. O PF (Figura 3b) equilibra eficiência e justiça, maximizando a vazão total com distribuição justa. Nesse cenário, o UE3 recebeu mais recursos, enquanto os UEs1 e 2 alternaram menores porções de RBs.. O Índice de Jain do PF é de 0,7362, refletindo um equilíbrio moderado entre eficiência e justiça. Com o MR, Figura 3c o UE 3, devido ao seu maior valor de MCS e, consequentemente, maior eficiência espectral, monopolizou a maior parte dos RBs durante a simulação. Como consequência, o UE 1 não obteve atribuição de recursos em 97,66% dos *slots*, enquanto o UE 2 não teve recursos alocados em 98,30% dos *slots*. Esse escalonador maximiza a vazão total, mas negligencia dispositivos com menores valores de MCS. O Índice de Jain do MR é 0,3411, refletindo baixa justiça na divisão de recursos. Por fim, com o FMR, Figura 3d, embora o UE 3 tenha recebido uma porção maior dos recursos devido à sua maior eficiência espectral, os UEs 1 e 2 não foram negligenciados. Em todos os *slots*, cada UE recebeu uma parcela de recursos suficiente para atender aos critérios mínimos de vazão estabelecidos pela função de recompensa. O Índice de Jain para o FMR é de 0,4696, indicando uma justiça intermediária, mas com a garantia de que nenhum dispositivo sofreu com inanição no processo de divisão.

A Figura 4 demonstra a vazão individual alcançada por cada UE com os escalonadores avaliados. Como consequência da divisão de recursos, o RR, Figura 4a, obteve as menores taxas de vazão para cada UE, devido à divisão uniforme dos RBs. O PF, Figura 4b, aumenta ligeiramente a vazão do UE de maior MCS ao passo que reduz a vazão dos UEs de menor MCS, em relação ao RR por ter um índice de justiça menor do que o do RR. Com o MR, Figura 4c, a vazão do UE de maior MCS é a maior entre todas a vazões fornecidas pelos escalonadores. O UE 3 recebe a maior parte dos recursos, pois possui alta eficiência espectral. Por outro lado, os UEs 1 e 2 sofrem inanição, uma vez que o UE 1 não obteve RBs em 97,66% dos *slots* e o UE 2 em 98,30% dos *slots*. O FMR, Figura 4d, sacrifica parcialmente a eficiência total para priorizar uma distribuição mais equitativa de RBs. O FMR garante que nenhum UE fique em estado de inanição. Assim, o UE 3 apresenta uma redução em sua vazão em comparação com MR, enquanto os UEs 1 e 2 obtêm uma maior vazão. Em todos os *slots*, cada UE recebeu uma parcela de recursos suficiente para atender aos critérios mínimos de vazão estabelecidos pela função de recompensa.



(a) Vazão por UE com RR. **(b) Vazão por UE com PF.**

(c) Vazão por UE com MR. **(d) Vazão por UE com FMR.**

Figura 4. Comparação da vazão entre UEs por escalonador.

A Figura 5 compara a vazão agregada de cada escalonador. O MR apresenta a maior vazão agregada, seguido pelo FMR, enquanto o PF e o RR possuem valores inferiores aos dois primeiros. Considerando o equilíbrio entre vazão e equidade na divisão de recursos, é possível afirmar que o FMR apresenta uma divisão de recursos mais justa,

mesmo que para isso não alcance uma vazão agregada igual ou superior à do MR. Além disso, embora os escalonadores RR e PF sejam mais equitativos na divisão, o FMR obteve um desempenho de vazão superior ao RR, com um aumento médio de 24,73%, e ao PF, com um aumento médio de 11,03%, enquanto apresentou uma redução de apenas 10,50% em relação ao MR. Embora os gráficos apresentem um padrão consistente para todos os escalonadores, isso pode ser atribuído ao uso de dados idênticos, que visaram garantir igualdade de condições na avaliação.

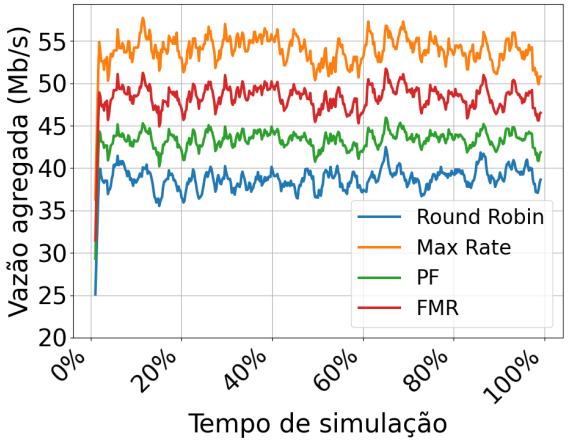


Figura 5. Comparação da vazão agregada por escalonador.

Observa-se que, no cenário analisado, a vazão é limitada pela largura de banda fixa de 10 MHz, mesmo com a utilização da numerologia μ=1 (SCS=30 kHz). Nesse contexto, a largura de banda disponível se configura como o principal fator determinante do desempenho. Embora numerologias mais altas, como μ=2 (SCS=60 kHz), reduzam a duração dos *slots*, esse benefício não se mostra relevante no cenário avaliado, uma vez que a análise não envolve situações de baixa latência, alta mobilidade ou utilização de grandes larguras de banda.

A Tabela 1 apresenta a comparação dos valores do Índice de Jain e a média da vazão agregada da rede para cada escalonador. A principal conclusão é que existe um compromisso entre vazão agregada e justiça. Quanto mais justo é o escalonador, menor é a vazão agregada proporcionada pelo escalonador. No entanto, é preciso levar em consideração que UEs não podem sofrer inanição. Por isso, a proposta do FMR atende seu objetivo: ter alta vazão agregada, mas sem negligenciar os UEs com MCSes mais baixos. Nenhum UE fica sem receber RBs em um *slot* de tempo com o FMR, ao preço de uma redução de cerca de 10% da vazão agregada em relação ao MR. O FMR é 27,2% mais justo do que o MR, segundo o Índice de Jain.

5. Trabalhos Relacionados

O uso de técnicas de aprendizado por reforço tem se destacado como uma solução promissora para a alocação de recursos em redes móveis 5G. Diversas abordagens têm sido exploradas na literatura para otimizar o uso dos recursos disponíveis, considerando diferentes métricas de desempenho, cenários e restrições operacionais.

Al-Tam *et al.* propõem o escalonador LEASCH, baseado em de aprendizado por reforço profundo (DDQN) proposto para otimizar o agendamento de recursos de rádio

Tabela 1. Resumo do Índice de Jain e vazão agregada para os resultados de distribuição de recursos utilizando diferentes escalonadores.

Escalonador	Índice de Jain	Vazão Agregada Média (Mb/s)
Round Robin (RR)	0,9247	38,48
Proportional Fair (PF)	0,7362	43,06
Max Rate (MR)	0,3411	53,62
Fair Max Rate (FMR)	0,4696	48,05

em redes 5G [Al-Tam et al. 2020]. O LEASCH considera diversos fatores, como elegibilidade e taxa de dados, para tomar decisões de divisão. Os resultados indicam melhorias significativas em relação aos escalonadores tradicionais RR e PF, mas o modelo ainda apresenta limitações em cenários com poucos usuários.

Gu *et al.* apresentam o K-DDPG, um algoritmo de aprendizado por reforço assistido por conhecimento, proposto para melhorar o agendamento em redes 5G com foco em tráfego sensível ao tempo [Gu et al. 2021]. Ao combinar o algoritmo DDPG com modelos teóricos, o K-DDPG reduz perdas de pacotes e apresenta convergência mais rápida. No entanto, o desempenho do algoritmo ainda não foi suficientemente validado em situações com rápidas variações do canal ou alta mobilidade dos usuários, indicando uma oportunidade para investigações futuras nesses cenários específicos.

Huang e Kadoch propõem uma estratégia de agendamento baseada em Q-learning para garantir baixa latência em redes 5G com recursos espectrais limitados [Huang and Kadoch 2020]. O modelo seleciona alocações de subportadoras que minimizam a latência, apresentando resultados promissores em termos de latência e eficiência espectral. A principal limitação é o desempenho em cenários com grande variabilidade de usuários ou condições de canal.

Saraiva *et al.* propõem o Deep Q-RA, que utiliza um enfoque multiagente para otimizar a divisão de recursos de rádio em redes multi-serviço [Saraiva et al. 2020]. Os agentes, treinados centralmente, tomam decisões de forma paralela e distribuída. Os resultados indicam um desempenho próximo ao ótimo em termos de taxa de transferência e baixas taxas de interrupção. No entanto, o modelo apresenta dependência de parâmetros iniciais e não considera a correlação do canal ao longo do tempo.

Yang *et al.* propõem o RDRL-RA, que combina DQN com uma função de recompensa específica para garantir a qualidade de serviço e melhorar a confiabilidade em cenários com alta mobilidade [Yang et al. 2022]. Os resultados mostram uma redução significativa na violação de QoS e um aumento na vazão. A principal limitação é o impacto de mudanças rápidas nas condições do canal no desempenho do algoritmo.

Este trabalho propõe o escalonador FMR, que utiliza o algoritmo PPO (*Proximal Policy Optimization*) para resolver o problema de divisão de recursos em redes 5G. O PPO foi escolhido por sua capacidade em lidar com espaços contínuos e pelo mecanismo de *clipping*, que contribui para uma maior estabilidade durante as atualizações da política, características adequadas ao cenário analisado neste estudo [Schulman et al. 2017]. Associado aos MCSes dos dispositivos e a uma função de recompensa projetada para valorizar a vazão e a justiça, o PPO se torna uma solução adequada para ambientes diversos.

6. Conclusões

Este artigo propôs um escalonador de recursos de rádio para redes 5G, denominado *Fair Max Rate* (FMR). O objetivo do FMR é equilibrar a vazão agregada e a justiça na distribuição de RBs entre os UEs. Diferente dos escalonadores tradicionais, que seguem políticas de alocação predefinidas e fixas ao longo do tempo, o FMR aprende uma política de alocação dinâmica a partir da interação com o ambiente, ajustando sua estratégia de divisão de RBs conforme as condições da rede. Para isso, o FMR implementa um agente de aprendizado por reforço, que responde pelas decisões de alocação de RBs.

Os resultados obtidos por simulação demonstram que o FMR se mostra mais justo do que o MR e apresenta maior vazão em comparação com o RR e o PF. E, principalmente, com o FMR, nenhum UE sofre inanição. Por isso, o FMR atende seu objetivo: ter alta vazão agregada, mas sem negligenciar os UEs com MCSes mais baixos. O FMR é 27,2% mais justo do que o MR, segundo o Índice de Jain, e aumenta a vazão agregada em cerca de 25% se comparado ao RR.

Trabalhos futuros incluem realizar simulações que considerem outros fatores da rede, incluindo latência e carga de processamento. Embora existam propostas na literatura com abordagens semelhantes, a comparação direta com esses trabalhos não foi viável, pois os cenários e métricas avaliadas diferem significativamente, sendo incompatíveis com o ambiente Python utilizado neste trabalho. Um próximo passo inclui a implementação do FMR em simuladores amplamente utilizados, como o NS-3, permitindo uma avaliação mais abrangente das métricas e possibilitando a comparação com trabalhos existentes na área.

Agradecimentos

Esta pesquisa é parte do INCT de Redes de Comunicação e Internet das Coisas Inteligentes (ICoNIoT), financiado por CNPq (proc. 405940/2022-0), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 88887.954253/2024-00. Também há financiamento parcial por parte da FAPERJ, RNP e PGC/UFF.

Referências

3GPP (2020a). Physical channels and modulation (3GPP TS 38.211 Version 16.2. 0 Release 16) ETSI. *Sophia Antipolis, France*.

3GPP (2020b). Physical layer procedures for data (3GPP TS 38.214 Version 16.2. 0 Release 16) ETSI. *Sophia Antipolis, France*.

Agiwal, M., Roy, A., and Saxena, N. (2016). Next Generation 5G Wireless Networks: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials*, 18(3):1617–1655.

Al-Tam, F., Correia, N., and Rodriguez, J. (2020). Learn to Schedule: LEASCH: A Deep Reinforcement Learning Approach for Radio Resource Scheduling in the 5G MAC Layer. *IEEE Access*, 8:143063–143076.

Andrews, J. G., Buzzi, S., Choi, W., Hanly, S. V., Lozano, A., Soong, A. C., and Zhang, J. C. (2014). What will 5G be? *IEEE Journal on selected areas in communications*, 32(6):1065–1082.

Bonati, L., D'Oro, S., Polese, M., Basagni, S., and Melodia, T. (2021). Intelligence and Learning in O-RAN for Data-driven NextG Cellular Networks. *IEEE Communications Magazine*, 59(10):21–27.

Dahlman, E., Parkvall, S., and Skold, J. (2020). *5G NR: The next generation wireless access technology*. Academic Press.

Gu, Z., She, C., Hardjawana, W., Lumb, S., McKechnie, D., Essery, T., and Vucetic, B. (2021). Knowledge-assisted deep reinforcement learning in 5G scheduler design: From theoretical framework to implementation. *IEEE Journal on Selected Areas in Communications*, 39(7):2014–2028.

Holma, H., Toskala, A., and Reunanen, J. (2015). *LTE Small Cell Optimization: 3GPP Evolution to Release 13*. Wiley.

Huang, Q. and Kadoch, M. (2020). 5G resource scheduling for low-latency communication: A reinforcement learning approach. In *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, pages 1–5. IEEE.

Jain, R. K., Chiu, D.-M. W., and Hawe, W. R. (1984). A quantitative measure of fairness and discrimination. *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA*, 21:1.

Patriciello, N., Lagen, S., Bojovic, B., and Giupponi, L. (2019). An E2E simulator for 5G NR networks. *Simulation Modelling Practice and Theory*, 96:101933.

Rodrigues, C. F. F., Lovisolo, L., and da Silva Mello, L. (2022). Alocação de Recursos da Interface Aérea 5G a partir de um Critério de Utilidade. In *Anais do XL Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT 2022)*.

Saraiva, J. V., Jr., I. M. B., Monteiro, V. F., Lima, F. R. M., Maciel, T. F., Jr., W. C. F., and Cavalcanti, F. R. P. (2020). Deep Reinforcement Learning for QoS-Constrained Resource Allocation in Multiservice Networks.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Vihriälä, J., Zaidi, A. A., Venkatasubramanian, V., He, N., Tiirola, E., Medbo, J., Lähetkangas, E., Werner, K., Pajukoski, K., and Cedergren, A. (2016). Numerology and frame structure for 5G radio access. In *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–5.

Yang, L., Jia, J., Lin, H., and Cao, J. (2022). Reliable dynamic service chain scheduling in 5G networks. *IEEE Transactions on Mobile Computing*, 22(8):4898–4911.

You, X., Zhang, C., Tan, X., Jin, S., and Wu, H. (2019). AI for 5G: research directions and paradigms. *Science China Information Sciences*, 62:1–13.

Zhang, C., Ueng, Y.-L., Studer, C., and Burg, A. (2020). Artificial Intelligence for 5G and Beyond 5G: Implementations, Algorithms, and Optimizations. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 10(2):149–163.