

# **Alocação de Recursos no Uplink Baseada em Aprendizagem por Reforço considerando Otimização de Potência e Retardo para Redes 5G com Comunicações D2D**

**Marcus V. G. Ferreira<sup>1</sup>, Leonardo A. Melo<sup>1</sup>, Flávio H. T. Vieira<sup>2</sup>**

<sup>1</sup>Escola de Engenharia Elétrica, Mecânica e de Computação (EMC),  
Universidade Federal de Goiás (UFG)

Av. Universitária, 1488, Quadra 86, Setor Leste Universitário, Goiânia, Goiás, Brasil

<sup>2</sup>Centro de Excelência em Redes Inteligentes Sem Fio e Serviços Avançados  
(CERISE), EMC, UFG

Av. Universitária, 1488, Quadra 86, Setor Leste Universitário, Goiânia, Goiás, Brasil

marcusviniciusbr@gmail.com, leoalvesdemelo@gmail.com, flavio\_vieira@ufg.br

**Abstract.** In this paper, the problem of resource allocation for devices is investigated in a multi-sharing uplink scenario within cyclic prefix orthogonal frequency-division multiplexing (CP-OFDM) and millimeter-waves (mmWaves)-based wireless networks, incorporating device-to-device (D2D) communications. Specifically, a power and delay optimization uplink resource allocation algorithm with reinforcement learning (PDO-URA-RL) is proposed, divided into two stages. First, network resources are allocated to cellular user equipments (CUEs) in terms of power and transmission rate through an approach aimed at maximizing throughput. Subsequently, idle resources are allocated with a focus on minimizing delay. Computational simulations are conducted in a 5G-oriented communication scenario, leveraging mmWaves frequencies above 6 GHz, while comparing the performance with other literature algorithms in terms of quality of service (QoS) parameters, such as throughput and delay.

**Resumo.** Neste artigo, o problema de alocação de recursos para dispositivos é investigado em um cenário de compartilhamento múltiplo no uplink de redes sem fio baseadas em multiplexação por divisão de frequência ortogonal de prefixo cílico (CP-OFDM) e ondas milimétricas (mmWaves), com comunicações dispositivo-a-dispositivo (D2D). Especificamente, é proposto um algoritmo de alocação de recursos no uplink baseado na otimização de potência e retardo e na aprendizagem por reforço (PDO-URA-RL), dividido em duas etapas. Primeiramente, os recursos da rede são alocados para os equipamentos celulares de usuário (CUEs) em termos de potência e taxa de transmissão por meio de uma abordagem que visa maximizar a vazão. Em seguida, os recursos ociosos são alocados considerando a minimização do retardo. Simulações computacionais são realizadas considerando um cenário de comunicação com características do 5G, como ondas milimétricas em frequências acima de 6 GHz, comparando também o desempenho com outros algoritmos da literatura em termos de parâmetros de qualidade de serviço (QoS), como vazão e retardo.

## 1. Introdução

O aumento exponencial da demanda por tráfego sem fio levou os operadores de redes de comunicação a buscar formas de aumentar a vazão do sistema e reduzir a latência, a fim de suportar aplicações relacionadas a conceitos e tecnologias como internet das coisas, comunicação máquina a máquina e realidade aumentada. À medida que cresce o número de dispositivos conectados às redes sem fio, torna-se necessário desenvolver sistemas de comunicação celular que empreguem técnicas avançadas para melhorar a eficiência e oferecer melhor cobertura e vazão.

Uma das técnicas estudadas na literatura para melhorar a eficiência espectral é a comunicação dispositivo-a-dispositivo (D2D). A comunicação D2D foi introduzida no LTE-A (*Long Term Evolution - Advanced*) e, atualmente, é ainda mais direcionada para as redes de comunicação de quinta geração (5G). O principal conceito das comunicações D2D envolve dispositivos que se comunicam entre si com baixa interferência do nó B evoluído (eNodeB), aliviando a sobrecarga de tráfego no eNodeB. A comunicação D2D toma proveito da característica assimétrica do tráfego de dados em redes de Internet, ou seja, o tráfego é muito maior no *downlink* do que no *uplink*. Assim, há uma maior disponibilidade de recursos ociosos no *uplink* em comparação com o *downlink* e, por esse motivo, a maioria das pesquisas sobre comunicação D2D se concentra no compartilhamento de recursos no *uplink*.

Diversos trabalhos na literatura abordam a comunicação D2D de compartilhamento múltiplo no *uplink*. Em [Sun et al. 2013], os autores propuseram uma heurística com o objetivo de garantir o número máximo de pares de dispositivos se comunicando entre si com a mínima interferência, denominado algoritmo de alocação de recursos ganancioso (GRA). Mais especificamente, eles apresentaram um algoritmo ótimo baseado no método húngaro para resolver o problema de alocação de recursos. O algoritmo proposto em [Sun et al. 2013] não foi desenvolvido para um grande número de dispositivos na rede, e os autores consideraram um cenário de compartilhamento único, onde cada bloco de recurso (RB) alocado pode ser reutilizado por apenas um par de dispositivos.

Em [Ciou et al. 2015], os autores propõem um algoritmo baseado em alguns resultados descritos em [Sun et al. 2013] com o objetivo de resolver o problema de alocação de recursos com compartilhamento múltiplo, aumentando a vazão do sistema e garantindo os requisitos de relação sinal-interferência-mais-ruído (SINR), denominado como algoritmo de maximização gananciosa de vazão mais (GTM+). Em [Ciou et al. 2015] os autores formularam o problema de alocação com múltiplo compartilhamento, provaram que se trata de um problema NP-difícil e propuseram um algoritmo que utiliza conceitos como grafos de conflito e peso máximo independente para encontrar uma solução rápida e eficiente.

Neste contexto de alocação de recursos, o parâmetro de retardo é considerado essencial, principalmente para aplicações em tempo real com taxa de transmissão variável e requisitos específicos de largura de banda, como serviços de VoIP (voz sobre protocolo de Internet) e videoconferência. Esses serviços estão cada vez mais em evidência à medida que o número de dispositivos móveis cresce exponencialmente. Também destacamos a necessidade de desenvolver novas técnicas para viabilizar o cumprimento de requisitos rigorosos de QoS para redes 5G, como taxa de dados de até 10 Gbps e latência de 1 milissegundo. As redes sem fio 5G foram propostas com o objetivo de atender às

atuais necessidades de intercomunicação entre bilhões de dispositivos. Para que isso seja possível, é essencial que os requisitos estabelecidos para as redes 5G sejam atendidos, permitindo a comunicação simultânea entre os diversos dispositivos com maior taxa de transmissão e menor latência.

Diferentemente dos trabalhos mencionados anteriormente, propomos neste artigo um algoritmo mais amplo que lida, primeiramente, com a otimização da comunicação entre os dispositivos com o eNodeB e, em seguida, com a comunicação de compartilhamento múltiplo entre dispositivos D2D, visando minimizar o parâmetro de retardo do sistema. Acrescentamos à proposta apresentada em [Ferreira et al. 2022], denominado algoritmo de alocação de recursos no *uplink* baseado na otimização de potência e retardo (*power and delay optimization based uplink resource allocation - PDO-URA*) o método de criação de links de comunicação D2D por meio de aprendizagem por reforço [Sreedevi and Rama Rao 2019].

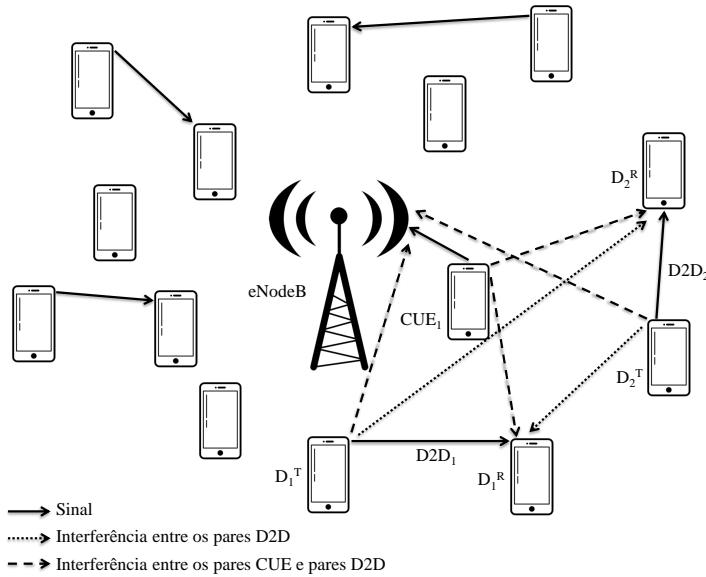
O algoritmo é dividido em duas etapas. Primeiramente, a alocação de recursos para os equipamentos celulares de usuário (CUEs) é realizada pelo eNodeB, levando em consideração a otimização do retardo do tráfego de dados e a potência alocada. Para isso, propomos formular o problema de otimização do retardo como um problema de maximização de vazão e introduzimos uma abordagem de alocação de potência e taxa de transmissão que envolve maximizar a métrica de *goodput*, expressa em termos da SINR. Em segundo lugar, os recursos ociosos são alocados para os dispositivos D2D. Nesse sentido, propomos utilizar uma abordagem baseada em aprendizagem por reforço para formação dos links D2D em conjunto com o algoritmo de alocação de recursos para comunicação de compartilhamento múltiplo, chamado algoritmo de grafos de conflito e minimização de retardo (*delay minimization conflict graph - DMCG*), introduzido em [Ferreira and Vieira 2021], para reduzir o retardo dos equipamentos de usuário D2D (DUEs) e aumentar o desempenho do sistema. Na abordagem DMCG, o valor do limite de retardo é estimado utilizando o processo envelope multifractal e a curva de serviço para os fluxos de tráfego no *uplink* por meio de conceitos como grafos de conflito e conjuntos de peso máximo independente. No entanto, o algoritmo DMCG difere dos algoritmos mencionados, pois tem como objetivo resolver o problema de alocação de recursos de compartilhamento múltiplo considerando a minimização do retardo. O algoritmo proposto neste artigo, além de considerar a alocação de potência para os usuários, aborda a alocação de recursos entre os CUEs e o eNodeB.

O cenário simulado neste artigo considera comunicações que ocorrem por meio da propagação por ondas milimétricas de 26 GHz, uma característica importante das redes 5G com o objetivo de melhorar a vazão e a latência para os usuários da rede. O algoritmo de alocação de recursos no *uplink* baseado na otimização de potência e retardo e na aprendizagem por reforço (*power and delay optimization based uplink resource allocation with reinforcement learning - PDO-URA-RL*) também é aplicável a outras faixas de frequência. No entanto, optamos por simular cenários com propagação por ondas milimétricas para aproveitar a maior disponibilidade de largura de banda para os usuários nessa faixa de frequência, possibilitando análises em taxas mais altas, específicas das redes 5G. Os resultados da simulação mostram que o algoritmo PDO-URA-RL proposto proporciona ganhos consideráveis em termos de vazão e retardo no cenário simulado de propagação por ondas milimétricas.

O restante do artigo está organizado da seguinte forma: A Seção 2 descreve o problema de alocação de recursos com compartilhamento múltiplo e o modelo do sistema. A Seção 3 apresenta o algoritmo de alocação proposto. A Seção 4 aborda a avaliação de desempenho dos algoritmos de alocação de recursos. A Seção 5 apresenta nossas considerações finais.

## 2. Modelo do Sistema

O modelo de sistema para comunicações D2D considerado neste trabalho é representado pela Figura 1. Nesse sistema, existem pares de DUEs que reutilizam os recursos de rádio alocados aos CUEs, comunicando-se diretamente entre si.



**Figura 1. Modelo de sistema para comunicações D2D com DUEs reutilizando recursos *uplink* de CUEs.**

Assumimos neste trabalho que os CUEs reais não utilizam RBs ociosos. Mais especificamente, consideramos que cada RB ocioso, que não está sendo usado por nenhum CUE real, está relacionado a um CUE virtual com potência de transmissão zero e sem a exigência de um valor mínimo para a SINR. O sistema D2D é composto por  $N$  DUEs,  $M$  CUEs, que inclui CUEs reais e virtuais,  $K$  RBs ociosos agendados para transmissão durante um intervalo de tempo de transmissão (TTI) e  $M - K$  CUEs reais. Um RB representa a unidade mínima de alocação no sistema de comunicação sem fio. Os CUEs são denotados por  $C_1, C_2, \dots, C_M$  e os DUEs são denotados por  $D_1, D_2, \dots, D_N$ . O receptor do par de DUE é denotado por  $D_{n,Rx}$  e o transmissor por  $D_{n,Tx}$ .

Um conjunto de RBs para cada CUE é pré-alocado antes de compartilhar os RBs com os DUEs durante a transmissão. A largura de banda alocada  $W_m$  para cada CUE  $m$  é proporcional ao número de RBs alocados. No modelo de sistema considerado neste trabalho, assume-se que um grande número de conexões simultâneas pode ocorrer entre DUEs devido à característica de cada CUE de poder compartilhar seus RBs com vários pares de DUEs, conforme o conceito de alocação de recursos de compartilhamento múltiplo.

O conjunto denotado como  $\Theta_n$  especifica a relação CUE e DUE e representa o conjunto de CUEs que compartilham seus RBs com o par de DUE  $n$ . Outro conjunto, denotado como  $\Delta_m$ , representa o conjunto dos pares de DUEs que reutilizam os RBs alocados para o CUE  $m$ . Os conjuntos  $\Delta_1, \Delta_2, \dots, \Delta_M$  são subconjuntos de  $\{1, 2, \dots, N\}$ .

Consideramos neste trabalho que a interferência restringe a alocação de RBs para os DUEs no *uplink*, ou seja, o conjunto de DUEs só pode reutilizar RBs do CUE  $m$  se a interferência na transmissão de  $m$  obedecer à exigência de SINR. O SINR recebido para qualquer CUE  $m \in \{1, 2, \dots, M\}$  deve ser superior ao seguinte limite:

$$\frac{P_m G_{mB}}{\sigma_m^2 + \sum_{n \in \Delta_m} P_n G_{nB}} \geq \gamma_m, \quad (1)$$

onde  $P_m$  é a potência de transmissão para o CUE  $m$ ,  $G_{mB}$  é o ganho de canal entre o CUE  $m$  e o eNodeB,  $\sigma_m^2$  é a potência de ruído para o CUE  $m$ ,  $P_n$  é a potência de transmissão para o par de DUE  $n$ ,  $G_{nB}$  é o ganho de canal entre o transmissor do par de DUE  $n$  e o eNodeB para SINR maior que zero e  $\gamma_m$  é o limiar de SINR exigido pelo CUE  $m$ . Os valores de ganho de canal  $G_{mB}$  e  $G_{nB}$  variam dependendo do modelo de canal utilizado e das condições de propagação do sinal. Neste trabalho, o modelo de canal *clustered delay line* D (CDL-D) é utilizado para gerar os valores desses ganhos [3gp 2017].

Analogamente, existem exigências de SINR para os pares de DUEs. Os RBs podem ser reutilizados por um par de DUE  $n$  somente se o SINR recebido for superior ao seguinte limite:

$$\frac{P_n G_{nn}}{\sigma_n^2 + P_m G_{mn} + \sum_{n' \in \Delta_m - \{n\}} P_{n'} G_{n'n}} \geq \gamma_n, \forall m \in \Theta_n, \quad (2)$$

onde  $G_{nn}$  é o ganho de canal entre os dois terminais do par DUE  $n$ ,  $\sigma_n^2$  é a potência de ruído para o par DUE  $n$ ,  $P_m$  é a potência de transmissão para o CUE  $m$ ,  $G_{mn}$  é o ganho de canal de CUE  $m$  para o par DUE  $n$ ,  $G_{n'n}$  é o ganho de canal de  $D_{n',Tx}$  para  $D_{n,Rx}$  para SINR maior que zero e  $\gamma_n$  é o limiar de SINR para o par DUE  $n$ . Os valores de ganho de canal  $G_{nn}$ ,  $G_{mn}$  e  $G_{n'n}$  variam dependendo da modelagem de canal utilizada e das condições de propagação do sinal. Os valores para esses ganhos de canal também são dados pela modelagem de canal CDL-D neste artigo [3gp 2017].

Neste trabalho, assume-se que o eNodeB conhece algumas propriedades do canal, como potência de transmissão, ganho de canal e potência de ruído do link de comunicação para os CUEs  $m$  e os pares DUE  $n$ . Ou seja, o eNodeB conhece a informação do estado do canal e o limiar de SINR para cada usuário. O conhecimento desses parâmetros pelo eNodeB é assumido em vários trabalhos relacionados à alocação de recursos em redes sem fio, como [Sun et al. 2013], [Ciou et al. 2015] e [Ferreira and Vieira 2021], com o propósito de simplificar cenários de simulação. Neste artigo, focamos na parte de alocação de recursos, sem perda de generalidade, e considerando o conhecimento das informações de ganho de canal para os usuários da rede. Em outras palavras, assumimos que os parâmetros de ganho de canal usados para simular o algoritmo proposto poderiam ser estimados pelo eNodeB utilizando técnicas de estimativa de canal conhecidas na literatura [Belgiovine et al. 2021].

Em um sistema de comunicação D2D, o objetivo do problema de alocação de recursos de compartilhamento múltiplo consiste em determinar quais pares DUE reutilizam RBs dos CUEs para maximizar o rendimento total do sistema. O problema de alocação de recursos de compartilhamento múltiplo para um sistema de comunicação D2D é NP-difícil (não-determinístico polinomial de tempo difícil), como provado em [Sun et al. 2013] e [Ciou et al. 2015]. A dificuldade NP do problema reforça a necessidade de trabalhos como o apresentado neste artigo com o objetivo de encontrar um algoritmo rápido e eficiente para resolver o problema de alocação.

### **3. Algoritmo de Alocação de Recursos no Uplink Baseado na Otimização de Potência e Retardo e na Aprendizagem por Reforço (PDO-URA-RL)**

Nesta seção, propomos um algoritmo de alocação de recursos para os dispositivos em redes sem fio de compartilhamento múltiplo no *uplink* utilizando multiplexação por divisão de frequência ortogonal de prefixo cílico (CP-OFDM) e comunicação D2D. O algoritmo proposto é dividido em duas etapas. Primeiramente, os recursos da rede são alocados para os CUEs em termos de potência e taxa de transmissão por meio de uma abordagem proposta que visa maximizar a vazão. Em seguida, os recursos ociosos são alocados considerando a minimização do retardo e um método de aprendizagem por reforço com abordagem Q-learning para formação dos links D2D. Para isso, propomos o uso de outro algoritmo, denominado grafos de conflitos e minimização de retardo (DMCG), onde os recursos ociosos são compartilhados com os DUEs na rede, considerando os grafos de conflitos formados e as informações de retardo estimado por meio de conceitos de Cálculo de Rede, como o processo de envelope e a curva de serviço.

#### **3.1. Aprendizagem por Reforço para Redes de Comunicação D2D**

O método de aprendizagem por reforço tem como objetivo construir uma rede de comunicação D2D com melhor conectividade e latência mínima. Esse algoritmo se baseia na abordagem de aprendizado por reforço Q-learning, uma ferramenta eficiente de aprendizagem por máquina para tomada de decisões em condições desconhecidas da rede. A principal vantagem do algoritmo é que ele simplifica o projeto, pois abrange a maioria dos fatores que afetam o desempenho, e não requer conhecimento prévio da rede de comunicação D2D.

Os principais componentes desse modelo são: agente, estado, ação, recompensa e ambiente. Na rede de comunicação D2D, os dispositivos que realizam a tarefa de criar um link de comunicação D2D atuam como agentes. O estado do ambiente  $S$  de cada dispositivo  $d_i$  é definido pelo conjunto de vizinhos de  $d_i$  e pelo número de pacotes em sua fila de *buffer*. Esse estado serve como fator de decisão para o algoritmo. Assim,  $S = \{(n_1, p_1), (n_2, p_2), \dots (n_M, p_M)\}$ , onde  $n_1, n_2, \dots n_M$  são os números de dispositivos vizinhos,  $p_1, p_2, \dots p_M$  são os números de pacotes na fila de *buffer* dos dispositivos  $d_1, d_2, \dots d_M$ , respectivamente. Em cada estado  $s \in S$ , o dispositivo executa uma ação  $a$ , que determina o link de comunicação D2D para o dispositivo correspondente. O estado e a recompensa podem ser alterados ou afetados pela ação escolhida pelo agente. O agente realiza uma ação, que interage com o ambiente, modificando a recompensa  $RW$  [Sreedevi and Rama Rao 2019].

O desempenho da rede de comunicação ao executar uma ação em um estado no instante anterior é calculado pela função de recompensa. O conjunto de vizinhos de um

dispositivo na rede depende muito da cobertura desse dispositivo, e maximizar a recompensa melhora o desempenho da rede. Nesta abordagem, para calcular a função de recompensa em cada dispositivo, o algoritmo utiliza o número de dispositivos vizinhos  $n_i$  do dispositivo  $d_i$ , denominado grau de vizinhança. O grau de vizinhança e o número de pacotes no *buffer* juntos formam o estado  $s$ . O alcance de comunicação  $r$  define a ação  $a$ . A função de recompensa do dispositivo  $d_i$ ,  $RW_{t,i}$  no tempo  $t$ , possui três componentes independentes [Sreedevi and Rama Rao 2019]:

$$RW_{t,i} = 1 - \omega_n \frac{n_{t,i}}{N_{max}} - \omega_r \frac{r_{t,i}}{R_{max}} - \omega_p \frac{p_{t,i}}{P_{max}}. \quad (3)$$

O algoritmo proposto utiliza a abordagem Q-learning, uma técnica popular de aprendizagem por reforço. Nessa abordagem, cada dispositivo emprega o método  $\epsilon$ -greedy para selecionar seu alcance de comunicação, levando em conta o alcance de comunicação e o tamanho do *buffer*. O número de pacotes no *buffer* é considerado uma medida de latência. Quanto menos pacotes houver no *buffer*, menor será a latência ao longo do link. O dispositivo escolhe o alcance de comunicação e o tamanho do *buffer* com o maior valor  $Q$  com alta probabilidade  $(1 - \epsilon_t)$ . Da mesma forma, os demais alcances de comunicação são considerados com uma probabilidade igualmente baixa. A abordagem Q-learning adota um mecanismo de seleção  $\epsilon$ -greedy, onde  $\epsilon = 0, 1$ . Ao analisar o maior valor  $Q$  para todas as ações possíveis  $A(s)$ , a probabilidade  $\pi(s, a)$  de selecionar uma ação  $a$  é definida como [Sreedevi and Rama Rao 2019]:

$$\pi(s, a) = \begin{cases} 1 - \frac{\epsilon}{|A(s)|}, & \text{se } a = a*, \text{ onde } a* = \max_{A(s)} Q(s, a), \\ \frac{\epsilon}{|A(s)|}, & \text{em outro caso,} \end{cases} \quad (4)$$

onde o valor  $Q$  é demarcado como  $Q(s, a) = R(s, a) + \gamma \max_{A(s)} Q(s', a)$ , sendo  $\gamma = 0, 5$  considerada a constante de aprendizagem para melhor adaptabilidade a ambientes não estacionários. O Q-learning é calculado pela soma da recompensa esperada nos estados com a ação  $a$  e o maior valor de  $Q$  para o próximo estado com todas as ações possíveis no subconjunto  $A(s)$  [Sreedevi and Rama Rao 2019].

A rede de comunicação D2D considerada é uma rede com  $M$  dispositivos comunicantes. Cada dispositivo na rede transmite uma mensagem com a potência de transmissão existente, o alcance máximo de comunicação e o número de pacotes na fila do *buffer* e, em seguida, coleta as informações de seus vizinhos. Com base nas informações recebidas, cada dispositivo executa o algoritmo de reforço e decide o link de comunicação D2D com máxima cobertura, utilizando seu alcance e o número de pacotes na fila do *buffer*.

### 3.2. Algoritmo Geral para Alocar Recursos para CUEs e DUEs

Para resumir o algoritmo proposto PDO-URA-RL, que considera as comunicações de CUEs e DUEs em uma rede CP-OFDM, um pseudo-código é apresentado no **Algoritmo 1**. O **Algoritmo 1** fornece soluções que visam maximizar a taxa de transferência e minimizar o retardo para os usuários, dividindo em problemas de otimização separados. Primeiramente, a potência é alocada para os usuários a fim de minimizar a taxa de erro de bits (*bits error rate* - BER) do link de transmissão. Em seguida, o número de blocos de

recursos (taxa de transmissão) para os CUEs é determinado considerando a minimização do retardo. Finalmente, os blocos de recursos que não foram alocados para os CUEs são distribuídos de forma otimizada para os DUEs, considerando a abordagem por aprendizagem por reforço descrita anteriormente e a minimização do retardo por meio do algoritmo de grafos de conflito (**Algoritmo DMCG** [Ferreira et al. 2022]).

---

**Algoritmo 1:** Algoritmo de Alocação de Recursos no Uplink Baseado na Otimização de Potência e Retardo e na Aprendizagem por Reforço (PDO-URA-RL)

---

- 1 Receba os parâmetros iniciais:  $M$  CUEs,  $N$  pares de DUEs,  $K$  RBs ociosos e  $M - K$  CUEs reais;
  - 2 Para calcular a taxa para os CUEs a fim de minimizar o retardo e otimizar a alocação de potência:  
/\* Executa **Algoritmo de Otimização de Potência e Retardo** [Ferreira et al. 2022] \*/
  - 3 Para alocar recursos ociosos para os DUE considerando a minimização de retardo e a aprendizagem por reforço:  
/\* Executa **Algoritmo DMCG** [Ferreira et al. 2022]  
para formar os grafos de conflito \*/  
/\* Forma os links D2D tendo por base a função de recompensa (3) através do método descrito na Seção 3.1 \*/
- 

#### 4. Simulações e Resultados

Esta seção apresenta o modelo de canal e os valores utilizados para os parâmetros do sistema sem fio, além dos resultados numéricos obtidos por meio das simulações computacionais. Os resultados de simulação do algoritmo de alocação proposto são comparados com os resultados do algoritmo DMCG [Ferreira and Vieira 2021] e GTM+ [Ciou et al. 2015].

Os cenários de simulação apresentados neste artigo consideram a propagação por ondas milimétricas, uma das características das redes 5G. Os níveis de potência para os sinais de radiofrequência utilizados nas simulações deste artigo estão de acordo com os limites de exposição humana especificados pela Comissão Internacional para a Proteção contra Radiações Não Ionizantes (ICNIRP) [Karipidis et al. 2021]. O modelo de múltiplos percursos CDL-D é considerado neste trabalho para verificar os impactos da modelagem do canal por meio de simulações. A frequência portadora de 26 GHz foi escolhida com base em estudos recentes da Agência Nacional de Telecomunicações (ANATEL) do Brasil, que também deve ser adotada em outros países da Europa [tec ].

As funções e rotinas de simulação foram implementadas manualmente utilizando o software MATLAB versão R2018a, a fim de obter maior controle sobre a configuração do cenário de simulação e os parâmetros de modelagem do canal em comparação com as ferramentas de simulação de rede disponíveis.

Cinco séries de tráfego real do protocolo de controle de transmissão/protocolo da Internet (TCP/IP) foram utilizados para representar o tráfego de dados dos usuários

durante a simulação dos algoritmos. Essas séries foram agregadas em uma escala de tempo de 1 ms atribuídas aleatoriamente aos usuários e representam o tráfego TCP/IP real da Universidade de Waikato [wai ].

As simulações foram baseadas em um sistema *uplink* CP-OFDM cuja configuração consiste na numerologia  $\nu = 3$ , prefixo cílico (CP) normal, duração de *slot* de  $125 \mu\text{s}$  e espaçamento entre subportadoras de 120 kHz, conforme descrito em [3gp 2018b] [3gp 2018a].

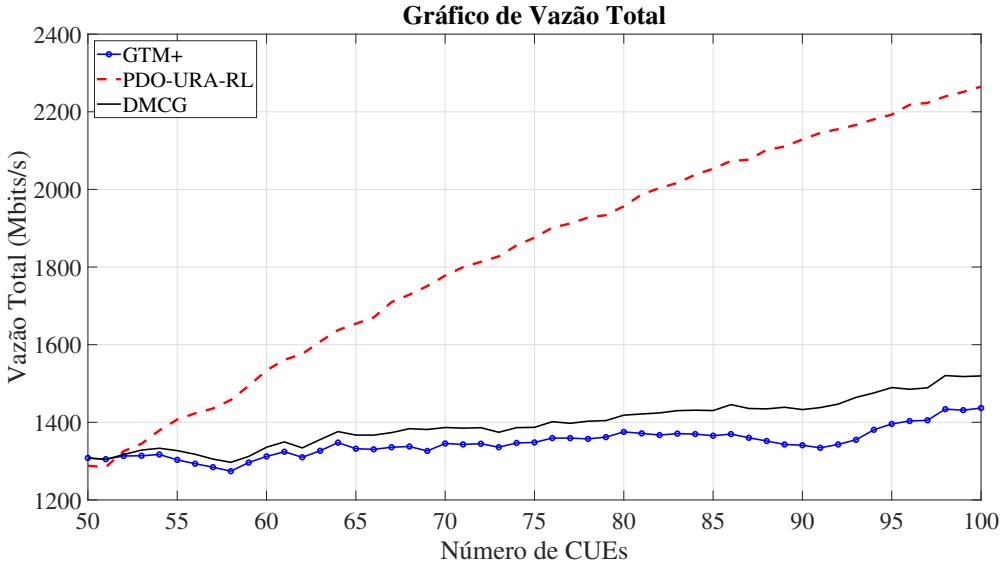
Os parâmetros de simulação do sistema sem fio e as condições do canal são apresentados na Tabela 1. A escolha dos parâmetros de modelagem de canal utilizados nas simulações teve como objetivo representar cenários de propagação por ondas milimétricas mais próximos de situações realistas, onde as condições do canal tendem a apresentar maior degradação. Os valores dos parâmetros foram definidos de acordo com as referências [3gp 2017], [Sun et al. 2013] e [Ciou et al. 2015].

**Tabela 1. Parâmetros de Simulação**

Modelo multipercorso	CDL-D
Cenário	UMa
Potência de transmissão do CUE	23 dBm
Potência de transmissão do DUE	10 dBm
Requisito de SINR para cada CUE	7 dB
Requisito de SINR para cada par DUE	4.7 dB
Raio de cobertura do eNodeB	500 m
Distância entre cada par DUE	15 m
Densidade de potência do ruído branco	-174 dBm/Hz
Modelo de perda de percurso para CUE e DUE	$128.1 + 37.6 \log_{10}(d_{dis}[\text{km}])$
Modelo de perda de percurso para pares DUE	$148 + 40 \log_{10}(d_{dis}[\text{km}])$
Largura de banda do sistema	396 MHz
Número de RBs	275
Frequência da portadora	26 GHz
Número de CUEs reais	40
Número de CUEs	No intervalo entre 50 e 100
Número de pares DUE	160
Número de TTIs	1000
Duração do <i>slot</i>	$125 \mu\text{s}$

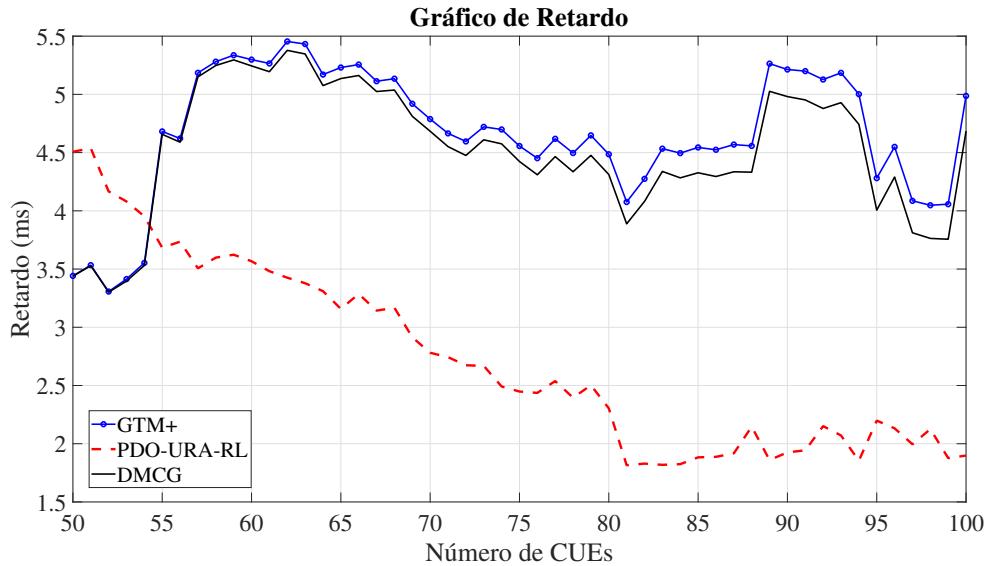
Os valores de taxa de vazão total do sistema são apresentados na Figura 2. O algoritmo de alocação proposto apresenta os maiores valores considerando o cenário simulado. Tanto o algoritmo proposto quanto o DMCG superam o GTM+ porque incorporam a estratégia de grafos de conflitos em conjunto com a minimização do retardo, embora todos os algoritmos simulados apresentem a mesma ordem de complexidade computacional,  $O(n^4)$ . Para reduzir o retardo do sistema, ambos os algoritmos (proposto e DMCG) precisam aumentar a taxa de vazão do sistema ou encontrar uma solução que otimize o processo de alocação. A taxa de vazão total proporcionada pelo algoritmo de alocação proposto melhora à medida que o número de CUEs na rede aumenta. Este fato demonstra que a estratégia de alocação otimizada de recursos para os dispositivos D2D tem grande

impacto no aumento da vazão do algoritmo proposto.



**Figura 2. Taxa de vazão total para diferentes números de CUEs.**

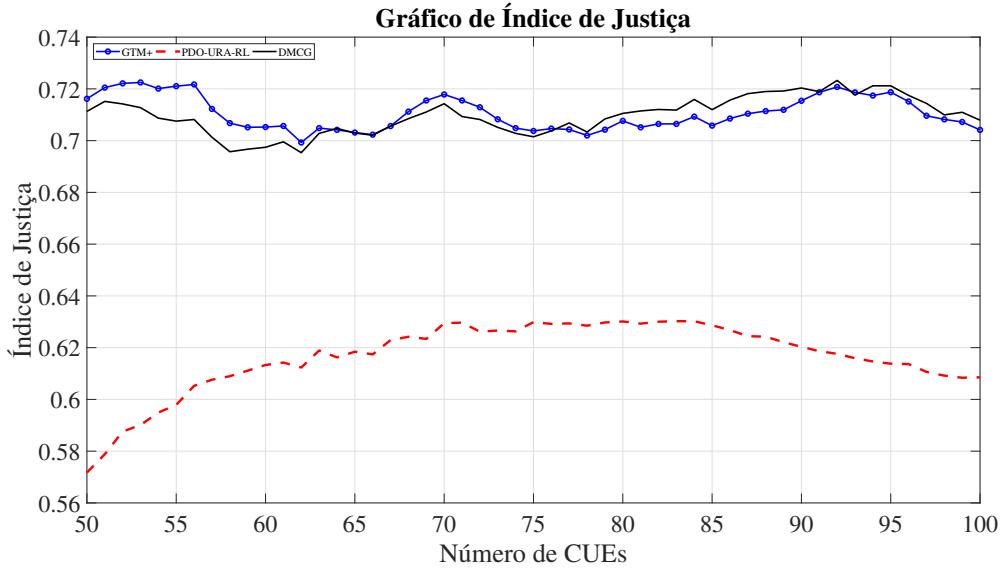
Os valores de retardo médio apresentados pelo algoritmo de alocação proposto são os menores valores, conforme pode ser verificado na Figura 3. Este resultado comprova que o algoritmo proposto é eficiente em sua estratégia de utilizar aprendizagem por reforço para formação dos links e o retardo estimado como função de utilidade no sistema de alocação. O desempenho superior do algoritmo de alocação proposto em comparação com os outros algoritmos considerados deve-se principalmente à sua característica de alocar potência de forma a melhorar o *goodput*. A redução do retardo médio do sistema tem como consequência direta a melhoria da taxa de vazão total do sistema.



**Figura 3. Retardo médio para diferentes números de CUEs.**

Em relação ao índice de justiça (*fairness*), mostrado na Figura 4, verifica-se que o algoritmo proposto apresenta os menores valores dentre os algoritmos considerados,

comprovando que as melhorias significativas em termos de vazão e retardo são efetivadas às custas de uma distribuição menos igualitária dos recursos entre os usuários. Em outras palavras, os usuários com condições de canais mais favoráveis são contemplados com mais recursos em rede.



**Figura 4.** Índice de justiça dos dispositivos D2D para diferentes números de CUEs.

## 5. Conclusões

Este artigo apresentou um algoritmo de alocação de recursos no *uplink* baseado na otimização de potência e retardo e na aprendizagem por reforço (PDO-URA-RL) para redes de comunicação D2D. O algoritmo proposto compreende duas etapas. Primeiramente, a alocação de recursos para CUEs é realizada utilizando um algoritmo de otimização de alocação de potência e retardo que considera a seleção de pesos de atributos para a taxa de vazão calculada para os CUEs. Em segundo lugar, aplica-se o algoritmo DMCG em conjunto com uma técnica de aprendizagem por reforço, para decidir sobre a alocação de recursos ociosos na rede para dispositivos D2D. O algoritmo proposto é avaliado em um cenário de rede CP-OFDM com propagação por ondas milimétricas para aproveitar a maior disponibilidade de largura de banda para os usuários.

Os resultados apresentados demonstram que o algoritmo de alocação proposto oferece os maiores valores de taxa de vazão e os menores valores de retardo quando comparado aos algoritmos GTM+ e DMCG, especialmente quando considerado um número maior de CUEs na simulação. Simulações computacionais confirmaram que a alocação de potência e taxa para usuários CUE, em conjunto com o uso de aprendizagem por reforço e das informações de retardo estimado para identificar grafos de conflito e conjuntos de peso máximo independente melhora o desempenho da alocação de recursos. Isso resulta em melhorias tanto na taxa de vazão quanto no retardo, quando comparado aos algoritmos DMCG e GTM+, em um cenário de comunicação D2D com ondas milimétricas.

Como trabalho futuro, pretendemos aplicar uma variação da abordagem proposta envolvendo transmissão de dados em *downlink*, comparando seu desempenho com outras

estratégias de alocação de recursos.

## 6. Agradecimentos

Os autores agradecem ao Centro de Excelência em Redes Inteligentes Sem Fio e Serviços Avançados (CERISE), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG) pelo apoio e financiamento à pesquisa.

## Referências

- mmwave: o que são as ondas milimétricas que farão o 5g funcionar em altíssimas frequências. <https://tecnoblog.net/270324/5g-nr-mmwave-altas-frequencias-ondas-milimetricas/>. Jun 30, 2021.
- University of waikato - wand network research group: Series of real tcp/ip traffic. <http://wand.net.nz/wits/waikato/8/>. Jun 24, 2021.
- (2017). 3gpp tr 38.901 version 14.0.0 release 14. 5g; study on channel model for frequencies from 0.5 to 100 ghz.
- (2018a). 3gpp ts 38.104 version 15.2.0 release 15. 5g; nr; base station (bs) radio transmission and reception.
- (2018b). 3gpp ts 38.211 version 15.2.0 release 15. 5g; nr; physical channels and modulation.
- Belgiovin, M., Sankhe, K., Bocanegra, C., Roy, D., and Chowdhury, K. R. (2021). Deep learning at the edge for channel estimation in beyond-5g massive mimo. *IEEE Wireless Communications*, 28(2):19–25.
- Ciou, S., Kao, J., Lee, C. Y., and Chen, K. (2015). Multi-sharing resource allocation for device-to-device communication underlaying 5g mobile networks. In *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1509–1514.
- Ferreira, M. V., Vieira, F. H., and Álisson A. Cardoso (2022). Power and delay optimization based uplink resource allocation for wireless networks with device-to-device communications. *Computer Communications*, 194:226–239.
- Ferreira, M. V. G. and Vieira, F. H. T. (2021). Delay minimization based uplink resource allocation for device-to-device communications considering mmwave propagation. *PeerJ Computer Science*, 7:e462.
- Karipidis, K., Mate, R., Urban, D., Tinker, R., and Wood, A. (2021). 5g mobile networks and health - a state-of-the-science review of the research into low-level rf fields above 6 ghz. *Journal of Exposure Science and Environmental Epidemiology*, 31:585–605.
- Sreedevi, A. and Rama Rao, T. (2019). Reinforcement learning algorithm for 5g indoor device-to-device communications. *Transactions on Emerging Telecommunications Technologies*, 30(9):e3670.
- Sun, H., Sheng, M., Wang, X., Zhang, Y., Liu, J., and Wang, K. (2013). Resource allocation for maximizing the device-to-device communications underlaying lte-advanced networks. In *2013 IEEE/CIC International Conference on Communications in China - Workshops (CIC/ICCC)*, pages 60–64.