

ESCALONAMENTO DINÂMICO DE RECURSOS EM REDES 5G

Eduardo Fabricio Notari e Carlos Marcelo Pedroso

Resumo—Entre as novas aplicações suportadas pelo 5G New Radio destaca-se o *massive Machine Type Communication* (mMTC), que exige baixo atraso de propagação, baixa probabilidade de perda de pacotes e uma alta densidade de dispositivos, com o *throughput* normalmente muito baixo. Ao mesmo tempo, a rede tem que atender aplicações que exigem grandes volumes de dados como por exemplo a transmissão de vídeo. A convivência destes dois tipos de aplicação é difícil, uma vez que os requisitos envolvidos são opostos. Este artigo propõe o algoritmo *Dynamic Priority Scheduler* (DPS), capaz de identificar automaticamente e priorizar a transmissão de aplicações mMTC. Resultados simulados mostram que o método proposto permite que um maior número de aplicações mMTC possa existir na rede rádio.

Palavras-Chave—5G NR, Escalonamento, Internet das Coisas

Abstract—Among the new applications supported by 5G New Radio stands out the *massive Machine Type Communication* (mMTC), which requires low latency, low packet loss ratio, and high device density, commonly with low throughput. At the same time, the network has to serve applications that require the transmission of large volumes of information, especially video streaming. The coexistence of these two types of applications is difficult since the requirements involved are opposites. This article proposes the *Dynamic Priority Scheduler* (DPS) algorithm, capable of automatically identify and prioritize the transmission of mMTC applications. Simulated results show that the proposed method allows a greater number of mMTC applications in the radio area network.

Keywords—5G NR, Scheduler, Internet of Things

I. INTRODUÇÃO

Com a evolução da indústria e a necessidade de atender a demanda das novas aplicações, desenvolveu-se o *5G New Radio* (5G NR), projetado para atender um número massivo de dispositivos, suportar as tecnologias anteriores e ter um melhor aproveitamento do espectro [1]. Suporta aplicações como *enhanced Mobile Broadband* (eMBB), que necessitam de grandes taxas de transmissão (ex. vídeo streaming), proveniente do *Long Term Evolution* (LTE), e habilita novas aplicações como o *massive Machine Type Communication* (mMTC), que representa aplicações IoT (*Internet of Things*) e aplicações de missão-crítica com requisitos exigentes chamado *ultra-Reliable Low-Latency Communication* (uRLLC) [2]. A característica típica da comunicação mMTC envolve a transmissão de pequenos pacotes de dados, de forma periódica ou assíncrona. As aplicações eMBB são menos sensíveis à variação da qualidade do serviço [3]. A convivência entre aplicações mMTC e eMBB é o alvo deste trabalho. O suporte à

aplicações uRLLC [4] exige o uso de escalonadores projetados para promover qualidade de serviço e não foi alvo deste estudo.

O 5G utiliza o *Orthogonal Frequency-Division Multiple Access* (OFDMA) que organiza as sub-portadoras de maneira ortogonal, evitando a interferência inter-canal. A técnica de modulação é definida de acordo com a estimativa da relação sinal-ruído do canal. A modulação trabalha com símbolos que representam de 1 à 8 bits, dependendo da técnica de modulação usada, que por sua vez é escolhida em função da estimação da qualidade do canal, que vai afetar também a eficiência do código [5]. O sistema 5G NR utiliza *frames* de transmissão de período 10 ms subdivididos em *sub-frames* de 1 ms, que, como no sistema LTE, é o menor intervalo de transmissão chamado de *TTI* (*Time Transmission Interval*) [6].

Um *Resource Block* (RB) é o menor elemento de alocação de recursos na rede de rádio. Cada RB carrega 14 símbolos em um TTI, em 12 sub-portadoras de largura de banda variável definida pela numerologia escolhida [7]. Desta forma, a eficiência espectral de um RB varia dependendo da qualidade do canal estimada, o que é fundamental para os algoritmos de escalonamento [8]. O número de RBs em um TTI varia com a largura de banda, a numerologia e a banda de guarda [9].

De forma a transmitir os dados dos usuários, as UEs (*User Equipments*) competem pelos RBs disponíveis, que são alocados pela estação rádio-base (chamada de gNodeB). Os algoritmos escalonadores [10] determinam como os RBs serão distribuídos para as UEs [11]. Os algoritmos de escalonamento mais usados são o *Round Robin* (RR), o *Proportional Fair* (PF) e o *Maximum Throughput* (MT) [12]. Neste trabalho será abordado um cenário onde existem aplicações eMBB e mMTC competindo por recursos.

A coexistência entre as aplicações 5G foi pesquisada Kumar et. al. [13], onde diversos algoritmos escalonadores são descritos. Zhang et. al. [14] demonstra um algoritmo que deriva a probabilidade de perdas e maximiza o *throughput* na coexistência de serviços eMBB com as restrições dos serviços uRLLC. Katila et al. [15] em um algoritmo complexo chamado N-PF, utiliza diversos parâmetros das transmissões para escalar usuários mMTC e eMBB coexistentes, demonstrando a superioridade do algoritmo em relação ao PF. Em outro estudo, Liu et. al. [16] sugere um algoritmo que otimiza a eficiência energética e espectral através de um processo de decisão de Markov com aplicações mMTC e eMBB concorrentes, porém com alta complexidade computacional.

Este artigo propõe um novo algoritmo de escalonamento capaz de identificar e priorizar aplicações mMTC na rede. Os resultados simulados mostram que o uso do método proposto

resulta em um incremento no número de dispositivos mMTC suportados na rede rádio.

Este artigo está organizado da seguinte forma. A Seção 2 apresenta os principais algoritmos de escalonamento. A Seção 3 descreve o DPS, com respectiva avaliação de desempenho mostrada na Seção 4. Finalmente as conclusões são apresentadas na Seção 5.

II. ESCALONAMENTO DE RECURSOS NO 5G

O algoritmo de escalonamento é responsável por determinar qual UE pode utilizar cada RB disponível. O mapa de utilização pode ser atualizado a cada milissegundo e enviado para as UEs utilizando um canal de sinalização [17].

A decisão do escalonador pode ser realizada com informações coletadas pelo sistema, como o estado da fila de transmissão, a qualidade do canal reportada pelo *Channel Quality Indicator* (CQI), o histórico de alocação, o estado do *buffer*, entre outros [12].

A. Principais Algoritmos de Escalonamento

A estratégia de escalonamento pode ser definida formalmente através da métrica $m_{j,k} = \max_i \{m_{i,k}\}$, onde j é o índice da UE que receberá a alocação do RB de índice k . O índice i varia entre todas as UEs que desejam receber recursos [18]. Desta forma, o RB_k será alocado para a UE_j que possuir o maior valor para a métrica $m_{i,k}$.

1) Round Robin (RR)

Este método atribui os RBs em esquema de fila circular de acordo com $m_{i,k}^{RR} = t - T_i$, onde T_i representa o último instante de tempo que a UE_i recebeu a alocação de um RB. A alocação não depende da qualidade de canal e pode gerar situações onde *throughput* dos fluxos fique muito diferente entre as UEs, uma vez que a taxa de transmissão possível com um RB varia de acordo com a estimativa da qualidade de canal.

2) Maximum Throughput (MT)

O objetivo do *Maximum Throughput* (MT) é obter o maior *throughput* médio na célula rádio, mesmo que isto sacrifique a justiça na distribuição de recursos. A distribuição de recursos é realizada conforme $m_{i,k}^{MT} = \log[1 + \text{SNIR}_k^i(t)]$, onde $\text{SNIR}_k^i(t)$ é a relação sinal ruído mais interferência (*signal noise plus interference ratio*) estimada para a UE_i em relação ao RB_k .

Neste caso, a UE que possuir melhor qualidade de canal irá receber a alocação do RB disponível. Como a UE que possui melhor qualidade de canal estará usando a técnica de modulação com maior número de bits por símbolo, o *throughput* médio será o maior possível. No entanto, no pior caso, é possível que determinadas UEs nunca consigam alocar recursos, gerando grandes injustiças.

3) Proportional Fair (PF)

O objetivo do PF é obter justiça na divisão do *throughput* entre as UEs, utilizando $m_{i,k}^{PF} = \log[1 + \text{SINR}_k^i(t)/\overline{R_i}(t), \overline{R_i}(t) = \gamma * \overline{R_i}(t-1) + (1-\gamma)r_i(t), 0 \leq \gamma \leq 1$, onde $r_i(t)$ é a taxa de transmissão instantânea da UE_i , $\overline{R_i}(t)$ representa uma média móvel exponencial ponderada de $r_i(t)$ com parâmetro γ . No numerador, $\text{SNIR}_k^i(t)$ representa a relação sinal ruído mais interferência da RB_k para a UE_i .

Com o uso do PF as UEs que possuem qualidade de canal mais baixa tendem a receber a alocação de um número maior

de RBs para obter um *throughput* médio (considerando a média móvel exponencial ponderada) equivalente às UEs que possuem uma melhor qualidade de canal.

Nenhuma das estratégias descritas é capaz de priorizar o tráfego mMTC em relação ao eMBB. Para que isto ocorra, deveria ser usado um escalonador projetado para suportar qualidade de serviço. No entanto, nestes casos, a operadora do sistema deve identificar explicitamente quais UEs geram cada tipo de tráfego, implicando em uma relação comercial específica, com a cobrança de valores e reserva de recursos na rede. Neste trabalho o objetivo é produzir um algoritmo capaz de identificar e priorizar o tráfego mMTC sobre o eMBB de forma automática, sem exigir a implementação de um escalonador projetado para qualidade de serviço. Não é de nosso conhecimento a existência de algoritmos de escalonamento projetados com este propósito na literatura.

III. MÉTODO PROPOSTO

O algoritmo proposto é o chamado *Dynamic Priority Scheduler* (DPS), inspirado nos algoritmos de escalonamento utilizados em sistemas operacionais de tempo compartilhado. Nestes sistemas o escalonador de processos identifica automaticamente processos consumidores de CPU e interativos. O objetivo é garantir que processos interativos sejam escalonados o mais brevemente possível para não impactar na qualidade percebida da aplicação. São exemplos de processos interativos as interfaces com usuários, servidores de rede e servidores de banco de dados. A ideia central é que o benefício obtido ao priorizar a aplicação interativa é muito superior ao custo, que é o aumentar o tempo de execução de uma tarefa consumidora de CPU. Neste contexto as aplicações eMBB seriam os processos consumidores de CPU e as aplicações mMTC os processos interativos.

O DPS utiliza duas prioridades: a prioridade base e a prioridade dinâmica. A prioridade base representa a maior prioridade que pode ser atribuída a um UE. Inicialmente a prioridade dinâmica é igual a prioridade base. Um RB será reservado para a UE que possuir a maior prioridade dinâmica. Quando uma UE receber um RB, sua prioridade dinâmica será decrementada em um valor fixo D até o mínimo de 0. Caso uma UE não receba recursos em um dado TTI, sua prioridade dinâmica será incrementada em um valor fixo I , respeitando o limite máximo da prioridade base.

Com o uso do DPS, espera-se obter os seguintes efeitos:

- 1) Aplicações eMBB demandam grande quantidade de recursos e terão sua prioridade dinâmica quase sempre baixa, só recebendo recursos caso não hajam aplicações mMTC solicitando recursos.
- 2) Aplicações mMTC raramente solicitam recursos, de forma que a tendência é que sua prioridade dinâmica seja quase sempre alta. Desta forma, quando uma aplicação mMTC solicitar recursos ela terá prioridade alta se comparada à aplicações eMBB.

Pode haver redução nos recursos alocados às aplicações eMBB para acomodar este tipo de priorização. No entanto, isto é aceitável e a maioria das aplicações eMBB é capaz de adaptar-se às condições de congestionamento como, por exemplo, o vídeo *streaming* adaptativo amplamente utilizado

atualmente. Desta forma, a métrica para reserva de recursos é definida como:

$$m_{i,k}^{\text{DPS}} = P_i^d + \frac{m_{i,k}^{\text{RR}}}{\max_i \{m_{i,k}^{\text{RR}}\}} \quad (1)$$

onde P_i^d é a prioridade dinâmica da UE_{*i*}, e $m_{i,k}^{\text{RR}}$ é a métrica para o *Round Robin*.

A prioridade dinâmica inicial de uma UE será igual à prioridade base, denotada por P_i^b . O Algoritmo 1 mostra os detalhes do incremento e decremento da prioridade dinâmica que deve ser executado a cada TTI pela gNodeB. Para cada tempo de transmissão TTI o algoritmo calcula a métrica proposta e determina qual UE_{*j*} que receberá o RB. A UE que recebeu alocação tem sua prioridade dinâmica decrementada por D e as UEs que não receberam alocação tem um contador incrementado por 1. Quando o contador atingir o limite I , sua prioridade dinâmica será incrementada por 1.

Algoritmo 1 Escalonamento com Prioridade Dinâmica

```

1: Runs every TTI
2: C[active UE] = {0}
3: for k ∈ available RB do
4:   for i ∈ active UE do
5:     mj,k = maior{mi,k, mi,kDPS};
6:     Ci = Ci + (i == j?0; 1);
7:   end for
8:   Pjd = Pjd - (Pjd > 0?D; 0);
9: end for
10: for i ∈ active UE do
11:   if Ci ≥ I then
12:     Pid < Pib?Pid = Pid + 1; Pid = Pib
13:     Ci = 0
14:   end if
15: end for
    
```

Caso haja um empate na prioridade dinâmica em (1) é utilizada a métrica para o algoritmo Round Robin. Para utilizar o PF como critério de desempate, a métrica pode ser escrita como:

$$m_{i,k}^{\text{DPS}} = P_i^d + \frac{m_{i,k}^{\text{PF}}}{\max_i \{m_{i,k}^{\text{PF}}\}} \quad (2)$$

onde P_i^d é a prioridade dinâmica e $m_{i,k}^{\text{PF}}$ a métrica para o *Proportional Fair*.

Caso não hajam dispositivos mMTC na rede rádio, o comportamento no limite converge para o PF. Para avaliação de desempenho foi utilizado (2).

IV. AVALIAÇÃO DO DESEMPENHO

O método proposto foi implementado no NS-3 (*Network Simulator*) [19] com o módulo 5G-LENA [20] [21].

A. Topologia

Foi configurado no NS-3 um sistema contendo uma gNodeB com UEs tipo eMBB e mMTC adicionados durante a execução dos testes [20], posicionados dentro de um raio de 40.000 metros, em um cenário de uso rural. O ambiente consiste em um servidor de aplicação mMTC e outro eMBB, conectados ao *User Plane Function* (UPF) e este conectado à gNodeB. Os servidores utilizam UDP, que faz a transmissão de dados no sentido do *downlink*, ou seja, do servidor para as UEs. Desta forma, o escalonador projetado será avaliado considerando o tráfego concorrente no *downlink* da rede rádio.

A conexão entre servidor de aplicação e a UPF foi realizada com um enlace de taxa muito alta (10 Gbps) e atraso desprezível (1 ns) para não influenciar nas medições de desempenho, uma vez que deseja-se avaliar o desempenho da RAN.

O posicionamento das UEs na RAN foi realizado de forma a posicionar as UEs entre 0 a 40.000 metros da gNodeB, com a variação de distâncias proporcional à quantidade de UEs no teste. Por exemplo, com 4 UEs, estas estarão posicionadas com distâncias respectivamente de 10.000, 20.000, 30.000 e 40.000 metros da gNodeB.

O modelo de propagação utilizado nos testes foi o de espaço livre. Este é um modelo simples, que supõem que os elementos se comunicando estejam em visada direta, sem obstáculos. O uso deste modelo não implica em perda de generalidade na avaliação do método proposto, uma vez que para a avaliação de desempenho do método proposto é necessário que haja variação da qualidade de canal, mas a forma como isto ocorre não impede ou invalida os resultados. No entanto, com o uso do modelo de propagação de espaço livre, a distância máxima entre a gNodeB e as UEs foi mantida em 40.000 metros para garantir que haja variação de qualidade de canal entre as UEs.

B. Parâmetros de Simulação

Para esta simulação foram utilizados os conceitos de testes descritos na norma TS-38.141 do 3GPP [22]. Foi definida a frequência central de 3,510 GHz, um canal com a largura de banda de 20 MHz e numerologia 0 com espaçamento da subportadora de 15 kHz. Este valor foi utilizado para reduzir a quantidade de UEs necessárias para saturar o sistema. Utilizou-se o RMa_LOS que é o cenário rural com visada direta entre a estação rádio-base e UEs. Para a gNodeB foi utilizada uma altura de 50 m e potência de 30 dBm e para os UEs uma altura de 1,5 m e potência de 28 dBm. Os testes foram realizados para os algoritmos RR, PF, MT e DPS com 600 s de duração.

C. Geradores de Tráfego

Para cada tipo de aplicação foi utilizada uma forma de gerar o tráfego esperado. No caso do mMTC foi realizada uma taxa constante de transmissão de 360 kbps [23].

Para o eMBB, foi utilizado o modelo ON/OFF [24], com o estado ON sendo modelado por uma distribuição de Pareto e o estado OFF por uma distribuição exponencial. Este é um dos modos mais populares para geração de tráfego de vídeo com características auto-similares. Neste caso, o parâmetro de *Hurst* é dado por $H = (3 - \alpha)/2$ [25], onde α é o parâmetro de forma da distribuição de Pareto. Deseja-se obter $H = 0,85$, que é um valor típico para transmissão de vídeo, logo α pode ser calculado como 1,3.

Considere que o tempo de permanência nos estados ON e OFF são representados respectivamente pelas variáveis aleatórias X e Y . A esperança para a distribuição de Pareto é dada por $E[X] = \alpha\beta/(\alpha - 1)$. Deseja-se imitar o tráfego de vídeo a 30 fps, sendo definido $E[X] = 1/60$ (tempo ON + tempo OFF = $1/30$). Com isso foi obtido o parâmetro β , com valor de 0,003846154. Para o estado ON, foi definido $E[Y] = 1/60$. Como a esperança da distribuição exponencial é dada por $E[Y] = \mu$, então $\mu = 1/60$. A taxa média resultante pode ser obtida através da taxa de transmissão dividida por 2, pois a esperança do tempo ON foi definida igual à esperança do

tempo OFF. A taxa de geração do fluxo eMBB foi configurada em 5,8 Mbps, o que leva a uma taxa média de transmissão de 2,9 Mbps por fluxo.

Para o parâmetro D foi atribuído o valor 1 para o decremento no cálculo da prioridade dinâmica. A prioridade base foi fixada em 20 para todos os UEs. Para a variável I foram realizados diversos testes com os valores de 10, 12, 20, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 48, 50, 51, 52, 53, 54, 60 e 72. Após verificar os resultados obtidos, o valor de 53 foi escolhido por apresentar melhor desempenho.

D. Resultados

1) Teste 1

As Figuras 1 e 2 apresentam as respostas das simulações para os dispositivos mMTC e eMBB respectivamente. Nesta simulação, 3 dispositivos eMBB estão ativos simultaneamente, transmitindo a taxas de 2,8 Mbps. Com a introdução de 0 a 40 dispositivos mMTC, pode-se notar que o algoritmo proposto tem decaimento do *throughput* menor dos dispositivos mMTC em relação aos concorrentes.

Observa-se também que o atraso médio para os dispositivos mMTC foi melhor em relação aos concorrentes. O atraso médio do MT foi baixo, mas isso ocorre porque diversos fluxos não realizaram transmissão alguma, o que melhora o atraso médio artificialmente.

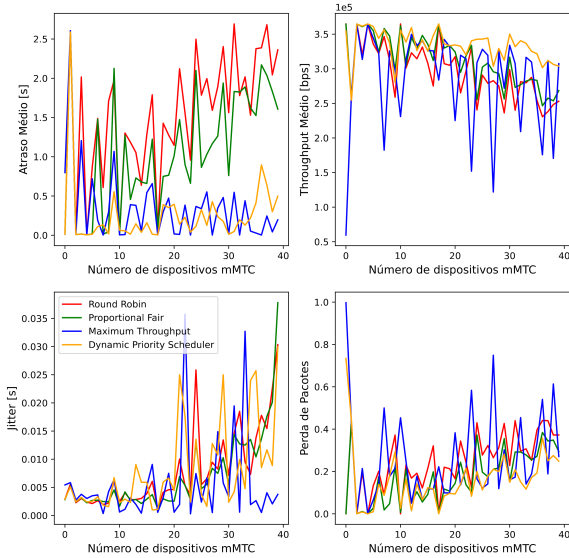


Fig. 1: Resultados do teste 1 para usuários mMTC

2) Teste 2

Nas Figuras 3 e 4, são apresentados os resultados simulados para as aplicações mMTC e eMBB respectivamente. Foram utilizados 5 usuários eMBB, cada um transmitindo 2,8 Mbps, totalizando 14 Mbps. Na Figura 3, nota-se a diminuição do *throughput* nos algoritmos concorrentes, enquanto o DPS continua constante, suportando uma quantidade maior de aplicações com as taxas de transmissão esperadas. Nota-se que ao atingir 40 usuários mMTC o *throughput* médio está 300 Kbps para o DPS, 250 Kbps para o PF e 200 Kbps para o MT e RR.

Os resultados também mostram que o *throughput* das aplicações eMBB não foi reduzido severamente em função da inserção de dispositivos mMTC. O *jitter* apresenta similaridade entre todos os algoritmos, indicando a baixa variação

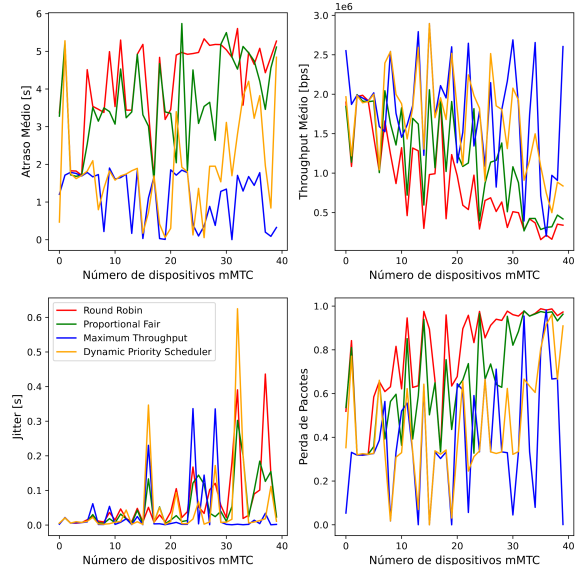


Fig. 2: Resultados do teste 1 para usuários eMBB

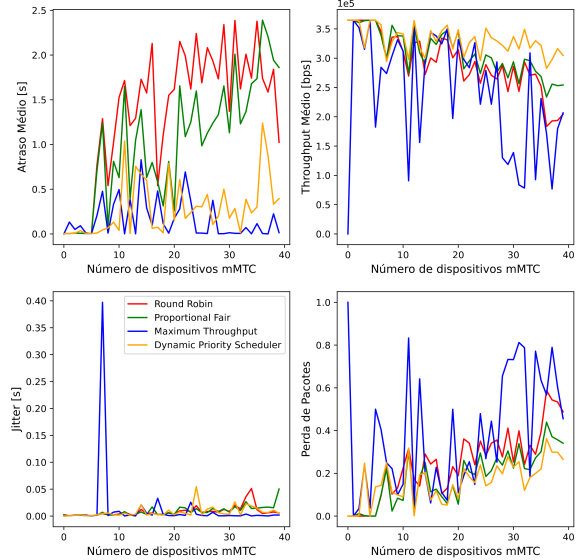


Fig. 3: Resultados do teste 2 para usuários mMTC

no atraso dos pacotes transmitidos.

Além disso, foi avaliada a justiça na distribuição de recursos e determinou-se uma diferença significativa de desempenho entre os diversos fluxos dependendo do algoritmo de escalonamento usado. Foi utilizado o índice de Jain [26], definido por $J = (\sum_{j=1}^n x_j)^2 / n \sum_{j=1}^n x_j^2$, onde J é o índice de justiça, n é o número de medidas e x_j representa uma medida. O valor de J deve estar entre 0 e 1, sendo 1 representando a maior justiça possível e 0 a menor.

O índice de Jain do *throughput* das aplicações mMTC foi calculado para os testes 1 e 2, e apresentados na Figura 5. O algoritmo DPS apresenta uma justiça de 0,9, enquanto o PF 0,73, e o MT e RR 0,6 aproximadamente, resultando em um ganho de 23% e 50% respectivamente, comprovando-se a superioridade do DPS em relação aos algoritmos concorrentes.

V. CONCLUSÕES

Durante as simulações foi verificado que o algoritmo proposto suportou uma taxa de transmissão maior para as aplicações

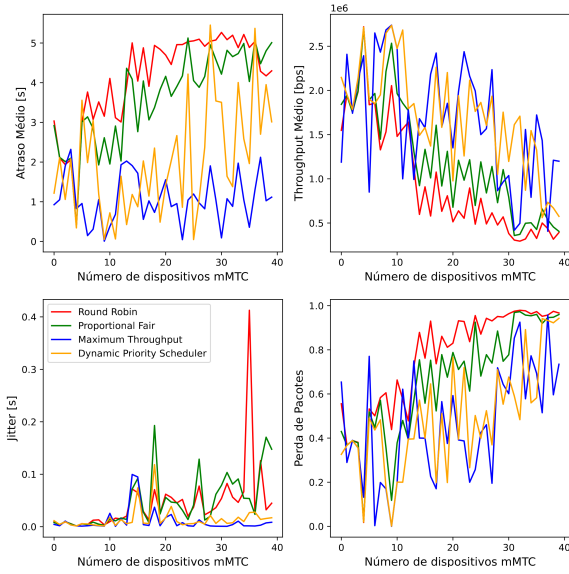


Fig. 4: Resultados do teste 2 para usuários eMBC

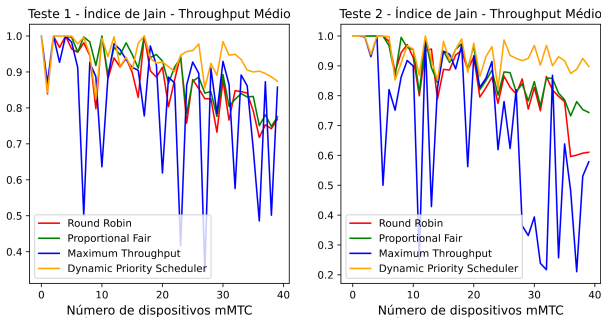


Fig. 5: Índice de Jain calculado para os testes 1 e 2 respectivamente.

mMTC em comparação com os algoritmos concorrentes em situação de congestionamento. Os resultados indicam que o menor atraso observado nas aplicações mMTC foi obtido a um custo do aumento no atraso das aplicações eMBC. No entanto, a penalização imposta às aplicações eMBC não foi grande e pode ser absorvida pelas aplicações, que são majoritariamente de *video streaming* e transferência de hipertexto. O impacto ao usuário eMBC seria uma pequena redução da qualidade em situações de congestionamento, o que se configura numa boa troca, considerando o ganho com o suporte à um número maior de aplicações mMTC.

O algoritmo de escalonamento desenvolvido permite a convivência de aplicações eMBC e mMTC sem a necessidade de implementar mecanismos de qualidade de serviço em redes 5G. O algoritmo proposto tem capacidade de identificar o tráfego e priorizá-lo na utilização dos recursos disponíveis na estação rádio-base. Nos testes realizados no simulador NS-3, verifica-se a superioridade de 23% ao PF, e 50% ao MT e RR na justiça entre os usuários mMTC envolvidos no sistema.

REFERÊNCIAS

[1] CISCO, “Global - 2021 Forecast High- lights, VNI Complete Forecast Highlights.” https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_2021_Forecast_Highlights.pdf, 2021. Acessado em 30/11/2021.

[2] M. Sauter, *From GSM to LTE-Advanced Pro and 5G: An Introduction to Mobile Networks and Mobile Broadband*. Wiley, 2021.

[3] N. Javaid, A. Sher, H. Nasir, and N. Guizani, “Intelligence in IoT-Based

5G Networks: Opportunities and Challenges,” *IEEE Communications Magazine*, vol. 56, no. 10, pp. 94–100, 2018.

[4] ITU, “Minimum requirements related to technical performance for IMT-2020 radio interface(s).” ITU-R Reports 2410-0, (ITU) - International Telecommunication Union, 2017.

[5] M. Abu-Rgheff, *5G Physical Layer Technologies*. IEEE Press, Wiley, 2019.

[6] 3GPP, “5G NR. Technical Specification Group Radio Access Network, NR, and NG-RAN Overall Description (Release 15),” Technical Specification (TS) 38.300, 3rd Generation Partnership Project (3GPP), 2018.

[7] A. Zaidi, F. Athley, J. Medbo, U. Gustavsson, G. Durisi, and X. Chen, *5G Physical Layer: Principles, Models and Technology Components*. USA: Academic Press, Inc., 1st ed., 2018.

[8] A. Mamane, M. Fattah, M. El Ghazi, M. Bekkali, B. Younes, and S. Mazer, “Scheduling algorithms for 5g networks and beyond: Classification and survey,” *IEEE Access*, vol. 10, pp. 1–1, 01 2022.

[9] 3GPP, “5G NR. Physical channels and modulation (Release 15),” Technical Specification (TS) 38.211, 3rd Generation Partnership Project (3GPP), 2019.

[10] S. Basilashvili, Grigol & Gogokhia, “Customized Packet Scheduling Algorithm for LTE Network,” *American Scientific Research Journal for Engineering, Technology, and Sciences.*, vol. 40, pp. 203–216, 2018.

[11] 3GPP, “5G NR. User Equipment (UE) radio transmission and reception (Release 15),” Technical Specification (TS) 38.101-1, 3rd Generation Partnership Project (3GPP), 2018.

[12] T. P. F. Maryam Imran Sheik Mamode, “Survey of Scheduling Schemes in 5G Mobile Communication Systems,” *Journal of Electrical Engineering, Electronics, Control and Computer Science – JEECCS*, vol. 6, no. 20, pp. 21–30, 2020.

[13] R. Kumar, D. Sinwar, and V. Singh, “Qos aware resource allocation for coexistence mechanisms between eMBB and urllc: Issues, challenges, and future directions in 5g,” *Computer Communications*, vol. 213, pp. 208–235, 2024.

[14] J. Shen, Y. Zhao, Y. Zhang, D. Xu, T. Song, and W. Wang, “Dynamic traffic scheduling strategy for the coexistence of urllc and eMBB services in power communication,” in *2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC)*, vol. 6, pp. 1334–1338, 2022.

[15] C. J. Katila, C. Buratti, M. D. Abrignani, and R. Verdone, “Neighbors-aware proportional fair scheduling for future wireless networks with mixed MAC protocols,” *EURASIP Journal on Wireless Communications and Networking*, no. 93, 2017.

[16] Z. Liu, X. Chen, Y. Chen, and Z. Li, “Deep reinforcement learning based dynamic resource allocation in 5G ultra-dense networks,” in *2019 IEEE International Conference on Smart Internet of Things (SmartIoT)*, pp. 168–174, 2019.

[17] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, “Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey,” *IEEE Communications Surveys and Tutorials*, vol. 15, no. 2, pp. 678–700, 2013.

[18] R. Subramanian, P. Ghosal, S. Barua, S. Xing, S. C. Lam, H. Al Kim, and K. Sandrasegaran, “Survey of LTE Downlink Schedulers Algorithms in Open Access Simulation Tools NS-3 and LTE-SIM,” *International Journal of Wireless & Mobile Networks*, vol. 7, pp. 01–16, 04 2015.

[19] G. F. Riley and T. R. Henderson, “The ns-3 network simulator,” in *Modeling and Tools for Network Simulation* (K. Wehrle, M. Günes, and J. Gross, eds.), pp. 15–34, Springer, 2010.

[20] K. Koutlia, B. Bojovic, Z. Ali, and S. Lagen, “Calibration of the 5G-LENA System Level Simulator in 3GPP reference scenarios,” 2022.

[21] N. Patriciello, S. Lagen, B. Bojovic, and L. Giupponi, “An E2E Simulator for 5G NR Networks,” 2019.

[22] 3GPP, “5G NR; Base Station (BS) conformance testing Part 1: Conducted conformance testing Release 15,” Technical Specification (TS) 38.141-1, 3rd Generation Partnership Project (3GPP), 2019.

[23] F. Mehmeti and T. F. La Porta, “Modeling and Analysis of mMTC Traffic in 5G Base Stations,” in *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*, pp. 652–660, 2022.

[24] O. A. Adeleke, N. Bastin, and D. Gurkan, “Network traffic generation: A survey and methodology,” *ACM Comput. Surv.*, vol. 55, jan 2022.

[25] V. Rusev and A. Skorikov, “The Asymptotics of Moments for the Remaining Time of Heavy-Tail Distributions,” *Computer Sciences & Mathematics Forum*, vol. 7, no. 1, 2023.

[26] R. Jain, *The art of computer systems performance analysis - techniques for experimental design, measurement, simulation, and modeling*. Wiley professional computing, Wiley, 1991.