

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E**  
**INFORMÁTICA INDUSTRIAL**

**FABIAN MAURICE MALHEIROS FRANCO**

**LATENCY-RATE DOWNLINK PACKET SCHEDULER PARA REDES**  
**LTE**

**DISSERTAÇÃO**

**CURITIBA**

**2018**

**FABIAN MAURICE MALHEIROS FRANCO**

**LATENCY-RATE DOWNLINK PACKET SCHEDULER PARA REDES  
LTE**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial da Universidade Tecnológica Federal do Paraná como requisito parcial à obtenção do título de “Mestre em Ciências” – Área de Concentração: Telemática.

Orientador: Prof. Dr. Emilio C. G. Wille

Coorientador: Prof. Dr. Augusto Foronda

**CURITIBA**

**2018**

#### Dados Internacionais de Catalogação na Publicação

F825L Franco, Fabian Maurice Malheiros  
2018 Latency-Rate Downlink Packet Scheduler para redes LTE /  
Fabian Maurice Malheiros Franco.-- 2018.  
77 p.: il.; 30 cm.

Disponível também via World Wide Web.

Texto em português com resumo em inglês.

Dissertação (Mestrado) - Universidade Tecnológica  
Federal do Paraná. Programa de Pós-Graduação em Engenharia  
Elétrica e Informática Industrial. Área de concentração:  
Telecomunicações e Redes, Curitiba, 2018 .

Bibliografia: p. 73-75.

1. Evolução a longo prazo (Telecomunicações). 2. Sistemas de  
comunicação móvel. 3. Qualidade de serviço (Redes de  
computação). 4. Comutação de pacotes (Transmissão de dados).  
5. Alocação de recursos. 6. Agendamento. 7. Multiplexação de  
divisão de frequência ortogonal. 8. Desempenho de rede  
(Telecomunicações). 9. Telecomunicações - Tráfego. 10.  
Simulação (Computadores). 11. Engenharia elétrica - Dissertações.  
I. Wille, Emílio Carlos Gomes, orient. II. Foronda, Augusto,  
coorient. III. Universidade Tecnológica Federal do Paraná.  
Programa de Pós-Graduação em Engenharia Elétrica e Informática  
Industrial. IV. Título.

CDD: Ed. 23 -- 621.3

## TERMO DE APROVAÇÃO DE DISSERTAÇÃO Nº 802

A Dissertação de Mestrado intitulada “**Latency-Rate Downlink Packet Scheduler para redes LTE**” defendida em sessão pública pelo(a) candidato(a) **Fabian Maurice Malheiros Franco**, no dia 30 de agosto de 2018, foi julgada para a obtenção do título de Mestre em Ciências, área de concentração Telecomunicações e Redes, e aprovada em sua forma final, pelo Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial.

BANCA EXAMINADORA:

Prof(a). Dr(a). Emilio Carlos Gomes Wille - Presidente (UTFPR)

Prof(a). Dr(a). Mauro Sergio Pereira Fonseca - (UTFPR)

Prof(a). Dr(a). Luiz Carlos Pessoa Albini - (UFPR)

A via original deste documento encontra-se arquivada na Secretaria do Programa, contendo a assinatura da Coordenação após a entrega da versão corrigida do trabalho.

Curitiba, 30 de agosto de 2018.

Dedico este trabalho à minha mãe, Nadir Malheiros (*In Memoriam*), e à minha irmã, Karla Adriana Malheiros Franco.

## AGRADECIMENTOS

Agradeço aos meus orientadores, Prof. Dr. Emilio C. G. Wille e Prof. Dr. Augusto Foronda pela oportunidade, apoio e incentivo em todas as fases deste trabalho. Em especial agradeço a paciência que tiveram durante esta jornada, os conhecimentos compartilhados e as contribuições que foram essenciais para a realização desta dissertação de mestrado.

Aos professores membros da banca examinadora, Prof. Dr. Mauro Sergio Pereira Fonseca - (UTFPR) e Prof. Dr. Luiz Carlos Pessoa Albini - (UFPR), por aceitarem o convite para participação na defesa da dissertação e pelas sugestões que aprimoraram o trabalho.

Aos funcionários e colaboradores da UTFPR – Curitiba, aos professores do CPGEI, pelos ensinamentos, em especial ao saudoso e inspirador Prof. Dr. Walter Godoy Júnior (*in memoriam*), e à UTFPR por proporcionar a chance de expandir os meus horizontes e me tornar uma pessoa melhor.

Aos colegas da UTFPR – Curitiba, Bruno Coutinho, Euclides Farias, Marcos Talau, Rodrigo Silva, Eden Dosciatti, Joilson Alves, Kleber Nabas, Acácio Zimbico e Charles Gentilin.

Aos colegas de trabalho, Eliezer e Elias Joaquim da UTFPR – Ponta Grossa por assumirem minhas funções quando estive ausente e ao Prof. Dr. Antônio Augusto de Paula Xavier pelo incentivo e apoio.

À minha amada mãe (*in memoriam*), onde quer que esteja, deixo meu agradecimento especial, por sua incansável dedicação em minha criação e educação, e por todas as lições de amor, companheirismo, amizade, caridade, compreensão e perdão.

À minha querida e amada irmã Karla Adriana Malheiros Franco, por todo amor, dedicação, por sempre ter lutado pela minha educação e por nunca desistir de mim.

À minha esposa Emanuelle Ferreira Machado por todo amor, carinho, compreensão e apoio incondicional em tantos momentos difíceis desta caminhada.

À minha sogra Reni, por todo o carinho, amor e dedicação em todos os momentos.

Aos amigos, André e Ana, Bruno e Diana, Cleber e Léia, Danillo, Diniz, Eduardo e Ivonete, Elias, Gemada (Márcio) e Larissa, Jarbas (Alcides), Jacaré (Emerson), Joelcio, Zé (Elson) e Pâmela, pelas palavras de incentivo e momentos de descontração.

*Give me a place to stand, and a lever long  
enough, and I will move the world.*

Archimedes

## RESUMO

MALHEIROS FRANCO, Fabian Maurice. **Latency-Rate Downlink Packet Scheduler para redes LTE**. 2018. 77f. Dissertação (Mestrado em Engenharia Elétrica e Informática Industrial) - Universidade Tecnológica Federal do Paraná. Curitiba, 2018.

O padrão *Long Term Evolution* (LTE) é a principal tecnologia utilizada em redes móveis 4G atualmente em todo o mundo. Sua arquitetura consiste basicamente em três componentes principais: um núcleo de rede, estação base e dispositivos de usuário. Os dispositivos de usuário podem requisitar serviços disponíveis na Internet, tais como serviços de voz, vídeo entre outros serviços de dados de transmissão não contínua como páginas Web. O acesso a esses serviços são gerenciados pela estação base, a qual tem o papel de realizar o escalonamento de recursos, por meio de uma tecnologia de rede de acesso múltiplo, garantindo a qualidade de serviço ao usuário. Assim, um de seus principais desafios está relacionado ao atendimento de serviços que exigem baixas latências e largura de banda. Entre estes, destaque-se os serviços vídeo por seu constante crescimento em termos de consumo de dados em redes móveis. Neste contexto este trabalho apresenta o *Latency-Rate Downlink Packet Scheduler* (LR-DPS) para o escalonamento de recursos de tráfego *downlink* visando atender os requisitos de atraso máximo para tráfegos de taxa variável. Considerando que o escalonamento de recursos em redes LTE é um problema de otimização complexo, a principal contribuição do trabalho é uma nova arquitetura de escalonamento com provisão de QoS para delay máximo. A proposta e formulação do problema são apresentadas em três etapas hierárquicas. Na primeira etapa, o tráfego é condicionado por um balde de fichas. Na segunda etapa, por meio de um modelo analítico, o problema de atribuição de tempo total (ATT) é solucionado e tem-se um tempo total para transmissão de dados e as taxas individuais para atendimento ao *delay* máximo solicitado. Na terceira, os dados são alocados em blocos de recursos de forma equilibrada atendendo os requisitos da etapa anterior. A avaliação de desempenho utiliza o ns-3 como simulador de rede, com o qual são criados cenários realísticos que exploram características da rede LTE e do modelo LR-DPS. Como parâmetro de comparação, são utilizados os escalonadores *Round Robin* e *Channel and QoS Aware*. Os resultados das simulações realizadas com diferentes tráfegos de vídeo com taxa variável (VBR) mostram que o LR-DPS atendeu aos requisitos de QoS quando os demais escalonadores excederam em até 90% o *delay* máximo solicitado.

**Palavras-chave:** *Long Term Evolution* (LTE). Qualidade de Serviço (QoS). Escalonamento de Recursos. *Downlink*.



## ABSTRACT

MALHEIROS FRANCO, Fabian Maurice. **Latency-Rate Downlink Packet Scheduler para redes LTE. 2018.** 77f. Dissertação (Mestrado em Engenharia Elétrica e Informática Industrial) - Universidade Tecnológica Federal do Paraná. Curitiba, 2018.

The Long Term Evolution (LTE) standard is the currently leading technology used in 4G mobile networks around the world. Its basic architecture consists of three main components: a network core, base station, and user devices. User devices can request services available on the Internet, such as voice services, video and other not continuous data transmission services as Web pages. The access to these services are managed by the base station, which has the function of performing the resource scheduling, with a multiple access network technology and guarantee the quality of service to the user. Therefore, one of its main challenges is the support to services that require low latency and bandwidth. Among these, we highlight the video services for their constant growth related to the data consumption in mobile networks. Inside this context, this work presents the Latency-Rate Downlink Packet Scheduler (LR-DPS) for the traffic scheduling of downlink resources to guarantee the maximum delay requirements for variable rate traffic. Considering that the resource allocation in LTE networks is a complex optimization problem, the main contribution of the work is a new scheduling architecture with QoS for maximum delay. Three hierarchical steps introduce the proposal and formulation of the problem. In the first step, a token bucket bound the traffic. In the second step, the total time allocation (TTA) problem is solve, resulting in a total time for data transmission and the individual rates for guarantee the requested maximum delay. The third step allocates the data to resource blocks into a balanced way, meeting the requirements of the previous step. The performance evaluation carried out in the network simulator ns-3, with realistic scenarios that explore LTE network features and the LR-DPS model. In order to evaluate the performance, was employed the Round Robin and Channel and QoS Aware schedulers. The results of the simulations performed with different variable rate video (VBR) traffic presenting that the LR-DPS met the QoS requirements when the other schedulers exceeded the requested maximum delay by up to 90%.

**Keywords:** Long Term Evolution (LTE). Quality of Service (QoS). Resource allocation. Downlink.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Principais tecnologias em redes móveis.....	21
Figura 2 – Arquitetura de rede LTE .....	21
Figura 3 – Pilha de protocolos LTE.....	23
Figura 4 – Multiplexação de dados em interface aérea. ....	24
Figura 5 – Multiplexação de símbolos em subportadoras .....	25
Figura 6 – Inserção de prefixo cíclico .....	25
Figura 7 – Estrutura de frame LTE.....	26
Figura 8 – Grade de recursos LTE no tempo e frequência utilizando prefixo cíclico normal .	27
Figura 9 – Multiplexação de dados em interface aérea .....	30
Figura 10 – Modelo do escalonador <i>Latency Rate</i> .....	37
Figura 11 – Representação do algoritmo do balde furado.....	38
Figura 12 – Tráfego de entrada modelado por um balde de fichas .....	39
Figura 13: Alocação por tempo de quadro em redes WiMAX.....	40
Figura 14: Alocação de recursos em redes LTE.....	41
Figura 15: Alocação de recursos em redes LTE.....	42
Figura 16: Arquitetura <i>Latency-Rate Downlink Packet Scheduler</i> .....	43
Figura 17: Aplicação do modelo em etapas.....	44
Figura 18: Aplicação do balde de fichas .....	45
Figura 19: Conjunto de TRBGs para alocação .....	47
Figura 20: Exemplo de conjunto de TRBGs .....	49
Figura 21: Alocação de RBGs por meio do Algoritmo 1 .....	50
Figura 22 – Topologia de simulação para validar o modelo proposto .....	52
Figura 23 – Tráfego variável utilizado .....	54
Figura 24 – Tráfego VBR tratado pelo LR-DPS .....	56
Figura 25 – Delay máximo para tráfego de vídeo utilizando LR-DPS .....	58
Figura 26 – Tráfegos VBR utilizados para o cenário 2 .....	59
Figura 27 – <i>Delay</i> máximo para dois tráfegos de vídeo .....	61

Figura 28 – <i>Throughput</i> mínimo para dois tráfegos de vídeo.....	62
Figura 29 – <i>Throughput</i> médio para dois tráfegos de vídeo .....	63
Figura 30 – <i>Delay</i> máximo em banda de 10 MHZ .....	64
Figura 31 – <i>Delay</i> máximo em banda de 15 MHZ .....	64
Figura 32 – Quantidade de UEs alocados em diferentes larguras de banda .....	65
Figura 33 – Quantidade de UEs atendidos em diferentes proporções de tráfegos.....	66
Figura 34 – Índice de Justiça para <i>delay</i> individual em diferentes proporções de tráfegos.....	67
Figura 35 – Quantidade de UEs atendidos em diferentes requisitos de <i>delay</i> .....	68

## LISTA DE TABELAS

Tabela 1 – Larguras de banda suportadas em redes LTE .....	27
Tabela 2 – Esquema de modulação em redes LTE.....	28
Tabela 3 – Relação entre a largura de banda e o tamanho do RBG .....	29
Tabela 4 – Propostas de escalonamento em redes LTE.....	33
Tabela 5 – Parâmetros de Simulação.....	53
Tabela 6 – Parâmetros do escalonador LR-DPS .....	55
Tabela 7 – Tráfego VBR de entrada.....	56
Tabela 8 – Tráfego de saída.....	57
Tabela 9 – Delay máximo medido.....	58
Tabela 10 – Parâmetros do balde de fichas .....	60
Tabela 11 – Quantidade de estações por tráfego requerido.....	60
Tabela 12 – Parâmetros de simulação para dois tráfegos distintos .....	66
Tabela 13 – Parâmetros de simulação para dois <i>delays</i> distintos.....	67
Tabela 15 – Comparação entre as propostas (Cenário 1).....	77
Tabela 16 – Comparação entre as propostas (Cenário 2).....	77
Tabela 17 – Comparação entre as propostas (Cenário 3).....	77

## LISTA DE ABREVIATURAS, SIGLAS E ACRÔNIMOS

16-QAM	16-Quadrature Amplitude Modulation
1G	Redes Móveis de Primeira Geração
2G	Redes Móveis de Segunda Geração
3G	Redes Móveis de terceira Geração
3GPP2	3rd Generation Partnership Project 2
3GPP	3rd Generation Partnership Project
4G	Redes Móveis de Quarta Geração
64-QAM	64-Quadrature Amplitude Modulation
AMPS	Analogue Mobile Phone System
CDMA	Code Division Multiple Access
CQA	Channel and QoS-Aware
CQI	Channel Quality Indicator
C-RNTI	Cell Radio Network Temporary Identifier
CSD	Circuit Switched Data
CTTC	Centre Tecnològic de Telecomunicacions de Catalunya
DL-SCH	Downlink Shared Channel
EDGE	Enhanced Data Rates for GSM Evolution
ENODEB	Evolved NodeB
EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
FDD	Frequency Division Duplexing
FDMA	Frequency Division Multiple Access
FFT	Fast Fourier Transform
FM	Frequency Modulation
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communication
HSPA	High Speed Packet Access
IFFT	Inverse Fast Fourier Transform
IP	Internet Protocol
ISI	Inter Symbol Interference

LENA	LTE-EPC Network simulAtor
LTE	Long-Term Evolution
MAC	Medium Access Control
MCS	Modulation Coding Scheme
MIMO	Multiple Input Multiple Output
MME	Mobility Management Entity
NMT	Nordic Mobile Telephone System
NS-3	Network Simulator 3
OFDMA	Orthogonal Frequency Division Multiple Access
PDCCCH	Physical Downlink Control Channel
PDU	Protocol Data Unit
P-GW	PDN Gateway
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RB	Resource Block
RBG	Resource Block Group
RLC	Radio Link Control
RR	Round Robin
S-GW	Serving Gateway
SMS	Short Message Service
SNR	Signal to Noise Ratio
SPS	Semi-Persistent Scheduling
TACS	Total Access Communications System
TCP	Transmission Control Protocol
TTI	Transmission Time Interval
UE	User Equipment
UTMS	Universal Mobile Telecommunication System
VoIP	Voice over IP
WCDMA	Wideband Code Division Multiple Access

## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>15</b>
1.1 CONTEXTO GERAL .....	15
1.2 OBJETIVOS .....	16
1.2.1 Objetivo Geral .....	16
1.2.2 Objetivos Específicos .....	17
1.3 ESTRUTURA DO TRABALHO .....	17
<b>2 REDES MÓVEIS LTE .....</b>	<b>18</b>
2.1 REDES MÓVEIS DE PRIMEIRA GERAÇÃO (1G) .....	18
2.2 REDES MÓVEIS DE SEGUNDA GERAÇÃO (2G) .....	18
2.3 REDES MÓVEIS DE TERCEIRA GERAÇÃO (3G).....	19
2.4 REDES MÓVEIS DE QUARTA GERAÇÃO (4G).....	20
2.5 ARQUITETURA DE REDE LTE .....	21
2.5.1 Camada Física de <i>Downlink</i> .....	23
2.5.2 OFDMA.....	23
2.5.3 Subportadoras .....	24
2.5.4 Inserção de Prefixo Cíclico.....	25
2.6 ESTRUTURA DO QUADRO LTE.....	26
2.7 GRADE DE RECURSOS DOWNLINK.....	26
2.8 LARGURA DE BANDA.....	27
2.9 INDICADOR DE QUALIDADE DE SINAL .....	28
2.10 ALOCAÇÃO DE RECURSOS .....	28
2.10.1 Taxa Física de Transmissão .....	29
2.11 ESCALONAMENTO DE RECURSOS .....	29
<b>3 CONDICIONAMENTO DE TRÁFEGO.....</b>	<b>31</b>
3.1 TRABALHOS RELACIONADOS .....	31
3.2 ESCALONAMENTO <i>ROUND ROBIN</i> .....	34
3.3 ESCALONAMENTO <i>CHANNEL AND QOS AWARE</i> (CQA).....	34
3.4 LATENCY RATE (LR).....	35
3.5 CONDICIONAMENTO DE TRÁFEGO .....	38
<b>4 MODELO DE ESCALONAMENTO COM GARANTIA DE <i>DELAY</i> MÁXIMO ....</b>	<b>40</b>
4.1 DESCRIÇÃO DA PROPOSTA.....	41
4.2 MODELO DO ESCALONADOR LATENCY-RATE DOWNLINK PACKET SCHEDULER .....	43
4.3 MODELO ANALÍTICO DO ESCALONADOR LATENCY-RATE DOWNLINK PACKET SCHEDULER.....	43
<b>5 AVALIAÇÃO DE DESEMPENHO .....</b>	<b>51</b>
5.1 CONFIGURAÇÃO DO AMBIENTE DE SIMULAÇÃO .....	52
5.2 CENÁRIO 1 – VALIDAÇÃO DO MODELO .....	54
5.3 CENÁRIO 2 – ANÁLISE DE DESEMPENHO .....	59

5.1.1 <i>Delay</i> Máximo.....	60
5.1.2 <i>Throughput</i> Mínimo.....	61
5.1.3 <i>Throughput</i> Médio .....	62
5.1.4 <i>Bandwidth</i> .....	63
5.1.5 UEs Alocados em Diferentes Larguras de Banda.....	65
5.1.6 <i>Delay</i> Máximo com Diferentes Proporções de Tráfegos.....	65
5.1.7 UEs Alocados com Diferentes Requisitos de <i>Delay</i> Máximo .....	67
<b>6 CONCLUSÕES E TRABALHOS FUTUROS.....</b>	<b>69</b>
6.1 TRABALHOS FUTUROS.....	71
<b>REFERÊNCIAS .....</b>	<b>73</b>
<b>APÊNDICE A – O PROBLEMA MINIMAX.....</b>	<b>76</b>
<b>APÊNDICE B – COMPARAÇÃO ENTRE PROPOSTAS.....</b>	<b>77</b>



# 1 INTRODUÇÃO

## 1.1 CONTEXTO GERAL

A utilização de dispositivos móveis de comunicação cresce ano a ano em todo o mundo e a quantidade de aparelhos já excede a população em muitos países. Com o aumento do número de aplicações de dados disponíveis por meio da Internet, aumenta também o número de assinaturas para utilização de rede móvel de banda larga. Assim, as chamadas de voz, que eram a principal forma de tráfego utilizada em redes de telecomunicações móveis, perderam espaço para uma demanda por novos serviços. Agora, serviços como voz sobre IP, navegação web, vídeo chamada e vídeo sob demanda, impõem novos desafios em relação à utilização de banda e latência para essas redes.

Novas tecnologias de comunicação são necessárias diante de um novo panorama de utilização de dispositivos móveis. Neste sentido, a atual tecnologia utilizada para implementação de redes de quarta geração, conhecida como *Long Term Evolution* (LTE), busca apresentar uma experiência satisfatória ao usuário final. A parceria entre organizações de desenvolvimento de telecomunicações, intitulada *3rd Generation Partnership Project* (3GPP), desenvolve especificações e normas para expandir a capacidade das redes móveis. O projeto abrange as tecnologias de redes de telecomunicações celulares, incluindo o acesso por rádio e a rede central de transporte (3GPP, 2017).

A rede de acesso LTE tem como base um esquema de acesso múltiplo por divisão de frequências ortogonais (OFDMA) e suporta uma ampla gama de serviços de multimídia e Internet, mesmo em cenários de alta mobilidade. Ela foi projetada para fornecer altas taxas de dados, baixa latência e uma eficiência espectral melhorada em relação às redes de gerações anteriores. Para alcançar esses objetivos, explora uma combinação de funções avançadas nas camadas MAC e física, como compartilhamento de recursos, relatórios de indicador de qualidade de canal, adaptação de links através de modulação e codificação adaptativa e pedido de retransmissão automatizado (CAPOZZI et al, 2013).

Em redes móveis, independente da tecnologia utilizada, o tráfego de vídeo apresenta um crescimento anual de cerca de 50%. Este tipo de tráfego representou mais da metade do consumo de dados, entre o terceiro trimestre de 2016 e o terceiro trimestre de 2017. Neste mesmo cenário, as redes sociais também devem crescer, mas sua participação relativa do tráfego diminuirá de 12% em 2017 para cerca de 8% em 2023, como resultado do forte

crescimento do consumo de vídeo. Outras categorias de aplicações têm taxas de crescimento menores, e, portanto, estão diminuindo em proporção em relação ao tráfego total. O uso de vídeo incorporado em redes sociais e páginas da web também está crescendo, alimentado por telas de dispositivos maiores, maior resolução e novas plataformas que suportam transmissão em tempo real (ERICSSON, 2017).

O surgimento de novas aplicações e as mudanças no comportamento do consumidor podem mudar os volumes de tráfego na rede. O *streaming* de vídeos em diferentes resoluções pode impactar o consumo de tráfego de dados em um alto grau. O vídeo de alta definição (1080p), geralmente aumenta o volume de tráfego de dados em torno de 4 vezes em relação ao mesmo vídeo com uma resolução padrão (480p). Uma tendência emergente com o aumento da transmissão de formatos de vídeo imersivos, como o vídeo de 360 graus, também afetaria o consumo de tráfego de dados. Por exemplo, um vídeo de 360 graus do YouTube consome 4 a 5 vezes mais largura de banda que um vídeo normal na mesma resolução (ERICSSON, 2017).

Um dos principais desafios para garantir qualidade de serviço em redes móveis de alta densidade e recursos limitados está a cargo da disciplina de escalonamento, que consiste em gerenciar as filas de transmissão e recepção de pacotes. Neste sentido, o presente trabalho propõe um escalonador para tráfegos de taxa variável, dentre os quais está contido o tráfego de vídeo, no sentido *downlink* para a arquitetura de rede LTE. A principal função deste escalonador é prover a garantia de um *delay* máximo para tráfegos com taxa variável tratados por uma estação base. Para isso o escalonador é composto por três etapas hierárquicas de processamento. A etapa 1, tem a função de condicionar o tráfego de entrada; a etapa 2 tem a função de calcular uma taxa de dados para garantir um *delay* máximo requisitado. Por fim, a etapa 3 faz a alocação do tráfego em blocos de recurso.

## 1.2 OBJETIVOS

### 1.2.1 Objetivo Geral

Desenvolver um escalonador para tráfegos de taxa variável em sentido *downlink* para redes LTE com garantia de *delay* máximo.

### 1.2.2 Objetivos Específicos

- Modificar o modelo proposto em (DOSCIATTI, 2015) para redes WiMAX, que utiliza alocação por divisão de tempo, e aplicá-lo em uma arquitetura de rede LTE, utilizando de alocação de recursos de acesso múltiplo no domínio de tempo e frequência;
- Otimizar a utilização do espectro visando maior quantidade de usuários alocados com requisitos de *delay* satisfeitos;
- Avaliar o desempenho do novo modelo proposto para verificar a eficiência frente a outros escalonadores presentes na literatura;
- Realizar estudos de simulação computacional explorando diferentes cenários em redes LTE;
- Modelar o tráfego de entrada utilizando um balde de fichas;
- Calcular o tempo total de transmissão que atenda um *delay* máximo solicitado para os tráfegos concorrentes;
- Alocar usuários de maneira justa buscando reduzir o *delay* entre todos os usuários do sistema.

## 1.3 ESTRUTURA DO TRABALHO

Esta dissertação está organizada em seis capítulos. Este capítulo apresenta o contexto geral que motivou o problema em estudo, assim como os objetivos geral e específicos. O Capítulo 2 apresenta uma revisão das gerações de redes móveis e os detalhes da arquitetura de rede LTE utilizada. O Capítulo 3 apresenta os trabalhos relacionados ao tema, descreve os escalonadores utilizados para fins de comparação e os mecanismos utilizados para a proposta do modelo de escalonamento. O Capítulo 4 descreve em detalhes o modelo de escalonamento proposto. O Capítulo 5 apresenta a validação do modelo e os resultados obtidos por meio de simulações. Finalmente, o Capítulo 6, apresenta as conclusões e os trabalhos futuros.

## 2 REDES MÓVEIS LTE

### 2.1 REDES MÓVEIS DE PRIMEIRA GERAÇÃO (1G)

Os sistemas móveis de primeira geração tornaram-se populares no início dos anos 80 e eram limitados à utilização de serviços de voz. As principais tecnologias eram AMPS (*Advanced Mobile Phone System*), NMT (*Nordic Mobile Telephone*) e TACS (*Total Access Communications System*) e utilizavam transmissão analógica, similares às utilizadas por um aparelho de rádio analógico comum. Os principais sistemas utilizavam bandas entre 12,5 kHz e 30 kHz e, em sua maioria, operavam em frequências entre 800 e 900 MHz. Todos os sistemas utilizavam a técnica de modulação em frequência (FM – *Frequency Modulation*) para transmissão de rádio e o tráfego era multiplexado para um sistema FDMA (*Frequency Division Multiple Access*) (SHARMA, 2013).

Apesar da boa qualidade de voz obtida, estes sistemas apresentavam baixa eficiência espectral. Além disso, os dispositivos móveis eram grandes e onerosos. Desta forma, surgiu a motivação para o desenvolvimento de uma nova rede móvel (COX, 2014).

### 2.2 REDES MÓVEIS DE SEGUNDA GERAÇÃO (2G)

No final dos anos 80 e no início dos anos 90, os sistemas móveis de segunda geração já eram suficientes para apoiar sistemas comerciais. Inicialmente, esses sistemas foram desenvolvidos apenas para serviços de voz. Entretanto, foram aprimorados e passaram a dar suporte a serviços de dados em comutação de circuitos CSD (*Circuit Switched Data*). Entre estes serviços, e um dos primeiros a ser concebido, está o serviço de mensagens curtas SMS (*Short Message Service*) (GHOSH et al, 2011).

Os sistemas de segunda geração usavam tecnologia de acesso múltiplo digital, tais como acesso múltiplo por divisão de tempo (TDMA - *Time Division Multiple Access*) ou acesso múltiplo por divisão de código (CDMA - *Code Division Multiple Access*). Desta forma, em comparação com os sistemas de primeira geração, os sistemas 2G não só ofereciam maior eficiência do espectro, como também melhores serviços de dados e *roaming* (SHARMA, 2013).

A comunicação 2G geralmente está associada ao GSM (*Global System for Mobile Communication*), o qual utilizava o método de transmissão TDMA e foi projetado por um

conjunto de empresas que trabalhavam sob a orientação da ETSI (*European Telecommunications Standards Institute*). Entretanto, sistemas como o IS-95 CDMA e IS-136 TDMA também estavam em desenvolvimento nos Estados Unidos (SHARMA, 2013).

Dispositivos de segunda geração menores e com baterias de longa duração, levaram esta tecnologia ao mundo todo. Ao mesmo tempo, a rede mundial de computadores apresentava um crescimento promissor. Assim, promoveu-se a junção destes dois conceitos e serviços de Internet passaram a ser oferecidos em dispositivos móveis. Este avanço foi possível com a implementação do GPRS (*General Packet Radio Service*), um sistema adicional de comutação de pacotes que podia prover taxas entre 56 e 115 kbps. Em seguida, esta taxa foi aprimorada com a introdução da tecnologia EDGE (*Enhanced Data Rates for GSM Evolution*) que possibilitou taxas de até 384 kbps para as redes GSM com sistema GPRS (BHALLA; BHALLA, 2010).

Com o avanço da tecnologia de segunda geração, além dos serviços de voz, era possível utilizar serviços de mensagem multimídia e de comunicação via Internet, como e-mail e acesso à páginas web (COX, 2014).

## 2.3 REDES MÓVEIS DE TERCEIRA GERAÇÃO (3G)

Assim como nas gerações anteriores, haviam diversos estudos para o desenvolvimento de um sistema de terceira geração. Dentre eles, o que teve maior destaque foi o UMTS (*Universal Mobile Telecommunication System*), o qual mantém o núcleo da rede GSM com uma interface de comunicação de rádio modificada. Este sistema foi desenvolvido pela 3GPP e utiliza tecnologia de acesso múltiplo por divisão de código WCDMA (*Wide-Band Code-Division Multiple Access*). Pouco depois, a organização paralela 3GPP2 foi formada para desenvolver a tecnologia concorrente cdma2000, uma evolução da tecnologia de segunda geração IS-95 (DAHLMAN; PARKVALL; SKÖLD, 2016).

A 3GPP adicionou tecnologias para melhoria da eficiência espectral, que juntas são conhecidas como HSPA (*High Speed Packet Access*). Assim, utilizando uma banda de 5MHz, é possível atingir taxas de transferência em *downlink* de até 14.4Mbps e taxas de *uplink* de até 5.8Mbps. Por fim, o HSPA+ (*Evolved HSPA*) trouxe o método *Multiple Input/Multiple Output* (MIMO) e modulação 16-QAM para *uplink* e 64-QAM para *downlink*, multiplicando a capacidade de transmissão (BHALLA; BHALLA, 2010).

A terceira geração de redes móveis, além de possuir maior flexibilidade de tecnologias de transmissão de rádio, tem como principal característica maiores taxas de transferência, possibilitando serviços de chamadas de vídeo e dados de banda larga (COX, 2014).

## 2.4 REDES MÓVEIS DE QUARTA GERAÇÃO (4G)

Com o sucesso da geração anterior, ocorreu a disseminação dos *smartphones* e *tablets*, com aplicações atrativas e de fácil utilização. As chamadas de voz deixaram de ser a principal fonte de tráfego, e a transmissão de dados cresceu exponencialmente nas redes móveis. Segundo (ERICSSON, 2017) o crescimento no ano de 2016, em relação ao ano anterior, foi de 70%, atingindo 7,2 EB<sup>1</sup> de tráfego mensal *Uplink/Downlink*. Desta forma, o principal desafio das redes de quarta geração é aumentar a capacidade de transmissão de dados.

Segundo (COX, 2014) existem três modos principais para aumentar a capacidade do sistema: a utilização de células menores, o que significa menos clientes por estação base e maior custo de operação da rede. Aumento da banda disponível, em um espectro de rádio finito e sob regulações diferentes por país. E, finalmente, o desenvolvimento de novas tecnologias visando maximizar a razão sinal-interferência-mais-ruído (SINR) e eficiência do espectro de rádio.

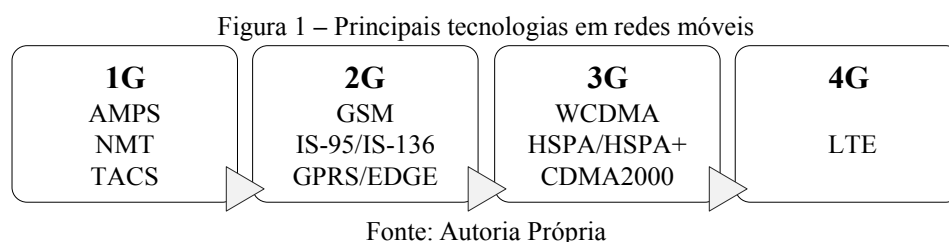
O aprimoramento das tecnologias de comunicação móvel é uma das principais motivações para o desenvolvimento da rede LTE. Projetada para ser um modelo totalmente IP, busca altas taxas de transferência, baixa latência e alta capacidade (DAHLMAN; PARKVALL; SKOLD, 2016). Para atender tais objetivos de desempenho, as redes LTE empregam a técnica de acesso OFDMA (*Orthogonal Frequency Division Multiple Access*) e uma solução de largura de banda variável, o que torna a rede mais flexível em termos de operação. Os intervalos de banda disponíveis variam entre 1,4MHz e 20MHz e permitem taxas de até 100 Mbps para tráfego *downlink* e 50 Mbps para tráfego *uplink* (CAPOZZI et al., 2013).

Por fim, um aspecto importante das redes LTE que deve ser destacado é referente a sua aceitação mundial. Diferente das gerações anteriores, onde tecnologias competiam entre si, a tecnologia LTE é única, podendo ser utilizada em modo TDD (*Time Division Duplex*) ou FDD (*Frequency Division Duplex*). Este fator promoveu sua disseminação e acelerou o desenvolvimento de novos serviços (DAHLMAN; PARKVALL; SKOLD, 2016).

---

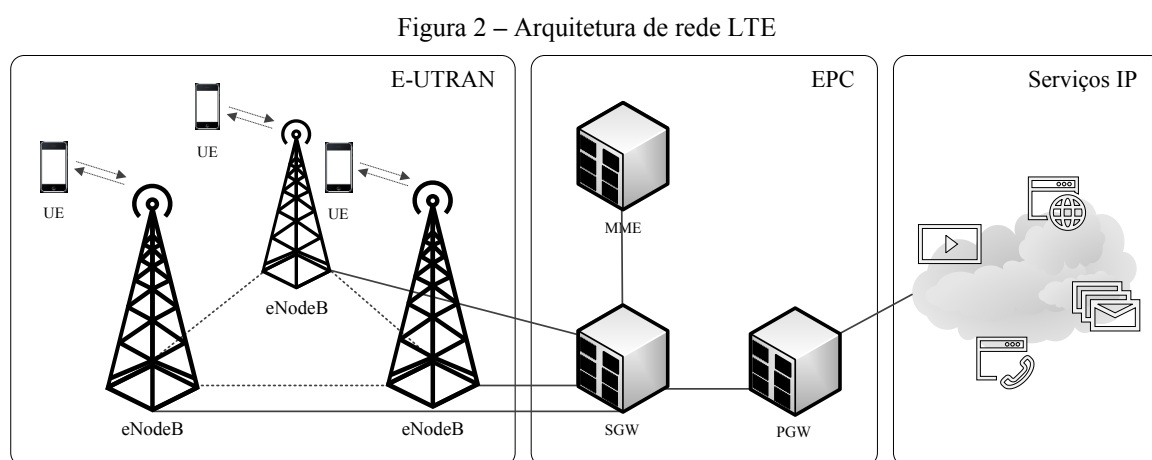
<sup>1</sup> 1 EB = 1 exabyte, equivalente a 1000<sup>6</sup> bytes.

As principais tecnologias empregadas em redes móveis destacadas acima são relacionadas na Figura 1.



## 2.5 ARQUITETURA DE REDE LTE

A arquitetura de rede LTE é formada por componentes de controle, estações base e dispositivos móveis de usuários, conforme mostrado na Figura 2. A rede de acesso é formada apenas pela estação base eNodeB (*Evolved NodeB*), que pertence a E-UTRAN (*Evolved UTMS Terrestrial Radio Access Network*) e fornece acesso por meio de uma rede sem fio ao equipamento do usuário (UE - *User Equipment*) que consomem serviços de dados.



Fonte: Autoria Própria

A eNodeB é responsável pela comunicação entre o UE e o EPC (*Evolved Packet Core*). O EPC é o núcleo da rede LTE e além de fornecer acesso aos serviços da rede externa, possui elementos responsáveis pelo controle de funções como autenticação, segurança, mobilidade, autorização de serviços, endereçamento IP e roteamento. Igualmente, gerencia e armazena informações de registro e perfil do usuário. Os componentes lógicos do EPC são descritos a seguir (COX, 2014):

- **Serving Gateway (SGW):** Suas principais funções incluem o roteamento e encaminhamento de pacotes, assim como o suporte necessário para que os UEs se movimentem em áreas atendidas por diferentes eNodeBs;
- **Packet Data Network Gateway (PGW):** Suas principais funções incluem o controle da troca de tráfego com a rede externa, seja ela a Internet ou um subsistema multimídia sobre IP, serviços de endereçamento IP e filtro de pacotes;
- **Mobility Management Entity (MME):** Suas principais funções incluem as funções de sinalização e controle para gerenciar o acesso do UE às conexões de rede, como o registro de recursos de rede e a função de gerenciamento de mobilidade.

A E-UTRAN é formada por uma série de estações base eNodeB distribuídas de forma não hierárquica e geralmente conectadas entre si. A eNodeB é responsável pelo gerenciamento dos recursos de rádio, o que inclui o controle de portadora, controle de admissão, mobilidade de conexão e alocação de recursos para os usuários. A alocação de recursos ocorre no sentido *uplink*, para recepção de dados transmitidos pelo UE, e no sentido *downlink* para transmissão de dados para o UE.

A comunicação entre os UEs e o eNodeB é tratada por uma pilha de protocolos na camada dois do modelo OSI, conforme mostra a Figura 3. Essa camada é composta de três subcamadas, sendo a PDCP (*Packet Data Convergence Protocol*) a de nível mais alto, seguida da subcamada RLC (*Radio Link Control*) e MAC (*Medium Access Control*) como subcamada de nível mais baixo.

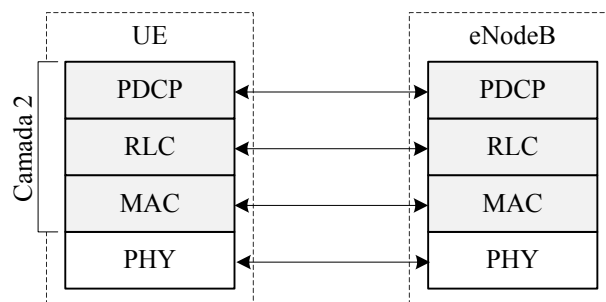
As principais funções de cada subcamada presentes na camada dois são descritas resumidamente a seguir (DAHLMAN; PARKVALL; SKOLD, 2016):

- **PDCP:** Suas principais funções incluem os procedimentos de compressão e descompressão dos dados de cabeçalho proporcionando menor *overhead*. Ademais, provê a mobilidade do UE em meio a diferentes eNodeBs, integridade e criptografia;
- **RLC:** Oferece funções de segmentação, remontagem e reordenação de PDUs (*Protocol Data Unit*). Além disso, fica a cargo da RLC gerenciar os processos de retransmissão para a camada física;
- **MAC:** Tem como principais funções a multiplexação e escalonamento de dados que serão trocados entre o UE e a estação base eNodeB na camada física.



Em síntese, pode-se dizer que a camada PDCP trata os dados trafegados, e as camadas RLC e MAC gerenciam os processos de transmissão e recepção dos dados entre o equipamento do usuário e a estação base.

Figura 3 – Pilha de protocolos LTE



Fonte: Autoria Própria

### 2.5.1 Camada Física de *Downlink*

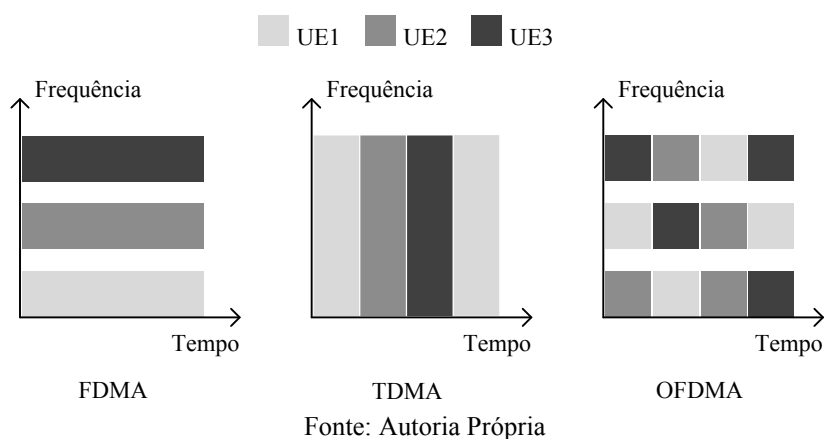
As transmissões no sentido de *downlink* escalonadas no eNodeB consistem na multiplexação de dados da camada MAC para a camada física. A multiplexação do sinal de *downlink* é facilitada pela estrutura de Acesso Múltiplo de Divisão de Frequência Ortogonal (OFDMA - *Orthogonal Frequency Division Multiple Access*), que permite que o fluxo de *downlink* seja subdividido em pequenas unidades de tempo e frequência.

### 2.5.2 OFDMA

O OFDMA é um método que atribui diferentes subportadoras a diferentes usuários de tal forma que mais de um usuário pode acessar a interface sem fio, ao mesmo tempo, para transmitir dados (YANG, 2010). Considerado uma versão multiusuário de um esquema de modulação digital OFDM, o sinal é desmembrado em subportadoras ortogonais independentes que são utilizadas para transportar os dados por meio de símbolos. Assim, a largura de banda disponível é dividida em larguras de banda menores.

Ao combinar tecnologias de acesso múltiplo por divisão de tempo (TDMA) e acesso múltiplo por divisão de frequência (FDMA), o OFDMA permite a alocação rápida de recursos de rádio e multiplexação multiusuário no domínio da frequência (PRASAD; SHUKLA; CHISAB, 2012). Conforme mostra a Figura 4, é possível verificar a diferença entre as tecnologias de multiplexação FDMA, TDMA e OFDMA, a partir de um exemplo de alocação de recursos para três estações clientes.

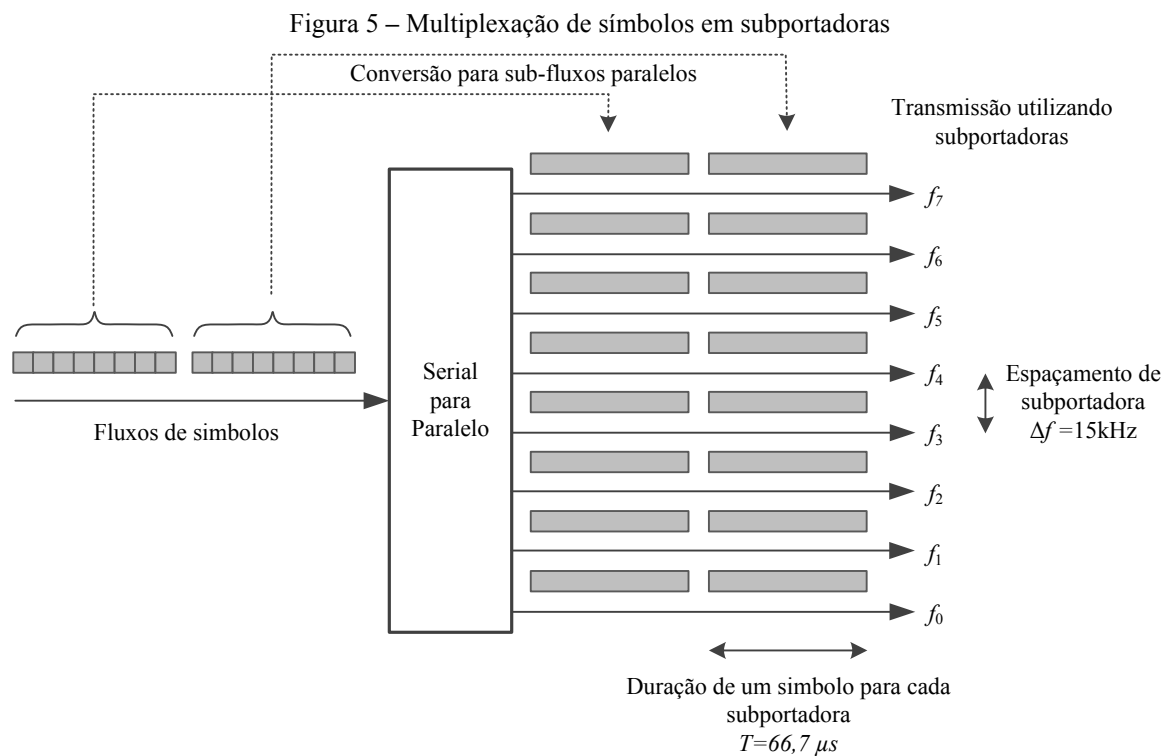
Figura 4 – Multiplexação de dados em interface aérea.



### 2.5.3 Subportadoras

Quando um transmissor OFDM recebe um bloco de símbolos, transmite cada símbolo em uma radiofrequência diferente, as quais são chamadas de subportadoras independentes. Desta forma, é possível que cada subportadora seja modulada de forma distinta, sendo que entre as técnicas utilizadas estão QPSK, 16QAM e 64QAM. A largura de banda de cada subportadora é pequena, portanto, pode suportar uma baixa taxa de símbolos. Sua divisão é espaçada de forma ortogonal, o que significa que quando a amplitude de uma subportadora é amostrada, a amplitude de outra é igual a zero. Esta característica é preservada por meio do emprego da transformada rápida de Fourier (FFT - *Fast Fourier Transform*) e sua inversa (IFFT - *Inverse Fast Fourier Transform*). A principal motivação desta forma de divisão de banda é suavizar os problemas de interferência intersimbólica (ISI - *Inter Symbol Interference*), associados às transmissões de banda larga disponíveis em canais seletivos em frequência (YANG, 2010).

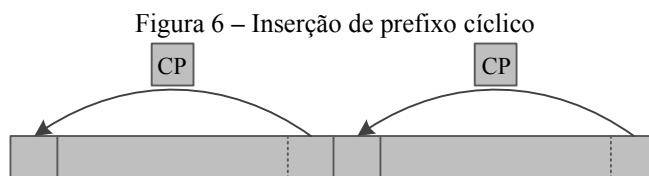
Em redes LTE, um símbolo tem a duração de  $66.7\mu\text{s}$  e, tendo em vista que o espaço para cada subportadora é dado por  $\Delta f = 1/T$ , este espaçamento corresponde a 15 kHz. Desta forma, espera-se uma taxa de símbolos de 15ksps. Na prática, esta taxa normalmente cai para 14ksps, uma vez que símbolos consecutivos são separados por um prefixo cíclico (CP - *Cyclic Prefix*) (COX, 2014). O processo de multiplexação de símbolos em subportadoras pode ser verificado na Figura 5.



Fonte: Adaptado (COX, 2014)

#### 2.5.4 Inserção de Prefixo Cíclico

Uma técnica adicional que permite livrar completamente o ISI consiste na ideia básica de inserir um período de guarda, no qual nada é transmitido antes de cada símbolo. O LTE usa uma técnica ligeiramente mais complexa conhecida como inserção do prefixo cíclico. Aqui, o transmissor começa inserindo um período de proteção antes de cada símbolo, como antes. No entanto, para preencher o período de guarda ele copia uma parte dos dados do final do símbolo seguinte (COX, 2014). Este processo é mostrado na Figura 6, correspondente a inserção dos prefixos cíclicos em dois símbolos OFDM.

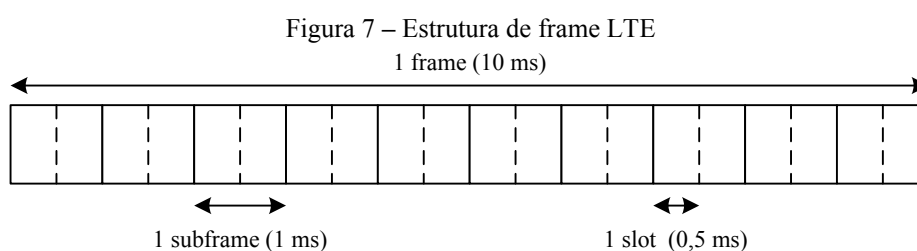


Fonte: Autoria Própria

A inserção do prefixo cíclico é benéfica no sentido de que torna o sinal OFDM insensível à dispersão, desde que a extensão da dispersão não exceda o comprimento do prefixo cíclico. Por outro lado, a inserção do prefixo cíclico implica em uma perda correspondente em termos de banda, uma vez que a taxa de um símbolo OFDM é reduzida sem que exista uma redução da largura de banda total do sistema.

## 2.6 ESTRUTURA DO QUADRO LTE

Os recursos de tempo e frequência são subdivididos por antenna de acordo com a seguinte estrutura: a maior unidade de tempo é de 10 ms para um quadro, que é subdividida em dez subquadros de 1 ms ou 1 intervalo de tempo de transmissão (TTI - *Transmission Time Interval*), cada um dos quais é dividido em dois slots de 0,5 ms. Essa estrutura é mostrada na Figura 7, onde cada slot compreende sete símbolos OFDM no caso do comprimento de prefixo cíclico normal, ou seis, se o prefixo cíclico estendido estiver configurado na célula.



Fonte: Adaptado (COX, 2014)

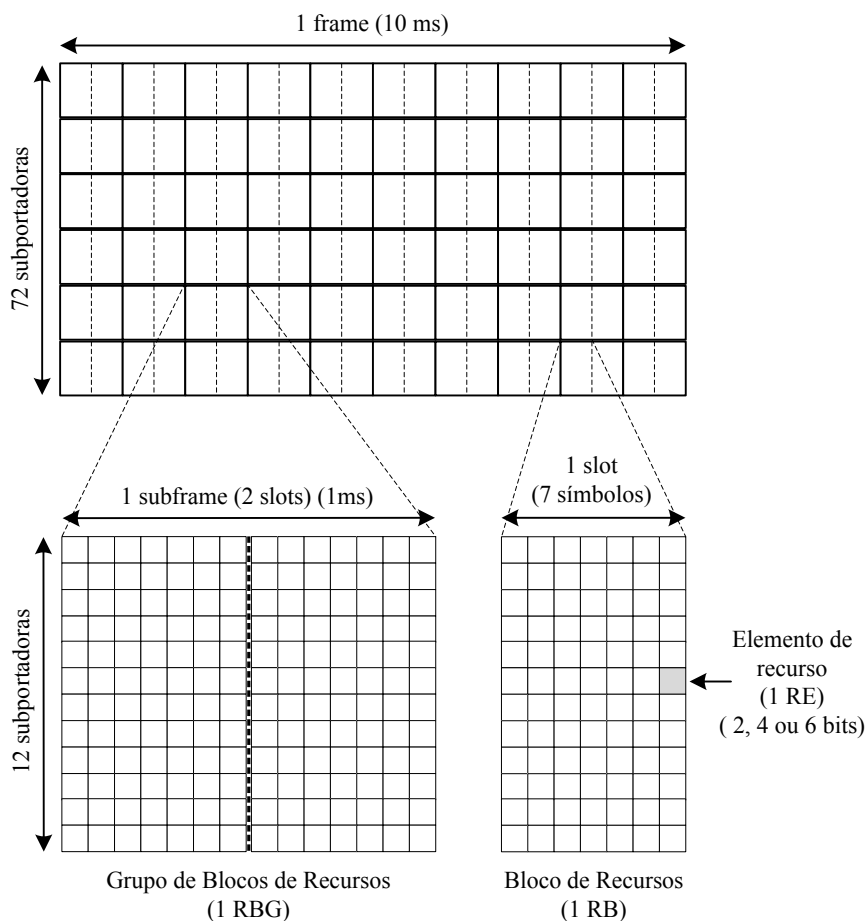
## 2.7 GRADE DE RECURSOS DOWNLINK

Conforme visto anteriormente, em redes LTE os recursos são organizados em função da frequência assim como em função do tempo, desta forma utiliza-se uma grade para representar a alocação dos recursos disponíveis. A Figura 8 mostra a grade de recursos para a utilização do prefixo cíclico normal, o qual é utilizado neste trabalho.

A menor unidade de recurso da grade é o elemento de recurso (RE – *Resource Element*), e consiste em uma subportadora por duração de um símbolo OFDM. Cada elemento de recurso normalmente transporta dois, quatro ou seis bits de canal físico, dependendo se o esquema de modulação adotado é QPSK, 16-QAM ou 64-QAM (COX, 2014).

Os elementos de recurso são organizados em grupos denominados bloco de recurso (RB – *Resource Block*). Um RB compreende 84 REs no caso do comprimento de prefixo cíclico normal e 72 REs no caso do prefixo cíclico estendido. No domínio do tempo, um RB tem duração de 0,5 ms, o que corresponde a um slot com 7 REs por subportadora. No domínio da frequência, um RB utiliza 12 subportadoras com espaçamento de 15 kHz, ocupando assim um total de 180 kHz. Esta unidade é utilizada pela eNodeB para a alocação de recursos aos UEs durante as transmissões de dados. Entretanto, a alocação é feita em grupo de blocos de recursos (RBG – *Resource Block Group*).

Figura 8 – Grade de recursos LTE no tempo e frequência utilizando prefixo cíclico normal



Fonte: Autoria Própria

## 2.8 LARGURA DE BANDA

Uma célula LTE pode ser configurada com uma das seis opções de largura de banda, as quais estão relacionadas na Tabela 1. Neste arranjo duas bandas de proteção são reservadas nas bordas superior e inferior da banda utilizada, o que minimiza a interferência com a próxima banda. As duas bandas de guarda geralmente são da mesma largura, mas o operador de rede pode ajustá-las, se necessário, deslocando a frequência central em unidades de 100 kHz.

Tabela 1 – Larguras de banda suportadas em redes LTE

Banda disponível	Blocos de recurso	Subportadoras	Banda utilizada	Banda de guarda
1.4MHz	6	72	1.08MHz	2×0.16MHz
3MHz	15	180	2.7MHz	2×0.15MHz
5MHz	25	300	4.5MHz	2×0.25MHz
10MHz	50	600	9MHz	2×0.5MHz
15MHz	75	900	13.5MHz	2×0.75MHz
20MHz	100	1200	18MHz	2×1MHz

Fonte: (COX,2014)

## 2.9 INDICADOR DE QUALIDADE DE SINAL

Para as transmissões de dados no sentido *downlink* em LTE, o eNodeB seleciona um esquema de modulação e taxa dependendo de uma previsão das condições do canal. Para este processo de seleção a eNodeB utiliza o indicador de qualidade do canal (CQI - *Channel Quality Indicator*) transmitido pelo UE no sentido *uplink*. O CQI é uma indicação da taxa de dados que pode ser suportada pelo canal, levando em consideração a razão sinal-interferência-mais-ruído (SINR – *signal to interference plus noise ratio*) e as características do receptor do UE (SESIA; TOUFIK; BAKER, 2011). Portanto, um terminal com um receptor mais avançado pode receber dados com uma taxa maior, pois informará um CQI maior do que um terminal com um receptor menos avançado, permanecendo idênticas todas as outras condições. A Tabela 2 mostra como o CQI é interpretado em relação ao esquema de modulação *downlink* e taxa de codificação.

Tabela 2 – Esquema de modulação em redes LTE

Índice CQI	Modulação	Taxa de codificação x 1024
0		
1	QPSK	78
2	QPSK	120
3	QPSK	193
4	QPSK	308
5	QPSK	449
6	QPSK	602
7	16-QAM	378
8	16-QAM	490
9	16-QAM	616
10	64-QAM	466
11	64-QAM	567
12	64-QAM	666
13	64-QAM	772
14	64-QAM	873
15	64-QAM	948

Fonte: (3GPP, 2013)

## 2.10 ALOCAÇÃO DE RECURSOS

A eNodeB possui várias maneiras de alocar os blocos de recursos para os UEs no sentido *uplink* e *downlink*. No *downlink*, ela pode usar dois formatos de alocação de recursos flexíveis que são conhecidos como tipo 0 e 1 e um formato compacto conhecido como tipo 2. Ao selecionar-se a alocação de recursos de *downlink* tipo 0, a estação base utiliza os blocos de recursos em grupos de blocos de recursos (RBGs). O tamanho do RGB, ou seja, o número de

blocos contidos no RBG, varia de acordo com a largura de banda do sistema. A relação entre o tamanho do RBG, e a largura de banda do sistema é mostrada na Tabela 3.

Tabela 3 – Relação entre a largura de banda e o tamanho do RBG

Largura de Banda	Tamanho do RBG
1,4 MHz	1
3 MHz	2
5 MHz	2
10 MHz	3
15 MHz	4
20 MHz	4

Fonte: (3GPP, 2013)

Com a alocação de recursos tipo 1, a eNodeB pode atribuir blocos de recursos individuais dentro de um grupo para diferentes UEs, mas tem menos flexibilidade sobre a atribuição dos próprios grupos. Por outro lado, no formato tipo 2 a eNodeB aloca um conjunto de RBs contíguos em múltiplos de 2 ou 4, dependendo da largura de banda. Entretanto, mesmo que a camada MAC aloque os RBs de forma adjacente, eles podem não estar alinhados quando são transmitidos na camada física.

### 2.10.1 Taxa Física de Transmissão

A partir dos dados apresentados acima é possível estimar a taxa de dados em 1 bloco de recurso. Considerando o prefixo cíclico normal, 7 símbolos OFDM por slot de 0,5 ms, modulação 64-QAM com 6 bits por símbolo e 12 subportadoras, são 504 bits por bloco de recurso ou 1008 kbps. Para uma banda de 20 MHz, onde tem-se 100 blocos de recursos disponíveis para alocação, chega-se a uma taxa física total de 100,8 Mbps.

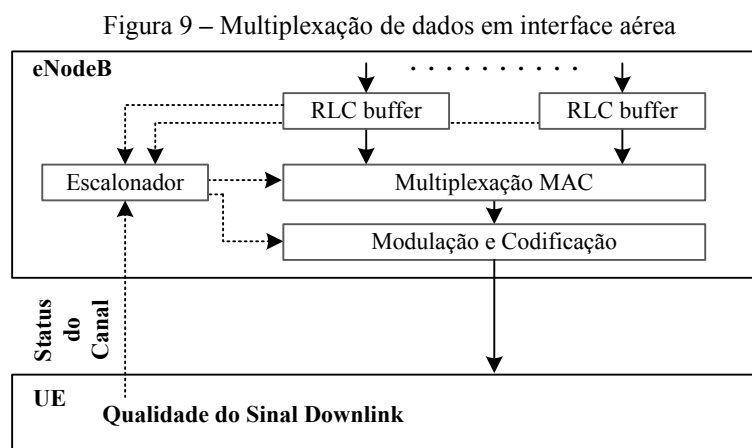
## 2.11 ESCALONAMENTO DE RECURSOS

O protocolo de controle de acesso ao meio (MAC) agenda as transmissões que são realizadas na interface sem fio e controla a camada física da eNodeB e no UE. Na eNodeB, a camada MAC possui uma entidade controladora que é responsável por uma série de funções, incluindo o procedimento de escalonamento de dados. O escalonador no eNodeB é responsável por distribuir os recursos de rádio disponíveis entre os UEs. A sinalização para suportar o escalonamento é padronizada, mas os detalhes do algoritmo são deixados para a implementação do eNodeB. Assim, fornecedores diferentes podem escolher estratégias diferentes em vários cenários para corresponder às necessidades do usuário.

A eNodeB aloca recursos de rádio no sentido *downlink* e *uplink* para cada UE com base, respectivamente, nos dados de *downlink* armazenados em *buffer* no eNodeB e em notificações de status de *buffer* recebidos do UE (SESIA, 2011). No sentido downlink, o modo usual de escalonamento é dinâmico e ocorre por meio de mensagens de atribuição para a alocação de recursos de *downlink*. As mensagens de controle são trocadas pelo canal físico de controle denominado PDCCH (*Physical Downlink Control Channel*) e os dados do usuário são multiplexados no canal de transporte compartilhado DL-SCH (*Downlink Shared Channel*). Para que as mensagens cheguem até os UEs pretendidos, um identificador de rádio temporário e único é utilizado, o qual é chamado de C-RNTI (*Cell Radio Network Temporary Identifier*).

Além do escalonamento dinâmico, o escalonamento semi persistente (SPS - *Semi-Persistent Scheduling*) pode ser usado. O SPS permite que os recursos de rádio sejam semi estaticamente configurados e alocados para um UE por um período de tempo mais longo. É útil para serviços como o VoIP, pois neste tipo de serviço o tempo e a quantidade de recursos de rádio necessários são previsíveis. Entretanto, o estudo proposto neste trabalho aborda o tráfego de vídeo com taxa variável, então, visando maior flexibilidade para a alocação de recursos com as características do tráfego pretendido, utiliza-se apenas o escalonamento dinâmico.

A Figura 9 apresenta o relacionamento do escalonador com os demais componentes envolvidos no processo de escalonamento. A partir dos dados armazenados em buffers na camada RLC, o escalonador fica responsável em selecionar quais dados serão transmitidos no próximo intervalo conforme sua disciplina de escalonamento. Para isso, verifica a qualidade do canal de transmissão para cada UE de destino com os quais irá se comunicar. Em seguida, os dados são enviados para o processo de multiplexação da camada MAC. Finalmente os dados passam por processos de modulação e codificação, até que sejam transmitidos pelo meio físico de acesso sem fio.





### 3 CONDICIONAMENTO DE TRÁFEGO

Este capítulo descreve os escalonadores *Round Robin* e *Channel and QoS Aware*, os quais são utilizados neste trabalho para fins de comparação de desempenho. O motivo da escolha de tais algoritmos se justifica pelo fato do primeiro se tratar de um algoritmo clássico, e do segundo possuir uma proposta que se assemelha ao trabalho proposto em termos de objetivos, em que se deseja um escalonador que respeite um *delay* máximo solicitado. Outro fator importante na escolha, é o fato de que ambos possuem implementação na ferramenta utilizada selecionada para as simulações.

O escalonamento é uma parte central de um sistema LTE. Para cada instante de tempo, o agendador determina a quais usuários o recurso compartilhado de tempo/frequência deve ser atribuído e determina a taxa de dados a ser usada para a transmissão. As transmissões de *uplink* e de *downlink* são escalonadas e conseqüentemente a eNodeB possui um escalonador para cada sentido de transmissão (DAHLMAN; PARKVALL; SKOLD, 2016).

O escalonamento multiusuário é responsável por distribuir recursos disponíveis entre usuários ativos para satisfazer suas necessidades de QoS. O algoritmo de escalonamento deve possuir baixa complexidade e escalonar as transmissões dos usuários atendendo restrições de qualidade de serviço (CAPOZZI et al, 2013). Considerando a comunicação de aplicações de voz e vídeo, tais restrições podem ser rigorosas, especialmente em relação ao *delay* máximo admissível para a entrega de pacotes (SESIA; TOUFIK; BAKER, 2011).

#### 3.1 TRABALHOS RELACIONADOS

O trabalho de Guan et al. (2011) propõe um algoritmo de escalonamento no sentido *downlink* com garantia de QoS. Inicialmente calcula-se a qualidade média de sinal (CQI) para cada usuário do sistema e com base nestes valores, são calculados os RBs necessários para atender seus requisitos de taxa de transmissão. Em seguida, os usuários são organizados de forma descendente por ordem de prioridade para a alocação de recursos.

Em (SU; WANG; LIU, 2012) é apresentada uma solução para provisão de QoS utilizando enxame de partículas (PSO) para tráfego *downlink* em redes LTE. Avaliando a qualidade de sinal para cada UE, a otimização busca garantir diferentes taxas requisitadas pelos usuários e resulta em uma matriz de alocação.

Em (Zhao et. al., 2012) utiliza-se o algoritmo de balde de fichas para provisão de QoS em tráfego *downlink* para redes LTE. A solução de escalonamento proposta visa atender requisições de tráfego de tempo real de vídeo e VoIP por meio de um processo dividido em três etapas. A primeira etapa classifica os fluxos de entrada para distinção dos serviços de tempo real. Na segunda etapa, um gerenciador de buffer armazena os dados e controla a taxa de dados. Na terceira etapa, o escalonador utiliza um balde de fichas para alocação de tráfegos de tempo real com diferentes taxas para tráfego de vídeo e VoIP, enquanto os demais tráfegos são tratados pelo algoritmo de escalonamento *Proportional Fair*.

No trabalho de Lai e Tang (2013), um algoritmo de escalonamento de pacotes busca atender restrições QoS de *delay* e largura de banda para aplicações em tempo real. O modelo proposto utiliza um mecanismo de previsão de pacotes (PPM – *Packet Prediction Mechanism*) que consiste em três fases. A primeira considera o domínio da frequência para utilização efetiva de largura de banda, a segunda, gerencia filas de transmissão e calcula atrasos esperados para os pacotes. Finalmente, na terceira fase que ocorre no domínio do tempo assim como a anterior, inicia um processo de corte para atender os requisitos de *delay*.

No trabalho de Ferdosian et al. (2015) busca-se atender as restrições de QoS padronizadas pelo grupo 3GPP para as classes QCI ao mesmo tempo em que se maximiza o desempenho do sistema em termos de justiça e *throughput* no sentido *downlink*. Para isso adota o algoritmo *Knapsack* (algoritmo da mochila) no domínio do tempo sobre os padrões de sobrecarga de tráfego.

Em (GHALUT; LARIJANI; SHAHRABI, 2016), é proposta uma otimização visando maximizar a qualidade da experiência do usuário (QoE - *Quality of Experience*) para escalonamento de tráfego do tipo vídeo *stream* no sentido *downlink*. A proposta aplica uma nova integração de Redes Neurais Aleatórias (RNN – *Random Neural Networks*) com uma heurística de Algoritmos Genéticos. O processo é iniciado com a identificação dos parâmetros de entrada da rede e da aplicação, como por exemplo taxa de transmissão e *delay*, assim como seus respectivos limites. Desta forma a RNN deve retornar uma função de adequação (*fitness*) para maximizar o QoE. Em seguida uma possível solução passa por um processo de evolução utilizando o AG e o processo continua, até que um critério de parada seja atingido ou um número máximo de gerações tenha sido executado.

Em (MATA, 2017), é proposto um processo de escalonamento com provisão de QoS para tráfego *uplink* utilizando a meta-heurística Algoritmos Genéticos. A solução de escalonamento é dividida em três passos. No primeiro passo os usuários são organizados em uma lista por prioridade e urgência de seus pacotes. No segundo passo, são alocados recursos

primeiramente aos usuários com pacotes próximos a um limite de atraso e em seguida outros usuários são selecionados para alocação. Finalmente, no terceiro passo, os recursos são alocados utilizando a solução de alocação encontrada a partir de um algoritmo genético.

A Tabela 4 sumariza os trabalhos descritos acima, destacando seus objetivos, pontos chave e parâmetros medidos. Verifica-se então que os trabalhos aqui relacionados não fazem a medição do *delay* máximo, o qual é utilizado como parâmetro de QoS para este trabalho.

Tabela 4 – Propostas de escalonamento em redes LTE

Proposta	Objetivo	Pontos Chave	Parâmetros Medidos
(GUAN et al., 2011)	Garantir QoS e maximizar <i>throughput</i> .	Atendimento de requisitos de taxa de transmissão com base na qualidade média de sinal (CQI) dos usuários.	<i>Throughput</i> e usuários satisfeitos.
(SU; WANG; LIU, 2012)	Garantir QoS e maximizar <i>throughput</i> .	Utilização de enxame de partículas (PSO) para atendimento de requisitos de taxa de transmissão a partir de uma matriz de alocação.	<i>Throughput</i> e usuários satisfeitos.
(ZHAO et al., 2012)	Atender requisitos de QoS para tráfegos de voz e vídeo.	Processo dividido em três etapas; Utilização de balde de fichas; Utilização do Algoritmo Proportional Fair demais tráfegos.	<i>Throughput</i> médio, <i>delay</i> médio e perda de pacote.
(LAI; TANG, 2013)	Atender requisitos de <i>delay</i> para aplicações de tempo real.	Proposta de um mecanismo de previsão de pacotes (PPM – <i>Packet Prediction Mechanism</i> ) dividido em três fases.	Taxa de pacotes inválidos, <i>goodput</i> e <i>delay</i> médio.
(FERDOSIAN et al., 2015)	Atender restrições de QoS e maximizar justiça e <i>throughput</i> .	Utiliza o <i>Knapsack</i> (algoritmo da mochila) e uma média do <i>throughput</i> já experimentado pelo usuário.	<i>Throughput</i> médio.
(GHALUT; LARIJANI; SHAHRABI, 2016)	Maximizar a percepção do usuário sobre a qualidade de serviço de vídeo.	Utiliza Redes Neurais Aleatórias (RNN – <i>Random Neural Networks</i> ) com uma heurística de Algoritmos Genéticos.	MOS ( <i>Mean Opinion Score</i> ), <i>throughput</i> e justiça.
(MATA; GUARDIEIRO, 2017)	Atender requisitos de QoS em sentido <i>uplink</i> .	Processo para escalonamento em sentido <i>uplink</i> dividido em três etapas e utilização de uma heurística de Algoritmos Genéticos.	<i>Throughput</i> , <i>delay</i> médio, índice de justiça, usuários satisfeitos, PSNR e tempo de execução.

Fonte: Autoria Própria

### 3.2 ESCALONAMENTO *ROUND ROBIN*

O escalonamento *Round Robin* (RR) é um dos primeiros algoritmos suportados para redes de comunicação de dados e tem como princípio servir, a seu turno, cada estação de forma igualitária. Neste sentido, cada estação terá a chance de transmitir seus dados por unidade tempo ou pacote de dados, determinados antecipadamente. Essa característica reduz a complexidade computacional envolvida no processo de alocação e torna o procedimento simples (NAGLE, 1987). No entanto, devido à natureza estocástica do canal sem fio, não é possível fornecer garantias de QoS exigentes (SIMON; LEUS, 2009).

Em redes LTE, este algoritmo atribui intervalos de tempo iguais a cada processo de transmissão e todo processo é tratado sem qualquer prioridade. Desta forma, promove justiça na distribuição dos blocos de recursos físicos entre os usuários sem distinção ao tipo de tráfego. Em termos práticos, no sentido *downlink*, a cada iteração a eNodeB aloca a mesma quantidade de blocos de recursos para transmitir os dados armazenados em um buffer RLC. Cada UE possui um buffer na eNodeB e caso esteja vazio o buffer do próximo UE é selecionado. Tal procedimento ocorre de forma cíclica até que todos os dados sejam encaminhados na forma de blocos de recursos (CAPOZZI et al, 2013).

O escalonamento das transmissões de forma circular faz com que o *delay* para o atendimento dos usuários permaneça constante se a quantidade de usuários a serem servidos não sofra alterações. Entretanto, a transmissão de um pacote de dados pode sofrer segmentação, sendo necessário mais de uma iteração para transmissão deste pacote. Assim, pacotes com tamanhos distintos podem apresentar *delays* proporcionais ao seu tamanho. Da mesma forma, canais com baixa qualidade de transmissão podem sofrer *delays* maiores em relação a canais com melhores condições. Isto ocorre pois a capacidade de transmissão de dados para um mesmo intervalo de tempo ou bloco de recurso, está diretamente ligada à razão sinal ruído do canal.

### 3.3 ESCALONAMENTO CHANNEL AND QoS AWARE (CQA)

O algoritmo de escalonamento *Channel and QoS Aware* (CQA) proposto em (BOJOVIC; BALDO, 2014) é específico para o sentido *downlink* em redes LTE. Sua disciplina prioriza o tráfego considerando o *delay* experimentado pelo UE, uma taxa de bits garantida e a qualidade do canal em diferentes sub-bandas. O CQA executa o escalonamento de acordo com critérios diferentes no domínio do tempo (TD) e no domínio da frequência (FD), visando

alcançar maior eficiência espectral ao mesmo tempo em que busca satisfazer os requisitos de *delay* do tráfego.

No domínio do tempo, a cada TTI, o escalonador seleciona dentre todos os usuários  $j=1, \dots, N$ , aqueles que não atingiram uma taxa máxima de bits. Em seguida, os usuários são agrupados em função do maior *delay* HOL (*Head of Line*) calculando uma métrica  $m_{td}$ , dada pela equação:

$$m_{td}^j(t) = \left[ \frac{d_{HOL}^j(t)}{g} \right], \quad (1)$$

onde  $d_{HOL}^j(t)$  é o *delay* HOL do tráfego  $j$ . O tamanho dos grupos é representado por  $g$  e representa a quantidade de fluxos a ser considerada para alocação. Assim, são selecionados os grupos de maior urgência considerando uma iteração no domínio da frequência. Os grupos de fluxos selecionados na iteração no domínio do tempo são encaminhados para o escalonamento FD, até que todos os RBGs sejam atribuídos no TTI correspondente, a partir dos fluxos com valores mais altos da métrica. No entanto, este parâmetro deve ser configurado de acordo com a capacidade da rede e a média esperada de usuários no sistema. No FD, para cada RBG  $k = 1, \dots, K$ , o CQA atribui o RBG  $k$  ao usuário  $j$  que possui o valor máximo da métrica FD definida da seguinte maneira:

$$m_{fd}^{(k,j)}(t) = d_{HOL}^j(t) \cdot m_{GBR}^j(t) \cdot m_{ca}^{k,j}(t), \quad (2)$$

onde a métrica  $d_{HOL}^j(t)$  em função da frequência, considera o *delay* HOL, que é o tempo que um pacote leva para ser efetivamente transmitido; a métrica  $m_{GBR}^j(t)$  é usada para garantia de taxa de transferência e a métrica  $m_{ca}^{k,j}(t)$  para adicionar conhecimento do canal.

### 3.4 LATENCY RATE (LR)

A teoria de servidores *Latency Rate* (LR), proposta em (STILIADIS; VARMA, 1998), permite o cálculo de limites para o *delay* máximo em redes de comunicação de dados. A nomenclatura *servidor* é utilizada para retratar a combinação de um escalonador e um transmissor existente em uma porta de saída de uma estação base ou roteador. Estes servidores podem suportar diferentes disciplinas de escalonamento e diferentes modelos de tráfego. No caso de um escalonador LR, seu comportamento é determinado por dois parâmetros, a latência

e a taxa alocada. Todos os servidores que garantem taxas aos seus clientes exibem essa propriedade e, portanto, podem ser modelados como servidores LR. A latência de um servidor LR pode ser considerada como o pior caso de *delay* do primeiro pacote no período ocupado por um fluxo.

O parâmetro de latência, em geral, depende do algoritmo de escalonamento utilizado, bem como da taxa alocada e dos parâmetros de tráfego da sessão que está sendo atendida. Para um algoritmo de escalonamento específico, parâmetros como a taxa de transmissão no link de saída, número de sessões que compartilham o link e as taxas atribuídas podem influenciar a latência.

Seja  $A_i(\tau, t)$  a representação para a chegada de dados da sessão  $i$  durante o intervalo de tempo  $(\tau, t)$  e  $W_i(\tau, t)$  a quantidade de serviço recebida pela sessão durante o mesmo intervalo. Assume-se então que  $A_i(\tau, t)$  aumenta apenas quando o último bit de um pacote é recebido pelo servidor. Da mesma forma,  $W_i(\tau, t)$  aumenta apenas quando o último bit do pacote no serviço sai do servidor. Assim, considera-se um *delay*  $Q_i(t)$  de uma sessão  $i$  no instante  $t$  dado pela equação:

$$Q_i(t) = A_i(0, t) - W_i(0, t), \quad \text{se } Q_i(t) > 0. \quad (3)$$

Desta forma, o *delay* máximo de ponta a ponta experimentado por um pacote, pode ser calculado a partir das latências de  $k$  escalonadores em série e dos parâmetros de tráfego da sessão que gerou o pacote. Mais especificamente se este tráfego for regulado por um balde de fichas, o *delay* de ponta a ponta  $D_i^{(p)}$  é limitado superiormente conforme a equação:

$$D_i^{(p)} \leq \frac{\sigma_i}{\rho_i} + \sum_{j=1}^k \theta_i^{S_j}, \quad (4)$$

onde  $\sigma_i$  e  $\rho_i$  são, respectivamente, parâmetros de tamanho máximo e taxa do balde de fichas, e  $\theta_i^{S_j}$  a latência do fluxo  $i$  no servidor  $S_j$  (STILIADIS; VARMA, 1998).

O *delay* máximo na entrega de um pacote de dados é medido entre o momento em que este pacote é recebido pela servidor LR, e sua transmissão. Este *delay*, considera o tempo de recebimento do primeiro bit de um pacote de dados até a transmissão de seu último bit. A Equação (4) pode ser aprimorada considerando-se a diferença entre a inclinação superior e inferior notada na Figura 10 (STILIADIS; VARMA, 1998). Assim, o *delay*  $D_i$  é limitado superiormente conforme a equação:

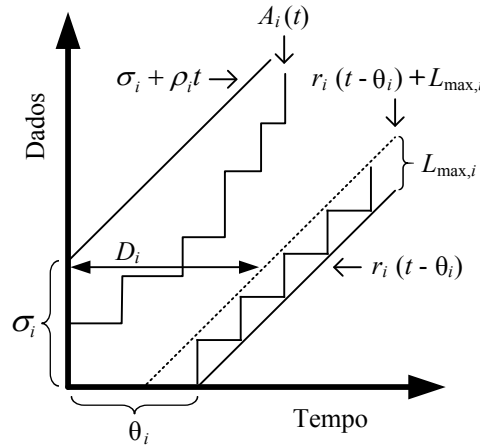
$$D_i \leq \frac{\sigma_i}{r_i} + \theta_i - \frac{L_{max,i}}{r_i}, \quad (5)$$

onde, a partir de uma taxa de dados  $r_i$ , para um pacote de tamanho máximo  $L_{max,i}$ , o servidor LR pode prover um *delay* máximo solicitado  $Dmax_i$  para cada tráfego atendendo a equação:

$$\frac{\sigma_i}{r_i} + \theta_i - \frac{L_{max,i}}{r_i} \leq Dmax_i, \quad (6)$$

Assim um escalonador de pacotes LR pode garantir um limite de *delay* se o tráfego de entrada for regulado por um balde de fichas conforme mostra a figura 10.

Figura 10 – Modelo do escalonador *Latency Rate*



Fonte: Figura Adaptada (STILIADIS; VARMA, 1998)

O cálculo utilizado para verificar o tempo necessário para transmitir os dados é obtido pela equação (7), considerando uma taxa de transmissão do meio físico dada por  $R$  para uma estação cliente  $i$ .

$$\theta_i = T_q + \frac{L_{max,i}}{R}, \quad (7)$$

onde  $T_q$  é o tempo para alocar todos os usuários. Conforme Higuchi, Foronda, et al., a restrição quanto ao *delay* máximo é dada pela equação (8).

$$\frac{T_q(\sigma_i - L_{max,i})}{r_i \times T_q - L_{max,i}} + T_q + \frac{L_{max,i}}{R} \leq Dmax_i, \quad (8)$$

Assim, isolando  $r_i$  obtém-se:

$$\frac{\sigma_i - Lmax_i}{Dmax_i - \frac{Lmax_i}{R} - T_T} + \frac{Lmax_i}{T_T} \leq r_i \quad (9)$$

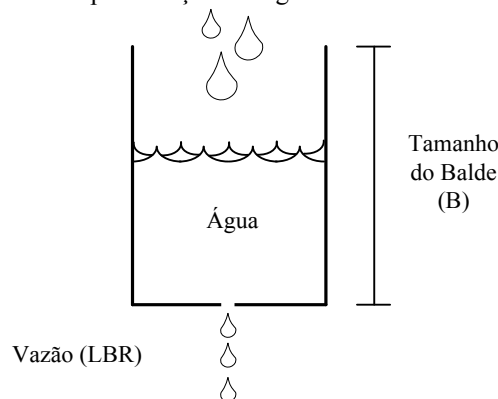
Além disso a equação (10) possibilita calcular a taxa alocada pelo servidor e o tempo para alocar todos os usuários. Verifica-se então que a taxa do balde de fichas somada a taxa para transmitir um pacote de tamanho  $Lmax_i$  deve ser menor que a taxa alocada pelo servidor.

$$\rho_i + \frac{Lmax_i}{T_T} \leq r_i. \quad (10)$$

### 3.5 CONDICIONAMENTO DE TRÁFEGO

O condicionamento de tráfego é uma técnica utilizada para regular a taxa média dos fluxos de dados de entrada em uma rede. Com o objetivo de garantir a qualidade de serviço para aplicações com diferentes necessidades, pode-se utilizar um mecanismo para especificar a largura de banda e limitar a alocação de recursos aos usuários. Neste sentido, em (TURNER, 2002) é proposto o método “balde furado” (*Leaky Bucket*). Para exemplificar seu funcionamento, considera-se um balde com um furo no fundo, representado na Figura 11.

Figura 11 – Representação do algoritmo do balde furado



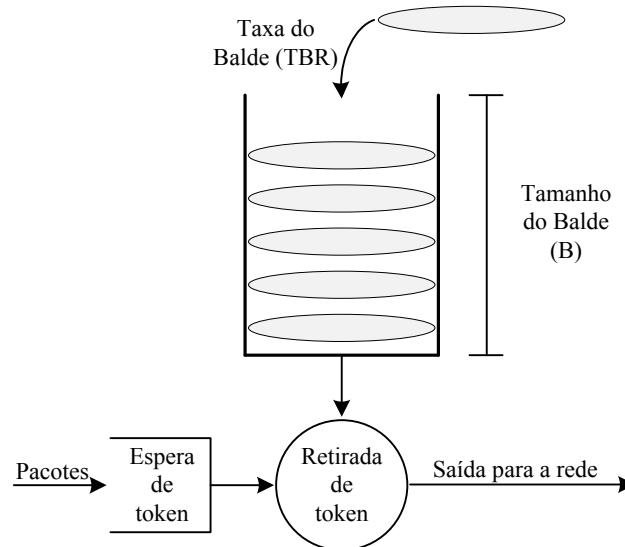
Fonte: Autoria Própria

Este balde pode receber água a qualquer velocidade, mas a saída da água será limitada por um fluxo constante LBR quando houver qualquer quantidade de água no balde e zero quando o balde estiver vazio. Caso a capacidade B do balde esteja completa, qualquer novo fluxo de água irá transbordar (TANENBAUM, 2011).



Uma abordagem equivalente é chamada de balde de fichas (*Token Bucket*), representada na Figura 12, onde considera-se um balde de tamanho  $B$  que recebe fichas (*tokens*) a uma taxa  $TBR$ . As fichas podem ser retiradas na medida em que estiverem disponíveis no balde. Caso o balde esteja cheio novas fichas geradas serão descartadas e não havendo fichas no balde é preciso aguardar até que novas fichas sejam depositadas.

Figura 12 – Tráfego de entrada modelado por um balde de fichas



Fonte: Autoria Própria

## 4 MODELO DE ESCALONAMENTO COM GARANTIA DE *DELAY* MÁXIMO

O modelo de escalonamento proposto nesta dissertação, intitulado *Latency-Rate Downlink Packet Scheduler* (LR-DPS), adota como base de estudo o trabalho de Dosciatti (2015), o qual sugere uma nova arquitetura para provisão de QoS em redes WiMAX fixas. Uma vez que se trata de uma tecnologia com outras características em termos de arquitetura e transmissão de dados, o modelo foi adaptado para redes móveis LTE. A arquitetura considerada em (DOSCIATTI, 2015) atende uma rede padrão IEEE802.16 com esquema de transmissão TDD e atribui tempos de quadro modelados por um balde de fichas em um canal de transmissão no sentido *uplink*, ou seja, o fluxo considerado ocorre da estação cliente em direção a estação base.

Neste trabalho, considera-se o escalonamento no sentido *downlink* para um canal de transmissão FDD em uma arquitetura de rede LTE. Para ilustrar o funcionamento do modelo utilizado por Dosciatti (2015), os tempos atribuídos para as estações clientes contidas na arquitetura WiMAX são representados na Figura 13. Nesta representação, a título de exemplo, para um intervalo de quadro de 6,5 ms, resultante do modelo matemático adotado na arquitetura considerada, as estações clientes são representadas como cliente1, cliente2 e cliente3. A alocação dos clientes ocorre apenas em função do tempo, onde os períodos atribuídos para cada estação são, respectivamente, de 2,8 ms, 1,5 ms e 2,2 ms.

Figura 13: Alocação por tempo de quadro em redes WiMAX



Fonte: Autoria Própria

Tendo em vista que a alocação de recursos usual do LTE ocorre de forma diferente, a abordagem da proposta contida neste trabalho utiliza a alocação em função da frequência e do tempo. Assim, o modelo que resulta em tempo de quadro para redes WiMAX, resulta em RBGs, para redes móveis LTE. A Figura 14 exemplifica a forma de alocação de recursos em uma rede LTE para um intervalo de 5 ms. Nota-se então, que a alocação dos fluxos de dados dos UEs são realizadas em intervalos fixos de 1 ms ou TTI, e que para o UE 1 foram alocados um total de 4 RBGs, enquanto que para o mesmo intervalo de tempo apenas 1 RBG foi alocado para o UE 4.

Figura 14: Alocação de recursos em redes LTE

		ue5	ue11	ue13	ue19	ue2	
1 RBG	}	ue4	ue10	ue13	ue18	ue1	Frequência
		ue3	ue9	ue13	ue17	ue1	
		ue2	ue8	ue12	ue16	ue50	
2 RBG	}	ue1	ue7	ue12	ue15	ue49	
		ue1	ue6	ue12	ue14	ue20	
Tempo (ms)		1	2	3	4	5	

Fonte: Autoria Própria

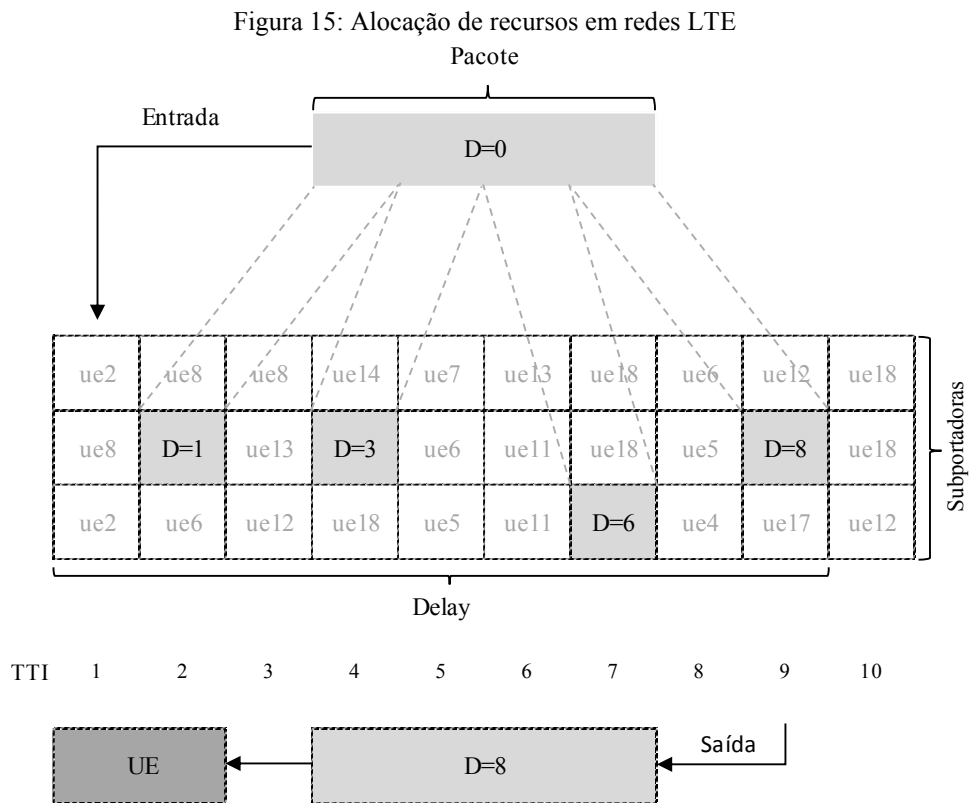
#### 4.1 DESCRIÇÃO DA PROPOSTA

A qualidade de serviço (QoS) em uma rede LTE é fornecida pelo método de escalonamento presente na eNodeB, o qual tem a função de selecionar os dados que serão transmitidos pelo meio físico a cada TTI em atendimento às requisições de seus enlaces. A abordagem de escalonamento, as características de tráfego e a banda disponível para transmissão, são fatores determinantes para garantir qualidade nas transmissões de dados.

O modelo de escalonamento proposto neste trabalho visa atender a qualidade de serviço de transmissões individuais. Para isso, utiliza como base o modelo *Latency-Rate* (STILIADIS; VARMA, 1998) em conjunto com um mecanismo de balde de fichas. Esta abordagem permite calcular limites restritos no atraso de ponta a ponta de sessões individuais. Assim, utiliza-se como parâmetro de medição para a qualidade de serviço o atendimento a um atraso máximo para cada enlace concorrente em uma eNodeB.

O atraso para a entrega de um pacote pode ocorrer por diversos fatores, sejam estes previsíveis ou não. Para este trabalho, considera-se o atraso decorrente do processo de escalonamento, que pode ser previsto e limitado mediante a aplicação do modelo proposto. O atraso produzido por um processo de escalonamento em uma rede LTE pode ser verificado por meio da Figura 15. Com o propósito de exemplificá-lo, considera-se um pacote segmentado na eNodeB para a transmissão pelo meio físico no sentido *downlink* com destino a uma UE. Neste exemplo, considera-se que ao ser recebido na eNodeB o pacote possui atraso igual a 0, e neste

caso é representado como  $D=0$ . Na eNodeB este pacote é segmentado em quatro RBGs para a transmissão. Não havendo oportunidade de alocação para os segmentos no primeiro intervalo de tempo, o primeiro deles é alocado no próximo intervalo e acumula um atraso  $D=1$ . O segundo segmento é alocado no quarto intervalo de tempo, o terceiro é alocado no sétimo intervalo e o quarto segmento é alocado após a passagem de oito intervalos de tempo. Acumula-se então um atraso  $D=8$ , ou seja, um atraso total de oito intervalos de tempo é gerado durante o escalonamento devido ao procedimento de alocação em tempo e frequência.



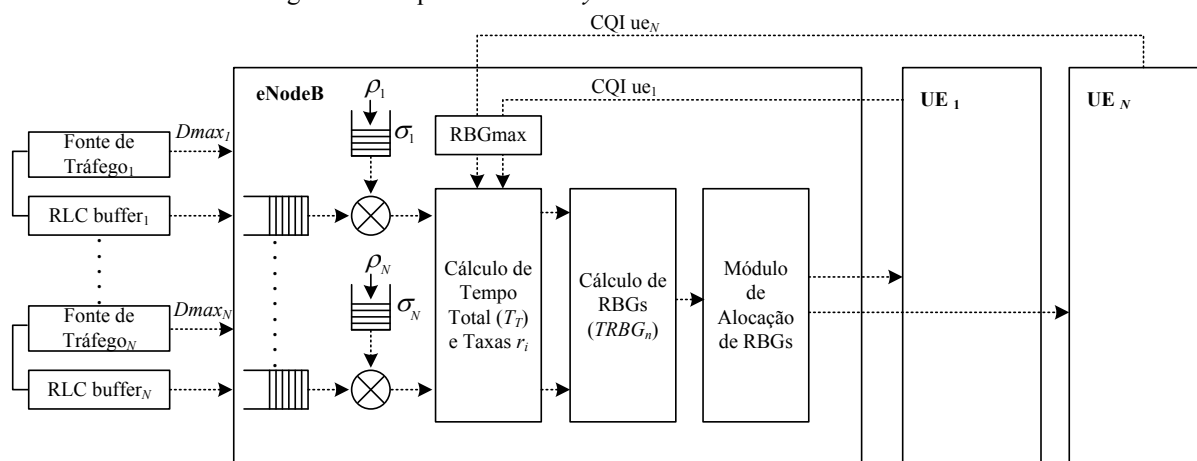
Fonte: Autoria Própria

Desta forma, percebe-se que a maneira adotada para alocação dos segmentos tem influência direta no atraso decorrente do processo de escalonamento para que o pacote seja totalmente enviado pelo meio físico. Neste sentido a proposta de trabalho aqui apresentada, visa propor um modelo que limite o atraso máximo acumulado para a entrega de um pacote de dados em redes LTE.

## 4.2 MODELO DO ESCALONADOR LATENCY-RATE DOWNLINK PACKET SCHEDULER

A arquitetura do escalonador *Latency-Rate Downlink Packet Scheduler* (LR-DPS) para redes LTE é apresentado na Figura 16. A partir das requisições dos UEs, fontes de tráfegos encaminham dados para a rede de acesso E-UTRAN e um atraso máximo deve ser atendido para cada tipo de tráfego. Neste caso, o atraso máximo requerido para um tráfego é identificado por  $D_{max,i}$  e deve ser conhecido antecipadamente pela eNodeB. Quando os dados são recebidos na eNodeB, são armazenados no buffer da camada RLC e submetidos a um algoritmo de balde de fichas. Em seguida, determina-se o tempo total ( $T_T$ ) para a transmissão de todos os tráfegos, com base na configuração de banda da eNodeB e da capacidade de recepção dos UEs, a qual é verificada a partir da informação do respectivo CQI. Assim, um cálculo é realizado para determinar a quantidade total de RBGs (TRBGs) a serem transmitidos no intervalo  $T_T$ . Finalmente, um módulo de alocação seleciona os tráfegos que serão encaminhados e os aloca em RBGs no espaço tempo/frequência para a transmissão via interface de rede sem fio. Maiores detalhes sobre a arquitetura LR-DPS são abordados a seguir.

Figura 16: Arquitetura *Latency-Rate Downlink Packet Scheduler*



Fonte: Autoria Própria

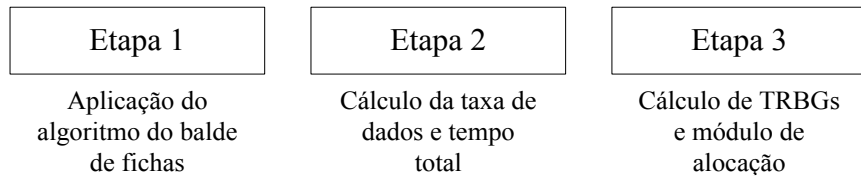
## 4.3 MODELO ANALÍTICO DO ESCALONADOR LATENCY-RATE DOWNLINK PACKET SCHEDULER

O funcionamento do escalonador, tendo como base o modelo utilizado por Dosciatti, é determinado por dois parâmetros, são eles latência ( $\theta_i$ ) e taxa de dados ( $r_i$ ). Neste modelo um

algoritmo de balde de fichas atua sobre o tráfego de entrada e o modelo LR provê uma taxa de dados para um atraso máximo requerido. O tráfego de entrada recebido pela eNodeB é requisitado por um UE. A requisição, neste caso específico, é atendida por uma fonte de tráfego disponível na rede externa, a qual faz a transmissão de dados por meio do núcleo da rede LTE (EPC).

Conforme ilustrada a Figura 17, a aplicação do modelo proposto pode ser segmentada em três etapas, onde a primeira etapa é a aplicação de um algoritmo de balde de fichas, a segunda se refere ao cálculo da taxa de dados ( $r_i$ ) e tempo total ( $T_T$ ) e a terceira ao módulo de alocação de RBGs. Dentre as etapas mencionadas, a segunda e a terceira referem-se especificamente à aplicação do modelo LR-DPS em redes LTE.

Figura 17: Aplicação do modelo em etapas



Fonte: Autoria Própria

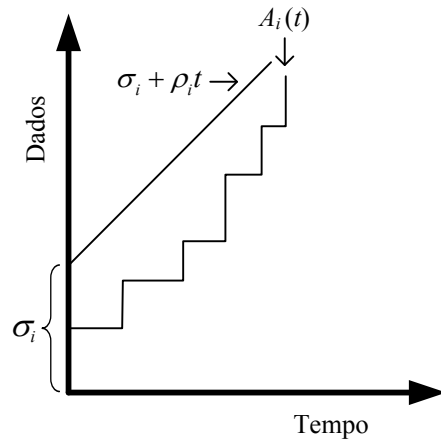
#### a) Etapa 1 – Aplicação do Balde de Fichas

Ao ser recebido pela eNodeB o tráfego de entrada  $A_i(t)$  é remodelado, sendo limitado pelo tamanho e pela taxa do balde de fichas. Assim, a primeira etapa se refere ao tratamento do tráfego de entrada pelo mecanismo do balde de fichas e está sujeito a:

$$A_i(t) \leq \sigma_i + \rho_i t, \quad (11)$$

onde  $\sigma_i$  corresponde ao tamanho do balde de fichas e  $\rho_i$  representa a taxa do balde de fichas no instante  $t$  para cada tráfego. A aplicação do balde de fichas é representada pela Figura 18, onde para cada tráfego de entrada  $A_i$  temos um conjunto  $\sigma_i + \rho_i t$  distinto, o qual representa um limite máximo para o tráfego de entrada. O tamanho do balde de fichas deve comportar os dados recebidos a partir da fonte de tráfego e a taxa deve ser suficiente para encaminhar os dados sem que haja descarte de pacotes.

Figura 18: Aplicação do balde de fichas



Fonte: Figura adaptada (STILIADIS; VARMA, 1998)

### b) Etapa 2 - Atribuição de Tempo de Total

Cada UE deve satisfazer à duas restrições quanto à taxa de transmissão, e, além disso, se as taxas de transmissão forem mínimas, um número maior de UEs poderão ser alocadas no espectro. Desta forma, neste trabalho, a Etapa 2 é formulada como um problema de otimização. O Problema de Atribuição de Tempo Total (ATT) tem por objetivo determinar o tempo total ( $T_T$ ) que minimiza a somatória das taxas de dados ( $r_i$ ) dos UEs. Seja  $N$  o número de UEs. O problema ATT é formulado da seguinte maneira:

$$r_{ATT} = \min_{T_T} \sum_{i \in N} r_i \quad (12)$$

Sujeito a:

$$\rho_i + \frac{RBGmax_i}{T_T} \leq r_i, \quad i \in N, \quad (13)$$

$$\frac{\sigma_i - RBGmax_i}{Dmax_i - \frac{RBGmax_i}{R} - T_T} + \frac{RBGmax_i}{T_T} \leq r_i, \quad i \in N, \quad (14)$$

$$0 < T_T < D_l. \quad (15)$$

A equação (12) é a função objetivo. As equações (13) e (14) correspondem às restrições quanto à taxa de dados das estações. A equação (15) é a restrição quanto ao tempo total para alocar todos os UEs com tráfegos ativos, onde  $D_l = \min_{i \in N}(Dmax_i)$ . Observa-se que  $r_{ATT} \leq R$ , onde  $R$  é a taxa física do sistema. Sendo a função objetivo linear e o conjunto de restrições convexo, o problema ATT pode ser classificado com um problema de otimização

convexa multi-variável restrito; a condição convexa garante a existência de uma única solução, ou seja, um ótimo global. Em geral trata-se de um problema de difícil solução analítica.

Em particular, se todas as UEs possuem as mesmas características, a solução do problema ATT é encontrada igualando as restrições (13) e (14). Desta forma o tempo total ótimo é dado pela equação:

$$T_T = Dmax - \frac{RBGmax}{R} - \frac{\sigma - RBGmax}{\rho} \quad (16)$$

e a taxa ótima de dados é produzida por:

$$r_{ATT} = \rho_i + \frac{RBGmax}{T_T}. \quad (17)$$

Nota-se que o número máximo de estações que podem ser alocadas corresponde a:  $\lfloor R/r_{ATT} \rfloor$ . Entretanto, numa situação geral (prática), é de fundamental importância que o problema ATT seja resolvido em tempo polinomial. Neste sentido, este trabalho sugere uma heurística simples para obter uma solução aproximada para o problema. Esta heurística, denominada *heurística da decomposição* (proposta em [Wille et al. 2009]), consiste em decompor o problema original em  $N$  problemas de fácil resolução e posterior composição final, conforme descrito a seguir.

Seja  $T_T^{(j)}$  a solução para o tempo total considerando que apenas a fonte  $j$  está ativa. Este valor é dado pela equação:

$$T_T^{(j)} = Dmax_j - \frac{RBGmax_j}{R} - \frac{\sigma_j - RBGmax_j}{\rho_j}, \quad j \in N. \quad (18)$$

Seja  $r_i^{(j)}$  a taxa de dados da fonte  $i$  obtida ao se considerar o tempo total  $T_T^{(j)}$ . Estes valores são dados pela equação:

$$r_i^{(j)} = \max \left\{ \rho_i + \frac{RBGmax_i}{T_T^{(j)}}, \frac{\sigma - RBGmax_i}{Dmax_i - \frac{RBGmax_i}{R} - T_T^{(j)}} + \frac{RGBmax_i}{T_T^{(j)}} \right\}, \quad i \in N, j \in N, \quad (19)$$

Desta forma, o tempo total ótimo ( $T_T^*$ ) será aquele que produz a menor taxa total de dados conforme a equação:

$$T_T^* = \left\{ T_T^{(j)} \mid \arg \min_{j \in N} \sum_{i \in N} r_i^{(j)} \right\}. \quad (20)$$



Pode-se provar que o esforço computacional da heurística da decomposição para a solução do problema ATT é  $\mathcal{O}(N)$ . Adicionalmente, ressalta-se que o problema ATT corresponde a uma forma equivalente de um problema Minimax (ver Apêndice).

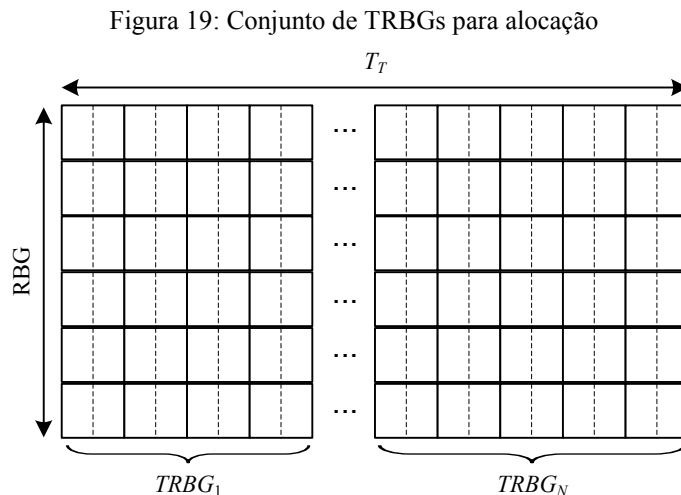
### c) Etapa 3 – Alocação de Blocos de Recursos

Na arquitetura de rede LTE os serviços utilizados no sentido *downlink* geram fluxos de dados que são encaminhados do EPC para a eNodeB. Ao receber estes dados, o escalonador é encarregado de alocá-los em forma de grupos de blocos de recursos (RBGs). A configuração dos RBGs atende a definição de Alocação de Recursos Tipo 0 (*Resource Allocation Type 0*), a qual determina a quantidade de blocos por grupo em função da banda de passagem disponível para transmissão dos dados. Em seguida, os RBGs são alocados para o UE pela camada física, respeitando uma divisão por tempo e frequência definida nesta etapa por um módulo de alocação de recursos.

Inicialmente, a taxa  $r_i$  e o intervalo  $T_T$  calculados na Etapa 2 são convertidos em TRBG (*Total Resource Block Group*). Este valor representa a quantidade total de RBGs alocados para um UE em um intervalo  $T_T$  e é definido por:

$$TRBG_i = \frac{r_i}{RBGmax_i} \times T_T, \quad (21)$$

onde  $RBGmax_i$  representa a capacidade de transmissão em bits por RBG para o UE $_i$ . Tem-se então um conjunto de TRBGs distribuídos no domínio do tempo e frequência para o intervalo  $T_T$  que garante a taxa  $r_i$  e, por conseguinte, o atraso solicitado  $Dmax_i$ . Este conjunto é mostrado na Figura 19.



Fonte: Autoria Própria

Uma vez encontrado o conjunto de TRBGs, o processo de alocação, descrito no Algoritmo 1, é realizado de forma sequencial e cíclica visando promover justiça em relação ao atraso máximo para a entrega dos pacotes. A execução deste algoritmo ocorre até que todo o tráfego seja transmitido.

A partir do conjunto de TRBGs, identifica-se entre os tráfegos solicitados o menor valor TRBG para alocação no tempo total  $T_T$ , ou seja, o que necessita menor quantidade de RBGs para transmissão. Este valor, identificado como TR, é utilizado como referência para a alocação de todos os tráfegos. Desta forma, para cada iteração do escalonador, todos os UEs tem seus dados alocados em RBGs e, para cada UE, é alocada uma quantidade de RBGs proporcional ao valor de TR. Assim, quando o valor  $TRBG_i$  é igual ao valor de TR, 1 RBG é alocado para o UE selecionado, tendo em vista que a proporção será de 1:1.

---

Algoritmo 1 – Módulo de Alocação

---

**Parâmetros do algoritmo:**

$S_{RBG}$ : Saldo RBG para alocação.

$A_{RBG}$ : Alocação de RBGs contíguos para a iteração.

$N_{UE}$ : Quantidade de UEs a serem alocados.

**Dados de Entrada:**

$CTRBG []$  = Conjunto de TRBGs.

**Algoritmo:**

1:  $TR$  = Menor TRBG do conjunto de TRBGs

2:  $S_{RBG,i}^{(0)} = 0$

3: **para cada** iteração  $i = 1$  até TR

4:     **para cada** UE <sub>$i$</sub>  até  $N_{UE}$

5:          $A_{RBG,i}^{(n)} = \text{int} \left[ \frac{TRBG_i}{TR} + S_{RBG,i}^{(n-1)} \right]$

6:          $S_{RBG,i}^{(n)} = \text{mod} \left[ \frac{TRBG_i}{TR} + S_{RBG,i}^{(n-1)} \right]$

7:         Aloca  $A_{RBG,n,i}$  para o UE selecionado

8:     **fim para cada**

9: **fim para cada**

**Retorno:**

Alocação de UEs no espaço tempo/frequência para  $T_T$ .

---

Fonte: Autoria Própria

A quantidade proporcional de RBGs alocada em cada iteração é identificada por  $A_{RBG}$  (RBG Alocado). Entretanto, deve-se considerar que a proporção pode resultar em valores não inteiros e para que a alocação seja feita de forma eficiente os RBGs devem ser ocupados por completo sempre que possível. Neste caso, os valores inteiros são alocados a cada iteração e, havendo tráfego na fila de transmissão, os valores decimais são atribuídos a um indicador de saldo  $S_{RBG,i}^{(n)}$ . Desse modo, a cada nova iteração, o saldo é incrementado com os valores não inteiros até que o seu valor seja igual ou maior que 1, isto é, que o saldo possa ocupar um RBG por completo. Não havendo mais tráfego na fila de transmissão, os dados do saldo são enviados na última iteração. Por fim, observa-se que o valor de  $TR$  corresponde também ao número de

iterações necessárias para alocação de todos os RBGs no intervalo  $T_T$ , no entanto, o Algoritmo 1 é executado sempre que houver dados para transmissão.

A título de exemplo, a Figura 8 mostra a distribuição de RBGs no espaço tempo/frequência a partir da aplicação do modelo proposto para três tráfegos distintos. Para simplificar o entendimento, o cenário de exemplo é composto por apenas três UEs com TRBGs diferentes e uma largura de banda que corresponde a 6 RBGs ou 1,4MHz. Considera-se que a Etapa 1 e Etapa 2 foram concluídas e são fornecidos os valores de  $r_i$  para cada UE assim como o tempo total de alocação  $T_T = 10$  ms. Neste exemplo, o processo de alocação foi dividido em dois passos. Para o primeiro passo, o conjunto de TRBGs foi calculado a partir da Equação (20) e, conforme mostra a Figura 20, correspondem a 12 RBGs para o UE1, 30 RBGs para o UE2 e 18 RBGs para o UE3. Assim temos os seguintes valores:  $T_T = 10$  ms,  $TRBG_1 = 12$ ,  $TRBG_2 = 30$  e  $TRBG_3 = 18$ . Identifica-se então que o valor de TR, que corresponde ao menor valor TRBG, é igual 12 e as proporções em relação a este valor são de 1:1 para o UE1, 1:2,5 para o UE2 e 1:1,5 para o UE3.

Figura 20: Exemplo de conjunto de TRBGs

6	ue1	ue1	ue2	ue2	ue2	ue2	ue2	ue3	ue3	ue3	
5	ue1	ue1	ue2	ue2	ue2	ue2	ue2	ue3	ue3	ue3	
4	ue1	ue1	ue2	ue2	ue2	ue2	ue2	ue3	ue3	ue3	
3	ue1	ue1	ue2	ue2	ue2	ue2	ue2	ue3	ue3	ue3	
2	ue1	ue1	ue2	ue2	ue2	ue2	ue2	ue3	ue3	ue3	
1	ue1	ue1	ue2	ue2	ue2	ue2	ue2	ue3	ue3	ue3	
		1	2	3	4	5	6	7	8	9	10
		Tempo (ms)									

Fonte: Autoria Própria

Para o segundo passo, considera-se a aplicação do módulo de alocação descrito no Algoritmo 1 e seu resultado é mostrado na Figura 21. Neste caso, na primeira iteração, são alocados 1 RBG para o UE1, 2 RBGs para o UE2 e 1 RBG para o UE3. Na segunda iteração, tendo em vista um saldo remanescente de 0,5 RBG para o UE2 assim como para o UE3, são alocados 1 RBG para o UE1, 3 RBGs para o UE2 e 2 RBGs para o UE3. Portanto, a cada nova iteração, a quantidade de RBGs é alternada garantindo ao final do processo que a quantidade de RBGs determinada no primeiro passo é alocada por completo. O processo de alocação ocorre em 12 iterações para um tempo  $T_T$  e se repetirá até que não haja mais dados para transmissão.

Figura 21: Alocação de RBGs por meio do Algoritmo 1

6	ue2	ue2	ue2	ue3	ue3	ue2	ue2	ue2	ue3	ue3
5	ue1	ue1	ue2	ue2	ue3	ue1	ue1	ue2	ue2	ue3
4	ue3	ue3	ue2	ue2	ue2	ue3	ue3	ue2	ue2	ue2
3	ue2	ue3	ue1	ue1	ue2	ue2	ue3	ue1	ue1	ue2
2	ue2	ue2	ue3	ue3	ue2	ue2	ue2	ue3	ue3	ue2
1	ue1	ue2	ue2	ue3	ue1	ue1	ue2	ue2	ue3	ue1
	1	2	3	4	5	6	7	8	9	10

Fonte: Autoria Própria.

Nota-se que a forma de alocação mostrada na Figura 20 pode prover o atraso solicitado uma vez que o conjunto de TRBGs são alocados no intervalo  $T_T$ , entretanto, percebe-se que no pior caso, o UE1 terá um atraso menor que o UE2, que por sua vez terá um atraso menor que o UE3. Pelo contrário, na Figura 21 verifica-se que no pior caso, os três UEs alocados apresentam atrasos iguais. Embora, quando aplicado o módulo de alocação, atrasos iguais não sejam uma regra, uma vez que há dependência da quantidade de UEs no sistema, pode-se garantir que haverá justiça entre os UEs alocados. Ou seja, quando aplicado o módulo de alocação, os atrasos observados serão próximos entre os UEs alocados.

## 5 AVALIAÇÃO DE DESEMPENHO

Existem três técnicas para avaliar o desempenho de sistemas e redes: análise matemática, medição e simulação computacional. Dentre elas, a simulação computacional ou simulação de eventos discretos, é a mais usual para redes de comunicação de dados. Um simulador computacional é uma abstração de um sistema de interesse e pode possuir diferentes níveis de detalhamento, tornando mais simples a validação de novas soluções e sua compreensão (GROSS; GÜNEŞ, 2010).

Neste trabalho o método de escalonamento *Latency-Rate Downlink Packet Scheduler* é avaliado por meio de um simulador computacional. Tendo em vista a existência de mais de um simulador para redes LTE, as características determinantes para a escolha da ferramenta foram as seguintes:

- **Simplicidade:** deve possibilitar a configuração de uma rede LTE de forma simples, assim como oferecer o nível de detalhamento necessário para o objeto de estudo;
- **Confiabilidade:** deve ser amplamente testado além de ser aceito na comunidade acadêmica para a validação de novas propostas;
- **Documentação:** deve ficar claro por meio da documentação disponível qual o nível de abstração e detalhamento do simulador, ou seja, para quais fins o simulador é útil;
- **Customização:** deve permitir o acesso ao código fonte para estudo e desenvolvimento.

Considerando os fatores mencionados acima, o *Network Simulator 3* (ns-3), versão 3.26, foi escolhido como instrumento para validar o modelo proposto. O ns-3 é um simulador de eventos discretos para sistemas de rede, voltado principalmente para uso educacional e pesquisa, uma vez que aborda várias tecnologias de comunicação de dados tais como ethernet, LTE, Wi-Fi, WiMAX, entre outras. É um software livre escrito em C++ e está disponível publicamente sob a licença GNU GPLv2 para uso, pesquisa e desenvolvimento (NS-3, 2017). Além disso, o ns-3 foi desenvolvido para substituir o *Network Simulator 2* (ns-2) e um dos objetivos fundamentais do projeto foi melhorar o realismo dos modelos, tornando as simulações mais próximas das implementações reais que elas representam (RILEY; HENDERSON, 2010).

Um dos módulos disponíveis na versão utilizada é o *LTE-EPC Network Simulator* (LENA), o qual foi desenvolvido por uma equipe de pesquisadores do *Centre Tecnològic de*

*Telecomunicacions de Catalunya* (CTTC) para a simulação de redes LTE. O módulo LENA é o principal recurso utilizado para a investigação do modelo proposto neste trabalho. Sua topologia de rede possui dois componentes principais que devem ser levados em consideração (BALDO, 2012):

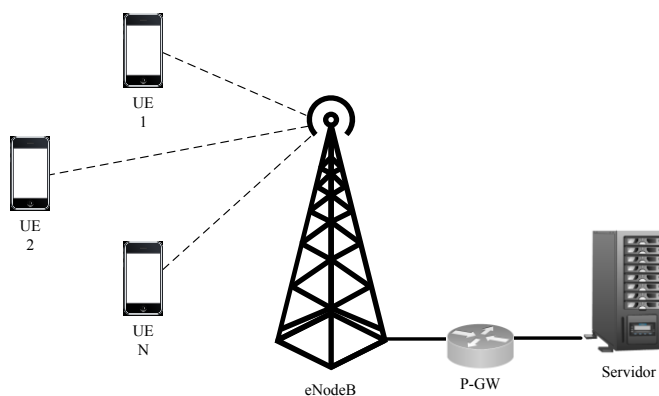
- **E-UTRAN:** inclui a pilha de protocolos de rádio (RRC, PDCP, RLC, MAC, PHY) que estão presentes nos UEs e eNodeBs. Este componente será utilizado para implementação do escalonador presente na camada MAC do eNodeB;
- **EPC:** compreende as interfaces do núcleo da rede, protocolos e entidades. Essas entidades e protocolos residem nos nós SGW, PGW e MME, e parcialmente dentro dos nós do eNodeB. Portanto, este componente será utilizado para prover acesso aos serviços de rede externa.

A seguir serão abordados os detalhes dos cenários de simulação criados para a avaliação do modelo proposto neste trabalho.

## 5.1 CONFIGURAÇÃO DO AMBIENTE DE SIMULAÇÃO

Nesta seção é apresentada a configuração do ambiente de simulação utilizado para validar e testar o desempenho do modelo proposto. Para isso, um cenário característico para uma rede LTE foi criado utilizando o módulo LENA, o qual tem sua topologia mostrada na Figura 22.

Figura 22 – Topologia de simulação para validar o modelo proposto



Fonte: Autoria Própria

A topologia presente na Figura 22 é composta por UEs conectados a uma eNodeB, a qual possui conexão com um servidor por meio do componente PGW. As conexões de interface

sem fio entre os UEs e a eNodeB possuem as mesmas características, permitindo que todos os UEs possam atingir as mesmas taxas de transferência. Considerando que todos os UEs solicitam o mesmo tráfego com a mesma restrição de delay, o processo de validação do modelo proposto é simplificado, uma vez que a quantidade de RBs alocados será igual para todos os UEs.

Para verificar o delay gerado no processo de escalonamento considera-se o intervalo de tempo que a eNodeB consome para entrega de um determinado pacote a um UE. Esta medição é feita pela análise dos dados do cabeçalho PDCP, considerando o pior caso, ou seja, considera-se o maior delay ocorrido em toda a transmissão para verificar se o delay requerido pelo UE foi atendido.

O tráfego utilizado para a simulação foi adquirido de um vídeo com taxa de bits variável (VBR), gerado com base no *EvalVid* (EVALVID, 2017), um *framework* para a avaliação da qualidade de vídeo transmitido através de uma rede de comunicação real ou simulada. Para utilizá-lo no ns3 foi necessária a configuração do módulo *evalvid-ns3*, o qual possibilita a transmissão de vídeo a partir de um modelo cliente/servidor e está disponível em (GERCOM, 2017). Por fim, os parâmetros utilizados para configuração do primeiro cenário de simulação são mostrados na Tabela 5. Alguns destes valores são determinados pelo padrão e outros são usuais em diversos trabalhos relacionados.

Tabela 5 – Parâmetros de Simulação

eNodeB	1
UE	30, 40, 50
eNodeB TX Power	46 dBm
Modelo de Atenuação	EPA 3km/h
Distância	500 m
Tipo de Tráfego	Vídeo (VBR)
Tempo de Simulação	30 s
Quantidade de Simulações	50
Largura de Banda	5 MHz
MCS	28
Blocos de Recursos	25
Tamanho do RBG	2 RBs
TTI	1 ms
RB Bandwidth	180 kHz
Modulação	64QAM
Modo de Transmissão	FDD

Fonte: Autoria Própria

Conforme pode ser verificado pela tabela de parâmetros utilizados para os testes iniciais, a largura de banda selecionada entre as disponíveis para o padrão LTE foi a de 5 MHz. Esta largura de banda é segmentada em 25 RBs de 180 kHz que são organizados em 12 grupos (RBG) de 2 blocos para cada TTI. Assim, considerando que um RB não é alocado, estão

disponíveis 24 RBs no sentido *downlink*. Neste caso a banda efetiva disponível é de 4,32 MHz, desconsiderando o bloco não alocado e 10% de utilização para banda de guarda.

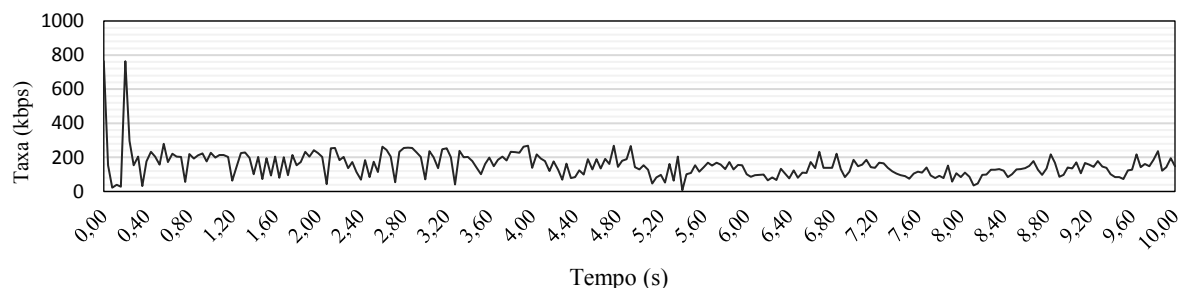
A escolha da largura de banda tem influência direta no tempo da simulação, uma vez que maior largura de banda significa maior quantidade de RBs e assim mais UEs podem ter suas restrições de delay satisfeitas. Portanto, larguras de banda de 10, 15 e 20 MHz não foram utilizadas nos testes iniciais exclusivamente por exigirem maior esforço computacional. Contudo, diferentes larguras de banda são utilizadas nos testes posteriores.

Para tornar o cenário mais realístico utiliza-se um modelo de atenuação de sinal que está disponível no ns-3 por meio de arquivos de rastreamento previamente calculados. Existem dois modelos disponíveis no simulador: *Extended Vehicular A* (EVA) e *Extended Pedestrian A* (EPA). O modelo de atenuação EVA disponível simula a variação de sinal para um tráfego veicular onde os usuários se movimentam a uma velocidade de 60 km/h, enquanto o modelo EPA simula a movimentação de pedestres a uma velocidade de 3 km/h. Nesta proposta o modelo EPA é utilizado em todos os cenários de simulação.

## 5.2 CENÁRIO 1 – VALIDAÇÃO DO MODELO

A primeira avaliação do modelo tem como objetivo verificar se o delay máximo requerido pelo usuário é atendido de acordo com os dados obtidos pelo modelo analítico. Para isso utiliza-se um tráfego de taxa variável (VBR) gerado a partir de um arquivo de vídeo de 30 segundos. A taxa em função do tempo é mostrada a partir da Figura 23 para o trecho de 10 segundos iniciais do tráfego utilizado nas simulações.

Figura 23 – Tráfego variável utilizado



Fonte: Autoria Própria

O tráfego de entrada deve ser conhecido com antecedência para que os parâmetros de tamanho e taxa do balde de fichas sejam definidos para cada UE. O tamanho do balde possui relação direta com o tráfego, assim o valor selecionado é igual a maior quantidade de dados



gerada em um intervalo de tempo  $e$ , neste caso, se refere aos primeiros milissegundos do tráfego representado na Figura 23. Uma vez definido o tamanho do balde busca-se então a taxa ideal do balde para o tráfego, a qual deve atender as restrições do meio além de ser suficiente a ponto de garantir que não haja descarte de dados. Esta configuração, de tamanho e taxa do balde, é realizada para cada tráfego tratado e resultará em valores diferentes para tráfegos distintos.

Em seguida são calculados os valores de  $r_i$ ,  $T_T$  e TRBG para atender um *delay* máximo ( $D_{max}$ ) de 150 ms, considerando a taxa do meio físico de valor  $R=17.760$  kbps e RBGmax de 1480 bits, onde os valores de  $R$  e RBGmax são determinados pela escolha da largura de banda e MCS. O valor  $D_{max}$  foi selecionado para atender as recomendações presentes na tabela QCI do padrão LTE. Os valores de entrada e os valores calculados nas etapas 2 e 3 do LR-DPS, a partir das restrições (16) e (17) (Etapa 2) e da equação (21) (Etapa 3) são mostrados na Tabela 6.

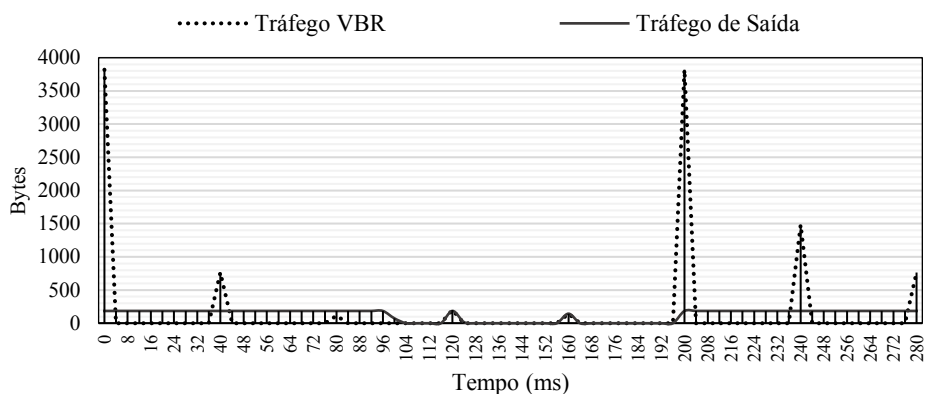
Tabela 6 – Parâmetros do escalonador LR-DPS

Parâmetros LR-DPS	
Valores de Entrada	
Tamanho do Balde ( $\sigma$ ) (bits)	30560
Taxa do Balde ( $\rho$ ) (kbps)	306
$D_{max}$ (ms)	150
RBGmax (bits)	1480
R (kbits)	17760
Valores de Calculados	
$T_T$ (ms)	55
$r_i$ (kbps)	333,280
TRBG	12,39
Quantidade de UEs	50

Fonte: Autoria Própria

A partir da análise da transmissão de dados realizada pela eNodeB é possível verificar a atuação do modelo no padrão de tráfego enviado. Assim, ao analisarmos graficamente um trecho de alguns milissegundos, conforme mostra a Figura 24, nota-se que o tráfego passa a se comportar de forma constante e não mais em forma de rajadas. No instante 0 a eNodeB recebe 3820 bytes para serem encaminhados a um determinado UE que solicitou o tráfego ao servidor externo. No escalonador LR-DPS, a eNodeB encaminha os dados recebidos a partir do módulo de alocação respeitando os valores calculados que resultam em 185 bytes ou 1 RBG a cada 4 ms. Assim, percebe-se que o tráfego de entrada é amortizado até o instante de 96 ms. No instante de 120 ms o pacote recebido pela UE é encaminhado sem que haja a necessidade de amortização da taxa de entrada e o mesmo ocorre para 160 ms. Entretanto no instante de 200 ms nota-se que o tráfego é modificado novamente pela atuação do escalonador.

Figura 24 – Tráfego VBR tratado pelo LR-DPS



Fonte: Autoria Própria

Para promover um melhor entendimento da Figura 24, o trecho do tráfego tratado é mostrado na Tabela 7 onde pode-se conferir o identificador do pacote recebido, o tempo de chegada na eNodeB e seu tamanho. Verifica-se que no instante 0 foram recebidos três pacotes pela eNodeB, dos quais os pacotes identificados como 1 e 2 possuem 1490 bytes e o pacote de identificador 3 possui 840 bytes, totalizando 3820 bytes de informação. Novos pacotes são recebidos em intervalos constantes de 40 ms. Tem-se então que o próximo pacote recebido possui 766 bytes, identificador de valor 4 e é recebido pela eNodeB no instante de 40 ms.

Tabela 7 – Tráfego VBR de entrada

Identificador	Tempo (ms)	Tamanho do Pacote (bytes)
1	0	1490
2	0	1490
3	0	840
4	40	766
5	80	112
6	120	201
7	160	144
8	200	1490
9	200	1490
10	200	840
11	240	1484
12	280	766

Fonte: Autoria Própria

Tendo em vista que o identificador é um valor sequencial e único para cada pacote enviado a partir do servidor, pode-se verificar a atuação do balde de fichas por meio da Tabela 8, a qual relaciona os três primeiros pacotes recebidos pela eNodeB. Após o recebimento do tráfego na eNodeB este deve ser encaminhado ao UE de destino e para isso deve ser tratado pelo LR-DPS. Assim, nota-se que o tratamento do primeiro pacote ocorre no intervalo de tempo entre 0 e 30 milissegundos, onde os dados são alocados em segmentos de 185 bytes a cada 4 ms. No instante de 28 ms o escalonador completa a alocação de 1480 bytes referentes ao

primeiro pacote, restando para a próxima alocação 10 bytes. Estes serão alocados no mesmo segmento utilizado para alocação dos dados iniciais do segundo pacote e assim tem-se alocados no instante 32 os 10 bytes referentes ao pacote 1 e 175 bytes referentes ao pacote 2. Por fim, nota-se por meio da Tabela 8 que a alocação dos três primeiros pacotes recebidos pela eNodeB ocorre nos primeiros 80 ms.

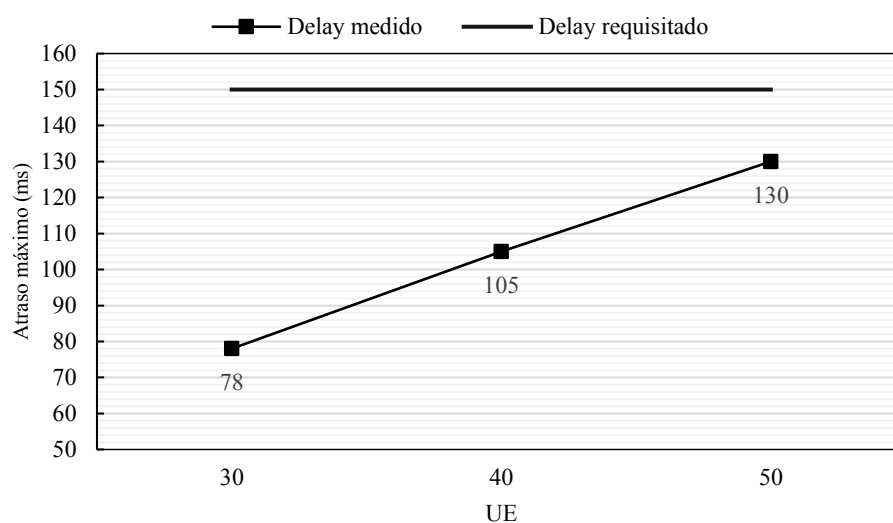
Tabela 8 – Tráfego de saída

Identificador do Pacote	Tempo (ms)	Saída LR-DPS (bytes)
1	0	185
1	4	185
1	8	185
1	12	185
1	16	185
1	20	185
1	24	185
1	28	185
1	32	10
2	32	175
2	36	185
2	40	185
2	44	185
2	48	185
2	52	185
2	56	185
2	60	185
2	64	20
3	64	165
3	68	185
3	72	185
3	76	185
3	80	120

Fonte: Autoria Própria.

Após a análise da alocação, verifica-se o comportamento do escalonador LR-DPS com relação ao *delay* suportado para a entrega dos pacotes de dados aos UEs. A fim de verificar o atendimento das restrições de *delay* solicitadas pelos UEs, foram realizadas simulações para diferentes densidades de usuários conectados a uma única eNodeB. Utilizando uma largura de banda de 5MHz para o tráfego *downlink*, verifica-se que a eNodeB entrega o tráfego VBR solicitado atendendo a restrição de *delay* máximo. Conforme mostrado no gráfico da Figura 25, as densidades de usuários utilizadas foram de 30, 40 e 50 UEs. Desta forma, percebe-se que para o conjunto de simulações iniciais utilizando um tráfego VBR de 30 segundos, os pacotes com maior *delay* na entrega acumulam períodos de 78 ms para 30 UEs, 105 ms para 40 UEs e 130 ms para 50 UEs. Assim, nota-se que o *delay* máximo de 150 ms foi atendido considerando o meio compartilhado por até 50 UEs.

Figura 25 – Delay máximo para tráfego de vídeo utilizando LR-DPS



Fonte: Autoria Própria

Os valores para delay máximo são medidos em milissegundos e foram obtidos por meio de 50 simulações para cada cenário. Cada simulação foi executada com sementes distintas a fim de aumentar o grau de confiabilidade dos resultados. Para cada uma das simulações executadas foi considerado o *delay* de pico para uma UE, ou seja, o valor considerado foi o mais alto entre os *delays* de entrega para uma UE.

Tratando-se de um cenário simplificado onde o tráfego solicitado é o mesmo para todos os UEs, as distâncias são equivalentes e o modelo de atenuação utilizado é único, espera-se que não haja grande variação entre os valores de *delay* máximo medidos para os UEs. Desta forma os valores obtidos por meio das simulações são mostrados em síntese na Tabela 9.

Tabela 9 – Delay máximo medido

UEs	30	40	50
Desvio Padrão	0,570	1,649	2,413
Intervalo de Confiança	99%	99%	99%
Limite Inferior (ms)	76,655	101,884	125,726
Média (ms)	76,923	102,556	126,605
Limite Superior (ms)	77,191	103,227	127,484

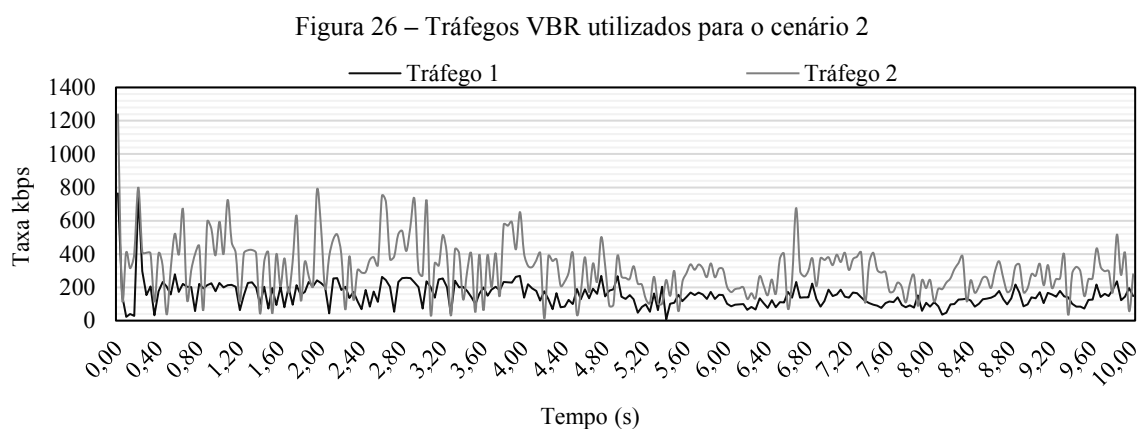
Fonte: Autoria Própria

Nota-se então que mesmo para um intervalo de confiança elevado a variação é pequena para os *delays* de pico. Observa-se ainda que na medida em que a quantidade de UEs utilizados na simulação aumenta, os limites inferiores e superiores se distanciam. Finalmente, com base nos resultados apresentados acima, pode-se concluir que o escalonador proposto atende as restrições de *delay* máximo de 150 ms para até 50 UEs utilizando um tráfego variável aplicando o escalonador LR-DPS configurado com os dados provenientes do modelo analítico.

### 5.3 CENÁRIO 2 – ANÁLISE DE DESEMPENHO

A partir do primeiro cenário selecionado é possível verificar que o escalonador LR-DPS foi eficaz e garantiu o *delay* máximo requerido para um tráfego VBR. Busca-se então variar os parâmetros de entrada para verificar o desempenho do algoritmo de escalonamento proposto em relação a outros escalonadores. Para isso, optou-se em selecionar os escalonadores disponíveis no simulador ns-3 garantindo assim que todos compartilhem da mesma topologia e das características proporcionadas pelo ambiente simulado. Assim, tendo em vista a proposta do trabalho, foram selecionados os escalonadores *Round Robin* (RR), devido a sua simplicidade e ampla utilização e *Channel And Qos-Aware* (CQA), por possuir a particularidade que busca minimizar o *delay* na entrega de pacotes para tráfegos de voz e vídeo.

Para o segundo cenário são utilizados dois tráfegos VBR distintos, sendo que um destes foi utilizado para o primeiro cenário. Ambos possuem tempo total de 30 segundos e suas taxas em função do tempo são mostradas na Figura 26, onde verifica-se que o segundo tráfego possui taxas maiores em relação ao primeiro.



A partir da análise dos tráfegos 1 e 2, os parâmetros do balde de fichas são definidos para um *delay* máximo de 150 ms onde uma UE solicita o tráfego 2 e as demais solicitam o tráfego 1. Considerando taxas distintas para os dois tráfegos, encontra-se dois conjuntos de valores para o LR-DPS, os quais estão relacionados na Tabela 10. Para isso, utiliza-se a função *fmincon()* do MATLAB® para resolução do problema ATT (Etapa 2), com retorno para os valores exatos de  $T_T$  e  $r_i$ . A título de exemplo, o Apêndice B mostra os resultados obtidos, a partir da heurística da decomposição e da função do MATLAB®, para um número variável de UEs com tráfegos diferentes.

Tabela 10 – Parâmetros do balde de fichas

Parâmetros de Entrada		
Tráfego VBR	Tráfego 1	Tráfego 2
Tamanho do Balde ( $\sigma$ ) (bits)	30560	49472
Taxa do Balde ( $\rho$ ) (kbps)	306	680
$D_{max}$ (ms)	150	150
RBGmax(bits)	1480	1480
Taxa do meio (kbps)	17760	17760
Valores Calculados		
$T_T$ (ms)	55	55
$r_i$ (kbps)	333,280	706,910
TRBG	12,39	26,27

Fonte: Autoria Própria

Uma vez definidos os parâmetros do balde de fichas, estes são configurados para cada UE de acordo com o tráfego solicitado. Assim, para as simulações desta seção, uma única UE recebe o tráfego 2 e utiliza um balde de fichas com capacidade para 49472 bits e taxa de 680 kbps. As demais UEs utilizam individualmente um balde com 30560 bits de capacidade e uma taxa de 306 kbps e o conjunto de todas as UEs recebem os dados dos tráfegos por meio do canal *downlink* da mesma eNodeB. A quantidade de estações que solicitam tráfego 1 ou tráfego 2, utilizadas para as medições dos três escalonadores, pode ser verificada na Tabela 11.

Tabela 11 – Quantidade de estações por tráfego requerido

	UEs	UEs	UEs
Tráfego 1	29	39	49
Tráfego 2	1	1	1

Fonte: Autoria Própria

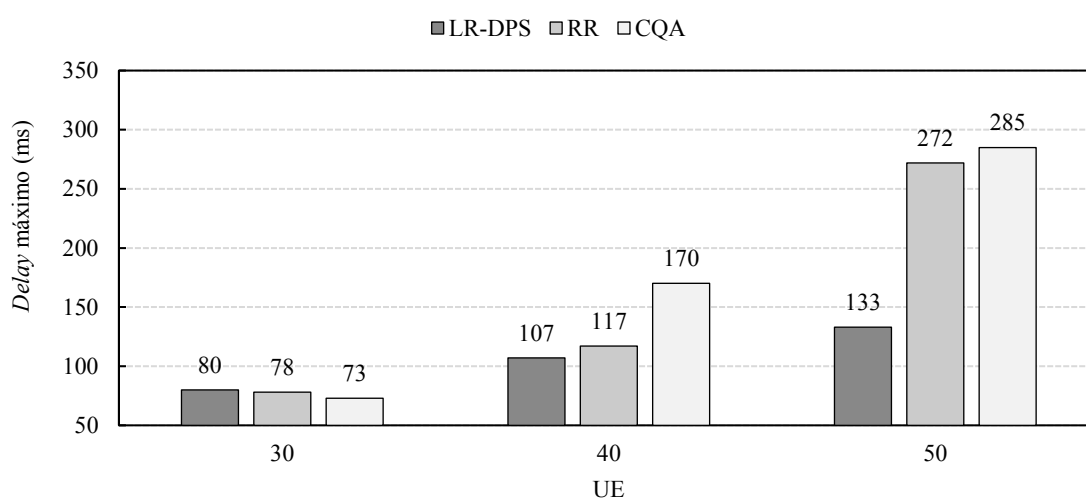
### 5.3.1 Delay Máximo

A Figura 27 mostra que a partir da utilização do modelo proposto foi possível garantir o *delay* máximo solicitado para dois tráfegos VBR distintos. Os resultados demonstram que para 30 UEs, onde 1 UE solicita o tráfego 2 e 29 solicitam o tráfego 1 o *delay* máximo medido foi de 80 ms para o LR-DPS, 78 ms para o escalonador RR e 73 para o escalonador CQA. Portanto, nota-se uma desvantagem do modelo proposto em relação ao CQA, entretanto entende-se que este comportamento ocorre quando a quantidade de UEs está abaixo da quantidade máxima UEs suportada em uma eNodeB para o *delay* requisitado pelos usuários.

Para o cenário com 40 UEs, onde temos 1 UE que solicita o tráfego 2 e 39 que solicitam o tráfego 1, percebe-se que para o modelo LR-DPS ocorre o atendimento do *delay* requisitado de 150 ms e certa vantagem em relação aos outros escalonadores verificados. Os *delays* medidos são de 107 ms para o modelo proposto, 117 ms para o escalonador RR e 170 ms para o escalonador CQA.

Por fim, os resultados de *delay* máximo para 1 UE com tráfego 2 e 49 UEs com tráfego 1, demonstram que o escalonador RR excedeu o limite de 150 ms em 81% e o escalonador CQA em 90%. Desta forma, nota-se que há uma vantagem expressiva do escalonador proposto em relação aos escalonadores RR e CQA na medida em que a densidade de UEs aumenta para 40 e 50 UEs e que o atendimento do *delay* máximo é suportado para os casos apresentados pela Figura 27.

Figura 27 – *Delay* máximo para dois tráfegos de vídeo



Fonte: Autoria Própria

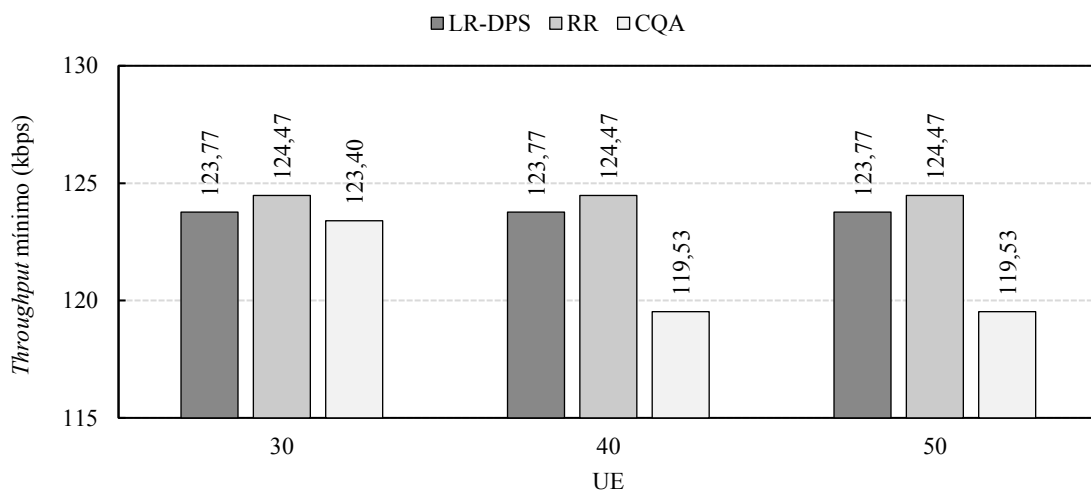
A vantagem verificada para o modelo proposto está na forma como o tráfego 2 é tratado no processo de escalonamento. Uma vez que o tráfego de entrada é conhecido e modelado por um balde de fichas parametrizado por um modelo matemático, tem-se como resultado prático a reserva de mais RBs para UEs com mais dados a serem alocados no mesmo período de tempo. Por outro lado, se o processo é realizado de forma agnóstica em relação ao tráfego de entrada, tráfegos diferentes terão a mesma taxa de dados, resultando em maior *delay* para tráfegos com taxas maiores.

### 5.3.2 *Throughput* Mínimo

Observando os dados recebidos na camada PDCP dos UEs foi medido o valor de *throughput* mínimo, que corresponde a menor vazão de dados verificada no sentido downlink em um determinado período. Neste cenário, o cálculo é feito para intervalos de 1 segundo onde para cada UE são realizadas 30 medições referentes aos 30 segundos de transmissão. Em seguida seleciona-se o menor valor dentre as medições.

Na Figura 28, nota-se que o escalonador LR-DPS, assim como o escalonador RR, apresentam *throughput* mínimo com valores próximos a 123 kbps e 124 kbps para três densidades de UEs. Para o escalonador CQA verifica-se que o valor mínimo para 30 UEs fica em torno de 123 kbps e quando a densidade de UEs aumenta observa-se uma pequena queda do *throughput* para cerca de 119 kbps. Diante destes resultados, pode-se dizer que os três escalonadores não apresentam grande diferença de vazão mínima, quando comparados entre si e quando um escalonador é verificado com diferentes densidades de UEs.

Figura 28 – *Throughput* mínimo para dois tráfegos de vídeo



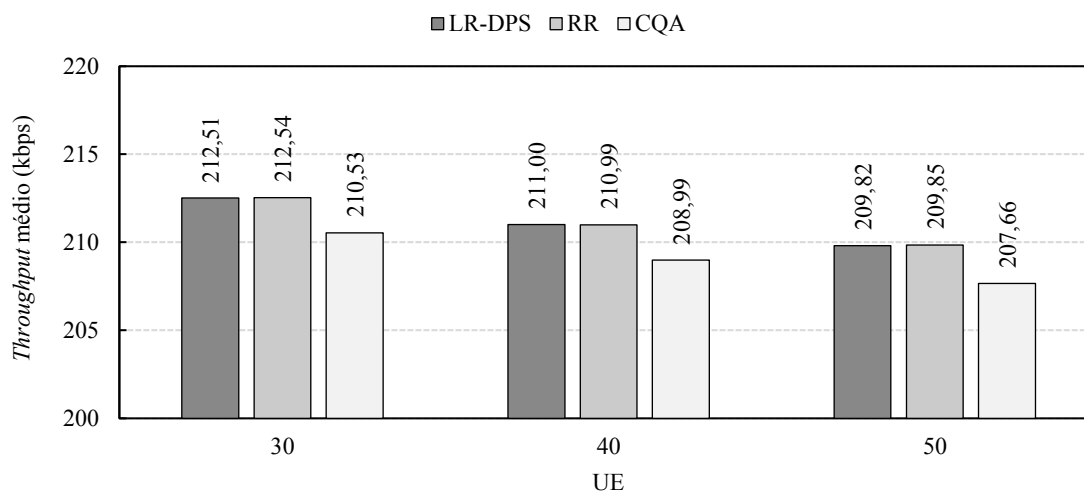
Fonte: Autoria Própria

### 5.3.3 *Throughput* Médio

A partir da mesma simulação, foram extraídos os valores médios para a vazão com três densidades diferentes. O cálculo, assim como no *throughput* mínimo, considera a vazão medida a partir de intervalos de 1 segundo, a partir dos quais são calculados os valores médios encontrados para cada densidade de UE. Os valores verificados a partir do gráfico da Figura 29, mostram que os valores médios de vazão permaneceram próximos, da mesma forma que os valores mínimos verificados.

Os valores observados para a vazão média apresentam uma pequena queda na taxa na medida em que a quantidade de usuários aumenta, considerasse então, que o sistema está subutilizado até 50 UEs. Utilizando-se de recursos finitos para alocação, menos recursos são alocados por usuário no mesmo período de tempo se houverem mais UEs disputando o meio de transmissão. Assim, em uma situação de estresse da rede, tanto o *delay* máximo não será atendido como o *throughput* médio irá reduzir consideravelmente.



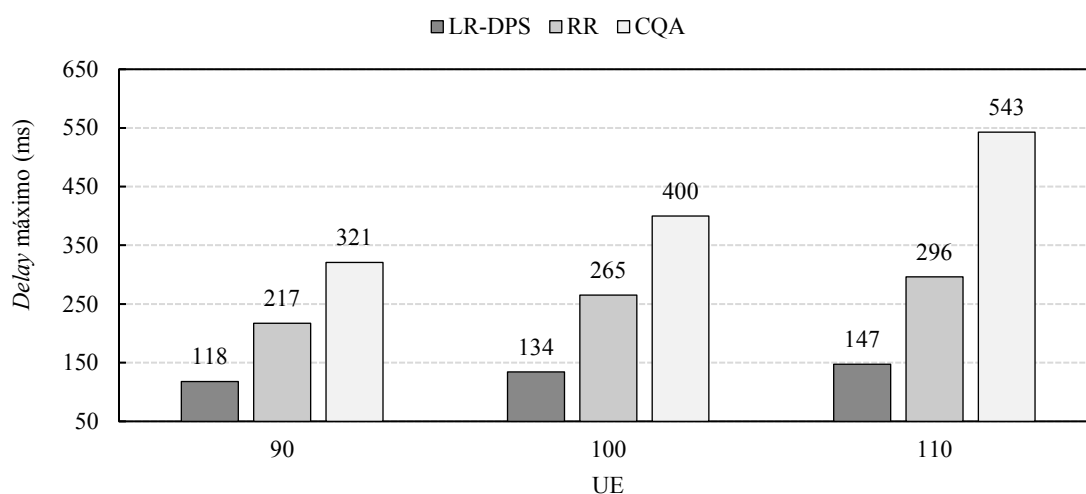
Figura 29 – *Throughput* médio para dois tráfegos de vídeo

Fonte: Autoria Própria

#### 5.3.4 Bandwidth

Diante das possibilidades de largura de banda disponíveis para redes LTE, relacionadas no Capítulo 2, busca-se verificar a quantidade de usuários alocados em um cenário utilizando uma banda de 10MHz em um limite de *delay* máximo de até 150 ms. Uma vez que uma largura de banda maior é selecionada, mais recursos estão disponíveis para alocação e por consequência mais UEs podem ser alocadas. A largura de banda de 10MHz possui 50 RBs por TTI, os quais são agrupados em RBGs de 3 RBs. Os demais parâmetros permanecem os mesmos da configuração inicial considerando dois tráfegos distintos.

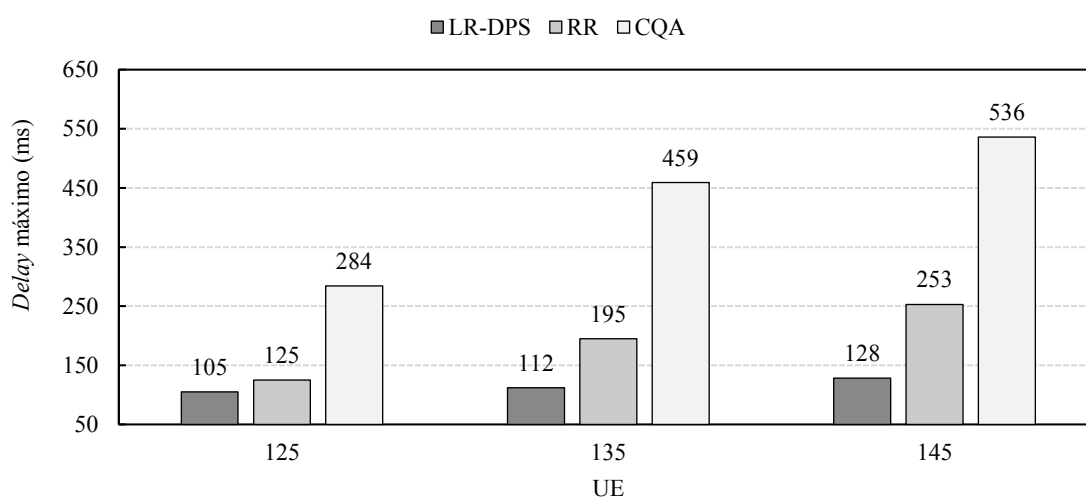
Os resultados obtidos por meio das simulações utilizando 10MHz como largura de banda são mostrados na Figura 30. Para um cenário com 1 UE solicitando o tráfego 2 e 89 solicitando o tráfego 1, foram percebidos *delays* de 118 ms para o escalonador LR-DPS, 217 ms para o RR e 321 ms para o CQA. Para o cenário onde 1 UE solicita o tráfego 2 e 99 solicitam o tráfego 1, foram verificados *delays* de pico de 134 ms para o modelo proposto, 265 ms para o RR e 400 ms para o CQA. Por fim para 1 UE que solicita o tráfego 2 e 109 que solicitam o tráfego 1, foram verificados *delays* máximos de 147 ms para o LR-DPS, 296 ms para o escalonador RR e 543 ms para o escalonador CQA.

Figura 30 – *Delay* máximo em banda de 10 MHz

Fonte: Autoria Própria

Os valores apurados na Figura 30, demonstram que com o escalonador proposto e uma largura de banda de 10 MHz, foi possível alocar até 110 usuários em uma eNodeB sem que o *delay* máximo de 150 ms fosse ultrapassado. Nota-se ainda, que para as demais disciplinas de escalonamento avaliadas, os valores de *delay* máximo ultrapassam os 150 ms mesmo em cenários de menor densidade, com 90 ou 100 UEs.

A Figura 31 apresenta os valores de *delay* máximo obtidos para uma largura de banda de 15 MHz, variando de 125 para 145 usuários. Para esta largura de banda, 75 blocos de recursos estão disponíveis em RBGs de tamanho 4.

Figura 31 – *Delay* máximo em banda de 15 MHz

Fonte: Autoria Própria

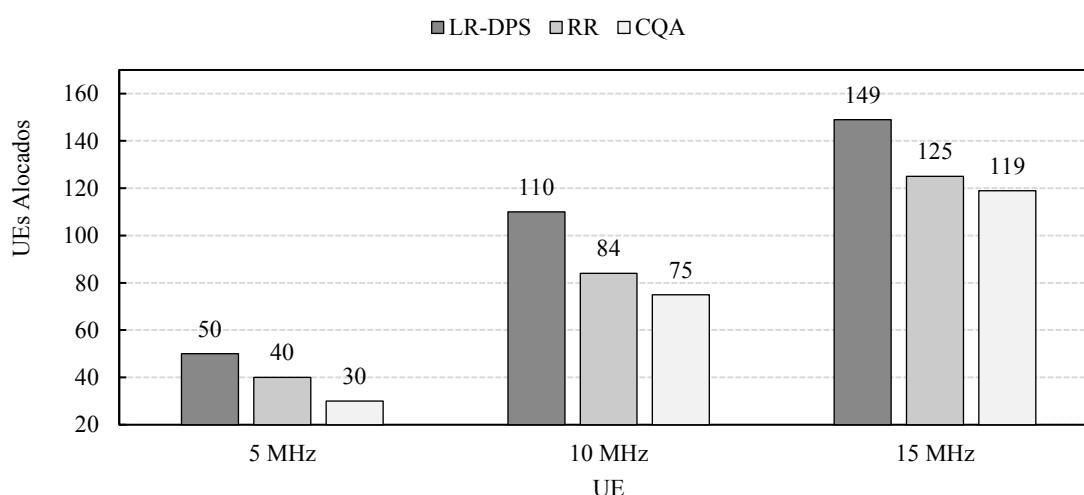
Verifica-se então, que o comportamento em relação ao *delay* máximo se manteve vantajoso para o escalonador proposto. Este desempenho pode ser explicado pelo fato de que os outros escalonadores tratam de forma igual os dois tráfegos que possuem taxas diferentes.

Na medida em que a densidade de usuários aumenta, comparando o evolução do *delay* máximo para cada escalonador, percebe-se que o escalonador proposto possui menor variação em relação aos demais. Esta característica é proporcionada primordialmente pelo *delay* suportado para a entrega dos pacotes originados por meio do tráfego 2. Com mais usuários para alocar, as filas de entrega para usuários que solicitam o tráfego 2 aguardam mais tempo para serem alocadas se não forem priorizadas. Desta forma, a verificação do *delay* HOL da disciplina de escalonamento CQA, proposta por (BOJOVIC; BALDO, 2014), não foi suficiente para amenizar o *delay* máximo verificado para os tráfegos VBR utilizados nas simulações. Por outro lado, nota-se que, para dois tráfegos VBR, a disciplina de escalonamento LR-DPS produz um resultado promissor em relação aos demais escalonadores verificados. Entretanto, para o escalonador LR-DPS, vale ressaltar que é necessário o conhecimento prévio do tráfego de entrada, para que os parâmetros do balde sejam configurados de forma adequada.

### 5.3.5 UEs Alocados em Diferentes Larguras de Banda

A partir dos resultados obtidos para diferentes larguras de banda é possível analisar os limites de usuários atendidos, levando em conta um *delay* máximo de 150 ms para 5, 10 e 15 MHz. O gráfico da Figura 32 mostra que para as três larguras de banda testadas o escalonador proposto alocou mais usuários em relação aos escalonadores RR e CQA.

Figura 32 – Quantidade de UEs alocados em diferentes larguras de banda



Fonte: Autoria Própria

### 5.3.6 *Delay* Máximo com Diferentes Proporções de Tráfegos

Considerando que tráfegos diferentes resultam em parâmetros de configuração distintos para o escalonador LR-DPS, busca-se verificar o desempenho do mesmo em três

cenários de teste com diferentes proporções de tráfego. Os valores utilizados para três os testes propostos podem ser verificado na Tabela 12.

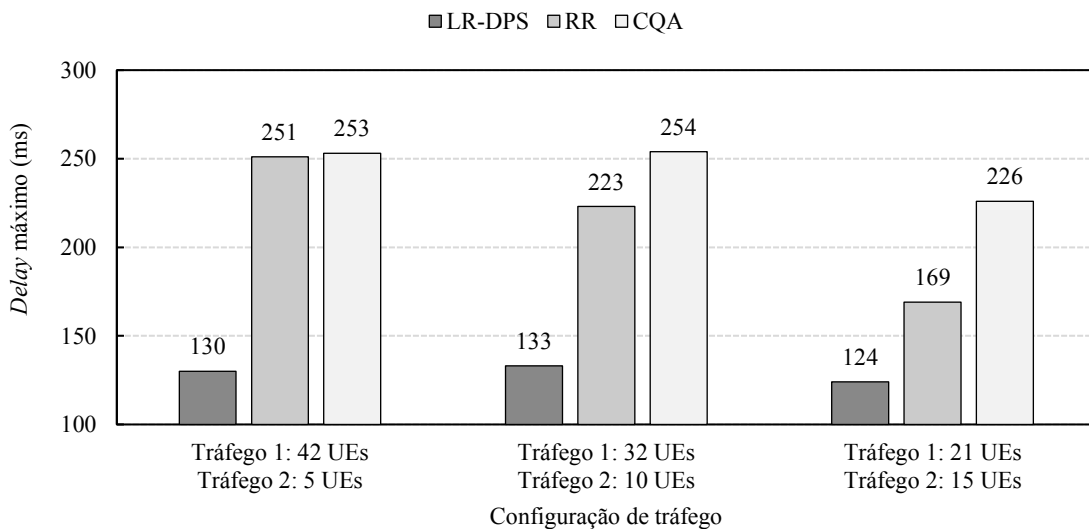
Tabela 12 – Parâmetros de simulação para dois tráfegos distintos

Teste	1		2		3	
Tráfego	1	2	1	2	1	2
UEs	42	5	32	10	21	15
$r_i$ (kbps)	333,280	706,910	333,280	706,910	333,280	706,910
TRBG	12,39	26,27	12,39	26,27	12,39	26,27
$T_T$ (ms)	55		55		55	
R (kbps)	17.532		17.734		17.603	

Fonte: Autoria Própria

O desempenho do LR-DPS é comparado com os escalonadores RR e CQA. No primeiro teste, 42 UEs solicitam o tráfego 1 enquanto 5 UEs requisitam o tráfego 2. No segundo teste, 32 UEs solicitam o tráfego 1 e 10 UEs solicitam o tráfego dois. Para o terceiro teste, 21 UEs são atendidos com o tráfego 1 enquanto 15 UEs são atendidos com o tráfego 2. A Figura 33 mostra que o *delay* máximo de 150 ms foi atendido apenas com o escalonador LR-DPS nos três testes propostos.

Figura 33 – Quantidade de UEs atendidos em diferentes proporções de tráfegos

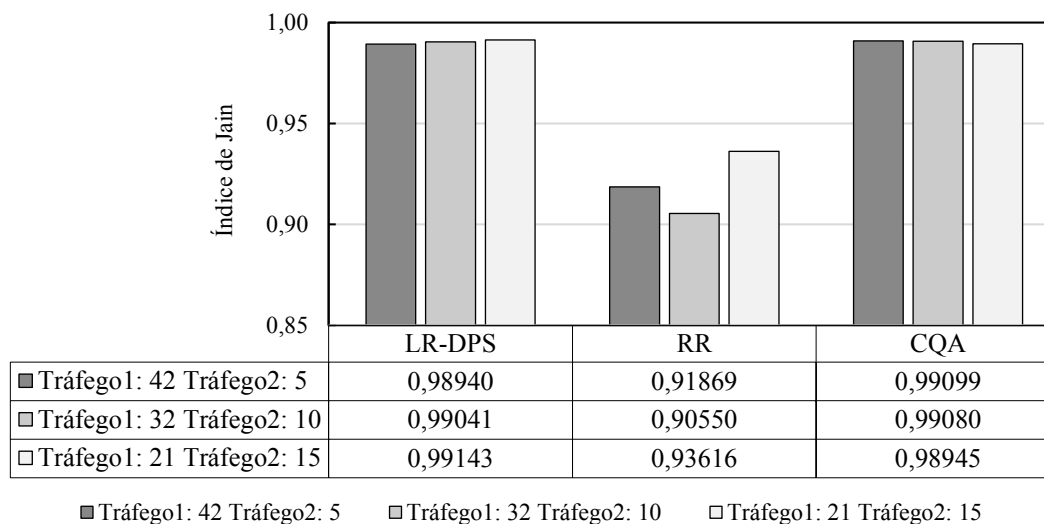


Fonte: Autoria Própria

O LR-DPS busca uma melhor distribuição dos *delays* máximos individuais por meio da forma de distribuição dos RBGs na Etapa 3. Assim, este mesmo cenário de teste é utilizado para verificar a justiça entre os três escalonadores e validar um dos propósitos da Etapa 3. Para tanto, utiliza-se o índice de Jain (JAIN; DAH-MING; HAWK, 1984) que varia entre 0 e 1, onde 1 representa o caso de maior justiça. A partir dos *delays* máximos individuais, medidos em cada cenário de teste, verifica-se que os escalonadores LR-DPS e CQA apresentaram melhores

resultados, onde os índices de justiça se mantiveram em torno de 0,99. Para o escalonador RR os índices variam entre 0,91 e 0,93, o que representa menor justiça entre os UEs em relação ao LR-DPS e CQA. Estes resultados são mostrados na Figura 34.

Figura 34 – Índice de Justiça para *delay* individual em diferentes proporções de tráfegos



Fonte: Autoria Própria

### 5.3.7 UEs Alocados com Diferentes Requisitos de *Delay* Máximo

Por fim, busca-se verificar o desempenho do escalonador LR-DPS para diferentes requisitos de *delay*, sendo eles 100 ms e 150 ms para *delay* máximo solicitado. Assim, três cenários de teste são examinados, onde a quantidade de UEs que solicitam atraso máximo de 100 ms são de 10, 15 e 20, e as quantidade de UEs que podem ser atendidas com 150 ms são calculadas a partir da Etapa 2 do modelo. Os valores utilizados para os três testes com o tráfego 1 podem ser verificados na Tabela 13.

Tabela 13 – Parâmetros de simulação para dois *delays* distintos

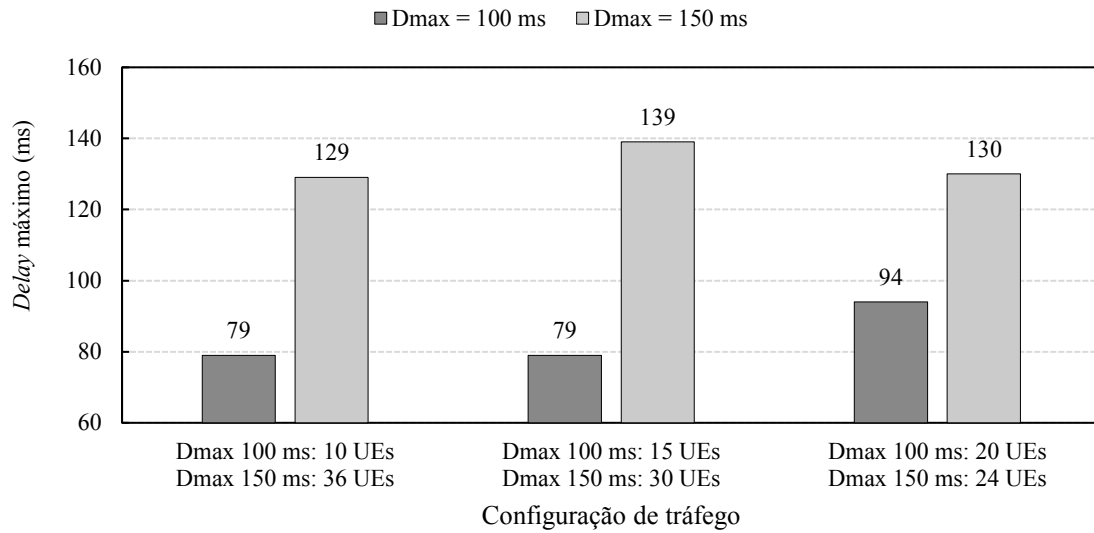
Teste	1		2		3	
<i>Dmax</i> (ms)	100	150	100	150	100	150
UEs	10	36	15	30	20	24
$r_i$ (kbps)	474,420	352,250	461,090	357,030	450,340	362,920
TRBG	10,26	7,62	9,03	7,00	7,91	6,38
$T_T$ (ms)	32		29		26	
R (Kbps)	17.425		17.627		17.717	

Fonte: Autoria Própria

A Figura 35 apresenta os resultados obtidos por meio das simulações com dois requisitos de *delay* máximo solicitados. Os resultados mostram que o escalonador LR-DPS atendeu os requisitos de *delay Dmax* de 100 ms e 150 ms para os três testes propostos. Considerando que a taxa do meio é de 17.760 kbps para a largura de banda de 5MHz, verifica-

se que foi possível atender 46 UEs para o Teste 1, 45 UEs para o Teste 2, e 44 UEs para o Teste 3.

Figura 35 – Quantidade de UEs atendidos em diferentes requisitos de *delay*



Fonte: Autoria Própria

## 6 CONCLUSÕES E TRABALHOS FUTUROS

As redes LTE são o padrão adotado para as redes móveis celulares atuais de quarta geração, as quais possuem um notório crescimento anual em termos de utilização. Dentre os principais serviços utilizados nessas redes, destacam-se aqueles com base em tráfego de vídeo, o qual pode exigir garantias de QoS que devem ser providas pelas eNodeBs. Entretanto, os modelos de escalonamento verificados não atendem de forma satisfatória requisitos de *delay*. Neste sentido, tendo em vista as características de um tráfego de taxa variável, assim como o fato de que as redes LTE não definem um procedimento de escalonamento, este trabalho apresentou o LR-DPS como uma proposta para provisão de QoS referente ao *delay* gerado durante o escalonamento de recursos. Para isso, foi utilizada uma técnica de condicionamento de tráfego com garantia de *delay* máximo, solicitado pela aplicação, para tráfegos de taxa variável.

O modelo de escalonamento LR-DPS garante o *delay* máximo para tráfego no sentido *downlink* em redes LTE e inspira-se no trabalho de Dosciatti (2015), o qual apresenta uma nova arquitetura para provisão de QoS no sentido *uplink* em redes WiMAX. Entretanto, o modelo aplicado em redes WiMAX utiliza alocação por divisão de tempo, enquanto para redes LTE utiliza-se a alocação por divisão de tempo e frequência, o que acarreta em modificações no modelo. Assim, este trabalho contribui não somente com uma nova abordagem para redes LTE, mas também com uma nova visão do modelo anterior onde, agora, os recursos para alocação são divididos no espaço tempo/frequência.

O modelo de escalonamento LR-DPS foi proposto de forma modular em três etapas hierárquicas utilizadas pela eNodeB para o processo de alocação de RBGs. Na Etapa 1, considerou-se apenas o tratamento do tráfego de entrada por um balde de fichas. Esta etapa pode ser fonte de um atraso extra, devido a necessidade de informações sobre as características de tráfego, as quais devem ser calculadas para ajuste dos parâmetros do balde de fichas ou informadas de fonte externa. Na Etapa 2, são calculados os valores correspondentes às taxas de cada UE para atendimento aos requisitos de *delay* máximo exigido. Nesta etapa são conhecidos os parâmetros de configuração do balde de fichas, o *delay* máximo requisitado, a taxa do meio físico e a capacidade de transmissão por RBG. A taxa do meio físico, assim como a capacidade de transmissão por RBG são calculadas a partir da banda disponível para a eNodeB, e a qualidade de sinal verificada segundo as possibilidades para redes LTE expostas no Capítulo 2. Esta etapa, calcula o tempo ideal necessário para alocação dos recursos ( $T_T$ ), além das taxas de

dados de cada usuário ( $r_i$ ). Tratando-se de um problema de difícil resolução, optou-se pela utilização da função *fmincon* do software MATLAB® para resolver o problema ATT. Por fim, na Etapa 3, considera-se um algoritmo de alocação de recursos para atendimento dos parâmetros obtidos na Etapa 2, o qual foi estruturado para promover maior justiça entre os UEs com dados para transmissão. Assim, a partir de um total de RBGs calculados para cada UE (TRBG), os dados na fila de transmissão são alocados de forma sequencial e cíclica. Este procedimento ocorre até que todos os usuários sejam atendidos no intervalo  $T_T$  e que toda a transmissão seja concluída.

Para verificar o desempenho do LR-DPS, foram montados cenários utilizando o simulador de rede ns-3, que com a configuração de módulos adicionais tornou possível a simulação de transmissão de tráfego de vídeo VBR, trazendo mais realismo as simulações. Inicialmente foram realizados testes em um cenário simplificado para validar o modelo LR-DPS e comprovar a eficiência do modelo para redes LTE. Em seguida foram escolhidos os escalonadores *Round Robin* (RR) e *Channel and QoS Aware* (CQA) para fins de comparação, sendo o primeiro deles como uma referência amplamente conhecida e o segundo por possuir características mais próximas da proposta deste trabalho. As simulações realizadas exploraram diversas possibilidades de configuração do ambiente de simulação, garantindo assim a avaliação do modelo proposto em todas as suas possibilidades.

No cenário 1 foram realizadas simulações com 30, 40 e 50 UEs, uma eNodeB e uma única fonte de tráfego VBR, onde foi possível atender um *delay* máximo de 150 ms em largura de banda de 5 MHz para em todos os casos. Os valores de *delay* máximo medidos foram de 78 ms, 105 ms e 130 ms, comprovando assim a ação do modelo para redes LTE. No cenário 2, um segundo tráfego VBR com taxas de pico maiores foi acrescentado ao ambiente de simulação, tornando possível encontrar dois conjuntos de parâmetros para execução do LR-DPS. Foi possível verificar que o LR-DPS atendeu os requisitos de *delay* máximo de 150 ms com dois tráfegos distintos e os escalonadores RR e CQA excederam o limite em 80% e 90%, respectivamente. Verificou-se ainda que os resultados de *throughput* mínimo e médio foram bastante próximos entre os três escalonadores, demonstrando assim que o atendimento em termos de *delay*, verificado para o escalonador LR-DPS, não comprometeu a vazão dos tráfegos.

Dentre as possibilidades de configuração da rede LTE, foram exploradas configurações de banda maiores como 10 MHz e 15 MHz, além da banda de 5 MHz utilizada nas demais simulações. Neste caso, além de garantir o *delay* máximo para todas as bandas, observou-se maior capacidade de alocação de usuários do LR-DPS frente aos escalonadores



RR e CQA. Em banda de 5 MHz, 10MHz e 15MHz, o LR-DPS alocou pelo menos 10 usuários a mais que os demais, sendo seu melhor resultado com diferença de 24 usuários alocados frente ao RR e 30 frente ao CQA.

Tendo em vista os dois tráfegos utilizados nas simulações, verificou-se o atendimento ao *delay* máximo de 150 ms em diferentes proporções de tráfego solicitado. Assim, foram calculadas as quantidades máximas de usuários que poderiam ser atendidas com o tráfego 1 quando 5, 10 e 15 usuários solicitaram o tráfego 2. Para esses limites de usuários calculados, o LR-DPS não somente garantiu o *delay* requisitado para todos os UEs, mas também apresentou melhor desempenho nos três casos verificados.

Simulações realizadas com diferentes requisitos de *delay* máximo, demonstraram que o LR-DPS foi capaz de garantir *delays* de 100 ms e 150 ms de forma concorrente. Para um *delay* máximo de 100 ms foram observados valores de pico entre 79 ms e 94 ms, enquanto para o *delay* de 150 ms solicitado foram verificados valores de pico entre 129 ms e 139 ms. De forma coerente, observou-se que o número total de usuários atendidos foi menor quando mais usuários solicitaram um *delay* máximo de 100 ms.

Analisando os resultados obtidos por meio das simulações, constata-se que o modelo LR-DPS atendeu ao seu propósito e garantiu o *delay* máximo solicitado em diversas situações. Do mesmo modo, verificou-se que a quantidade de usuários no sistema foi maior em vista aos outros dois escalonadores analisados (RR e CQA).

## 6.1 TRABALHOS FUTUROS

Abaixo são relacionados alguns tópicos que podem ser explorados em trabalhos futuros relacionados a esta dissertação:

- A partir do LR-DPS, desenvolver um sistema genérico de fácil adaptação para outras tecnologias de redes de comutação de pacotes que possam utilizar a técnica de transmissão OFDMA;
- Desenvolver um modelo para ajuste dinâmico dos parâmetros da Etapa 1 e Etapa 2 do LR-DPS;
- Utilizando um ajuste dinâmico na Etapa 1, analisar como o tamanho do balde de fichas pode afetar o processo de alocação utilizando diferentes intervalos de tempo para ajuste de parâmetros durante as transmissões de dados;

- Analisar diferentes meta-heurísticas para resolução do problema ATT;
- Verificar o desempenho do LR-DPS em um ambiente multi-células e multi-antenas (MIMO 2X2, MIMO 4X4);
- Ampliar as simulações com garantia para tráfegos concorrentes de VoIP e de taxa constante (CBR);
- Alterar o Algoritmo 1 para a possibilidade de uma distribuição aleatória uniforme de RBGs para promover maior justiça entre os UEs.

## REFERÊNCIAS

- 3GPP, About 3GPP. Disponível em: <http://www.3gpp.org/about-3gpp>. Acesso em 12 jan. 2018.
- BALDO, N. et al. New model for the simulation of the LTE-EPC data plane. In: Proc. of Workshop on ns-3 (WNS3 '12), 2012.
- BHALLA, M. R.; BHALLA, A. V. Generations of Mobile Wireless Technology: A Survey. *International Journal of Computer Applications*, v. 5, n., 4 p. 26–32, 2010.
- BOJOVIC, B.; BALDO, N. A new channel and QoS aware scheduler to enhance the capacity of voice over LTE systems, *IEEE 11th International Multi-Conference on Systems, Signals & Devices (SSD14)*, Barcelona, 2014, p. 1-6.
- CAPOZZI, F. et al. Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey. *IEEE Communications Surveys and Tutorials*, v. 15, n. 2, p. 678–700, 2013.
- COX, C. *An Introduction to LTE: LTE, LTE-Advanced, SAE, VoLTE and 4G Mobile Communications*. 2st Edition. Wiley, 2014.
- DA MATA, S.; GUARDIEIRO, P.R. Resource Allocation for the LTE Uplink Based on Genetic Algorithms in Mixed Traffic Environments. *Computer Communications*. v. 107, p. 125-137, 2017.
- DAHLMAN, E.; PARKVALL, S.; SKÖLD, J. *4G, LTE – Advanced Pro and The Road to 5G*. Third Edition. Academic Press, 2016.
- DOSCIATTI, E. R. Uma nova arquitetura para provisão de QoS utilizando enxame de partículas em rede WiMAX fixas. 2015. Tese (Doutorado em Engenharia Elétrica e Informática Industrial) – Universidade Tecnológica Federal do Paraná, Curitiba, 2015.
- ERICSSON, Ericsson Mobility Report, Ericsson 2017. Disponível em: <https://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-november-2017-central-and-eastern-europe.pdf>. Acesso em 12 jan. 2018.
- EVALVID. Department of Telecommunication Systems Telecommunication Networks Group. EvalVid. Disponível em: <http://www.tkn.tu-berlin.de/menue/research/evalvid>. Acesso em: 12 jul. 2017.

FERDOSIAN, N. et al. Throughput-aware Resource Allocation for QoS Classes in LTE Networks. *Procedia Computer Science*. 59. p. 115 - 122. 2015.

GERCOM, NS3 Evalvid module. Disponível em: <<https://github.com/gercom/evalvid-ns3>>. Acesso em: 25 jun. 2017.

GHOSH, A. et al. *Fundamentals of LTE*. 1st Edition. Prentice Hall Press, 2011.

GROSS, J.; GÜNEŞ, M. Introduction. In: WEHRLE, K.; GÜNEŞ, M.; GROSS, J. *Modeling and Tools for Network Simulation*. Berlin Heidelberg: Springer-verlag, 2010, p. 1-10.

GUAN, Na. et al. QoS guaranteed resource block allocation algorithm for LTE systems. In: *IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Wuhan, 2011, p. 307-312.

HIGUCHI, Y.; FORONDA, A. et al. Delay guarantee and service interval optimization for HCCA in IEEE 802.11 e WLANs. In: *Proc. IEEE Wireless Communications & Networking Conference*. p. 2080-2085, 2007.

JAIN, R.; DAH-MING, W.; HAWKES, W. R. A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. *Computer Systems, ACM Transactions on*, 1984.

LAI, W. K.; TANG, C. L. QoS-aware downlink packet scheduling for LTE networks. *Computer Networks*. v. 57, p. 1689–1698, 2010.

NAGLE, J. B. On packet switches with infinite storage. In: *Innovations in Internetworking*. Artech House, Inc., 1988 p. 136-139.

NS-3. What is ns-3. 2011-2017. Disponível em: <<https://www.nsnam.org/overview/what-is-ns-3/>>. Acesso em: 28 out. 2017.

PRASAD, S. S.; SHUKLA, C. K.; CHISAB, R. F. *Performance Analysis of OFDMA in LTE*. ICCCNT. Coimbatore, India, 2012.

RILEY, G. F.; HENDERSON, T. R. The ns-3 Network Simulator. In: WEHRLE, K.; GÜNEŞ, M.; GROSS, J. *Modeling and Tools for Network Simulation*. Berlin Heidelberg: Springer-verlag, 2010, p. 15-33.

SEZIA, S.; TOUFIK, I.; BAKER, M. LTE - the UMTS Long Term Evolution: From Theory to Practice. 2nd Edition, Wiley, 2011

SHARMA, P. Evolution of Mobile Wireless Communication Networks-1G to 5G as well as Future Prospective of Next Generation Communication Network. International Journal of Computer Science and Mobile Computing, v. 2, Issue. 8, p.47-53, 2013.

SIMON, C.; LEUS, G. Round-robin scheduling for time-varying channels with limited feedback. In: IEEE 10th Workshop on Signal Processing Advances in Wireless Communications, Perugia, 2009, p. 231-234.

STILIADIS, D.; VARMA, A. Latency-Rate Servers: A General Model for Analysis of Traffic Scheduling Algorithms. IEEE-ACM Transactions on Networking, vol.5, pp.611-624, 1998.

SU, L.; WANG P.; LIU, F. Particle swarm optimization based resource block allocation algorithm for downlink LTE systems. In: 18th Asia-Pacific Conference on Communications (APCC), 2012, p. 970-974.

TANENBAUM, A. S.; WETHERALL, D. J. Redes de Computadores. 5<sup>a</sup> ed. Pearson Education do Brasil, 2011.

TURNER, J. S. New directions in communications (or which way to the information age?). In: IEEE Communications Magazine, vol. 40, no. 5, p. 50-57, 2002.

WANG, L.; Luo, Z. A Simple SQP Algorithm for Constrained Finite Minimax Problems. The Scientific World Journal, 2014.

WILLE, E. C. G. et al. A Lagrangean relaxation approach for QoS networks CFA problems. AEU-International Journal of Electronics and Communications, v. 63, p. 743-753, 2009

YANG, S. C. OFDMA System Analysis and Design. Norwood: Artech House, 2010.

ZHAO, L. et al. QoS guaranteed resource block allocation algorithm in LTE downlink. In: 7th International Conference on Communications and Networking in China, 2012, p. 425-429.

## APÊNDICE A – O PROBLEMA MINIMAX

O problema Minimax é um dos mais importantes problemas de otimização não diferenciáveis, e corresponde a minimizar o máximo valor de um número de variáveis de decisão (WANG; LUO, 2014). O problema Minimax estendido pode ser formulado como sendo o seguinte problema de otimização:

$$\min \left( \max_i f_i(x) + \dots + \max_k h_k(x) \right) \quad (21)$$

Sujeito a:

$$lb \leq x \leq ub \quad (22)$$

onde  $f_i(x), \dots, h_k(x)$  são diferenciáveis.

Como a função objetivo não é diferenciável, não se pode, diretamente, usar métodos clássicos para solução de problemas de otimização restritos. Entretanto, pode-se converter apropriadamente sua formulação. Desta forma, um problema Minimax pode ser expresso de forma equivalente pela minimização de variáveis auxiliares que correspondem a um upper bound para cada função individual ( $f_i(x), \dots, h_k(x)$ ), como segue:

$$\min \sum_{j \in N} z_j \quad (23)$$

Sujeito a:

$$\begin{aligned} f_i(x) &\leq z_1, \forall i \\ &\dots \\ g_k(x) &\leq z_N, \forall k \\ lb &\leq x \leq ub \end{aligned} \quad (24)$$

## APÊNDICE B – COMPARAÇÃO ENTRE PROPOSTAS

As tabelas 15, 16 e 17, mostram os resultados obtidos, com a função *fmincon()* MATLAB® e com a heurística da decomposição, para simulações com dois tráfegos, onde, para cada cenário de teste, alterna-se as proporções de tráfegos utilizadas.

Tabela 15 – Comparação entre as propostas (Teste 1).

Cenário	1 - MATLAB		1 - Heurística	
	1	2	1	2
Tráfego				
UEs	42	5	42	5
$r_i$ (kbps)	333.280	706.910	332.965	706.965
TRBG	12,38	26,27	12,37	26,27
$T_T$ (ms)	54,9		54,9	
R (kbps)	17.532		17.519	

Fonte: Autoria Própria

Tabela 16 – Comparação entre as propostas (Teste 2).

Cenário	2 - MATLAB		2 - Heurística	
	1	2	1	2
Tráfego				
UEs	32	10	32	10
$r_i$ (kbps)	333.280	706.910	332.965	706.965
TRBG	12,38	26,27	12,37	26,27
$T_T$ (ms)	54,9		54,9	
R (kbps)	17.734		17.724	

Fonte: Autoria Própria

Tabela 17 – Comparação entre as propostas (Teste 3).

Cenário	3 - MATLAB		3 - Heurística	
	1	2	1	2
Tráfego				
UEs	21	15	21	15
$r_i$ (kbps)	333.280	706.910	332.965	706.965
TRBG	12,38	26,27	12,37	26,27
$T_T$ (ms)	54,9		54,9	
R (kbps)	17.603		17.596	

Fonte: Autoria Própria