



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Alocação de blocos de recurso em redes LTE sensível à heterogeneidade dos fluxos de dados

Guilherme David Branco

Dissertação apresentada como requisito parcial para
conclusão de Mestrado em Informática

Orientador
Prof. Dr. Jacir Luiz Bordim

Brasília
2020

Dedicatória

Dedico o presente trabalho aos meus pais, Maria Luiza David e Geraldo de Souza Branco, assim como à minha namorada, Aline Albuquerque dos Angelos, que auxiliaram na confecção, revisão, e acima de tudo, suportaram e deram apoio em todas as dificuldades enfrentadas na jornada desta pesquisa.

Agradecimentos

Agradeço ao meu orientador, Dr. Jacir Luiz Bordim, por compartilhar seus conhecimentos. Assim como aos membros da banca de qualificação, doutores Marcelo Antonio Marotta e Eduardo Alchieri, por avaliarem e proporcionarem uma melhoria sem igual a essa pesquisa. Agradeço, também, aos membros do projeto 5G Range, alunos e professores da Universidade de Brasília, quanto aos parceiros do projeto ao redor do mundo. Por fim, agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo suporte oferecido por meio de acesso a diversos periódicos e bolsas de pesquisa.

Resumo

A alocação de blocos de recurso constitui uma área marcante dentro de redes sem fio com modulação *Orthogonal Frequency Division Multiple Access* (OFDM), em especial a LTE 4G. Os blocos de recurso definem como acomodar os usuários da rede, permitindo maior vazão e atendimento aos usuários da célula, a partir de uma boa alocação dos recursos disponíveis. A proliferação de dispositivos para a Internet das Coisas pode aumentar a demanda das redes LTE, trazendo à tona a necessidade de mecanismos mais adaptados aos tráfegos com requisitos voltados para esses dispositivos. Esta pesquisa apresenta a proposta de alocação para redes LTE sensível à heterogeneidade dos tipos de fluxos de dados, o mecanismo possui duas camadas e apoia-se em teoria de controle e priorização por latência, obtendo um melhor desempenho para vazão, índice de justiça e taxa de perda de pacotes quando comparado a outras soluções presentes na literatura. A saber, para o cenário em que ocorre uma proporção de 75% de tráfego classificado como tempo real, contra 25% como não tempo real, refletindo um caso comum para ambientes reais, atingiu-se um desempenho de vazão 46 vezes maior para o tráfego *web* quando comparado ao escalonador *Frame Level Scheduler*, mantendo níveis equiparados para o tráfego de vídeo, já comparado à estratégia *Exponential Rule* obtém um resultado 75% maior, menos expressivo, mas ainda considerável.

Palavras-chave: Redes móveis, Alocação de blocos de recurso, *Long Term Evolution* 4G

Abstract

Resource block scheduling composes a memorable area within OFDM modulated wireless networks, such as LTE 4G. A good distribution of resource blocks enables a better fit for the network, enhancing metrics like throughput, fairness, packet loss ratio and others. Internet of Things (IoT) devices are experiencing an unending increase, and are demanding more of the networks, such a rapid growth makes need for newer mechanisms that are better suited for such traffic. This work presents a resource block allocation mechanism in two levels, supported by control theory and latency prioritization, intended to better accommodate both types of traffic and obtaining better results for those cited metrics in comparison to other solutions commonly available. For instance, in a scenario where a proportion between real time traffic and non real time traffic assume values of 75% and 25%, respectively, a 46 times increase in throughput has been achieved in comparison to Frame Level scheduler, and about 75% increase in relation to Exponential Rule scheduler, for non real time traffics, whilst maintaning similar metrics to other schedulers for real time.

Keywords: Mobile networks, Resource block scheduling, Long Term Evolution 4G

Sumário

1	Introdução	1
2	Fundamentação Teórica	4
2.1	Redes móveis de celulares	4
2.2	<i>Long Term Evolution</i> (LTE)	5
2.2.1	Disposição da rede	6
2.2.2	Pilha de rede	7
2.2.3	Tráfego	9
2.2.4	Blocos de recurso	11
2.3	Escalonadores	12
2.3.1	Políticas de alocação de recursos em redes LTE	15
2.3.2	Métricas de avaliação	21
2.4	Simuladores	22
2.4.1	LTE-Sim	23
2.5	Considerações finais	29
3	Revisão do Estado da Arte	31
3.1	Políticas de alocação de blocos de recurso	31
3.2	Discussão	37
4	Escalonadores de blocos de recurso: avaliação comparativa	39
4.1	Ambiente de simulação	40
4.2	Resultados comparativos	44
4.3	Discussão	48
5	Alocação de blocos de recurso sensível à presença de fluxos NTR	50
5.1	Proposta de alocação sensível à heterogeneidade dos fluxos de dados	50
5.2	Resultados	53
5.2.1	Ambiente estático	54
5.2.2	Ambiente com mobilidade	67

5.3 Discussão	69
6 Conclusão	71
6.1 Considerações finais	71
6.2 Trabalhos Futuros	72
Referências	73

Lista de Figuras

2.1	Evolução das redes móveis.	5
2.2	Disposição 4G LTE.	6
2.3	Pilha plano de usuário.	8
2.4	Pilha plano de controle.	8
2.5	Portadoras LTE.	10
2.6	Transmissão OFDM.	12
2.7	Classificação Padrão Escalonadores LTE.	13
2.8	Fluxo genérico de um escalonador.	17
2.9	Funcionamento escalonador <i>Frame Level</i>	19
2.10	Exemplo de alocação FLS.	21
2.11	Direção aleatória.	25
2.12	Passeio aleatório.	26
2.13	LTE-SIM pilha do plano de usuário.	28
4.1	Vazão para Vídeo/ <i>Web</i>	45
4.2	Latência para Vídeo/ <i>Web</i>	46
4.3	Taxa de perda de pacotes para Vídeo/ <i>Web</i>	47
4.4	Índice de justiça para Vídeo/ <i>Web</i>	47
5.1	Fluxograma da proposta para política de alocação de recursos.	51
5.2	Vazão para <i>Web</i>	55
5.3	Latência e <i>jitter</i> para <i>Web</i>	55
5.4	Índice de justiça e perda de pacotes para <i>Web</i>	56
5.5	Vazão para Vídeo/ <i>Web</i>	57
5.6	Latência e <i>jitter</i> para Vídeo.	57
5.7	Latência e <i>jitter</i> para <i>Web</i>	58
5.8	Índice de justiça e taxa de perda de pacotes para Vídeo.	59
5.9	Índice de justiça e taxa de perda de pacotes para <i>Web</i>	59
5.10	Vazão para Vídeo/ <i>Web</i>	60
5.11	Latência e <i>jitter</i> para Vídeo.	61

5.12	Latência e <i>jitter</i> para <i>Web</i>	61
5.13	Índice de justiça e taxa de perda de pacotes para Vídeo.	62
5.14	Índice de justiça e taxa de perda de pacotes para <i>Web</i>	62
5.15	Vazão para Vídeo/ <i>Web</i>	63
5.16	Latência e <i>jitter</i> para Vídeo.	64
5.17	Latência e <i>jitter</i> para <i>Web</i>	64
5.18	Índice de justiça e taxa de perda de pacotes para Vídeo.	65
5.19	Índice de justiça e taxa de perda de pacotes para <i>Web</i>	65
5.20	Vazão para Vídeo.	66
5.21	Latência e <i>jitter</i> para Vídeo.	67
5.22	Índice de justiça e taxa de perda de pacotes para Vídeo.	67
5.23	Vazão para Vídeo/ <i>Web</i>	68
5.24	Índice de justiça e taxa de perda de pacotes para <i>Web</i>	69

Lista de Tabelas

2.1	Classes de Qualidade de serviço.	9
2.2	Quantidade de blocos de recurso por largura de banda.	11
2.3	Indicadores de qualidade do canal.	14
2.4	Resumo de variáveis comuns às políticas de alocação.	17
3.1	Comparativo estado da arte.	37
4.1	Parâmetros do cenário.	42
4.2	Atendimento de usuários para tráfego <i>Web</i>	48
5.1	Parâmetros do cenário.	54
5.2	Atendimento de usuários tráfego <i>Web</i>	56
5.3	Atendimento de usuários tráfego <i>Web</i>	60
5.4	Atendimento de usuários tráfego <i>Web</i>	63
5.5	Atendimento de usuários tráfego <i>Web</i>	66
5.6	Atendimento de usuários tráfego <i>web</i> (Cenário 4 (75% TR - 25% NTR) com mobilidade).	69

Lista de Abreviaturas e Siglas

3GPP *3rd Generation Partnership Project.*

ACK *Acknowledged Signal Message.*

AMC *Adaptive Modulation and Coding.*

BCQI *Best CQI.*

BET *Blind Equal Throughput.*

BIBO *Bound Input Bound Output.*

BLER *Block Error Rate.*

CBR *Constant Bit Rate.*

CDMA *Code Division Multiple Access.*

CIF *Common Intermediate Format.*

CQI *Channel Quality Indicator.*

DCI *Downlink Control Information.*

DP *Delay Priority.*

E-RAB *EPS Radio Access Bearer.*

E-UTRAN *Evolved Universal Terrestrial Radio Access Network.*

EBCQA *Earliest Deadline First Queue Aware Channel Based.*

EDGE *Enhanced Data GSM Evolution.*

eNB *Evolved NodeB.*

EPC *Evolved Packet Core.*

EXP *Exponential PF.*

EXPR *Exponential rule.*

FIFO *First in First Out.*

FLS *Frame Level Scheduler.*

FTP *File Transfer Protocol.*

GBR *Guaranteed Bit Rate.*

GPF *Generalized PF.*

GPRS *General Packet Radio Service.*

GSM *Global System for Mobile communications.*

HARQ *Hybrid automatic repeat request.*

HSS *Home Subscriber Service.*

HTTP *Hypertext Transfer Protocol.*

IdC *Internet das Coisas.*

IP *Internet Protocol.*

IWFQ *Inverse Weighted Fair Queuing.*

LBCQA *Logarithmic Queue Aware Channel Based.*

LOGR *Logarithmic rule.*

LTE *Long Term Evolution.*

LTE-Sim *LTE-Simulator.*

MAC *Media Access Control.*

MCS *Modulation Coding Scheme.*

MLWDF *Modified largest weighted delay first.*

MME *Mobility Management Entity.*

MMS *Multimedia Message Services.*

MOS *Mean Opnion Score.*

ms *microsegundos.*

MT *Maximum Throughput.*

NACK *Not Acknowledged Signal Message.*

NAS *Non-Access Stratum.*

NS3 *Network Simulator 3.*

NTR *não tempo real.*

OFDM *Orthogonal Frequency Division Multiple Access.*

P-GW *Packet Data Network Gateway.*

PDCCH *Physical Downlink Control Channel.*

PDCP *Packet Data Convergence Protocol.*

PDSCH *Physical Downlink Shared Channel.*

PF *Proportional and fair.*

PIWFQ *Proportional Inverse Weighted Fair Queuing.*

PSNR *Peak Signal to Noise Ratio.*

PUSCH *Physical Uplink Shared Channel.*

QCI *QoS Class Identifier.*

QH-MLWDF *Queue Head of line MLWDF.*

QoS *Qualidade de Serviço.*

RAN *Radio Access Network.*

RB *Resource Block.*

REST *Representational State Transfer.*

RLC *Radio Link Control.*

RR *Round Robin.*

S-GW *Serving Gateway.*

SIM *Subscriber Identification Module.*

SMS *Short Message Service.*

SNR *Signal to Noise Ratio.*

TR *tempo real.*

TTI *Transmission Time Interval.*

UE *User Equipment.*

VoIP *Voice Over IP.*

VT-MLWDF *Virtual Token MLWDF.*

WEB *World Wide Web.*

Capítulo 1

Introdução

Nas redes 4G, os mecanismos fundamentais compreendem os *gateways* e as *Evolved Node B* (eNB) [1]. O primeiro, *gateways*, trabalham a parte de autorização da operadora, de tributação do tráfego, de conexão com a Internet, bem como outras funcionalidades. Já, o segundo, as torres base, permitem a comunicação entre os usuários e o resto da arquitetura 4G, com a função de estabelecer a comunicação com outras torres para transferências de usuários (do Inglês, *handover*), promove também a acomodação de pacotes atrasados por meio de *buffer* [2]. Por fim, realizam uma funcionalidade muito importante, a distribuição e alocação dos recursos de rádio, cuja ação consiste em estabelecer qual usuário transmitirá ou receberá dados a cada intervalo de transmissão [3].

Grande parte do volume dos dados presentes em redes 4G recaem sobre os fluxos de dados considerados de tempo real (TR), tais como vídeo, *Voice Over IP* (VoIP) ou serviços de *streaming* [4]. Porém, o rápido crescimento no quantitativo de dispositivos para a Internet das Coisas (IdC) traz consigo uma quantidade muito maior que a atual para fluxos com classificação não tempo real (NTR), como requisições *Representational State Transfer* (REST), transferências de arquivos etc., que não se encontram priorizadas nas atuais políticas de alocação de blocos de recurso, fazendo com que possam ocorrer gargalos e maiores taxas de perda de pacotes, tornando alguns mecanismos ineficientes. Os diferentes requisitos de qualidade de serviço nos tráfegos das redes LTE representam a heterogeneidade dos fluxos de dados, grande parte dos escalonadores encontrados na literatura possuem priorização ao tráfego de tempo real, devido aos requisitos de serviço que normalmente encontram-se mais restritos, por meio de técnicas de priorização por latência. Entretanto, para os fluxos de dados NTR, não ocorre nenhum tipo de priorização por meio das características de qualidade de serviço, e além disso há um grande crescimento no quantitativo de dispositivos que irão trazer consigo maior número de requisições a outra classificação, o que pode acarretar um certo nível de negligência dos fluxos. Portanto, o objetivo principal dessa pesquisa encontra-se alinhado à questão da

melhora da priorização dos fluxos de dados, não necessariamente priorizando o tráfego não tempo real, porém utilizando da latência e do tempo máximo de espera, presente nos requisitos de *Qualidade de Serviço* (QoS) para melhor alocar os blocos de recurso, desta forma, dar-se-á oportunidade para os tráfegos que já estão próximos ao limite da qualidade de serviço, o que pode contemplar de melhor forma os tráfegos NTR.

Dado o objetivo principal de desenvolver um modelo de alocação de blocos de recurso que visa melhor priorização de tráfegos por meio da latência, objetivos específicos foram traçados para possibilitar o seu cumprimento. Dentre eles, encontram-se: identificação dos fundamentos teóricos para as políticas de alocação de blocos de recursos, verificação dos benefícios para a alocação e limitações das estratégias mais abordadas na literatura, buscando mecanismos alternativos para melhor equiparação de prioridade entre as classificações de tráfego TR e NTR, definição do ambiente de simulação para a solução proposta e, por fim, desenvolvimento da proposta em um simulador de redes LTE, determinando o desempenho quando comparado aos mecanismos encontrados no estado da arte.

A metodologia aplicada neste trabalho consistiu da revisão do estado da arte, abrangendo a arquitetura da rede LTE 4G e, especialmente, trabalhos com relação à alocação de blocos de recurso. Abordou-se os benefícios e limitações dos mecanismos, no que tange os objetivos traçados, com base nas informações e avaliações dos escalonadores comumente presentes na literatura. Buscou-se um mecanismo alternativo que possibilitasse a equiparação de prioridade quanto às classificações de tráfegos, TR e NTR, a fim de melhor acomodar fluxos de dados com a segunda classificação, considerando a possibilidade de um rápido crescimento do fluxo devido aos dispositivos IdC. Para averiguar o mecanismo proposto, o simulador escolhido para a implementação foi o LTE-Sim, devido ao quantitativo de trabalhos avaliados que o utilizam [5], [6], [7], com grande parte dos parâmetros utilizados na maioria das pesquisas, como configurações de mobilidade do usuário, diferentes tipos de aplicação, opções de largura de banda utilizada entre outros requisitos.

A presente pesquisa abordou uma proposta de política de alocação de blocos de recursos, sensível à heterogeneidade dos fluxos de dados, permitindo uma equiparação quanto à prioridade de qualquer tipo de tráfego. A proposta desenvolvida consistiu na ampliação do estudo de Piro *et al.* [8], quanto aos pontos desfavoráveis em sua política de alocação de blocos de recursos, especialmente quando o tráfego classificado como não tempo real (NTR) torna-se mais requisitado que o tráfego tempo real (TR). O mecanismo proposto explorou um modelo de duas camadas, em que a primeira camada reorganiza a quantidade de dados a serem enviados por meio de teoria de controle, e a segunda realiza a alocação dos blocos de recurso, propriamente dita, a partir de leves modificações nas políticas dos escalonadores *Exponential rule* (EXPR) ou *Logarithmic rule* (LOGR) [9]. Os resultados

obtidos mostraram melhorias para casos em que tráfegos classificados como não tempo real apresentam maior proporção com relação ao tráfego não tempo real. Mantendo, também, níveis aceitáveis e, em certos pontos, atingindo melhor desempenho que os originais para cenários convencionais, em que existe maior tráfego de tempo real. Vale ressaltar que um artigo referente a presente pesquisa já está em fase de preparação para futura publicação em conferências relacionadas ao tema.

A pesquisa encontra-se organizada com a seguinte estrutura: Capítulo 2, fundamentos teóricos, abrangendo a arquitetura 4G, seus dispositivos, tipos de conexão, pilha de rede, escalonadores e mecanismos de alocação. Capítulo 3, compreende a revisão do estado da arte, demonstrando soluções distintas frente ao problema de alocação de blocos de recursos em redes LTE 4G, assim como detalhes a respeito das desvantagens na utilização de algumas das soluções. Capítulo 4, dispõe os resultados de alguns mecanismos da literatura para comparativo a partir de um cenário de simulação estático. Com base nesses resultados, discute-se as motivações para a proposta da pesquisa, por meio de indicativos das métricas, tais como: justiça, vazão, latência e taxa de perda de pacotes. Capítulo 5, retrata uma nova proposta, baseada em priorização por latência assistida de teoria de controle para alocação de blocos de recurso, apoiada no que foi apresentado nos capítulos anteriores, bem como os resultados obtidos por esta estratégia. Por fim, o Capítulo 6, relata as considerações finais e os trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Este capítulo tem como objetivo situar o leitor quanto aos assuntos a serem abordados na presente pesquisa, a fim de possibilitar um entendimento nivelado dos tópicos necessários. Dentre os assuntos presentes nesta seção, destacam-se a arquitetura de redes móveis *Long Term Evolution* (LTE), os tipos de tráfego e o sistema de escalonamento de blocos de recurso que representa o foco da pesquisa.

2.1 Redes móveis de celulares

A telefonia móvel proporcionou um grande avanço na tecnologia de comunicação sem fio. Cada geração define um conjunto de padrões que detalham a implementação das tecnologias utilizadas nas redes de telefones móveis, que, atualmente, caminha para a tecnologia da quinta geração (5G) [10]. Junto ao desenvolvimento da tecnologia ocorre também evolução nas taxas de transferência de dados, melhoria de transferência de dispositivos entre as torres e outras funcionalidades das redes.

A primeira geração obteve maior sucesso na década de 1980, em geral, os dispositivos não dispunham de boa eficiência energética, qualidade de voz e segurança, muitas vezes, os usuários experienciavam o encerramento aleatório das chamadas [11]. Já a segunda geração, passou de sinal analógico para digital, utilizando redes sem fio com *Code Division Multiple Access* (CDMA) e *Global System for Mobile communications* (GSM), provendo novas interfaces de dados como *Short Message Service* (SMS) e *Multimedia Message Services* (MMS), além das chamadas de voz convencionais. A taxa de transferência de dados, na segunda geração, alcançou cerca de 50Kb/s no sistema *General Packet Radio Service* (GPRS), atingindo a casa de 1Mb/s perto do lançamento do 3G por meio de *Enhanced Data GSM Evolution* (EDGE) [12].

A terceira geração, objetivou inicialmente uma conexão com taxas de transferência de 200Kb/s para Internet, além disso, com chamadas mais claras e maior cobertura ge-

ográfica, torna-se possível, por exemplo, efetuar ligações em um veículo a 100km/h, sem perda de conexão com as torres, devido às várias estações base, estrategicamente distribuídas [1]. A quarta geração, que será abordada de maneira mais aprofundada nessa pesquisa, evidencia larguras de banda de 5 a 20MHz, taxas de transmissão vislumbrando até 12,5Mb/s, em sua versão original, e diversos controles de qualidade de serviço, expandindo de Voz, SMS ou MMS, para vídeos, acesso web (*Hypertext Transfer Protocol* (HTTP), *File Transfer Protocol* (FTP)) e jogos. Vale ressaltar que a 5G, quinta geração, já apresenta estudos, porém existem poucos simuladores, emuladores ou *test beds* retratando esse novo formato, além disso muito da quinta geração está atrelado a melhorias do LTE. A Figura 2.1, baseada em [13], resume a evolução das redes e suas principais contribuições, indo ao encontro do exposto nos parágrafos anteriores.

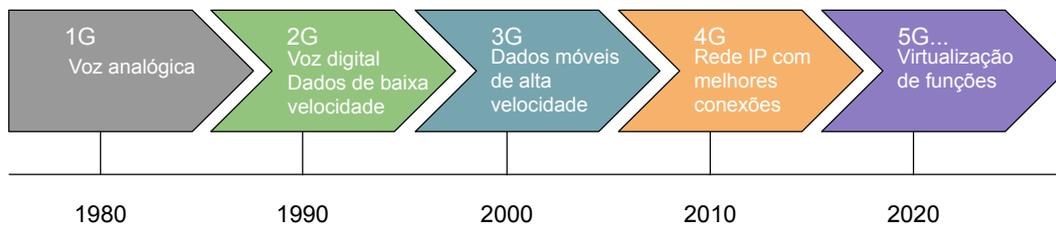


Figura 2.1: Evolução das redes móveis.

2.2 *Long Term Evolution* (LTE)

A padronização do LTE pela *3rd Generation Partnership Project* (3GPP), uma empresa organizadora de protocolos para telefonia móvel, foi alcançada em busca de maior desempenho para redes móveis [3]. O LTE constitui um sistema que utiliza *Orthogonal Frequency Division Multiple Access* (OFDM) para o espectro de banda larga, técnica de modulação que alcança maiores taxas de transmissão que as tecnologias de rádios anteriores. Ao contrário das redes 3G, pensadas com foco principal em dados de voz, as redes 4G foram tratadas com requisitos ambiciosos, tais como: aumento da eficiência espectral para praticamente o dobro, ampliação da cobertura da rede, melhoria de taxa de transmissão para usuários nas bordas da célula. Essa geração oportunizou taxas de até 100Mb/s para *downlink* e 50Mb/s para *uplink*, na últimas versões, e mantendo-se estável durante a mobilidade do usuário. O suporte a requisitos de qualidade de serviço, por meio de técnicas sofisticadas de gerenciamento de recursos de rádio [14], constitui uma das principais melhorias propostas no novo padrão LTE 4G.

2.2.1 Disposição da rede

A arquitetura LTE 3GPP inclui alguns módulos obrigatórios, o núcleo de rede, o *Evolved Packet Core* (EPC) e a interface aérea *Evolved Universal Terrestrial Radio Access Network* (E-UTRAN) [12]. O EPC pode ser subdividido em outras entidades: *Mobility Management Entity* (MME), *Serving Gateway* (S-GW), *Packet Data Network Gateway* (P-GW) e *Home Subscriber Service* (HSS). Há uma divisão em dois planos lógicos, usuário e controle. O MME faz parte do plano de controle, responsável por gerenciar os sinais e as conexões com a rede de acesso a rádio (do Inglês, *Radio Access Network* (RAN)) [2], enquanto o S-GW representa a base do plano do usuário, responsável por enviar e receber os pacotes de dados propriamente ditos. Já P-GW, constitui um ponto de terminação, conectando a rede LTE com o mundo. Por fim, o HSS gerencia autenticação e cobrança dos usuários inscritos na rede. Cada entidade possui sua interface de conexão, tais como S1 e X2, que serão explanadas no decorrer desta seção. A Figura 2.2, baseada em [14], ilustra uma possível disposição da rede junto aos dispositivos que a compõem.

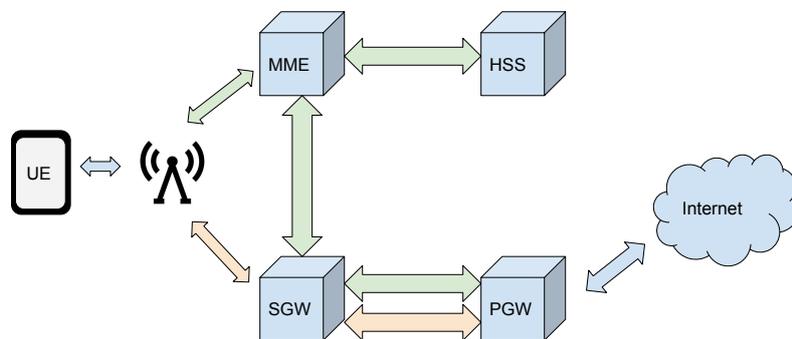


Figura 2.2: Disposição 4G LTE.

Equipamento de usuário

Refere-se a cliente todo dispositivo que pode se conectar à rede de Equipamento de usuário (do Inglês, *User Equipment* (UE)). A conexão fica registrada por meio de um cartão *Subscriber Identification Module* (SIM) de uma operadora de rede móvel [1] e os usuários na rede 4G podem ser cobrados pela utilização dos serviços providos.

Estação base

No padrão 4G LTE, as estações base recebem o nome de *Evolved NodeB* (eNB), esta entidade atua como intermediária da conexão dos UE para o S-GW. Por meio da interface LTE-Uu, o rádio UE comunica-se com as torres próximas, e estas, com o S-GW por meio da interface S1. As eNB podem também se comunicar com outras torres para controle ou transferência de célula de um usuário, tal comunicação recai sobre a interface X2. Devido a interlocução de duas entidades que utilizam tecnologias de comunicação diferentes, as torres possuem duas pilhas de rede [2], tópico a ser detalhado na Seção 2.2.2.

Mobility Management Entity

A responsabilidade desta entidade gestora de mobilidade consiste no controle de sinais entre o UE e o núcleo da rede a partir do protocolo *Non-Access Stratum* (NAS), o que inclui gerenciamento da portadora de dados e da conexão, paginação e autenticação de UE, entre outros procedimentos.

Serving Gateway

O S-GW atua como intermediador entre o UE e o P-GW, coletando informações para cobrança, exemplo *download* e *upload*, realizando *buffering* de dados, ou seja, armazenando temporariamente os dados, enquanto o MME gerencia a portadora e outras funcionalidades. Atua também como gerenciador de mobilidade para tecnologias legadas 2G e 3G [12].

Packet Data Network Gateway (P-GW)

O P-GW realiza a alocação de endereços IP para os UE, efetua cumprimento das qualidades de serviço (do Inglês, *Qualidade de Serviço* (QoS)) e da taxa de dados necessária para tal, aplica cobranças baseada no tráfego de dados e filtra as informações em diferentes padrões, como vídeo, voz, web etc. Esta entidade também pode servir como gerenciador de mobilidade para tecnologias não 3GPP, como WiMAX [15].

2.2.2 Pilha de rede

Como citado previamente, a arquitetura LTE possui dois tipos de planos, usuário e controle, cada um com duas pilhas de rede, uma para o rádio (utilizada pelos equipamentos de usuário (UE) e estações base (eNB)) e outra para conexões a cabo (estações base para *gateways*). O plano de controle contém a parte de configuração e de gerenciamento da

rede, tal como autenticação, cobrança e conexão dos usuários. Enquanto o plano de usuários, também chamado de plano de dados, comprime e transmite o tráfego dos usuários [16]. As Figuras 2.3 e 2.4, baseadas em [14], ilustram a pilha de rede para ambos os planos, incluindo as interfaces de conexão. O pacote *Internet Protocol* (IP) encapsulado em

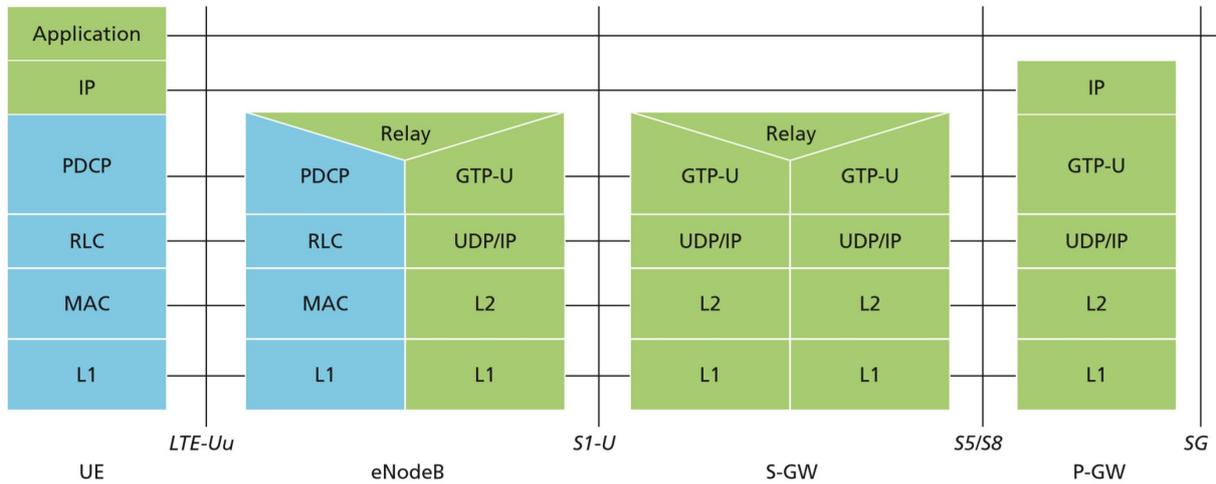


Figura 2.3: Pilha plano de usuário.

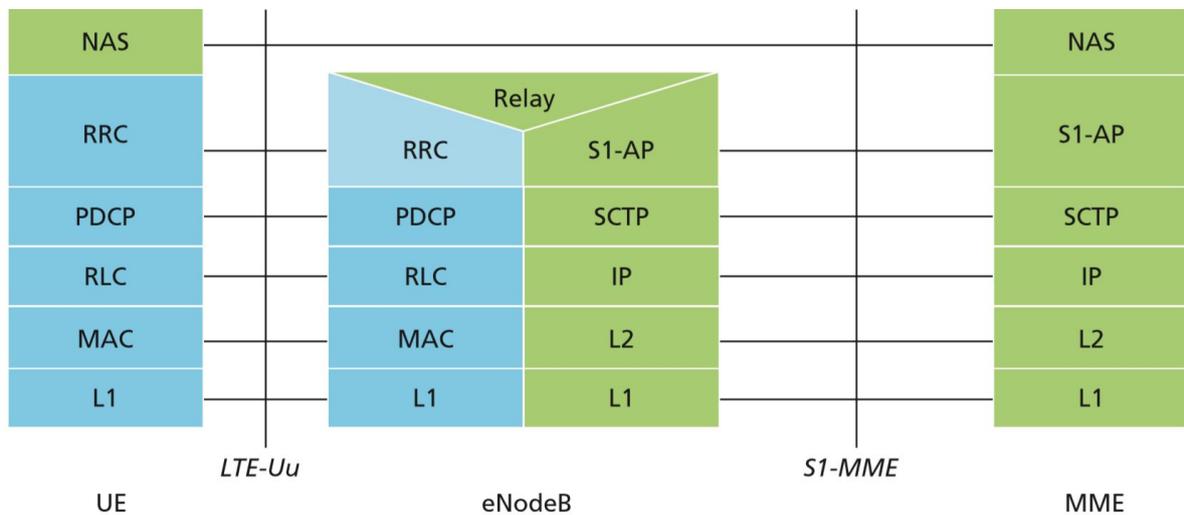


Figura 2.4: Pilha plano de controle.

um protocolo pelo EPC será transmitido entre o P-GW e a eNB, utilizando do protocolo de tunelamento *General Packet Radio Service* (GPRS), chamado de portadora no padrão LTE [1], para enviar dados na interface S1, S5/S8 e CN. Outros protocolos podem ser utilizados, porém a responsabilidade de compatibilidade fica a cargo da operadora da rede móvel. No segmento E-UTRAN, existem as camadas *Packet Data Convergence Protocol* (PDCP), *Radio Link Control* (RLC) e *Media Access Control* (MAC).

PDCP encontra-se na camada de topo da pilha, esse protocolo processa mensagens do plano de controle e pacotes IP no plano de usuário. As funções principais do PDCP incluem compressão, segurança (integridade e criptografia), reordenação de pacotes e retransmissão durante transferência de EB [14]. O RLC objetiva a segmentação e a reconstrução dos pacotes da camada de cima, isso deve ser realizado para adaptar o tamanho do pacote a ser transmitido na interface de rádio. Por fim, a camada MAC se encarrega de gerenciar o escalonamento de blocos de recurso e (de)multiplexação.

2.2.3 Tráfego

O tráfego nas redes 4G LTE pode ser identificado por meio da classe de identificador de qualidade de serviço (do Inglês, *QoS Class Identifier* (QCI)), dividida em 8 prioridades, a fim de melhor diferenciar os tipos de pacotes transmitidos na rede [17]. Para cada identificador existe um correspondente valor máximo de prioridade, latência e taxa de perda de pacotes definidos pela 3GPP.

O valor de qualidade de serviço ocorre por meio de um indicador da classe de qualidade, caracterizada por um tipo base *Guaranteed Bit Rate* (GBR) ou não-GBR, um nível de prioridade, a latência máxima de atraso e uma taxa aceitável de perda de pacotes [18]. A Tabela 2.1 indica as classificações de qualidade de serviço, em que *QCI* se refere à classe da qualidade, *P* à prioridade, *LM* à latência média esperada, *TPE* à taxa de perda de pacotes esperada e *TB* ao tipo base.

Tabela 2.1: Classes de Qualidade de serviço.

QCI	TB	P	LM (ms)	TPE	Serviço
1	GBR	2	100	10^{-2}	Voz de conversação
2	GBR	4	150	10^{-3}	Vídeo de conversação (stream ao vivo)
3	GBR	5	300	10^{-6}	Vídeo (stream com buffer)
4	GBR	3	50	10^{-3}	Jogos em tempo real
5	NGBR	1	100	10^{-6}	Sinais IMS
6	NGBR	7	100	10^{-3}	Voz, Vídeo (stream ao vivo), jogos interativos
7	NGBR	6	300	10^{-6}	Vídeo (stream com buffer)
8	NGBR	8	300	10^{-6}	TCP (WWW, e-mail)

De acordo com a empresa Ericsson [19], o número de usuários LTE crescerá para 9 bilhões por volta de 2022, o que significa dizer que mesmo com o advento da tecnologia 5G, 4G ainda será muito explorada por diversas áreas. Além disso, espera-se um aumento do tráfego em torno de 50% por ano até 2022. Dentre os tipos de tráfego, acredita-se que Vídeo cubra 75% de toda transferência da rede, sendo os outros 25% divididos entre áudio, *web*, Redes Sociais e compartilhamento de arquivos. Ademais, existem planos para

atualizar pontos de rede 2G para 3G/4G, o que estimula a comunidade acadêmica e a empresarial a financiar ou manter pesquisas nessas tecnologias.

Portadoras

Conexões virtuais que realizam o transporte dos dados e das informações de qualidade do serviço entre as camadas da rede, efetuando serviço fim a fim do cliente à Internet. Entre os dispositivos de rádio encontra-se a portadora de rádio, entre a torre e o S-GW, a portadora da interface S1, juntas estas duas formam a *EPS Radio Access Bearer* (E-RAB).

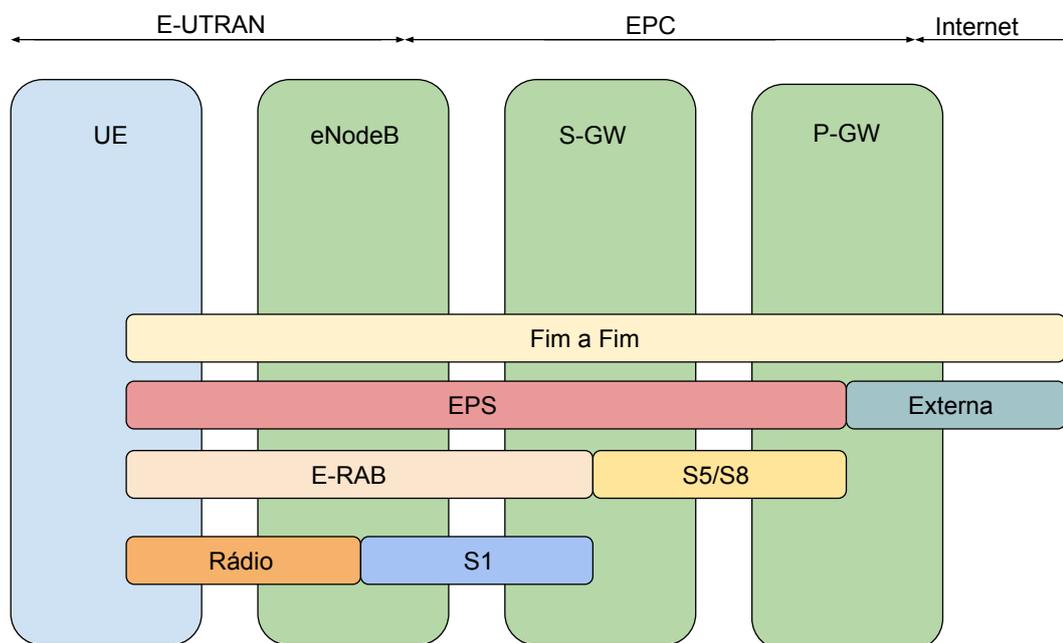


Figura 2.5: Portadoras LTE.

Concatenando a portadora da interface S5, resultante da junção do S-GW ao P-GW, tem-se a EPS, que liga o UE ao P-GW. Por fim, a portadora externa agrega o P-GW à Internet, fazendo uma conexão fim a fim [12]. A Figura 2.5 ilustra cada fase e seu posicionamento no fluxo de dados.

Um conjunto de parâmetros de qualidade de serviço permanece associado à portadora a depender dos dados da aplicação a qual ela pertence, desta forma, pode-se diferenciar os fluxos de dados [16]. Esta associação de uma portadora EPS para cada tráfego ao

se conectar com o P-GW possui um valor de QoS baixo. Porém, após a conexão, uma portadora dedicada pode ser instanciada a partir dos valores específicos do tráfego, os serviços de qualidade da conexão variam de acordo com a operadora.

2.2.4 Blocos de recurso

O sistema LTE foi projetado de maneira flexível para comportar diversas configurações de bandas, entre 1,4MHz a 20MHz. O acesso ao espectro emprega a técnica *Orthogonal Frequency Division Multiple Access* (OFDM) [2]. A 3GPP define o número de blocos associados de acordo com a largura de banda [17], o que impacta a performance do LTE, pois determina o número de recursos compartilhados pelos usuários, como ilustra a Tabela 2.2.

Tabela 2.2: Quantidade de blocos de recurso por largura de banda.

Banda do canal (MHz)	1,4	3	5	10	15	20
Blocos de Recurso	6	15	25	50	75	100
Sub-bandas ocupadas	72	180	300	600	900	1200
Taxa de amostragem (MHz)	1,92	3,84	7,68	15,36	23,04	30,72
Amostras por compartimentos	960	1920	3840	7680	11520	15360

Os recursos do espectro podem ser alocados no domínio do tempo ou da frequência. No tempo, distribuem-se a cada intervalo de transmissão (do Inglês, *Transmission Time Interval* (TTI)), sendo 1ms a duração de cada um. Cada TTI pode ser subdividido em 2 compartimentos de 0,5ms, cada um contendo 7 símbolos OFDM [16]. Já no domínio da frequência, divide-se a banda em subcanais de 180kHz, possuindo 12 subportadoras espaçadas em 15kHz. A Figura 2.6 ilustra a técnica, na qual um ponto k tem-se uma das subportadoras transmitindo e as outras, neste determinado ponto, com interferência próxima de zero (0).

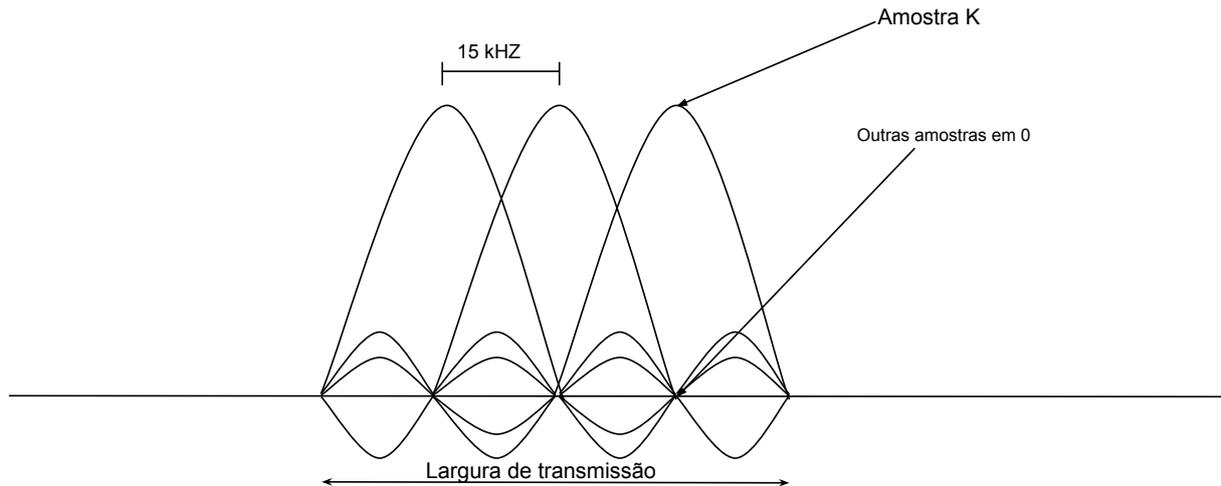


Figura 2.6: Transmissão OFDM.

Aos recursos que se estendem por meio TTI, denominam-se de bloco de recurso (do Inglês, *Resource Block* (RB)) correspondendo à menor unidade de recurso que pode ser alocada para um dispositivo de usuário. Com o tamanho do subcanal fixado obtém-se um número dinâmico de blocos de recurso, de acordo com a configuração da banda do sistema. Vale ressaltar que o uso dos blocos de recursos deve ocorrer com base em tráfegos que possuem diferentes características, de modo a atender os distintos usuários que acessam a rede. A seção subsequente aborda algumas das estratégias mais recorrentes na literatura.

2.3 Escalonadores

Em uma extensa revisão da literatura, Piro *et al.* [3] apresentam os principais aspectos e desafios de se construir um escalonador de blocos de recurso para *downlink*. O artigo descreve as principais técnicas que, em grande parte, seguem em uso ou serviram de base para o desenvolvimento de novas políticas de alocação de recursos. Na revisão, os escalonadores foram divididos em estratégias: sem consciência da qualidade do canal e sem qualidade do serviço, com consciência da qualidade do canal e sem da qualidade do serviço e consciência de ambos, como ilustra a Figura 2.7, com base no trabalho de Piro *et al.* [3]. Podemos ainda considerar a consciência energética do escalonador, que possui como objetivo minimizar os gastos energéticos do sistema ou de alguma classe de dispositivo, podendo ser englobada às outras. Para Piro *et al.*, o projeto de um escalonador deve levar em conta diferentes necessidades, logo, a política de escalonamento deve contemplar e/ou focar em certos objetivos, tais como: equidade, taxa de transmissão, latência (tempo de resposta), número de usuários atendidos, balanceamento da utilização de recursos entre outros.

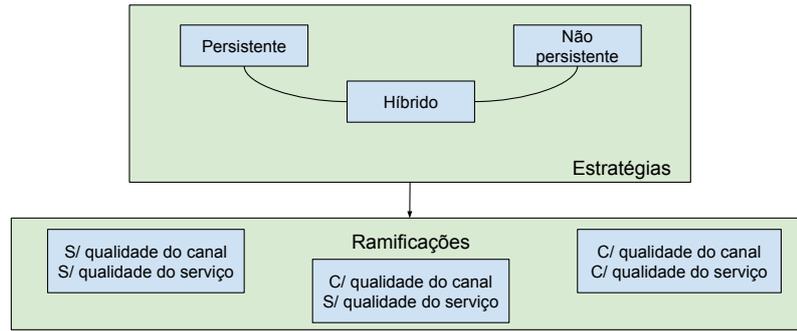


Figura 2.7: Classificação Padrão Escalonadores LTE.

No LTE, em cada TTI, o escalonador computa uma métrica, que depende do algoritmo de escalonamento utilizado, para cada tráfego a ser enviado. O escalonador funciona atribuindo um bloco de recurso j ao tráfego com maior métrica w [20], podendo ser sumarizado como apresentando a seguir:

- Torre cria a lista de tráfego *downlink*, ou seja, a lista que pode ser escalonada no subquadro corrente;
- Para cada tráfego, o tamanho da fila na camada MAC, os parâmetros QoS e as respostas de qualidade de canal se mantêm;
- De acordo com o algoritmo escolhido, computa-se a métrica w para cada tráfego;
- A torre atribui os subcanais aos tráfegos com as maiores métricas, guardando um total de dados enviados por cada tráfego, de modo que ao chegar em seu tamanho final ele seja eliminado da lista;
- Para cada tráfego, a torre calcula o tamanho do bloco de transporte, o quanto será transmitido durante o TTI corrente. Em mais detalhes, o módulo *Adaptive Modulation and Coding* (AMC) mapeia a qualidade do canal para uma codificação de modulação, que pode ser utilizada na obtenção do tamanho do bloco de transporte.

Para obter a métrica, algoritmos normalmente utilizam a taxa de transmissão média \bar{R}_i do tráfego i , e também a taxa de transmissão instantânea do receptor no subcanal j . Essas informações se mostram úteis para garantir justiça por meio da performance obtida no último TTI. Utilizando a métrica por meio da taxa de transmissão obtém-se uma equidade relacionada a esta taxa, porém, pode-se utilizar uma métrica com o objetivo de otimizar o gasto energético, o tempo utilizado por usuário, entre outros fatores. Nota-se então um equilíbrio entre os diferentes métodos de obtenção da métrica, não sendo possível otimizar todos ao mesmo tempo.

Cálculos por métricas diferentes podem ser classificados entre os seguintes tipos de escalonador: persistente, não persistente e semipersistente. Esses modos de operação podem ser ramificados entre as subestratégias, conforme apontado no artigo de Piro *et al.* [3]. O escalonador persistente atua enviando informações de controle para N TTIs, iniciando com ao menos 3, desta forma, necessita-se de menos dados de controle para os usuários saberem quando poderão baixar informações. Por outro lado, esse modo de operação não fornece suporte para mudanças dinâmicas de qualidade de canal, pois o recurso alocado não pode ser trocado. Da mesma maneira, não dá suporte para *Hybrid automatic repeat request* (HARQ) [21], técnica utilizada para detecção de erros no pacote que objetiva a retransmissão em casos de erro no recebimento, uma vez que requisições não recebidas ou não enviadas pelo usuário serão reportadas após o período de persistência.

O modo não persistente, em geral, mostra-se mais utilizado na literatura, devido aos problemas expostos acima. Por fim, o modo semipersistente busca um ponto médio entre os dois anteriores, utilizando o mesmo bloco de recurso para um tráfego em que se sabe que ocorrerá uma transmissão contínua, como por exemplo para dados de voz, enquanto que outros tipos de tráfego se mantêm não persistentes.

Tabela 2.3: Indicadores de qualidade do canal.

Índice CQI	Modulação	Eficiência
0	N/A	N/A
1	QPSK	0,1523
2	QPSK	0,2344
3	QPSK	0,3770
4	QPSK	0,6016
5	QPSK	0,8770
6	QPSK	1,1580
7	16QAM	1,4766
8	16QAM	1,9141
9	16QAM	2,4063
10	64QAM	2,7305
11	64QAM	3,3223
12	64QAM	3,9023
13	64QAM	4,5234
14	64QAM	5,1152
15	64QAM	5,5547

A qualidade do canal referenciado nos últimos tópicos especifica o nível da qualidade percebida por cada usuário, sendo que o 4G LTE segue o padrão estabelecido pela 3GPP. O *Channel Quality Indicator* (CQI) constitui uma informação codificada, enviada, periodicamente, pelo UE à estação base, a cada 5 intervalos de transmissão. O cálculo dessa informação ocorre a partir do fator sinal ruído (do Inglês, *Signal to Noise Ratio* (SNR)),

cada nível de qualidade se relaciona a uma modulação esperada, a fim de não passar do limite da taxa de erro por bloco (do Inglês, *Block Error Rate* (BLER)). A Tabela 2.3 indica modulações, taxa de codificação e eficiência para cada nível. Um fator de sinal ruído alto representa aumento na qualidade do canal, o que garante um serviço com maiores taxas de transmissão para os usuários. Já usuários com condições ruins de canal, em geral nas bordas de uma célula, possuem conexões ativas com baixa taxa de transmissão. A taxa de codificação possui um limite máximo, assim, melhorias no fator sinal ruído após o último nível não garantem ganhos ao usuário. Esta limitação existe para obter melhores consumos energéticos, enquanto se mantêm taxas de transmissão constantes.

Por fim, no canal físico, reserva-se uma parte do quadro de envio de dados para controle, dentro deste encontra-se o *Physical Downlink Shared Channel* (PDSCH). Este canal serve para compartilhar informação com todos os usuários da rede definindo como os recursos de rádio podem ser utilizados pelas camadas superiores. Dados de controle sobre *downlink* ocorrem, principalmente, por meio do *Physical Downlink Control Channel* (PDCCH), que carrega em si as autorizações de uso de recursos de *downlink*. A parte de *uplink* não será explorada por esta pesquisa, contudo, vale ressaltar que os canais de *uplink* também existem, e quando não há transmissão de dados ocorrendo nesses veículos, eles podem ser usados para sinalização de transmissão de *downlink* (*Acknowledged Signal Message* (ACK)/*Not Acknowledged Signal Message* (NACK)), ou para reportar CQI ou ainda para requisição de transmissão.

2.3.1 Políticas de alocação de recursos em redes LTE

Os escalonadores apresentados nessa seção fazem parte do conjunto não persistente e possuem, também, grande número de referências na literatura, como [22], [23], [8], [5] e [24]. A Tabela 2.4 apresenta as diferentes variáveis dos escalonadores e suas definições de maneira unificada, indicando o que cada uma significa com relação às métricas observadas pelas políticas de alocação de blocos de recursos. Como já citado anteriormente, os escalonadores em sistemas LTE computam uma métrica w com o objetivo de alocar um bloco de recurso ao usuário com maior métrica, satisfazendo

$$m_{j,k} = \max m_{i,k}, \quad (2.1)$$

em que j representa um usuário da rede e k um bloco de recurso. As métricas podem ser calculadas utilizando-se de informações relacionadas a cada tráfego, para auxiliar a decisão de alocação, têm-se:

Qualidade do canal - Valores de CQI podem ser utilizados para atender usuários que

experienciam melhor qualidade do canal, aumentando a taxa de transmissão esperada para o usuário;

Estado do *buffer* - Enviar a informação do *buffer* de recebimento junto ao CQI para buscar amenizar condições de estouro de *buffer*, usuários com mais espaço livre podem possuir maior prioridade;

Histórico de alocação - Informação sobre desempenho anterior pode ser utilizada para aumentar a equidade entre os usuários, desta forma usuários com baixa vazão ou baixo atendimento de pacotes podem ter mais prioridade;

Requisitos de serviço - As classificações de tráfego possuem diferentes valores de latência total, tempo de espera em *buffer* e outros requisitos que podem ser utilizados para garantir diferentes níveis de qualidade de serviço;

Estado da fila de transmissão - Utilizado para minimizar a latência da entrega de pacotes, caso a fila esteja muito grande uma maior prioridade será dada.

Em cada TTI, o escalonador faz uma decisão de alocação, enviando informações aos usuários por meio do PDCCH. Os UEs recebem informações contidas em mensagens de *Downlink Control Information* (DCI) para saber quais blocos de recursos podem ser utilizados na transmissão de dados no PDSCH, recebem também recursos dedicados no *Physical Uplink Shared Channel* (PUSCH), e devolvem a informação de CQI para a torre. A torre aplica as informações que o usuário enviou para definir os blocos de recurso que serão utilizados e quem os utilizarão. A técnica de AMC seleciona a configuração de transmissão, e a torre volta a informar aos usuários sobre as novas informações no PDCCH [2]. A Figura 2.8 exemplifica o fluxo descrito acima para o funcionamento genérico de um escalonador de pacotes no ambiente de redes LTE, utilizando informações de qualidade de canal para auxiliar a alocação de blocos de recurso.

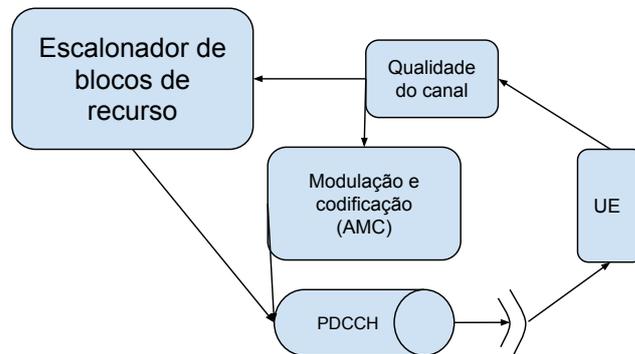


Figura 2.8: Fluxo genérico de um escalonador.

Tabela 2.4: Resumo de variáveis comuns às políticas de alocação.

Variável	Definição
$m_{i,k}$	Métrica sobre a priorização do i -ésimo pacote no k -ésimo RB
$r_{i,k}$	Taxa de transmissão instantânea do i -ésimo pacote no k -ésimo RB
\bar{R}_i	Taxa de transmissão média do i -ésimo tráfego
$D_{HOL,i}$	Tempo de espera no <i>buffer</i>
τ_i	Limite de latência do i -ésimo pacote
δ_i	Limite de perda de pacote para o i -ésimo tráfego
Γ_k^i	Eficiência espectral do i -ésimo pacote no k -ésimo RB

Vazão máxima - O escalonador de vazão máxima (do Inglês, *Maximum Throughput* (MT)) [25], se mostra um dos mais simples explicitados. Utiliza somente da condição do canal para encontrar a maior taxa de transmissão instantânea ($r_{i,j}$) e a usa para atribuir os recursos de modo a priorizar os que possuem maior taxa. Não busca equidade ou qualquer forma de controle com a qualidade de serviço.

Proportional and fair (PF) - O escalonador proporcional e justo [26] atribui os recursos levando em consideração a qualidade do canal experienciado pelo usuário, assim como a taxa de transmissão média anterior. Tem como objetivo maximizar a taxa de

transmissão do sistema, enquanto garante equidade entre os tráfegos. Para tal, define-se a métrica w por

$$w_{i,j} = \frac{r_{i,j}}{\bar{R}_i}, \quad (2.2)$$

em que $r_{i,j}$ refere-se à taxa de transmissão instantânea e \bar{R}_i à média da taxa de transmissão, sendo i o i -ésimo tráfego e j o j -ésimo bloco de recurso.

Modified largest weighted delay first (MLWDF) - Apresenta duas métricas, uma para tráfegos de tempo real (TR) e outra para não tempo real (NTR). Sendo que o NTR funciona como o PF. A nova métrica busca priorizar tráfegos de tempo real com maior espera na fila e melhor condição de canal, definida em [27] como

$$w_{i,j} = \alpha_i D_{HOL,i} \frac{r_{i,j}}{\bar{R}_i}, \quad (2.3)$$

em que $\alpha_i = -\frac{\log \delta_i}{\tau_i}$, sendo δ_i a probabilidade do atraso $D_{HOL,i}$ atingir o limite máximo definido pela qualidade do serviço. E τ_i , um limite de latência.

Exponential PF (EXP) - Desenvolvido para aumentar a prioridade de fluxos de tempo real contra fluxos não tempo real. Apresenta métrica semelhante ao MLWDF, porém com fator exponencial, conforme demonstrado por Basukala *et al.* [28]. Definido por

$$w_{i,j} = \exp\left(\frac{\alpha_i D_{HOL,i} - \chi}{1 + \sqrt{\chi}}\right) \frac{r_{i,j}}{\bar{R}_i}. \quad (2.4)$$

Os símbolos conhecidos possuem os mesmos significados previamente demonstrados, e $\chi = \frac{1}{N_{rt}} \sum_{i=1}^{N_{rt}} \alpha_i D_{HOL,i}$, em que N_{rt} representa o número de tráfegos de tempo real no *downlink*. E para não tempo real a métrica fica definida como a do PF.

Exponential rule (EXPR) - Neste método, os pacotes que não foram servidos até o momento de expiração podem ser removidos da fila, além disso, a latência dos usuários permanece normalizada a partir da latência total do sistema, garantindo maior estabilidade sem necessitar da informação do canal. Para aplicações não tempo real, este escalonador possui funcionamento semelhante ao PF, já para as que apresentam necessidade de tempo real [9], a métrica será calculada por

$$w_{i,j} = \exp\left(\frac{\alpha_i D_{HOL,i}}{1 + \sqrt{\chi}}\right) \Gamma_k^i. \quad (2.5)$$

Logarithmic rule (LOGR) - A partir de uma função logarítmica em cima da latência, este escalonador prioriza tráfegos com maior sensibilidade à latência [9]. Calcula-se sua

métrica por

$$w_{i,j} = \log(\alpha_i D_{HOL,i}) \Gamma_k^i \quad (2.6)$$

Apresenta capacidade de reduzir em até 80% a latência média, tornando o escalonador útil para momentos de muita carga na rede.

Frame Level Scheduler (FLS) - Definido por três etapas, este escalonador mostra-se entre os melhores quanto à performance em meio à comunidade acadêmica. Utiliza teoria de controle em tempo discreto para estabelecer uma lei de controle com o objetivo de filtrar os tráfegos [8]. A Figura 2.9 ilustra seu comportamento em alto nível, explicitando a camada superior e inferior, conforme especificado por Piro *et al.* [8].

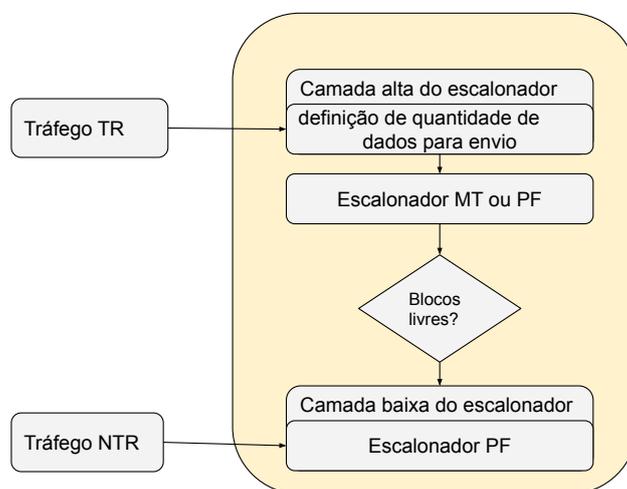


Figura 2.9: Funcionamento escalonador *Frame Level*.

A teoria de controle constitui um subdomínio da matemática, que opera sobre a decisão de sistemas dinâmicos. O controlador monitora uma variável e a compara com limites de referência [29]. Existem dois tipos de controle, laços abertos e fechados. No primeiro tipo, a ação do sistema pode ocorrer com independência da saída do sistema, já no segundo existe dependência.

Observando-se o artigo de Piro *et al.* [8], percebe-se que esta lei ocorre no começo de todo quadro, utilizando o volume de dados que cada tráfego deve transmitir. Busca ainda estabilidade *Bound Input Bound Output* (BIBO), que representa o máximo de dados trafegados (*output*) por um UE num determinado quadro, limitado por um topo B relacionado à sua entrada de dados (*input*). Por meio dessa estabilidade, encontra-se quais tráfegos

já enviaram o máximo estabelecido nos últimos subquadros, passando, assim, a ficar fora da alocação, dando chance de prioridade aos próximos tráfegos. Possuir a propriedade BIBO não garante qualquer melhora de desempenho, porém, estabelece que o sistema não irá tentar alocar uma quantidade infinita de recursos para apenas um elemento. Piro *et al.* [8] propõe um modelo que satisfaz requisitos de latência dos fluxos a serem servidos. Para definir o quanto de dados será transmitido utiliza-se a seguinte equação

$$u_i(k) = q_i(k) + \sum_{n=2} M_i [q_i(k-n+1) - q_i(k-n+2) - u_i(k-n+1)] c_i(n), \quad (2.7)$$

em que $q_i(k)$ corresponde ao tamanho da fila no tempo $t_{k,i}$, $u_i(k)$ ao total de dados transmitidos no k -ésimo quadro, $d_i(k)$ à quantidade de dados que irá preencher a fila durante o k -ésimo quadro e M_i à quantidade de intervalos de amostragem. Desta forma, o filtro consegue garantir, como mostrado pelos autores, limites de latência em até M_i intervalos consecutivos.

Já na segunda etapa do FLS, há uma dependência entre a nova entrada com a correção feita no sistema. Desta forma, filtrando o sinal e aproximando o desvio dos limites a zero, a alocação nesta etapa resulta da aplicação das estratégias MT ou PF, porém nada impede que outras estratégias sejam adaptadas para utilização. Por fim, na terceira etapa, caso ainda existam blocos de recurso livres, a estratégia *Proportional and Fair* aloca o tráfego não tempo real. Os blocos ficam por último, pois a qualidade de serviço permite maior tempo de espera total, utilizando o PF para garantir maior justiça para estes tráfegos. A Figura 2.10 demonstra uma possível alocação desta estratégia, indicando blocos utilizados por qualidades de serviços prioritárias e não prioritárias, sendo o colorido pertencente a tráfegos TR e o branco, disponível à camada de baixo para tráfegos NTR.

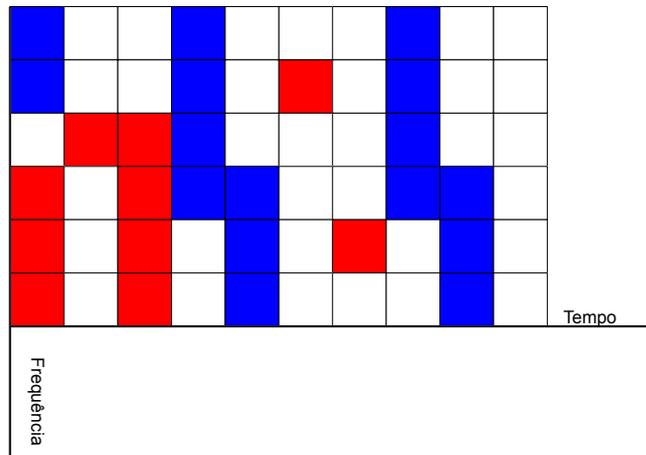


Figura 2.10: Exemplo de alocação FLS.

2.3.2 Métricas de avaliação

As métricas de avaliação possuem um papel de extrema importância na medição do desempenho atingido por cada escalonador. Por isso, determina-se medidas em que as estratégias de escalonamento se diferenciam, primariamente, no balanço entre a complexidade computacional e a otimalidade da configuração. Assim, métricas como fração de perda de pacotes, taxa de transmissão, latência, variação da latência (*jitter*) e índice de justiça (equidade) denotam os principais fatores que medem o impacto de desempenho entre as políticas de alocação de blocos de recurso [3].

A fração de perda de pacotes representa o montante de pacotes enviados pelos remetentes em relação ao recebido pelos destinatários, $\frac{N}{T}$, em que N corresponde à quantidade de pacotes recebidos e T ao total de pacotes. Uma porcentagem alta de perda de pacotes pode indicar instabilidades na rede, locais de baixa conectividade e retransmissões desnecessárias. Por outro lado, uma baixa fração indica uma rede em pleno funcionamento e em boa capacidade, possivelmente capaz de manter uma quantidade maior de usuários.

A taxa de transmissão representa a quantidade de *bits* transmitida por segundo. A medição pode ocorrer individualmente para cada dispositivo em cada torre, alcançando-se, assim, a taxa por célula. A partir daí, obtém-se a taxa para a rede inteira, somando a vazão de cada dispositivo [30], $\sum_{i=1}^n \bar{R}_i$, em que n representa o número de usuários e \bar{R}_i a vazão média por usuário.

A latência de transmissão se refere à quantidade de tempo necessária para enviar um pacote do remetente ao destinatário. Sua medição, geralmente, ocorre em microssegundos (ms) [31]. O *jitter* constitui uma medida que possui correlação com a latência *jitter*, representando a variabilidade da latência de transmissão, uma aplicação de taxa de transmissão constante, por exemplo, apresentaria *jitter* 0 ou muito próximo desse valor. Por ser a variância do tempo em que ocorre, a transmissão pode ser medida por meio da fórmula da variância da população

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2, \quad (2.8)$$

em que N representa o tamanho da população, y_i um elemento e μ a média da propriedade. Caso necessário, pode também ser feita a variância da amostra

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n n(x_i - \bar{x})^2, \quad (2.9)$$

de maneira semelhante n representa o tamanho da amostra, x_i o elemento e \bar{x} a média da propriedade dentro da amostra de tamanho n .

Por fim, o índice de equidade indica se os usuários competem pelos recursos de maneira equivalente. Revela-se um requerimento de extrema importância, pois um baixo nível de equidade implica debilidade dos usuários quanto aos recursos. Esta medida, em geral, decorre da taxa de transmissão média dentre os usuários, mas também pode ser medida por outros fatores, como por exemplo, o tempo médio de conexão de cada usuário ou a qualidade da conexão. O índice de equidade mais utilizado constitui o índice de Jain [32], obtido por

$$\frac{(\sum_{i=1}^n x_i)^2}{n \sum_{i=1}^n x_i^2}, \quad (2.10)$$

em que n corresponde ao número de usuários e x_i à taxa de transmissão da i -ésima conexão. A gama deste coeficiente varia de $\frac{1}{n}$ a 1, sendo 1 o melhor resultado, por representar a situação em que todos os usuários receberam uma alocação equivalente.

2.4 Simuladores

Atualmente, em se tratando de simulação em nível de sistema, existem algumas alternativas para o LTE, dentre elas encontram-se Network Simulator 3 (NS3) [33], LTE-Simulator (LTE-Sim) e soluções ad hoc [5]. Estas três possibilidades compreendem as mais utilizadas na literatura atual. As soluções ad hoc permitem ao pesquisador maior controle e poder de customização das características, porém, requerem uma curva de aprendizado

maior, uma vez que esse tipo de solução foi construída pelos pesquisadores com o intuito específico do trabalho realizado.

O NS3 e o LTE-Sim consistem em dois simuladores mais robustos e *open source*, desenvolvidos com o objetivo de facilitar ao pesquisador a realização de simulações, permitindo, assim, foco no tópico de pesquisa com maior facilidade. O NS3 apresenta, além do modelo LTE, suporte a diversos outros protocolos e arquiteturas, demandando tempo mais elevado para o aprendizado, pois se faz necessário notar como determinados protocolos interagem entre si por meio das classes em comum. Por outro lado, o LTE-Sim apresenta somente a arquitetura LTE, o que promove maior facilidade de entendimento dos códigos, facilitando aos pesquisadores a realização e o foco no tema de pesquisa e causando menores efeitos colaterais em mudanças no sistema.

2.4.1 LTE-Sim

Disponibilizado em 2012, o LTE-Sim se mantém em evidência em trabalhos recentes, vide [23] ou [34]. Quando do seu lançamento não havia outros simuladores de código aberto para LTE e os pesquisadores ou operadoras desenvolviam seus próprios simuladores [5]. As poucas alternativas disponíveis necessitavam da compra de uma licença, porém como não existia acesso ao código, as funcionalidades mantinham-se limitadas quanto ao mapa de atualizações da empresa mantenedora do simulador.

Nesse mesmo período, 2012, a demanda para operadoras que disponibilizam 4G LTE crescia, existindo mais de 20 empresas com tecnologia para pronta entrega. A tecnologia 4G LTE representava um grande tópico de estudo no polo industrial e no meio acadêmico. Atualmente, conforme esclarece a empresa alemã “Statista”, provedora de dados estatísticos e estratégias comerciais, a quantidade de inscritos no serviço LTE gira em torno de 3,5 bilhões de usuários, com expectativa de crescimento para 5,3 bilhões por volta de 2021 [4]. Neste diapasão, o estudo do 4G ainda se mantém em alta, com vários dos resultados obtidos podendo ser empregados à tecnologia 5G que engatinha para uma padronização estável por meio de novas pesquisas.

Em seu lançamento, o LTE-Sim englobava diversos aspectos das redes LTE, como ambientes de célula única e múltiplas, gestão de qualidade de serviço, multiusuários com mobilidade e funcionalidade de transferência de célula. Três nós foram modelados, UE, eNB e MME/GWs e diversos geradores de tráfego na camada de aplicação, disponibilizados, possuindo gestão de portadora. Implementou-se, ainda, estratégias de escalonamento bem conhecidas na literatura, como PF, MLWDF e EXP. Por fim, modelos para reportar a qualidade do canal de cada usuário, técnicas de reuso de frequência e também a camada física foram desenvolvidos. Estas características proporcionaram a pesquisadores e ope-

radores testar e melhorar as técnicas para a quarta geração de redes móveis em diversas áreas, disponibilizando gratuitamente o conhecimento sob a licença “GPLv3”.

Arquitetura do sistema

Um cenário de simulação apresenta diversos objetos para modelar o sistema LTE. Cada um deles pode, se necessário, enviar uma chamada para gerar um novo evento por meio do método *Simulator::Schedule*. O “Calendar” organiza os eventos gerados em diversas áreas em ordem cronológica. No começo de cada simulação, o calendário possui três eventos: *StartSimulation*, *StartFrameManager* e *StopSimulation*. De acordo com a configuração de simulação, outros eventos como *Application::CreatePacket*, *NetworkNode::SendPacketBurst* e *ENodeB::ResourceAllocation* podem ser adicionados para preencher o cenário criado.

Os três tipos de nós, UE, eNB e MME/GWs, encontram-se controlados pelo gerenciador de rede (*Network Manager*). Cada nó existente pode ser um remetente ou destinatário dos tráfegos de dados definidos pela quinta-upla clássica (*Source IP*, *Destiny IP*, *Source Port*, *Destiny Port*, *Transport Protocol*). O simulador provê suporte à alocação de recursos tanto no domínio do tempo quanto no da frequência, em consonância com as especificações da tecnologia LTE apresentadas anteriormente na Seção 2.2.

O gerenciador de quadros (*Frame Manager*) se torna responsável por escalonar corretamente os quadros e subquadros (TTIs) e por sincronizar as torres (eNBs). O modelo físico representa um objeto, em cada UE e eNB, que guarda os parâmetros correspondentes ao modelo do canal, à qualidade do canal, aos níveis de interferência. Outras informações referentes à camada física, como frequência da portadora, banda disponível, lista dos RBs para *downlink* e *uplink* e configurações de reuso de frequência encontram-se no gerenciador de banda (*Bandwidth Manager*).

Topologia da rede

A topologia constitui um conjunto de células e nós (eNBs, um ou mais MME/GW e UEs) distribuídos entre as células. O gerenciador de rede (*Network Manager*) provê os métodos para controle da topologia da rede. Uma célula LTE, implementada pela classe *Cell*, possui identificador único (ID) e atributos como o raio da célula e a posição no sistema cartesiano. Para cada nó da rede há uma entidade dedicada que estende o nó base (*NetworkNode*), ou seja, *ENodeB*, *UserEquipment*, *MME-GW*, também verificados por um identificador único e sua posição.

Diversas funcionalidades do plano de usuário e de controle se apresentam no simulador, providas pela entidade *ProtocolStack*, que contém o RRC, PDCP e MAC da camada de protocolos LTE discutidos na Seção 2.2 da presente pesquisa. A torre (eNB) efetua o gerenciamento de recursos do rádio, já as estratégias de *downlink* e *uplink* encontram-se

implementadas e definidas nas variáveis $m_downlinkScheduler$ e $m_uplinkScheduler$. O objeto para simulação física m_phy foi definido para prover interface de comunicação entre o dispositivo LTE e o canal, este modelado pela entidade *Channel*, oferecendo meios de acesso ao canal de rádio. Quanto à qualidade de canal, os dispositivos de usuário estimam a qualidade e convertem em um conjunto de respostas enviadas à torre, o simulador possui capacidade de utilizar o modelo periódico e aperiódico. Durante a simulação, cada torre guarda as informações dos UEs associados, e para cada um deles, seu ID e o último conjunto de qualidade de canal recebido.

Mobilidade dos usuários

Dois modelos de mobilidade se encontram implementados no simulador, direção aleatória e passeio aleatório [35]. Uma entidade dedicada que estende *MobilityModel* foi gerada para cada modelo, sendo elas *RandomDirection* e *RandomWalk* [5]. A entidade *MobilityModel* apresenta variáveis para controle de velocidade e direção. Para cada UE, um elemento $m_mobility$ indica o gerenciamento do modelo de mobilidade. De acordo com o 3GPP [36], as velocidades de 0, 3, 30 e 120 km/h correspondem a cenários imóveis (estáticos), pedestres (caminhada) e veiculares, respectivamente.

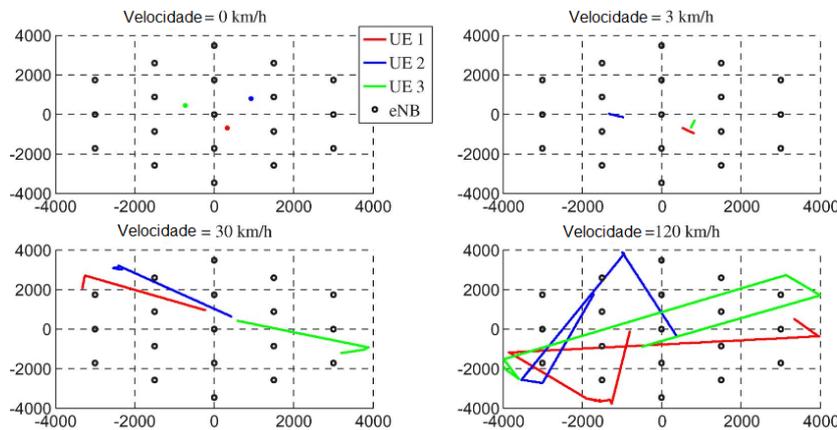


Figura 2.11: Direção aleatória.

No primeiro modelo, direção aleatória, o UE decide uma direção e move-se em velocidade constante pré-determinada até atingir o limite da área de simulação, em que decide uma nova direção. Já para o segundo modelo, passeio aleatório, o UE escolhe uma direção e uma distância, alterando a escolha caso atinja a distância ou o limite da área de simulação. A distância percorrida padrão se encontra atrelada à velocidade de movimento, sendo 200, 400 e 1000m para 3, 30 e 120km/h, respectivamente. As Figuras 2.11, 2.12, baseadas em [5], ilustram o movimento de dois modos nas três velocidades, exemplificando a diferença entre eles.

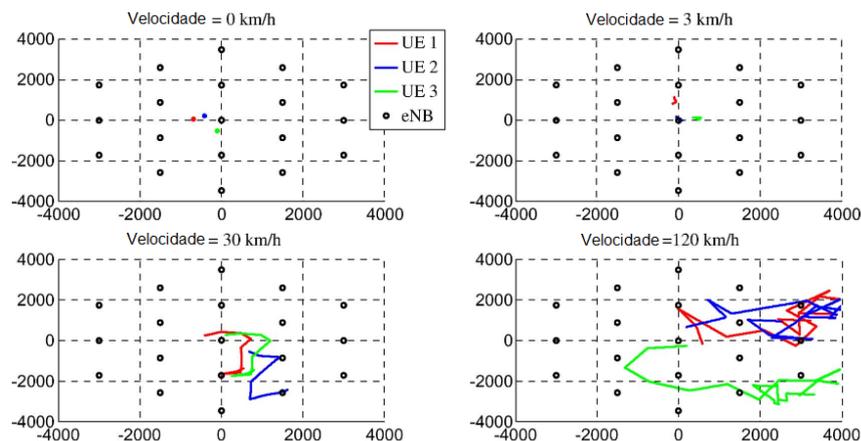


Figura 2.12: Passeio aleatório.

A gerência da mobilidade situa-se sob responsabilidade da entidade *NetworkManager*, esta, por sua vez, atualiza a posição dos usuários a cada TTI, e realiza a transferência de célula se houver necessidade. Durante uma transferência, o UE ingressa em um estado no qual nenhum tráfego encaminhado a ele possa ser escalonado por um determinado período de tempo, sendo o padrão de 30ms.

Tráfego da rede

A entidade *Bearer* modela uma portadora de rádio dedicada, podendo ser classificada em GBR ou non-GBR. Quando um tráfego de *download* se inicia, ativa consigo uma portadora dedicada entre a torre e o UE e vice-versa. Além disso, entre o UE e a torre pode-se ativar mais de uma portadora. O objeto *QosParameters* provê, para cada portadora, a definição dos requisitos de qualidade de serviço para o tráfego, tais como o identificador de QoS, a prioridade de alocação e retenção, o GBR, a taxa máxima de transferência, e a latência de espera máxima.

Um classificador baseado em IP mapeia os pacotes que chegam da camada IP para uma determinada portadora, de acordo com os parâmetros na entidade *ClassifierParameters*. A classificação se baseia na quinta-upla de transporte IP clássica. Esta funcionalidade encontra-se implementada na entidade *Classifier*, definida para todo elemento da rede.

Existem alguns tipos de geradores de tráfego implementados no LTE-Sim, tais como *Trace Based*, em que a partir de um arquivo no formato *Common Intermediate Format* (CIF) pode-se enviar pacotes de vídeo, *Voice Over IP* (VoIP), *World Wide Web* (WEB), *Constant Bit Rate* (CBR) e *Infinite Buffer* (*Best Effort*), todas as comunicações conduzidas no simulador ocorrem por UDP, o padrão TCP não se encontra implementado no simulador.

A aplicação de Vídeo utiliza rastros de arquivos de vídeo reais, cada arquivo possui quatro colunas, a primeira representa o índice do quadro, a segunda indica o tipo do quadro (I - *Intra coded picture*, P - *Predicted picture* e B - *Bidirectional predicted picture*), a terceira explicita o tempo em que o quadro foi gerado pelo *encoder* e a quarta, o tamanho do quadro em *bytes*, o simulador contém três vídeos intitulados *Foreman*, *Highway* e *Mobile* em diferentes *bitrates*, especificamente 128, 242 e 440 Kb/s [5]. Modela-se a aplicação *web* por tamanhos de pacotes em uma distribuição de Pareto, enviando-os seguindo uma distribuição geométrica, para os períodos ligado/desligado estilo cadeia de Markov, podendo modificar o tamanho máximo de cada pacote. A aplicação VoIP gera fluxos G.729 modelados por um interruptor estilo liga/desliga em uma cadeia de Markov. O período ligado se mostra distribuído exponencialmente com média de 3s, e o desligado, com um período máximo de 6,9s e média de 3s. Enquanto ligado, o remetente envia pacotes de 20 bytes a cada 20ms (mantendo uma taxa de 8kb/s). Estes parâmetros podem ser modificados por valores definidos na entidade VoIP. A aplicação CBR gera pacotes de taxa constante, podendo modificar o tamanho do pacote e o intervalo de envio. Por fim, a *infinite buffer* trabalha de maneira voraz, sempre portando pacotes para enviar durante toda a simulação.

Cada aplicação herda as características da entidade *Application*, tal como o método de enviar pacotes, *Application::Send* entrega o pacote ao dispositivo de rede por meio de *NetworkNode::EnqueuePacket*. O datagrama será passado pelo plano de usuário (dados) para que seja adicionado os cabeçalhos do protocolo e então remetido à camada MAC para ser associado a uma portadora, por meio de seus parâmetros classificadores. A Figura 2.13, baseada em [5], representa o plano de usuário encontrado no LTE-Sim.

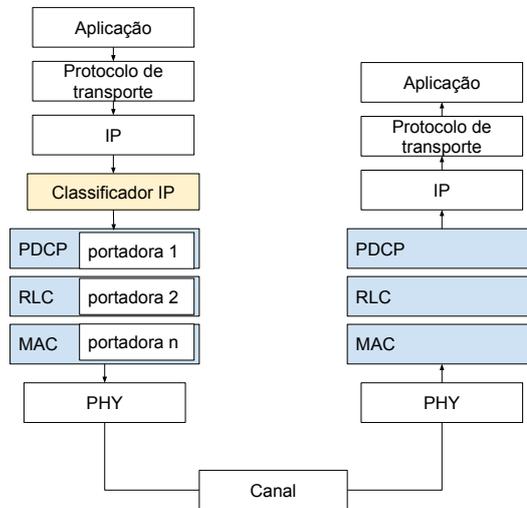


Figura 2.13: LTE-SIM pilha do plano de usuário.

O simulador implementa diversas funcionalidades dos planos apresentados na Seção 2.2.2 desta pesquisa, a entidade *ProtocolStack* contém o RRC, PDCP e MAC. Uma instância desta entidade existe para cada dispositivo da rede, além disso, há uma entidade RLC para cada portadora. Para as torres, eNBs, a entidade *EnBMACEntity* adiciona funcionalidades para escalonadores de *uplink* e *downlink*.

Os pacotes pertencem a entidade *Packet*, que contém três importantes variáveis para o funcionamento, $m_timeStamp$, m_size e $m_packetHeaders$. A primeira, representa o instante de criação do pacote na camada de aplicação, utilizada para computar a latência de transferência. A segunda e a terceira representam o tamanho do pacote e a lista de protocolos de cabeçalho, sendo o tamanho atualizado de acordo com o cabeçalho. Cada portadora mantém sua lista *First in First Out* (FIFO) a partir da entidade *MACQueue*, e adiciona-se o cabeçalho PDCP que provê compreensão dos dados [37].

Gerenciamento de recursos

Os recursos de rádio encontram-se distribuídos entre os usuários, no domínio do tempo e no da frequência, já a estação base separa os recursos para *uplink* e *downlink* no começo de cada subquadro. O simulador suporta as seis bandas de frequência especificadas pelo 3GPP [17] 1, 4, 3, 5, 10, 15 e 20MHz.

O gerenciador de banda se torna necessário para todos os dispositivos, assim há uma entidade dedicada *BandwidthManager* para cada objeto físico (PHY), desse modo dispositivos pertencentes a uma mesma célula possuem a mesma informação de banda.

O reuso de frequência implementado segue Rappaport [38], a ideia consiste em uma torre utilizar somente um subconjunto dos canais disponíveis, de modo que torres adjacentes utilizem outros subconjuntos. Denomina-se um grupo de células de agrupamento (*cluster*) que pode ser replicado para cobrir a área de serviço.

O método *NetworkManager::RunFrequencyReuse* pode receber os seguintes parâmetros: tamanho do *cluster*, número de células, bandas para *downlink* e *uplink*. E distribui a banda para que as células de um mesmo agrupamento não tenham canais sobrepostos. O método retorna uma lista de objetos *BandwidthManager* indicando as novas bandas a serem utilizadas.

Sobre escalonamento de recursos, o LTE objetiva satisfazer os requisitos de qualidade de serviço descritos na Seção 2.2 deste trabalho, buscando encontrar um balanço entre justiça e utilização dos canais, seja por eficiência ou por taxa de transmissão, utilizando-se para tal tarefa os seguintes algoritmos presentes no simulador: *Proportional and Fair*, *Maximum Throughput*, *Modified largest weighted delay first*, *Exponential PF*, *Exponential Rule*, *Logarithmic rule* e *Frame Level Scheduler*, todos bem conhecidos na literatura e já abordados em seções anteriores. Por fim, criar um novo algoritmo no simulador pode ser facilmente desenvolvido a partir dos seguintes passos:

- gerar uma nova entidade para o escalonador ou estender de algum existente;
- adicionar os métodos obrigatórios de acordo com a entidade pai;
- definir os métodos para implementar as funcionalidades do novo escalonador.

A estimativa \bar{R}_i (taxa média de transmissão) ocorre por

$$\bar{R}_i(k) = 0,8\bar{R}_i(k-1) + 0,2\bar{R}_i(k), \quad (2.11)$$

em que a taxa média do tráfego i para o TTI k e $k-1$ representa o TTI anterior, já a métrica w varia de acordo com o algoritmo de escalonamento escolhido.

2.5 Considerações finais

Neste capítulo, discutiu-se a fundamentação necessária ao entendimento dos tópicos principais a serem abordados na presente proposta, em especial o escalonador de blocos de recurso no *downlink*, foco desta pesquisa, posto o escalonador não depender da tecnologia utilizada pelo rádio, o que possibilita estudos futuros em 5G ou outras redes, uma vez

que LTE 4G aos poucos perde espaço nas pesquisas e futuramente no próprio mercado consumidor de tecnologia.

Os blocos de recursos representam a menor unidade de recurso que pode ser alocado para um usuário, possibilitando o envio de dados requisitados pelos usuários por meio da torre. Para determinar qual bloco será alocado, a camada MAC utiliza uma estratégia de alocação para definir uma métrica w , reservando um bloco para os n usuários que possuem maior valor desta métrica, em que n representa a quantidade de blocos disponíveis de acordo com a largura da banda.

Além disso, conta com uma explicação dos simuladores de rede em nível de sistema, com foco no simulador LTE-Sim, devido à escolha de seu uso nesta pesquisa. Este simulador conta com a simulação completa do sistema LTE, possuindo parâmetros de mobilidade do usuário, transferência de célula, variação da latência e, o mais importante, diversas estratégias de alocação de blocos de recurso, incluindo todas as presentes na Seção 2.3 e apresentando fácil extensão das funcionalidades para implementação de novas políticas ou protocolos. O próximo capítulo, apresentará estudos relacionados ao estado da arte na área de políticas de alocação de blocos de recurso em redes móveis (4G LTE), buscando-se, assim, o devido embasamento teórico à pesquisa.

Capítulo 3

Revisão do Estado da Arte

Este capítulo tem como objetivo demonstrar o estado da arte e os principais desafios para a elaboração de uma política de alocação de blocos de recurso. A Seção 3.1 apresenta a revisão de diferentes estudos que exploram estratégias de alocação e a Seção 3.2 compreende as considerações que abordam tópicos de relevância para este trabalho.

3.1 Políticas de alocação de blocos de recurso

Para acomodar os diferentes tipos de fluxo de tráfego, a política de alocação de blocos de recurso deve ser flexível o suficiente para manter um conjunto de métricas de qualidade que se relacionam com os requisitos do usuário. A fim de atingir este objetivo, vários pesquisadores propõem soluções diversas que exploram táticas de gerenciamento de recursos. Porém, algumas destas soluções podem trazer consigo limitações em certos pontos, como detrimento ou exaustão de tipos de tráfego, baixa taxa de transmissão ou alta perda de pacotes, a depender dos cenários em que as simulações se encontram.

A proposta de Suganya *et al.* [25] emprega as classes de qualidade de serviços e suas prioridades para utilizar métodos de pesos diferentes para cada fluxo de dado, desta forma pode-se mesclar a técnica de *Inverse Weighted Fair Queuing* (IWFQ) com o PF (Seção 2.3), resultando na política *Proportional Inverse Weighted Fair Queuing* (PIWFQ) e mudar a forma de priorização dos pacotes. Os resultados revelam-se promissores, entretanto, apesar de recente, o estudo utiliza estratégias defasadas, como *Round Robin* (RR) e *Blind Equal Throughput* (BET), e apenas PF entre as estratégias mais consolidadas na literatura atual, podendo ter um impacto inferior quando comparada a estratégias mais novas e eficientes. No trabalho, os autores utilizaram apenas 3 usuários, e fizeram a comparação de vazão por meio dos compartimentos de 0,5ms para um TTI, atingindo melhorias de até 600% em relação ao RR. Apesar de ser um bom resultado possui comparações muito

curtas com relação ao tempo de duração da aplicação, apenas um TTI, e pouca variedade de usuários.

O trabalho de Akl [39] inclui modificações sutis no mecanismo da estratégia *Proportional and fair*, utilizando-a em conjunto à *Best CQI* (BCQI). Caso o número de usuários que desejam transmitir seja maior que a quantidade total de blocos de recurso utiliza-se PF, do contrário, significa que a quantidade de blocos se torna suficiente para todos os usuários ativos, utilizando assim o *Best CQI*. No trabalho, o autor utiliza apenas o *Round Robin* na comparação, atingindo vazão média maior em até 260% com média de ao menos 200%. Novamente, encontra-se um estudo que utiliza poucos intervalos, neste caso, 50 TTI para duração do tempo de simulação e há apenas a comparação de uma política de alocação de blocos de recurso. Verificou-se uma extensão deste trabalho em [40], no qual foram utilizadas novas estratégias de alocação, *Maximum Throughput* e PF, para outra comparação de desempenho, além de aumentarem o quantitativo de usuários chegando a 90 no total. Na extensão do estudo, o novo algoritmo proposto encontra-se próximo ao PF original. Observou-se que o aumento do número de usuários resultou em melhor performance, mantendo um aumento de vazão média a uma melhora de 33%, entretanto ainda houve perda de cerca de 13% em relação ao desempenho do MT (Seção 2.3) para a mesma métrica. A definição da métrica de alocação de blocos de recursos ($\Gamma_{i,k}$) se dá pela razão entre a taxa esperada de vazão (T_E) sobre a vazão máxima de um determinado UE ($T_{i,\max}$), somada a razão da vazão média ($T(AVG)_i$) sobre, também, a vazão máxima do UE, resultando em

$$\Gamma_{i,k} = [((1 - (\frac{T(AVG)_k}{T_{i,\max}}))(\frac{T(t-1)}{T_{\max}})) + ((1 - (\frac{T(t-1)}{T_{\max}}))(\frac{T_E}{T_{i,\max}}))]. \quad (3.1)$$

O estudo [40] continua sem apresentar desempenhos de outras métricas relevantes como latência ou índice de justiça.

Aramide *et al.* [41] apresentam uma proposta de generalização do PF, com o intuito de aproximar os resultados do PF original ou do *Best CQI*, dado fatores de ajuste β e Γ , expressada por

$$w_{i,j} = \frac{r_{i,j}^\beta}{R_i^\Gamma}, \quad (3.2)$$

desta forma, conforme se altera os fatores se utiliza mais da métrica da taxa de transmissão instantânea ou da média histórica da taxa para os fluxos de dados de cada aplicação. Para os autores, os valores de 0,6 e 0,7 representam pontos de maior desempenho, atingindo melhorias de 6% com relação ao PF e 13% com *Round Robin*, porém perdendo por 17% da política *Best CQI*. Os autores não realizaram comparações em relação à latência, entretanto evidenciaram melhorias em relação à eficiência energética de maneira seme-

lhante à métrica de vazão, perdendo também para o *Best CQI*. Apesar dos pontos fracos quando comparado ao escalonador de melhor qualidade do canal, a proposta de Aramide *et al.* [41] exhibe vantagem no índice de justiça, pois neste caso o BCQI apresenta pontuação de 0,26 enquanto o proposto *Generalized PF* (GPF) marca 0,86, tornando-o mais justo.

No trabalho de Piro *et al.* [8], há a proposta de um algoritmo de alocação denominado *Frame Level Scheduler* (FLS), na qual, considerando os resultados obtidos, alcançou um dos melhores desempenhos em contextos propostos pelos pesquisadores. Os testes envolveram fluxos de tráfego em MPEG-4 (vídeo) e G.729 (voz), em cenários de 15 a 25 usuários. Os resultados obtidos neste escalonador, em geral, relevaram-se melhores que os já existentes em determinados cenários, os autores desta pesquisa utilizaram testes de *Mean Opinion Score* (MOS) para medir VoIP e também *Peak Signal to Noise Ratio* (PSNR) para estimar a qualidade do vídeo recebido, demonstrando aumento do PSNR em 30dB em relação às estratégias EXPR/LOGR e pontuação MOS de 3,6, o que corresponde à satisfação de grande parte dos usuários. A estratégia FLS divide o escalonamento de blocos de recurso em dois níveis: Alto e Baixo. A camada baixa funciona como um escalonador convencional, utilizando a estratégia *Proportional and Fair*. Já a camada Alta, serve para avaliar quadro a quadro as limitações de qualidade de serviço, a fim de determinar a quantidade de dados (*bytes*) a ser enviado. Feito por meio de teoria de controle em tempo discreto [42], ao recalculer a quantidade ideal do tamanho da transmissão, nota-se um aumento na métrica de vazão, sem prejudicar outras, neste ponto se encontra o grande diferencial da pesquisa. Entretanto, esta solução possui uma grande priorização de pacotes de tempo real, alocando para o outro tipo de tráfego apenas quando ainda há blocos de recurso livres após a primeira etapa. Por isso, torna-se possível identificar cenários nos quais o desempenho da proposta estará defasado em relação a qualquer outra estratégia, por exemplo, em um cenário com grande volume de dados não tempo real e pouco tempo real encontra-se uma taxa de perda de pacotes demasiada para pacotes não priorizados, o que prejudica a rede como um todo, chegando até mesmo a exaurir fluxos deste contexto. Diante do exposto, na proposta que será descrita no Capítulo 5 da presente pesquisa, busca-se amenizar os efeitos da perda de pacotes e vazão em cenários nos quais o volume de dados não tempo real cresce mais que o seu concorrente.

Teixeira e Timóteo [7] trabalham em uma solução que possui o objetivo de melhorar a estimativa da taxa média de transmissão, presente na Equação 2.11, por meio de um filtro de Kalman. Tal solução utiliza o simulador LTE-Sim e realiza a modificação nas políticas FLS, EXP, EXPR, LOGR e MLWDF. O filtro de Kalman emprega duas variáveis de controle Q e R , para determinar o peso associado ao medido no ambiente e ao estimado pelo filtro. Desta forma, torna-se possível obter estimativas mais próximas da real ou da

filtragem, ocorrendo a calibração. No trabalho, os autores conseguiram mostrar que a utilização de Kalman nesta etapa pode ser eficiente, obtendo um aumento de vazão, de no máximo de 50% para a estratégia EXP, mas ainda demonstrando melhorias quando comparada às outras, assim como uma redução de taxa de perda de pacotes, em torno dos 30% para o EXP. Observa-se um aumento da latência devido aos cálculos extras que a camada MAC deve realizar para terminar a alocação, girando entorno dos 52%, porém ainda abaixo do limite máximo de 40ms escolhido pelos autores. No trabalho, os referidos autores utilizaram cenários contendo apenas tráfego de vídeo, contudo, os outros tipos de tráfego costumam influenciar nos resultados obtidos, logo, novos trabalhos necessitam explorar o efeito que a mistura de aplicações pode causar no filtro proposto.

Baseado na otimização de estratégias conhecidas na literatura, Gungor e Uyan [43] propõem um algoritmo embasado no PF [26], modificado para prover qualidade de serviço por meio da latência entre os pacotes. Nesse algoritmo, usuários nas bordas de célula ganham prioridade quando o envio de seus pacotes não ocorrem devido à latência. Resultados revelam que o desempenho da política proposta se torna muito maior quando comparado a algoritmos que não possuem estratégias de qualidade de serviço. Apesar de apresentar melhorias quando comparado ao PF original, os autores não realizaram comparações com estratégias que possuem qualidade de serviço, como EXP, EXPR ou LOGR [9], o que revelaria desempenhos menos drásticos, pois propõem modificações similares aos já existentes na literatura. Quantificadamente os resultados obtidos pelos pesquisadores possuem vazão média com desempenho 16% superior em relação ao PF e 300% com relação a *Round Robin*.

Latência também configura o estudo realizado por Li *et al.* [44], no qual propõem uma política de prioridade, baseada no algoritmo MLWDF [27], que objetiva a melhoria da qualidade de serviços para tráfegos em tempo real. O estudo apresentou quatro escalonadores que a partir do MLWDF utilizaram a prioridade por latência, obtendo melhorias tanto na taxa de transmissão quanto no índice de justiça, nomeados *Virtual Token MLWDF* (VT-MLWDF), *Queue Head of line MLWDF* (QH-MLWDF), MLWDF2 e DP-VT-MLWDF, sendo os últimos dois propostos no estudo dos autores. O trabalho dos referidos autores também explana um modelo de prioridade por latência (do Inglês, *Delay Priority* (DP)), que pode ser acoplado às estratégias de alocação para, possivelmente, melhorar o seu desempenho. O primeiro dos escalonadores abordados no estudo, VT-MLWDF, utiliza um sistema que verifica a fila do *buffer* de cada fluxo de dados, escolhendo o que possui maior quantidade de elementos na fila, seguindo

$$m_{i,j} = \alpha_i Q_i \frac{r_{i,j}(t)}{R_i(t)}, \quad (3.3)$$

as variáveis aqui representam os mesmos significados indicados pela Tabela 2.4, sendo Q_i o *buffer*. O segundo, QH-MLWDF, utiliza de *Head of Line Delay* ($D_{HoL,i}$) para saber a latência da primeira requisição em espera de cada fila, adicionando um passo de priorização. Propõem então a estratégia MLWDF2, terceiro escalonador, que possui mais um aprimoramento, no qual utiliza um mecanismo de priorização de latência, aplicando

$$DP_i(t) = \frac{1}{\tau_i - D_{HoL,i}}. \quad (3.4)$$

Por fim, modificação do VT-MLWDF para acrescentar à priorização por latência descrita previamente. Os resultados obtidos pelos autores demonstram uma clara vantagem na utilização da prioridade por latência, melhorando não só o desempenho NTR em torno de 8% em termos de vazão, 200% com relação a taxa de perda de pacotes, como também proporcionando pequenas melhorias na latência e na vazão de tráfego RT, em torno de 14%. Entretanto, o primeiro algoritmo MLWDF2 proposto não se sobressai com relação aos demais, possuindo maior latência, e pior desempenho nas outras métricas nos cenários descritos. A estratégia que realmente se sobressaiu ao utilizar o mecanismo de DP foi a modificação do VT-MLWDF para o DP-VT-MLWDF. Diante disso, a presente pesquisa busca incorporar medidas de latência, a fim de obter desempenho semelhante ao apresentado em [44], mas confrontando a nova proposta com políticas de alocação de recursos baseada em priorização de latência com outras que possuam características semelhantes, tais como EXP, EXPR, LOGR ou FLS.

Jansang *et al.* [45] propõem um algoritmo que monitora a latência dos pacotes tempo real para cada UE, organizando a fila por tempo de espera, e comparando a um valor base C , que serve como um indicativo crítico, antes de atingir o valor máximo de espera D , o que causaria a perda do pacote. Desta forma, a partir de indicativos para os quais os pacotes foram alertados, pode-se manipular a fila e dar maior prioridade a esses pacotes com níveis ruins. A partir do momento que certa quantidade de pacotes se encontra neste estado crítico, pode-se transferir uma porcentagem de recursos destinados ao não tempo real para os de tempo real e vice-versa. Após isso, efetua-se a alocação por meio da estratégia *Proportional and Fair* convencional [26]. A etapa de classificação realiza, essencialmente, a diferenciação de tráfegos próximos de seu tempo de espera máximo, observando o $D_{HoL,i}$ ao entrar na denominada “zona crítica”, na qual atinge um primeiro nível C . Então, conforme sua classificação em TR ou NTR, determina-se um tratamento para o tráfego. Da mesma forma, dependendo da quantidade de requisições em níveis críticos, aumenta-se a quantidade de recursos alocados para cada tipo de tráfego. Os resultados foram interessantes, pois reduziu-se a latência de ambos os tráfegos quando comparado ao PF puro, sendo a maior diferença de 100% e a menor de 5%, assim como

na taxa de perda de pacotes, sendo a maior em 300%. De outra forma, aumentando levemente a vazão em torno de 30%, há também cenários nos quais a proporção entre os tipos de tráfego mudam, mostrando a eficiência da solução. Entretanto, não se utiliza cenários em que o tráfego não tempo real se torna maior que tempo real, assim como não apresenta comparações com outros algoritmos presentes na literatura, além de PF.

No trabalho de Aiyetoro e Takawira [20], há a proposta de uma estratégia de alocação, nomeada *Earliest Deadline First Queue Aware Channel Based* (EBCQA) baseada no EXP. O modelo utiliza o sistema padrão do *Proportional and Fair* multiplicado por um fator que busca atingir níveis de qualidade de serviço aceitáveis por meio de uma taxa oriunda da diferença do tempo máximo de latência pelo tempo esperado no *buffer*. Representado por

$$\exp \frac{\mu(T_{k,d} - W_k(n))}{W_k(n)}, \quad (3.5)$$

em que $T_{k,d}$ constitui o tempo máximo de latência fim a fim (*delay deadline*), $W_k(n)$ o tempo de espera no *buffer* e μ um fator de ajuste para determinar se tende mais à vazão ou à qualidade de serviço, definido como $\mu = \frac{e}{0,99T_{k,d}}$, sendo e variável de 5 a 10 de acordo com os autores. Este esquema obteve melhora de latência de 24% e leve melhora de vazão, chegando a 15%. Os resultados foram obtidos com base em uma simulação de 500s, mobilidade de usuários a uma velocidade de 30km/h, testados somente com tráfego de tempo real (Vídeo e VoIP), dentre os testes, a variável de controle obteve melhores resultados quando seu valor foi de 5.

Há uma extensão do estudo de Aiyetoro e Takawira presente em [6], no qual se modifica a estratégia original para conter otimizações, nomeada *Logarithmic Queue Aware Channel Based* (LBCQA). μ agora fica definido por $\mu = -\log_n \frac{\delta_k}{0,99T_{k,d}}$, em que n corresponde à base do log, podendo ser 2, 10 ou e , de acordo com os autores, e δ_k a probabilidade de ultrapassar a latência máxima. As mesmas condições de simulação foram realizadas para esta nova política, os resultados finais mostraram-se semelhantes ao da pesquisa original, ainda não realizando cenários em que exista a heterogeneidade das tipos de fluxo de dados, nem a utilização de novas estratégias na comparação, como EXPR ou LOGR que contemplam o estado da arte em alocação de blocos de recursos para redes LTE.

Nesta seção, percorreu-se alguns dos trabalhos relacionados ao tema da presente pesquisa, principalmente no que se refere à alocação de blocos de recurso, mostrando o desempenho de soluções recentes. Na próxima seção, delinear-se-ão alguns dos problemas encontrados nos trabalhos para melhor suportar a proposta encontrada no Capítulo 5.

3.2 Discussão

Como pôde ser visto, diversos artigos abordaram a diferenciação entre os tráfegos de tempo real e não tempo real, isso devido às restrições de latência para o primeiro grupo se encontrarem muito menores em relação ao segundo grupo. De modo geral, trata-se primeiramente o TR, utilizando a maior parte dos recursos para este e os que sobram para o NTR. Entretanto, isso pode causar um efeito de perda de pacotes para o grupo menos prioritário, e dependendo do cenário pode causar, até mesmo, a exaustão do tipo de tráfego na rede.

Notou-se, também, que apesar de alguns trabalhos serem relativamente recentes, 2016 e 2017, trazem comparativos com políticas de alocação de blocos de recurso precursoras. Mesmo demonstrando resultados interessantes e muitas vezes fortes quando comparados a certos escalonadores, como RR, BCQI, BET e PF, ao se olhar para técnicas mais novas como EXP, EXPR ou LOGR, a eficácia das soluções propostas nos estudos pode tornar-se menos expressiva. A presente pesquisa, propõe otimizações de políticas de alocação de blocos de recurso mais recentes e constantes na literatura, utilizando métodos de priorização em busca de maior eficiência, seja com relação à vazão, ao índice de justiça ou a outras métricas.

Outro método comumente utilizado ocorre por meio de um fator de peso para cada tipo de tráfego, o que contribui para facilitar a priorização a ser dada, pois um bom balanceamento dos pacotes garante maior equidade, atingindo assim um maior número de usuários e, logo, maiores taxas de transferência na célula. Este balanceamento se mostra essencial, uma vez que os parâmetros de qualidade do serviço se revelam mais restritivos para certos tipos de aplicações e mais flexíveis para outras.

Tabela 3.1: Comparativo estado da arte.

Política	Simulador	Comparações	Principal contribuição
PIWFQ [25]	MATLAB	BET, MT, PF	Pesos aos tráfegos por meio de <i>Inverse Weighted Fair Queuing</i>
[39], [40]	MATLAB	RR, MT, PF	Limitação de blocos de recurso por classificação de tráfego e modelo de cálculo de métrica baseado na taxa de vazão histórica dos UEs
GPF [41]	Vienna LTE Simulator	PF, BCQI, RR	Generalização do PF para obter performance tendendo à vazão ou ao índice de justiça dependendo do valor dos fatores
FLS [8]	LTE-Sim	PF, MLWDF, EXP, EXPR, LOGR	Utilização de teoria de controle para redefinir a quantidade de dados enviados em cada frame
Filtro de Kalman [7]	LTE-Sim	EXP, EXPR, LOGR	Utilização de filtro de Kalman para previsão da taxa de transmissão no próximo TTI
[43]	Vienna LTE Simulator	PF, RR	Utilização de dados de latência para prover melhor qualidade de serviço ao PF
DP-MLWDF [44]	LTE-Sim	VT-MLWDF, QH-MLWDF	Modelo <i>Delay Priority</i> para otimização do MLWDF
[45]	NS-3 Simulator	PF	Modelo de priorização dada a quantidade de pacotes em estado crítico, aumentando ou diminuindo a quantidade de RBs para a classificação do tráfego
EBCQA [20], LBCQA [6]	LTE-Sim	MLWDF, EXP/PF	Definição de novas métricas de alocação baseadas em funções de controle exponenciais e logarítmicas

A Tabela 3.1 possui um referencial rápido com resumo sobre as soluções descritas na seção anterior, esclarecendo os tipos de tráfego utilizados, o simulador que rodou os cenários, as estratégias de alocação de blocos de recurso comparadas e as principais contribuições de cada trabalho. Note que a maior parte dos trabalhos relacionados utilizaram o LTE-Sim como simulador, um dos motivos que incentivou o uso desse simulador na presente pesquisa. Apesar de alguns artigos, como [3], [8], [9], possuírem comparações com base em mais políticas, em geral a maioria não abrange os mesmos parâmetros, o que influencia as métricas e os resultados. Elementos como proporção entre classes de tráfego, largura da banda, velocidade dos usuários, quantidade de usuários, latência máxima de espera para cada tráfego, latência de recebimento de dados, *jitter* e outros, impactam diretamente os resultados de cada escalonador. Devido a grande gama de parâmetros utilizados nas simulações, não há um cenário pré-determinado aceito e utilizado por pesquisadores, desta forma a comparação entre os estudos torna-se menos clara. Diante disso, a presente pesquisa, busca um comparativo que expõe qualidades e limitações de alguns dos mecanismos da literatura, bem como determina os cenários nos quais cada estratégia será utilizada, com a finalidade de obter uma base de comparação encontra-se no próximo capítulo algumas políticas de alocação selecionadas exemplificando pontos fortes e fracos de cada mecanismo de alocação. Será também apresentada uma proposta de política de alocação de blocos de recurso para redes 4G LTE, apoiada em teoria de controle e prioridade por latência com características semelhantes à de políticas expostas.

Capítulo 4

Escalonadores de blocos de recurso: avaliação comparativa

A proliferação de dispositivos IdC [46], que via de regra manipulam tráfegos com prioridade não tempo real, trazem à tona a obrigatoriedade de haver mecanismos que deem vazão a esses tipos de fluxos, uma vez que estudos recentes indicam um aumento expressivo de dispositivos *smart*, sistemas de monitoração, comunicação entre máquinas e outros similares [47]. O crescimento do uso de dispositivos IdC incentiva pesquisas a exibirem cenários de simulação nos quais ocorrem uma porcentagem maior de outros tipos de tráfego, como por exemplo tráfego *web*, não necessariamente com um maior volume de dados, mas com aumento do quantitativo de usuários que utilizam aplicações de fluxo NTR, pois dispositivos de monitoração possuem baixo volume de tráfego e grande quantidade de chamadas.

Esse capítulo possui como objetivo principal a comparação das estratégias de alocação de blocos de recurso comumente utilizadas na literatura, abordadas nos Capítulos 2 e 3, denotando o comportamento das soluções em ambientes de simulação com maior quantitativo de usuários empregando aplicações de tráfego NTR, visando a determinação de pontos fortes, assim como desfavoráveis aos escalonadores. Ressalta-se que a literatura, em geral, possui critérios específicos para atender tráfegos TR, porém para NTR utiliza mecanismos menos sofisticados. Além disso, os parâmetros de configuração encontrados no estado da arte muitas vezes se mostram divergentes, não contendo os mesmos valores para as variáveis, desta forma, como objetivo secundário, torna-se necessário delinear um cenário controlado para submeter os mecanismos e se obter um desempenho inicial, e a partir deste comparar novos mecanismos de alocação de blocos de recurso, demonstrando seus resultados.

4.1 Ambiente de simulação

O Simulador LTE-Sim foi o escolhido para o ambiente de simulação desta pesquisa devido à quantidade de trabalhos aos quais se encontra aplicado, tais como [3], [7] e [20], conforme demonstrado no capítulo anterior, bem como devido à curva de aprendizagem para sua utilização quando comparado com outras soluções de simulação. A definição de cenário do simulador será aqui representada por uma função estática de arquivo de cabeçalho na linguagem C++, preferencialmente, guardada no diretório *simulation/scenarios*, permitindo ao usuário configurar o contexto desejado no programa principal. Um cenário simples pode ser gerado a partir dos passos a seguir, sendo que os componentes relatados encontram-se definidos no Capítulo 2 da presente pesquisa:

- Gerar uma célula, eNB e UE utilizando métodos do *NetworkManager*. Para cada objeto, diversos parâmetros podem ser modificados a fim de melhor se adequar à pesquisa.
- Elaborar as aplicações, definindo seus remetentes e destinos, parâmetros de qualidade do serviço, classificador de IP, tempo de início e duração.
- Definir a duração total da simulação.

Para simplificação, alguns cenários pré-determinados acompanham o simulador com parâmetros configuráveis, como por exemplo, número de células, usuários, classificação dos tráfegos, entre outros. Os resultados da simulação podem ser obtidos na saída padrão do terminal ou transferidos para um arquivo, sendo as informações contidas na saída uma visão global, incluindo as transmissões da torre e recebimento pelos usuários. Além disso, existem opções de *debug* que habilitam algumas mensagens extras, como criação de nós, sobre a alocação propriamente dita e outras. Um exemplo de resultado encontra-se a seguir:

```
1 2 3 4 5 6 7 8 9
TX WEB ID 3 B 5 SIZE 87 SRC -1 DST 6 T 0.575779 0
RX WEB ID 0 B 5 SIZE 87 SRC -1 DST 6 D 0.00122109 0
RX WEB ID 1 B 5 SIZE 87 SRC -1 DST 6 D 0.00222109 0
RX WEB ID 2 B 5 SIZE 87 SRC -1 DST 6 D 0.00222109 0
RX WEB ID 3 B 5 SIZE 87 SRC -1 DST 6 D 0.00322109 0
TX VIDEO ID 4 B 4 SIZE 1490 SRC -1 DST 5 T 1.20002 0
TX VIDEO ID 5 B 4 SIZE 1490 SRC -1 DST 5 T 1.20002 0
TX VIDEO ID 6 B 4 SIZE 840 SRC -1 DST 5 T 1.20002 0
RX VIDEO ID 4 B 4 SIZE 1490 SRC -1 DST 5 D 0.00997886 0
RX VIDEO ID 5 B 4 SIZE 1490 SRC -1 DST 5 D 0.0159789 0
```

```

RX VIDEO ID 6 B 4 SIZE 840 SRC -1 DST 5 D 0.0189789 0
TX VIDEO ID 7 B 4 SIZE 766 SRC -1 DST 5 T 1.24002 0
RX VIDEO ID 7 B 4 SIZE 766 SRC -1 DST 5 D 0.00497886 0
TX VIDEO ID 8 B 4 SIZE 112 SRC -1 DST 5 T 1.28002 0
RX VIDEO ID 8 B 4 SIZE 112 SRC -1 DST 5 D 0.00197886 0
TX VIDEO ID 9 B 4 SIZE 201 SRC -1 DST 5 T 1.32002 0
RX VIDEO ID 9 B 4 SIZE 201 SRC -1 DST 5 D 0.00297886 0
TX VOIP ID 10 B 1 SIZE 32 SRC -1 DST 2 T 1.34297 0
RX VOIP ID 10 B 1 SIZE 32 SRC -1 DST 2 D 0.00103307 0
TX VIDEO ID 11 B 4 SIZE 144 SRC -1 DST 5 T 1.36002 0
RX VIDEO ID 11 B 4 SIZE 144 SRC -1 DST 5 D 0.00197886 0

```

Os campos encontrados no registro da saída do simulador LTE-Sim possuem significados relevantes para a obtenção das métricas citadas na Seção 2.3.2. O arquivo pode ser formatado por sequências de símbolos separados por espaço, entre eles encontram-se: coluna 1 com valores de TX, RX e DROP, possuindo associação com os pacotes Transmitidos, Recebidos ou Perdidos (devido a *Block Error Rate* (BLER)), respectivamente; coluna 2 com VIDEO, CBR, VOIP, WEB, INF_BUF referem-se ao tipo da aplicação do pacote; coluna 3, ID, compreende o identificador único do pacote; coluna 4, B, refere-se à portadora que mapeia o tráfego de uma aplicação; coluna 5, SIZE, remete ao tamanho do pacote em *bytes*; coluna 6, SRC, refere-se ao remetente, enquanto a coluna 7, DST, informa o destino; coluna 8, T, refere-se ao momento em que o pacote foi gerado, já D, compreende a latência ao receber o pacote; por fim, há um valor binário, coluna 9, que indica se o dispositivo situa-se dentro ou fora de uma área fechada, como um prédio por exemplo, caso na simulação exista esse tipo de configuração, o que não ocorre para essa pesquisa.

A partir desses campos, pode-se analisar diversas outras métricas como, por exemplo, taxa de transmissão, fração de pacotes perdidos, índice de justiça e atraso fim a fim. Com o uso de alguma ferramenta para facilitar a extração das informações consegue-se obter as métricas descritas por meio das equações providas na Seção 2.3.2. Entretanto, para se obter as métricas, torna-se necessária a adaptação de algumas das equações, de modo a utilizar as informações contidas na saída do simulador. A taxa de transmissão pode ser calculada filtrando os dados marcados com RX, inclusive por tipo de tráfego, e assim calculando a média. Dada por

$$\bar{T}t = \frac{\sum_{i=0}^{N_{RX}-1} (SIZE8)}{T}, \quad (4.1)$$

em que T referencia o tempo de duração da aplicação em segundos. A taxa de pacotes perdidos

$$Tp = \frac{N_{TX} - N_{RX}}{N_{TX}} 100, \quad (4.2)$$

em que N_{TX} se refere à quantidade de pacotes transmitidos e N_{RX} à de recebidos. O índice de justiça

$$I = \frac{\bar{T}t^2}{\sum_{i=0}^{N_{RX}-1} (SIZE8)^2 N_{Uapp}}, \quad (4.3)$$

utilizando a equação de Jain, apenas alterando para as variáveis obtidas na saída da simulação, sendo N_{Uapp} o número de usuários presentes na rede por aplicação. O *Jitter* pode ser calculado diretamente por meio da variância da latência (*delay*) sem necessidade de alterações.

No LTE-Sim, o cenário refere-se à configuração da rede, incluindo posicionamento e mobilidade das entidades que a compõem. A posição inicial encontra-se constituída de um ponto aleatório dentro do raio da célula, por meio de uma distribuição uniforme, com o intuito de obter uma quantidade próxima de usuários entre a área da torre.

O simulador possui diversos cenários padrões para testes de diferentes métricas, como qualidade do canal, proporção sinal ruído, testes de escalabilidade, modelos de mobilidade, femto células e performance de escalonamento. Alguns dos artigos apontados no embasamento teórico e estado da arte utilizaram o cenário padrão de desempenho do escalonamento para obter os comparativos entre as propostas dos pesquisadores perante os escalonadores implementados, já em outros trabalhos foi desenvolvido um cenário próprio. O cenário pré-determinado pelos idealizadores do simulador [5] não utiliza tráfego *web* ou algumas das características em que esta pesquisa possui interesse de explorar. Diante disso, desenvolveu-se um novo cenário para obtenção dos resultados relacionados com o objetivo deste trabalho. As parametrizações referentes ao presente estudo encontram-se descritas na Tabela 4.1.

Tabela 4.1: Parâmetros do cenário.

Parâmetros	Valor	Descrição
Largura de banda	1400	Faixa de frequência em Hertz.
# Blocos de recurso	6	Quantidade de blocos de recurso da torre.
# Usuários	10,20,30,40,50	Quantidade de usuários.
Tipos de Tráfego	-	Tempo real (Vídeo e VoIP), Não tempo real (<i>web</i> e CBR).
Porcentagem de tráfego	50/50	Porcentagem dos tipos de tráfego.
Latência máxima	Vídeo - 60, VoIP - 40, <i>web</i> - 80, CBR - 100	Latência máxima para cada pacote em milissegundos.
Velocidade	0	Velocidade em que os usuários se movem em km/h.
Raio da célula	500	Tamanho do raio da célula da torre em metros.
Posição dos usuários	-	Posição aleatória dentro do raio da célula para cada usuário.
Duração de simulação	100	Duração do tempo de simulação em segundos.
Taxa de processamento vídeo	242	Bitrate do vídeo a ser transmitido em kb/s

Para utilizar este cenário algumas melhorias tiveram de ser implementadas no simulador, como por exemplo o balanceamento de tipos de tráfego, explicitando a porcentagem de usuários que utilizarão cada tipo, assim como a utilização da aplicação *web*, que não havia sido considerada originalmente no LTE-Sim. Essas alterações foram necessárias para tratar cenários mais complexos, demonstrando de forma mais clara momentos desfavorecidos para as políticas de alocação explicitadas nos Capítulos 2 e 3, e visando a descoberta de espaços que proporcionem melhorias para uma nova estratégia de alocação de blocos de recurso. Para isso, descartou-se a mobilidade dos usuários, em um primeiro momento, a fim de averiguar com mais certeza o desempenho obtido. As latências máximas toleradas para cada tipo de aplicação foram escolhidas de modo a obter o mesmo tipo de prioridade apresentada pela 3GPP, vide Tabela 2.1, entretanto os valores foram reduzidos, tornando-os ainda mais severos do que a 3GPP apresenta originalmente, esse procedimento ocorreu em conformidade com o trabalho de Piro *et al.* [8], para fins de comparação com a literatura. Com o propósito de obter simulações coerentes em todos os mecanismos de alocação, manteve-se a largura de banda em 1,4MHz o que permitiu, como já mencionado na Tabela 2.2, 6 blocos de recursos a serem utilizados na camada de escalonamento, possibilitando testar todos os casos em tempo hábil. A quantidade pequena de blocos de recurso ofertada facilitou o seu consumo, possibilitando maior rapidez na saturação da banda da torre e permitindo uma menor variação de quantidade de usuários para atingir casos similares aos com muito tráfego de dados. Diante disso, pôde-se utilizar testes entre 10 a 50 usuários, revelando o padrão das estratégias de alocação. Os resultados apresentados possuem 95% de nível de confiabilidade, definindo o intervalo de confiança a partir da distribuição *t-student* com 20 simulações para cada escalonador e quantidade de usuários.

Este capítulo tratará somente o cenário em que ocorre a proporção 50% para cada classificação de tráfego, e mais a frente, no Capítulo 5, abordar-se-á outras proporções, assim como o ambiente que trata de mobilidade. O resultado a ser apresentado nesse capítulo possui 50% de usuários alocados para cada classificação (TR e NTR). Considerando o subcaso do cenário quando há 50 usuários utilizando a rede, 25 destes estarão destinados ao tráfego de tempo real, sendo divididos, ainda, entre as aplicações de vídeo e voz. Existe o mesmo tipo de segregação dos usuários para o tráfego não tempo real, porém para as aplicações de *web* e CBR. Vale ressaltar que essa segregação ocorre para todos os cenários, variando a depender da proporção definida no escopo.

Durante as simulações empregou-se os tipos de aplicações presentes no simulador, classificados como tráfegos de tempo real (Vídeo e VoIP) e não tempo real (*web* e CBR). As aplicações foram utilizadas como descritas no Capítulo 2, sendo as mais importantes, *web* e vídeo, em que a primeira segue uma distribuição de Pareto para o tamanho do

pacote a ser enviado, possuindo um valor máximo de MAX_PACKET_SIZE , e a segunda utiliza pacotes baseados em arquivos de vídeo realísticos [5].

Por fim, alterou-se a latência máxima para os pacotes com o objetivo de testar o comportamento dos mecanismos de alocação quando a priorização encontra-se também dentro dos requisitos de qualidade de serviço. Isso torna a alocação dos blocos de recurso mais restrita, o que pode aumentar a quantidade de perda de pacotes, pois o mecanismo deve tomar decisões sobre qual tipo de tráfego será alocado, dado o tempo de atraso de cada requisição no *buffer*.

4.2 Resultados comparativos

A comparação principal recai sobre os tipos de tráfego vídeo e *web*, pertencentes às classificações TR e NRT, respectivamente. Para todos os cenários voz e CBR ocorreram com baixa saturação, gerando impactos mínimos em qualquer um dos escalonadores, devido ao tamanho reduzido de suas requisições em comparação com as outras aplicações, por esse motivo os resultados de voz e CBR não serão apresentados. As métricas coletadas encontram-se entre as apresentadas na Seção 2.3.2, a saber, taxa de transmissão (vazão), taxa de perda de pacotes, latência, *jitter* e índice de justiça. O objetivo aqui reside em demonstrar as capacidades e a coerência do simulador perante as estratégias de alocação já existentes na literatura para ambientes em que a quantidade de usuários empregando aplicações com tráfego NTR torna-se maior que TR. Os mecanismos escolhidos, tais como PF, FLS, EXPR ou LOGR, apresentam maior número de citações entre os trabalhos encontrados no embasamento teórico, presentes nos capítulos 2 e 3 desta pesquisa.

Na Figura 4.1, o eixo X refere-se ao quantitativo de usuários presentes na aplicação e o Y ao valor da vazão em MB/s, nela encontra-se a vazão referente ao cenário para requisições acatadas de vídeo e *web*, respectivamente. Notou-se que os níveis de taxa de transmissão para cada escalonador encontram-se de acordo com o esperado e encontrado na literatura, uma vez que o PF obtém o pior desempenho, enquanto que o FLS possui a melhor performance para vídeo. O crescimento do FLS para vídeo apenas se satura acima de 10 usuários, enquanto os demais, acima de 8; já o PF satura com apenas 5 usuários. Com relação à vazão em si, o FLS apresenta mais de 1,5MB/s, representando em torno de 9% de melhoria comparado ao EXPR ou LOGR em uma amostra de 13 usuários utilizando aplicações de vídeo entre os 50 presentes na rede.

Por outro lado, experiencia-se o padrão contrário para aplicações *web*, uma vez que a priorização do FLS foca somente em tráfegos TR, o que acaba negligenciando quase completamente o tráfego NTR da rede LTE, prejudicando usuários com este tipo de aplicações. Os outros escalonadores possuem desempenho médio entre PF e FLS, com um

certo nível de balanceamento. Quando comparado ao PF, o LOGR possui 95% da performance do primeiro, enquanto o EXPR possui 55% da métrica de vazão, para aplicação *web*. Note que o LOGR obteve melhores taxas no tráfego NTR, enquanto o EXPR em fluxos TR, indicando uma certa tendência de priorização em seus mecanismos. Este padrão foi observado também nos outros subcasos, sendo melhor visualizado no próximo capítulo, em que todos os cenários serão abordados.

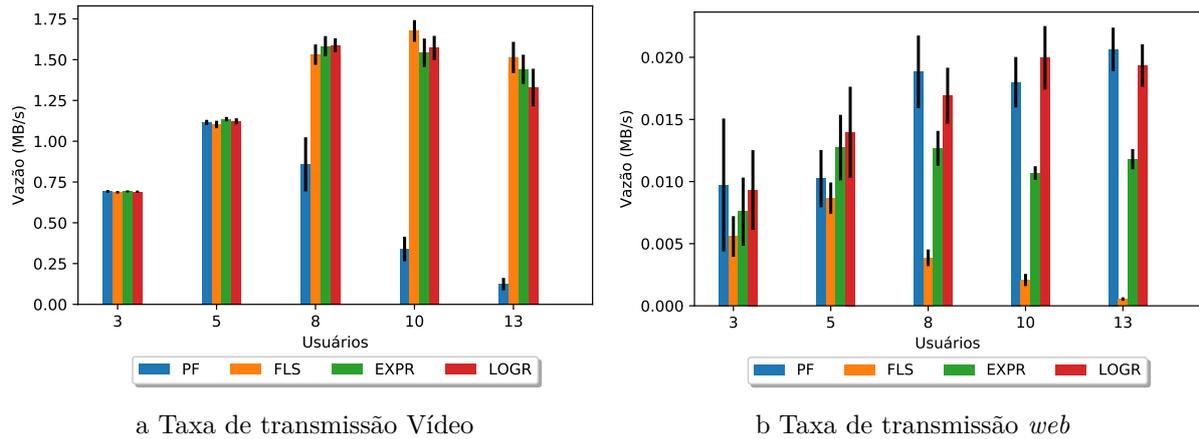


Figura 4.1: Vazão para Vídeo/*Web*.

Com relação à latência, obteve-se um valor alto para o PF, quando comparado às demais políticas de alocação em requisições de vídeo, chegando à marca de 5s com 13 usuários. Para melhor visualização da Figura 4.2, não serão apresentados graficamente os valores do PF para latência em vídeo, devido a demasiada diferença entre os demais dificultarem o reconhecimento dos níveis, lembrando que o eixo Y representa o tempo de latência em segundos e X o quantitativo de usuários da aplicação. Os resultados demonstrados referem-se à latência de vídeo e *web*, respectivamente. Novamente ocorrem situações semelhantes para os escalonadores FLS e PF, em que um se torna o contrário do outro a depender do tipo de tráfego, demonstrando a primordialidade de melhoria de ambos os mecanismos de alocação, pois há a necessidade de manutenção do nível de qualidade de serviço para as aplicações. Ainda que não atenda todos os usuários, ao menos um grau mínimo de atendimento deve coexistir entre os tipos de tráfego.

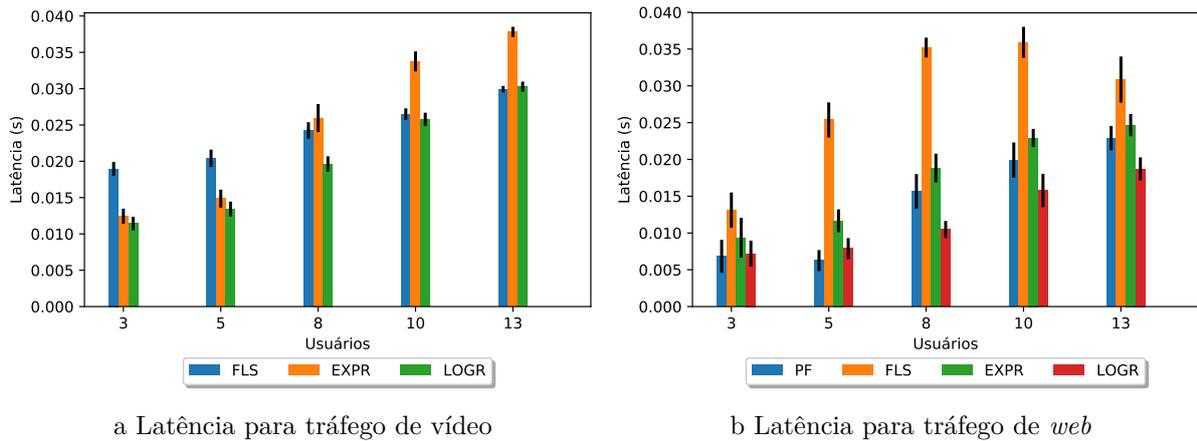
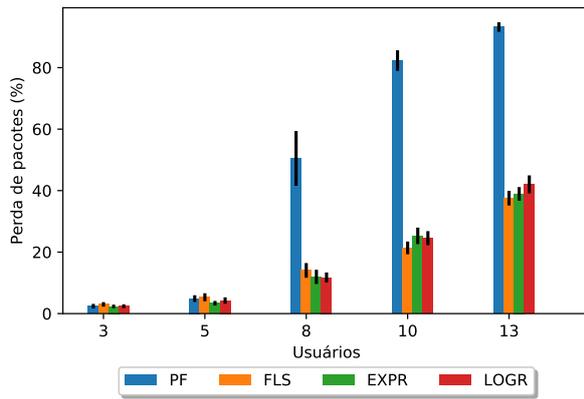


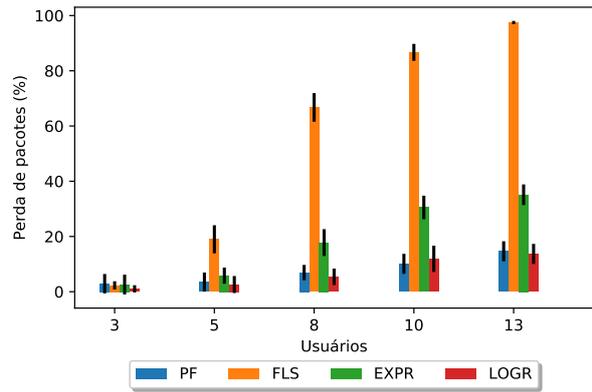
Figura 4.2: Latência para Vídeo/*Web*.

Quanto à taxa de perda de pacotes, pode ser observado o motivo das vazões ocorrerem contrárias para ambos os mecanismos desses escalonadores, conforme explicitado na Figura 4.3, em que o eixo Y representa a porcentagem de perda de pacotes para a aplicação e X o quantitativo de usuários da aplicação. O FLS apresenta quase 100% de perda para *web*, enquanto o PF possui mais de 80% para vídeos, com 13 usuários para ambas as aplicações, ressaltando que a proporção se mantém para outros valores de balanço de carga.

O FLS separa a alocação NTR de TR objetivando fornecer blocos de recurso livres para tráfegos que julga prioritários, entretanto, se não houver mais blocos livres, a alocação das outras requisições serão desconsideradas, desta forma não ocorre a mistura entre as filas das requisições, o que resulta em uma alta taxa de perda de pacotes para NTR em cenários com heterogeneidade de requisições entre os tipos de fluxos de dados. Por outro lado, o PF não possui distinção entre as filas, tampouco existe em sua estratégia mecanismos que busquem priorização por latência. Utilizando apenas a qualidade do canal de cada usuário como fator para determinar o peso da alocação, resulta em perdas de pacotes maiores para usuários que requisitam muitos dados e não estão em localidade favorável. Os outros escalonadores possuem métricas entre os extremos das estratégias PF e FLS, e, portanto, mostram-se capazes de manter um nível tolerável para ambas as classificações de tráfego.



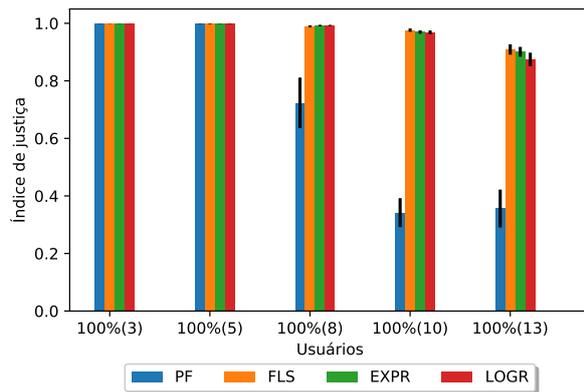
a Perda de pacotes para tráfego de vídeo



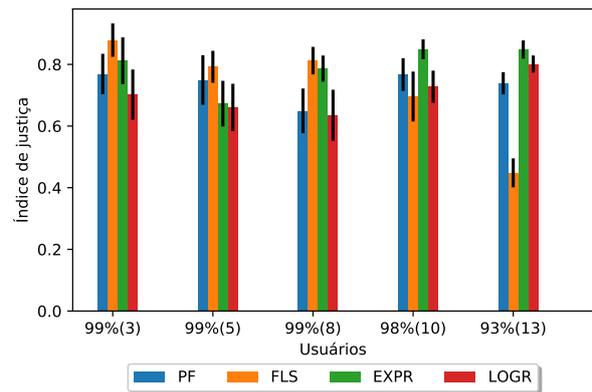
b Perda de pacotes para tráfego de *web*

Figura 4.3: Taxa de perda de pacotes para Vídeo/*Web*.

Para a métrica de índice de justiça, observou-se um valor relativamente próximo para todos os escalonadores, exceto, novamente, para PF no caso de vídeo e FLS para a aplicação *web*, demonstrando que os usuários atendidos possuem uma condição próxima de qualidade de serviço para a maioria dos escalonadores utilizados no comparativo. A Figura 4.4 ilustra o comportamento desta última métrica para os fluxos de vídeo e *web*, as marcações do eixo *X* compreendem a média da porcentagem dos usuários atendidos entre todos os escalonadores, seguido pelo total de usuários para aquele fluxo entre parênteses, já o eixo *Y* refere-se ao valor do índice de justiça (Jain), Equação 2.10, obtido para o mecanismo de alocação.



a Índice de justiça para tráfego de vídeo



b Índice de justiça para tráfego de *web*

Figura 4.4: Índice de justiça para Vídeo/*Web*.

Apesar do índice de justiça ser, na maioria dos casos, acima de 0,6, ele se encontra definido de acordo com a vazão média por usuário, assim, pode ocorrer de alguns usuários terem apenas alguns de seus pacotes alocados, entretanto, a média de vazão ser semelhante

aos que tiveram todos os pacotes atendidos. Para melhor visualizar esse comportamento, deve-se levar em conta a perda de pacotes para o fluxo, pois um bom índice de justiça com grande perda de pacotes pode ocultar um problema de vazão na rede. No gráfico, para 13 usuários, percebe-se um atendimento médio de 93%, entretanto, no caso do FLS este valor cai para apenas 70%, evidenciando que em torno de 4 usuários não receberam sequer algum pacote, o que corrobora a alta taxa de perda de pacotes. Os valores de atendimento de usuários podem ser melhor observados na Tabela 4.2, denotando a quantidade total de usuários e a porcentagem representada em relação aos atendidos. Os valores obtidos para a proporção evidenciada nessa seção, relacionada à aplicação de vídeo, mostram que 100% dos usuários receberam ao menos um pacote, entretanto ao se checar a métrica para a aplicação *web*, observa-se a existência de usuários que sequer receberam algum pacote.

Tabela 4.2: Atendimento de usuários para tráfego *Web*.

Escalonador	Subcasos									
	10		20		30		40		50	
	Total	Percentual	Total	Percentual	Total	Percentual	Total	Percentual	Total	Percentual
PF	3	98%	5	99%	8	100%	10	99%	13	97%
FLS	3	100%	5	99%	8	95%	10	93%	13	70%
EXPR	3	100%	5	100%	8	98%	10	98%	13	96%
LOGR	3	100%	5	100%	8	99%	10	99%	13	98%

4.3 Discussão

A comunidade acadêmica conta com diversas soluções e propostas para o problema de alocação dos blocos de recurso, porém, na maioria dos trabalhos as comparações encontram-se entre os algoritmos clássicos, tais como *Proportional and Fair*, *Maximum Throughput* e *M-LWDF* [3]. Além disso, existem muitos fatores, parâmetros e métricas que alteram o desempenho de um escalonador, o que torna a comparação entre artigos de difícil precisão, pois nem todos permeiam todas as estratégias de alocação ou utilizam cenários semelhantes.

Dentre as políticas, uma das abordagens que se mostra mais promissora intercorre pelo FLS. Em seu artigo Piro *et al.* [8] demonstram um cenário configurado com grande parte do tráfego de tempo real, o que proporciona vantagem à estratégia *Frame Level Scheduler*, devido a seus métodos de funcionamento, priorizando a classificação deste tipo de tráfego e postergando a alocação do tipo não real. Devido a esse mecanismo, a eficiência da política em um contexto no qual o tráfego de tempo real torna-se menor ou, em caso extremo, nulo, observa-se um desempenho semelhante ao *Proportional and Fair* [45] para tráfego de tempo real, que possui um dos piores resultados devido a sua simplicidade.

A abordagem de postergar tráfegos menos prioritários revela-se útil, porém deve ser utilizada de maneira mais eficiente, como no caso da abordagem dos algoritmos *Exponential Rule* ou *Logarithmic Rule* [9], que utilizam métricas de latência contidas na portadora para definir qual fluxo será alocado. Todavia, apesar de possuírem mecanismos que priorizam as métricas de latência, ainda fazem com que tráfegos não tempo real como *web* ou CBR transcorram por outro tipo de escalonador. Logo, as vantagens contidas nos algoritmos apresentam-se inadequadas para fluxos NTR, acarretando em um desempenho, possivelmente, inferior, pois o escalonador utilizado para a segunda classificação de tráfego consiste do *Proportional and Fair*, que pode ser suficiente para fluxos sem grandes restrições de latência.

Com as informações aqui apresentadas, compreende-se que os escalonadores divergem os seus resultados conforme a estratégia de alocação utilizada em seu núcleo. Além disso, apesar de poderem obter desfechos semelhantes, existem cenários que permitem métricas de melhor desempenho para cada um, por exemplo o FLS pode não se sair tão bem em momentos com muito tráfego não tempo real, mesmo possuindo o melhor resultado para a outra classificação. Observou-se que a grande maioria dos artigos presentes na subseção 3.1 possuem como objetivo melhorar o desempenho da política de alocação de blocos de recurso por meio da priorização de tráfego de tempo real, porém mesmo este tipo de tráfego requisitando, em geral, a maior fatia percentual de vazão de dados, pode não ser o maior em relação à quantidade de usuários utilizando-o simultaneamente. O crescimento de IdC [47], em especial sua utilização por serviços REST ou outros meios HTTP, simboliza um aumento do tráfego *web*, que por sua vez não contempla a priorização abordada na maioria dos trabalhos, pois representa um aumento na parcela de tráfego NTR, possuindo um volume e requisitos distintos. No próximo capítulo, analisa-se uma nova abordagem, cujo objetivo reside em não exaurir os fluxos de dados NTR, a fim de obter resultados semelhantes em boa parte dos casos, além de melhorar o desempenho em certos subcasos, como exemplificado neste capítulo, em que o FLS atingiu 100% de perda de pacotes.

Capítulo 5

Alocação de blocos de recurso sensível à presença de fluxos NTR

Na presente pesquisa, busca-se uma nova política de alocação não persistente que contenha parâmetros para qualidade de canal e serviço, baseada em prioridade por latência, com o intuito de melhorar o atendimento a usuários que requisitam fluxos de dados NTR. Neste capítulo, sugere-se uma nova estratégia de alocação de recursos que utiliza teoria de controle para redefinição da quantidade de dados enviados por cada *frame*, baseada no FLS [8], associada a uma modificação na política de alocação proveniente do EXPR e LOGR, escalonadores abordados na Seção 2.3. Realiza-se, também, a comparação da proposta às outras, demonstrando pontos favoráveis e desfavoráveis de cada estratégia por meio de cenários pré-configurados em diferentes condições. A avaliação dos mecanismos correspondem às métricas explanadas na Seção 2.3.2, que abrangem diferentes atributos contidos na alocação de blocos de recurso e na transferência de pacotes em qualquer tipo rede.

5.1 Proposta de alocação sensível à heterogeneidade dos fluxos de dados

Ao se avaliar o modelo do FLS e os resultados obtidos no capítulo anterior, notou-se que a lei de controle reorganizadora da quantidade de dados a ser transmitidos encontra-se utilizada somente na classificação de tráfego tempo real, por este motivo aplicações de vídeo, por exemplo, alcançam uma vazão superior quando comparadas ao mesmo fluxo em outros escalonadores. Observou-se, também, que o tráfego NTR sofre de uma grande taxa de perda de pacotes, devido à estratégia utilizada na segunda camada, em que se prioriza a alocação da primeira classificação, distribuindo os blocos de recurso apenas

se ainda existirem recursos livres após alocação prioritária. Em análise ao modelo das estratégias EXPR ou LOGR, notou-se um esquema de priorização também focado em tráfegos tempo real, utilizando as métricas de latência apenas para este tipo de tráfego. Para NTR empregou-se a pesagem do *Proportional and Fair*, que não considera métricas além da qualidade do canal, prejudicando levemente o tráfego NTR, apesar de atender de maneira razoável.

A fim de otimizar o FLS para certos cenários, como o presente no Capítulo 4, no qual existe uma quantidade de usuários requisitando dados do tipo NTR, maior ou igual a TR, se faz necessário modificar o mecanismo que promove a exaustão do tráfego não tempo real na rede LTE, fato que prejudica a satisfação e fere a qualidade de serviço para os usuários. Diante disso, propõe-se uma nova política de alocação de recursos baseada em priorização por latência, auxiliada de teoria de controle para ajustar a quantidade de dados enviados em cada quadro LTE, sem separar as classificações TR e NTR em diferentes filas ou mecanismos de alocação, baseando-se, principalmente, na latência dos fluxos para o atendimento.

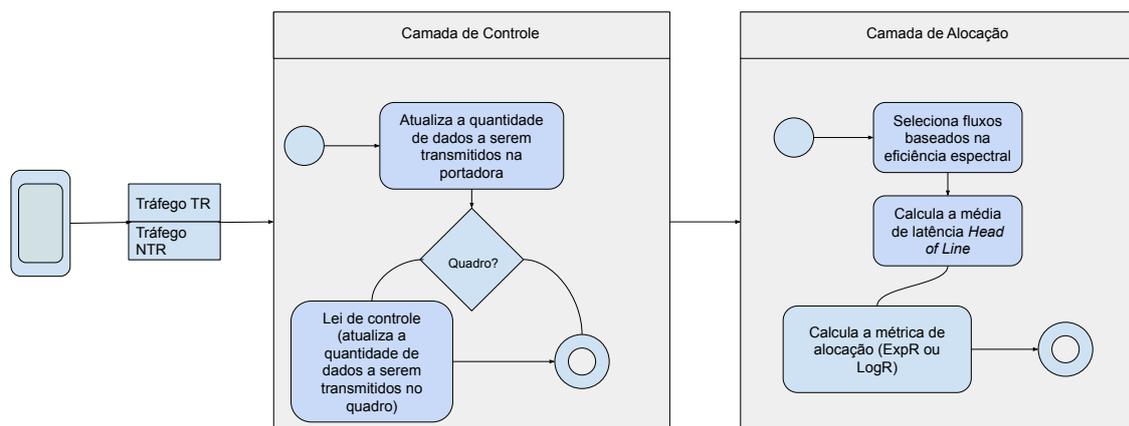


Figura 5.1: Fluxograma da proposta para política de alocação de recursos.

A implementação, representada na Figura 5.1 por meio de um fluxograma de sistema, retrata as etapas necessárias para a utilização da proposta em questão. Nesta política de alocação de blocos de recurso, os tráfegos passam pelos mesmos fluxos dentro do escalonador, promovendo uma alocação mais heterogênea, uma vez que a priorização depende da latência e da qualidade do canal experienciada por cada requisição do usuário. Esta implementação revela-se capaz de eliminar a escassez de recursos para o tráfego não tempo real, mantendo os níveis de qualidade de serviço estipulados para a latência. Os resultados obtidos encontram-se demonstrados na Seção 5.2.

A proposta em questão apresenta-se dividida em duas camadas principais, Controle e Alocação, cada uma com seus próprios objetivos. A camada superior utiliza a ideia proveniente do FLS [8], utilizando teoria de controle e implementada como especificado por Piro *et al.*, conforme exposto previamente no Capítulo 2. A redefinição da quantidade de dados a ser transmitidos de cada requisição ocorre no começo de cada quadro LTE, dessa forma, para executar o passo da atualização, basta saber se o intervalo de transmissão (TTI) atual pode ser divisível por 10. A camada superior tem como objetivo esta adaptação (atualização) para aproveitar da capacidade dos blocos de recurso, assim como aumentar a garantia da qualidade de serviço da latência dos pacotes. Esta camada se mostra bem semelhante ao FLS original, exibido na Figura 2.9, porém modifica o sistema para incluir os tráfegos NTR durante o momento de execução da lei de controle. Piro *et al.* [8] apresentam algumas restrições ao modelo e suas provas. O filtro do sistema h , filtra cada tráfego i para sobrescrever a quantidade de dados a preencher a fila de tráfego além disso, a fila filtrada não pode conter mais dados que a entrada (um pulso de largura 1), satisfazendo

$$h(k) = \sum_{n=0}^M c_i(n)d(k-n), \quad (5.1)$$

representando a resposta em até M passos, para que seja possível obter tráfegos limitados pela latência em até $M + 1$ intervalos, sendo d a quantidade de dados a preencher a fila de tráfego, os coeficientes c_i modelam a forma como os tráfegos serão distribuídos nos intervalos.

O segundo nível, camada inferior, possui a alocação dos blocos de recurso em si. As políticas escolhidas para os testes desta proposta foram EXPR e LOGR, sendo possível utilizar outras estratégias com algumas modificações. Utilizar estratégias que busquem o balanceamento da fila ou latência melhora a diversidade dos usuários, permitindo maiores oportunidades de aumento na taxa de transmissão, aumentando a justiça entre os usuários e entregando os requisitos de qualidade de serviço. De acordo com Sadiq *et al.* [9], ambos os mecanismos selecionados possuem as seguintes propriedades:

$$D_{HOL,i} \leq \tau_i, \quad (5.2)$$

de modo que o tempo de espera do primeiro pacote da fila $D_{HOL,i}$ do i -ésimo fluxo deve ser menor que o tempo limite de latência τ imposto pela qualidade de serviço, assim como

$$\sum_{i=0}^N r_{i,k} \leq \sum_{k=0}^M R_k, \quad (5.3)$$

de modo que o somatório da vazão instantânea $r_{i,k}$ dos fluxos alocados em cada bloco de

recurso deve ser menor que a capacidade de vazão do canal considerando cada bloco de recurso com o melhor *Modulation Coding Scheme* (MCS), pacotes que não respeitam estas limitações serão descartados. Apesar de possuírem as mesmas limitações, o LOGR tende a diminuir o balanço da fila em favor ao aumento da taxa de transmissão, satisfazendo:

$$q(t) \in \arg \max_{1 \leq i \leq N} \log (\alpha_i D_{HOL,i}) \Gamma_t^i, \quad (5.4)$$

lembrando que $\alpha_i = -\frac{\log \delta_i}{\tau_i}$, em que δ_i representa a probabilidade do atraso $D_{HOL,i}$ atingir o limite de latência e $q(t)$ um fluxo da fila em determinado tempo t . Já o EXPR faz o tratamento inverso, dando maior ênfase ao balanço da fila, satisfazendo:

$$q(t) \in \arg \max_{1 \leq i \leq N} \exp \left(\frac{\alpha_i D_{HOL,i}}{1 + \sqrt{\chi}} \right) \Gamma_k^i, \quad (5.5)$$

ressalta-se que $\chi = \frac{1}{N} \sum_{i=1}^N \alpha_i D_{HOL,i}$, e denota-se por N o número de fluxos de dados em ambos os mecanismos. Após cada intervalo de transmissão a fila será decrementada de acordo com os tráfegos alocados e as latências $D_{HOL,i}$ ajustadas para refletir a alocação e servirem de entrada para o próximo intervalo. Note que fez-se necessário uma modificação para remover o *Proportional and Fair* de ambos os escalonadores, esta alteração ocorreu com o propósito de unificar as prioridades de alocação de blocos de recurso por meio de latência para qualquer tipo de fluxo, promovendo maior heterogeneidade ao escalonador e oportunizando alocação para pacotes NTR quando próximos ao limite de qualidade de serviço.

5.2 Resultados

A partir da implementação dos métodos citados previamente, utilizou-se o cenário encontrado no capítulo anterior, com o objetivo de averiguar a capacidade e a eficiência de cada política existente, comparando-as aos mecanismos recém elaborados. A velocidade de movimento dos usuários foi fixada em 0 km/h e 30 km/h, para o ambiente estático e com mobilidade, respectivamente, a fim de confirmar a eficácia das soluções em diferentes situações. O modo de mobilidade escolhido para o ambiente móvel foi o *Random Waypoint*, no qual o usuário começa em uma posição aleatória como descrito e caminha até um ponto pré-determinado dentro do raio da célula, escolhendo um novo destino ao finalizar o movimento.

Dentro dessa seção, criou-se cinco subcenários, nos quais os tráfegos variam em porcentagem de uso na rede, iniciando em 0% tempo real e 100% não tempo real e finalizando no inverso, com incrementos/decrementos de 25%. A Tabela Tabela 5.1 apresenta um re-

ferencial para todos os parâmetros utilizados, expandindo a Tabela 4.1 para as seguintes proporções de tráfego: 0/100, 25/75, 75/25 e 100/0, além disso inclui os parâmetros para o cenário com mobilidade, alterando somente a configuração de velocidade em relação ao ambiente estático.

Tabela 5.1: Parâmetros do cenário.

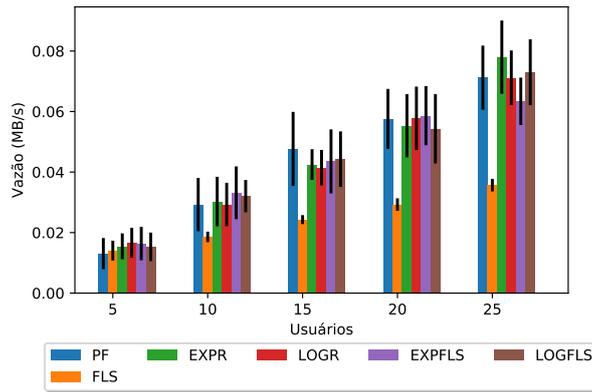
Parâmetros	Valor	Descrição
Largura de banda	1400	Faixa de frequência em Hertz.
# Blocos de recurso	6	Quantidade de blocos de recurso da torre.
# Usuários	10,20,30,40,50	Quantidade de usuários.
Tipos de Tráfego	-	Tempo real (Vídeo e VoIP), Não tempo real (<i>web</i> e CBR).
Porcentagem de tráfego	0/100,25/75,50/50,75/25,100/0	Porcentagem dos tipos de tráfego.
Latência máxima	Vídeo - 60, VoIP - 40 <i>web</i> - 80, CBR - 100	Latência máxima para cada pacote em milissegundos.
Velocidade	0, 30	Velocidade em que os usuários se movem em km/h.
Raio da célula	500	Tamanho do raio da célula da torre em metros.
Posição dos usuários	-	Posição aleatória dentro do raio da célula para cada usuário.
Duração de simulação	100	Duração do tempo de simulação em segundos.
Taxa de processamento vídeo	242	<i>Bitrate</i> do vídeo a ser transmitido em kb/s

5.2.1 Ambiente estático

Nessa seção, dividiu-se o cenário do capítulo anterior em subcasos distintos de acordo com a porcentagem dos tipos de tráfego presentes na simulação, assim pôde-se ver a eficácia dos mecanismos de alocação de blocos de recursos para diferentes cargas na rede. Nesse ambiente, foram considerados subcasos enumerados de 1 a 5, cada cenário estima uma proporção de tráfegos de tempo real e não tempo real, seguindo os valores 0%, 25%, 50%, 75% e 100%.

Cenário 1 (0% TR - 100% NTR)

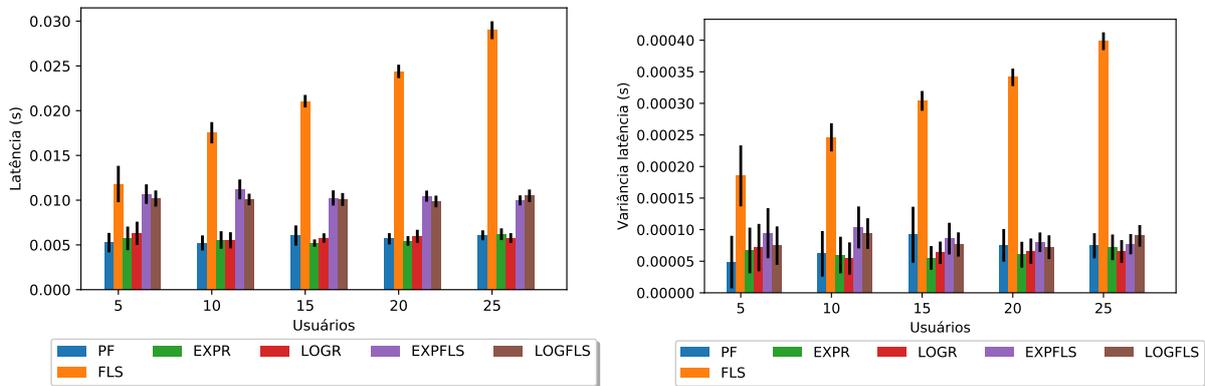
Neste cenário inicial, a maior parte das políticas apresentou desempenho semelhante, entretanto a estratégia FLS exibiu um resultado inferior mesmo quando o tráfego de tempo real não se encontrava presente na rede. Analisando as outras políticas, notou-se um crescimento não saturado com relação à vazão, como revelado na Figura 5.2, com certas variações ao longo do caminho devido à qualidade do canal experienciada pelos usuários, possuindo performances equivalentes. Apesar de todos os escalonadores ainda possuírem uma certa curva de crescimento, o FLS apresentou uma curva menos acentuada. As implementações propostas, EXPFLS e LOGFLS, possuem níveis próximos aos demais, variando em torno de 5%. Observou-se, ainda, que o LOGFLS se sobressaiu em relação ao EXPFLS para o fluxo *web*. Nesse ambiente, em específico, a posição dos usuários para o EXPR demonstrou-se melhor, promovendo um aumento no indicador da qualidade do canal, proporcionando um pequeno aumento (8%) em relação ao LOGR.



a Taxa de transmissão *web*

Figura 5.2: Vazão para *Web*.

Em análise à quantidade de pacotes perdidos, ficou evidente o motivo pelo qual o FLS possui métricas tão abaixo dos outros escalonadores para este subcaso, atingindo uma taxa de perda de pacotes de mais de 16%, enquanto os seus concorrentes mal passaram de 2%, representados na Figura 5.4. Percebeu-se que apesar de, inicialmente, o valor ter aumentado para as novas estratégias, logo se estabilizou e manteve-se com desempenho mais próximo dos anteriores. Com relação à latência, representada pela Figura 5.3, notou-se casos semelhantes à taxa de perda de pacotes, com as novas políticas apresentando pouca elevação comparada às demais, e ainda bem abaixo do FLS.



a Latência

b Jitter

Figura 5.3: Latência e *jitter* para *Web*.

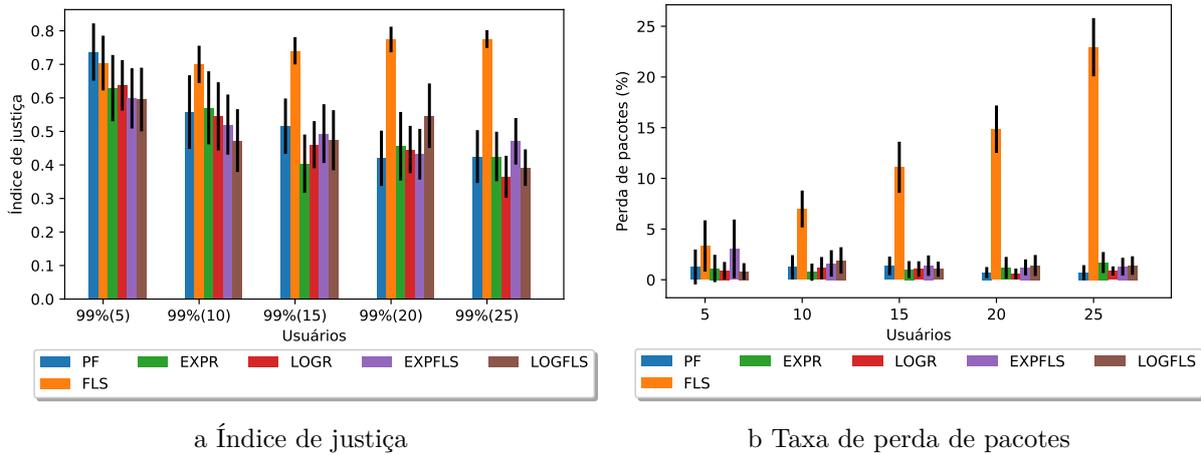


Figura 5.4: Índice de justiça e perda de pacotes para *Web*.

Como já citado no capítulo anterior, o índice de justiça de Jain não se refere à única medida de equidade que pode ser analisada. A Tabela 5.2 exibe o percentual de usuários atendidos para o tráfego *web*. Alerta-se que para o fluxo de vídeo não será apresentada essa métrica, para esse ou qualquer outro cenário, posto que ao menos um pacote de cada usuário sempre foi atendido. Assim, no cenário 1, observou-se que, com 100% de classificação não tempo real, ocorreu um mínimo de 97% de atendimento aos usuários, denotando que em média aproximadamente 24 dos 25 usuários foram atendidos para o fluxo em questão, mas vale lembrar que em relação ao quantitativo de pacotes atendidos há mais de 20% de pacotes perdidos. As novas políticas se mostraram equiparadas quanto a esta métrica em relação aos outros escalonadores.

Tabela 5.2: Atendimento de usuários tráfego *Web*.

Escalonador	Subcasos									
	10		20		30		40		50	
	Total	Percentual	Total	Percentual	Total	Percentual	Total	Percentual	Total	Percentual
PF	5	99%	10	100%	15	99%	20	100%	25	100%
FLS	5	100%	10	99%	15	98%	20	97%	25	97%
EXPR	5	100%	10	100%	15	100%	20	99%	25	99%
LOGR	5	100%	10	100%	15	100%	20	99%	25	99%
EXPFLS	5	99%	10	100%	15	99%	20	99%	25	99%
LOGFLS	5	100%	10	99%	15	99%	20	99%	25	99%

Cenário 2 (25% TR - 75% NTR)

Neste cenário, tornou-se mais evidente o decaimento do tráfego NTR para o escalonador FLS, mantendo padrões semelhantes ao primeiro, demonstrando pior vazão, latência e

perda de pacotes. Constatou-se também o comportamento inverso no FLS e PF, em que PF apresentou um comportamento parecido, porém para o fluxo contrário de tempo real.

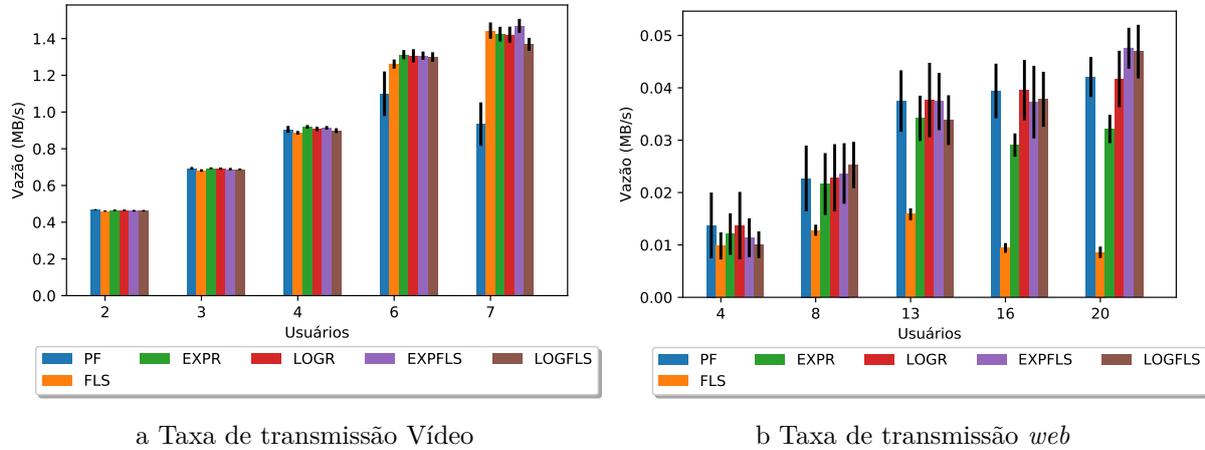


Figura 5.5: Vazão para Vídeo/*Web*.

Na Figura 5.5, verificou-se a taxa de transmissão média das políticas, de acordo com a quantidade de usuários. Observou-se a ocorrência de algumas discrepâncias devido à posição dos usuários, posto a localização ocorrer de maneira aleatória, promovendo uma qualidade do canal experienciada maior, gerando aumento da vazão. Todavia, a diferença entre as novas estratégias se revelou menor, apesar de ainda não superar completamente o estado da arte. Verificou-se, para o segundo cenário, melhores desempenhos de vazão em vídeo e em algumas situações de *web*. A melhoria em questão reflete na vazão, girando por volta de apenas 1% comparando o EXPFLS ao FLS para vídeo, e 500% quando comparado ao mesmo escalonador para tráfego *web*. Quanto aos resultados perante as políticas de alocação LOGR ou EXPR, apesar de menos expressivos ainda os superam, com desempenho médio acima de 30% para *web*.

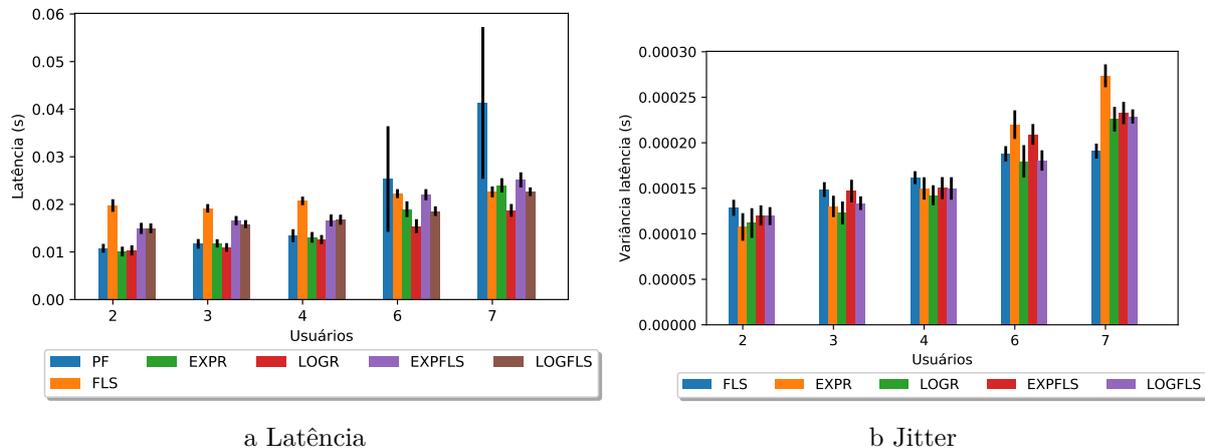


Figura 5.6: Latência e *jitter* para Vídeo.

Quanto à latência e sua variância, representadas nas Figuras 5.6 e 5.7, assim como quanto ao índice de justiça e à perda de pacotes, representados nas Figuras 5.8 e 5.9, verificou-se métricas muito próximas para todos os escalonadores. A maior diferença recaiu entre EXPFLS e LOGR (desconsiderando o PF) no caso de vídeo, atingindo um aumento de quase 40%, porém ainda satisfazendo o requerimento da qualidade de serviço (60ms). Já para *web*, desconsiderando o FLS, ocorreu uma situação similar, sendo o aumento, agora, em torno de 200%, porém muito abaixo do valor de 100ms da qualidade de serviço de serviço, com o maior valor médio atingindo cerca de 0,02ms.

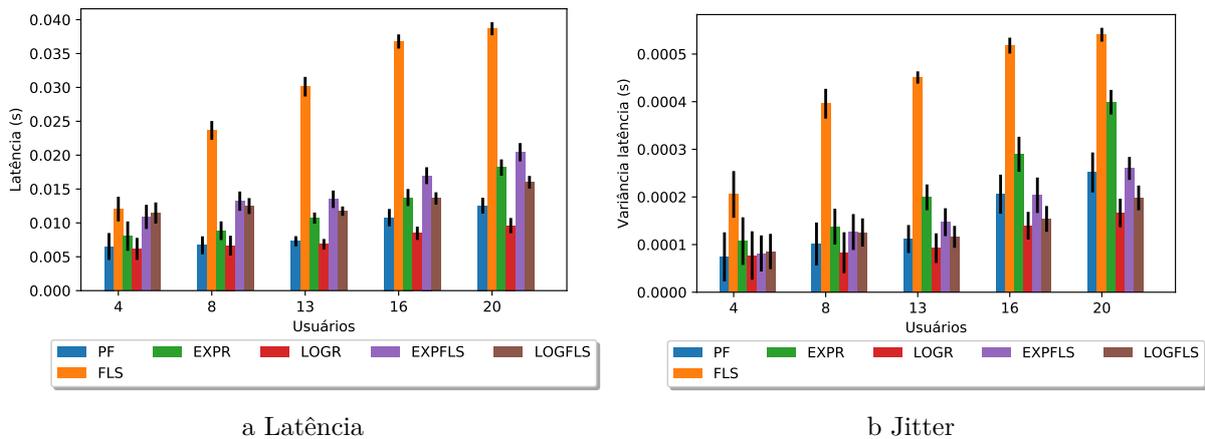


Figura 5.7: Latência e *jitter* para *Web*.

Por fim, quanto à taxa de perda de pacotes e ao índice de justiça, percebeu-se situações contrárias, mas esperadas. Enquanto o FLS se mostrou o melhor na perda de pacotes de Vídeo, revelou-se o pior para *web*. Além disso, observou-se que as novas estratégias apresentaram uma taxa de perda de pacotes semelhante a do EXP e do LOGR, por volta dos 15% para vídeo e 5% para *web*.

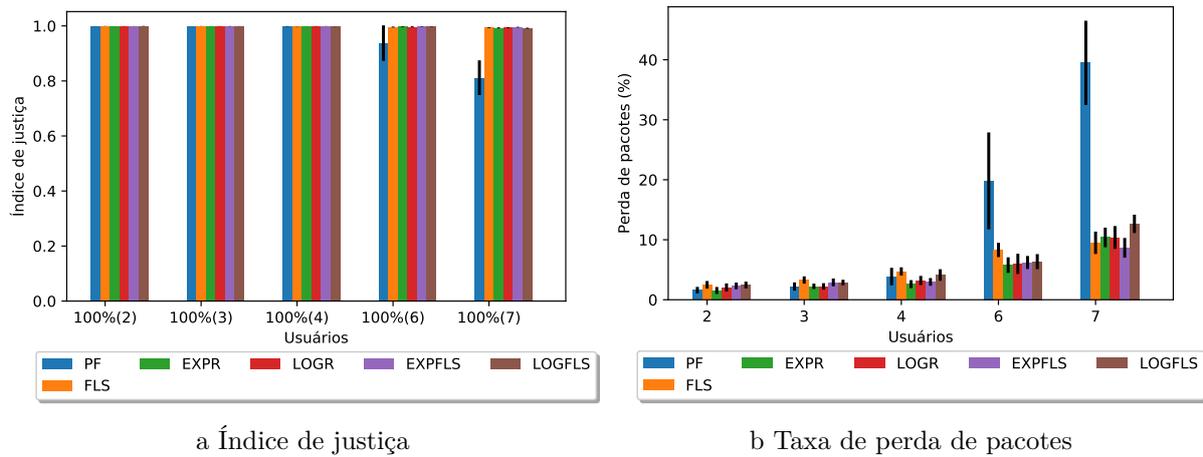


Figura 5.8: Índice de justiça e taxa de perda de pacotes para Vídeo.

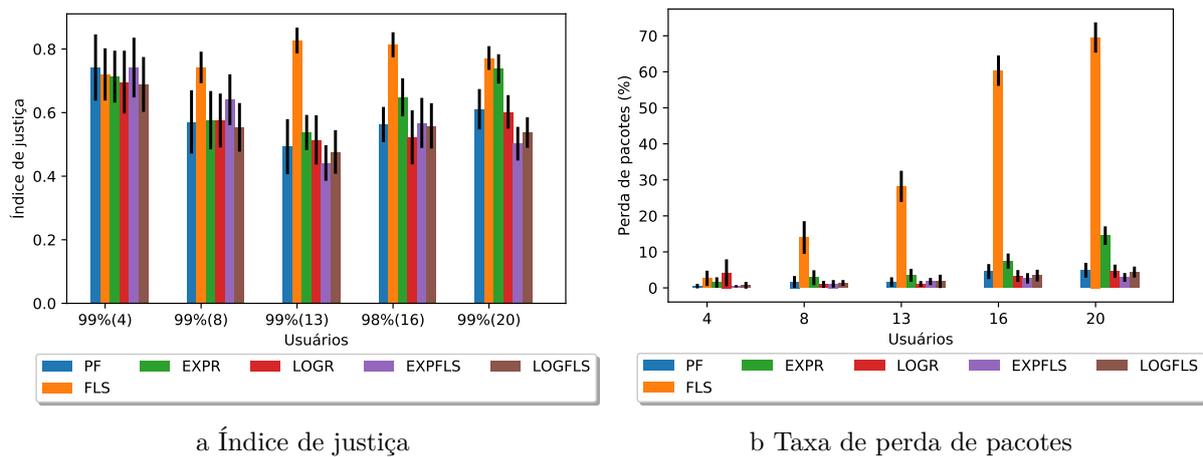


Figura 5.9: Índice de justiça e taxa de perda de pacotes para Web.

Para o segundo cenário, não ocorreram grandes diferenças entre os percentuais de usuários atendidos, abaixando somente 1% com relação ao cenário 1. As novas políticas se comportaram de maneira eficiente, mantendo-se equiparadas ou levemente melhor que os outros escalonadores, conforme as informações apresentadas na Tabela 5.3.

Tabela 5.3: Atendimento de usuários tráfego *Web*.

Escalonador	Subcasos									
	10		20		30		40		50	
	Total	Percentual	Total	Percentual	Total	Percentual	Total	Percentual	Total	Percentual
PF	4	100%	8	100%	13	100%	16	99%	20	99%
FLS	4	100%	8	98%	13	97%	16	96%	20	96%
EXPR	4	100%	8	99%	13	100%	16	100%	20	99%
LOGR	4	97%	8	100%	13	100%	16	98%	20	100%
EXPFLS	4	100%	8	100%	13	100%	16	99%	20	100%
LOGFLS	4	100%	8	100%	13	99%	16	99%	20	99%

Cenário 3 (50% TR - 50% NTR)

Com 50% de usuários utilizando cada classificação de fluxo de dados ocorreu um decaimento mais rápido, por exemplo, com 10 a 13 usuários já se evidenciou uma saturação na rede para o tráfego de vídeo. A Figura 5.10, ilustra a vazão para Vídeo e *web* respectivamente, verificou-se a ocorrência de declínio um pouco mais agravado para as novas estratégias ao atingir 13 usuários para vídeo. Entretanto, nesse cenário, uma mesma quantia de dispositivos utilizaram ambas as classificações de tráfego, notou-se, então, a vantagem das propostas para o tráfego NTR, especificamente para *web*, em que há um crescimento maior quando comparado às outras políticas, um crescimento médio atingindo 200% comparado ao EXPR, e em torno de 15% quanto ao LOGR.

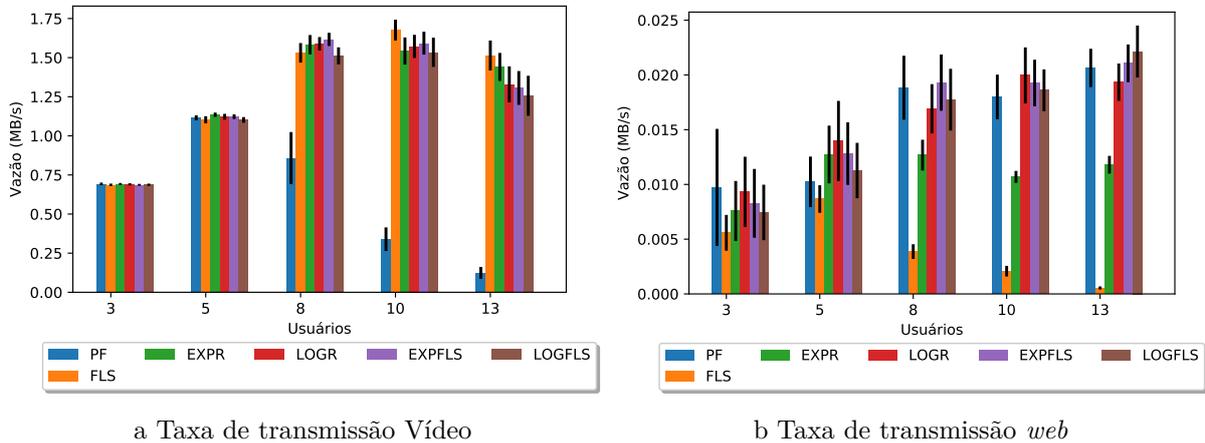


Figura 5.10: Vazão para Vídeo/*Web*.

Nesse cenário, encontrou-se uma leve vantagem referente às novas estratégias para o tráfego *web*, em geral elas permaneceram similares às outras e mantiveram-se melhores para vídeo, demonstrando o ganho por utilizar a técnica de teoria de controle do FLS

sobre os tráfegos não tempo real, o que ocasionou maior visibilidade para o mecanismo elaborado neste capítulo.

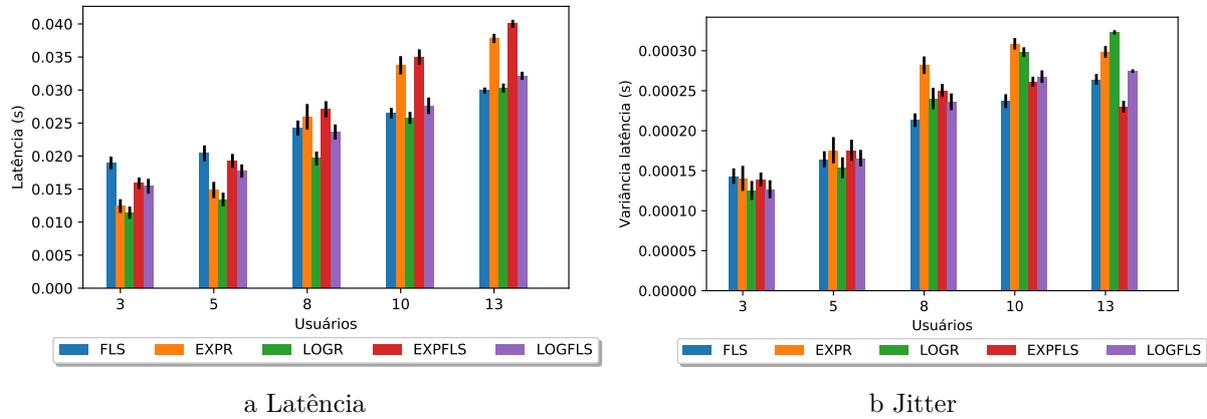


Figura 5.11: Latência e *jitter* para Vídeo.

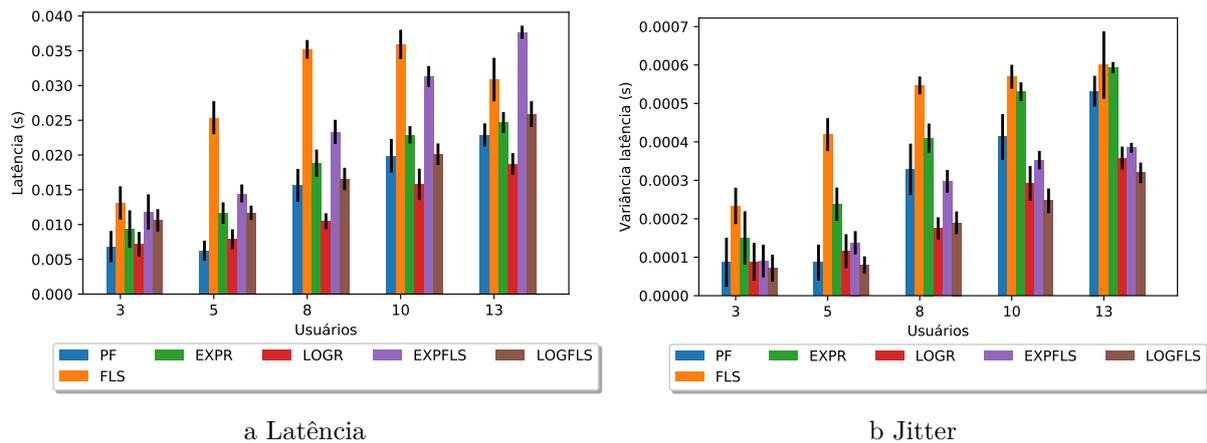


Figura 5.12: Latência e *jitter* para Web.

A variância se manteve baixa, atingindo um máximo de 0,0006 para *web*, o que a tornou quase que insignificante no impacto das outras métricas, a latência respeitou os valores indicados pela qualidade de serviço para ambos os casos, sendo que o LOGFLS manteve-se, em geral, abaixo do FLS original, comportamento ilustrado nas Figuras 5.13 e 5.14.

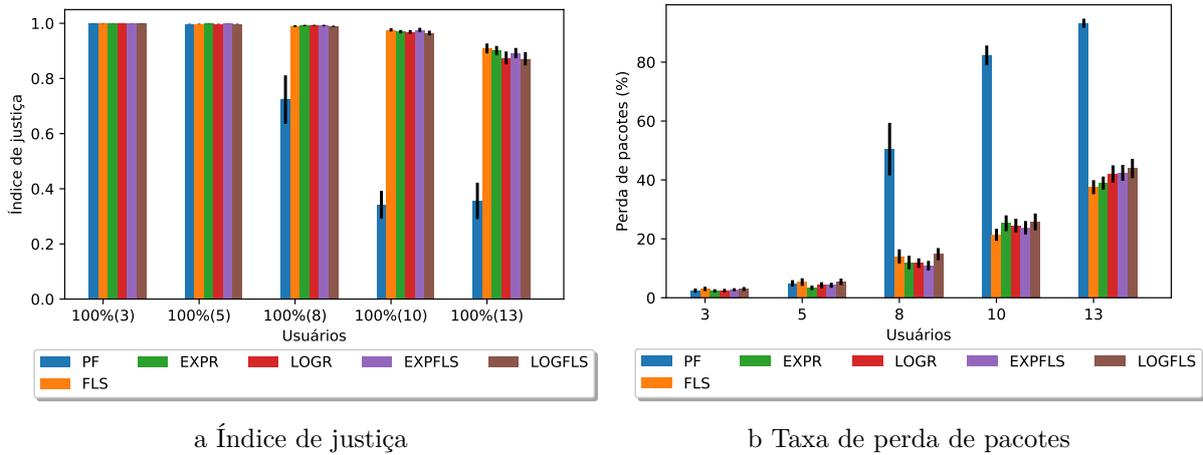


Figura 5.13: Índice de justiça e taxa de perda de pacotes para Vídeo.

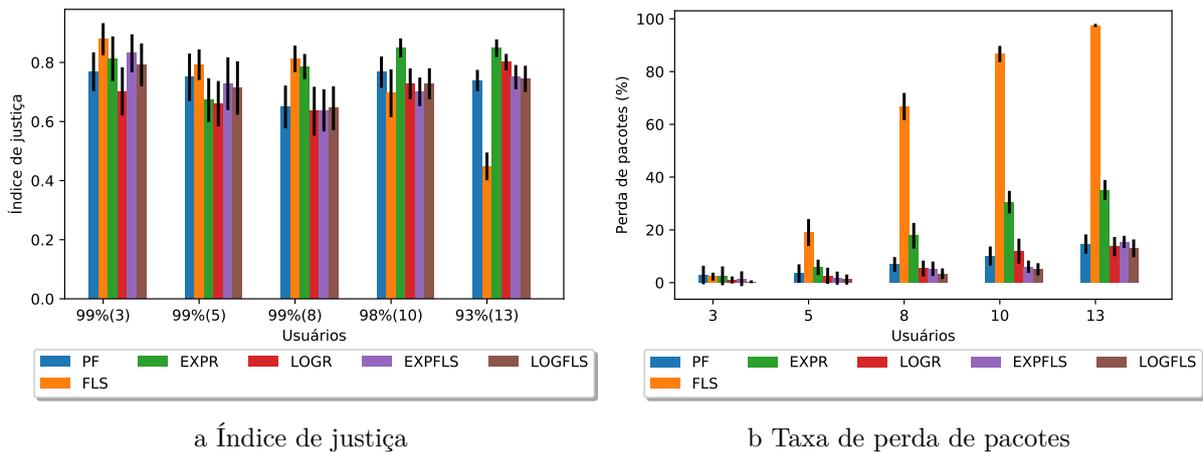


Figura 5.14: Índice de justiça e taxa de perda de pacotes para Web.

Para o índice de justiça e percentual de perda de pacotes, verificou-se valores semelhantes aos cenários anteriores, evidenciando o mesmo padrão para FLS e PF. As novas políticas de alocação se mostraram levemente piores em relação à justiça para o tráfego *web*, porém com menor taxa de perda de pacotes, conseguindo, assim, atingir um maior número de requisições e aumentando levemente a vazão deste fluxo na rede, como demonstrado nas Figuras 5.11 e 5.12.

A partir do cenário 3, o FLS começou a demonstrar uma maior desvantagem em relação ao percentual de atendimento aos usuários, aqui para *web* já ocorreu de apenas 70% receberem ao menos um pacote, representando aproximadamente 9 usuários do universo dos 13. Os outros escalonadores ainda mantiveram-se em uma média de 97%, como visualizado na Tabela 5.4.

Tabela 5.4: Atendimento de usuários tráfego *Web*.

Escalonador	Subcasos									
	10		20		30		40		50	
	Total	Percentual	Total	Percentual	Total	Percentual	Total	Percentual	Total	Percentual
PF	3	98%	5	99%	8	100%	10	99%	13	97%
FLS	3	100%	5	99%	8	95%	10	93%	13	70%
EXPR	3	100%	5	100%	8	98%	10	98%	13	96%
LOGR	3	100%	5	100%	8	99%	10	99%	13	98%
EXPFLS	3	100%	5	99%	8	100%	10	99%	13	97%
LOGFLS	3	100%	5	100%	8	100%	10	99%	13	99%

Cenário 4 (75% TR - 25% NTR)

O penúltimo subcaso, representado pelo cenário 4, caracterizou-se por um declínio agravado na métrica de vazão para os 13 usuários, pois o fluxo de vídeo se mostrou muito mais rígido em relação a suas métricas quando comparado ao fluxo *web*, necessitando de requisitos de qualidade de serviço mais restritos. Apesar disso, a vazão *web* para as novas políticas de alocação de blocos de recursos ainda se mostrou melhor que as anteriores, com exceção do PF, que se demonstrou como o melhor para tráfego NTR, como visto na Figura 5.15. Vale ressaltar que nesse caso, a redução da vazão no tráfego de vídeo perante o aumento de usuários foi levemente mais lenta para as novas políticas quando comparado às outras. Por outro lado, houve uma certa estabilização da taxa de vazão para o fluxo *web* em um patamar mais elevado com relação ao EXPR ou LOGR, uma melhoria por volta de 60% em média.

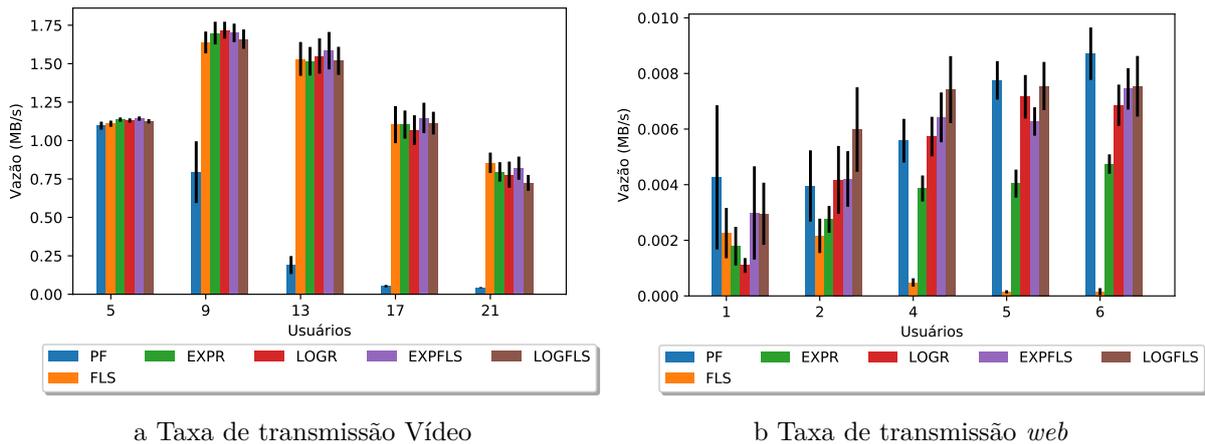


Figura 5.15: Vazão para Vídeo/*Web*.

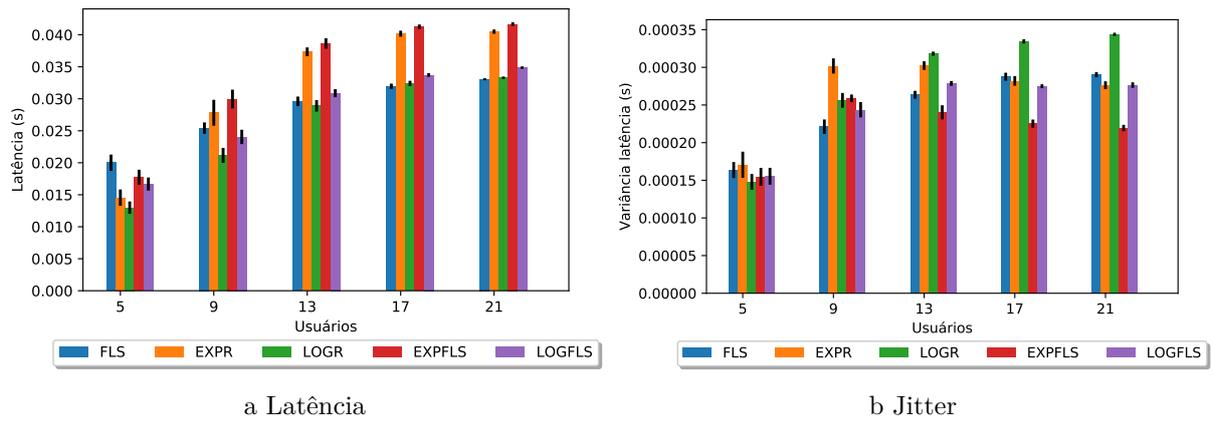


Figura 5.16: Latência e *jitter* para Vídeo.

Novamente verificou-se uma latência levemente superior para as novas políticas, mas se mantendo dentro dos níveis esperados pela qualidade de serviço, revelando sua capacidade de entregar as requisições com eficiência. Observou-se que as variâncias para ambos os fluxos continuam baixas, não impactando as outras métricas, conforme demonstrado nas Figuras 5.16 e 5.17.

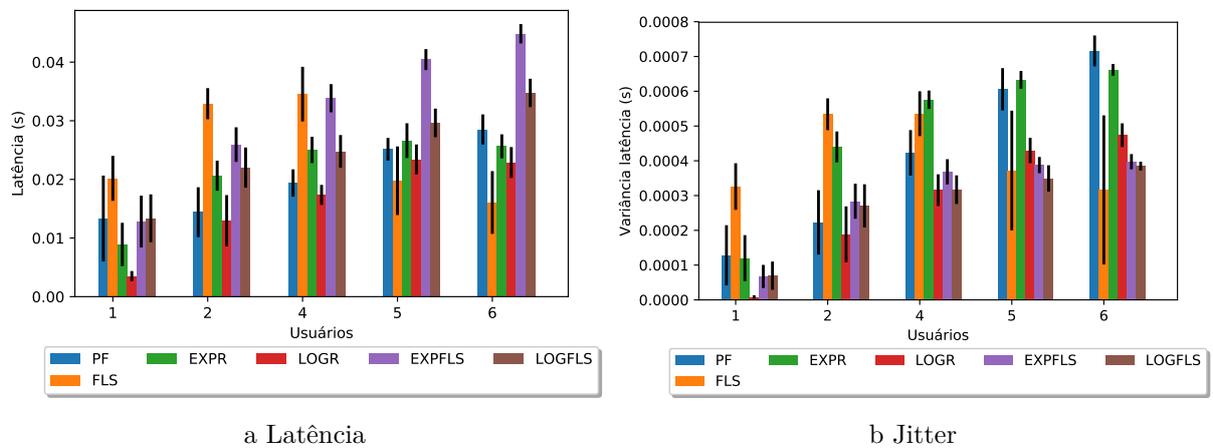


Figura 5.17: Latência e *jitter* para Web.

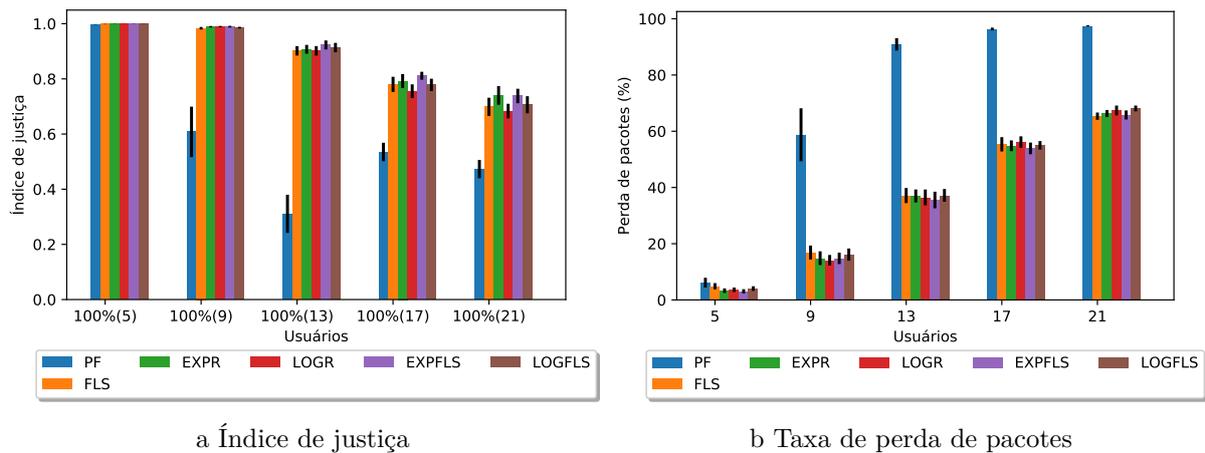


Figura 5.18: Índice de justiça e taxa de perda de pacotes para Vídeo.

Por fim, para o índice de justiça e a taxa de perda de pacotes, os valores seguiram os padrões já encontrados em subcenários anteriores, com as novas políticas mantendo-se extramente perto dos indicadores dos outros mecanismos de alocação e, em certas quantidades de usuários, mostrando-se melhores, como pode ser visto no caso de 21 usuários para vídeo, Figura 5.18, assim como para *web* na Figura 5.19.

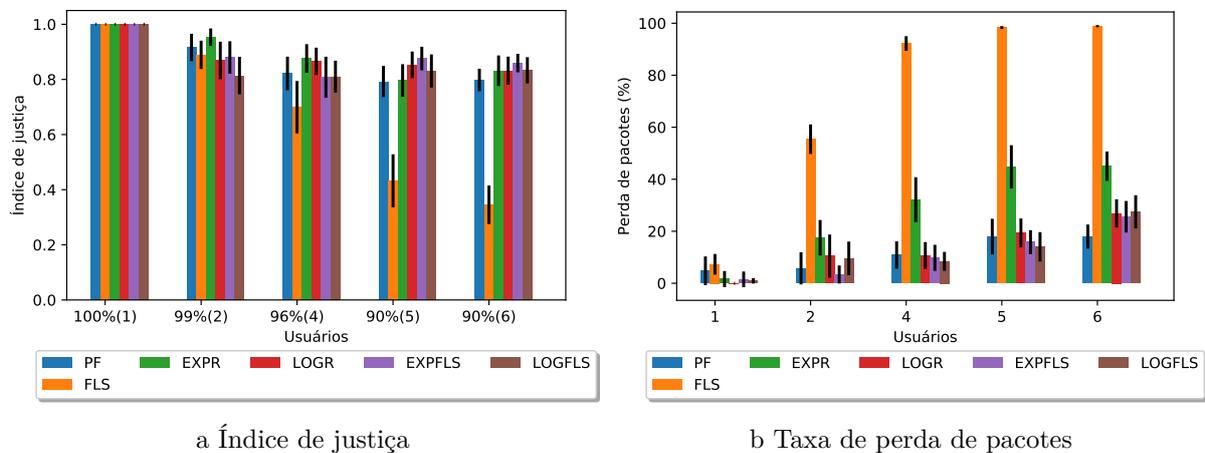


Figura 5.19: Índice de justiça e taxa de perda de pacotes para *Web*.

Percebeu-se um decaimento mais agravado em relação ao percentual de usuários atendidos com ao menos um pacote para o FLS. Fato esperado, devido à demasiada taxa de perda de pacotes, chegando a 45% de atendimento, ou seja, menos de 3 usuários para esse cenário. Apesar disso, os outros escalonadores ainda se mantiveram próximos, e em geral, as novas políticas ficaram levemente acima, assegurando uma boa posição perante às soluções na grande maioria dos subcasos.

Tabela 5.5: Atendimento de usuários tráfego *Web*.

Escalonador	Subcasos									
	10		20		30		40		50	
	Total	Percentual	Total	Percentual	Total	Percentual	Total	Percentual	Total	Percentual
PF	1	100%	2	100%	4	96%	5	100%	6	97%
FLS	1	100%	2	100%	4	85%	5	56%	6	45%
EXPR	1	100%	2	100%	4	97%	5	94%	6	95%
LOGR	1	100%	2	97%	4	97%	5	98%	6	95%
EXPFLS	1	100%	2	100%	4	98%	5	99%	6	96%
LOGFLS	1	100%	2	100%	4	100%	5	99%	6	94%

Cenário 5 (100% TR - 0% NTR)

Este último cenário demonstrou apenas uma saturação extremamente rápida para todas as estratégias de alocação, resultando em taxas de perda de pacotes próximas a 80% para a maioria dos casos. Indicando que, de fato, vídeo revelou-se o tipo de tráfego mais pesado para a alocação de recursos na célula, corroborando os mecanismos que priorizam esse tipo de fluxo. Entretanto, vale ressaltar que os demais fluxos também devem ser verificados para que não ocorra escassez das classificações, pois conforme demonstrando em testes anteriores, há a possibilidade de manter-se um bom resultado para vídeo, assim como para os outros fluxos, e ainda melhorar as métricas para certos casos. Como já explicitado previamente, nenhum dos usuários TR ficou sem receber pacotes, logo a tabela com o percentual de atendimento não faz sentido para esse cenário.

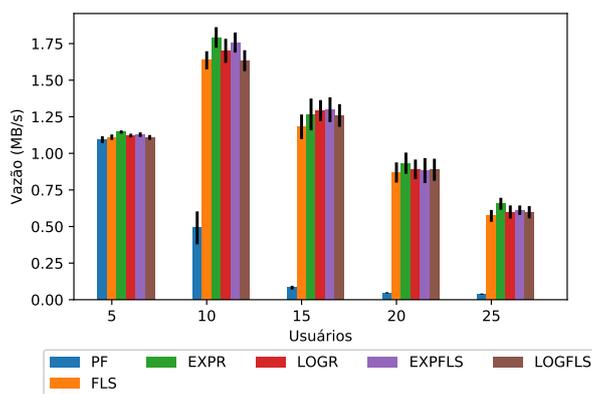


Figura 5.20: Vazão para Vídeo.

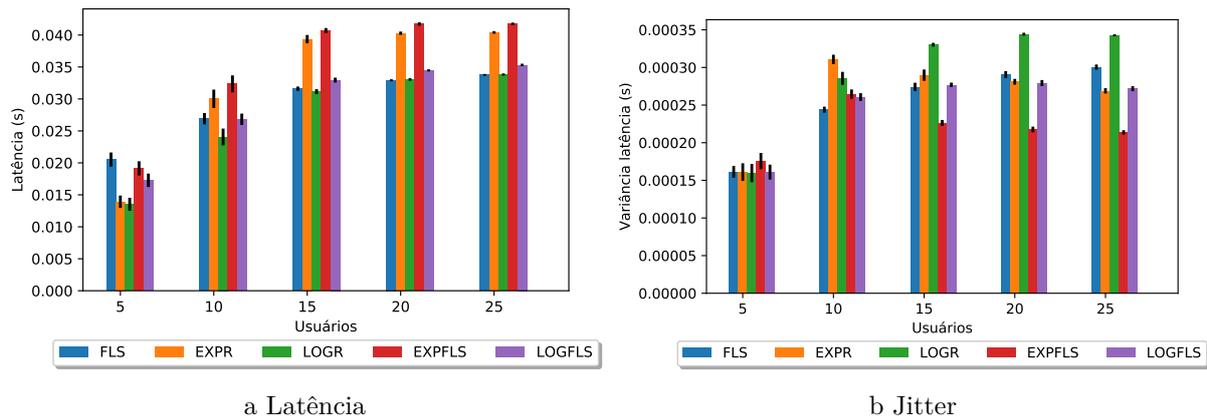


Figura 5.21: Latência e *jitter* para Vídeo.

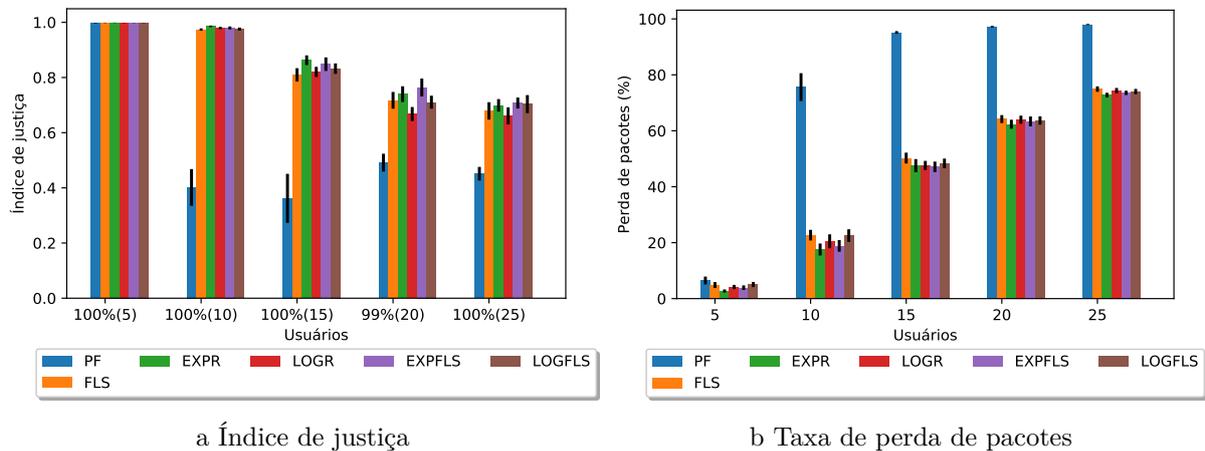


Figura 5.22: Índice de justiça e taxa de perda de pacotes para Vídeo.

5.2.2 Ambiente com mobilidade

A mobilidade constitui o modo *Random waypoint*, no qual o usuário escolhe um determinado ponto no mapa, ao redor do raio da torre, movimenta-se ao seu encontro em linha reta e seleciona uma nova localização, repetindo a ação até o fim da simulação. Assim como o ambiente estático, o ambiente com mobilidade também foi dividido em cinco cenários. Porém, nesse ambiente não serão apresentados todos os cenários da simulação, evidenciando-se apenas cenários nos quais sucederam diferenças consideráveis, pois os padrões de desempenho se mantiveram relativamente próximos aos do ambiente estático, ocorrendo divergências em pontos como rapidez de saturação da rede, índices de justiça mais baixos devido a flutuação da qualidade do canal ou níveis de latência mais altos. Da mesma forma que no ambiente estático, os resultados aqui apresentados possuem 95% de

nível de confiança. O número de execuções efetuadas foi de 20 e o intervalo de confiança calculado a partir da distribuição *t-student*.

Devido à mobilidade do usuário, os valores da qualidade do canal experienciada apresentam maior flutuação, isso ocorre à medida que o usuário se aproxima ou se afasta da torre, permitindo, assim, uma melhora ao escalonador FLS quanto à vazão do tráfego *web*. Como visto na Figura 5.23, houve menor decaimento da métrica de vazão para a maior parte dos cenários quando comparado ao ambiente estático, sendo apresentado somente o Cenário 2 (25% TR - 75% NTR). Ainda que sucedido um menor decaimento, o FLS não se manteve tão bom, pois os outros escalonadores apresentaram melhor priorização para essa classificação de tráfego.

Com relação às novas políticas de alocação de blocos de recurso, notou-se uma curva em crescimento para os dois tipos de aplicação, vídeo e *web*. Fato ocorrido somente para a política LOGR, entre os mecanismos avaliados na literatura.

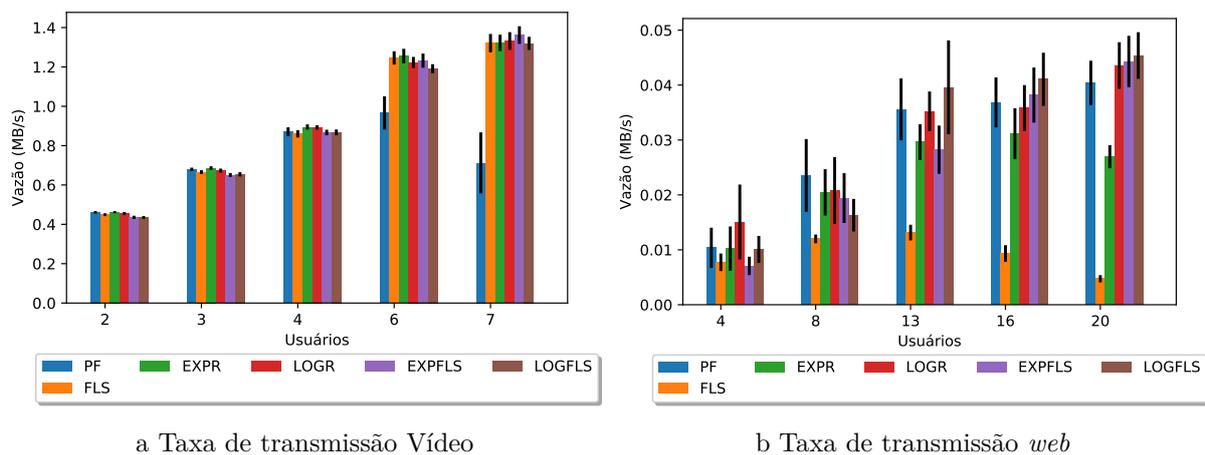


Figura 5.23: Vazão para Vídeo/ Web.

Outro ponto interessante, observado pela movimentação do usuário, consiste na maior variabilidade do índice de justiça frente à diminuição da taxa de perda de pacotes, indicando o motivo do leve aumento da vazão para o FLS, conforme demonstrado na Figura 5.24.

Novamente as novas políticas atingiram um maior número de usuários atendidos. Com um índice de justiça próximo a 0,6, os escalonadores que possuíam maior valor para o índice perderam muitos dos pacotes enviados, deixando o índice maior, pois apesar de menos usuários serem atendidos, a média de vazão por usuário apresentou mais semelhança.

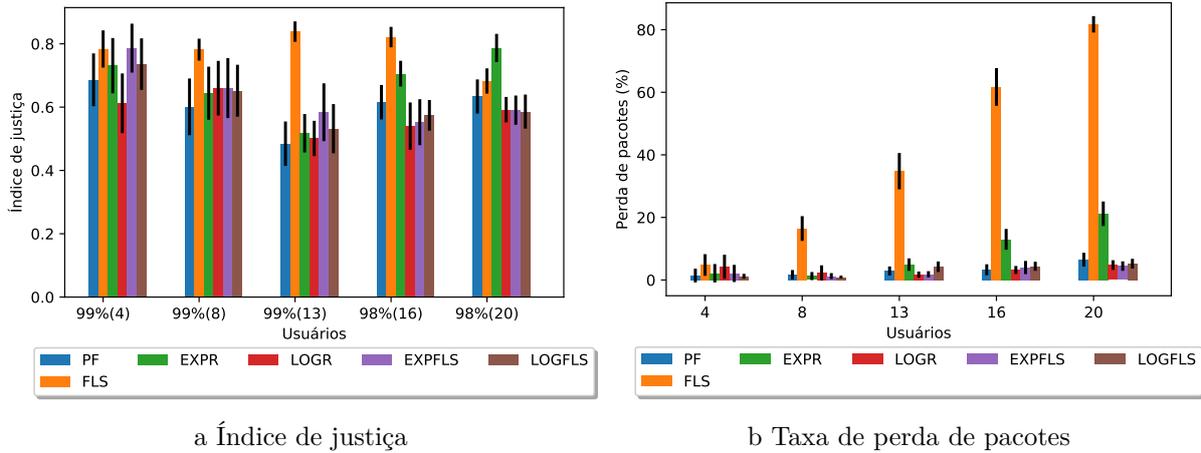


Figura 5.24: Índice de justiça e taxa de perda de pacotes para *Web*.

Mesmo com a mobilidade, o desempenho entre os escalonadores revelou baixa variação, representando um certo grau de flexibilidade para os parâmetros selecionados. A Tabela 5.6 indica o percentual de atendimento de usuários que receberam ao menos um pacote, ressaltando que a performance obtida ainda se encontra próxima aos resultados do ambiente estático. Repetidamente, o FLS se encontrou em último lugar para aplicações de fluxo *web*, caindo para 40% de atendimento e ainda mantendo o mais baixo nível de vazão. Já os escalonadores propostos se mostraram com maior vazão, tanto para aplicação *web*, quanto para vídeo, o que revelou uma boa performance quando comparado aos outros mecanismos.

Tabela 5.6: Atendimento de usuários tráfego *web* (Cenário 4 (75% TR - 25% NTR) com mobilidade).

Escalonador	Subcasos									
	10		20		30		40		50	
	Total	Percentual	Total	Percentual	Total	Percentual	Total	Percentual	Total	Percentual
PF	1	100%	2	100%	4	100%	5	98%	6	95%
FLS	1	100%	2	100%	4	75%	5	46%	6	40%
EXPR	1	100%	2	100%	4	97%	5	97%	6	91%
LOGR	1	100%	2	100%	4	100%	5	98%	6	96%
EXPFLS	1	100%	2	100%	4	100%	5	96%	6	97%
LOGFLS	1	100%	2	100%	4	97%	5	97%	6	99%

5.3 Discussão

Os resultados obtidos neste capítulo sintetizam o desempenho dos novos mecanismos de alocação de blocos de recurso, EXPFLS e LOGFLS, abordando uma implementação com

EXPR e outra com LOGR, respectivamente. Comparando as propostas a políticas de alocação constantemente presentes na literatura, especificamente, PF, FLS, EXPR e LOGR, todos os resultados apresentados foram obtidos por meio de repetição das simulações, sendo os valores representados nos gráficos uma média de 20 execuções e o intervalo de confiança obtido por meio da distribuição *t-student* com 95% de nível de confiança.

A resolução entre os ambientes estático ou móvel não apresentou variações que chamassem muita atenção. Revelou um grau de flexibilidade quanto aos parâmetros dos escalonadores e possibilitou um melhor atendimento para os novos mecanismos em certos cenários, uma vez que apresentou maiores taxas de vazão, menores ou semelhantes taxas de perda de pacotes e índice de justiça, além de demonstrar maior índice de atendimento aos usuários.

O FLS, como visto, possui um desempenho abaixo da média para fluxos NTR, pois sua priorização vai contra este tipo de tráfego. A otimização proposta manteve o sistema baseado em teoria de controle, priorizando os tipos de tráfegos por latência e mostrou-se uma boa alternativa a políticas de alocação já existentes na literatura, dado o objetivo de permitir maior heterogeneidade dos fluxos de dados. O modelo proposto, na presente pesquisa, revelou desempenho semelhante ao EXPR e ao LOGR, obtendo, em certos casos, resultados melhores que estes. Conseguiu ainda, superar completamente o FLS nos cenários nos quais este se mostrou ineficiente. Dessa forma, os modelos propostos possuem melhor equilíbrio quanto à priorização de fluxos TR ou NTR, obtendo melhores níveis de atendimento e configurando-se em boas alternativas para o uso em redes com maior utilização de tráfego não tempo real, cumprindo, assim, o objetivo estipulado inicialmente para a pesquisa. Assumindo o cenário 4 do ambiente estático como exemplo, a fim de reforçar a execução do objetivo principal, elabora-se um tráfego de 75% TR contra 25% NTR, refletindo um caso comumente encontrado na realidade, em que grande parte dos usuários utilizam aplicações como vídeo. Neste cenário, verifica-se a eficiência do modelo proposto, obtendo equiparação de métricas perante o tráfego de tempo real, e ainda atingindo melhorias quanto ao tráfego não tempo real, com desempenhos para vazão com aumento em torno de 60% quando comparado ao EPXR ou LOGR, e por volta de 46 vezes maior quando comparado ao FLS original. Além disso, verifica-se também maior percentual de usuários atendidos com ao menos um pacote, dobrando o quantitativo do FLS, passando da média de 2,75 para 5,8, e equiparando-se aos outros escalonadores.

Capítulo 6

Conclusão

Dado o crescimento do tráfego em dispositivos IdC, nos quais, em geral, ocorrem chamadas REST - HTTP, existe um aumento do tipo de tráfego não tempo real, em especial o tráfego *web*. Desta forma, apesar de a grande maioria dos fluxos de dados se originarem dos tráfegos de tempo real, como Vídeo (*stream* ou *buffered*), torna-se indispensável o tratamento do tráfego não tempo real. Assim, políticas que priorizam exclusivamente vídeos, VoIP entre outros, mostram-se ineficazes em cenários cuja a quantidade não priorizada se torna maior, gerando uma enorme taxa de perda de pacotes e mínima vazão, como exemplificado no Capítulo 5.

A grande maioria das políticas de alocação de blocos de recursos atuais utilizam diferenciação entre as classificações de tráfego, a fim de obter melhores resultados de vazão para os fluxos priorizados. Por meio de estratégias elaboradas, que utilizam latência, tempo de espera na fila, vazão média esperada entre outros atributos, os escalonadores conseguem melhorar o desempenho dos tráfegos de tempo real. Em contrapartida, a classificação não tempo real utiliza, na maioria dos casos, o mecanismo mais simples, *Proportional and Fair*, ocasionando em uma performance abaixo do esperado.

6.1 Considerações finais

Esta pesquisa buscou estudar estratégias que permitem o balanço dos fluxos de dados, independente da proporção dos tipos de tráfego na rede, por meio da mesclagem da lei de controle presente no *Frame Level Scheduler* e das estratégias de pesagem do *Exponential Rule* ou *Log Rule*, sem distinguir os tipos de tráfego na utilização dos mecanismos dessas políticas. Buscou também, comprovar os resultados perante os demais escalonadores, atingindo boas métricas na maioria dos cenários, em especial, cenários nos quais o FLS se comporta de modo desfavorável, por exemplo, quando o tráfego *web* revela-se maior que o tráfego de vídeo em relação à quantidade de usuários utilizando as aplicações. As

políticas de alocação de blocos de recurso propostas na presente pesquisa se revelaram suficientes, mantendo um nível de qualidade de serviço e não exaurindo o tráfego *web* como antes ocorria, além disso, evidenciaram resultados acima da média para alguns dos cenários apresentados no decorrer do Capítulo 5, atingindo o objetivo principal do trabalho, obtendo uma alternativa às políticas de alocação de blocos de recurso comumente utilizadas na literatura. Encontram-se nos capítulos 3 e 4 da presente pesquisa, descrições dos cenários empregados a fim de obter ambientes de simulação padronizados, facilitando a comparação do desempenho dos escalonadores, cumprindo objetivos secundários, uma vez que muitos trabalhos encontrados no estado da arte possuem divergências entre os cenários aplicados.

6.2 Trabalhos Futuros

Para trabalhos futuros, sugere-se a utilização de outras métricas, além das obtidas para essa pesquisa, tais como a eficiência espectral e energética, comparações com outros escalonadores, exposição à interferência do sinal, aplicação de ambientes com topologias maiores aproveitando-se de *handover* ou reuso de frequência, entre outras. Com relação a estratégia em si, sugere-se aprimoramento por meio de algumas técnicas encontradas no Capítulo 3, como a utilização do filtro de Kalman, conforme demonstrado por Teixeira e Timóteo [7] durante a estimativa da taxa de transmissão; a técnica de *Delay Priority* desenvolvida por Li *et al.* [44] concebida originalmente para o MLWDF, porém podendo ser modificada para a implementação no EXPR ou LOGR utilizados na presente pesquisa.

Pondera-se, também, quanto à padronização e normalização das métricas de peso para os fluxos de dados, habilitando um modo de alocação que utiliza mais de um tipo de escalonador para diferentes aplicações, aproveitando das características de cada estratégia, pois como já revelado na pesquisa, alguns mecanismos possuem maiores vantagens para determinados fluxos de dados, por exemplo o EXPR possui vantagens para TR (vídeo), enquanto o LOGR para NTR(*web*).

Referências

- [1] Sesia, Stefania, Issam Toufik e Matthew Baker: *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley Publishing, 2009, ISBN 0470697164, 9780470697160. 1, 5, 6, 8
- [2] Holma, Harri e Antti Toskala: *LTE for UMTS: Evolution to LTE-Advanced*. Wiley Publishing, 2nd edição, 2011, ISBN 0470660007, 9780470660003. 1, 6, 7, 11, 16
- [3] Capozzi, F., G. Piro, L. A. Grieco, G. Boggia e P. Camarda: *Downlink packet scheduling in lte cellular networks: Key design issues and a survey*. IEEE Communications Surveys Tutorials, 15(2):678–700, Second 2013, ISSN 1553-877X. 1, 5, 12, 14, 21, 38, 40, 48
- [4] *Number of lte subscriptions worldwide from 2018 to 2023 (in billions)*. <https://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-june-2017.pdf>. Acessado em: 2019-04-04. 1, 23
- [5] Piro, G., L. A. Grieco, G. Boggia, F. Capozzi e P. Camarda: *Simulating lte cellular systems: An open-source framework*. IEEE Transactions on Vehicular Technology, 60(2):498–513, Feb 2011, ISSN 0018-9545. 2, 15, 22, 23, 25, 27, 42, 44
- [6] Aiyetoro, G. e F. Takawira: *A packet scheduling scheme in s-lte networks using logarithmic function as a control parameter*. Em *2017 Global Wireless Summit (GWS)*, páginas 98–103, Oct 2017. 2, 36, 37
- [7] Teixeira, M. J. e V. Salvador Timóteo: *Using a kalman filter to improve schedulers performance in mobile networks*. Em *2019 15th International Wireless Communications Mobile Computing Conference (IWCMC)*, páginas 853–858, June 2019. 2, 33, 37, 40, 72
- [8] Piro, Giuseppe, Luigi Alfredo Grieco, Gennaro Boggia, Rossella Fortuna e Pietro Camarda: *Two-level downlink scheduling for real-time multimedia services in lte networks*. IEEE Transactions on Multimedia, 13:1052–1065, 2011. 2, 15, 19, 20, 33, 37, 38, 43, 48, 50, 52
- [9] Sadiq, Bilal, Ritesh Madan e Ashwin Sampath: *Downlink scheduling for multiclass traffic in lte*. EURASIP J. Wirel. Commun. Netw., 2009:14:9–14:9, março 2009, ISSN 1687-1472. <http://dx.doi.org/10.1155/2009/510617>. 2, 18, 34, 38, 49, 52
- [10] Gohil, A., H. Modi e S. K. Patel: *5g technology of mobile communication: A survey*. Em *2013 International Conference on Intelligent Systems and Signal Processing (ISSP)*, páginas 288–292, March 2013. 4

- [11] Khare, Varsha e Shubhanshi Garg: *Comparative study of 1g, 2g, 3g and 4g sapna shukla, student, electronics and communication, rkgitw, ghaziabad.* 4
- [12] Dahlman, Erik, Stefan Parkvall e Johan Skold: *4G: LTE/LTE-Advanced for Mobile Broadband.* Academic Press, Inc., Orlando, FL, USA, 1st edição, 2011, ISBN 012385489X, 9780123854896. 4, 6, 7, 10
- [13] *Ciena 5g bootcamp.* <https://www.ciena.com/insights/articles/Welcome-to-Ciena-5G-Boot-Camp.html>. Acessado em: 2019-01-15. 5
- [14] Lucent, Alcatel: *The lte network architecture a comprehensive tutorial.* Strategic WhitePaper, 2009. 5, 6, 8, 9
- [15] Remy, Jean Gabriel e Charlotte Letamendia: *LTE Standards and Architecture*, capítulo 1, páginas 1–112. John Wiley & Sons, Ltd, 2014, ISBN 9781119043508. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119043508.ch1>. 7
- [16] Anritsu: *Future technologies and testing for fixedmobile convergence, sae and lte in cellularmobile communications.* Strategic WhitePaper, 2012. 8, 10, 11
- [17] Schwarz, Rohde &: *Lte-advanced (3gpp rel. 12) technology introduction.* Strategic WhitePaper, 2014. 9, 11, 28
- [18] Ali, S. G. A., M. Dani Baba e M. A. Mansor: *Rb preserver downlink scheduling algorithm for real-time multimedia services in lte networks.* Em *2015 IEEE Symposium on Computer Applications Industrial Electronics (ISCAIE)*, páginas 21–26, April 2015. 9
- [19] *Ericsson mobility report.* <https://www.statista.com/statistics/206615/forecast-of-the-number-of-global-hspa-lte-subscriptions-up-to-2014/>. Acessado em: 2019-01-12. 9
- [20] Aiyetoro, G. e F. Takawira: *An exponential based packet scheduling scheme for real time traffic in satellite lte networks.* Em *2017 IEEE AFRICON*, páginas 215–220, Sep. 2017. 13, 36, 37, 40
- [21] Alsharbaty, Firas: *The effect of harq procedure on the performance of lte.* AL Rafdain Engineering Journal, 22:112–122, julho 2013. 14
- [22] Benchaabene, Y., N. Boujnah e F. Zarai: *Comparative analysis of downlink scheduling algorithms for lte femtocells networks.* Em *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, páginas 1957–1961, June 2017. 15
- [23] Kayali, M. O., Z. Shmeiss, H. Safa e W. El-Hajj: *Downlink scheduling in lte: Challenges, improvement, and analysis.* Em *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, páginas 323–328, June 2017. 15, 23

- [24] Kumar, S., D. P. Goswami, A. Sarkar e A. Sur: *Buffer aware three level scheduler for video streaming over lte*. Em *2017 9th International Conference on Communication Systems and Networks (COMSNETS)*, páginas 222–228, Jan 2017. 15
- [25] S. Maheshwari and e Y. S. Latha and: *Resource scheduling algorithms for lte using weights*. Em *2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, páginas 264–269, July 2016. 17, 31, 37
- [26] Wengertter, C., J. Ohlhorst e A. G. E. von Elbwart: *Fairness and throughput analysis for generalized proportional fair frequency scheduling in ofdma*. Em *2005 IEEE 61st Vehicular Technology Conference*, volume 3, páginas 1903–1907 Vol. 3, May 2005. 17, 34, 35
- [27] Ramli, H. A. M., R. Basukala, K. Sandrasegaran e R. Patachaianand: *Performance of well known packet scheduling algorithms in the downlink 3gpp lte system*. Em *2009 IEEE 9th Malaysia International Conference on Communications (MICC)*, páginas 815–820, Dec 2009. 18, 34
- [28] Basukala, R., H. A. M. Ramli e K. Sandrasegaran: *Performance analysis of exp/pf and m-lwdf in downlink 3gpp lte system*. Em *2009 First Asian Himalayas International Conference on Internet*, páginas 1–5, Nov 2009. 18
- [29] Ogata, Katsuhiko: *Modern Control Engineering*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 4th edição, 2001, ISBN 0130609072. 19
- [30] Tanenbaum, Andrew: *Computer Networks*. Prentice Hall Professional Technical Reference, 4th edição, 2002, ISBN 0130661023. 21
- [31] Kurose, James F. e Keith W. Ross: *Computer Networking: A Top-Down Approach (6th Edition)*. Pearson, 6th edição, 2012, ISBN 0132856204, 9780132856201. 22
- [32] Jain, Raj, Dah Ming Chiu e W. Hawe: *A quantitative measure of fairness and discrimination for resource allocation in shared computer systems*. CoRR, cs.NI/9809099, 1998. <http://dblp.uni-trier.de/db/journals/corr/corr9809.html#cs-NI-9809099>. 22
- [33] *Network simulator 3*. <https://www.nsnam.org/>. Acessado em: 2019-01-17. 22
- [34] Mehmeti, F. e T. Spyropoulos: *Performance analysis of mobile data offloading in heterogeneous networks*. *IEEE Transactions on Mobile Computing*, 16(2):482–497, Feb 2017, ISSN 1536-1233. 23
- [35] Camp, Tracy, Jeff Boleng e Vanessa Davies: *A survey of mobility models for ad hoc network research*. WIRELESS COMMUNICATIONS & MOBILE COMPUTING (WCMC): SPECIAL ISSUE ON MOBILE AD HOC NETWORKING: RESEARCH, TRENDS AND APPLICATIONS, 2:483–502, 2002. 25
- [36] *Physical layer aspect for evolved universal terrestrial radio access (utra)*. <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=1247>. Acessado em: 2019-01-12. 25

- [37] Jonsson, Lars Erik, Kristofer Sandlund e Ghyslain Pelletier: *RObust Header Compression (ROHC): ROHC over Channels That Can Reorder Packets*. RFC 4224, janeiro 2006. <https://rfc-editor.org/rfc/rfc4224.txt>. 28
- [38] Rappaport, Theodore: *Wireless Communications: Principles and Practice*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edição, 2001, ISBN 0130422320. 29
- [39] Alotaibi, S. e R. Akl: *Packet scheduling bandwidth type-based mechanism for lte*. Em *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON)*, páginas 363–368, Oct 2017. 32, 37
- [40] Alotaibi, S. e R. Akl: *Radio resource management in lte femtocell networks*. Em *2017 IEEE 16th International Symposium on Network Computing and Applications (NCA)*, páginas 1–5, Oct 2017. 32, 37
- [41] Aramide, S. O., B. Barakat, Y. Wang, S. Keates e K. Arshad: *Generalized proportional fair (gpf) scheduler for lte-a*. Em *2017 9th Computer Science and Electronic Engineering (CEECE)*, páginas 128–132, Sep. 2017. 32, 33, 37
- [42] A.Astrom, Karl J. e Born Wittenmark: *Computer-controlled Systems (3rd Ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1997, ISBN 0-13-314899-8. 33
- [43] Gungor, V. C. e O. G. Uyan: *Qos-aware downlink scheduling algorithm for lte networks: A case study on edge users*. Em *2017 25th Signal Processing and Communications Applications Conference (SIU)*, páginas 1–4, May 2017. 34, 37
- [44] Li, Y., B. Hu, H. Zhu, Z. Wei e W. Gao: *A delay priority scheduling algorithm for downlink real-time traffic in lte networks*. Em *2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference*, páginas 706–709, May 2016. 34, 35, 37, 72
- [45] Kaewmongkol, K., A. Jansang e A. Phonphoem: *Delay-aware with resource block management scheduling algorithm in lte*. Em *2015 International Computer Science and Engineering Conference (ICSEC)*, páginas 1–6, Nov 2015. 35, 37, 48
- [46] Prehofer, C.: *Models at rest or modelling restful interfaces for the internet of things*. Em *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, páginas 251–255, Dec 2015. 39
- [47] Tandale, U., B. Momin e D. P. Seetharam: *An empirical study of application layer protocols for iot*. Em *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, páginas 2447–2451, Aug 2017. 39, 49