



DISSERTAÇÃO DE MESTRADO

**ESCALONAMENTO DE TRÁFEGO
EM REDES WiMAX
NO MODO PMP**

Carino Andrade Rodrigues

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA



Universidade de Brasília
Faculdade de Tecnologia
Departamento de Engenharia Elétrica

ESCALONAMENTO DE TRÁFEGO EM REDES WiMAX NO
MODO PMP

Carino Andrade Rodrigues

Orientador: Paulo Roberto de Lira Gondim

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA
ELÉTRICA

PUBLICAÇÃO: PPGENE.DM - 391/09
Brasília / DF: Julho/2009

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

**ESCALONAMENTO DE TRÁFEGO EM REDES
WiMAX NO MODO PMP**

Carino Andrade Rodrigues

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE.

APROVADA POR:

Dr. Paulo R. de Lira Gondim, ENE/UnB _____

Orientador

Dr. Georges Amvame Nze, UnB Gama _____

Examinador externo

Dr. Paulo Henrique P. Carvalho, ENE/UnB _____

Examinador interno

BRASÍLIA(DF), 10 DE JULHO DE 2009.

FICHA CATALOGRÁFICA

Rodrigues, Carino Andrade

Escalonamento de tráfego em redes WiMAX no modo PMP. [Distrito Federal] 2009. xviii, 109p., 210 x 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2009). Dissertação de Mestrado, Universidade de Brasília. Faculdade de Tecnologia.) Departamento de Engenharia Elétrica.

1. Escalonamento, 2. WiMAX e 3. Qualidade de Serviço

Publicação PPGENE.DM - 391/09

REFERÊNCIA BIBLIOGRÁFICA

Carino Andrade Rodrigues (2009). Escalonamento de tráfego em redes WiMAX no modo PMP. Dissertação de Mestrado, Publicação PPGENE.DM - 391/09 10 de Julho de 2009, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília-DF, 109 p.

CESSÃO DE DIREITOS

NOME DO AUTOR: Carino Andrade Rodrigues

TÍTULO DA DISSERTAÇÃO: Escalonamento de tráfego em redes WiMAX no modo PMP

GRAU/ANO: Mestre/2009.

É concedida à Universidade de Brasília permissão para reproduzir cópias desta Dissertação de Mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação de mestrado pode ser reproduzida sem a autorização por escrito do autor.

Carino Andrade Rodrigues (carinor@yahoo.com.br)

SHIN CA 5, Conj. M2, Ed. Res. Enzo, ap. 110 - Lago Norte

CEP 71503-505 - Brasília - DF - Brasil

Dedicatória

Dedico este trabalho à memória da minha "dinda", Ustana Rodrigues Silva.

Carino Andrade Rodrigues

Agradecimentos

Ao Prof. Dr. Paulo Gondim, representando a UNB, pela oportunidade inicial e por sua orientação durante o trabalho;

Ao Banco do Brasil, pela liberação parcial para participar do Programa de Mestrado, em especial aos administradores: Pagotti, Angelino, Domingo, Edi, Márcia, Milton, Moacyr e Edilson;

Aos colegas da Gerência de Infra-estrutura Tecnológica II, que tornaram as minhas ausências imperceptíveis, em especial aos colaboradores dos núcleos 23 e 33;

Aos colegas de curso pela ajuda mútua;

Aos amigos e familiares por acreditarem sempre no meu sucesso;

A minha sogra, Sra. Joselita Fonseca Ribeiro Lima, pela torcida em todos os momentos;

Em especial, aos meus pais, Sr. José Rodrigues da Silva e Sra. Mirian Andrade Rodrigues, principais responsáveis por cada uma das minhas conquistas;

E por último, a minha esposa, Sra. Jaqueline Fonseca Lima Rodrigues, pela revisão ortográfica, por seu apoio nos momentos mais difíceis, por seu incentivo e certeza da minha vitória.

Carino Andrade Rodrigues

Escalonamento de tráfego em redes WiMAX no modo PMP**Autor: Carino Andrade Rodrigues****Orientador: Paulo Roberto de Lira Gondim****Programa de Pós-graduação em Engenharia Elétrica****Brasília, mês de Julho (2009)**

O padrão IEEE 802.16 define uma rede de acesso, local e metropolitana, de banda larga sem fio conhecida como WiMAX (*worldwide interoperability for microwave access*). A norma estabelece requisitos de qualidade de serviço para tráfego multimídia. Para atingir estes objetivos, a BS (*base station*) deve controlar as requisições de largura de banda das SSs (*subscriber station*) utilizando uma estratégia de escalonamento. A técnica de escalonamento não é descrita na norma, permitindo aos fabricantes a implementação de um algoritmo próprio. Tendo em vista a grande capacidade da tecnologia OFDMA (*orthogonal frequency division multiple access*), os algoritmos de escalonamento devem estar adaptados para funcionar com uma camada física que utilize esta tecnologia. Este trabalho propõe uma nova estratégia de camada MAC (*medium access control*), simples e eficaz, para o tráfego de *uplink* no modo PMP (*point-to-multipoint*), com suporte à camada física em WirelessMAN - OFDMA, utilizando o método PUSC (*partial usage of subchannels*) para associação de subcanais. O comportamento da técnica foi testado utilizando-se o simulador NS-2 (*Network Simulator 2*), permitindo a verificação da justa distribuição dos recursos da rede entre estações do mesmo tipo de tráfego e o atendimento a requisitos de qualidade de serviço entre os diferentes tipos de tráfego. Adicionalmente, verificou-se a adaptabilidade da estratégia de escalonamento aos possíveis perfis de transmissão de uma rede WiMAX.

Palavras chaves: Escalonamento, WiMAX, Qualidade de Serviço.

ABSTRACT

Traffic scheduling in WiMAX networks on the PMP mode.

Author: Carino Andrade Rodrigues

Supervisor: Paulo Roberto de Lira Gondim

Programa de Pós-graduação em Engenharia Elétrica

Brasília, month of July (2009)

The standard IEEE 802.16 defines fixed and mobile broadband wireless access systems for local and metropolitan area networks, also called WiMAX (worldwide interoperability for microwave access). The standard establishes requirements of quality of service for multimedia traffic. To achieve these objectives, the BS (base station) shall monitor the requests of bandwidth from the SS's (subscriber station) using a scheduling strategy. The technique of scheduling is not described in the standard, allowing manufacturers to implement an algorithm by themselves. Considering the large capacity and flexibility of OFDMA (orthogonal frequency division multiple access), the scheduling algorithms must be adapted to work with a physical layer using this technology. This work proposes a new strategy for MAC (medium access control) layer, simple and objective, for uplink traffic on the PMP (point-to-multipoint) mode, supporting PHY WirelessMAN - OFDMA layer, using PUSC (partial usage of subchannels) method for assigning logical subchannels. The technique was simulated using the NS-2 (Network Simulator 2), allowing the verification of the fair distribution of network resources between the stations of the same traffic type and the achievement of quality of service requirements between different traffic types. Additionally, it was observed the adaptability of the scheduling strategy to possible transmission profiles of a WiMAX network.

Keywords: Scheduling, WiMAX, Quality of Service.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	MOTIVAÇÕES	2
1.2	APRESENTAÇÃO DO PROBLEMA	3
1.3	OBJETIVOS	3
1.4	ORGANIZAÇÃO DA DISSERTAÇÃO	4
2	SERVIÇOS DE COMUNICAÇÕES MULTIMÍDIA SOBRE REDES SEM FIO	5
2.1	INTRODUÇÃO	5
2.2	FLUXO DE ÁUDIO	6
2.3	SERVIÇOS DE VOIP	7
2.4	FLUXO DE VÍDEO	10
2.5	SERVIÇOS DE IPTV	11
2.5.1	ARQUITETURA DE IPTV	12
3	WIMAX	17
3.1	INTRODUÇÃO	17
3.2	EVOLUÇÃO DO PADRÃO IEEE 802.16	19
3.3	CAMADA FÍSICA (PHY)	20
3.3.1	WIRELESSMAN-SC	22
3.3.2	WIRELESSMAN-SCA	22
3.3.3	WIRELESSMAN-OFDM	23
3.3.4	WIRELESSMAN-OFDMA	27
3.3.5	WIRELESSHUMAN	30
3.4	CAMADA DE ENLACE (MAC)	30
3.4.1	SUBCAMADA DE CONVERGÊNCIA	32
3.4.2	SUBCAMADA DE PARTE COMUM	33
3.4.3	SUBCAMADA DE SEGURANÇA	51
4	ESTRATÉGIAS DE ESCALONAMENTO	53
4.1	INTRODUÇÃO	53
4.2	ESTRATÉGIAS DE ESCALONAMENTO	54

4.2.1	ESTRATÉGIAS DE ESCALONAMENTO PARA SISTEMAS OPERACIONAIS E REDES FIXAS COMUTADAS	55
4.2.2	ESTRATÉGIAS DE ESCALONAMENTO PARA O WiMAX.....	65
5	ALGORITMO PROPOSTO, SIMULAÇÕES E RESULTADOS.....	81
5.1	INTRODUÇÃO	81
5.2	ADAPTAÇÃO DA ESTRATÉGIA DE ESCALONAMENTO.....	81
5.3	ESCOLHA DO SIMULADOR.....	83
5.4	ADAPTAÇÃO NO MÓDULO WiMAX DO NDSL	84
5.5	AMBIENTE DE SIMULAÇÃO.....	86
5.6	SIMULAÇÕES E ANÁLISE DOS RESULTADOS	87
5.6.1	CENÁRIO 1 - ESCALONAMENTO DE ESTAÇÕES <i>Best Effort</i>	87
5.6.2	CENÁRIO 2 - ESCALONAMENTO DE SSS COM DIFERENTES CLASSES DE SERVIÇO	90
5.6.3	CENÁRIO 3 - ESCALONAMENTO DE SSS COM DIFERENTES MODU- LAÇÕES DE SINAL.....	97
6	CONCLUSÕES.....	101
	REFERÊNCIAS BIBLIOGRÁFICAS	103

LISTA DE FIGURAS

2.1	Arquitetura Funcional para serviços VoIP	8
2.2	Pilha de protocolos para serviços VoIP utilizando SIP	8
2.3	Pilha de protocolos para serviços VoIP utilizando H323	8
2.4	Arquitetura de VoIP sobre WiMAX	9
2.5	Arquitetura Genérica de IPTV	12
2.6	Arquitetura de IPTV sobre WiMAX	14
2.7	Pilha de protocolos para transmissão de IPTV	15
2.8	Arquitetura de QoS de camada MAC para transmissão de IPTV e VoIP	16
3.1	Operação em TDD.	21
3.2	Operação em FDD.....	21
3.3	Representação em frequência do canal OFDM	25
3.4	Representação do símbolo OFDM	25
3.5	<i>Frame</i> para camada física WirelessMAN-OFDM operando em PMP.....	26
3.6	Codificação de canal obrigatória por esquema de modulação.....	26
3.7	<i>Frame</i> para camada física WirelessMAN-OFDMA operando em PMP.....	28
3.8	Múltiplas zonas permitidas no <i>frame</i> OFDMA	29
3.9	Diagrama esquemático da função do <i>scheduler</i> na alocação de subcanais lógicos	29
3.10	<i>Payload</i> por esquema de modulação e taxa de codificação.	30
3.11	Modelo de referência da estrutura em camadas	32
3.12	Formato do MAC PDU	36
3.13	Cabeçalho MAC Genérico	37
3.14	Cabeçalho de sinalização sem <i>payload</i> - Tipo I.....	38
3.15	Cabeçalho de sinalização sem <i>payload</i> - Tipo II	39
3.16	Cabeçalho sem <i>payload</i> - Requisição de Banda	40
3.17	Regras de requisição/transmissão.....	48
4.1	Servidor FIFO ou FCFS	56
4.2	Servidor Round-Robin.....	57
4.3	Funcionamento RR (a) e WRR (b)	59
4.4	Gráfico Delay x Taxa de saída das estratégias FCFS e WFQ.....	61
4.5	Funcionamento da janela deslizante do PWFQ	63
4.6	Arquitetura de QoS, proposta por Chu <i>et al.</i>	66

4.7	Arquitetura de QoS, proposta em Wongthavarawat <i>et al.</i>	67
4.8	Estratégia DPFQ	68
4.9	Funções dos <i>schedulers</i> para garantia de QoS, propostas por Cicconetti <i>et al.</i>	69
4.10	Arquitetura de QoS, proposta por Lilei <i>et al.</i>	70
4.11	Algoritmo QBMS	70
4.12	Exemplo da diferença de <i>jitter</i> com base na ordem de <i>slots</i>	75
5.1	Estrutura da rede para o cenário 1	90
5.2	Gráfico da Taxa de transmissão X tempo, para o Cenário 1 em Sayenko <i>et al.</i>	91
5.3	Gráfico da Taxa de transmissão X tempo, deste trabalho, para o Cenário 1	91
5.4	Estrutura da rede para o cenário 2.....	92
5.5	Resultado do cenário 2 - DL/UL - 1:1.....	93
5.6	Resultado do cenário 2 de Sayenko <i>et al.</i>	94
5.7	Resultado do cenário 2 - DL/UL - 3:1.....	95
5.8	Gráfico de Jitter (ms) x tempo (s) em Sayenko <i>et al.</i>	95
5.9	Gráfico de Jitter (ms) x tempo (s)	96
5.10	Estrutura da rede para o cenário 3.....	98
5.11	Gráfico da Taxa de transmissão X tempo, para o Cenário 3 em Sayenko <i>et al.</i>	99
5.12	Gráfico da Taxa de transmissão X tempo, deste trabalho, para o Cenário 3	100

LISTA DE TABELAS

3.1	Principais características PHY WirelessMAN-SC	23
3.2	Principais características PHY WirelessMAN-SCa	24
3.3	Principais características PHY WirelessMAN-OFDM	27
3.4	Principais características PHY WirelessMAN-OFDMA	31
3.5	Principais características PHY WirelessHUMAN	31
3.6	Descrição dos campos do Cabeçalho MAC Genérico	37
3.7	Códigos dos subheaders do GMH	38
3.8	Códigos do campo <i>Type</i> do cabeçalho de sinalização MAC - Tipo I	39
3.9	Formato do GMS	41
3.10	Descrição dos campos do GMS	42
3.11	Formato das mensagens DL-MAP	44
3.12	Formato das mensagens UL-MAP	45
4.1	Taxonomia de estratégias de escalonamento	56
5.1	Parâmetros fixos do módulo	87
5.2	Parâmetros de Simulador utilizado em Sayenko <i>et al.</i>	88
5.3	Modulação, Codificação e Tamanho de <i>Slot</i>	88
5.4	Modulação, Codificação e Tamanho de Símbolo OFDMA, utilizados neste trabalho	88
5.5	Controle de ativação do tráfego	89
5.6	Tipos de tráfego e Parâmetros de QoS para o cenário 2	92
5.7	Tipos de tráfego e Parâmetros de QoS para o cenário 3	98
5.8	Controle da Modulação Utilizada	98

LISTA DE SÍMBOLOS

Siglas

AAS	<i>adaptive antenna system</i>
ARQ	<i>automatic repeat request</i>
ATM	<i>asynchronous transfer mode</i>
BR	<i>bandwidth request</i>
BS	<i>base station</i>
CBR	<i>constant bit rate</i>
CPS	<i>common part sublayer</i>
CQI	<i>channel quality information</i>
CQICH	<i>channel quality information channel</i>
CRC	<i>cyclic redundancy check</i>
CS	<i>service-specific convergence sublayer</i>
DHCP	<i>dynamic host configuration protocol</i>
DIUC	<i>downlink interval usage code</i>
DL	<i>downlink</i>
ESF	<i>extended subheader field</i>
FTP	<i>file transfer protocol</i>
FUSC	<i>full usage of subchannels</i>
GMH	<i>generic MAC header</i>
GMS	<i>grant management subheader</i>
HUMAN	<i>high-speed unlicensed metropolitan area network</i>
HT	<i>header type</i>
IE	<i>information element</i>
IFFT	<i>inverse fast fourier transform</i>
IP	<i>internet protocol</i>
ITU	<i>International Telecommunications Union</i>
LOS	<i>line-of-sight</i>
LSB	<i>least significant bit</i>
MAC	<i>medium access control layer</i>
MAN	<i>metropolitan area network</i>
ML	<i>maximum latency</i>

MRTR	<i>minimum reserved traffic rate</i>
MS	<i>mobile station</i>
MSB	<i>most significant bit</i>
MSTR	<i>maximum sustained traffic rate</i>
NLOS	<i>non-line-of-sight</i>
NTP	<i>network time protocol</i>
OFDM	<i>orthogonal frequency division multiplexing</i>
OFDMA	<i>orthogonal frequency division multiple access</i>
PMP	<i>point-to-multipoint</i>
PBR	<i>piggyback request</i>
PDU	<i>protocol data unit</i>
PHS	<i>payload header suppression</i>
PHY	<i>physical layer</i>
PKM	<i>privacy key management</i>
PM	<i>poll-me bit</i>
PS	<i>physical slot</i>
PSH	<i>packing subheader</i>
PUSC	<i>partial usage of subchannels</i>
QAM	<i>quadrature amplitude modulation</i>
QoS	<i>quality of service</i>
QPSK	<i>quadrature phase-shift keying</i>
REQ	<i>request</i>
RNG	<i>ranging</i>
RSP	<i>response</i>
RTP	<i>real time transport protocol</i>
RTG	<i>receive/transmit transition gap</i>
SC	<i>single carrier</i>
SDU	<i>service data unit</i>
SF	<i>service flow</i>
SFID	<i>service flow identifier</i>
SIP	<i>session initiation protocol</i>
SNMP	<i>simple network management protocol</i>
SNR	<i>signal-to-noise ratio</i>
SS	<i>subscriber station</i>
STC	<i>space time coding</i>
TCP	<i>transmission control protocol</i>
TDD	<i>time division duplex or duplexing</i>
TDM	<i>time division multiplexing</i>
TDMA	<i>time division multiple access</i>
TJ	<i>tolerated jitter</i>
TP	<i>Traffic Priority</i>

TTG	<i>transmit/receive transition gap</i>
UDP	<i>user datagram protocol</i>
UL	<i>uplink</i>
VBR	<i>variable bit rate</i>
VCI	<i>virtual channel identifier</i>
VPI	<i>virtual path identifier</i>
WiFi	<i>wireless fidelity</i>
WiMAX	<i>worldwide interoperability for microwave access</i>

Variáveis

B_i	requisito de largura de banda para a conexão i , em bps
BW	largura de banda do canal em Hz
C_i	serviço de escalonamento da conexão i
CP	<i>Cyclic Prefix</i>
Δf	espaçamento entre subportadoras
F_s	frequência de amostragem
FPS	<i>frames per second</i>
G	razão entre o tempo de guarda e o tempo útil de símbolo
i	identificador de cada conexão
n	fator de amostragem
N_{FFT}	menor potência de dois superior a N_{used}
N_{used}	número de subportadoras
NS_i	número de subcanais por <i>subframe</i>
R_i	solicitação em <i>bytes</i> da conexão i por <i>subframe</i>
Sb_i	número de <i>bytes</i> por símbolo OFDMA, para a conexão i
T_b	tempo útil de símbolo
T_g	tempo de guarda ou tempo do CP
T_s	tempo de símbolo
TS	número de símbolos OFDMA por <i>subframe</i>

Capítulo 1

Introdução

A demanda dos usuários por serviços *triple play* (termo que se refere ao tráfego de dados, voz e vídeo) e *quadruple play* (termo que se refere ao triple play com mobilidade) apresenta-se como um grande desafio para as operadoras de telecomunicações. Atualmente, as redes celulares já disponibilizam serviços de acesso à WEB com mobilidade e a transmissão de vídeo-chamadas. Estas redes não foram concebidas para transmissão de dados, mas de voz [1]. A mobilidade dos usuários e o fornecimento de serviços de voz e vídeo sobre IP, aumentam os requisitos das redes sem fio.

Outra questão, de substancial importância, é o empenho dos governos dos países em desenvolvimento para inclusão digital das suas populações de baixa renda. Os custos de levar acesso à Internet, por exemplo, para áreas rurais em regiões remotas destes países é extremamente elevado. Tecnologias como WiFi (do inglês, *Wireless Fidelity*), DSL (*Digital Subscriber Line*), GSM (*Global System for Mobile communications*), ISDN (*Integrated Services Digital Network*) e as relativamente novas tecnologias 3G não estão sendo capazes de fornecer uma solução completa para extinguir a *Digital Divide* [2] (dívida digital, em português). Algumas vantagens do WiMAX podem fazer com que esta tecnologia apresente-se como uma solução para a dívida digital: menor custo de infra-estrutura do que as atuais tecnologias metropolitanas cabeadas e sem fio, alta segurança dos dados, arquitetura flexível, normatização internacional que permite o desenvolvimento de equipamentos com baixo custo e interoperáveis e a garantia de qualidade de serviço de acordo com as diversas aplicações permitidas [2].

O padrão IEEE 802.16-2001, revisado em 2004 [3], foi criado com o objetivo de transportar dados multimídia, em redes locais e metropolitanas fixas. Na versão IEEE 802.16e [4], a norma passou a permitir também o acesso móvel. Estes padrões são conhecidos comercialmente como WiMAX, nome dado pelo WiMAX Forum [5], entidade responsável pela certificação destes sistemas. O WiMAX é capaz de prover serviços de redes locais e metropolitanas, sejam elas fixas, portáteis ou móveis.

Os padrões [3] e [4] definem diversas características de camada física, como o formato

do sinal a ser transmitido, faixas de frequência de operação, tecnologia de multiplexação, modulação e codificação e define características da camada de enlace, como os parâmetros de qualidade de serviço, a tecnologia de múltiplo acesso e os serviços de escalonamento. Mas apesar de definir as linhas gerais que os serviços de escalonamento devem possuir, as normas não definem um algoritmo específico, deixando a implementação para os fabricantes de equipamentos. Um dos objetivos deste trabalho é então propor uma estratégia de escalonamento para camada de enlace do WiMAX.

1.1 Motivações

Desde o primeiro semestre do Programa de Mestrado, o autor se dedica ao tema de alocação de recursos. Inicialmente, participou de um estudo em alocação de canais para redes de comunicações móveis, principalmente redes celulares [6]. Este trabalho resultou em uma publicação no PIMRC 2008, importante simpósio na área de telecomunicações, avaliado com Qualis A1 pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Em seguida, iniciou-se o estudo de alocação de recursos para rede de distribuição de conteúdo, que culminou com a definição do tema desta dissertação na alocação de recursos na camada de enlace de redes WiMAX.

Além da fase inicial em que se encontra a utilização do WiMAX no Brasil e no mundo, o grande número de recentes trabalhos publicados na área, como é possível notar através das referências bibliográficas ([7] - [8]), envidando esforços para desenvolver técnicas capazes de alocar recursos de forma rápida e precisa, serviu de motivação para a realização deste trabalho.

As redes WiMAX permitem a inclusão de novas aplicações de redes metropolitanas como sistemas de localização (a custo menos elevado do que os que utilizam satélites), redes privadas de abrangência metropolitana, permitindo mobilidade aos agentes das empresas com garantia de QoS, disponibilização de acesso a público Internet (com um custo de infraestrutura menor do que com as redes WiFi), além de ser uma opção para interatividade na TV digital (canal de *uplink*).

Adicionalmente, observa-se que, nas redes de quarta geração, cresce de importância o suporte à camada física em OFDMA (*orthogonal frequency division multiple access*). Esta técnica de múltiplo acesso ganhou destaque com a sua escolha como camada física de tecnologias de comunicações sem fio recentes, como *Long Term Evolution*, *Ultra Mobile Broadband*, IEEE 802.20, além do WiMAX ([9], [10], [11] e [12])

Esta dissertação tem o objetivo de fazer um estudo do padrão IEEE 802.16, concentrando-se na camada de enlace e nos serviços de escalonamento, realizando um estudo sobre as técnicas existentes, remontando à origem histórica dos mesmos, bem como propor uma nova estratégia de escalonamento através da adaptação de uma estratégia existente.

1.2 Apresentação do problema

Uma estratégia de escalonamento deve ser capaz de decidir, com base nas requisições de tráfego de *uplink* recebidas dos usuários, qual a melhor forma de distribuir os recursos da rede entre as estações. O algoritmo é executado na estação rádio base e pertence à camada de parte comum, subcamada da camada de enlace.

A camada de enlace e a estratégia de escalonamento devem possuir suporte à camada física utilizada. Métodos de acesso múltiplo por divisão do tempo requerem uma estratégia de alocação de recursos unidimensional, enquanto o acesso múltiplo por portadoras ortogonais demanda uma técnica bidimensional.

A técnica de escalonamento deve ser capaz de distribuir recursos de forma justa entre as estações de mesmo tipo de tráfego e atender aos requisitos de qualidade de serviço para classes de tráfego distintas. Esta distribuição deve ser efetuada da forma mais simples possível, pois uma estação rádio base WiMAX deve ser capaz de despachar até 400 *frames* por segundo [13].

Para esta distribuição, considera-se a necessidade de atendimento a demandas a serem caracterizadas em cenários que incluam serviços de VoIP com supressão de silêncio, IPTV, FTP e navegação WEB. Adicionalmente, com base nesses cenários, deverá ser avaliada a característica de *fairness* para estações com serviço *best effort*, bem com a adaptabilidade aos perfis de transmissão do WiMAX (característica necessária para permitir a mobilidade do usuário).

1.3 Objetivos

Após a análise de estratégias de escalonamento para redes WiMAX, vislumbrou-se a possibilidade de produzir uma nova técnica de escalonamento para a camada de enlace, que forneça suporte à camada física em OFDMA. Pode-se citar como objetivos desta dissertação:

- realização de um *survey* sobre as estratégias de escalonamento;
- adaptação do módulo do NDSL [14], *Networks & Distributed Systems Laboratory*, para consideração de parâmetros de qualidade de serviço;
- adaptação do módulo do NDSL [14] para inclusão de uma nova técnica de escalonamento;
- adaptação da estratégia de escalonamento proposta em [13] para suporte à camada física em OFDMA PUSC;

- verificação de atendimento às demandas de tráfego associadas aos serviços mencionados na seção 1.2, com base na realização de simulações no *Network Simulator 2*.

1.4 Organização da dissertação

O restante da dissertação contém cinco capítulos. No capítulo 2, serão abordadas as necessidades de comunicação atuais, envolvendo tráfego de tempo real de voz e vídeo e seu tratamento em redes IP. No capítulo 3, é realizado um estudo dos aspectos mais importantes dos padrões IEEE 802.16 para continuidade deste trabalho, com maior detalhamento para a camada de enlace. No capítulo 4, é realizada uma revisão sobre algoritmos de escalonamento, com referências clássicas sobre o tema e trabalhos mais atuais. No capítulo 5, apresenta-se uma proposta de estratégia de escalonamento para redes WiMAX, os resultados das simulações, fazendo a análise dos mesmos e, por último, o capítulo 6 é destinado à conclusão e à sugestão de trabalhos futuros.

Capítulo 2

Serviços de comunicações multimídia sobre redes sem fio

2.1 Introdução

O atendimento a requisito das redes de comunicação multimídia representa um desafio, inclusive para redes cabeadas. Aplicações como a transmissão de vídeo em tempo real, com exigência de altas larguras de banda e baixa tolerância a *jitter* continuam sendo alvo de pesquisas, envolvendo diferentes aspectos e alternativas de redes de distribuição de conteúdo (veja, por exemplo, [15], [16] e [17]).

Do ponto de vista da utilização da rede, cada vez mais usuários buscam utilizar um ou mais equipamentos para se manter conectados a provedores de serviços, podendo efetuar a leitura de correios eletrônicos, receber e efetuar ligações telefônicas, assistir à programação da TV, navegar na WEB, dentre outros serviços. Por outro lado, *handsets* e computadores móveis, como notebooks e PDA's (*personal digital assistants*), contituem um dos segmentos de mais rápido crescimento da indústria de informática [1].

As redes móveis sem fio são uma opção para atender às demandas deste tipo de usuário. No caso de tais redes, muitas aplicações surgiram com a popularização de redes locais sem fio, conhecidas como *WiFi*, como o acesso WEB para visitantes em empresas, restaurantes, hotéis, entre outros, bem como com a utilização de RFID para triagem de produtos em indústrias e *chats* para envio de mensagens em conferências e congressos. Quando a extensão necessária ultrapassa algumas centenas de metros ou os requisitos de qualidade de serviço são elevados, uma *WiFi Zone* deixa de atender à demanda [1].

Com o advento das tecnologias de terceira geração, as redes celulares, se apresentam também como uma opção para serviços multimídias, como a vídeo-chamada e a transmissão de sinais de TV ao vivo. As redes celulares possuem abrangência em regiões metropolitanas, ao contrário das redes *WiFi*. Entretanto, estas redes não foram projetadas para transmissão

de dados, mas de voz [1]. A inclusão de requisitos de qualidade de serviço, através de acordos de nível de serviço, também é um grande desafio para as redes celulares [2].

Precisa-se de uma rede que tenha vocação para o tráfego multimídia, de tempo real ou não, podendo transmitir sinais de voz e sinais áudio e vídeo encapsulados em diversos protocolos mantendo a qualidade dos serviços. O WiMAX é uma rede sem fio que pode possuir maior abrangência, com base no padrão IEEE 802.16 [3]. Em especial, a versão baseada no padrão IEEE 802.16e [4] permite mobilidade do usuário em região metropolitana respeitando os critérios de qualidade especificados. A garantia de interoperabilidade e baixo custo dos equipamentos, maior rapidez na implantação, flexibilidade de arquitetura e elevados critérios de segurança são outras vantagens das redes WiMAX frente as redes celulares [2]. A adoção do WiMAX como uma opção para o IMT2000 do ITU aumentou a competitividade desta tecnologia, frente as redes celulares [2]. Nas demais seções deste capítulo faremos um estudo sobre os serviços de VoIP e IPTV, partindo dos fluxos de áudio e vídeo até a apresentação de uma proposta de arquitetura de QoS para os mesmos sobre uma rede WiMAX.

2.2 Fluxo de áudio

Devida à exigência de alta largura de banda, já é difícil fazer uma comunicação multimídia funcionar sobre conexões fixas. Em [1] considera-se a distribuição de vídeos com qualidade VHS, em uma rede sem fio, algo distante. Multimídia literalmente quer dizer duas mídias ou mais como, por exemplo, os textos e os gráficos desta dissertação. Mas quando as pessoas se referem à multimídia, em geral, elas querem se referir à combinação de dois meios contínuos simultaneamente. Na prática, os dois meios mais comuns são áudio e vídeo [1]. Apesar desta definição, aplicações com áudio puro, em tempo real, associados à Internet, têm sido referenciadas como multimídia (um termo mais correto seria mídia de fluxo [1]).

Um sinal sonoro é uma onda mecânica que se propaga longitudinalmente em qualquer meio, exceto no vácuo. O ouvido humano é capaz de captar frequências entre 20 e 20.000 Hz. A fala apresenta uma amplitude em frequência entre 600 e 6.000 Hz. Para transmissão dos sinais, através de uma de transmissão de dados, é necessária a conversão analógica/digital dos sinais sonoros. A modulação de código por pulso utilizada nos sistemas de telefonia tradicional utiliza 8.000 amostras por segundo, captando frequências até 4.000 Hz. Um CD de áudio com taxa de 44.100 amostras por segundo pode capturar até 22.050 Hz. Para transmitir um som estéreo, um CD de 44.100 amostras de 16 bits cada precisa de uma largura de banda de 1,411 Mbps [1].

Para possibilitar a transmissão dos sinais digitalizados através de redes de computadores como a Internet, é fundamental o processo de compactação. Arquivos de áudio digitalizado

se tornaram populares na Internet com a adoção do formato MP3 (MPEG *audio layer 3*).

Esse fluxo de áudio (que pode ser, por exemplo, a transmissão de rádio ou comunicação de voz - VoIP) pode então ser transmitido em tempo real por servidores de mídia utilizando os protocolos RTP e UDP através da Internet. O receptor deve ser capaz de lidar com a flutuação do atraso (*jitter*) utilizando *buffers* para diminuição dos seus efeitos.

Atualmente, porém, são os serviços de VoIP que despertam maior interesse dos usuários domésticos e das corporações. Se este serviço for prestado em região metropolitana utilizando meios não cabeados, o interesse poderá ser ainda maior. Para este fim, abordaremos o fornecimento de serviços VoIP sobre uma rede WiMAX na seção 2.3.

2.3 Serviços de VoIP

A maior parte da informação que trafega nas redes das operadoras de telecomunicações é voz [18]. Esta era a realidade em 2001, é a realidade hoje e deverá continuar sendo a realidade no futuro próximo [19]. As redes comutadas por circuitos são capazes de garantir alta qualidade para o transporte da voz, mas podem ser extremamente ineficientes do ponto de vista da utilização dos recursos. Por outro lado, as rede comutadas por pacotes apesar de apresentarem maior eficiência podem ser desastrosas para a qualidade de tráfego de voz. A tecnologia VoIP (do inglês, *Voice over IP*) pretende alcançar a eficiência das redes comutadas por pacote com a qualidade de voz percebida nas redes comutadas por circuitos.

As aplicações de voz possuem característica de tempo real, não sendo tolerantes a atrasos demasiados, perda de pacotes, pacotes fora de ordem e *jitter*. As redes sem fio tendem a intensificar estes problemas em relação às redes cabeadas, pois possuem maior latência, maior frequência de perda de pacotes e maior *jitter* [18].

Em termos de arquitetura, o provedor de serviços VoIP deve ter componentes em sua rede de no mínimo quatro categorias: codificação / decodificação, sinalização, transporte e um *media gateway* (conversor de mídia para entrada e saída). A figura 2.1 apresenta uma arquitetura básica para serviços VoIP [19]. Os codificadores (o mais utilizado é o G.711) são os responsáveis pela quantização, formatação e compactação dos sinais de áudio para transmissão; a sinalização, realizada por diversos protocolos (como SIP, H.323, SS7), é responsável pelo estabelecimento da sessão e controle das chamadas; a rede de transporte é responsável por encapsular os dados com cabeçalhos de controle e endereçamento (UDP, RTP, IP, etc.) e os *gateways* fazem a transição entre o mundo comutado por circuitos e o mundo comutado por pacotes.

As figuras 2.2 e 2.3, retiradas de [19], apresentam pilhas de protocolos para serviços VoIP, com a utilização de SIP ou H.323. As camadas de enlace e física podem ser acrescentadas às pilhas de protocolos apresentadas, a exemplo da figura 2.7 para o IPTV.

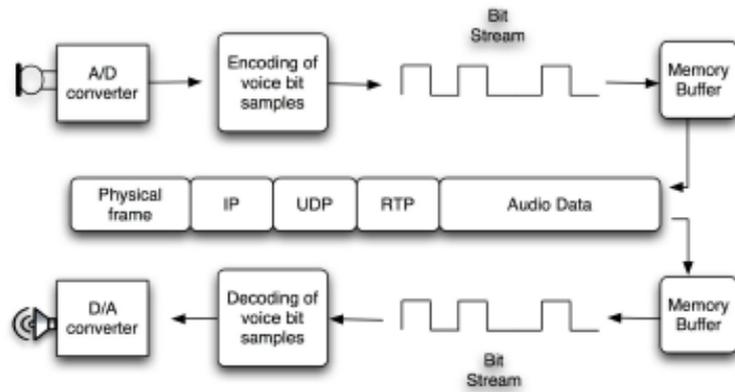


Figura 2.1: Arquitetura Funcional para serviços VoIP

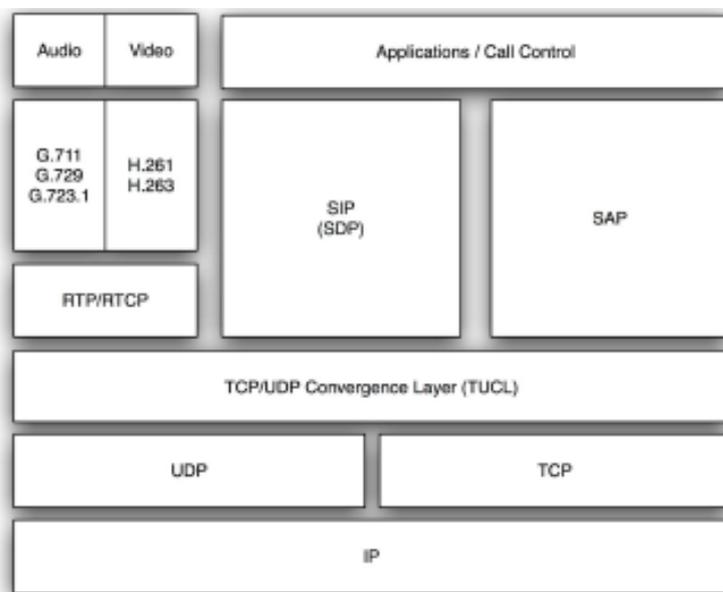


Figura 2.2: Pilha de protocolos para serviços VoIP utilizando SIP

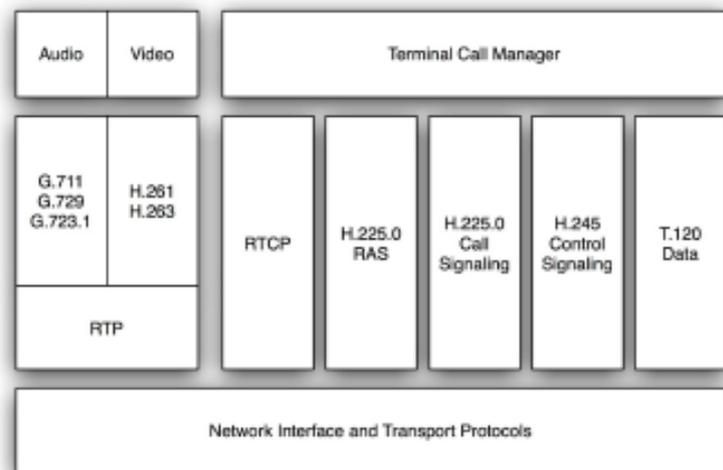


Figura 2.3: Pilha de protocolos para serviços VoIP utilizando H323

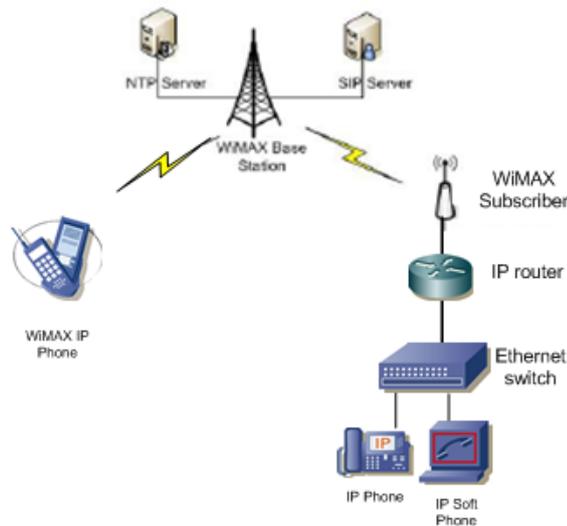


Figura 2.4: Arquitetura de VoIP sobre WiMAX

A utilização de IP para roteamento da voz em redes corporativas está se tornando uma prática cada vez mais comum. Companhias conseguem diminuir seus custos de serviços de telecomunicações utilizando o *backbone* existente, também, para interligação de centrais telefônica via VoIP [20]. Desta forma, empresas podem realizar ligações entre os ramais utilizando o seu canal de dados já contratado. Nestas corporações, já existe larga utilização de redes *WiFi* para permitir a mobilidade dos funcionários entre os andares de um grande prédio ou entre prédios próximos de uma mesma empresa. Parece fazer sentido que este acesso sem fio também permita o tráfego de voz. O baixo nível de QoS e os problemas da camada de múltiplo acesso das redes WiFi, entretanto, praticamente inviabilizam o uso de VoIP sem adaptações para solução destas questões [20].

A rede WiMAX foi criada com o objetivo de suportar o tráfego multimídia e possui características para camada de enlace, como níveis de prioridade e serviços de escalonamento específicos para o tráfego VoIP, com a vantagem de permitir a mobilidade de agentes daquela mesma empresa, por exemplo, dentro de uma região metropolitana, com garantia de qualidade de serviço. A figura 2.4, adaptada do trabalho [21], apresenta uma arquitetura genérica para serviços VoIP sobre uma rede WiMAX baseada no protocolo SIP. A arquitetura propõe um NTP (*Network Time Protocol*) *Server* para sincronismo de tempo considerando que os usuários podem em regiões distantes uma da outra.

Em termos de qualidade de serviço, a arquitetura de QoS para camada MAC apresentada na figura 2.8, inicialmente proposta para IPTV, também pode ser utilizada para serviços VoIP sobre WiMAX. A figura 2.8 foi complementada a partir do trabalho [22]. Nesta arquitetura, o tráfego VoIP é escalonado pelo serviço eRTPS, para o tráfego de DL. A arquitetura também é válida para o escalonamento do tráfego no canal de UL, com a verificação das requisições das estações usuárias do mesmo.

2.4 Fluxo de vídeo

Da mesma forma como o ouvido humano, a nossa visão também possui limitações. Uma imagem projetada a uma velocidade de 50 quadros ou mais por segundo é incapaz de ser diferenciada de uma imagem contínua pelo olho humano. Os sistemas de vídeo são baseados neste princípio.

A origem destes sistemas de vídeo é a televisão em preto-e-branco. As imagens são formadas no monitor através da rápida varredura horizontal e a lenta varredura na vertical registrando a intensidade da luz durante o percurso. Os parâmetros da varredura variam de país para país. No sistema americano, existem 525 linhas de varredura, uma relação entre eixos de 4:3 e são reproduzidos 30 quadros por segundo, enquanto no sistema europeu, a varredura é de 625 linhas, com a mesma relação entre eixos e uma velocidade de reprodução de 25 quadros por segundo [1].

A uma velocidade de 25 quadros por segundo muitas pessoas poderiam notar falhas na reprodução. Adotou-se a técnica do entrelaçamento reproduzindo linhas ímpares e em seguida as linhas pares. Cada um destes conjuntos de linhas recebe o nome de campo. O olho humano é incapaz de perceber as falhas a uma velocidade de 50 campos por segundo [1].

A evolução para a imagem colorida é resultado da combinação de 3 varreduras, uma para cada cor primária: vermelho, verde e azul, o padrão RGB. No entanto, outros sistemas mais eficientes foram propostos como o NTSC (*National Television Standards Committee*), o SECAM (*Sequentiel Couleur Avec Memorie*) e o PAL (*Phase Alternating Line*). Para permitir que as transmissões a cores pudessem ser recebidas por aparelhos em preto-e-branco, os três sistemas combinam os sinais RGB com um sinal de luminância (brilho) e com dois de crominância (cores). Nos últimos anos, a HDTV (High Definition Television) apresenta-se como a evolução da televisão, com imagens mais nítidas em razão do maior número de linhas de varredura e razão entre eixos de 16:9.

Para a digitalização das imagens, foi criado o conceito de elemento de imagem ou *pixel*. Um sinal de vídeo digital é formado por um conjunto de quadros, e cada quadro é formado por uma grade retangular de elementos de imagem e reproduzido a velocidade 25 quadros por segundo (assim como os analógicos). Quanto maior o número de bits utilizados para representar um *pixel*, maior a qualidade da imagem. Oito bits para cada cor geram uma imagem de alta qualidade, possibilitando 24 bits por *pixel* e mais de 16 milhões de cores [1].

Monitores digitais de boa qualidade repetem a varredura a partir de imagens memorizadas a uma frequência de 75 vezes por segundo, dispensando o entrelaçamento. Mas a largura de banda necessária para transmissão de 24bits por *pixel* a 25 quadros por segundo, com uma relação entre eixos de 1.024 x 768, é superior a 471 Mbps. Ou seja, para que pos-

sibilita a transmissão de sinais de vídeo digitais, é necessário a compactação das imagens [1].

A compactação dos vídeos surgiu da compressão de imagens estáticas. O padrão JPEG (*Joint Photographic Experts Group*) surgiu para compactação de imagens fotográficas, resultado do esforço conjunto do ITU, ISO e IEC. O MPEG (*Motion Pictures Experts Group*) é apenas a codificação JPEG aplicada quadro a quadro com recursos extras de compactação entre os quadros.

Nos filmes, a dois tipos de redundância: espacial e temporal. O MPEG faz o uso de ambas. A redundância espacial pode ser utilizada pela própria codificação JPEG. Uma compactação adicional pode ser realizada verificando a diferença do quadro atual para o quadro antecessor, compactando com o JPEG apenas a diferença. Porém, para imagens em que o pano de fundo é alterado rapidamente ou que os movimentos são bruscos, esta técnica é pouco eficaz. A grande vantagem do MPEG sobre o JPEG é a existência de diferentes tipos de quadros [1]:

- quadros I (*Intracoded*) - imagens estáticas, independentes codificadas com o JPEG;
- quadros P (*Predictive*) - diferença bloco a bloco em relação ao último quadro;
- quadros B (*Birectional*) - diferenças entre o último quadro e o quadro seguinte; e
- quadros D (*DC-coded*) - médias de blocos usadas para avanço rápido.

Os blocos e macroblocos são úteis para delimitar espaços na tela onde ocorrem alterações, permitindo a codificação apenas da diferença da imagem anterior. Demais detalhes do processo de codificação estão além do escopo deste trabalho. Desta forma, é possível reduzir a largura de banda necessária para transmissão do sinal de vídeo digital para algo em torno de 4 a 8 Mbps.

A redução da largura de banda necessária para transmissão de vídeos de boa qualidade possibilitou o fornecimento de serviços como vídeo sob demanda e distribuição de vídeo em tempo real. Estes vídeos podem ser distribuídos através da Internet ou por outras redes de computadores, encapsulados por outros protocolos que não o IP. Na seção 2.5 apresentaremos arquitetura para fornecimento de serviços de IPTV sobre uma rede WiMAX.

2.5 Serviços de IPTV

Assistimos desde a década de 60, o fenômeno digitalização e eventual entrega empacotada da voz sobre o Internet Protocol (IP), chamado VoIP (do inglês, *Voice over IP*) [23]. Neste novo milênio, poderemos assistir a um fenômeno parecido com os sinais de TV. Os

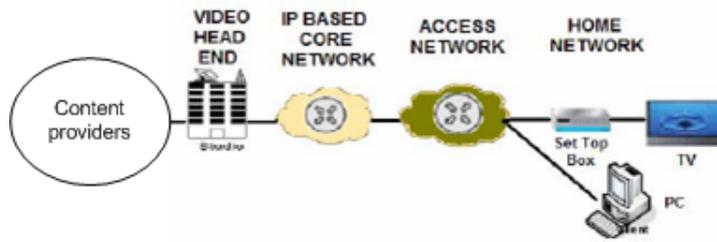


Figura 2.5: Arquitetura Genérica de IPTV

avanços nas tecnologias de acesso a Internet e em técnicas de codificação de vídeo tornaram possível a utilização do IPTV (do inglês, Internet Protocol Television) [24].

Muitos fornecedores de serviço de dados, telefonia e TV a cabo têm migrados seus serviços para *triple-play* (dados, voz e vídeo) ou até mesmo *quadruple-play* (*triple-play* com mobilidade). Paralelamente percebe-se o avanço tecnológico para miniaturização dos equipamentos tornando-os cada vez mais portáteis, melhorias no consumo de energia dos equipamentos e em tecnologias de comunicação sem fio, como WiMAX [24].

Levando-se em consideração estas informações, a transmissão de IPTV sobre WiMAX apresenta uma integração entre os desenvolvimentos tecnológicos mencionados e a disponibilização de um serviço completo ao usuário. Como o IPTV é um serviço que consome altas quantidades de largura de banda e, para transmissão ao vivo, também possui requisitos de tempo bastante rígidos, é fundamental que a infra-estrutura por trás deste serviço possa garantir qualidade para o usuário.

Para resolver os desafios do IPTV sobre o WiMAX, é necessário, primeiramente, conhecer os componentes básicos da arquitetura de IPTV. Esta seção tem o objetivo de apresentar estes componentes baseando-se nos trabalhos de [22], [24] e [25].

2.5.1 Arquitetura de IPTV

Uma arquitetura genérica para entrega de IPTV ao usuário deve possuir no mínimo os seguintes blocos de componentes: consumidor (*customer premises*), *video headend* e a rede de transporte (*transport networks*), ilustrada na figura 2.5, adaptada de [25]. Descreveremos, nas próximas subsecções, alguns componentes destes blocos.

2.5.1.1 Consumidor

Chamaremos de Set-Top Box (STB) o dispositivo do consumidor que fará a interface entre o terminal (televisão, PC, entre outros) e a rede do provedor de serviço. No STB é instalado um *middleware* que recebe a programação, decodifica vídeos MPEG2 e MPEG4 e mostra na tela. Em um STB, podem ser integrados modems DSL ou via cabo, acesso WiFi

e futuramente WiMAX para fornecimento de acesso a Internet.

2.5.1.2 *Headend*

Um *Headend* de vídeo é composto dos seguintes componentes: codificador de vídeo, servidores de vídeo (ao vivo e sob demanda) e gerenciamento de conteúdo e usuários [24]. Os codificadores de vídeo são responsáveis por receber vídeos dos provedores de conteúdo, digitalizá-los caso analógicos e comprimí-los utilizando as tecnologias MPEG2 e MPEG4. Os vídeos recebidos podem ser transmissões ao vivo ou conteúdos para serem armazenados e transmitidos sob demanda, após os processos de direitos autorais digitais e de encriptação.

Servidores de vídeo ao vivo para transmissão *broadcast* são os responsáveis por reformatar e encapsular os *streams* de vídeo, dos mais diferentes formatos, e transmití-los à rede de núcleo (em inglês, *Core Network*) para serem entregues às redes de acesso (em inglês, *Access Networks*). Servidores de conteúdo sob demanda para transmissão *unicast* possuem as mesmas características, mas normalmente são dotados de grande capacidade de armazenamento, tendo em vista que o conteúdo pode ficar armazenado por grandes períodos.

Os gerenciadores de conteúdo e de usuários são fundamentais para o bom funcionamento do serviço de IPTV. O controle do faturamento, perfis e pacotes de serviços dos usuários, autenticação são algumas das funções do gerenciador de clientes ou usuários. O gerenciador de conteúdo deve possuir controle sobre a data de expiração de conteúdos, controle dos direitos autorais e de parâmetros de segurança para o conteúdo e controle de entrega do serviço aos clientes.

2.5.1.3 Rede de Transporte

A rede de transporte para entrega do serviço de IPTV aos usuários se divide em duas partes: redes de núcleo (em inglês, *Core Network*) e rede de acesso (em inglês, *Access Network*). A rede de núcleo é a conexão entre o *Headend* e a rede de acesso, normalmente constituída por uma conexão nacional em Gigabit Ethernet. O conteúdo gerenciado normalmente é centralizado e processado para depois ser distribuído à rede de acesso. Já o conteúdo não gerenciado, pode ser enviado, por meio da rede de núcleo, diretamente para redes de acesso para consumidores de toda Internet.

As redes de acesso são as responsáveis por fazer a conexão da rede de transporte com o usuário final, alcançando cada STB individualmente. Entre as tecnologias que executam esta função atualmente estão: xDSL (do inglês, Digital Subscriber Line, onde o x representa todas as tecnologias DSL), HFC (do inglês, coaxial hybrid fiber cable) e FTTN (do inglês, fiber-to-the-node). Tendo em vista a limitação de largura de banda das redes de acesso, a tecnologia de *multicasting* está sendo cada vez mais utilizada, evitando a duplicação

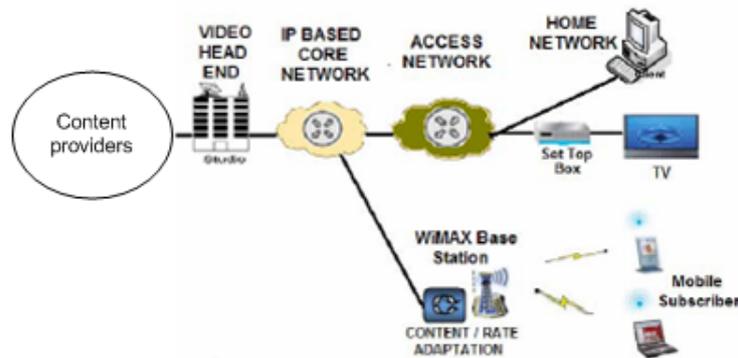


Figura 2.6: Arquitetura de IPTV sobre WiMAX

desnecessária de pacotes como ocorre com o formato *unicast* ao transmitir um mesmo canal em uma mesma rede de acesso ou até mesmo em toda rede de transporte. Desta forma, pacotes são duplicados apenas quando realmente necessário, por exemplo, em consumidores de regiões distantes uma da outra.

Novas redes de acesso estão surgindo no mercado. Operadoras de telefonia móvel já estão oferecendo o serviço de TV móvel utilizando tecnologias das gerações 2.5G e 3G. Avançadas tecnologias de compressão permitem a transmissão de vídeos sob demanda, mas com capacidade limitada, utilizando as redes celulares existentes. Neste cenário, muitos provedores de serviços esperam uma solução *quadruple-play* (TV, telefonia, Internet, com mobilidade) para as redes de acesso, em uma única infra-estrutura IP.

Em [25] é proposta uma arquitetura para entrega de serviço de IPTV utilizando uma rede de acesso em WiMAX. Esta arquitetura é apresentada na figura 2.6, adaptada de [25]. A rede WiMAX é uma opção com vocação natural para o tráfego multimídia com a garantia de qualidade de serviço. Em [22] é apresentada a pilha de protocolos para transmissão de IPTV sobre uma rede de transporte WiMAX, conforme figura 2.7. O fluxo de bits gerado por servidores de conteúdo, formado por sinais de áudio e vídeo, é formatado e codificado (utilizando MPEG-2 ou MPEG-4) e encapsulado pelos protocolos RTP e UDP (para vídeos ao vivo) ou RTP e TCP (para vídeos sobre demanda). Em seguida, estes sinais são encapsulados pelo protocolo IP e pelos cabeçalhos de camada MAC e PHY da rede WiMAX [22].

A camada MAC WiMAX possui uma subcamada de convergência responsável pelo tratamento específico de protocolos baseado em pacotes, como o IP por exemplo, e uma subcamada de parte comum responsável por alocar os recursos da rede. Mais detalhes sobre estas camadas serão apresentados no capítulo 3. O grande desafio atual é encontrar os mecanismos para disponibilizar os recursos da rede, tendo em vista os diferentes tipos de demanda. Uma rede WiMAX é suficientemente flexível para o tratamento da pilha de protocolos proposta.

Em [22] é apresentada uma arquitetura de QoS para a camada MAC com foco na

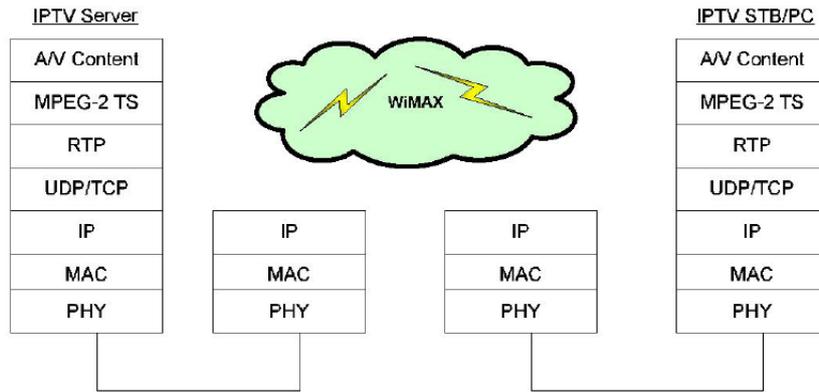


Figura 2.7: Pilha de protocolos para transmissão de IPTV

transmissão de IPTV, conforme figura 2.8 (retirada de [22]). É proposta a divisão em filas para cada serviço de escalonamento, considerando quatro tipos existentes em [3]. Fazem parte da arquitetura o classificador de tráfego, as políticas de QoS, o controle de conexão, mensagens de gerenciamento da camada MAC e a estratégia de escalonamento.

O classificador de tráfego é a estrutura responsável por distribuir os pacotes entre as filas correspondentes. Um conjunto de políticas de QoS deve ser associado a cada fluxo de serviço que por sua vez é associado a uma conexão. Para um serviço de IPTV, os parâmetros *Minimum Reserved Traffic Rate* (MRTR), *Maximum Sustained Traffic Rate* (MSTR), *Maximum Latency* e *Request/Transmission Policy* são obrigatórios [22]. Os parâmetros de QoS dos padrões [3] e [4] serão detalhados no capítulo 3. O controle de conexão é responsável por gerenciar as conexões de gerenciamento e as conexões de transporte. O gerenciamento de conexões será detalhado no capítulo 3. As mensagens de gerenciamento da camada MAC têm a função, entre outras, de transmitir aos usuários quais recursos foram reservados para sua utilização. A estratégia de escalonamento é a responsável por definir esta reserva de recursos. A estrutura da figura 2.8, adaptada de [22], é proposta para o tráfego de *downlink*.

Incluindo a última categoria de serviços de escalonamento proposta em [4], esta estrutura pode ser utilizada para o escalonamento do tráfego de *uplink*, recebendo as requisições de largura de banda dos usuários. De posse destas requisições, a estação rádio-base deve ser capaz de alocar os recursos da rede para atendê-las, respeitando os parâmetros de qualidade de serviço acordados.

Um dos objetivos deste trabalho é fornecer uma estratégia de escalonamento para subcamada de parte comum, atendendo ao bloco de *scheduling* da arquitetura da figura 2.8. Esta estratégia de escalonamento deverá respeitar os parâmetros de qualidade obrigatórios das normas [3] e [4], para todas as classes de serviços, inclusive para um tráfego de tempo real, considerando cabeçalhos presentes em vídeos IPTV ao vivo.

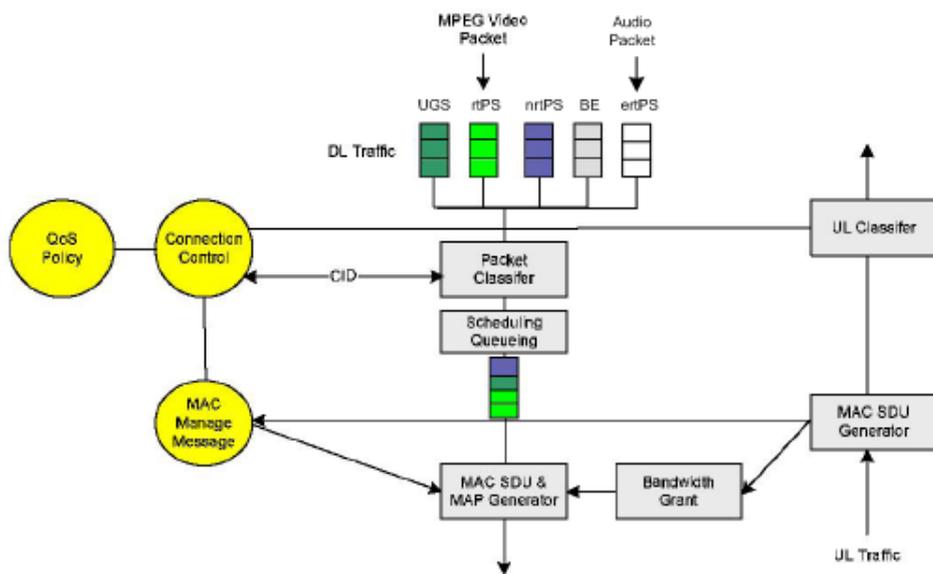


Figura 2.8: Arquitetura de QoS de camada MAC para transmissão de IPTV e VoIP

Capítulo 3

WiMAX

3.1 Introdução

A tecnologia *WiMAX* (do inglês, *Worldwide Interoperability for Microwave Access*) é o nome dado por um grupo de indústrias, reunidas sob o nome de *WiMAX Forum*, ao padrão IEEE 802.16. O padrão IEEE 802.16-2004 - *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems* ([3]) - define as características de camada física (PHY) e de camada de enlace (MAC) para Redes Locais e Metropolitanas de Banda Larga, sem fio. Este padrão suporta Acessos Fixos e Nomádico [26]:

- acesso fixo (*fixed access*) - o dispositivo do usuário é suposto fixo em uma determinada posição geográfica durante sua subscrição na rede. O dispositivo do usuário pode conectar e desconectar da rede. Poderá selecionar a melhor estação base no momento da sua entrada na rede. O dispositivo do usuário será associado tipicamente com o mesmo setor ou estação base.
- acesso nomádico (*nomadic access*) - o dispositivo do usuário é suposto fixo em uma mesma posição geográfica, pelo menos enquanto uma sessão de serviço está em operação. Se o dispositivo do usuário for movido para uma posição diferente na mesma rede, a subscrição do dispositivo do usuário é reconhecida e uma nova sessão de serviço pode ser estabelecida. O dispositivo do usuário pode selecionar a melhor estação base no momento da sua entrada na rede.

Em 2005, o padrão foi revisado e passou a permitir também comunicações móveis, através do adendo *IEEE 802.16e-2005 -IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1* - ([4]) - possibilitando o Acesso

Portátil(*Portable Access*), Acesso de Mobilidade Simples(*Simple Mobility Access*) e Acesso de Mobilidade Completa(*Full Mobility Access*):

- acesso portátil - o dispositivo do usuário manterá uma sessão de serviço de dados operando, enquanto se move em velocidade de pedestre, nos limites de cobertura de rede. *Handovers* de capacidade limitada são permitidos durante a operação de uma sessão de serviço dos dados, enquanto o dispositivo do usuário move-se de um setor para outro ou entre as células.
- acesso de mobilidade simples - o dispositivo do usuário manterá uma sessão de serviço de dados operando para aplicações de tempo não-real, enquanto move-se em velocidade veicular dentro da área da cobertura da rede. *Handovers* entre setores e estações bases fornecem continuidade do serviço para todas as aplicações de tempo não-real.
- acesso de mobilidade completa - o dispositivo do usuário manterá uma sessão de serviço de dados operando, enquanto move-se, com velocidade veicular elevada, dentro da área de cobertura da rede. O desempenho garantido do *handover* entre setores e estações bases fornece a continuidade do serviço para todas as aplicações.

Em [4], também foi inserido o conceito de MS (do inglês, *mobile station*): uma estação para ser utilizada em serviços móveis. Em [4], toda MS é considerada uma SS (do inglês, *subscriber station*), a menos que seja expressamente dito o contrário.

O padrão IEEE 802.16 inclui as diversas características da camada PHY como, por exemplo, a modulação dos sinais, as faixas de frequências de operação, licenciadas ou não e esquemas de multiplexação. Também define detalhes da camada MAC como, por exemplo, mapeamento de serviços específicos para determinados protocolos (como o IP, por exemplo), definição dos tipos de conexões (básica, gerenciamento e transporte) e, principalmente, os serviços de *scheduling* e os parâmetros de qualidade de serviço.

São permitidas três formas de operação: ponto-multiponto(PMP), obrigatoriamente, *mesh* centralizado e *mesh* distribuído, opcionalmente ([7] e [27]). No modo de operação PMP, todas as estações recebem a mesma transmissão da BS (do inglês, *Base Station*). As transmissões feitas pelas SSs são direcionadas para a BS. Já no modo *mesh* centralizado, a BS controla a comunicação entre as SSs, recebendo informações sobre a topologia da rede em canais de controle. No modo *mesh* distribuído, existe comunicação direta entre as SSs sem controle da BS [7].

No modo PMP, o acesso ao recurso para *downlink* (DL), sentido da BS para as SSs, é exclusivo da BS, enquanto o acesso de *uplink* (UL), sentido das SSs para a BS, é disputado entre as SSs. É fundamental a existência de uma estratégia de reserva de recursos para ambos os canais e principalmente para o UL.

No restante deste capítulo, faremos uma descrição da evolução da norma IEEE 802.16, bem como das principais características das camadas física e de acesso ao meio.

3.2 Evolução do padrão IEEE 802.16

Nesta seção, forneceremos uma breve descrição, em ordem cronológica, das publicações realizadas pelo grupo de trabalho IEEE 802.16. Este grupo de trabalho está vinculado às Sociedades de Comunicação e de Técnicas e Teorias de Microondas do IEEE.

A primeira edição do padrão foi publicada em 2001, *IEEE Std. 802.16-2001* [28], projetada para o acesso fixo (estacionário), permitindo a operação nas frequências entre 10-66 GHz exigindo linha de visada. Existia apenas uma especificação de camada física com portadora única e era permitido o tipo de rede ponto-multiponto. Em 2002, foi publicada uma extensão do padrão, *IEEE 802.16c-2002* [29], alterando características e funções relacionadas diretamente às implementações, sem alterar os fundamentos do padrão.

Em 2003, foi publicada nova extensão da norma, *IEEE 802.16a-2003* [30], que promoveu alterações nas especificações de camada MAC e PHY. A partir desta extensão passou a ser permitida a operação em frequências entre 2-11 GHz, foram incluídas especificações de camada física permitindo tecnologias de múltiplas portadoras. Foi incluída a possibilidade de construções de redes do tipo MESH. Em 2004, foi publicada uma revisão e consolidação das normas [28], [29] e [30], *IEEE 802.16-2004* [3]. Não houve acréscimos de modos de operação ou faixas de frequências, mas alterações relativas aos detalhes de implementação das camadas físicas. Em 2005, uma nova extensão foi publicada, *IEEE 802.16f-2005* [31], com o objetivo, entre outros, de incluir um capítulo específico sobre a estrutura da MIB (do inglês, *management information base*) para coletar as informações do protocolo SNMP relativas ao gerenciamento das estações.

Uma grande contribuição para o padrão ocorreu em 2005 com a publicação da extensão *IEEE 802.16e-2005* [4]. A partir desta extensão, o padrão passou a permitir mobilidade das estações dos assinantes. Tendo em vista estas novas funcionalidade, algumas características relativas ao *handoff* foram incluídas, novos critérios de segurança foram inseridos e a camada MAC também foi revisada, incluindo, entre outros atributos, um novo serviço de *scheduling*, o *Extended Real-Time Polling Service*. Em 2007, duas novas extensões foram publicadas, *IEEE 802.16k-2007* [32] e *IEEE 802.16g-2007* [33], nesta ordem. Em [32], são ajustados alguns objetos para serviços internos da camada MAC entre subcamadas. Em [33], são propostas alterações nas funções, interfaces e protocolos de gerenciamento. Encontra-se em desenvolvimento o padrão IEEE 802.16m, chamado de WiMAX 2.0, que deverá propor avanços ao padrão IEEE 802.16e para obter taxas de transmissão de 100Mbps para usuários móveis e 1Gbps para usuários fixos [34].

3.3 Camada Física (PHY)

O texto das próximas seções deste capítulo é baseado nas normas [3] e [4] e em outras fontes a serem citadas oportunamente. O objetivo principal da camada PHY é o transporte físico dos dados gerados nas camadas superiores. Existem cinco especificações para a camada física WiMAX: WirelessMAN - SC, WirelessMAN - SCa, WirelessMAN - OFDM, WirelessMAN - OFDMA e WirelessHUMAN. A utilização de cada uma das especificações varia de acordo com os seguintes parâmetros:

- faixas do espectro eletromagnético - é permitida a operação em faixas de 10-66 GHz licenciada, faixas abaixo de 11GHz licenciadas e faixas abaixo de 11GHz não-licenciadas (inicialmente limitada a 5-6 GHz), com restrição de potência e utilização de seleção dinâmica de frequências (DFS, do inglês Dynamic Frequency Selection) para evitar interferências;
- propagação de sinais - é possível a operação em sistemas LOS (do inglês, *Line Of Sight*) ou NLOS (do inglês, *Non-Line Of Sight*), dependendo do comprimento de onda e das aplicações;
- tipo de aplicação - dependendo da aplicação será necessária a utilização de determinada tecnologia, como por exemplo, para redes *Mesh* são possíveis apenas as especificações WirelessMAN - OFDM e WirelessHUMAN.

Para a transmissão dos dados (e das informação de inicialização e gerenciamento) é utilizada uma estrutura de *frame* (quadro), que por sua vez é dividida em *subframes* de DL e UL. Em cada *subframe* existe uma quantidade de PSs (do inglês, *Physical Slot*). A utilização dos PSs depende do modo de direcionamento de tráfego utilizado.

Quatro das cinco especificações de camada física permitem o direcionamento da comunicação por divisão de tempo e de frequência, conhecidas pelas siglas em inglês, TDD(time-division duplex) e FDD(frequency-division duplex), respectivamente. A especificação WirelessHUMAN permite apenas a operação no modo TDD.

No modo TDD, os *subframes* de DL e UL ocorrem em diferentes instantes de tempo, geralmente na mesma frequência. No modo FDD, os *subframes* ocorrem simultaneamente, mas em frequências distintas. As figuras 3.1 e 3.2 [3] ilustram os dois modos de operação. Na configuração FDD, são suportadas comunicações *full-duplex* e *half-duplex(H-FDD)*. Na primeira, as SSs transmitem e recebem simultaneamente, na segunda não.

Para perfeita operação nas configurações TDD e H-FDD, são necessários intervalos entre a transmissão/recepção, TTG (do inglês, *Transmit/receive Transition Gap*) e entre recepção/transmissão, RTG (do inglês, *Receive/transmit Transition Gap*). O TTG é inserido após o *subframe* de DL e antes do *subframe* de UL subsequente para permitir a

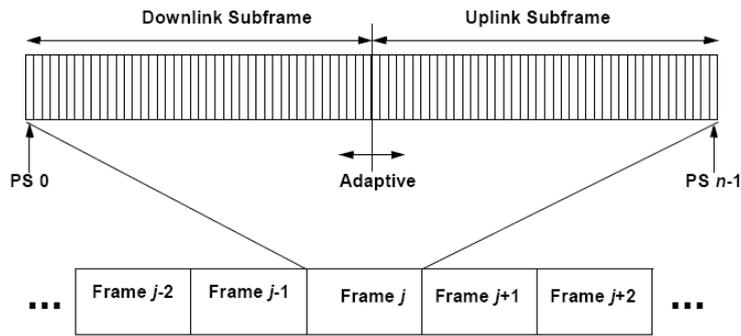


Figura 3.1: Operação em TDD.

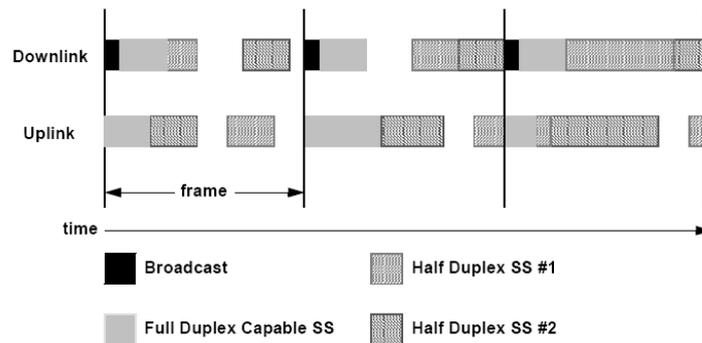


Figura 3.2: Operação em FDD.

alternância da BS do modo de transmissão para recepção e das SSs do modo de recepção para transmissão [3]. O RTG é inserido após o *subframe* de UL e antes do *subframe* de UL subsequente para permitir a alternância da BS do modo de recepção para transmissão e das SSs do modo de transmissão para recepção. Os dois intervalos são definidos como quantidades de PSs.

Em cada *subframe*, são definidos perfis de transmissão (*Downlink Burst Profile* e *Uplink Burst Profile*) que são associados aos códigos DIUC (do inglês, *Downlink Interval Usage Code*) e UIUC (do inglês, *Uplink Interval Usage Code*), respectivamente para o DL e UL. Estes códigos são enviados através de mensagens de camada MAC, DCD (*Downlink Channel Descriptor*) e UCD (*Uplink Channel Descriptor*) responsáveis por descrever cada canal de comunicação [3]. A indicação da alteração do DCD ou do UCD é realizada pelas mensagens MAC de mapeamento DL-MAP e UL-MAP que serão estudadas na seção 3.4. Cada especificação de camada física tem um formato de quadro que possui, geralmente, um cabeçalho e dados úteis. No cabeçalho estão presentes, geralmente, um preâmbulo de sincronismo e os perfis de transmissão (as mensagens DL-MAP e UL-MAP). Os dados úteis estão multiplexados de acordo com a tecnologia empregada, entre as SSs que compõem a rede.

As especificações de camada física estão divididas de acordo com a tecnologia de portadora utilizada: SC (do inglês, *Single Carrier*) ou OFDM (do inglês, *Orthogonal Frequency*

Division Multiplexing). Como o próprio nome diz, as especificações baseadas na tecnologia SC utilizam uma única portadora, enquanto os padrões baseados em OFDM utilizam várias portadoras ortogonais entre si. Além disso, outras tecnologias são utilizadas como diversos esquemas de múltiplo acesso, modulação e codificação de sinal.

3.3.1 WirelessMAN-SC

Esta especificação de camada física foi desenhada para operação em frequências entre 10 e 66 GHz, LOS, com alto grau de flexibilidade, permitindo às operadoras de telecomunicações otimização dos seus sistemas de acordo com o planejamento das células, o custo, a capacidade e os serviços a serem oferecidos ([3], seção 8.1.1).

A especificação permite a operação nas configurações TDD e FDD. Em ambas as configurações, os parâmetros de transmissão, como os esquemas de modulação e codificação, por exemplo, podem ser definidos para cada SS, a cada novo frame.

Para o *subframe* de UL, é utilizada uma combinação de duas técnicas de múltiplo acesso: TDMA e DAMA. Deve ser definida pela camada MAC a quantidade de *slots* destinada ao *subframe*, incluindo todo o tipo de tráfego (registro, intervalos, gerenciamento e dados). Esta quantidade de *slots* pode variar no tempo para melhorar a performance ([3], seção 8.1.2).

Para o *subframe* de DL, o método de multiplexação utilizado é o TDM, onde a comunicação para todos os SSs é realizada em um único fluxo de dados, recebido por todos os SSs de um determinado setor ou célula. Para suportar os SSs na configuração FDD full-duplex, uma porção do *subframe* de DL também é separada para multiplexação TDMA. A camada PHY possui uma subcamada de transmissão convergente (TCS, do inglês *Transmission Convergence Sublayer*), responsável por incluir um apontador (um *byte*) para orientar os SSs quanto ao local de início do correspondente MAC PDU (do inglês, *MAC Protocol Data Unit*). O fluxo de bits proveniente da TCS é mapeado em QPSK ou constelações QAM: obrigatoriamente, 16-QAM e opcionalmente, 64-QAM. A tabela 3.1 resume as principais características da especificação PHY WirelessMAN-SC.

3.3.2 WirelessMAN-SCa

Esta especificação de camada física foi desenhada para faixas de frequências inferiores a 11 GHz, NLOS, baseada na tecnologia de portadora única. A largura de banda dos canais pode variar em divisões base 2 da faixa de frequência total. No entanto, a largura de banda de cada canal não deve ser inferior a 1,25 MHz ([3], seção 8.2).

Esta especificação deve suportar configuração TDD e FDD. A multiplexação utilizada para o UL é TDMA. TDM e TDMA são utilizados na multiplexação do DL ([3], seção 8.2).

Tabela 3.1: Principais características PHY WirelessMAN-SC

Característica	Tecnologia Utilizada	Observações
Portadora	<i>Single Carrier</i>	-
Faixa de frequência (GHz)	10 - 66 GHz	Licenciada
Visada	LOS	-
Configuração <i>Duplex</i>	TDD e FDD	<i>Full e half-duplex</i> para FDD
Multiplexação <i>Uplink</i>	TDMA e DAMA	-
Multiplexação <i>Downlink</i>	TDM e TDMA	TDMA para SSs <i>half-duplex</i> FDD
Modulação de sinal	QPSK e 16-QAM	Opcionalmente 64-QAM

Alguns serviços opcionais são possíveis nesta especificação. São definidos parâmetros e mensagens de sinalização (entre as camadas MAC e PHY) para possibilitar a utilização de AAS (do inglês, *Adaptive Antenna System*), de ARQ (do inglês, *Automatic Repeat Request*) para controle de erros e STC (do inglês, *Space Time Coding*).

A tabela 3.2 resume as principais características da especificação PHY WirelessMAN-SCa.

3.3.3 WirelessMAN-OFDM

A especificação de PHY WirelessMAN-OFDM é baseada na tecnologia de múltiplas portadoras, OFDM, e desenhada para faixa de frequência inferior a 11 GHz com operação NLOS ([3], seção 8.3.1).

Esta camada deve suportar configuração TDD e FDD. As SSs FDD podem utilizar H-FDD. No entanto, para faixas de frequência não-licenciadas a configuração deve ser TDD. A tecnologia utilizada para o acesso de UL é TDMA.

Na multiplexação OFDM, o fluxo de dados de cada canal de comunicação é dividido em subportadoras de frequências ortogonais entre si e depois é transmitido. No receptor, ocorre o processo inverso. São utilizadas 256 portadoras divididas em ([3], seções 8.3.1.1.1 e 8.3.1.1.2):

- subportadoras de dados, responsáveis pelo transporte dos dados;
- subportadoras pilotos, responsáveis por funções de estimação das condições do canal;

Tabela 3.2: Principais características PHY WirelessMAN-SCa

Característica	Tecnologia Utilizada	Observações
Portadora	<i>Single Carrier</i>	-
Faixa de frequência (GHz)	<11 GHz	Licenciada
Visada	NLOS	-
Configuração <i>Duplex</i>	TDD e FDD	<i>Full e half-duplex</i> para FDD
Multiplexação <i>Uplink</i>	TDMA	-
Multiplexação <i>Downlink</i>	TDM e TDMA	TDMA para SSS <i>half-duplex</i> FDD
Modulação de sinal	BPSK, QPSK e 16,64-QAM	Opcionalmente 256-QAM
Serviços opcionais	AAS, ARQ, STC e Mobilidade	-

- subportadoras nulas, utilizadas como banda de guarda e para portadoras DC.

No domínio da frequência o canal é representado pela figura 3.3, conhecida como *brick wall*, em inglês. Podemos notar que não existe subcanalização; todo o canal (e todas as portadoras) é utilizado por um único transmissor. No domínio do tempo, a transformada inversa rápida de Fourier (IFFT) permite a criação de uma forma de onda, cuja duração T_b é mostrada na figura 3.4.

Alguns parâmetros são importantes para a definição de um símbolo OFDM, são eles :

- BW , largura de banda do canal em Hz;
- N_{used} , o número de subportadoras utilizadas;
- n , o fator de amostragem. Este parâmetro, em conjunto com os dois anteriores, define o espaçamento entre as subportadoras e o tempo útil de um símbolo. Na seção 8.3.2.4, de [3] e de [4], existem valores pré-definidos para n em função de BW ;
- G , razão entre o T_g ou CP e o tempo útil T_b , podendo assumir os valores 1/4, 1/8, 1/16 ou 1/32.

Dos parâmetros acima é possível encontrar outros valores característicos dos sistemas OFDM e OFDMA :

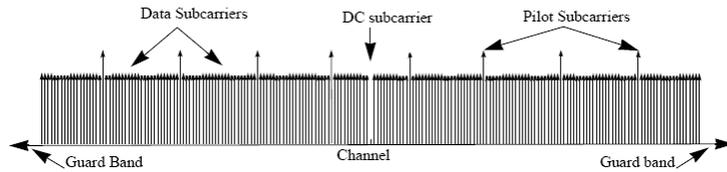


Figura 3.3: Representação em frequência do canal OFDM

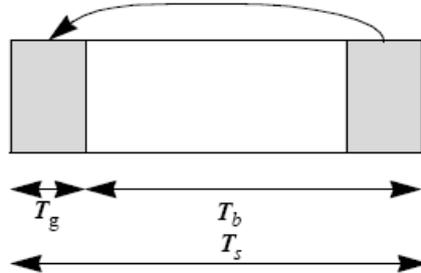


Figura 3.4: Representação do símbolo OFDM

- N_{FFT} , menor potência de dois maior que N_{used} ;
- F_s , frequência de amostragem, $F_s = (\lfloor \frac{n \times BW}{8000} \rfloor) \times 8000$;
- Δf , espaçamento entre subportadoras, $\Delta f = \frac{F_s}{N_{FFT}}$;
- T_b , tempo útil, $T_b = \frac{1}{\Delta f}$;
- T_g , tempo destinado ao *CP* (*cyclic prefix*), $T_g = G \times T_b$;
- T_s , tempo do símbolo OFDM, $T_s = T_g + T_b$; e
- a relação $\frac{T_b}{N_{FFT}}$ é o tempo de amostragem.

Os símbolos da camada física OFDM são organizados em uma estrutura chamada de *frame* (quadro). A estrutura do *frame* para operação PMP é apresentada na figura 3.5, retirada de [4]. Através da figura, pode-se perceber que o *frame* é dividido em *subframes* de DL e UL. Cada *subframe* possui uma quantidade inteira de *slots*. O *slot* corresponde a um intervalo de tempo do símbolo OFDM e é a unidade mínima para alocação de recursos na especificação PHY WirelessMAN-OFDM.

Para modulação de sinais pode ser utilizado QPSK, 16-QAM e 64-QAM. É possível a utilização de diferentes taxas de codificação para cada modulação de sinal, como 1/2, 2/3 e 3/4. A quantidade de bytes que pode ser inserida em cada *slot* OFDM de acordo com a modulação e a codificação utilizada é apresentada na figura 3.6, obtida de [3].

Para uso como camada primária da especificação WirelessHUMAN com modulação 64-QAM é opcional. A camada física em WirelessMAN-OFDM possui serviços adicionais como o suporte à mobilidade, AAS, ARQ, redes MESH e STC. A tabela 3.3 resume as principais características da especificação PHY WirelessMAN-OFDM.

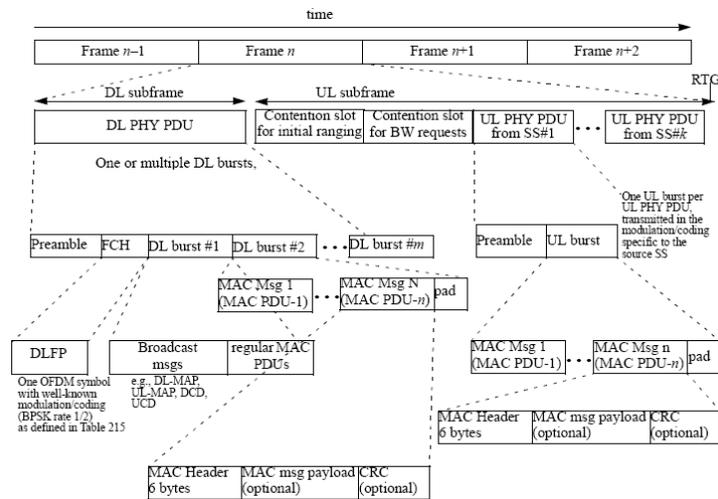


Figura 3.5: *Frame* para camada física WirelessMAN-OFDM operando em PMP.

Modulation	Uncoded block size (bytes)	Coded block size (bytes)	Overall coding rate	RS code	CC code rate
BPSK	12	24	1/2	(12,12,0)	1/2
QPSK	24	48	1/2	(32,24,4)	2/3
QPSK	36	48	3/4	(40,36,2)	5/6
16-QAM	48	96	1/2	(64,48,8)	2/3
16-QAM	72	96	3/4	(80,72,4)	5/6
64-QAM	96	144	2/3	(108,96,6)	3/4
64-QAM	108	144	3/4	(120,108,6)	5/6

Figura 3.6: Codificação de canal obrigatória por esquema de modulação.

Tabela 3.3: Principais características PHY WirelessMAN-OFDM

Característica	Tecnologia Utilizada	Obs.
Portadora	<i>Multiple carriers</i>	-
Faixa de frequência (GHz)	<11 GHz	Licenciada
Visada	NLOS	-
Configuração <i>Duplex</i>	TDD e FDD	<i>Full e half-duplex</i> para FDD
Multiplexação <i>Uplink</i>	OFDM	Não-licenciadas apenas TDD
Multiplexação <i>Downlink</i>	OFDM	acesso por TDMA
Modulação de sinal	BPSK, QPSK e 16,64-QAM	-
Serviços opcionais	AAS, ARQ, MESH, STC e Mobilidade	-

3.3.4 WirelessMAN-OFDMA

A especificação de PHY WirelessMAN-OFDMA é baseada na tecnologia de múltiplas portadoras, OFDM, e desenhada para faixa de frequência inferior a 11 GHz com operação NLOS. A largura de banda dos canais pode variar em divisões base 2 da faixa de frequência total. No entanto, a largura de banda de cada canal não deve ser inferior a 1 MHz (seção 8.4.1 de [3]).

Esta especificação agrega o conceito de subcanal ao OFDM. Os subcanais lógicos agrupam subportadoras reais após um dos processos de permutação de subportadoras: DSP ou ASP. Quando aplicado o método DSP, são permitidas duas formas de utilização dos subcanais lógicos: FUSC (*full usage of subchannels*) ou PUSC (*partial usage of subchannels*). No caso do FUSC, todos os subcanais lógicos são utilizados para um transmissor, enquanto no PUSC, conjuntos de subcanais lógicos são reservados aos transmissores.

Os padrões IEEE 802.16 ([3] e [4]) possuem algumas definições para a utilização dos subcanais e símbolos OFDMA, são eles:

- domínio do tempo - símbolo OFDMA;
- domínio da frequência - parâmetros do domínio da frequência, como N_{FFT} , n , G , entre outros;
- definição do *slot*; e

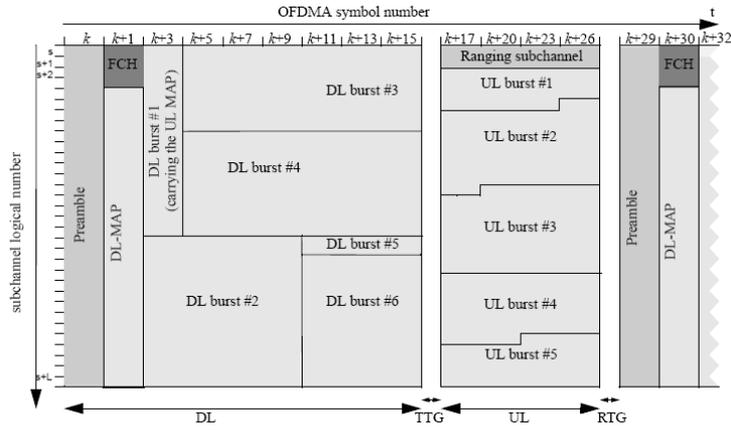


Figura 3.7: *Frame* para camada física WirelessMAN-OFDMA operando em PMP.

- a estrutura do *frame*.

No domínio do tempo, a estrutura do símbolo OFDM, mostrada na figura 3.4, também é válida para a especificação da camada física em OFDMA. Os parâmetros no domínio da frequência, apresentados na subseção 3.3.3, também são os mesmos, ressaltando apenas que o fator de amostragem pode possuir os valores definidos na seção 8.4.2.3 da norma [4].

A definição do *slot* OFDMA depende do sentido de transmissão e do tipo de utilização dos subcanais (seção de 8.4.3.1 [4]):

- para *downlink* FUSC ou *downlink* FUSC opcional usando permutação de subportadora distribuída, um *slot* é um subcanal por um símbolo OFDMA;
- para *downlink* PUSC usando permutação de subportadora distribuída, um *slot* é um subcanal por dois símbolos OFDMA;
- para PUSC *uplink* usando permutação de subportadora distribuída, um *slot* é um subcanal por três símbolos OFDMA; e
- para permutação de subportadora adjacente, um *slot* é um subcanal por dois, três ou seis símbolos OFDMA.

Uma característica importante para a estratégia de escalonamento que será proposta neste trabalho é a forma como os subcanais serão utilizados: total (FUSC, *full usage of subchannels*) ou parcialmente (PUSC, *partial usage of subchannels*).

A estrutura do *frame* OFDMA, devido à subcanalização, é bidimensional, diferentemente do *frame* OFDM que utiliza apenas a divisão do tempo para o múltiplo acesso (TDMA). Esta estrutura é apresentada na figura 3.7.

O *frame* pode possuir zonas de utilização distintas, como pode ser visto na figura 3.8, mas é obrigatória a existência de zonas PUSC no *subframe* de DL. No *subframe* de UL, utilizando

	QPSK		16 QAM		64 QAM		
Encoding rate	R=1/2	R=3/4	R=1/2	R=3/4	R=1/2	R=2/3	R=3/4
Data payload (bytes)	6						
		9					
	12		12				
	18	18		18	18		
	24		24			24	
		27					27
	30						
	36	36	36	36	36		

Figura 3.10: *Payload* por esquema de modulação e taxa de codificação.

Esta especificação possui serviços adicionais como o suporte a mobilidade, AAS, ARQ, HARQ e STC. A tabela 3.4 resume as principais características da especificação PHY WirelessMAN-OFDMA.

3.3.5 WirelessHUMAN

A especificação PHY WirelessHUMAN permite a utilização de duzentos canais, com as frequências centrais separadas por 5 MHz, na faixa espectral entre 5 e 6 GHz. Esta configuração pode optar pela utilização das especificações WirelessMAN-SCa, WirelessMAN-OFDM e WirelessMAN-OFDMA (seção 8.5.1 [3]).

No entanto, existem algumas limitações. A canalização para esta especificação deve ser de 10 ou 20 MHz, não permite FDD e a modulação de sinal obrigatória menos robusta é a 16-QAM. A tabela 3.5 resume as principais características da especificação PHY WirelessHUMAN.

3.4 Camada de Enlace (MAC)

A camada MAC tem a função de controlar a camada física através dos requisitos recebidos das camadas superiores. A camada MAC está dividida em três subcamadas: a Subcamada de Convergência (CS, do inglês, *Service-Specific Convergence Sublayer*), a Subcamada de Parte Comum (CPS, do inglês, *Common Part Sublayer*) e a Subcamada de Segurança (*Security Sublayer*).

A CS recebe todos os dados externos através do CS SAP (do inglês, *service access point*) e transforma-os ou mapeia-os em MAC SDUs (do inglês, *Service Data Unit*) que são repassados a CPS. A CS classifica os SDUs de acordo com o CID (do inglês, *Connection*

Tabela 3.4: Principais características PHY WirelessMAN-OFDMA

Característica	Tecnologia Utilizada	Obs.
Portadora	<i>Multiple carriers</i>	-
Faixa de frequência (GHz)	<11 GHz	Licenciada
Visada	NLOS	-
Configuração <i>Duplex</i>	TDD e FDD	<i>Full e half-duplex</i> para FDD
Multiplexação <i>Uplink</i>	OFDMA	-
Multiplexação <i>Downlink</i>	OFDMA	-
Modulação de sinal	BPSK, QPSK e 16, 64-QAM	-
Serviços opcionais	AAS, ARQ, HARQ, STC e Mobilidade	-

Tabela 3.5: Principais características PHY WirelessHUMAN

Característica	Tecnologia Utilizada	Obs.
Portadora	<i>Single/Multiple carrier(s)</i>	-
Faixa de frequência (GHz)	< 11 GHz	Inicialmente entre 5-6 GHz, não-licenciada
Visada	NLOS	-
Configuração <i>Duplex</i>	TDD	-
Multiplexação <i>Uplink</i>	TDMA e OFDMA	-
Multiplexação <i>Downlink</i>	TDM e OFDMA	-
Modulação de sinal	BPSK, QPSK e 16,64-QAM	64-QAM é opcional
Serviços opcionais	AAS, ARQ, STC e MESH	-

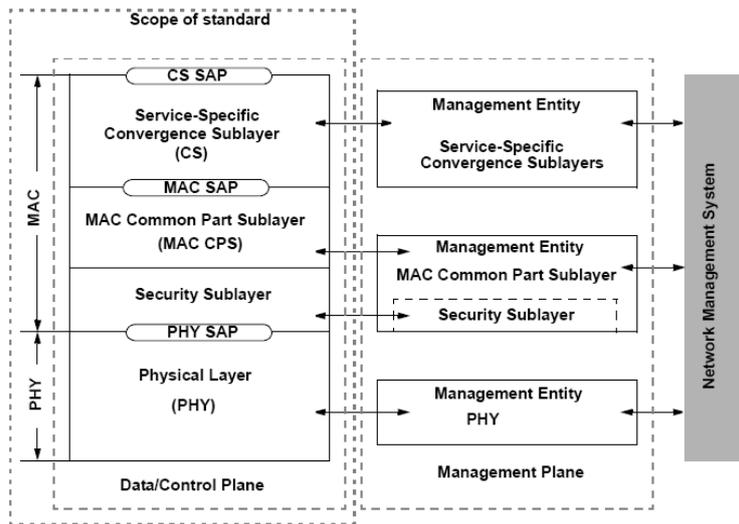


Figura 3.11: Modelo de referência da estrutura em camadas

Identifier) referenciado e os associa ao correto SFID (do inglês, *Service Flow Identifier*), identificador do SF (do inglês, *Service Flow*). Existem alguns tipos específicos de subcamada CS para mapear algumas tecnologias como, por exemplo, células ATM e redes baseada em pacotes.

A CPS recebe os MAC SDUs através do MAC SAP. A CPS provê os serviços mais importantes da camada MAC como: alocação de largura de banda, estabelecimento de conexão e manutenção das conexões em funcionamento. Nesta camada, os parâmetros de qualidade de serviço são aplicados à transmissão e ao escalonamento das conexões, repassando os dados para camada física, através do PHY SAP. A subcamada de segurança é responsável pelos serviços de autenticação, intercâmbio de chaves e encriptação. A figura 3.11 (seção 1.4 de [3]) ilustra a estrutura em camadas da norma IEEE 802.16.

A seguir, faremos um breve resumo das funções de cada subcamada.

3.4.1 Subcamada de Convergência

A CS situa-se no topo da camada MAC fazendo interface com as camadas superiores e utiliza os serviços da CPS, via MAC SAP, conforme figura 3.11. A CS executa as seguintes funções (seção 5 de [3]):

- aceita os PDUs das camadas superiores;
- classifica os PDUs das camadas superiores;
- processa (se necessário) os PDUs das camadas superiores, baseado na classificação;
- entrega os CS PDUs para o MAC SAP apropriado;

- recebe CS PDUs do seus pares de comunicação.

A CS possui duas subdivisões: CS ATM (células ATM, sigla em inglês para *asynchronous transfer mode*) e CS Pacotes (pacotes IP versões 4 e 6). A CS ATM é uma interface lógica que associa serviços ATM com a MAC CPS SAP. A CS ATM aceita células ATM da camada de rede ATM, realiza classificação, se especificado, realiza também a supressão de cabeçalho (*Payload Header Suppression* - PHS) e entrega os CS PDUs ao MAC SAP apropriado.

Uma conexão ATM é identificada inequivocamente por um VPI (do inglês, *virtual path identifier*) e um VCI (do inglês, *virtual channel identifier*). O classificador da CS ATM possui um conjunto de critérios de casamento como, por exemplo, o par VPI, VCI e um CID de referência. Quando uma célula atende aos critérios, MAC SDUs são repassados, através do MAC SAP, para a CPS, e em seguida é entregue a conexão identificada com aquele CID (seção 5.1 de [3]).

A CS Pacote realiza a classificação de todos os protocolos baseados em pacotes como IP (do inglês, *Internet Protocol*), por exemplo. Os MAC SDUs recebidos da camada de rede são mapeados em uma conexão de transporte entre pares MACs. Os MAC SDUs são entregues a CPS, através do CPS SAP. O mapeamento é capaz de levar em consideração o protocolo IEEE 802.1Q, havendo parâmetros específicos da CS Pacotes para este fim.

3.4.2 Subcamada de Parte Comum

Deixamos esta subcamada por último, por ser a mais importante para a camada MAC e para o foco deste trabalho; o escalonamento dos tráfegos de UL e DL é realizado nesta subcamada. O texto desta seção é baseado em [3] e [4]. Existem duas divisões dentro da CPS: as especificações para o modo de operação PMP (do inglês, *point-to-multipoint*) e para o modo de operação MESH. No modo de operação PMP, toda comunicação ocorre entre a BS e as SSs, enquanto no modo MESH é possível o encaminhamento do tráfego de uma SS para outra ou ainda a comunicação direta entre as SSs. No modo PMP, todo controle do tráfego é realizado pelas BSs. Neste trabalho, não abordaremos redes com o modo de operação MESH e, por isso, descreveremos apenas as especificações para o modo de operação PMP.

No modo PMP, o DL é realizado, geralmente, por *broadcast*. Todas as SSs que não encontrarem uma indicação de que parte do *subframe* de DL lhe pertence, deverá receber o *subframe* completo. Então, são verificados os CIDs existentes e cada SS retém apenas os PDUs que lhe pertencem (seção 6.1 de [3]).

O *subframe* de *uplink* para a BS é dividido entre as SSs baseado na classe de serviço e na demanda de cada uma. Certos tipos de tráfego possuem garantia contínua para transmissão e outros terão o direito de transmitir baseado na própria requisição de largura de banda

(seção 6.1 de [3]). Para certas classes de serviços, as SSs são questionadas individualmente sobre o seu interesse em usar o canal de UL. Algumas aplicações requerem o questionamento via *multicast*, como, por exemplo, as aplicações de distribuição de vídeo. Para outras, resta responder ao questionamento geral sobre requisição de largura de banda, via canal de *broadcast*.

Para garantir a qualidade de serviço desejada e controlar a disputa pelos canais de UL, estão previstos cinco mecanismos de *scheduling*, que podem ser implementados utilizando *unsolicited bandwidth grants*, *polling* e *contention procedures*. Os mecanismos são definidos de forma que os fabricantes tenham liberdade para otimizar a utilização dos recursos. Períodos de contenção podem ser utilizados para evitar o *polling* das SSs que estão há um longo período sem transmitir dados (seção 6.1 de [3]).

Mecanismos *unsolicited bandwidth grants* evitam que as SSs, que estão executando aplicações de alta sensibilidade ao atraso, deixem de ser servidas por um tempo maior que o requerido. O *polling* simplifica o acesso ao recurso e garante que as estações recebam serviços de forma determinística, quando requisitarem. Em geral as aplicações são tolerantes ao atraso e possuem baixa largura de banda, mas aplicações de vídeo em tempo real, por exemplo, podem necessitar de alta largura de banda durante um longo período de tempo e com uma tolerância ao atraso bastante restrita (seção 6.1 de [3]).

A camada MAC do WiMAX é orientada à conexão. A responsabilidade pelo controle das conexões é da CPS. Existem quatro tipos de conexões: conexão básica, conexão primária de gerenciamento, conexão secundária de gerenciamento e conexão de transporte. As três primeiras são destinadas à inicialização, à manutenção e à finalização das chamadas e a última à comunicação propriamente dita. Cada conexão possui um CID.

A cada conexão de transporte, assim que estabelecida, será associado um SF, que possui um SFID. A relação entre CID e SFID é de um para um. O objetivo do SF é mapear os parâmetros de QoS de cada conexão de transporte. O conceito de SF sobre uma conexão de transporte é central para camada MAC. Uma conexão de transporte depois de estabelecida deve ser mantida de acordo com os parâmetros de QoS estabelecidos no SF. Alguns tipos de conexão praticamente não precisam de manutenção, como tráfego de voz não canalizado, que possui largura de banda contínua e após estabelecido será apenas terminado (seção 6.1 de [3]). No tráfego da voz canalizada, apesar de pequenas, pode haver mudanças na largura de banda de transmissão. O tráfego de serviços IP, por outro lado, em razão da característica de rajada, possuem elevada quantidade de manutenção.

O encerramento das conexões de transporte podem acontecer por solicitação da rede ou das estações, quer por alteração de parâmetros de QoS ou por inexistência de tráfego a ser transmitido (seção 6.1 de [3]). Todos estes serviços requerem configurações, modificações e exclusões de SFs. No restante desta seção, faremos uma descrição dos principais componentes da CPS para a continuidade deste trabalho, principalmente os serviços de *scheduling*

e de alocação e requisição de largura de banda. Também fazem parte da camada MAC as definições e parâmetros de QoS.

3.4.2.1 Endereçamento e conexões

Cada interface aérea das SSs deve possuir um endereço MAC universal de 48 bits, de acordo com o padrão IEEE 802-2001. Este endereço identifica inequivocamente a interface aérea da SS em qualquer rede. Este endereço é utilizado durante o período de inicialização até o estabelecimento da conexão entre a SS e a BS.

As conexões são identificadas por um CID de 16 bits. No processo de inicialização, dois pares (UL e DL) de conexões de gerenciamento devem ser estabelecidos e um terceiro é opcional. Isto acontece em razão da existência de três níveis diferentes de QoS para as conexões de gerenciamento (seção 6.3.1.1 de [4]):

- as conexões básicas são utilizadas para o envio de mensagens urgentes entre as camadas MAC da BS e da SS;
- as conexões de gerenciamento primárias são utilizadas também para troca de mensagens entre as camadas MAC, mas que possuem maior tolerância ao atraso. Adicionalmente, este tipo de conexão especifica que tipo de mensagem de gerenciamento é enviada via conexão de Broadcast (*Broadcast connection*);
- as conexões de gerenciamento secundárias são utilizada para mensagens de sinalização tolerantes a atrasos, referentes a serviços padronizados, como por exemplo, DHCP e SNMP.

Mensagens carregadas nas conexões de gerenciamento secundárias podem ser empacotadas ou fragmentadas. As conexões secundárias são requeridas apenas para as SSs gerenciadas. Os CID para todas estas conexões são fornecidos pelas mensagens RNG-RSP e REG-RSP. São fornecidos três CIDs. Conexões de *uplink* e *downlink* de cada par utilizam o mesmo número de CID (seção 6.3.1.1 de [4]).

O registro de uma SS na BS ou a alteração das condições contratadas inicia a configuração de um *service flow*. Cada SF é associado unicamente com uma conexão de transporte e, conseqüentemente, com um CID. Mensagens de gerenciamento não podem ser enviadas por conexões de transporte e os dados não podem ser enviados por uma conexão básica ou de gerenciamento.

As requisições para transmissão são baseadas nos CIDs, permitindo que a largura de banda para transmissão seja diferente, mesmo para usuários utilizando um mesmo tipo de serviço, como o RTPS, por exemplo. Uma mesma SS servindo diferentes funcionários de um escritório deve fazer requisições para todos eles, dentro dos limites contratados, podendo os parâmetros de conexão diferir entre si.

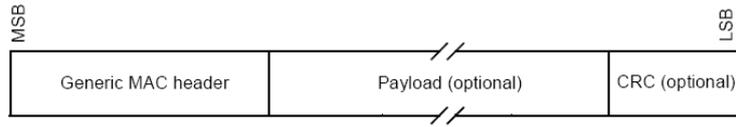


Figura 3.12: Formato do MAC PDU

Sessões de aplicações das camadas superiores podem funcionar em mesmo CID. Por exemplo, usuários de uma empresa utilizando um mesmo acesso WiMAX, comunicando-se através de TCP/IP para diferentes destinos, mas utilizando os mesmos parâmetros de qualidade, podem utilizar *polling* para requisição de oportunidades de transmissão. Como os endereços da fonte e destino originais estão encapsulados no *payload* da transmissão, não há problema para identificar os usuários das diferentes sessões (seção 6.3.1.1 de [3]).

3.4.2.2 Quadros, Cabeçalhos e Mensagens

Os MAC PDUs devem possuir o formato da figura 3.12 (seção 6.3.2 de [3]). Cada PDU deve possuir um cabeçalho genérico, podendo ser seguido por um *payload* e podendo ser finalizado por um campo CRC do inglês, *cyclic redundancy check*. Quando o *payload* está presente, poderá conter vários tipos de *subheaders* e MAC SDUs ou fragmentos destes. O tamanho do campo *payload* é variável. Isto permite que diversos tipos e formatos de dados, de diferentes tecnologias, sejam transportados.

Estão definidos três tipos de cabeçalho MAC, após as alterações publicadas em [4] (ver seção 6.3.2.1); dois para o UL e um para o DL. A figura 3.13 ilustra a estrutura do cabeçalho genérico (GMH sigla em inglês para *Generic MAC Header*). Para o DL, foi definido o GMH que inicia todo MAC PDU contendo mensagens de gerenciamento ou dados da CS. Para o UL, foram definidos dois tipos de cabeçalhos. O primeiro é o GMH que inicia todo MAC PDU contendo mensagens de gerenciamento ou dados da CS, no qual o campo HT (do inglês, *Header Type*) tem valor 0 (zero). O segundo é um formato de cabeçalho com o campo HT preenchido com o valor 1 (um) e sem os campos *payload* e CRC. Cabeçalhos com HT=1 não são encriptados. A tabela 3.6 descreve os campos do GMH. O campo ESF pode indicar a presença de ao menos um *subheader*. Os tipos de *subheader* são indicados na tabela 3.7.

Cabeçalhos que não possuem *payload* são divididos em dois tipos: cabeçalhos de sinalização MAC tipo I e II. As figuras 3.14 e 3.15 (seções 6.3.2.1.2.1 e 6.3.2.1.2.2 de [4]) ilustram os dois tipos. Entre os cabeçalhos de sinalização do tipo I existem cabeçalhos para informação da SNR, informação sobre a potência de transmissão de UL, informação sobre a relação portadora/interferência, requisição de alocação de CQICH, preferências da camada física, e principalmente BR (do inglês, *Bandwidth Request*). Os cabeçalhos de sinalização MAC do tipo I possuem o campo HT=1 e o campo EC=0 e os três bits do campo Tipo (*Type*, ver figura 3.14) preenchidos conforme tabela 3.8. Entre os cabeçalhos de sinalização

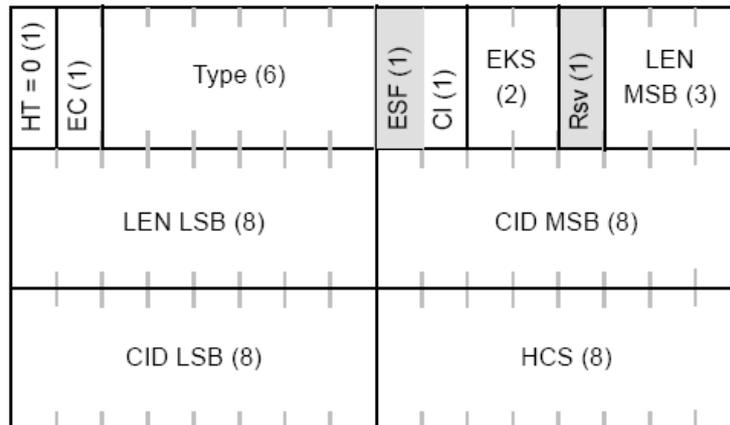


Figura 3.13: Cabeçalho MAC Genérico

Tabela 3.6: Descrição dos campos do Cabeçalho MAC Genérico

Campo	Tamanho(bits)	Descrição
CI	1	<i>CRC Indicator</i> . Quando CI=1, o campo CRC está presente.
CID	16	Identificador da conexão.
EC	1	<i>Encryption Control</i> . Quando EC=1 o <i>payload</i> está criptografado.
EKS	2	<i>Encryption Key Sequence</i> . Utilizado para PDUs com EC=1.
ESF	1	<i>Extended Subheader Field</i> . Quando ESF=1 existe pelo menos um <i>subheader</i> .
HCS	8	<i>Header Check Sequence</i> . Campo utilizado para detectar erros de cabeçalho.
HT	1	Deve ser preenchido com zero para o GMH.
LEN	11	Tamanho do MAC PDU incluindo o <i>payload</i> e o CRC, se presente.
Type	6	Indica os tipos de subheaders (tabela 3.7).

Tabela 3.7: Códigos dos subheaders do GMH

Tipo	Valor
#5 (bit mais significativo)	<i>Mesh subheader.</i> 1=presente, 0=ausente
#4	<i>ARQ Feedback Payload.</i> 1=presente, 0=ausente
#3	<i>Extended Type.</i> Indica a presença de <i>extended subheaders</i> de <i>Fragmentation</i> ou <i>Packing</i> . Com ARQ habilitado este bit deve ser 1.
#2	<i>Fragmentation subheader.</i> 1=present, 0=ausente.
#1	<i>Packing subheader.</i> 1=presente, 0=ausente.
#0 (bit menos significativo)	DL: <i>FAST-FEEDBACK Allocation subheader.</i> UL: <i>Grant Management subheader.</i> 1=presente, 0=ausente.

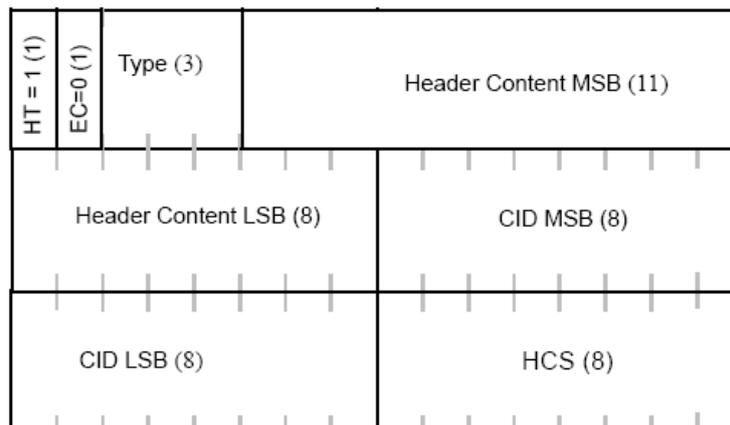


Figura 3.14: Cabeçalho de sinalização sem *payload* - Tipo I

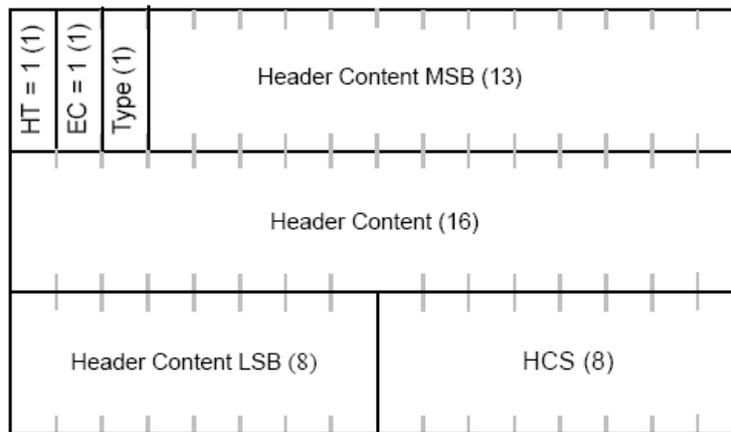


Figura 3.15: Cabeçalho de sinalização sem *payload* - Tipo II

Tabela 3.8: Códigos do campo *Type* do cabeçalho de sinalização MAC - Tipo I

Tipo	Valor
000	BR incremental
001	BR aggregate
010	PHY channel report
011	BR with UL Tx Power Report
100	Bandwidth request and CINR report
101	BR with UL sleep control
110	SN Report
111	CQICH allocation request

do tipo II existem cabeçalhos para *feedback*.

O BR PDU consiste apenas no cabeçalho, sem *payload* ou CRC, especificamente para UL. Cabeçalhos de requisição de banda recebidos pelas SSs no canal de DL devem ser descartados. O cabeçalho de requisição de largura de banda pertence ao tipo I e possui função fundamental para estratégia de scheduling da BS: transmite as solicitações de largura de banda das SSs para a BS. A figura 3.16 ilustra o formato deste cabeçalho (seção 6.3.2.1.2 de [3]). O campo CID indica a conexão para a qual está sendo solicitada a largura de banda. O campo BR indica a quantidade de bytes que está sendo solicitada. O campo *Type* é preenchido conforme tabela 3.8 com os valores (seção 6.3.2.1.1 de [4]):

- 000, para requisição de largura de banda incremental;
- 001, para requisição de largura de banda agregada.

A capacidade para BR incremental é opcional para a SS e obrigatória para a BS. A capacidade de BR agregada é obrigatória para ambas. As requisições de banda também

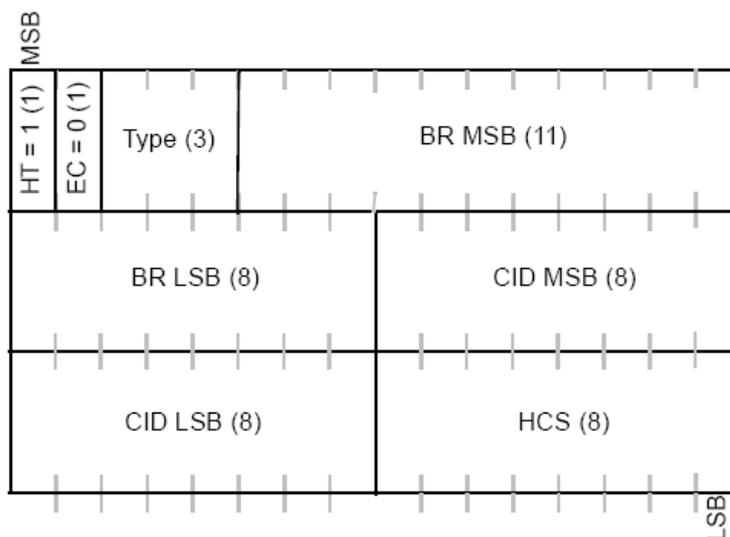


Figura 3.16: Cabeçalho sem *payload* - Requisição de Banda

podem ser feitas com os outros cabeçalhos, com os tipos 011, 100 e 101, de acordo com a tabela 3.8, em conjunto com informações e controles de camada física. Além da requisição de banda através deste cabeçalho, as SSs podem realizar a BR utilizando o mecanismo de PBR (*Piggyback Request*), utilizando o subcabeçalho GMS (do inglês, *Grant Management Subheader*)(seção 6.3.2.2 de [4]).

Existem alguns tipos de subcabeçalhos inseridos após o GMH, formando um cabeçalho estendido (o bit ESF do campo GMH é preenchido com 1), como, por exemplo, os subcabeçalhos de Fragmentação (em inglês, *Fragmentation subheader*) e de Empacotamento (em inglês, *Packing subheader*) (seção 6.3.2.2.1 e 6.3.2.2.3 de [4]). O subcabeçalho de Fragmentação define a fragmentação ou não do *payload* do PDU. Sua presença é indicada no campo *Type* do GMH, conforme tabela 3.7. O subcabeçalho de Empacotamento é utilizado para indicar a transmissão de mais de um SDU em um mesmo *payload* de um MAC PDU. Os subcabeçalhos de Empacotamento e Fragmentação são mutuamente exclusivos e não podem estar presentes em um mesmo MAC PDU.

O GMS é um subcabeçalho que possui papel importante para a estratégia de escalonamento, quando utilizado. O uso do GMS é opcional tanto para BS quanto para as SSs. A tabela 3.9 descreve o formato e a tabela 3.10 explica cada campo do GMS, após as alterações realizadas por [4] (ver seção 6.3.2.2.2).

Como é possível observar na tabela 3.9, este subcabeçalho, como os demais, não contém o CID. Sendo assim, suas requisições e informações estão associadas ao CID existente no cabeçalho principal [4]. A utilização deste subcabeçalho e do cabeçalho de BR será detalhada nas próximas seções.

O campo *payload* do MAC PDU, além dos dados da transmissão, pode conter também mensagens de gerenciamento. Algumas mensagens de gerenciamento que são fundamentais

Tabela 3.9: Formato do GMS

Sintaxe	Tamanho(bits)	Obs.
se (serviço == UGS)	-	-
{...	-	-
SI	1	-
PM	1	-
FLI	1	-
FL	4	-
Reservado	9	Preenchidos com zeros.
...}	-	-
senão se (serviço == eRTPS)	-	-
{...	-	-
<i>Extended piggyback request</i>	11	-
FLI	1	-
FL	4	-
...}	-	-
senão	-	-
{...	-	-
<i>Piggyback Request(PBR)</i>	16	-
...}	-	-

Tabela 3.10: Descrição dos campos do GMS

Campo	Tamanho (bits)	Descrição (regra de formação na tabela 3.9)
SI	1	<i>Slip Indicator</i> . É preenchido com 1, pela SS, para indicar que a fila de transmissão está ultrapassando o tamanho máximo permitido. É zerado quando a fila está dentro do limite.
PM	1	<i>Poll-me</i> . Se PM=1, a SS está solicitando para ser incluída no <i>polling</i> . Se PM=0, nenhuma ação é realizada.
Extended PBR	11	Número de bytes de UL requisitados pela MS.
FLI	1	<i>Frame latency indication</i> . Indica o preenchimento do campo FL, quando FLI=1.
FL	4	<i>Frame latency</i> . Indica há quantos <i>frames</i> , antes do atual, os dados estavam disponíveis para transmissão. Se > 15 é preenchido com 15.
Reservado	9	Reserva preenchida com zeros.
PBR	16	Número de bytes de UL requisitados pela SS.

para o funcionamento da camada MAC e para a estratégia de *scheduling*. Após as alterações de [4], o padrão passou a definir 67 tipos de mensagens diferentes, incluindo as reservadas. Não é do escopo deste trabalho o detalhamento de todas as mensagens, mas algumas são fundamentais para a estratégia de escalonamento, como por exemplo, as mensagens DL e UL-MAP.

Mensagens do tipo DL-MAP (do inglês, *Downlink map message*), enviadas em conexões de *broadcast*, definem o acesso ao canal de DL. Cada BS gera mensagens DL-MAP conforme a tabela 3.11, incluindo os seguintes parâmetros (seção 6.3.2.3.2 de [4]):

- *PHY Synchronization* - campo de sincronização dependente da especificação de camada física;
- *DCD Count* - contador para acusar alteração nas configurações do perfil de transmissão;
- *BS ID* - identificador, de 48 bits, para cada BS. Os primeiros 24 bits são utilizados como identificação da operadora. É utilizado em conjunto com o DCD para definição do limite das células e setores.

As mensagens DL-MAP devem possuir um número inteiro de *bytes*. Caso, após construída a mensagem, o número de bytes não seja inteiro, o tamanho é arredondado para cima, incluindo o campo *Padding Nibble* ao final da mesma. Os bits extras são descartados nas SSs.

Mensagens do tipo UL-MAP (do inglês, *Uplink map message*), enviadas em conexões de *broadcast*, definem o acesso ao canal de UL. Cada BS gera mensagens UL-MAP conforme a tabela 3.12, incluindo os seguintes parâmetros (seção 6.3.2.3.3 de [4]):

- *UCD Count* - contador para acusar alteração nas configurações do perfil de transmissão;
- *Allocation Start Time* - apontador para o instante (símbolo OFDMA) exato de início da mensagem UL-MAP, dentro do *subframe* de UL. Depende da especificação de camada física;
- *MAP IEs* - definem a alocação de largura de banda. O conteúdo dos IEs dependem da especificação de camada física.

A utilização da mensagem UL-MAP é fundamental para a alocação de recurso. A participação das mensagens UL-MAP será explicada nas próximas seções. Os contadores de DCD e UCD referem as mensagens de descrição dos canais de comunicação de DL e UL, respectivamente. Estas mensagens contêm especificações de camada física, vistas na seção anterior, notadamente os *burst profile* de DL e UL, informando o tipo de modulação e codificação de sinal, a cada período de tempo.

Tabela 3.11: Formato das mensagens DL-MAP

Sintaxe	Tamanho (bits)	Obs.
DL-MAP_Message_Format()	-	-
{...}	-	-
Management Message Type = 2	8	-
PHY Synchronization Field (campo de sincronização da camada PHY)	variável	Depende da camada física.
DCD Count	8	-
BS ID	48	-
Início da secção de camada PHY	-	Depende da camada física.
{...}	-	-
Se (<i>PHY = WirelessMAN – OFDMA</i>)	-	-
{...}	-	-
Número de símbolos OFDMA	8	Número de símbolos OFDMA no <i>subframe</i> de <i>downlink</i> .
...}	-	-
for (i = 1; i <= n; i++)	-	-
{...}	-	-
<i>DL – MAP_IE</i> () (para cada elemento de 1 até n)	variável	Depende da camada física.
...}	-	-
...}	-	-
Se diferente de (<i>byte boundary</i>)	-	-
{...}	-	-
Padding Nibble	4	Bits para arredondamento.
...}	-	-
...}	-	-

Tabela 3.12: Formato das mensagens UL-MAP

Sintaxe	Tamanho (bits)	Obs.
UL-MAP_Message_Format()	-	-
{...}	-	-
Management Message Type = 3	8	-
UCD Count	8	-
Allocation Start Time	32	-
Início da secção de camada PHY	-	Depende da camada física.
{...}	-	-
Se (<i>PHY = WirelessMAN – OFDMA</i>)	-	-
{...}	-	-
Número de símbolos OFDMA	8	Número de símbolos OFDMA no <i>subframe</i> de <i>uplink</i> .
...}	-	-
for (i = 1; i <= n; i++)	-	-
{...}	-	-
<i>UL – MAP_IE</i> () (para cada elemento de 1 até n)	variável	Depende da camada física.
...}	-	-
...}	-	-
Se diferente de (<i>byte boundary</i>)	-	-
{...}	-	-
Padding Nibble	4	Bits para arredondamento.
...}	-	-
...}	-	-

3.4.2.3 Serviços de *Scheduling*

Após a publicação do padrão [4] (ver seção 6.3.5), estão disponíveis cinco tipos de serviços de *scheduling*: UGS(do inglês, *Unsolicited Grant Service*), eRTPS (do inglês, *extended Real-Time Polling Service*), RTPS(do inglês, *Real-Time Polling Service*), nRTPS(do inglês, *non-Real-Time Polling Service*) e BE(do inglês, *Best Effort*). Cada conexão de transporte possui um SF associado que, por sua vez, está associado a um único tipo de serviço de *scheduling*. Cada tipo de *scheduling* possui um conjunto de parâmetros de qualidade de serviço que será determinante para o funcionamento do mesmo. Estes parâmetros serão definidos na seção 3.4.2.4, como por exemplo, o MSTR (do inglês, *Maximum Sustained Traffic Rate*), MRTR (do inglês, *Minimum Reserved Traffic Rate*), ML (do inglês, *Maximum Latency*), TJ (do inglês, *Tolerated Jitter*), TP (do inglês, *Traffic Priority*) e as regras de requisição/transmissão (*Request/Transmission Policy*).

O tipo de escalonamento UGS foi projetado para suportar SFs de UL em tempo real que transportem pacotes de tamanho fixo e períodos constantes como, por exemplo, canais E1/T1 e tráfego VoIP sem supressão de silêncio. O serviço oferece garantias periódicas de tamanho fixo para o UL das SSs eliminando o *overhead* e a latência da requisição de banda.

A garantia de *slots* é baseada no parâmetro MSTR. Caso o parâmetro MRTR esteja presente, deve ser preenchido com o mesmo valor do parâmetro MSTR. A garantia periódica de *slots* deve contemplar o tamanho fixo para transporte dos dados, do GMH e do GMS (quando presente). As regras de requisição e transmissão devem proibir que as SSs utilizando UGS participem do período de contenção (disputa por PSs para UL) (seção 6.3.5.2.1 de [3]).

A BS não deve alocar banda superior ao parâmetro MSTR, exceto quando o campo SI do GMS estiver ativo. Isto indica que a fila de transmissão da SS ultrapassou o limite desejado. Neste caso, é possível o acréscimo de 1% de garantia adicional. Os campos FLI e FL do GMS são usados para a MS informar a BS que a latência está superior a um determinado limite, por exemplo, a duração de um frame. Neste caso, a BS deve fornecer garantias em um período mais curto para aquele SF. Os parâmetros de qualidade de serviço obrigatórios para o UGS são: MSTR, ML, TJ e as regras de requisição e transmissão (seção 6.3.5.2.1 de [4]).

Já o tipo RTPS foi projetado para suportar SFs de UL em tempo real que transportem pacotes de tamanho variável e períodos constantes como, por exemplo, vídeos MPEG. Este serviço oferece oportunidades periódicas de requisição de banda, individualmente (*unicast request opportunities*), e permite às SSs informar a quantidade de largura de banda desejada. Este tipo necessita de maior *overhead* que o UGS, mas permite otimização no uso dos recursos para transmissão dos dados, adequando-se ao volume existente. As SSs são proibidas de participar de períodos de disputa por *slots*, pois recebem oportunidades periódicas individuais para o UL. A BS deve realizar o questionamento individual às SSs

mesmo que em oportunidade anterior a requisição tenha sido nula. As SSs utilizando RTPS também poderão realizar requisição de banda utilizando as oportunidades de transmissão dos dados. Os parâmetros de qualidade de serviço obrigatórios para o RTPS são: MRTR, MSTR, ML e as regras de requisição e transmissão (seção 6.3.5.2.2 de [4]).

O tipo de *scheduling* eRTPS, definido em [4], combina características do UGS e do RTPS. O serviço foi idealizado para suportar SFs de UL em tempo real que geram, periodicamente, pacotes de tamanho variável, como, por exemplo, VoIP com supressão de silêncio. Este mecanismo de *scheduling* deve garantir slots de UL, mesmo que não solicitados, como no UGS, mas o tamanho desta reserva não é fixo, mas dinâmico, como no eRTPS. A BS realiza alocações periódicas no UL para requisição de largura de banda ou para transmissão de dados das SSs utilizando eRTPS. As MSs podem solicitar alteração de largura de banda através do campo *extended PBR* do GMS, do cabeçalho de sinalização BR ou ainda em conjunto com o envio de informações de camada física. No caso de requisições de tamanho nulo, a BS poderá alocar recurso suficiente apenas para o subcabeçalho BR ou até mesmo não reservar nenhum *slot*. Neste caso, a MS poderá participar do período de disputa por oportunidades de requisição de largura de banda. Os parâmetros de qualidade de serviço obrigatórios para o eRTPS são: MRTR, MSTR, ML e as regras de requisição e transmissão (seção 6.3.5.2.2.1 de [4]).

O tipo de escalonamento nRTPS assegura oportunidades individuais de requisição de bandas para as SSs mesmo em períodos que a rede estiver congestionada. A BS questiona as SSs sobre a necessidade de banda, tipicamente, em um período não inferior a 1s. No entanto, as regras de requisição e transmissão devem permitir que as SSs utilizem os períodos de disputa por *slots* de UL. Este tipo de serviço é utilizado para aplicações com tolerância a atraso, mas que devem ter uma garantia mínima de largura de banda. Os parâmetros de qualidade de serviço obrigatórios para o nRTPS são: MRTR, MSTR, TP e as regras de requisição e transmissão (seção 6.3.5.2.3 de [4]).

Por fim, o serviço BE não possui qualquer requisito de largura de banda de transmissão ou de requisito de tempo real. Este tipo de serviço é utilizado, tipicamente, para a navegação na WEB. As regras para requisição e transmissão devem permitir que as SSs, utilizando o serviço BE, participem dos períodos de disputa por *slots* para requisição de largura de banda para transmissão. Desta forma, as SSs só conseguirão oportunidade de transmissão por requisição individual ou por períodos de contenção. Nenhum parâmetro de QoS é definido para este tipo de serviço (seção 6.3.5.2.4 de [4]).

3.4.2.4 Parâmetros de QoS

Cada serviço de escalonamento, definido na seção 3.4.2.3, possui parâmetros de QoS associados, são eles: MSTR, MRTR, ML, TJ, TP e as regras de requisição/transmissão. Explicaremos a seguir estes parâmetros.

Type	Length	Value
[145/146].12	4 <u>1</u>	Bit #0 – Service flow shall not use broadcast bandwidth request opportunities. (Uplink only) Bit #1 – Reserved; shall be set to zero Service flow shall not use multicast bandwidth request opportunities. (Uplink only) Bit #2 – The service flow shall not piggyback requests with data. (Uplink only) Bit #3 – The service flow shall not fragment data. Bit #4 – The service flow shall not suppress payload headers (CS parameter) Bit #5 – The service flow shall not pack multiple SDUs (or fragments) into single MAC PDUs Bit #6 – The service flow shall not include CRC in the MAC PDU. Bit #7 – Reserved; shall be set to zero

Figura 3.17: Regras de requisição/transmissão

Todos os parâmetros de qualidade de serviços são definidos através de codificações TVL (do inglês, *Type*, *Value and Length*) utilizadas em arquivos de configuração e em mensagens de gerenciamento. O campo *Type* possui tamanho de 1 byte e segue uma extensa lista de tipos que pode ser encontrada em [4] (ver seção 11). Os códigos do 145 são utilizados para SFs de UL e 146 para SFs de DL. Um item do tipo [145/146] significa que o mesmo é válido para ambos os sentidos de transmissão. Para o gerenciamento dos SFs, são definidos 46 subtipos como por exemplo, o [145/146].7, que contém a definição da MSTR.

O campo *Length* indica o tamanho da codificação. O campo *Value* contém o valor do parâmetro que está sendo enviado. O tamanho do campo *Length* varia com o campo *Value*, contendo no mínimo 1 byte, para conteúdos do campo *Value* inferiores a 127 bytes, variando para se adequar ao tamanho deste campo para valores maiores.

Os parâmetros MSTR e MRTR são obrigatórios para todos os tipos de escalonamento. A MSTR corresponde a taxa máxima de informação de um determinado serviço expressa em bps (*bit per second*). O campo *length* é de 4 bytes, o que permite valores da ordem de Gbps. A MSTR é medida na CS e não inclui, portanto, o *overhead* da camada MAC. A estratégia de escalonamento deve levar em consideração os cabeçalhos MAC e o CRC, quando utilizado, para permissão da taxa máxima total. Vale lembrar que esta taxa é permitida, mas nem sempre estará disponível (seção 11.13.6 de [4]).

A MRTR indica a taxa mínima que deve ser reservada para um SF. Esta taxa também é medida na CS e permite expressar valores em bps de até 4 bytes. A taxa agregada MRTR de todos os SFs pode exceder a taxa de transmissão disponível. Se este parâmetro for omitido, o valor padrão é zero, ou seja, não será garantida nenhuma largura de banda (seção 11.13.8 de [4]).

As regras de requisição/transmissão possuem uma codificação de oito bits, sendo o bit #7 reservado. O bit #0, por exemplo, define se a conexão de UL vinculada aquele SF deve ou não participar de períodos de oportunidade de BR em *broadcast*. A figura 3.17 apresenta um fragmento de uma tabela de [4] (ver seção 11.13.12) com todas as regras.

A ML é o tempo máximo permitido desde a entrada do pacote na CS até a entrega

do SDU referente para interface aérea; ou seja, é a latência da camada MAC. Este valor deve ser medido em milisegundos e expresso em um campo de até 4 bytes. Ressaltamos que BS e SSs deixam de ter compromisso com este parâmetro quanto a taxa de transmissão excede o valor da MRTR (seção 11.13.14 de [3]). O TJ é máxima variação do atraso (*jitter*) permitido para a conexão associada aquele SF. Este parâmetro é medido em milisegundos, podendo ser expresso por um número de até 4 bytes (seção 11.13.14 de [3]). Este parâmetro não poderá ser desconsiderado, mesmo que a MRTR ou MSTR tenham sido atendidas.

O TP é o parâmetro responsável por indicar a prioridade entre os SFs. Caso dois SFs possuam critérios idênticos de qualidade de serviço, o SF de maior valor de TP deve ter preferência para armazenamento em *buffer* e menor atraso. Para SFs com parâmetro distintos, permanecem os demais critérios de qualidade. Este parâmetro deve ser considerado para efeito de desempate por uma estratégia de escalonamento (seção 11.13.5 de [3]).

3.4.2.5 Requisição e alocação de largura de banda

Através dos três pares de conexões estudados, é possível realizar o controle do tráfego das SSs, inclusive com relação aos níveis de qualidade de serviço, notadamente a largura de banda de transmissão. Existem diversos métodos para que uma SS possa requisitar largura de banda a BS. Nesta seção, descreveremos estes métodos.

Uma requisição é o mecanismo pelo qual a SS indica à BS sua necessidade de alocação de banda no UL para transmissão. Esta requisição pode ser realizada através do cabeçalho de sinalização BR ou por PBR. A capacidade de PBR é opcional. Como as condições do canal de comunicação são variáveis e, conseqüentemente, os perfis de transmissão, a requisição deve levar em conta o número de bytes necessários para o cabeçalho e *payload* da camada MAC, sem fazer referência ao *overhead* da camada PHY. Uma SS só deve realizar requisições de banda quando possuir PDUs para transmissão (seção 6.3.6.1 de [3] e [4]).

As requisições podem ser incrementais ou agregadas. No caso de uma requisição incremental, a BS deve adicionar à sua percepção de largura de banda daquela conexão, a quantidade de bytes solicitada (seção 6.3.6.1 de [3]). No caso de uma requisição agregada, a BS deve substituir a sua percepção de largura de banda pela quantidade de bytes solicitada. O campo *Type* do cabeçalho de sinalização BR indica se a solicitação é incremental ou agregada (ver tabela 3.8) [4]. Solicitações através de PBR, como não possuem o campo para definição do tipo, devem ser sempre incrementais. Devido a possibilidade de colisão, as requisições provenientes de períodos de contenção devem ser agregadas.

A característica de auto-correção do protocolo de requisição/alocação de largura de banda requer que, periodicamente, as SSs possam realizar requisições de largura de banda agregada, como função da manutenção da qualidade de serviço e da qualidade do canal de comunicação. A capacidade de BR incremental é opcional para as SSs e obrigatória para a BS. A capacidade de BR agregada é obrigatória para ambas (seção 6.3.6.1 de [3]).

Para as SSs, as requisições são realizadas por conexão, mas a reserva dos recursos (em inglês, *grants*) é realizada para o CID de conexão básica da SS e não para cada CID. Portanto, como a SS não sabe para que conexão os *grants* foram realizados, a SS deve possuir um mecanismo de análise das requisições e distribuição dos recursos entre suas conexões (seção 6.3.6.2 de [3]). O algoritmo de escalonamento nos SSs é o responsável por esta função. Caso a alocação seja inferior a requisição realizada, a SS pode optar por descartar os SDUs ou realizar o *backoff* (pode ter havido colisão) e iniciar nova transmissão.

O *Polling* é o processo pelo qual a BS aloca recursos específicos para que as SSs possam requisitar largura de banda para transmissão. Estas alocações podem ser individuais ou para grupos de SSs. As alocações não são realizadas através de mensagens diretas, mas pelo envio dos IEs nas mensagens de UL-MAP. Existem três tipos de *polling*: *unicast*, *multicast* e *broadcast* (seção 6.3.6.3 de [3]).

No *unicast polling*, não são enviadas mensagens individuais para cada SS, mas garantida largura de banda para transmissão do cabeçalho de sinalização BR, alocando IE correspondente ao SS na mensagem UL-MAP. O *unicast polling* é feito com base no SS, utilizando o CID da conexão básica (seção 6.3.6.3.1 de [3]). Uma conexão do tipo UGS não participa do *polling*, a menos que o bit PM esteja ativo. Este procedimento economiza largura de banda.

Porém, é possível que não exista largura de banda suficiente para todas as SSs que desejam participar do *unicast polling*. As SSs não atendidas pelo *unicast polling* podem participar dos grupos de *multicast polling* ou do *broadcast polling* (seção 6.3.6.3.2 de [3]). Como no caso do *unicast*, a alocação é realizada na mensagem de UL-MAP. A diferença é que ao invés da alocação utilizar o CID da conexão básica da SS, são utilizados CIDs específicos de conexões de *multicast* e *broadcast*. Em caso de colisão, um algoritmo para resolução da disputa é executado apenas entre as SSs que enviaram a requisição de banda não nula.

A camada PHY WirelessMAN - OFDMA especifica os subcanais de *ranging* (*Ranging Subchannel*) e um conjunto de códigos de *ranging* (*Ranging Codes*) que são utilizados para os períodos de disputa por largura de banda de transmissão de UL (seção 6.3.6.5 de [3]). A BS deve prover alocação dos recursos de UL para os SSs enviando o CID de *broadcast* em combinação com o código de *ranging* especificando o instante de transmissão. Com este conjunto de mecanismos e mensagens, o algoritmo de escalonamento, executado na BS, deve alocar os recursos existentes entre as SSs para prover os serviços solicitados, dentro dos parâmetros de qualidade acordados.

3.4.2.6 Suporte à camada física

A camada MAC deve suportar os métodos de duplexação TDD e FDD. Esta dissertação utilizou-se apenas o TDD. Neste caso, as transmissões de UL e DL ocorrem em instantes

diferentes, normalmente na mesma frequência. O *frame* TDD possui duração fixa e é dividido em um *subframe* de UL e um *subframe* de DL. O *frame* é dividido em uma quantidade inteira de PSs, o que facilita a divisão dos recursos (seção 6.3.7.2 de [3]).

Apesar do tamanho fixo do *frame*: a razão entre os *subframes* pode variar de acordo com as características da rede. As razões são escritas na forma, porção do *frame* para DL : porção do *frame* para UL, sendo comum as taxas 3:1, 3:2 e 1:1 [35].

As mensagens de DL-MAP definem os intervalos para cada tipo de transmissão de DL da camada PHY. As mensagens UL-MAP definem o tempo de alocação inicial (*allocation start time*) para cada tipo de transmissão de UL da camada PHY. Este tempo inicial pode se referir ao mesmo *frame* ou ao *frame* seguinte (seção 6.3.7.4.1 de [3]). Para camada PHY WirelessMAN-OFDMA, a alocação de largura de banda de uplink (UL-MAP) é realizada em unidades de subcanais e símbolos (seção 6.3.7.4.2 de [3]). São as mensagens UL-MAP que conterão as reservas de banda da estratégia de escalonamento e as transmitirão aos SSs.

3.4.3 Subcamada de Segurança

A subcamada de segurança tem como objetivo prover autenticação, privacidade e confidencialidade aos assinantes da rede de comunicação sem fio (seção 7 de [3]). A BS protege o transporte dos dados de acessos não autorizados, fornecendo segurança aos SFs associados a sua rede.

A subcamada de segurança utiliza um protocolo de gerenciamento de chaves (PKM, do inglês, *key management protocol*) entre o cliente e o servidor, no qual a BS é o servidor, controlando a distribuição de chaves para as SSs (seção 7.1.2 de [3]). Adicionalmente, mecanismos de segurança básica são fortificados através da utilização, em conjunto com o PKM, de certificados digitais para autenticação dos dispositivos das SSs. Como não é o foco deste trabalho, não detalharemos as funcionalidades da subcamada de segurança.

Capítulo 4

Estratégias de Escalonamento

4.1 Introdução

Uma estratégia de escalonamento deve ser dimensionada a partir dos recursos disponíveis. O padrão IEEE 802.16 permite operação em TDD e FDD. A maior parte das estratégias de escalonamento utiliza TDD, por ser um método mais eficiente de alocação de recursos [36]. O modo TDD permite uma configuração assimétrica dos *subframes* de UL e DL para tratamento de tráfegos desbalanceados. Neste trabalho, avaliaremos uma técnica de escalonamento que utiliza o modo TDD. No modo TDD, uma estratégia de escalonamento é responsável por reservar os recursos para transmissão de UL das SSs para a BS.

A norma [3] permite a transmissão de dados multimídia utilizando tipos de escalonamento, associados a parâmetros de QoS; são eles: UGS, RTPS, nRTPS e BE. No adendo IEEE 802.16e [4], foi incluído mais um tipo, o eRTPS. Todos estes tipos de tráfegos e escalonamentos podem coexistir em uma SS e ambas, SS e BS, monitoram o tráfego das conexões entre os mesmos. As requisições de banda de *uplink* podem ser fornecidas utilizando dois mecanismos: o GPC (do inglês, *Grant per Connection*), no qual a reserva é realizada para cada conexão e o GPSS (do inglês, *Grant per Subscriber Station*), no qual a reserva é realizada para a SS agregando as requisições de todas as conexões pertencentes ao mesmo. Este último é considerado o mais eficiente deles ([37]-[38]).

Baseado nas requisições solicitadas e nas requisições garantidas, a BS deve ser capaz de reservar recursos para o tráfego de *uplink* de cada SS, de acordo com a qualidade de serviço acordada. A alocação realizada é transmitida às SSs através de mensagens UL-MAP, anexas ao *subframe* de *downlink*. No entanto, a norma de WIMAX não define uma estratégia de escalonamento. Os fabricantes estão livres para escolher uma técnica para esta finalidade, respeitando as classes de serviços e demais características do padrão.

4.2 Estratégias de Escalonamento

Para o funcionamento de uma rede WiMAX, é fundamental a existência de uma entidade responsável pela alocação dos recursos disponíveis nas estações rádio-base. Esta alocação ocorre da seguinte forma [39]:

- os pacotes da camada de aplicação são classificados pelas SSs de acordo com a classe da conexão;
- os pacotes são encaminhados para a respectiva fila;
- de acordo com os fluxos de serviços gerados, a SS faz uma requisição de banda, reportando a quantidade de bits requerida para cada conexão à BS;
- o algoritmo de escalonamento é responsável por receber as requisições das SSs e alocar recursos de acordo com a classe de serviço das várias conexões de cada SS;
- o momento de transmissão de cada SS, em cada conexão, é então informado através de mensagens de controle específicas;
- cada SS decodifica sua mensagem e aguarda o período informado para transmissão.

Aproveitando lacunas existentes na norma técnica, diversos trabalhos vêm sendo realizados para encontrar uma estratégia de escalonamento eficiente que atenda aos requisitos de qualidade de serviço. Nas seções seguintes, examinaremos algumas destas estratégias.

Ao término do estudo, procurou-se, na literatura, formas para melhor agrupar as estratégias de escalonamento de redes cabeadas e sem fio, com o objetivo de facilitar o entendimento. Desta busca, foram encontrados dois artigos ([40] e [41]) que propõem classificações para estratégias de escalonamento.

Em [40], é proposta uma classificação baseada em duas categorias: métodos tradicionais e métodos novos. Os métodos tradicionais, englobam estratégias propostas para redes cabeadas, algumas cuja origem remonta o escalonamento de processos em sistemas operacionais. Os métodos tradicionais são divididos em métodos simples e hierárquicos. As estratégias simples são aquelas que propõem uma técnica para tratamento dos pacotes recebidos como, por exemplo, *First In First Out*, *Fair Queuing* e *Round Robin*. As técnicas hierárquicas são métodos que utilizam diferentes estratégias tradicionais para classes de serviço distintas. Estas técnicas, no entanto, apresentam alta complexidade de adoção prática [40].

Os métodos novos propõem soluções que divergem das estratégias propostas anteriormente. Duas subcategorias dividem estas estratégias: métodos que consideram apenas uma

classe de serviço e métodos que consideram todas as classes. O trabalho [13] é citado como uma estratégia nova que engloba todas as classes de serviços.

Em [41], são propostas três categorias: estratégias homogêneas, heterogêneas e oportunistas. As estratégias homogêneas são aquelas que utilizam uma única técnica, como, por exemplo, os trabalhos [42] e [43]. As estratégias heterogêneas são aquelas que utilizam mais de uma técnica, como, por exemplo os trabalhos [44] e [37]. Apesar de não citar [40], o trabalho [41] possui semelhanças com a classificação proposta em [40], na medida em que a categoria das estratégias homogêneas pode ser entendida como a subcategoria simples dos métodos tradicionais vista em [40]. Por outro lado, a categoria das heterogêneas é semelhante à subcategoria hierárquica dos métodos tradicionais. A diferença fica por conta das categorias das estratégias oportunistas [41] e da categoria de métodos novos [40]. Em [41], são consideradas estratégias oportunistas aquelas que baseiam-se em parâmetros de camada física para o escalonamento dos recursos da rede, como a relação sinal ruído, relação portadora interferência, eficiência espectral, entre outros parâmetros. Esta dissertação não realiza o estudo deste tipo de estratégia. Em [40], são considerados métodos novos aqueles que abandonam as estratégias tradicionais buscando novas formas para tratamento do problema.

Alguns problemas são enfrentados ao se utilizar estas classificações. Estratégias que podem ser consideradas cronologicamente novas utilizam estratégias ditas tradicionais como é o caso do trabalho [27], o que inviabiliza a utilização da classificação proposta em [40], publicada no ano seguinte ao trabalho [27]. Por outro lado, existem estratégias como a proposta em [45] que não podem ser consideradas homogêneas, heterogêneas, nem oportunistas, o que inviabiliza a utilização da classificação proposta em [41].

Adotou-se uma classificação baseada no problema tratado pela estratégia de escalonamento. As estratégias são agrupadas em técnicas para resolução de escalonamento de processos de sistemas operacionais e de redes fixas comutadas e estratégias propostas especificamente para o escalonamento de redes WiMAX.

4.2.1 Estratégias de escalonamento para sistemas operacionais e redes fixas comutadas

As estratégias de escalonamento baseadas em servir primeiro o pacote mais antigo a chegar ao servidor ou FIFO (do inglês, *First-IN, First-OUT*) ou, ainda, FCFS (do inglês, *First-Come, First-Serve*), tornaram-se ineficientes para suportar a sobrecarga de pacotes [54] e as novas demandas por serviços multimídia com diferentes requisitos de qualidade de serviço. O problema surge quando as requisições de cada conexão são enfileiradas em uma única fila. Mesmo que a capacidade de armazenamento em memória seja infinita, o tempo de vida de um pacote pode, com alta carga de dados, resultar no descarte do mesmo. A Figura 4.1, adaptada de [54], ilustra a técnica FIFO.

Tabela 4.1: Taxonomia de estratégias de escalonamento

Trab.	Técnica	Sist. Op. e Redes fixas.	Redes WiMAX	Local de decisão
[42]	RR	X	-	Servidor
[46]	OCF	X	-	Servidor
[46]	EDF	X	-	Servidor
[43]	WFQ	X	-	Servidor
[47]	SFQ	X	-	Servidor
[48]	WRR	X	-	Servidor
[49]	PGPS	X	-	Servidor
[50]	DRR	X	-	Servidor
[51]	PWFQ	X	-	Servidor
[44]	WRR, WFQ e FIFO	-	X	BS e SS
[37]	EDF e WFQ	-	X	BS e SS
[52]	DFPQ	-	X	BS
[27]	-	-	X	BS e SS
[13]	-	-	X	BS
[45]	-	-	X	SS
[8]	PBFS	-	X	BS
[53]	QBMS	-	X	SS

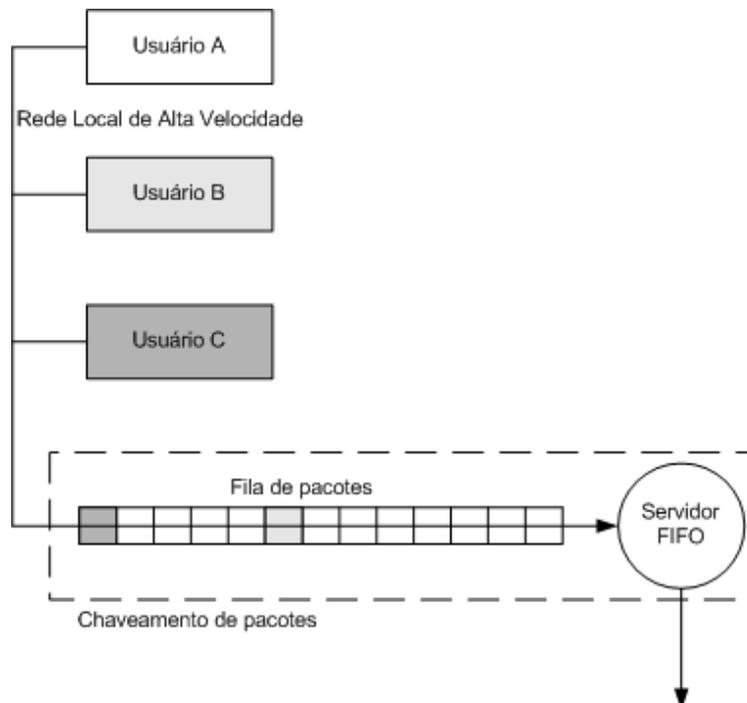


Figura 4.1: Servidor FIFO ou FCFS

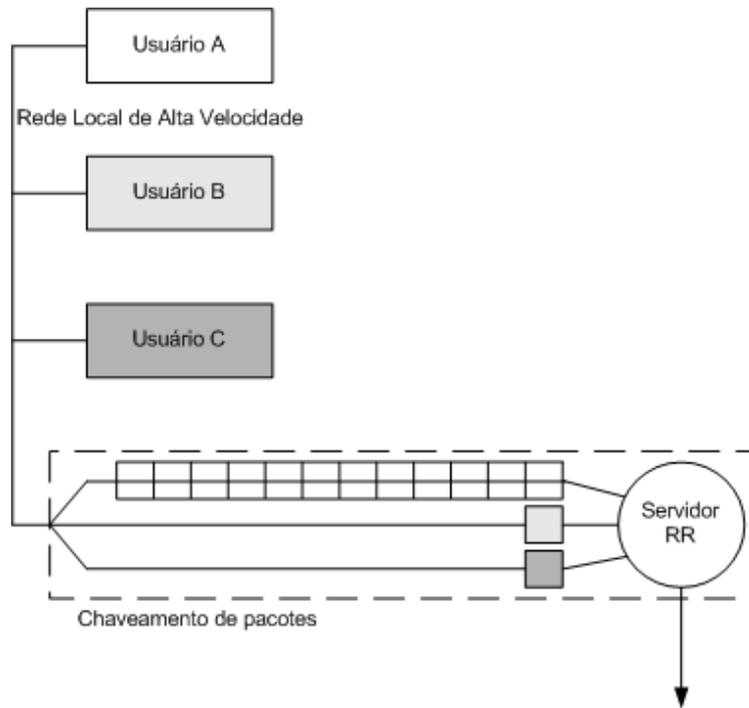


Figura 4.2: Servidor Round-Robin

Neste tipo de estratégia, as estações eram solicitadas a enviar o máximo de pacotes possível, porém o acúmulo de pacotes na fila pode ocasionar descarte dos mesmos, prejudicando estações mais lentas e o sistema como um todo. Estações que enviam um grande número de pacotes são capazes de monopolizar os recursos da rede.

O conceito de *fairness* (que pode ser traduzido como justiça) e a estratégia de escalonamento chamado *Round-Robin*(RR), inicialmente propostas em [42], foram conjuntamente a primeira tentativa inovadora de resolver o problema de escalonamento em redes fixas cabeadas de comutação de pacotes. A técnica divide requisições das diferentes conexões em filas distintas, servindo um pacote de cada fila a cada rodada. A Figura 4.2, adaptada de [54] ilustra esta situação.

Neste caso, as estações precisam transmitir exatamente um pacote para cada rodada da técnica, liberando a capacidade de processamento das mesmas para outras funções. O sistema, de maneira geral, trabalha de forma isonômica para com as requisições recebidas. Com esta técnica, estações que possuem elevadas taxas de transmissão (grande número de envio de pacotes por um determinado período de tempo) terão as mesmas oportunidades daquelas com taxa inferior. Obviamente, estas estações ainda poderão ter pacotes descartados, mas de forma mais tardia que na estratégia FIFO.

A técnica RR, no entanto, é injusta na visão do tamanho dos pacotes. Apesar de fazer o balanceamento dos pacotes, não existe um balanceamento da carga propriamente dita. Não é realizada uma verificação com relação à quantidade de bits ou da largura de banda que será necessária para transmitir cada pacote. Existe também o problema relacionado com

a capacidade do nó responsável pelo tratamento das conexões, pois são necessárias tantas filas quantas forem às conexões.

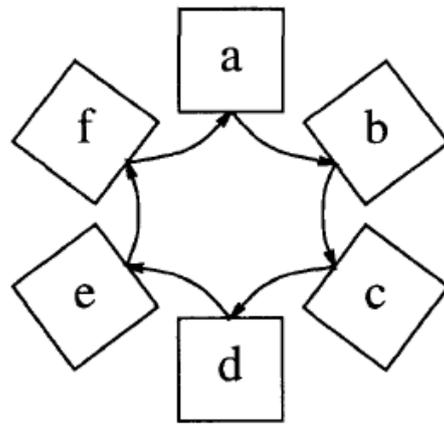
Outra técnica derivada do RR é o SFQ (do inglês, *Stochastic Fair Queuing*) [47]. Esta técnica resolve o problema de quantidade de filas, mas não trata o tamanho dos pacotes. A estratégia consiste na aplicação de uma função de *hash* a cada fluxo de dados, encontrando um valor dentro de intervalos previamente definidos. Cada um destes intervalos constituirá em uma fila para ser despachada. Filas muito longas continuam a ter pacotes descartados e o tamanho dos pacotes não é avaliado. Porém, o *hash* é executado em tempo $O(1)$ e a quantidade de filas é reduzida. A eficiência da técnica está atrelada à suposição de que o número de filas é menor do que o número de fluxos de dados. A justiça na transmissão dos pacotes é garantida de forma probabilística, daí a característica estocástica, e as filas são servidas por *round-robin*.

Outra estratégia baseada em *Round-Robin* é o WRR (do inglês, *Weighted Round-Robin*) [48]. Propõe-se servir as filas como RR, mas utilizando um conjunto de pesos W_i , associado às conexões para que o sistema seja mais justo. Estes pesos permitem que algumas conexões sejam servidas mais de uma vez em cada rodada. A figura 4.3 [48] ilustra o RR e WRR nas letras a e b, respectivamente. Desta forma, as conexões são servidas de acordo com o peso de cada uma delas. Apesar de [48] tratar o problema de alocação de VCs (do inglês, Virtual Circuits) em redes ATMs, a técnica pode ser usada para outras redes.

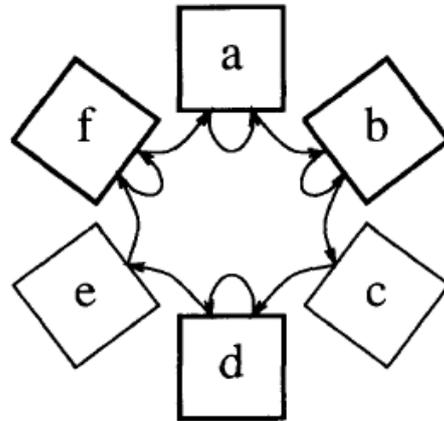
A técnica Bit-a-Bit Round-Robin (BR) procura resolver o problema relacionado a variabilidade do tamanho dos pacotes [43]. A idéia da estratégia BR é permitir o envio de um bit para cada conexão. Como, na prática, é impossível a execução desta idéia, a técnica busca simular este comportamento calculando o tempo aproximado que um pacote teria saído de um nó da rede em direção ao nó atual (processador ou servidor da fila). Em seguida, o pacote é inserido em uma lista ordenada, baseada no tamanho do pacote, encontrado a partir da diferença de tempo mencionada. A estratégia BR é simulada utilizando-se a rodada (*round*), como um tempo virtual em que o pacote, e não cada bit, começa e termina a ser servido.

A técnica é baseada nos seguintes parâmetros:

- α , identifica uma conversação (conexão entre fonte e destino);
- i^α , identifica cada pacote em uma conversação α ;
- $R(t)$, número de rodadas até um tempo t ;
- P_i^α , tamanho do pacote i na conexão α ;
- t_i^α , tempo de chegada do pacote i da conexão α ao nó processador da fila;
- S_i^α , valor de $R(t)$ quando o pacote i começou a ser processado;



(a)



(b)

Figura 4.3: Funcionamento RR (a) e WRR (b)

- F_i^α , valor de $R(t)$ quando o pacote i terminou de ser processado.

Desta forma, os valores de F_i^α são definidos como $F_i^\alpha = S_i^\alpha + P_i^\alpha$ e $S_i^\alpha = \text{Max}(F_{i-1}^\alpha, R(t_i^\alpha))$, sabendo-se que $R(t) = R(t_0) + P$ é a rodada de término de serviço do pacote de tamanho P . A emulação do BR ordena a saída dos pacotes através dos valores de F_i^α . Esta técnica passou a ser conhecida como *Packet-to-Packet Round-Robin* ou *Weighted Fair Queueing* (WFQ). Porém, o custo para inserir um pacote na lista já ordenada é elevado, $O(\log(n))$, sendo n o número de conexões.

O problema do número de filas não é resolvido, mas, pela primeira vez, é definida uma forma de tratar requisitos de tempo real em estratégias de *scheduling*. A estratégia original ordena o envio de pacotes baseado no número F_i^α . O menor número F_i^α é enviado primeiro. Assim, conexões α com pacotes menores podem ser escalonadas, impedindo que os de tamanhos maiores ocupem todo recurso.

Para o problema de *delay* (um dos requisitos de tempo real) foi inserida uma nova variável e um parâmetro de sintonia, B_i^α e δ , respectivamente, sendo que :

- $B_i^\alpha = P_i^\alpha + \text{Max}(F_{i-1}^\alpha, R(t_i^\alpha) - \delta)$

A variável B_i^α representa o valor de $R(t)$ quando o pacote i terminou de ser processado, corrigido pelo parâmetro δ . O envio dos pacotes processados é ordenado pelos valores de B_i^α , não mais por F_i^α . Quando $R(t_i^\alpha) \leq F_{i-1}^\alpha$, a conversação α está ativa. A técnica se comporta como o BR (existem bits na fila para processamento), independentemente do valor de δ . Quando $R(t_i^\alpha) \geq F_{i-1}^\alpha$, a fila estava inativa e a estratégia passa a depender do valor de δ . Para $\delta=0$, $B_i^\alpha = P_i^\alpha + R(t_i^\alpha)$, completamente independentemente do histórico de conversações α , mas do número de rodadas. Se $\delta = \infty$, $B_i^\alpha = P_i^\alpha + F_{i-1}^\alpha$, depende da rodada de término do penúltimo pacote da conversação α , independente de quantas tenham ocorrido anteriormente. Para valores intermediários de δ , a decisão de escalonamento de um pacote que chegou a uma fila inativa depende da rodada de término do último pacote e se ela ocorreu há muitas rodadas.

É possível, portanto, para elevados valores de δ , forçar o escalonamento de pacotes de determinada conversação que estão há algumas rodadas sem ter pacotes processados. Foi incluído um tratamento de requisito de tempo para filas que estavam inativas, no entanto, não existe um tratamento de *delay* referente aos pacotes já enfileirados, como uma mudança de prioridade por exemplo. Apesar disso, a estratégia descrita possui menor *delay* que estratégias FCFS e representou um marco no controle de filas com pacotes de tamanho variáveis de forma justa (FQ - fairness queueing), sendo amplamente citada na literatura ([47], [49], [51] e [37]). A figura 4.4 demonstra o ganho obtido com o WFQ.

Uma estratégia baseada em WFQ (também chamado GPS, Generalized Processor Sharing) é o PGPS (do inglês, Packet Generalized Processor Sharing) [49]. Esta técnica combina

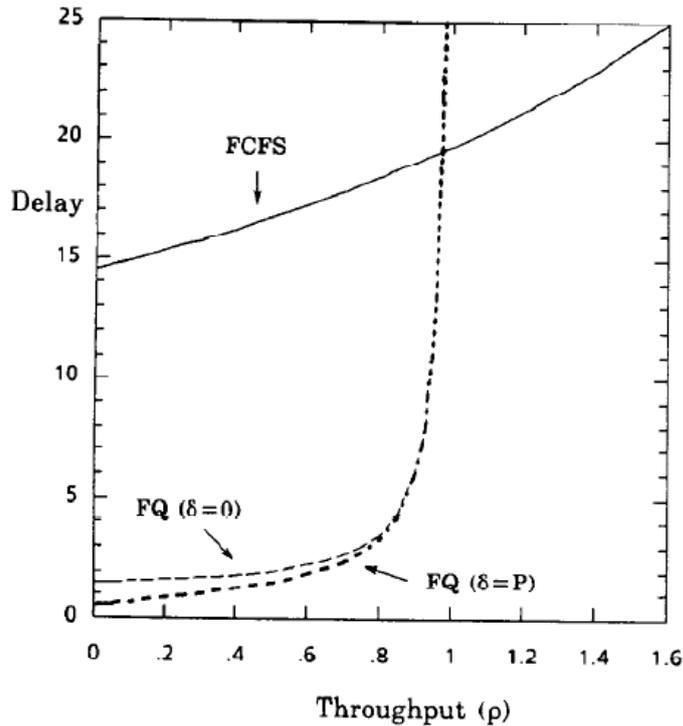


Figura 4.4: Gráfico Delay x Taxa de saída das estratégias FCFS e WFQ

a versão para pacotes do GPS com uma estratégia de restrição do tráfego entrante, conhecida como Leaky Bucket (LB). A estratégia LB restringe o tráfego, permitindo que seja entregue para a rede apenas a quantidade definida pela equação 4.1:

$$A_i(\tau, t) \leq \sigma_i + \rho_i(t - \tau), \forall 0 \leq \tau \leq t \quad (4.1)$$

onde,

- ρ_i é a taxa em que são garantidas as requisições de tráfego em cada conexão i ;
- σ_i é a quantidade de bits garantida para requisição, de cada conexão.

A combinação destas duas estratégias é utilizada também em [51], mas com tratamento especial para o *delay*.

A estratégia *Deficit Round-Robin*(DRR) [50] procura resolver o problema do tamanho do pacote e da quantidade de filas. Utiliza o algoritmo SFQ para atribuição de um fluxo de dados a uma fila, diminuindo o número de filas que deverão ser consumidas.

As filas são servidas utilizando-se a estratégia RR, mas com uma diferença: é mantida uma variável, o *quantum*, representando a quantidade de bits de um pacote não enviado em

uma rodada (por ser muito grande). Este valor é considerado um *deficit* para a próxima rodada. Assim, filas que foram prejudicadas em uma rodada são recompensadas na rodada seguinte.

De forma a evitar o exame de filas vazias (intervalos não utilizados gerados pelo *hasing*) é mantida uma variável apenas com as filas nas quais existe ao menos um pacote. A técnica DRR, apesar de resolver o problema de tamanho de pacotes, não resolve o problema de requisito de tempo. Outro ponto desfavorável à utilização desta técnica é a necessidade do conhecimento da quantidade de bits da fila.

Algumas técnicas foram construídas com o objetivo de tratar especificamente a prioridade. Em [46], é possível encontrar os princípios de funcionamento de duas técnicas: o *Oldest Customer First* (OCF) e o *Earliest Deadline First* (EDF).

Quando é utilizada a estratégia FCFS, pacotes servidos por filas maiores sofrerão maiores *delays* do que pacotes servidos por filas menores. A idéia por trás do OCF é dar prioridade aos pacotes que já estão há mais tempo trafegando na rede. A técnica pode ser implementada utilizando *timestamp* (tempo de marcação) do pacote. Os pacotes que chegam ao servidor (nó da rede que apresenta restrição de tráfego) são ordenados de forma decrescente em relação às suas "idades", em seguida processados. Em [46], prova-se um teorema que afirma que com a técnica OCF, aplicada a um servidor de filas com taxas de transmissão constantes, é possível:

- minimizar a idade máxima de envio de um pacote;
- maximizar a idade mínima de envio de um pacote; e
- minimizar a variação entre as idades de envio dos pacotes.

Por sua vez, a urgência para servir um pacote não pode ser baseada apenas no tempo de existência do mesmo na rede, mas também em outras informações como o seu tempo de expiração (*time expiration*), condições de tráfego de filas, rotas para atingir o destino, entre outros. A técnica utilizada em [46] possibilita a definição de um tempo limite para processamento de um pacote, admitindo-se um *delay* máximo tolerável, D_{max} . Supondo-se que um pacote chegou a um nó da rede em um tempo t , com idade a e possui um tempo estimado de roteamento até o destino d , o tempo de processamento limite (*deadline time*) deve ser $t + (D_{max} - d - a)$.

Em [46], prova-se o teorema que afirma ser possível com a técnica EDF:

- minimizar o atraso máximo;
- maximizar o atraso mínimo; e
- minimizar a variação do atraso.

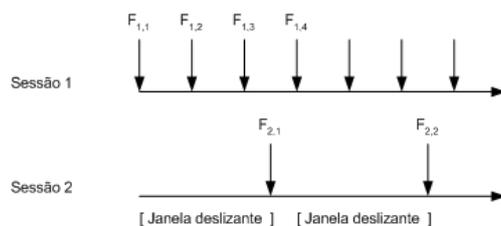


Figura 4.5: Funcionamento da janela deslizante do PWFQ

Por este motivo, a estratégia EDF foi escolhida, por muitos autores, para realização de escalonamento de tráfego em tempo real (como pode ser vistos nos trabalhos de [55], [37] e [52]).

A estratégia PWFQ (do inglês, *Priority-based Weighted Fair Queueing*) [51] combina o WFQ com o conceito de prioridade. Como já foi visto, o WFQ realiza o cálculo de um tempo virtual no qual um pacote terminará de ser servido. Aparentemente, incluir um fator de prioridade é incompatível com aquela técnica. A idéia que aparece em [49] e [43] é a de que quanto maior o percentual de recurso de largura de banda uma sessão utiliza, menor o *delay*. PWFQ propõe considerar ambos, balanceamento do uso da banda e prioridade para redução do *delay*. A idéia é escalonar pacotes que terminarão de ser processados em um mesmo tempo virtual de acordo com determinada prioridade de suas conexões, introduzindo o dispositivo de janela deslizante (*slide window*).

A janela deslizante começa no tempo atual e tem um tamanho fixo definido na estratégia. Os pacotes cujo tempo virtual de processamento termina dentro da janela são selecionados, ordenados de acordo com a prioridade mais alta (de suas conexões) e enviados. A figura 4.5, adaptada de [51], ilustra o funcionamento da estratégia PWFQ.

Suponha-se que a sessão 2 possui uma prioridade superior à da sessão 1, mas que esta ocupa maior parte do canal de comunicação. Supondo F_i^α como um tempo virtual, conforme definido anteriormente, a estratégia WFQ serviria os pacotes i da conversação α (i^α) na seguinte ordem: $1^1, 2^1, 3^1, 1^2, 4^1 \dots$. Por outro lado, se for utilizado o PWFQ, os pacotes $1^1, 2^1, 3^1, 1^2$ serão selecionados pela janela da figura 4.5. Em seguida, é verificada a prioridade de cada conversação (pode ser utilizado o termo conexão ou sessão) α . Como a conversação 2 possui prioridade superior a conversação 1, o pacote 1^2 é enviado primeiro e em seguida são enviados os pacotes da conversação 1. Do ponto de vista da divisão dos recursos, a conexão 1 recebe três vezes mais oportunidades que a 2, garantindo o balanceamento de carga. Mas a ordem de envio dos pacotes é alterada em razão da prioridade, ao invés da rodada de término de envio do pacote.

É possível notar que o tamanho da janela deslizante é um parâmetro fundamental para o sucesso da estratégia, prevenindo o *delay* excessivo, e deve ser sintonizado da forma mais eficiente possível. Um extremo seria estender a janela para todas as conexões e todos os pacotes. Neste caso hipotético, todas as conexões prioritárias teriam seus pacotes enviados

à frente das demais, transformando a estratégia em um tratamento puro da prioridade do tráfego. Uma janela nula transforma a técnica na estratégia WFQ original. Escolhendo-se uma janela que permita razoável balanceamento de tráfego e elevado tratamento da prioridade, as conexões terão oportunidade de serem atendidas e o requisito de tempo será respeitado (o *delay* obedecerá um limite aceitável).

O tamanho da janela deve ser definido baseado nas conexões que apresentam requisitos de tempo real. O máximo *delay* permitido para uma conexão i pode ser encontrado utilizando-se a equação 4.2 [49]:

$$D_i \leq \left(\frac{\sigma_i}{g_i} + \frac{L_{max}}{r} \right) \quad (4.2)$$

onde,

- σ_i é quantidade de bits entregue ao servidor proveniente da conexão i ;
- g_i é a taxa garantida de envio de pacotes da conexão i ;
- L_{max} é o tamanho máximo do pacote entre todas as conexões; e
- r é a taxa de serviço do servidor.

A taxa garantida de envio de pacotes de cada conexão pode ser definida de acordo com a equação 4.3 [49]:

$$g_i = \left(\frac{\phi_i}{\sum_{j=1}^N \phi_j} \right) \times r \quad (4.3)$$

onde,

- ϕ_i é o percentual de serviço da conexão i ; e
- ϕ_j é o percentual de serviço da conexão j .

No entanto, com a introdução da prioridade na estratégia, a definição do máximo *delay* é alterada, passando a vigorar a equação 4.4

$$D_i \leq \max \left(\frac{\phi_i}{g_i} - W, 0 \right) + d_i + \frac{L_{max}}{r} \quad (4.4)$$

onde,

- W é o tamanho da janela; e
- d_i é o *delay* provocado na conexão i por todas as outras conexões anteriores de prioridade superior.

Portanto, encontrar o tamanho da janela W não é tão simples e envolve o cálculo recursivo, visto que d_i depende de W , com complexidade de tempo, para o pior caso, de $O(N^2)$ [51].

Porém, a estratégia traz um fator novo ao cenário já estudado, a capacidade de priorizar determinada conexão sem afetar a justa distribuição da capacidade de tráfego.

4.2.2 Estratégias de escalonamento para o WiMAX

As estratégias descritas até aqui foram e ainda são utilizadas para redes de comunicação com fio. O WFQ, por exemplo, é utilizado em roteadores para redes IP comercializados atualmente. No entanto, outras estratégias surgiram especificamente para o tratamento do escalonamento de redes WiMAX, com o objetivo de atender às particularidades das mesmas. Algumas destas técnicas propõem utilizar estratégias já aplicadas para redes cabeadas e outras propõem estratégias distintas.

O padrão IEEE 802.16, como vimos na seção 4.1, definiu como política de QoS cinco classes de tráfego: UGS, eRTPS, RTPS, nRTPS e BE. Desta forma, a reserva para o tráfego e a prioridade no tratamento de cada conexão deve respeitar aos critérios definidos para cada uma destas classes.

Em [44], descrevem-se as necessidades deste novo sistema e se propõe uma arquitetura de QoS para a camada MAC do padrão IEEE 802.16. Nesta arquitetura, são obrigatórios algoritmos de controle de admissão e de *scheduling*. Na figura 4.6, adaptada de [44], existe uma ilustração da arquitetura proposta. Para [44], uma rede WiMAX, para fornecer qualidade de serviço, deve possuir as seguintes estruturas:

- um classificador de tráfego para as SSs;
- um algoritmo de escalonamento de *upstream* para as SSs;
- um algoritmo de escalonamento de requisições de *downstream* para as BSs;
- e estas devem estar conjugadas com a camada MAC.

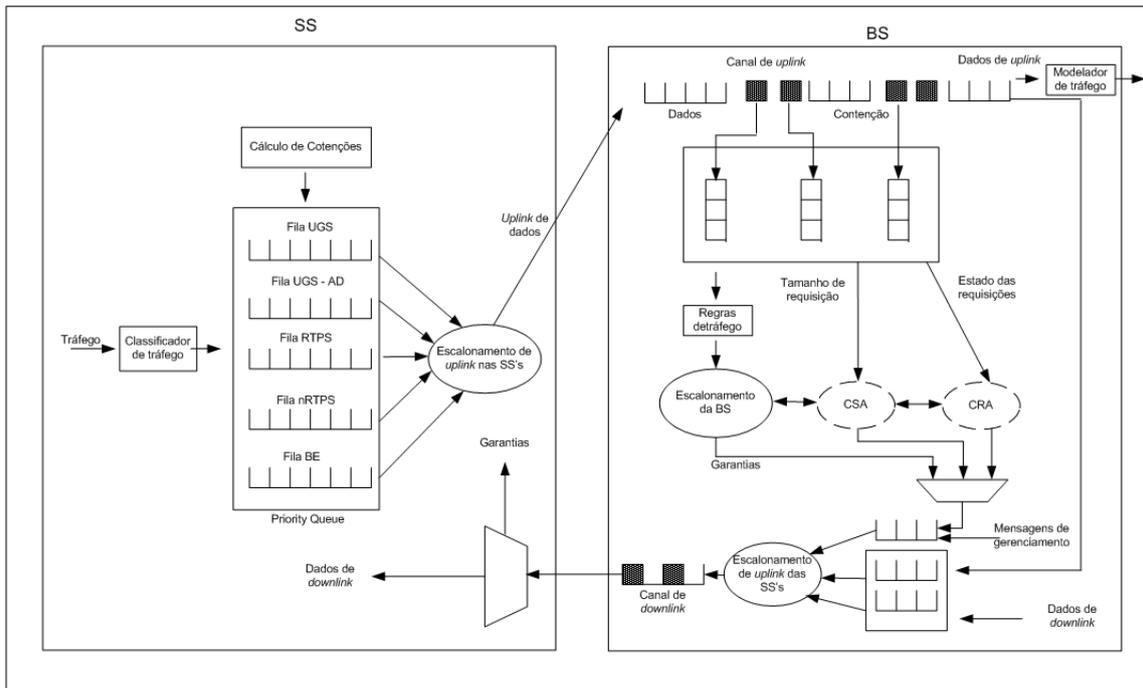


Figura 4.6: Arquitetura de QoS, proposta por Chu *et al.*

A seguir, são descritas as atividades do CSA (do inglês, Contention Slot Allocator), do CRA (do inglês, Contention Resolution Algorithm) e do US (do inglês, Upstream Scheduler). O CSA é responsável pela definição da quantidade de recursos que serão alocados para contenção e quantos para efetivo *upstream* de dados. A função do CSA é extremamente importante, pois, se forem alocados poucos recursos para contenção, podem ocorrer colisões; e recursos de contenção em excesso diminuem o espaço para transmissão de dados efetivos.

O CRA é o responsável por definir as regras de utilização dos recursos de contenção após o dimensionamento do CSA. Os USs, das BSs e das SSs, devem atuar de forma conjunta. A rede WiMAX, neste caso, opera no modo GPSS explicado anteriormente. De maneira que, o US de cada SS é responsável por reunir e escalonar as requisições de cada uma das suas conexões. O US da BS é responsável por receber as requisições das SSs e alocar recursos para *upstream* dos dados de cada conexão.

Em [37], é proposta uma arquitetura muito próxima, como é possível verificar pela figura 4.7 [37], apenas retirando o trabalho de alocação de recursos para a classe UGS do US, visto que a mesma possui requisito definido. A figura 4.7 evidencia as lacunas a serem preenchidas nas normas IEEE802.16 [3]. A norma não prevê a forma de escalonamento do tráfego de *uplink* das classes RTPS, nRTPS e BE, não define como será o Controle de Admissão na BS e, por último, não define o processo de *Traffic Policing* para as SSs.

Porém, nos trabalhos [44] e [37], nenhum novo algoritmo ou alteração dos existentes é proposto, mas é recomendada a utilização de estratégias baseadas em WFQ, WRR, EDF e FCFS (para baixa prioridade). No entanto, com a definição clara de classe de serviço e

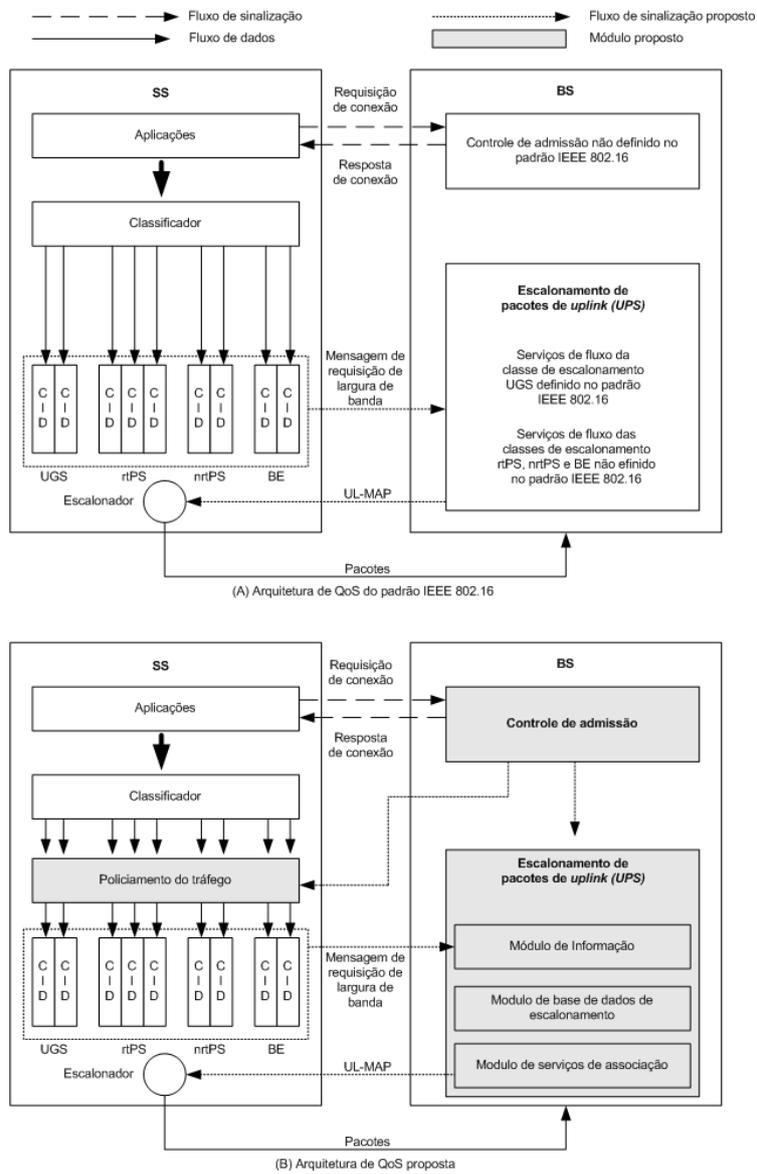


Figura 4.7: Arquitetura de QoS, proposta em Wongthavarawat *et al.*

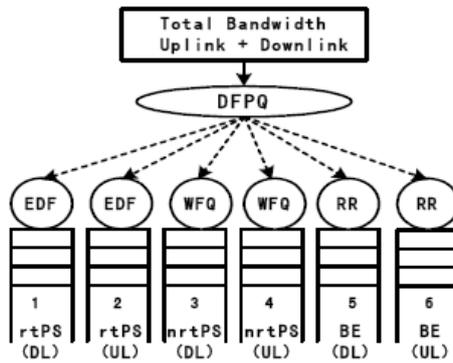


Figura 4.8: Estratégia DFPQ

com a demanda crescente por serviços de alto consumo de largura de banda e de requisitos de tempo real é necessário buscar estratégias mais simples para execução destas tarefas.

A estratégia proposta em [52] utiliza a arquitetura de [37] para propor uma técnica de escalonamento de duas camadas, o *Deficit Fair Priority Queue* (DFPQ). Trata-se de uma estrutura hierárquica em que o DFPQ recebe as requisições de escalonamento e em seguida, baseado na qualidade de serviço e na direção de tráfego, escolhe o algoritmo que será utilizado para a alocação dos recursos. A figura 4.8 ilustra a idéia apresentada em [52]. A proposta também é baseada em prioridade, de acordo com a seguinte ordem:

- RTPS, nRTPS e BE;
- *Downlink* e *Uplink*.

Desta forma, são utilizadas estratégias já conhecidas para tratar o problema, mas escolhendo a mais adequada para cada tipo de tráfego. O EDF é utilizado para as conexões do tipo RTPS, o WFQ para conexões nRTPS e o RR para conexões do tipo BE.

Em [27], a figura 4.9 ilustra a participação do *scheduler* para garantia de QoS em rede que atenda às normas IEEE 802.16 ([3] e [4]). Nesta arquitetura, é reafirmada a necessidade de algoritmos de alocação de recursos na BS (*uplink* e *downlink*) e no SS (*uplink*). Mais uma vez, é afirmada a liberdade para os fornecedores embutirem nas BSs e Ss a técnica de escalonamento de recursos.

No entanto, nenhum algoritmo novo é proposto. Em [27] sugere-se a utilização do DRR([50]) para o *downlink* da BS, em razão de possuir maior simplicidade de implementação. Por outro lado, o DRR não pode ser utilizado para o *uplink* devido à obrigatoriedade de conhecimento do tamanho da fila (tamanho de cada pacote na fila) a ser servida.

Foi então escolhido o WRR([48]) para o *uplink scheduler* da BS. O DRR foi o escolhido para o *uplink scheduler* dos Ss, pois cada SS conhece o tamanho dos pacotes em cada fila. Os trabalhos de [50] e [48] não abordaram as classes de serviços como [27] propõe, mas este relata apenas que os algoritmos foram especialmente configurados para tratamento de conexões do tipo RTPS.

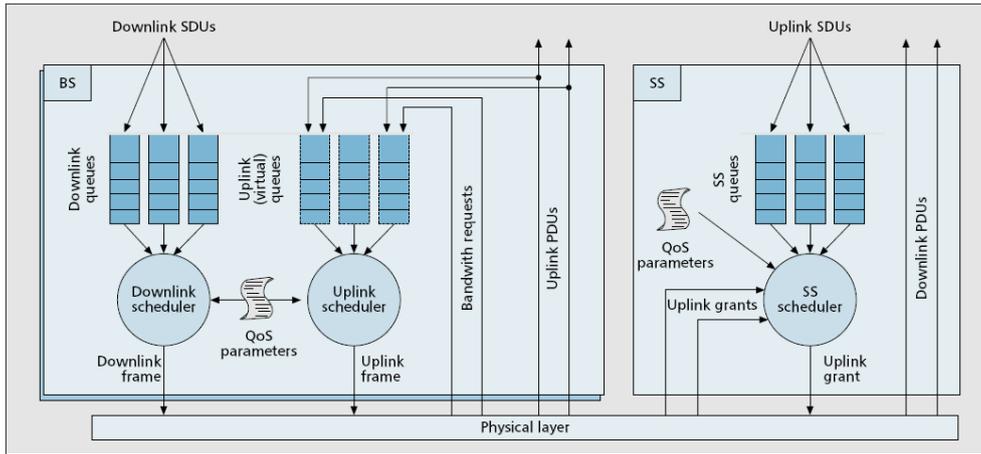


Figura 4.9: Funções dos *schedulers* para garantia de QoS, propostas por Cicconetti *et al.*

Além disso, em [27] são utilizadas apenas duas classes de serviço. Conexões de vídeo conferência, por exemplo, são transmitidas na classe de serviço RTPS, mas as chamadas VOIP também são atendidas nesta classe, ao invés de utilizar a classe eRTPS criada em 2005 pela norma [4].

Os trabalhos de [27], [52], [44] e [37], apesar de não apresentarem novos algoritmos de *scheduling*, propõem uma inovação: a utilização em paralelo de técnicas conhecidas até então.

Em [53], são propostas duas estratégias: o QBAC (do inglês, *QoS Based Admission Control*) e o QBMS (do inglês, *QoS Based Multi-Level Scheduling*). O primeiro é utilizado para propor o módulo de controle de admissão e o último é uma estratégia de *scheduling* baseada na técnica de dupla camada proposta em [52], conforme figura 4.11[53].

A figura 4.10 [53] demonstra o posicionamento dos módulos na arquitetura de QoS proposta, onde é possível perceber que o algoritmo QBMS é executado nas SSs. A estratégia responsável pela a BS, segundo o autor, é a WFQ, já estudada.

O trabalho [53] procura abordar as lacunas do padrão IEEE 802.16 de forma completa, fornecendo estratégias para o escalonamento e para o controle de admissão. Porém, não trata a nova classe de serviço, eRTPS, criada em [4].

Em [8] é proposta a estratégia *Priority-based Fair Scheduling (PBFS)*. A idéia da estratégia é reservar recursos primeiro para conexões RTPS, depois nRTPS e em seguida BE, visto que as conexões do tipo UGS já possuem recursos alocados previamente. Caso existam conexões do mesmo tipo, as mesmas são servidas baseadas em Round-Robin(RR).

Esta idéia já foi explorada em [51], sem apresentar as classes de serviços atuais, utilizando outros parâmetros para definição das prioridades. Em [13], como veremos, esta idéia já havia sido abordada com todas as classes de serviços definidas na norma IEEE 802.16([3] e [4]). O trabalho [8] não apresenta nenhum pseudo-código da estratégia proposta, não faz referência

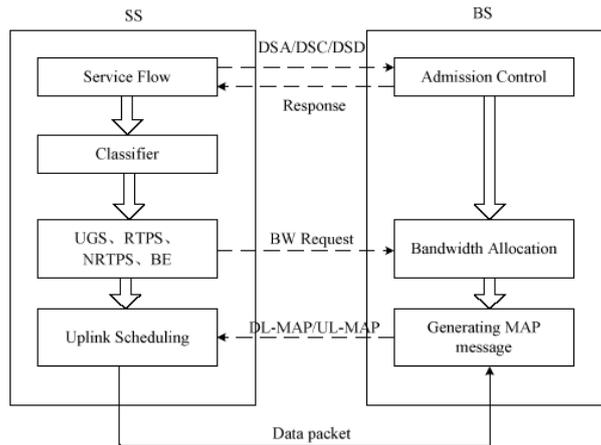


Figura 4.10: Arquitetura de QoS, proposta por Lilei *et al.*

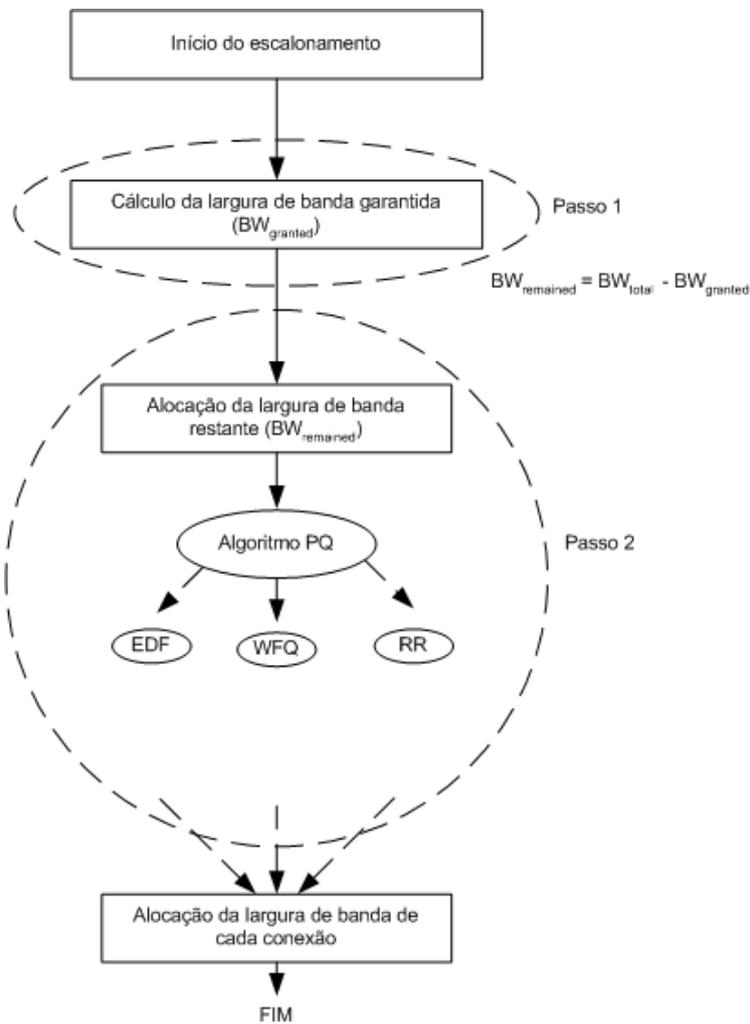


Figura 4.11: Algoritmo QBMS

ao tipo de RR utilizado em caso de conexões do mesmo tipo e também não apresenta qual estratégia é realizada para prevenir *delay* e *jitter* para estas conexões. Além disso, a classe de serviço eRTPS não é abordada.

Em [13], é proposta uma estratégia mais simples para o escalonamento de requisições de *upstream* em uma rede WIMAX. Uma BS deve ser capaz de alocar recursos à razão de 400 *frames* por segundo [13]. O padrão IEEE 802.16 define um tamanho fixo de *slot*, sendo desnecessária a preocupação com o tamanho das requisições, como na técnica RR. Outrossim, é necessário alocar um número inteiro de *slots* em cada *subframe* de (*downstream* e de *upstream*). Técnicas como WFQ e DRR que trabalham com valores fracionários teriam de ser adaptadas a esta realidade.

A estratégia está dividida em três grandes estágios:

- alocação do número mínimo de *slots*, garantindo o QoS;
- alocação de *slots* não utilizados, evitando o desperdício de recurso;
- seleciona a ordem dos *slots* para garantir QoS.

Destes estágios, apenas o primeiro é obrigatório, segundo [3]. No WIMAX, o comportamento natural, com a existência de um tamanho de *frame* fixo, é que *slots* não alocados sejam desperdiçados. O segundo estágio procura evitar este comportamento. O terceiro estágio, por sua vez, é responsável por ordenar os *slots* de modo a garantir parâmetros de QoS como *delay* e *jitter*.

A idéia descrita em [13] procura alocar o menor número de *slots* possível, garantindo o QoS acordado. Para tanto, define algumas grandezas:

- B_i é a largura de banda requerida pela conexão i ;
- S_i é o tamanho do *slot*, ou seja, o número de bits que uma conexão i pode enviar em um *slot*;
- FPS é a taxa de *frames* por segundo que a BS deve processar (no caso 400).

A partir destes parâmetros é calculado o N_i , número mínimo de *slots* em cada frame, por meio da equação 4.5. O cálculo de N_i possibilita o cumprimento do primeiro estágio da estratégia, garantido a largura de banda de acordo com cada classe de serviço prevista em [3] e [4].

$$N_i = \left\lceil \frac{B_i}{S_i \times FPS} \right\rceil \quad (4.5)$$

A idéia embutida na equação 4.5 é: primeiro calcular a quantidade inteira de *slots* (arredondada para cima) necessária para garantir o QoS de cada conexão e em seguida dividir pelo número de *frames* por segundo(FPS), encontrando o número de *slots* por frame. No entanto, para atender a [3] e [4] é necessário levar em consideração as classes de serviços e seus requisitos. É definido o identificador da classe de cada conexão, o C_i . Esta variável terá como conteúdo o tipo de cada conexão: UGS, eRTPS, RTPS, nRTPS ou BE.

A classe UGS não realiza requisições de largura de banda de transmissão e, portanto, não participa do período de contenção. É necessária alocação fixa de *slots* por frame, conforme equação 4.6, baseado no mínimo requerido para este tipo de conexão.

$$N_i^{min} = N_i^{max} = \left\lceil \frac{B_i}{S_i \times FPS} \right\rceil, \forall i | C_i = UGS. \quad (4.6)$$

No caso de conexões do tipo RTPS, será necessária alocação de recursos baseada na requisição de bytes, definida como R_i , e na largura de banda requerida. Caso uma conexão i , do tipo $C_i = RTPS$, não solicite nenhum *slot*, ou seja, $R_i = 0$, será alocado pela estratégia ao menos 1 *slot* para a mesma. As conexões do tipo RTPS não participam do período de contenção, possuem tamanho de pacotes variáveis e, portanto, devem ter a oportunidade (ao menos um *slot*) de solicitar maior taxa de transmissão quando necessário.

Caso R_i seja superior a zero, devem ser calculados o N_i^{min} e o N_i^{max} , de acordo com as equações 4.7 e 4.8, e comparados com a largura de banda máxima (B_i^{max}) e mínima (B_i^{min}) requeridas.

$$N_i^{min} = \begin{cases} 1, & \text{se } R_i=0 \\ \left\lceil \frac{B_i^{min}}{S_i \times FPS} \right\rceil, & \text{se } R_i > 0 \end{cases}, \forall i | C_i = RTPS. \quad (4.7)$$

$$N_i^{max} = \min \left\{ \left\lceil \frac{B_i^{max}}{S_i \times FPS} \right\rceil, \left\lceil \frac{R_i}{S_i} \right\rceil \right\}, \forall i | C_i = RTPS. \quad (4.8)$$

A norma [4] criou a classe de serviço eRTPS, que é uma combinação entre a classes UGS e RTPS, tipicamente usada para aplicações VoIP com supressão de silêncio. Quando a voz está ativa, um *codec* (do inglês *codification decodification*) traduzirá esta informação como uma taxa constante de dados, da mesma forma como ocorre com a classe de tráfego UGS. Por outro lado, esta classe de tráfego é capaz de fazer requisições de largura de banda. Caso R_i seja igual a zero, período de silêncio, será alocado pelo menos um *slot* para que a conexão tenha oportunidade de requisitar recursos quando necessário. As equações 4.9 e 4.10 são utilizadas para alocação de recursos para a classe de serviço eRTPS. No entanto, os valores de N_i^{min} e N_i^{max} não são fixos, ou seja, ou será alocado 1 *slot* ou será alocado um valor fixo para N_i , baseado no requisito de largura de banda para transmissão.

$$N_i^{min} = 1, \text{ se } R_i = 0, \forall i | C_i = eRTPS. \quad (4.9)$$

$$N_i^{max} = \left\lceil \frac{B_i}{S_i \times FPS} \right\rceil, \text{ se } R_i > 0, \forall i | C_i = eRTPS. \quad (4.10)$$

Com relação ao período de contenção, de acordo com [4], as conexões do tipo eRTPS participam da disputa por slots. Em [13], escolhe-se definir que ao menos um *slot* seja alocado para este tipo de conexão. Isto pode parecer correto com o objetivo de evitar possíveis *delays* para este tipo de tráfego, no entanto pode representar um desperdício de recursos importantes, pois a cada frame um *slot* será alocado, representando até 400 *slots* a cada segundo em uma BS WiMAX.

O tipo de conexão nRTPS segue a mesma lógica da classe RTPS. No entanto, existe uma pequena diferença, pois requisições de valor nulo não terão *slots* alocados, conforme equação 4.11. O terceiro estágio da estratégia poderá utilizar este *slot* não desperdiçado, em uma conexão que realmente esteja precisando do mesmo.

Outra diferença é que as conexões da classe nRTPS participam do período de contenção, podendo entrar da disputa por *slots*. Neste tipo de conexão, também é analisado o valor de R_i em relação ao valor de B_i , alocando recursos baseado, no mínimo, entre os dois valores. Desta forma, é possível alocar recursos de maneira mais eficiente, quando a conexão faz uma requisição inferior à largura de banda permitida, como é o caso do tráfego da navegação na WEB. N_i^{min} e N_i^{max} podem ser definidos de acordo com as equações 4.11 e 4.12.

$$N_i^{min} = \left\{ \begin{array}{ll} 0, & \text{se } R_i=0 \\ \left\lceil \frac{B_i}{S_i \times FPS} \right\rceil, & \text{se } R_i>0 \end{array} \right\}, \forall i | C_i = nRTPS. \quad (4.11)$$

$$N_i^{max} = \min \left\{ \left\lceil \frac{B_i^{max}}{S_i \times FPS} \right\rceil, \left\lceil \frac{R_i}{S_i} \right\rceil \right\}, \forall i | C_i = nRTPS. \quad (4.12)$$

Para a classe BE, como o próprio nome já diz, nenhum tipo de reserva é realizada e não existe uma quantidade de largura de banda requerida. Estas conexões participam do período de contenção e, se nada for solicitado, nenhum recurso é alocado. O máximo de *slots* respeitará a requisição realizada (R_i). Os valores de N_i^{min} e N_i^{max} podem ser definidos de acordo com as equações 4.13 e 4.14, respectivamente.

$$N_i^{min} = 0, \text{ se } R_i = 0, \forall i | C_i = BE. \quad (4.13)$$

$$N_i^{max} = \left\lceil \frac{R_i}{S_i} \right\rceil, \forall i | C_i = BE. \quad (4.14)$$

As equações de 4.6 a 4.14 podem ser utilizadas tanto para *uplink* quanto para *downlink*. Por meio das mesmas, é possível calcular o mínimo e o máximo número de *slots* para o *subframe*. A diferença é que, para o tráfego no sentido da BS para as SSs, o valor R_i será

o tamanho de cada fila e não o das requisições. As duas grandezas têm o mesmo sentido físico. Para este propósito, a BS pode alocar filas separadas para conexão ativa com tráfego de *downlink*.

Segundo [13], a BS deve ter a possibilidade de enviar pacotes independentes de cada conexão de acordo com os *slots* alocados pelo *scheduler*. Ainda segundo [13], filas separadas facilitam o processo de aferição do tamanho da fila.

Como visto na seção 4.2.1, [50] propõe utilizar um *hashing* para limitar o número de filas, diminuindo o trabalho a ser realizado pela BS. No entanto, deve-se comparar o benefício descrito acima com o tempo de processamento necessário para controlar um número maior de filas, o que deve ser percebido através *delay/jitter* experimentado pelas conexões.

Concluído o primeiro estágio, a estratégia segue para o segundo: alocação de *slots* sem uso. Após análise das classes de serviços, *slots* ainda livres podem ser utilizados para transmissão de dados das classes RTPS, nRTPS e BE, visto que não é previsto que as classes eRTPS e UGS tenham suas taxas de transmissão majoradas. A equação 4.15 verifica se existem *slots* livres para serem alocados. Apenas para $F > 0$, é possível alocar mais *slots* após o primeiro estágio.

$$F^{free} = F - \sum_i N_i^{min}, \quad (4.15)$$

Após admitida a possibilidade de alocação de *slots* extras, é preciso definir como fazê-la. Supondo um conjunto de conexões A ($|A|$ representando a quantidade de conexões), [13] propõe a alocação com base na ordem decrescente de N_i^{max} e de acordo com as classes de serviço. Primeiro o tipo de conexão RTPS, em seguida nRTPS e por último os *slots* restantes são divididos de forma equânime entre as conexões do tipo BE. Após cada atribuição, são atualizados os valores de N_i e F^{free} , de acordo com as equações 4.17 e 4.18 respectivamente, e retira-se a conexão i do conjunto de conexões A .

$$N_i^{add} = \min \left\{ N_i^{max} - N_i^{min}, \left\lfloor \frac{F^{free}}{|A|} \right\rfloor \right\} \quad (4.16)$$

$$N_i = N_i^{min} + N_i^{add} \quad (4.17)$$

$$F^{free} = F^{free} - N_i^{add} \quad (4.18)$$

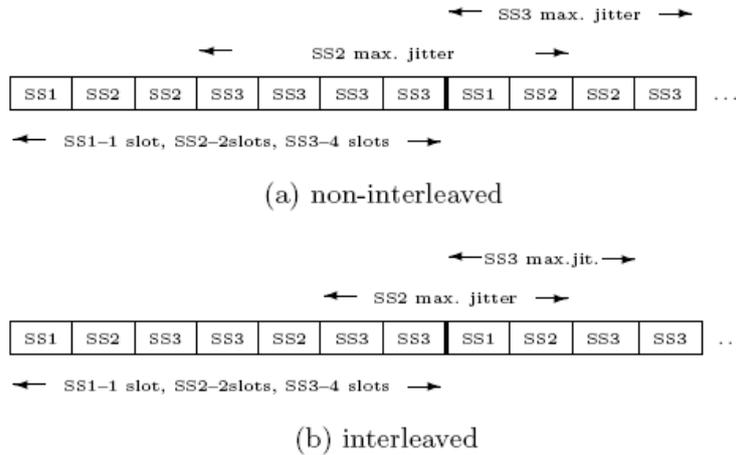


Figura 4.12: Exemplo da diferença de *jitter* com base na ordem de *slots*.

No entanto, a equação 4.16, extraída de [13], não representa a idéia de alocação por ordem de classe de serviço. Para tanto é necessário a consulta do tipo de conexão, ou seja, a variável C_i , após cada atribuição de cada *slot* extra. Só poderiam ser atribuídos *slots* a classe nRTPS após todos os pacotes enfileirados para a classe RTPS, $C_i = RTPS$, terem sido servidos. Ao invés disso, a equação 4.16 adiciona *slots* a uma conexão baseando-se no menor valor entre a quantidade média de *slots* disponíveis por conexão e a diferença de *slot* para atingir o limite de *slots*/por *frame* para aquela conexão.

O terceiro estágio da estratégia tem por objetivo principal a diminuição do *jitter* para as diferentes conexões, de acordo com a qualidade de serviço acordada. A figura 4.12 [13] demonstra o possível melhoramento apenas com a alteração na ordem que os *slots* são enviados, mantendo a taxa de transmissão constante.

A primeira estratégia possível é agrupar *slots* alocados para uma mesma conexão em seqüência. Outra estratégia é interpor *slots* de uma conexão entre outras conexões. Na parte superior da figura 4.12, temos uma alocação de *slot* não-interposta e na parte inferior uma alocação interposta. É possível notar que o *jitter* das SSs 2 e 3 são inferiores para alocação interposta do que não-interposta, permanecendo o mesmo para a SS 1.

O terceiro estágio da estratégia proposta em [13] utiliza os princípios abaixo:

- calcula a melhor opção (menor *jitter* e *delay*) para ordem de *slots*, para todas as conexões;
- começar a alocar slots de acordo com a ordem das classes de serviço mais rigorosas: UGS, eRTPS e RTPS;

- caso algum *slot* desejado já esteja alocado, o *slot* mais próximo é utilizado;
- os espaços restantes são preenchidos por conexões nRTPS e BE.

Esta parte da técnica pode ser considerada uma alteração de camada física e não mais uma estratégia de escalonamento para a camada de enlace.

Apesar dos princípios de funcionamento do algoritmo, nenhum pseudo-código é apresentado ao longo do trabalho de [13]. A estratégia proposta em [13] atende aos serviços de *scheduling*, definidos na última versão da norma IEEE 802.16, incluindo a nova classe eRTPS. A técnica é simulada na plataforma Network Simulator 2(NS-2) [56], utilizando uma camada física em OFDM e a camada de enlace aderente à normatização do WiMAX ([3] e [4]). No entanto, a estratégia proposta em [13] necessita de adaptações para o funcionamento com a camada PUSC OFDMA.

Em [45], apesar de citar o trabalho de [13], é proposta uma nova técnica para resolução do problema. Não é apresentado um pseudo-código do algoritmo, mas seu princípio de funcionamento. A solução proposta não considera as classes eRTPS e BE e tem por objetivo o escalonamento do tráfego nas SSs baseado na quantidade de *time slots* garantida pela BS, conforme equação 4.19:

$$B_{grt}^{tot} = \sum_{j=1}^k (B_j^{UGS}) + \sum_{j=1}^l (B_j^{RT}) + \sum_{j=1}^m (B_j^{nRT}) \quad (4.19)$$

onde k, l e m são os números de fluxos de tráfegos dos tipos UGS, RTPS, nRTPS, respectivamente, e B_j^{UGS} , B_j^{RT} e B_j^{nRT} são as quantidade de *time slots* garantidas para cada conexão.

A BS aloca uma parcela do *subframe* de *uplink* (duração, D_i) de acordo com o tráfego agregado de cada SS_i , indicando-a através da mensagem UL_MAP no *subframe* de *downlink*. Para cada SS, o tamanho do *subframe* de *uplink* ocupado, denominando a largura de banda de transmissão por B_w , em bits, é:

$$F^{tot} = D_i \times B_w \quad (4.20)$$

Cada SS é responsável por escalonar internamente o seu tráfego. Os tráfegos UGS são escalonados de acordo com máxima taxa possível, visto que os mesmos possuem taxa fixa, de acordo com a equação 4.21:

$$F_{UGS} = \sum_{j=1}^k (B_j^{UGS}) \times (C_j^t - C_j^{t-1}) \quad (4.21)$$

onde C_j^t é o tempo atual e C_j^{t-1} é o tempo em que ocorreu a última alocação de banda da conexão j do tipo UGS. O restante do frame, F_{poll} , é alocado para as demais classes de tráfego: RTPS e nRTPS, de acordo com a equação 4.22.

$$F_{poll} = F^{tot} - F_{UGS} \quad (4.22)$$

O tráfego das conexões do tipo RTPS e nRTPS é servido de acordo com o número de MPDUs (do inglês, Medium Access Control Protocol Data Unit) existentes em suas respectivas conexões no início do *subframe* de *uplink* [45]. A técnica proposta em [45] define o parâmetro α como a relação entre o tempo máximo para que os MPDUs das conexões RTPS e nRTPS podem aguardar na fila (*max_mpdu_delay*) e a latência máxima especificada para fluxos de tempo real (*max_latency_of_rtPS_flow*).

$$\alpha = \frac{\text{max_mpdu_delay}}{\text{max_latency_of_rtPS_flow}} \quad (4.23)$$

Considerando N_j^{rt} o número de MPDUs em uma conexão j do tipo RTPS em uma SS, no início do *subframe* de *uplink*, o número total de MBPDUs em todas as conexões do tipo RTPS em cada SS é dado pela equação 4.24:

$$N^{rt} = \sum_{j=1}^l (N_j^{rt}) \quad (4.24)$$

Considerando N_j^{nrt} o número de MPDUs em uma conexão j do tipo nRTPS em uma SS, no início do *subframe* de *uplink*, o número total de MBPDUs em todas as conexões do tipo nRTPS em cada SS é dado pela equação 4.25:

$$N^{nrt} = \sum_{j=1}^m (N_j^{nrt}) \quad (4.25)$$

De acordo com a estratégia, para as conexões do tipo RTPS é alocado um tempo de frame total, conforme a equação 4.26:

$$F_{tot}^{rt} = F_{poll} \times \left(\frac{N^{rt} \times \alpha}{N^{rt} \times \alpha + N^{nrt}} \right) \quad (4.26)$$

e para nRTPS, um tempo de frame total, conforme a equação 4.27:

$$F_{tot}^{nrt} = F_{poll} \times \left(\frac{N^{nrt}}{N^{rt} \times \alpha + N^{nrt}} \right) = F_{poll} - F_{tot}^{rt} \quad (4.27)$$

O parâmetro *alpha* torna o escalonamento dependente da largura de banda e da latência de cada de tipo de conexão. Para cada fluxo de tempo real, $v \in \{1, 2, \dots, l\}$, é alocado um tempo de *subframe* de acordo com a equação 4.28:

$$F_v^{rt} = F_{tot}^{rt} \times \frac{N_v^{rt}}{\sum_{j=1}^l (N_j^{rt})} \quad (4.28)$$

e cada fluxo não de tempo real, $w \in \{1, 2, \dots, m\}$, é alocado de acordo com a equação 4.29:

$$F_w^{nrt} = F_{tot}^{nrt} \times \frac{N_w^{nrt}}{\sum_{j=1}^m (N_j^{nrt})} \quad (4.29)$$

Este processo ocorre em todas as SSs da rede em todo início de *subframe* de *uplink*. A vantagem desta estratégia é que a alocação de recurso para conexões de tempo real é baseada na largura de banda necessária para transmissão, no *delay* tolerável e no tamanho da fila, de forma que o *delay* médio possa ser diminuído [45].

No entanto, o algoritmo proposto não poderia ser executado nas BSs. Estratégias que dependem do conhecimento do tamanho da fila de cada tipo de conexão não são aptas para o escalonamento de tráfego nas BSs [27]. Em [45], não é apresentado nenhuma técnica para o escalonamento das requisições nas BSs.

O problema de escalonamento de requisições nas BSs possui maior complexidade do que para as SSs, pois a quantidade de conexões e o volume de tráfego são superiores, o tempo para execução do algoritmo é menor, é necessário alocação de recursos de *polling* e uma decisão incorreta prejudicará todos os usuários vinculados aquela BS.

Em trabalhos posteriores, os co-autores de [13] dedicam-se a temas análogos, mas não propõem uma solução para este problema específico ([57], [58] e [59]). Em [57], realiza-se um estudo do *multicast polling*, funcionalidade opcional para uma rede WiMAX, especificamente para o tráfego VoIP. Em [58], é realizada uma avaliação da performance do mecanismo

ARQ. Em [59], o mecanismo H-ARQ também é avaliado como continuação ao trabalho de [58]. Os mecanismos ARQ e H-ARQ são funcionalidades opcionais para a norma [4].

O artigo de [60] cita o trabalho de [13] como uma estratégia que satisfaz o requerimento de largura de banda, mas para um camada física em OFDM no modo PMP. Em [60] é proposta uma técnica de escalonamento para redes WiMAX, operando em modo *Mesh*, permitindo múltiplos saltos entre as SSs.

Em [61], é proposta uma estratégia de escalonamento baseada em WFQ com duas camadas e programação dinâmica, apesar da referência à estratégia de [13]. Um dos princípios fundamentais para uma técnica de escalonamento para o WiMAX é a simplicidade, pois a mesma deverá ser capaz de tomar centenas de decisões por segundo, alocando os recursos frame por frame. No trabalho [61], foi utilizado o módulo proposto em [14], que também será utilizado nesta dissertação, mas não é esclarecido como a estratégia fornece o suporte à camada física em OFDMA PUSC, ou seja, como é resolvido o problema identificado em [13]. Além disso, o serviço de escalonamento eRTPS não é considerado por [61], estando em desacordo com a norma [4].

Encontramos apenas dois trabalhos sobre estratégias de escalonamento, ambos publicados em 2008, com suporte a uma camada física em OFDMA PUSC: [62] e [63]. No entanto, em ambos, a estratégia de escalonamento é utilizada apenas para o tráfego de *downlink*. Estratégias de escalonamento compatíveis com a camada física em OFDMA permitem o fornecimento de serviços móveis aos usuários da rede WiMAX [27]. Esta dissertação é o único trabalho de que temos conhecimento que simula uma estratégia escalonamento, para camada de enlace, adaptada a uma camada física OFDMA PUSC para o tráfego de *uplink*.

Alguns trabalhos dedicaram-se a realizar comparações da performance das estratégias de escalonamento existentes. Algumas das estratégias estudadas nesta dissertação foram comparadas nos trabalhos de [41] (estratégias [37], [64], [46], [48]-[50]) e [65] (trabalhos [49], [50] e [54]). Optou-se nesta dissertação em encontrar uma solução de escalonamento de UL da BS com suporte à camada física em OFDMA, baseada apenas em parâmetros de camada MAC. A proposta é uma adaptação da estratégia apresentada em [13], conforme descrito no capítulo 5.

Capítulo 5

Algoritmo Proposto, Simulações e Resultados

5.1 Introdução

No capítulo 4, foram analisadas diversas estratégias de escalonamento. Percebemos que a estratégia de [13] atende a serviços de *scheduling* definidos na última versão da norma IEEE 802.16, incluindo a nova classe eRTPS. A técnica é simulada na plataforma Network Simulator 2(NS-2) [56], utilizando a camada física em OFDM e a camada de enlace aderente à normatização do WiMAX.

Em suas conclusões, [13] cita que a estratégia é válida para as camadas física SC, OFDM, OFDMA-FUSC, mas que são necessárias adaptações para o funcionamento em OFDMA-PUSC. Isto inviabiliza o uso da técnica para utilização da estratégia de escalonamento para o WiMAX, pois, conforme visto no capítulo 3, o PUSC é obrigatório para o UL. Um dos objetivos deste trabalho foi a adaptação da estratégia proposta em [13] para permitir o funcionamento com uma camada física em OFDMA-PUSC. No restante deste capítulo, descreveremos a estratégia proposta, a escolha do simulador utilizado e as adaptações realizadas e os resultados obtidos.

5.2 Adaptação da estratégia de escalonamento

A estratégia proposta em [13] é baseada na definição de um número de *slots* que garanta taxa de transmissão adequada para cada tipo de serviço de *scheduling*, para uma camada PHY OFDM. Como vimos neste tipo de especificação de camada física, todas as subportadoras são alocadas para o transmissor envolvido. Utilizando-se a camada física em OFDMA-FUSC, cada *slot* é mapeado em um subcanal por um símbolo OFDMA e todos os subcanais são alocados para o transmissor (páginas 356 e 357 de [4]). Desta forma, a téc-

nica proposta pode ser aplicada diretamente ao OFDMA-FUSC. No entanto, são necessárias adaptações para utilização da camada física OFDMA-PUSC, conforme comentário do autor nas possibilidades de trabalhos futuros de [13].

Este trabalho não tem o objetivo de propor algoritmo para tratamento diretamente de subportadoras, mas da associação dos subcanais lógicos e símbolos OFDMA às SSs. Também serão necessárias adaptações na camada MAC do módulo desenvolvido em [14] para funcionamento de acordo com a estratégia proposta nesta dissertação, adaptada do trabalho de [13]. Estas adaptações serão explicadas na seção 5.4.

Uma estratégia de escalonamento, para suportar a camada física em OFDMA PUSC, deve permitir a alocação de símbolos por subcanais lógicos (domínio de tempo e frequência, respectivamente), visto que os mesmos são as unidades básicas para reserva de recursos das mensagens UL-MAP ([3] página 153). Em [63], é proposta uma técnica de escalonamento para uma camada física em OFDMA que utiliza o número de subcanais para reserva de recursos, baseada em parâmetros físicos, como a relação sinal ruído, mas apenas para o tráfego de DL.

Neste trabalho, realizamos uma adaptação na equação geral da estratégia vista em [13], para encontrar não o número de *slots*, mas o número de subcanais (NS_i) por *subframe*. As novas variáveis TS e Sb_i referem-se, respectivamente, ao número de símbolos OFDMA por *subframe* e ao número de *bytes* que a conexão i poderá enviar em um símbolo OFDMA, conforme equação 5.1. A variável B_i representa os parâmetros de qualidade MSTR e MRTR, enquanto R_i representa a quantidade de bytes solicitada pela estação i para o *subframe*.

$$NS_i = \left\lceil \frac{B_i}{Sb_i \times FPS \times TS} \right\rceil \quad (5.1)$$

As demais equações foram adaptadas seguindo o mesmo princípio. Desta forma, a equação 5.2 encontra um número fixo de subcanais necessários para este tipo de conexão.

$$NS_i^{min} = NS_i^{max} = \left\lceil \frac{B_i}{Sb_i \times FPS \times TS} \right\rceil, \quad \forall i | C_i = UGS. \quad (5.2)$$

As equações 4.7 e 4.8, referentes ao serviço de escalonamento do tipo RTPS, foram adaptadas para encontrar o número mínimo e máximo de subcanais necessários, resultando nas equações 5.3 e 5.4 respectivamente.

$$NS_i^{min} = \left\{ \begin{array}{ll} 1, & \text{se } R_i=0 \\ \left\lceil \frac{B_i^{min}}{Sb_i \times FPS \times TS} \right\rceil, & \text{se } R_i > 0 \end{array} \right\}, \quad \forall i | C_i = RTPS. \quad (5.3)$$

$$NS_i^{max} = \min \left\{ \left\lceil \frac{B_i^{max}}{Sb_i \times FPS \times TS} \right\rceil, \left\lceil \frac{R_i}{Sb_i \times TS} \right\rceil \right\}, \forall i | C_i = RTPS. \quad (5.4)$$

As equações 4.9 e 4.10, referentes ao tráfego da classe eRTPS, também foram adaptadas para encontrar o número mínimo e máximo de subcanais necessários, resultando nas equações 5.5 e 5.6, respectivamente. Como as estações do tipo eRTPS também podem participar da disputa por *slots* [4], optamos para economizar os recursos em não alocar nenhum subcanal para estação quando não existir requisição, diferentemente do procedimento adotado em [13].

$$NS_i^{min} = 0, \text{ se } R_i = 0, \forall i | C_i = eRTPS. \quad (5.5)$$

$$NS_i^{max} = \left\lceil \frac{B_i}{Sb_i \times FPS \times TS} \right\rceil, \text{ se } R_i > 0, \forall i | C_i = eRTPS. \quad (5.6)$$

As equações 4.11 e 4.12, referentes ao serviço de escalonamento do tipo nRTPS, também sofreram alterações para encontrar o número mínimo e máximo de subcanais necessários, resultando nas equações, 5.7 e 5.8, respectivamente.

$$NS_i^{min} = \begin{cases} 0, & \text{se } R_i = 0 \\ \left\lceil \frac{B_i}{Sb_i \times FPS \times TS} \right\rceil, & \text{se } R_i > 0 \end{cases}, \forall i | C_i = nRTPS. \quad (5.7)$$

$$NS_i^{max} = \min \left\{ \left\lceil \frac{B_i^{max}}{Sb_i \times FPS \times TS} \right\rceil, \left\lceil \frac{R_i}{Sb_i \times TS} \right\rceil \right\}, \forall i | C_i = nRTPS. \quad (5.8)$$

As equações 4.13 e 4.14, referentes ao tráfego da classe BE, também foram adaptadas para encontrar o número mínimo e máximo de subcanais necessários, resultando nas equações 5.9 e 5.10, respectivamente.

$$NS_i^{min} = 0, \text{ se } R_i = 0, \forall i | C_i = BE. \quad (5.9)$$

$$NS_i^{max} = \left\lceil \frac{R_i}{Sb_i \times TS} \right\rceil, \forall i | C_i = BE. \quad (5.10)$$

Com estas adaptações, é possível a alocação de subcanais às estações que utilizam a rede, respeitando parâmetros de qualidade de serviço.

5.3 Escolha do Simulador

Até a data do início das implementações, estavam disponíveis, na literatura, três módulos em plataforma aberta para simulação de redes WiMAX: o módulo criado pelo NIST (*National Institute of Standards and Technology*) [66], uma produção nacional da UNICAMP

[67] e o trabalho desenvolvido pelo NDSL(*Networks & Distributed Systems Laboratory*) [14]. Todos produziram módulos que são inseridos na estrutura do NS-2. A opção comercial existente seria o módulo 802.16 do simulador da OPNET [68]. Esta opção foi descartada em razão dos custos envolvidos e do tempo necessário para realização da compra.

Apesar da preferência do autor em utilizar o módulo nacional, era necessário uma análise das características de cada software, para verificar se os mesmos atendiam aos requisitos necessários. Os simuladores foram analisados a partir de quatro critérios: documentação, aceitação pela comunidade acadêmica, aderência à normatização vigente e, principalmente, a compatibilidade com o trabalho a ser realizado.

Nenhuma das opções apresenta uma documentação com o detalhamento do módulo, mas apenas uma descrição do relacionamento entre as classes e funções. Com relação à aceitação da comunidade acadêmica, todos os módulos possuem artigos publicados em congressos bem conceituados.

No tocante a normatização, todos módulos apresentam suficiente aderência à camada MAC do padrão IEEE 802.16e. No entanto, apenas os módulos [66] e [14] apresentam estratégias de escalonamento com compatibilidade com a camada PHY do padrão citado. E como último aspecto da análise, o módulo desenvolvido em [14] é o único que possui a camada MAC adaptada ao funcionamento sobre uma camada física em OFDMA-PUSC, permitindo a implementação adaptada da estratégia proposta em [13]. A implementação da camada MAC, adaptada a uma camada PHY em OFDMA, é objetivo futuro dos trabalhos de [66] e [67]. Por isto, escolhemos o módulo desenvolvido pelo NDSL para realização do trabalho.

5.4 Adaptação no módulo WiMAX do NDSL

Diversas adaptações foram realizadas para que o módulo WiMAX do NDSL pudesse ser utilizado para simular o algoritmo com as adaptações descritas na seção 5.2. Nesta seção, descreve-se as principais mudanças realizadas no módulo, com o objetivo de simular o funcionamento de uma rede IEEE 802.16, com a estratégia de escalonamento proposta.

Como a documentação sobre o módulo é praticamente inexistente, excetuando-se o artigo publicado ([14]), um grande esforço foi despendido para compreensão do módulo e de suas funções. Apesar do NS-2 permitir a implementação de funções através da linguagem TCL/TK, o módulo é implementado, em sua totalidade, em C++, com intensivo uso de ponteiros, o que dificulta bastante o entendimento inicial do *software*.

Não sabemos se propositadamente, ou não, diversos parâmetros estavam configurados de forma diferente do artigo. Para citar algumas das incorreções, identificamos que na função "*Priority()*", responsável pelo repasse dos pacotes das camadas superiores para a camada MAC, definia conexões BE e eRTPS com a mesma prioridade, embora, no artigo,

a última estivesse definida com maior prioridade; além disso, todas as requisições da classe BE estavam sendo multiplicadas por 10 ao serem recebidas na BS. Corrigimos estes e outros erros do módulo. Em seguida, passou-se para adaptação da função de escalonamento.

A idéia inicial era a de substituir o algoritmo de *scheduling*, o WRR, pelo algoritmo adaptado de [13]. No entanto, percebemos que a função "*Scheduler()*", existente no módulo, era responsável apenas pelo escalonamento de cada SS, tratando suas filas (dos cinco tipos de tráfego) sem verificar o total de conexões de *uplink*. De fato, foi necessário construir esta parte do módulo.

Construímos uma função com capacidade para verificar a quantidade de subcanais disponíveis em cada *frame* que, comparando as requisições das SSs e os parâmetros de QoS, realiza a alocação dos recursos de acordo com as equações adaptadas. O módulo original não previa nenhuma garantia de parâmetros de QoS, sendo necessária a inclusão da comparação entre as requisições das SSs e com a largura de banda requerida para cada conexão, conforme equações 5.2 a 5.10, considerando cada tipo de tráfego. A maneira com que o módulo associava subcanais lógicos às SSs foi alterada para funcionar de acordo com as equações mencionadas. Em [13], nenhum pseudo-código é apresentado. O pseudo-código abaixo apresenta o funcionamento do algoritmo de escalonamento.

Algoritmo de Escalonamento

```
1: Início
2: N=número de conexões ativas;
3: NS_total=número de subcanais de UL;
4: verifica garantias periódicas e de
   requisições solicitadas;
5: Enquanto (N > 0)
6:   verificar o tipo de conexão;
7:   verificar SFID associado;
8:   N = N-1;
9: fim enquanto
10: Para i=1 até n_UGS
11:   aplica-se a equação 5.2
     => NSi_UGS;
12:   NS_total = NS_total - NSi_UGS;
13: Para i=1 até n_eRTPS
14:   aplicam-se as equações 5.5 e 5.6
     => NSi_eRTPS;
15:   NS_total = NS_total - NSi_eRTPS;
16: Para i=1 até n_RTPS
17:   aplicam-se as equações 5.3 e 5.4
```

```

=> NSi_RTPS;
18: NS_total = NS_total - NSi_RTPS;
19: Para i=1 até n_nRTPS
20:   aplicam-se as equações 5.7 e 5.8
   => NSi_nRTPS;
21:   NS_total = NS_total - NSi_nRTPS;
22: Para i=1 até n_BE
23:   aplicam-se as equações 5.9 e 5.10
   => NSi_BE;
24:   NS_total = NS_total - NSi_BE;
25: Alocação de subcanais e símbolos
   OFDMA nas mensagens UL-MAP;
26: Fim

```

Para permitir a comparação com o trabalho de [13], todos os geradores de tráfego foram adaptados seguindo os parâmetros do referido artigo. Na próxima seção, detalharemos os geradores de tráfego para cada cenário de simulação.

5.5 Ambiente de simulação

Existem alguns parâmetros do módulo que não são passíveis de alteração e serão utilizados durante todos os cenários, como por exemplo, o tamanho do *frame* e o tempo de um símbolo OFDMA. Estes valores estão apresentados na tabela 5.1. O tamanho do *frame*, o mesmo utilizado em [35], documento do WiMax Forum, bem como nos recentes trabalhos de [63] e [62], permanece constante durante todas as simulações, mas a taxa de DL/UL pode ser alterada entre os cenários.

Em [13], as especificações do sistema não são explicitadas. Os únicos valores apresentados são: o número de *slots* por *frame*, o número de *frames* por segundo que a BS é capaz de tratar e a largura de banda do canal, conforme tabela 5.2.

Outros parâmetros são fundamentais para o funcionamento da simulação como a modulação dos sinais, a codificação do canal e a quantidade de *bytes* atribuída a um *slot*. As tabelas 5.3 e 5.4 apresentam, respectivamente, os parâmetros utilizados em [13] e os utilizados nesta dissertação. Estes valores serão mantidos fixos durante a simulação de todos os cenários.

Ressaltamos que três símbolos OFDMA, por subcanal, equivalem a um *slot* no UPLNK OFDMA PUSC. Desta forma, um *slot* de *uplink*, utilizando modulação 64 QAM e codificação 3/4, poderá enviar até 81 bytes de dados. Portanto, o módulo utilizado é menos

eficiente, no tocante ao número de bits/slot, do que o simulador utilizado em [13]. Porém, ambos estão corretos, pois atendem às especificações das camadas físicas respectivas, como visto no capítulo 3.

5.6 Simulações e Análise dos resultados

Para verificação da correção da nova técnica, montamos os mesmos cenários simulados em [13]. Vale ressaltar que, para todos os cenários, são utilizados os parâmetros das tabelas 5.1 e 5.4. Além do algoritmo de escalonamento, o módulo é capaz de enviar as mensagens obrigatórias do padrão WiMAX como UL-MAP, DL-MAP, UCD, DCD, *ranging*, *registration*, entre outras. Não são simulados erros na camada física, sendo a perda de pacote, porventura existente, responsabilidade da camada MAC. Os três cenários têm os seguintes objetivos: verificar a divisão justa de recursos para o mesmo serviço de *scheduling*, a garantia de recursos para serviço de qualidades distintas e, por último, a garantia de recursos com a alteração dos esquemas de modulação dos sinais.

Em todos os cenários, as estações disparam contra um nó destinatário, passando pela BS. A conexão entre a BS e o nó destinatário foi escolhida de modo a não enfileirar pacotes, permanecendo a BS como o gargalo da comunicação. A topologia de todos os cenários é de 1000m X 1000m. Em seguida apresentaremos os detalhes de cada cenário.

5.6.1 Cenário 1 - Escalonamento de estações *Best Effort*

A estrutura da rede do cenário 1 é montada distribuindo-se 10 estações do tipo BE, utilizando modulação 64 QAM 3/4, uniformemente ao redor da BS, conforme figura 5.1.

Tabela 5.1: Parâmetros fixos do módulo

Parâmetro	Valor
Número de símbolo de dados por frame	48
Tempo do símbolo OFDMA	100.8 μ s
Tempo de um frame OFDMA	5 ms
Largura de banda	20 MHz
Período de mensagens DCD/UCD	10 s
Intervalo de entrada na rede	1210.08 μ s
Intervalo de requisição de banda	1210.08 μ s
TTG	200 μ s
RTG	200 μ s

Tabela 5.2: Parâmetros de Simulador utilizado em Sayenko *et al.*

Parâmetro	Valor
Número de <i>frames</i> por segundo	400
Número de <i>slots</i> por <i>frame</i>	80
Largura de banda do canal	7MHz

Tabela 5.3: Modulação, Codificação e Tamanho de *Slot*

Modulação	Codificação	<i>Bytes</i> por <i>slot</i>
64 QAM	3/4	108
64 QAM	2/3	96
16 QAM	3/4	72
16 QAM	1/2	48
QPSK	3/4	36
QPSK	1/2	24

Tabela 5.4: Modulação, Codificação e Tamanho de Símbolo OFDMA, utilizados neste trabalho

Modulação	Codificação	<i>Bytes</i> por símbolo
64 QAM	3/4	27
64 QAM	2/3	24
16 QAM	3/4	18
16 QAM	1/2	12
QPSK	3/4	9
QPSK	1/2	6

Tabela 5.5: Controle de ativação do tráfego

SS	0-5s	5-10s	10-15s	15-20s
SS1 - SS3	X	X	X	
SS4 - SS5		X	X	
SS6 - SS8			X	
SS9 - SS10			X	X

Conectado à BS, por um canal de 1Gbps, com atraso de 2ms, existe o nó destinatário do tráfego de *uplink* de todos as SSs. Cada SS mantém duas conexões com a BS, uma de DL e uma de UL. O tráfego BE é simulado por um serviço FTP ativo em cada estação. Na camada de transporte, é utilizado o protocolo TCP, com tamanhos de pacote de 1500 *bytes*. O protocolo TCP procura atingir a maior taxa de transmissão possível, no caso de 24 Mbps, utilizando uma razão de DL:UL de 3:1. Utilizamos a taxa de 3:1, pois o total de taxa de transmissão com esta configuração é apenas um pouco maior do que a taxa máxima encontrada em [13], sendo possível encontrar taxas individuais também próximas. Para realizar o estudo da alocação, um serviço FTP é ativado em cada estação, em diferentes períodos de tempo, de acordo com a tabela 5.5.

O objetivo deste cenário é verificar se a estratégia de escalonamento é capaz de alocar os recursos de forma justa entre as estações de mesmas características, neste caso, utilizando a classe BE. É importante ressaltar que todas as estações realizam a entrada na rede no início da simulação, mas só recebem um subcanal lógico a partir da comunicação de requisição de banda. Neste momento, o algoritmo verifica quantas estações estão ativas e divide os subcanais entre as mesmas.

Em [13], o cenário 1 possui 15 estações do tipo BE e encontra os resultados mostrados na figura 5.2. Por ter 10 estações, a simulação do cenário 1, desta dissertação, possui menor número de estações concorrentes no intervalo de 10 a 15 s, mas possui a mesma quantidade de estações nos demais intervalos. Na figura 5.3, apresentamos o resultado encontrado. A figura 5.3 contém a taxa de transmissão de cada SS para a BS e a soma do tráfego na saída para o nó destinatário em função do tempo. É possível verificar que o algoritmo proposto divide de forma justa os recursos da BS, não obstante a alteração no número de estações.

Discordamos do argumento apresentado em [13] para o declínio da taxa total na saída da BS no intervalo entre 10 e 15s, associando o fato ao aumento no tamanho das mensagens DL-MAP e UL-MAP, em virtude do acréscimo no número das SSs. De fato, as mensagens aumentam, mas as mesmas são enviadas no *subframe* de DL e não interferem no *subframe* de UL, visto a taxa fixa de DL:UL de 1:1. No nosso entendimento, a queda da taxa total observada em menor escala no intervalo entre 5 e 10s e mais acentuada entre 10 e 15s, se deve ao fato da divisão inexata do *subframe* de UL entre as estações, o que não ocorre nos

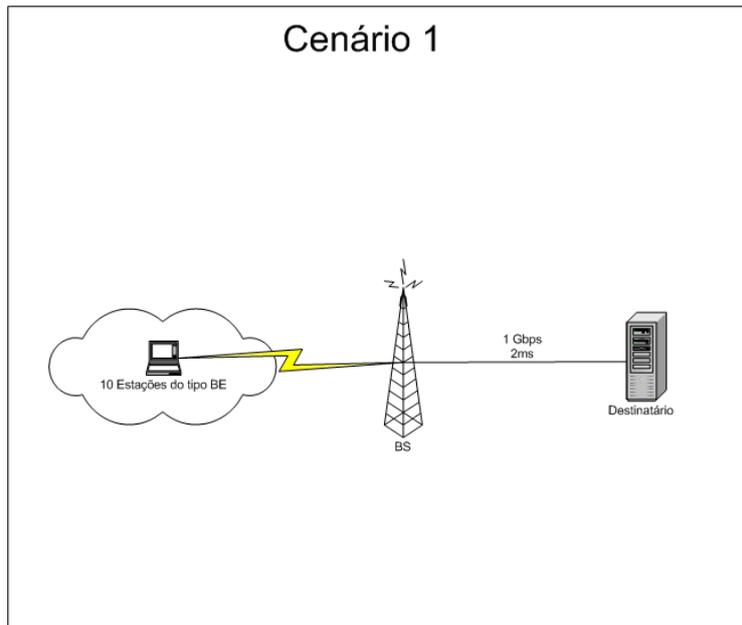


Figura 5.1: Estrutura da rede para o cenário 1

demais intervalos.

5.6.2 Cenário 2 - Escalonamento de SSs com diferentes classes de serviço

O objetivo do segundo cenário é verificar se a estratégia de escalonamento é capaz de diferenciar as classes de serviço e garantir os parâmetros de qualidade de cada uma delas. Para isso, são simulados sete fluxos de tráfegos de *uplink* de classes distintas, conforme tabela 5.6. A estrutura da rede é mostrada na figura 5.4. Cada SS possui apenas um serviço ativo e envia dados para um nó de destino, através da BS de acordo com o tipo de tráfego associado, utilizando codificação 64 QAM 3/4. O nó de destino é conectado à BS através de uma conexão de 1 Gbps e *delay* de 2ms, de forma que o gargalo do contexto seja a rede WiMAX e não a rede cabeada.

Com exceção das SS5 e SS6, os demais dados da tabela são idênticos às configurações do segundo cenário simulado em [13]. Em [13], são definidos valores diferentes para duas estações do mesmo tipo (banda mínima 3,5 Mbps e banda máxima 5,2 Mbps para a SS5 e banda mínima 3,5 Mbps e banda máxima 5,2 Mbps para o SS6). Consideramos os valores maiores (de mínimo e de máximo) para ambas as SSs, visto que trata-se do mesmo tipo de serviço de escalonamento, como no caso das SSs 3 e 4.

Os geradores de tráfego do módulo do NDSL foram adaptados para atender ao cenário proposto. Para simular a classe UGS, foi utilizada uma fonte de taxa de dados constante (CBR), de 4 Mbps, com tamanho de pacotes de 500 bytes e enviados através de UDP. A classe eRTPS é simulada utilizando uma fonte exponencial ON/OFF seguindo as especi-

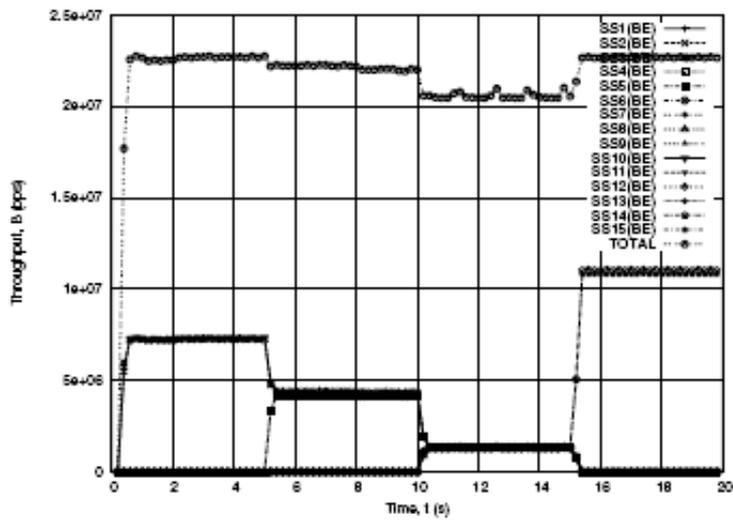


Figura 5.2: Gráfico da Taxa de transmissão X tempo, para o Cenário 1 em Sayenko *et al.*

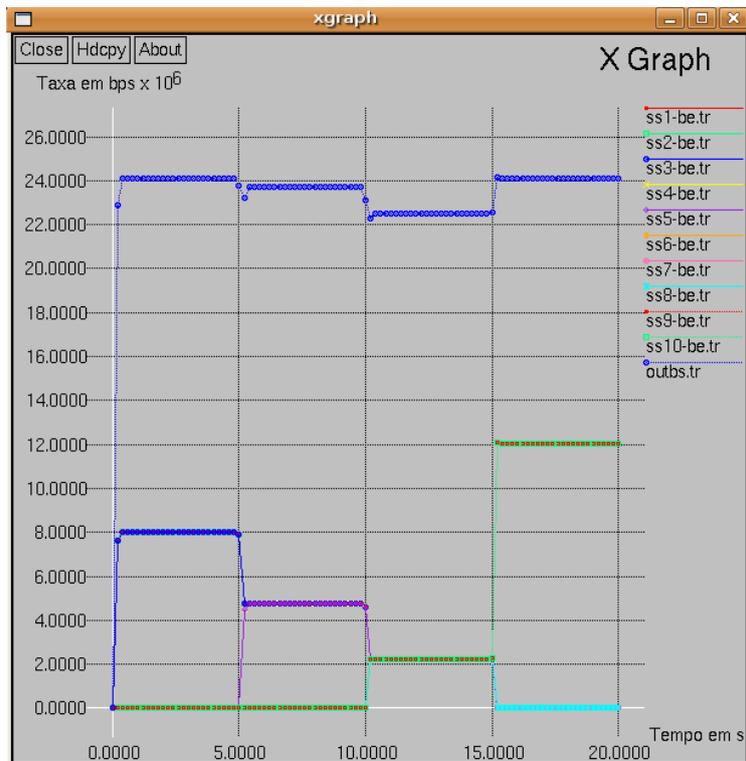


Figura 5.3: Gráfico da Taxa de transmissão X tempo, deste trabalho, para o Cenário 1

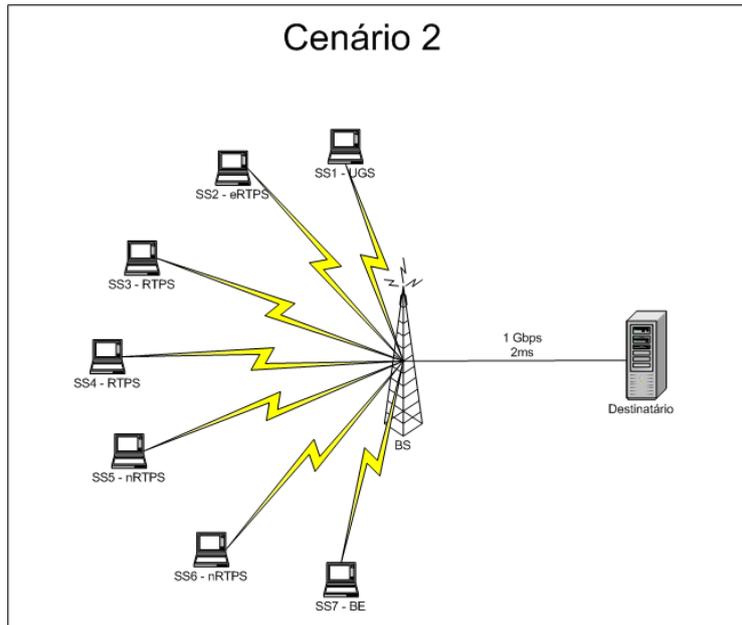


Figura 5.4: Estrutura da rede para o cenário 2

Tabela 5.6: Tipos de tráfego e Parâmetros de QoS para o cenário 2

SS	Tipo de tráfego	Tamanho do pacote	Banda mínima (bps)	Banda máxima (bps)
SS1	UGS	500	4.000.000	4.000.000
SS2	eRTPS	300	80.000	80.000
SS3	RTPS	1378	2.054.400	4.108.800
SS4	RTPS	1378	2.054.400	4.108.800
SS5	nRTPS	1060	3.524.000	5.200.000
SS6	nRTPS	1060	3.524.000	5.200.000
SS7	BE	200	-	-

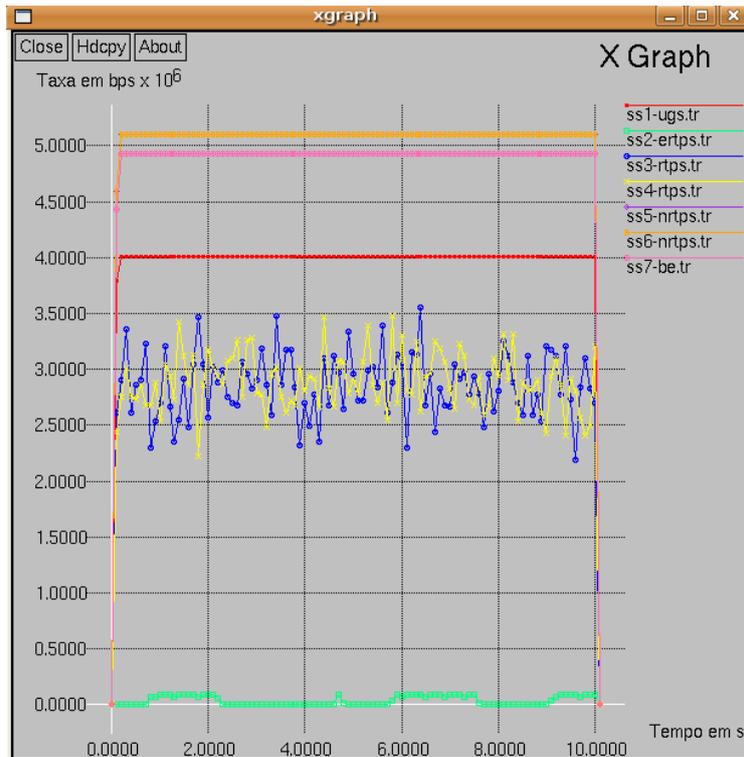


Figura 5.5: Resultado do cenário 2 - DL/UL - 1:1

ficações da recomendação P.59 do ITU-T [69], com períodos ativos (*talk spurt*) e de supressão de silêncio. A fase ativa segue os parâmetros da recomendação ITU G.711 e ocorre a uma taxa de 80Kbps com pacotes de 300 bytes, incluindo os tamanhos dos cabeçalhos IP/UDP/RTP. A classe de tráfego RTPS é simulada por um tráfego de taxa de dados variável (VBR), considerando *trace files* de vídeos reais contendo áudio e vídeo [70], adaptados para o formato aceito pelo NS-2, encapsulados pelos protocolos IP/UDP. As conexões nRTPS são simuladas por uma fonte FTP através do protocolo TCP. Esta conexão vai aumentando a taxa de transmissão até o máximo permitido pela rede, no caso 5,2 Mbps. A estação do tipo BE é simulada utilizando uma fonte CBR e protocolo UDP e tenta atingir uma taxa de 5 Mbps. Para os tráfegos dos tipos UGS e eRTPS, para garantia de uma taxa constante de determinado valor, é necessário considerar os bytes de *overhead* de cabeçalho dos protocolos, bem como o tamanho do PDU e dos cabeçalhos para a correta alocação dos recursos.

Um parâmetro importante da simulação é a taxa de DL/UL de cada *frame*. Em [13], está taxa é de 1:1. Para realizar a comparação configuramos a mesma taxa. Neste cenário, todas as estações transmitem dados para a BS do instante inicial até o fim da simulação. Obtivemos o resultado mostrado na figura 5.5. Os resultados obtidos em [13], para o mesmo cenário, são apresentados na figura 5.6.

Analisando a figura 5.5, é possível perceber que a estação SS1, com tráfego UGS, possui taxa de transmissão constante de 4 Mbps. A estação SS2 eRTPS possui taxa de transmissão

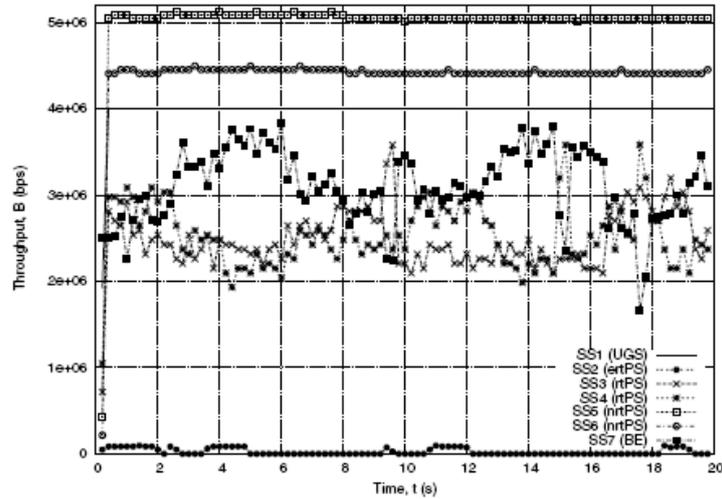


Figura 5.6: Resultado do cenário 2 de Sayenko *et al.*

de 80 Kbps, garantida na sua fase ativa. Os usuários SS 3 e 4 possuem o mínimo de taxa de transmissão acima de 2.054.000 bps e não ultrapassam 4.108.800 bps. As estações SS 5 e 6 atingem a taxa máxima de 5,2 Mbps. A SS7 alcança 4,9 Mbps, em virtude do *overhead*, que não é computado para a tráfegos do tipo BE. Podemos observar que a técnica aloca recursos de maneira a garantir o QoS dos usuários para todos os tipos de tráfego e que os recursos não se esgotam. Na figura 5.6, percebemos que apesar de garantir o QoS das estações de 1 a 6, a SS7 não consegue atingir a taxa de 5 Mbps, ou próximo disso, em virtude da escassez dos recursos.

Na prática, as operadoras de telecomunicações e tv a cabo que terão licença para prestação de serviços utilizando a tecnologia WiMAX, provavelmente definirão outras taxas entre os *frames* de DL e UL, visto que o tráfego no sentido da BS para as SSs será superior. Por isso, decidiu-se simular o comportamento da rede com uma relação DL/UL de 3:1. Os resultados obtidos são apresentados na figura 5.7.

Observando a figura 5.7, podemos notar que, com a redução do número de símbolos OFDMA disponíveis para o *uplink*, a rede passa a experimentar a escassez de recursos. Porém, todos os tipos de serviços que requerem determinada taxa de transmissão são atendidos. Mas a estação SS7 não mais atinge a taxa de 4,9 Mbps. A taxa útil passa a depender do quanto as estações RTPS (SS 3 e 4) utilizam da rede, já que as demais possuem taxa fixa. Portanto, mesmo em uma condição mais próxima da realidade, o algoritmo se comporta de maneira aceitável.

A segunda parte do cenário 2 refere-se à medição do *jitter* da estação SS1. Em [13], são encontrados valores pouco acima de 1 ms, conforme figura 5.8, utilizando o processo para alteração da ordem em que os *slots* são transmitidos. Na letra (a), os *slots* são ordenados consecutivamente: os *slots* para estação SS1, em seguida para estação SS2 e assim por diante, atingindo um *jitter* de, aproximadamente, 2ms. Na letra (b), o processo

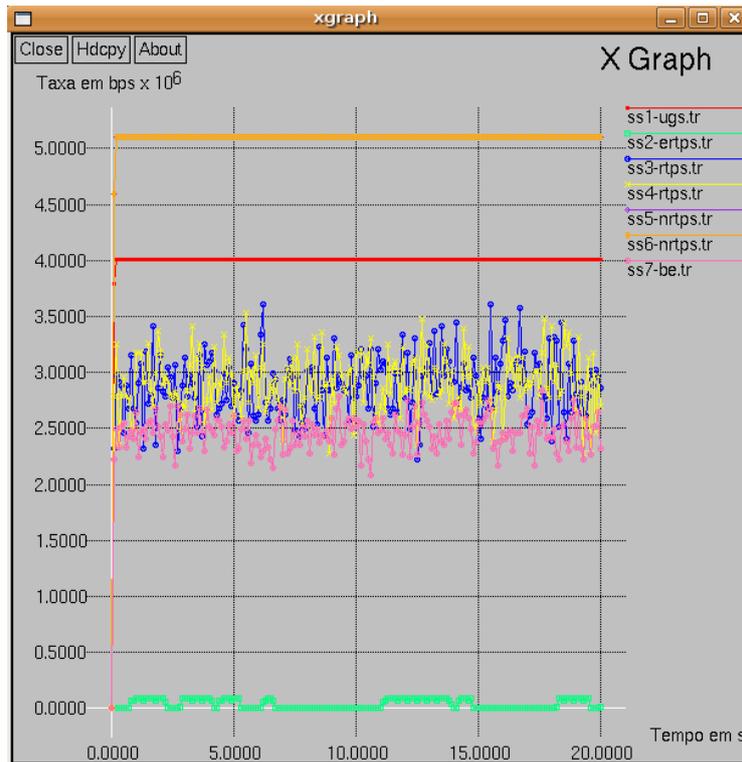


Figura 5.7: Resultado do cenário 2 - DL/UL - 3:1

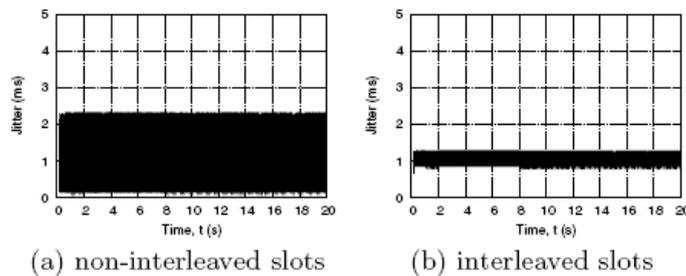


Figura 5.8: Gráfico de Jitter (ms) x tempo (s) em Sayenko *et al.*

nomeado, por aquele autor, de "*interleave*" é realizado. Esta estratégia permite reduzir o valor máximo de *jitter*. Esta alteração tem influência direta na camada física da rede, pois é necessário verificar, além da própria ordem, se os *bursts profiles* são atendidos, que, por sua vez, levam a verificação de parâmetros como relação sinal ruído e codificação. Este procedimento não pode ser aplicado diretamente ao OFDMA PUSC, pois a associação não é apenas de *slots* de tempo, mas também de subcanais lógicos.

A figura 5.9 apresenta a medição do *jitter* alcançado por esta dissertação. O *jitter* varia entre décimos de milissegundo até 4 ms ao longo de todo o período de simulação. Entre pacotes de um mesmo *subframe*, a variação do atraso é muito pequena, mas entre *subframes* distintos pode chegar até 4ms. Ou seja, o que leva a um maior *jitter* é o tempo necessário para BS tratar os *frames*.

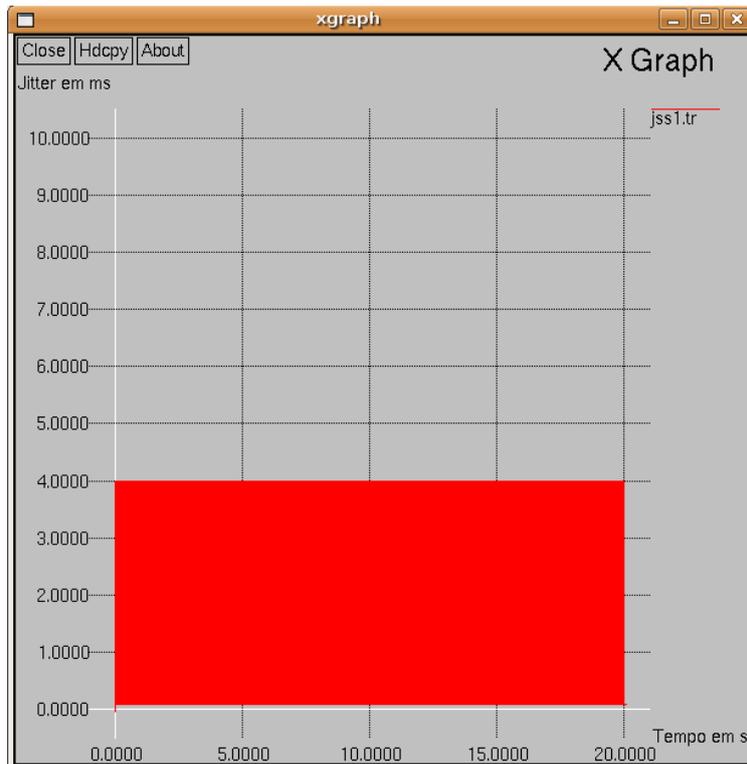


Figura 5.9: Gráfico de Jitter (ms) x tempo (s)

Neste ponto, é importante citar que, quanto maior o tamanho do *frame* utilizado na rede, maior a Acréscimo de *jitter* [13]. O tamanho do *frame* utilizado nesta dissertação é de 5ms, conforme descrito na tabela 5.1. Este parâmetro não foi explicitado em [13], mas é possível estimar o tamanho do *frame* utilizado através dos parâmetros apresentados na tabela 5.2 e dos coneitos de OFDM vistos na seção 3.3.3:

- $BW = 7MHz$, largura de banda do canal em Hz;
- $n = 8/7$, pois trata-se de canal múltiplo de 1,75MHz [4];
- G , pode assumir os valores 1/4, 1/8, 1/16 ou 1/32;
- $N_{FFT} = 256$, número da FFT para o OFDM;

Desta forma, é possível calcular os seguintes parâmetros:

- F_s , frequência de amostragem, $F_s = \left(\left\lfloor \frac{8/7 \times 7MHz}{8000} \right\rfloor \right) \times 8000 = 22,4MHz$;
- Δf , espaçamento entre subportadoras, $\Delta f = \frac{22,4 \times 10^6}{256}$;
- T_b , tempo útil, $T_b = \frac{1}{\Delta f} = 32 \mu s$;

Considerando os valores de G admissíveis, $1/4$, $1/8$, $1/16$ ou $1/32$, podemos encontrar os possíveis valores para T_s : 40, 36, 34 e 33 μs . Com isso, podemos estimar que o *frame* utilizado em [13] está entre 2,6 e 3,2 ms, considerando os 80 *slots* por *frame*. Isto permite com que a BS consiga tratar por segundo até, aproximadamente, o dobro do número de *frame* do módulo utilizado nesta dissertação. De fato, observamos durante as simulações que o número máximo de *frames* por segundo é de 200. Desta forma, é justificável o resultado apresentado na figura 5.9 em relação figura 5.8.

Além da referência [13], encontramos apenas um trabalho no qual o *jitter* é medido como característica de desempenho da estratégia de escalonamento. Em [27], são simulados um cenário residencial, com tráfego BE característico, e um cenário empresarial, com tráfego RTPS e BE, simulando chamadas VoIP, videoconferências e navegação, *download* e *upload* WEB. As estações com tráfego de UL em RTPS apresentam *jitter* de 30 ms. Ressaltamos que em [27] a camada física é OFDM e o tamanho do *frame* é 10 ms, o que reforça a relação do *jitter* com este parâmetro.

5.6.3 Cenário 3 - Escalonamento de SSs com diferentes modulações de sinal

A figura 5.10 apresenta o terceiro cenário a ser simulado, composto por seis estações com serviço de escalonamento do tipo UGS e uma BE. O objetivo deste cenário é verificar se a estratégia de escalonamento proposta possui o comportamento esperado quando existe a variação da modulação dos sinais para uma taxa mais ou menos robusta. Para isso, variou-se ao longo do tempo a modulação e codificação dos sinais das cinco SSs UGS, mantendo a estação BE com a modulação e codificação QPSK 1/2.

A tabela 5.7 define os tipos de tráfego e os parâmetros de QoS para as estações da figura 5.10. Desta forma, as variações da modulação e da codificação não devem causar alterações na taxa de transmissão de cada SS do tipo UGS. Para simular o tráfego do tipo UGS, é utilizado um fluxo de dados constantes a uma taxa de 1 Mbps com protocolo UDP e para simular a estação BE é utilizada uma aplicação FTP enviando dados através do protocolo FTP para o nó destinatário. A inclusão da SS com classe de serviço BE tem o objetivo de verificar se o *scheduler* aloca o recurso restante para a mesma, de acordo com a variação das taxas de modulação e codificação.

A tabela 5.8 define a variação da taxa de modulação das estações em função dos intervalos de tempo. O intervalo de 0 a 5 segundos é o que utiliza o menor número de subcanais para garantir a taxa de transmissão das estações UGS. O intervalo de 15 a 20 segundos, por sua vez, é o que exige mais recurso da BS, tendo em vista utilizar para todas as SSs UGS a segunda combinação (de modulação e codificação) menos eficiente, QPSK 3/4. Portanto, para o primeiro intervalo a estação BE deve conseguir a melhor taxa de transmissão e, para o último, a pior.

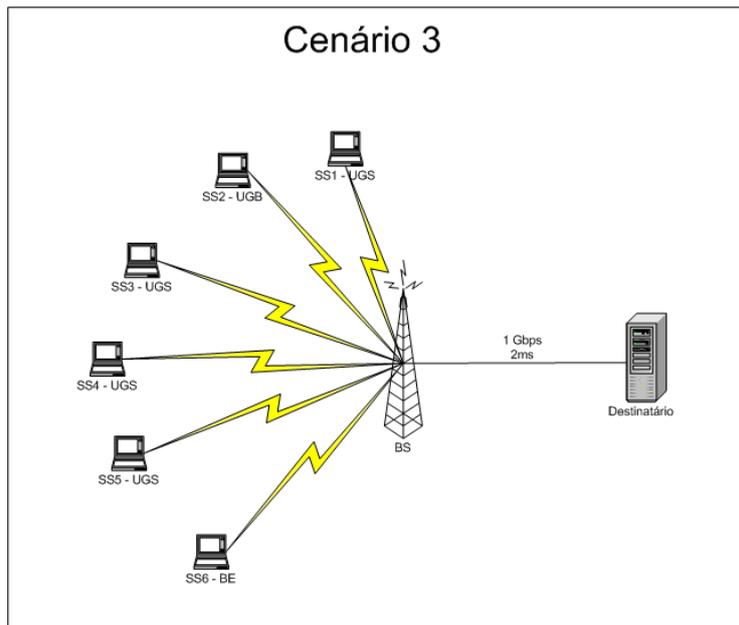


Figura 5.10: Estrutura da rede para o cenário 3

Tabela 5.7: Tipos de tráfego e Parâmetros de QoS para o cenário 3

SS	Tipo de tráfego	Tamanho do pacote	Banda (bps)
SS1	UGS	100	1.000.000
SS2	UGS	100	1.000.000
SS3	UGS	100	1.000.000
SS4	UGS	100	1.000.000
SS5	UGS	100	1.000.000
SS6	BE	1060	-

Tabela 5.8: Controle da Modulação Utilizada

SS	0-5s	5-10s	10-15s	15-20s
SS1	64-QAM 3/4	16-QAM 1/2	16-QAM 1/2	QPSK 3/4
SS2	64-QAM 2/3	16-QAM 1/2	16-QAM 1/2	QPSK 3/4
SS3	16-QAM 3/4	16-QAM 1/2	16-QAM 1/2	QPSK 3/4
SS4	QPSK 3/4	QPSK 3/4	16-QAM 1/2	QPSK 3/4
SS5	QPSK 3/4	QPSK 3/4	16-QAM 1/2	QPSK 3/4
SS6	QPSK 1/2	QPSK 1/2	QPSK 1/2	QPSK 1/2

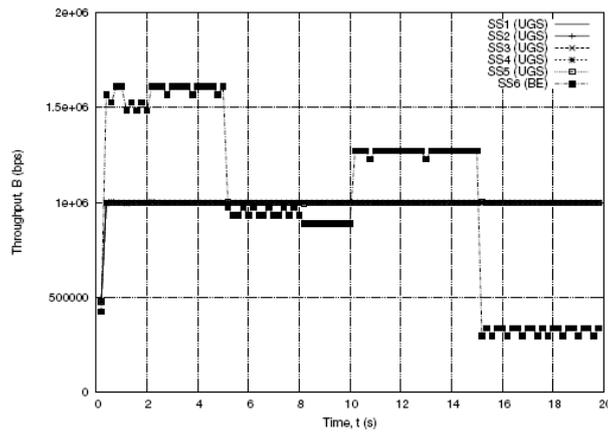


Figura 5.11: Gráfico da Taxa de transmissão X tempo, para o Cenário 3 em Sayenko *et al.*

A figura 5.11 apresenta o resultado obtido em [13] para o mesmo cenário, mas utilizando uma taxa de DL:UL de 1:1. A figura 5.12 contém o gráfico da taxa de transmissão de cada SS em função do tempo, obtido nesta dissertação. Podemos verificar que o comportamento esperado é atingido. A taxa de transmissão para as estações do tipo UGS é mantida durante todo o período de simulação, mesmo com as variações das taxas de modulação e codificação ao longo do tempo. É possível verificar que a estação BE tem sua maior taxa de transmissão no intervalo entre 0 e 5s e a pior no intervalo entre 15 e 20s.

Na figura 5.11 [13], é possível notar pequenas flutuações da taxa de transmissão da estação BE em contraste com a figura 5.12, resultado desta dissertação. Isto ocorre porque o recurso restante, após a garantia do tráfego UGS, é utilizado para a estação BE, mas também para períodos de requisição de banda. No módulo WiMAX, utilizado nesta dissertação, as oportunidades para requisição de banda estão reservadas, com a garantia de 12 símbolos OFDMA por *frame* para este fim. Este procedimento está em conformidade com a última versão do padrão, IEEE 802.16e [4], que manteve inalterado o item 6.3.6.5 do padrão IEEE Std 802.16-2004 [3]. Portanto, o recurso destinado a estação BE está livre do maior ou menor número de requisições necessárias, evitando a flutuação na taxa de transmissão.

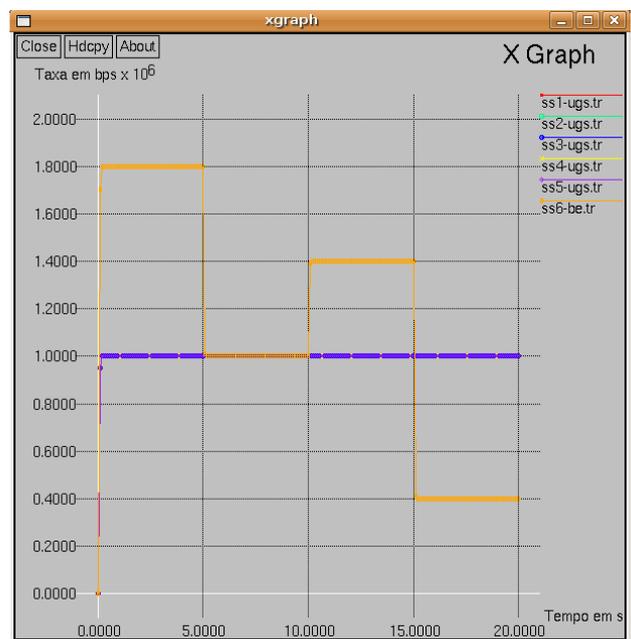


Figura 5.12: Gráfico da Taxa de transmissão X tempo, deste trabalho, para o Cenário 3

Capítulo 6

Conclusões

As redes de comunicação sem fio ganharam destaque com a popularização das redes WiFi e com as novas possibilidades de serviços de multimídia para as redes celulares de terceira geração. As redes WiMAX apresentam vantagens frente às redes anteriores, na medida em que permitem abrangência metropolitana e foram projetadas para propiciar tráfego multimídia com mobilidade e com garantia de qualidade de serviço a um custo inferior às redes celulares. Espera-se também que as redes WiMAX possam contribuir para o fim da dívida digital em países em desenvolvimento [2].

Uma rede WiMAX permite a operação nos modos PMP e Mesh, havendo expectativa de que o modo PMP seja o escolhido para prestação de serviço pelas operadoras de telecomunicações. Este modo permite total controle da comunicação, em ambos os sentidos (DL e UL), por parte da BS. Para este controle, é fundamental o papel de uma estratégia de escalonamento responsável por gerenciar os recursos da rede, atendendo aos parâmetros de qualidade de serviço de cada tipo de conexão.

Para camada de enlace, a normatização IEEE 80216 prevê diversas características como gerenciamento de conexões, definição dos fluxos de serviços, parâmetros obrigatórios de qualidade de serviço e serviços de escalonamento. A norma não prevê, no entanto, o algoritmo específico a ser utilizado no escalonamento.

Diversas propostas de algoritmos de escalonamento foram feitas ([7], [27], [52], [44], [37], [13] e [8]) a partir da lacuna existente no padrão IEEE 802.16 ([3] e [4]).

Em [13], é proposta uma técnica simples e eficaz para solução do problema aderente à camada MAC dos padrões IEEE 802.16. O autor menciona que a técnica é válida para a maioria das camadas físicas possíveis para o WiMAX, mas que para o uso de OFDMA PUSC são necessárias adaptações. A camada física em OFDMA tem sido avaliada como a tecnologia mais flexível e adequada aos diversos tipos de aplicações móveis [27]. A camada física em OFDMA ganhou reforço com a sua adoção como camada física de novas tecnologias de comunicação como como *Long Term Evolution, Ultra Mobile Broadband*, IEEE 802.20

([9], [10], [11] e [12]). Não encontramos nenhum outro trabalho que tenha complementado a proposta de [13].

Nesta dissertação, apresentamos uma estratégia de escalonamento MAC aderente às normas IEEE 802.16 com suporte à camada física WirelessMAN - OFDMA. Esta proposta é uma adaptação da técnica apresentada em [13], considerano o modo PMP, duplexação TDD e distribuição de subcanais PUSC, garantindo QoS dos diferentes tipos de tráfego definidos em acordo de nível de serviço.

Para avaliação desta técnica foram utilizados cenários que incluem comunicações multimídia, como os tráfegos VoIP e IPTV, avaliação da justa divisão dos recursos para o tráfego *best effort* e adaptabilidade às alterações nos perfis de transmissão.

No primeiro cenário, dez estações do tipo *best effort* disparam contra um nó destinatário conectado à BS procurando atingir a maior taxa de transmissão possível. A quantidade de estações é variada ao longo do tempo para verificação do comportamento da estratégia de escalonamento. No segundo cenário, estações de diferentes tipos de tráfego, com diferentes MRTR e MSTR exigem da estratégia de escalonamento a alocação de recursos que atendam a este requisito, com prioridade para as estações utilizando UGS, eRTPS e RTPS. No terceiro cenário, estações UGS disparando contra à BS com taxa de transmissão fixa, têm seus esquemas de modulação e suas taxas de transmissão alteradas ao longo do tempo, enquanto uma estação BE permanece utilizando QPSK 1/2. O objetivo é verificar se a estratégia de escalonamento é capaz de se adaptar as alterações mencionada revervando recursos para as estações UGS em detrimento da estação BE.

Os cenários de rede WiMAX foram simulados utilizando o módulo do NDSL [14] para o NS-2 com as adaptações necessárias, nos quais foi possível observar a correção da estratégia proposta. Os recursos da rede são divididos de forma igualitária entre as estações de mesmo tipo de tráfego. As estações dos diferentes tipos de tráfego têm seus parâmetros de QoS respeitados, em detrimento de estações do tipo BE, por exemplo. E por último, verificou-se que a técnica proposta é capaz de se adaptar aos diversos perfis de transmissão existente em uma rede WiMAX, suportando alterações dos mesmo ao longo do tempo.

Em estudos futuros, pretendemos incluir, na técnica, parâmetros específicos para o controle de *jitter*. Outra contribuição importante seria a adaptação do simulador para permitir aplicação de modelos de mobilidade, avaliando o comportamento da estratégia de escalonamento nestas situações. Outra extensão possível, é a utilização de técnicas de otimização, como a proposta em [6], definindo o número mínimo de subcanais necessários para uma determinada configuração da rede.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] TANENBAUM, A. S. Book. *Computer networks*. [S.l.]: Prentice-Hall, 2003. Quarta Ed., pp. : 912 p. ISBN 0130661023.
- [2] YARALI, A.; RAHMAN, S.; MBULA, B. WiMAX: The innovative broadband wireless access technology. *JCM*, v. 3, n. 2, p. 53–63, Abr 2008.
- [3] IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems. *IEEE Std 802.16-2004 (Revision of IEEE Std 802.16-2001)*, p. 1–857, 2004.
- [4] IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1. *IEEE Std 802.16e-2005 and IEEE Std 802.16-2004/Cor 1-2005 (Amendment and Corrigendum to IEEE Std 802.16-2004)*, p. 1–822, 2006.
- [5] WWW.WIMAXFORUM.ORG. WiMAX forum, 2008, visitado em 10/09/2008.
- [6] VIEIRA, C.; GONDIM, P.; RODRIGUES, C.; BORDIM, J. A new technique to the channel assignment problem in mobile communication networks. *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, p. 1–5, Sept. 2008.
- [7] XERGIAS, S.; PASSAS, N.; SALKINTZIS, A. Centralized resource allocation for multimedia traffic in IEEE 802.16 mesh networks. *Proceedings of the IEEE*, v. 96, n. 1, p. 54–63, Jan. 2008. ISSN 0018-9219.
- [8] WANG, Y.; CHAN, S.; ZUKERMAN, M.; HARRIS, R. Priority-based fair scheduling for multimedia WiMAX uplink traffic. *Communications, 2008. ICC '08. IEEE International Conference on*, p. 301–305, May 2008.
- [9] CHEN, H.-H.; ZHANG, X.; XU, W. Next-generation CDMA vs. OFDMA for 4G wireless applications. *Wireless Communications, IEEE*, v. 14, n. 3, p. 6–7, June 2007. ISSN 1536-1284.

- [10] SANCHEZ, J.; MORALES-JIMENEZ, D.; GOMEZ, G.; ENBRAMBASAGUAS, J. Physical layer performance of long term evolution cellular technology. *Mobile and Wireless Communications Summit, 2007. 16th IST*, p. 1–5, July 2007.
- [11] WANG, M. Ultra mobile broadband technology overview. *Communication Networks and Services Research Conference, 2008. CNSR 2008. 6th Annual*, p. 8–9, May 2008.
- [12] JALILI, A.; FAKHRAIE, S.; NADER-ESFAHANI, S. Performance evaluation of IEEE 802.20 PHY layer. *Computer Engineering and Technology, 2009. ICCET '08. International Conference on*, v. 2, p. 161–165, Jan. 2009.
- [13] SAYENKO, A.; ALANEN, O.; KARHULA, J.; HÄMÄLÄINEN, T. Ensuring the QoS requirements in 802.16 scheduling. *MSWiM '06: Proceedings of the 9th ACM international symposium on Modeling analysis and simulation of wireless and mobile systems*, ACM, New York, NY, USA, p. 108–117, 2006.
- [14] CHEN, J. et al. The design and implementation of WiMAX module for ns-2 simulator. In: *WNS2 '06: Proceeding from the 2006 workshop on ns-2: the IP network simulator*. New York, NY, USA: ACM, 2006. p. 5. ISBN 1-59593-508-8.
- [15] BING, H. et al. Digital content distribution based on hybrid cable and ethernet network. v. 3, p. –, 16-20 2006.
- [16] LIU, X. et al. Performance analysis and industrial practice of peer-assisted content distribution network for large-scale live video streaming. p. 568–574, March 2008. ISSN 1550-445X.
- [17] JIANG, H.; LIU, J.; LI, J.; LI, Z. A scalable system for large-scale multimedia distribution based on hybrid networks. p. 584–587, Nov. 2008.
- [18] MEHTA, P.; UDANI, S. Voice over IP. *Potentials, IEEE*, v. 20, n. 4, p. 36–40, Oct/Nov 2001. ISSN 0278-6648.
- [19] SOARES, V.; NEVES, P.; RODRIGUES, J. Past, present and future of IP telephony. *Communication Theory, Reliability, and Quality of Service, 2008. CTRQ '08. International Conference on*, p. 19–24, 29 2008-July 5 2008.
- [20] ADDA, M.; PEART, A.; ROCHFORD, L. *Voice over IP on Wireless Networks*. 0-0 2006. 2742-2747 p.
- [21] CANBERK, B.; OKTUG, S. Self similarity analysis and modeling of voip traffic under wireless heterogeneous network environment. p. 76–82, May 2009.
- [22] RETNASOTHIE, F. et al. Wireless IPTV over WiMAX: Challenges and applications. *Wireless and Microwave Technology Conference, 2006. WAMICON '06. IEEE Annual*, p. 1–5, Dec. 2006.

- [23] YARALI, A.; CHERRY, A. Internet protocol television (IPTV). *TENCON 2005 2005 IEEE Region 10*, p. 1–6, Nov. 2005.
- [24] SHE, J.; HOU, F.; HO, P.; XIE, L. IPTV over WiMAX: Key success factors, challenges, and solutions [advances in mobile multimedia]. *Communications Magazine, IEEE*, v. 45, n. 8, p. 87–93, August 2007. ISSN 0163-6804.
- [25] UILECAN, I.; ZHOU, C.; ATKIN, G. Framework for delivering IPTV services over WiMAX wireless networks. *Electro/Information Technology, 2007 IEEE International Conference on*, p. 470–475, May 2007.
- [26] WWW.WIMAXFORUM.ORG. Fixed, nomadic, portable and mobile applications for 802.16-2004 and 802.16e WiMAX networks, visitado em 16/09/2008. Nov. 2005.
- [27] CICCONE, C.; LENZINI, L.; MINGOZZI, E.; EKLUND, C. Quality of service support in IEEE 802.16 networks. *Network, IEEE*, v. 20, n. 2, p. 50–55, March-April 2006. ISSN 0890-8044.
- [28] IEEE Std. 802.16-2001 IEEE Standard for Local and Metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems. *IEEE Std 802.16-2001*, p. 1–322, 2002.
- [29] IEEE Standard for Local and metropolitan area networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems-Amendment 1: Detailed System Profiles for 10-66 GHz. *IEEE Std 802.16c-2002 (Amendment to IEEE Std 802.16-2001)*, p. 1–78, 2002.
- [30] IEEE Standard for Local and metropolitan area networks — Part 16: Air Interface for Fixed Broadband Wireless Access Systems— Amendment 2: Medium Access Control Modifications and Additional Physical Layer Specifications for 2-11 GHz. *IEEE Std 802.16a-2003 (Amendment to IEEE Std 802.16-2001)*, p. 1–292, 2003.
- [31] IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems- Amendment 1: Management Information Base. *IEEE Std 802.16f-2005 (Amendment to IEEE Std 802.16-2004)*, p. 1–245, 2005.
- [32] IEEE Standard for Local and Metropolitan Area Networks Media Access Control (MAC) Bridges Amendment 5: Bridging of IEEE 802.16. *802.16k-2007 (Amendment to IEEE Std 802.1D-2004)*, p. 1–14, 2007.
- [33] IEEE Standards for Local and metropolitan area networks - Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems - Amendment 3: Management Plane Procedure and Services. *IEEE Std 802.16g 2007 (Amendment to IEEE Std 802.16-2004)*, p. 1–202, 2007.

- [34] GARBER, L. Mobile wimax: The next wireless battle ground. *Computer*, v. 41, n. 6, p. 16–18, June 2008. ISSN 0018-9162.
- [35] WWW.WIMAXFORUM.ORG. Mobile WiMax part 1 A technical overview and performance evaluation, visitado em 16/09/2008. Agos. 2006.
- [36] FALLAH, Y.; NASIOPOULOS, P.; LEUNG, V. Fair scheduling in multirate wireless access networks. *Wireless Pervasive Computing, 2008. ISWPC 2008. 3rd International Symposium on*, p. 293–297, May 2008.
- [37] WONGTHAVARAWAT, K.; GANZ, A. IEEE 802.16 based last mile broadband wireless military networks with quality of service support. *Military Communications Conference, 2003. MILCOM 2003. IEEE*, v. 2, p. 779–784 Vol.2, Oct. 2003.
- [38] ALAVI, H.; MOJDEH, M.; YAZDANI, N. A quality of service architecture for IEEE 802.16 standards. *Communications, 2005 Asia-Pacific Conference on*, p. 249–253, Oct. 2005.
- [39] SAFA, H. et al. New scheduling architecture for IEEE 802.16 wireless metropolitan area network. *Computer Systems and Applications, 2007. AICCSA '07. IEEE/ACS International Conference on*, p. 203–210, May 2007.
- [40] KHALIL, A.; KSENTINI, A. Classification of the uplink scheduling algorithms in IEEE 802.16. *First International Workshop on Dynamic Networking, INSA Rennes*, Setembro 2007.
- [41] ALI, N. A.; DHRONA, P.; HASSANEIN, H. A performance study of uplink scheduling algorithms in point-to-multipoint WiMAX networks. *Computer Communications*, v. 32, n. 3, p. 511 – 521, 2009. ISSN 0140-3664. Adaptive Multicarrier Communications and Networks. Disponível em: <<http://www.sciencedirect.com/science/article/B6TYP-4TG9HVS-1/2/ce32ae7fee5c0368872dade5010c4c1c>>.
- [42] HAHNE, E. L.; GALLAGER, R. G. Round robin scheduling for fair flow control in data communication networks. p. 103–107, 1986.
- [43] DEMERS, A.; KESHAV, S.; SHENKER, S. *Analysis and simulation of a fair queueing algorithm*. New York, NY, USA: ACM, 1989. 1–12 p.
- [44] CHU, G.; WANG, D.; MEI, S. A QoS architecture for the MAC protocol of IEEE 802.16 BWA system. *Communications, Circuits and Systems and West Sino Expositions, IEEE 2002 International Conference on*, v. 1, p. 435–439 vol.1, June-1 July 2002.
- [45] RAGHU, K.; BOSE, S.; MA, M. Queue based scheduling for IEEE 802.16 wireless broadband. *Information, Communications & Signal Processing, 2007 6th International Conference on*, p. 1–5, Dec. 2007.

- [46] CHEN, T.; WALRAND, J.; MESSERSCHMITT, D. Dynamic priority protocols for packet voice. *Selected Areas in Communications, IEEE Journal on*, v. 7, n. 5, p. 632–643, Jun 1989. ISSN 0733-8716.
- [47] MCKENNEY, P. Stochastic fairness queueing. *INFOCOM '90. Ninth Annual Joint Conference of the IEEE Computer and Communication Societies. 'The Multiple Facets of Integration'. Proceedings., IEEE*, p. 733–740 vol.2, Jun 1990.
- [48] KATEVENIS, M.; SIDIROPOULOS, S.; COURCOUBETIS, C. Weighted round-robin cell multiplexing in a general-purpose ATM switch chip. *Selected Areas in Communications, IEEE Journal on*, v. 9, n. 8, p. 1265–1279, Oct 1991. ISSN 0733-8716.
- [49] PAREKH, A.; GALLAGER, R. A generalized processor sharing approach to flow control in integrated services networks: the single-node case. *Networking, IEEE/ACM Transactions on*, v. 1, n. 3, p. 344–357, Jun 1993. ISSN 1063-6692.
- [50] SHREEDHAR, M.; VARGHESE, G. Efficient fair queuing using deficit round-robin. *Networking, IEEE/ACM Transactions on*, v. 4, n. 3, p. 375–385, Jun 1996. ISSN 1063-6692.
- [51] WANG, S.; WANG, Y.-C.; LIN, K.-J. A priority-based weighted fair queuing scheduler for real-time network. *Real-Time Computing Systems and Applications, 1999. RTCSA '99. Sixth International Conference on*, p. 312–319, 1999.
- [52] CHEN, J.; JIAO, W.; WANG, H. A service flow management strategy for IEEE 802.16 broadband wireless access systems in TDD mode. *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*, v. 5, p. 3422–3426 Vol. 5, May 2005.
- [53] LILEI, W.; HUIMIN, X. A new management strategy of service flow in IEEE 802.16 systems. *Industrial Electronics and Applications, 2008. ICIEA 2008. 3rd IEEE Conference on*, p. 1716–1719, June 2008.
- [54] NAGLE, J. On packet switches with infinite storage. *Communications, IEEE Transactions on [legacy, pre - 1988]*, v. 35, n. 4, p. 435–438, Apr 1987. ISSN 0096-2244.
- [55] GEORGIADIS, L.; GUERIN, R.; PAREKH, A. Optimal multiplexing on a single link: delay and buffer requirements. *INFOCOM '94. Networking for Global Communications., 13th Proceedings IEEE*, p. 524–532 vol.2, Jun 1994.
- [56] WWW.ISI.EDU/NSNAM/NS/. The network simulator - ns-2, 2002, visitado em 10/08/2008.
- [57] ALANEN, O. Multicast polling and efficient voip connections in iee 802.16 networks. In: *MSWiM '07: Proceedings of the 10th ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems*. New York, NY, USA: ACM, 2007. p. 289–295. ISBN 978-1-59593-851-0.

- [58] SAYENKO, A.; TYKHOMYROV, V.; MARTIKAINEN, H.; ALANEN, O. Performance analysis of the IEEE 802.16 ARQ mechanism. In: *MSWiM '07: Proceedings of the 10th ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems*. New York, NY, USA: ACM, 2007. p. 314–322. ISBN 978-1-59593-851-0.
- [59] SAYENKO, A.; MARTIKAINEN, H.; PUCHKO, A. Performance comparison of HARQ and ARQ mechanisms in IEEE 802.16 networks. In: *MSWiM '08: Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems*. New York, NY, USA: ACM, 2008. p. 411–416. ISBN 978-1-60558-235-1.
- [60] GHOSH, D.; GUPTA, A.; MOHAPATRA, P. Scheduling in multihop WiMAX networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, ACM, New York, NY, USA, v. 12, n. 2, p. 1–11, 2008. ISSN 1559-1662.
- [61] CHENG, S.-T.; CHEN, B. Fairness-based scheduling algorithm for TDD mode IEEE 802.16 broadband wireless access systems. In: *ICAIT '08: Proceedings of the 2008 International Conference on Advanced Infocomm Technology*. New York, NY, USA: ACM, 2008. p. 1–6. ISBN 978-1-60558-088-3.
- [62] TSAI, T.; TSAI, Z. Design of a packet scheduling scheme for downlink channel in IEEE 802.16 BWA systems. *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*, p. 1453–1458, 31 2008-April 3 2008. ISSN 1525-3511.
- [63] ZHU, X. et al. QoS-guaranteed scheduling and resource allocation algorithm for IEEE 802.16 OFDMA system. *Communications, 2008. ICC '08. IEEE International Conference on*, p. 3463–3468, May 2008.
- [64] VINAY, K.; SREENIVASULU, N.; JAYARAM, D.; DAS, D. Performance evaluation of end-to-end delay by hybrid scheduling algorithm for QoS in IEEE 802.16 network. *Wireless and Optical Communications Networks, 2006 IFIP International Conference on*, p. 5 pp.–, April 2006.
- [65] LIN, J.-C.; CHOU, C.-L.; LIU, C.-H. Performance evaluation for scheduling algorithms in wimax network. p. 68–74, March 2008.
- [66] WWW.ANTD.NIST.GOV/SEAMLESSANDSECURE/PUBTOOL.SHTML#TOOLS. National institute of standards and technology, 2006, visitado em 5/08/2008.
- [67] FREITAG, J.; FONSECA, N. da. Wimax module for the ns-2 simulator. *Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on*, p. 1–6, Sept. 2007.
- [68] WWW.OPNET.COM/WIMAX/INDEX.HTML. Opnet, 2007, visitado em 5/08/2008.
- [69] WWW.ITU.INT. Artificial conversational speech. *ITU recommendation P.59*, visitado em 14/03/2009, Mar. 1993.

[70] TRACE.EAS.ASU.EDU/. Video traces research group, visitado em 10/01/2009. 2009.