

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ENGENHARIA ELÉTRICA
PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA



UMA PROPOSTA DE ESCALONADOR BASEADO NA
DISCIPLINA *PRIORITY QUEUING* (PQ) PARA REDES
IEEE 802.16

Flávio Alves Ferreira

Uberlândia

2011

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ENGENHARIA ELÉTRICA
PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

UMA PROPOSTA DE ESCALONADOR BASEADO NA
DISCIPLINA *PRIORITY QUEUING* (PQ) PARA REDES
IEEE 802.16

Flávio Alves Ferreira

Dissertação apresentada à Universidade Federal de Uberlândia para obtenção do título de Mestre em Engenharia Elétrica, aprovada em 27 de Janeiro de 2011 pela banca examinadora:

Paulo Roberto Guardieiro, Dr. – Orientador (UFU)

Pedro Frosi Rosa, Dr. (UFU)

Solange da Silva, Dra. (PUC)

**UMA PROPOSTA DE ESCALONADOR BASEADO NA
DISCIPLINA *PRIORITY QUEUING* (PQ) PARA REDES
IEEE 802.16**

Flávio Alves Ferreira

Dissertação apresentada por Flávio Alves Ferreira à Universidade Federal de Uberlândia
como parte dos requisitos à obtenção do título de Mestre em Engenharia Elétrica.

Prof. Paulo Roberto Guardieiro, Dr.
Orientador

Prof. Alexandre Cardoso, Dr.
Coordenador do Curso de Pós-Graduação

Dedicatória

*A meus pais Ricardo e Elisane,
Aos meus amigos,
A todos os meus professores ao longo de minha vida.*

Agradecimentos

Agradeço principalmente à minha família, pelo amor e incentivo.

Ao meu orientador, Prof. Dr. Paulo Roberto Guardieiro, pelos ensinamentos primorosos.

Aos meus amigos e companheiros de laboratório de Redes de Computadores, Neco e Eduardo, da qual pude compartilhar o conhecimento apresentado nesse trabalho, e também com quem pude aprender, e compartilhar momentos de descontração durante o curso do mestrado.

A todos os colegas de UFU que me apoiaram e a todos aqueles que me ajudaram de alguma forma para a realização deste trabalho.

Resumo

Ferreira, F. A., Uma Proposta de Escalonador Baseado na Disciplina *Priority Queuing* (PQ) para Redes IEEE 802.16, UFU, Uberlândia, Brasil, 2011, 118p.

O padrão IEEE 802.16, também conhecido como WiMAX (*Worldwide Interoperability for Microwave Access*), define uma rede de acesso banda larga sem fio (BWA - *Broadband Wireless Access*) com cobertura metropolitana, capaz de atingir altas taxas de transmissão de dados, com facilidade de implantação e manutenção. Sua principal vantagem é a presença inerente de uma arquitetura de QoS (*Quality of Service*). São especificadas duas camadas, a camada de Controle de Acesso ao Meio (MAC) e a camada física (PHY) [1] [2]. A camada MAC é orientada à conexão e o tráfego heterogêneo, nRT (*non Real Time*) e RT (*Real Time*), que é gerado pelas aplicações de rede é gerenciado como fluxos de serviço. Cada fluxo é definido por um sentido, *downlink* (DL) ou *uplink* (UL), e um conjunto de parâmetros para padronizar as operações entre as estações da rede. Os fluxos são mapeados em classes de serviço para a provisão de QoS, de acordo com seus requisitos. O principal mecanismo de QoS é o escalonador, que é responsável por definir a ordem de entrega dos pacotes armazenados nas filas das estações; porém, o padrão não define sua política de operação. Como a largura de banda nas redes WiMAX é um recurso limitado e o tráfego das classes RT apresenta requisitos diferentes das classes nRT como, por exemplo o atraso máximo limitado, a proposta deste trabalho é uma extensão do mecanismo de escalonamento para o tráfego UL apresentado em [3], baseado na disciplina *Priority Queuing* (PQ) para fornecer tratamento diferenciado com priorização de tráfego. Com o objetivo de aumentar a proteção e o isolamento de tráfego entre as classes RT e nRT foi criada uma nova fila separando as requisições dessas classes. Com o objetivo de aumentar a eficiência de utilização dos recursos foi proposto um novo esquema que pré-aloca os recursos que restam no UL, antes de receber o pedido de requisição de largura de banda, ou seja, de maneira antecipada, diretamente para a transmissão dos dados das estações de usuário. A avaliação desta proposta foi efetuada por meio de modelagem e simulação, e os resultados obtidos mostraram que o seu desempenho foi superior ao da proposta apresentada em [3], reduzindo o atraso médio da classe rtPS (*Real Time Polling Service*) e aumentando a vazão média e a utilização de recursos das classes nRT.

Palavras-chave: IEEE 802.16, WiMAX, BWA, QoS, escalonamento.

Abstract

Ferreira, F. A., A Proposal of Scheduler Based on Priority Queuing (PQ) Discipline for IEEE 802.16 Networks, UFU, Uberlandia, Brazil, 2011, 118p.

The IEEE 802.16 standard, also known as WiMAX (Worldwide Interoperability for Microwave Access), defines a broadband wireless access network (BWA) with metropolitan coverage, capable of achieving high data rates transmission, with easy deployment and maintenance. Its main advantage is the presence of an inherent QoS (Quality of Service) architecture. Two layers are specified, the Medium Access Control (MAC) layer and Physical Layer (PHY) [1] [2]. The MAC layer is connection-oriented. The heterogeneous traffic, nRT (non Real Time) and RT (Real Time), which is generated by network applications is managed as service flows. Each flow is defined by a direction, downlink (DL) or uplink (UL), and a set of parameters to standardize operations among stations. The flows are mapped into classes of service for QoS provisioning, according to your requirements. The main QoS mechanism is the scheduler, which is responsible for defining the order delivery of packets stored in the queues of the stations, but the standard does not define its policy of operation. As the bandwidth in WiMAX networks is a limited resource, and the traffic classes, RT and nRT, has different requirements such as the maximum latency, the purpose of this work is an extension of the scheduling mechanism for UL traffic presented in [3], based on Priority Queuing (PQ), to provide differentiated treatment with traffic prioritization. Aiming to increase the protection and isolation between traffic classes, RT and nRT, was created a new queue separating the requests of these classes. Aiming to increase the efficiency of resource utilization has been proposed a new scheme that pre-allocates the remaining resources in the UL, before receiving the bandwidth request, ie, in advance, straightly to the data transmission of the user stations. The assessment of this proposal was made by means of modeling and simulation, and the results showed that its performance was superior to the proposal presented in [3], reducing the average delay of rtPS (*Real Time Polling Service*) class and increasing the average throughput and resource utilization of nRT classes.

Keywords: IEEE 802.16, WiMAX, BWA, QoS, scheduling.

Sumário

1. INTRODUÇÃO	20
2. REDES DE ACESSO BANDA LARGA SEM FIO UTILIZANDO O PADRÃO IEEE 802.16	24
2.1. Introdução	24
2.2. Padrão IEEE 802.16	24
2.3. Modelo de Referência do Padrão IEEE 802.16	28
2.3.1. A Camada MAC	29
2.3.1.1. MAC PDU: O Quadro da Camada MAC	30
2.3.1.1.1. Cabeçalho Genérico da MAC PDU	31
2.3.1.1.2. Cabeçalho de Requisição de Banda da MAC PDU	33
2.3.1.2. Subcamadas da Camada MAC	34
2.3.1.2.1. Subcamada de Convergência Específica (CS)	34
2.3.1.2.2. Subcamada da Parte Comum (CPS)	35
2.3.2. A Camada Física	38
2.3.2.1. Quadro de Operação da Camada Física	39
2.3.2.2. Função de Adaptação do Enlace	41
2.4. Provisão de Qualidade de Serviço (QoS) em Redes Sem Fio	43
2.5. Provisão de Qualidade de Serviço (QoS) no Padrão IEEE 802.16	44
2.5.1. Classes de Serviço	45
2.5.2. Fluxos de Serviço	47
2.5.3. Mecanismos de Provisão de QoS no Padrão IEEE 802.16	48
2.6. Considerações Finais	52
3. ESCALONAMENTO EM REDES IEEE 802.16	54
3.1. Introdução	54
3.2. Disciplinas de Escalonamento	54
3.3. Estado da Arte dos Mecanismos de Escalonamento em Redes IEEE 802.16	59
3.3.1. Classificação dos Algoritmos de Escalonamento Utilizados em Redes IEEE 802.16	60
3.3.2. Propostas de Mecanismos de Escalonamento para as Redes IEEE 802.16	61
3.4. Considerações Finais	73

4. UMA PROPOSTA DE ESCALONADOR BASEADO NA DISCIPLINA <i>PRIORITY QUEUING</i> (PQ) PARA REDES IEEE 802.16	74
4.1. Introdução	74
4.2. Descrição do Problema	75
4.3. Solução Proposta	77
4.4. Trabalhos Relacionados	85
4.5. Considerações Finais	87
5. ANÁLISE DE DESEMPENHO DO ESCALONADOR PROPOSTO	89
5.1. Introdução	89
5.2. O Simulador para o Ambiente de Rede IEEE 802.16 (WiMAX)	89
5.3. Ambiente de Rede e Parâmetros de Simulação	90
5.4. Apresentação e Análise dos Resultados	92
5.4.1. Cenário <i>I</i>	93
5.4.2. Cenário <i>II</i>	97
5.4.3. Cenário <i>III</i>	99
5.4.4. Cenário <i>IV</i>	102
5.5. Considerações Finais	104
6. CONCLUSÕES GERAIS	105
7. REFERÊNCIAS BIBLIOGRÁFICAS	110

Lista de Figuras

Figura 2.1: Emprego das Redes BWA do Padrão IEEE 802.16 [6]	26
Figura 2.2: Topologia PMP	27
Figura 2.3: Topologia <i>Mesh</i>	28
Figura 2.4: Modelo de Referência do Padrão IEEE 802.16 [2]	28
Figura 2.5: Formato da MAC PDU [1] [2]	30
Figura 2.6: Composição do Cabeçalho Genérico da MAC PDU [1]	31
Figura 2.7: Composição Eventual da MAC PDU [10]	32
Figura 2.8: Composição do Cabeçalho de Requisição de Banda da MAC PDU [1]	33
Figura 2.9: Esquema de Transmissão TDD e FDD [1] [2]	39
Figura 2.10: Estrutura do Quadro TDD [10]	39
Figura 2.11: Utilização do MCS de Acordo com a Distância da SS à BS [14]	43
Figura 2.12: Arquitetura de QoS no padrão IEEE 802.16d [24]	44
Figura 2.13: Teoria do Modelo de Objetos de Operação [2]	45
Figura 2.14: <i>Polling Unicast</i> [25]	49
Figura 2.15: <i>Polling</i> Baseado em Contenção [25]	50
Figura 2.16: Mecanismos de Escalonamento da Rede IEEE 802.16	52
Figura 3.1: Encaminhamento de Pacotes de Acordo com o Escalonador FIFO	55
Figura 3.2: Encaminhamento de Pacotes de Acordo com o Escalonador PQ	56
Figura 3.3: Encaminhamento de Pacotes de Acordo com o Escalonador FQ	56
Figura 3.4: Comportamento Inadequado do Escalonamento WRR [35]	57
Figura 3.5: Arquitetura do Escalonador Proposto em [3]	62
Figura 3.6: Estrutura Hierárquica do UPS Proposto por [47]	64
Figura 3.7: Estrutura Modular do Mecanismo de Escalonamento Proposto por [48]	65
Figura 3.8: Arquitetura de Escalonamento Proposta por [49]	66
Figura 3.9: Dinâmica da Elegibilidade das SSs para Escalonamento em [50]	67
Figura 3.10: Mecanismo <i>Token Bucket</i> [51]	68
Figura 3.11: Escalonador Apresentado em [52]	69
Figura 3.12: Ajuste Dinâmico do Tamanho dos <i>Subframes</i> [53]	70
Figura 4.1: Estrutura do Mecanismo de Escalonamento Proposto	78
Figura 4.2: Vantagem do Esquema de Pré-Alocação	80

Figura 5.1: Cenário de Simulação de uma Rede IEEE 802.16 PMP TDD	91
Figura 5.2: Vazão das SSs nrtPS (PQ-4F vs. PQ-3F)	94
Figura 5.3: Vazão de Dez Conexões Concorrentes nrtPS (PQ-4F)	95
Figura 5.4: Percentual de Utilização (PQ-4F vs. PQ-3F)	96
Figura 5.5: Percentual de Pacotes de Gerenciamento no <i>Uplink</i> (PQ-4F vs. PQ-3F)	97
Figura 5.6: Vazão Média das Conexões nrtPS e BE (PQ-4F vs. PQ-3F)	98
Figura 5.7: Atraso Médio das SSs UGS e rtPS (PQ-4F vs. PQ-3F)	98
Figura 5.8: Percentual de Pacotes Descartados nos <i>Buffers</i> (PQ-4F)	100
Figura 5.9: Vazão Média das Conexões nrtPS (PQ-4F)	100
Figura 5.10: Percentual de Pacotes Descartados nos <i>Buffers</i> (PQ-4F vs. PQ-3F)	101
Figura 5.11: Vazão Média em Relação à Capacidade dos <i>Buffers</i> (PQ-4F vs. PQ-3F)	101
Figura 5.12: Atraso Médio das Conexões rtPS (PQ-4F vs. PQ-3F)	102
Figura 5.13: Vazão Média das Conexões rtPS e nrtPS (PQ-4F vs. PQ-3F)	103

Lista de Tabelas

Tabela 2.1: Família IEEE 802.16	25
Tabela 2.2: Interfaces Aéreas Definidas para a Camada Física [1]	38
Tabela 2.3: MCS Relacionada ao Patamar SNR	42
Tabela 2.4: Parâmetros de QoS das Classes de Serviço [1]	46
Tabela 5.1: Parâmetros de Configuração da Rede	91
Tabela 5.2: Características dos <i>Traces</i> de Vídeos Utilizados nas Simulações [3]	92
Tabela 5.3: Intervalo de Confiança de Vazão (Figura 5.3)	95
Tabela 5.4: Intervalo de Confiança de Vazão (Figura 5.4)	95

Lista de Abreviaturas

4G	Quarta Geração
ACQPS	<i>ActiveQoSParamSet</i>
AGT	Agente Gerador de Tráfego
AMC	<i>Adaptive Modulation and Coding</i>
AQPS	<i>AdmittedQoSParamSet</i>
ARQ	<i>Automatic Repeat Request</i>
ATM	<i>Asynchronous Transfer Mode</i>
BE	<i>Best Effort</i>
bps	<i>bits por segundo</i>
BPSK	<i>Binary Phase Shift Keying</i>
BR	<i>Bandwidth Request</i>
BS	<i>Base Station</i>
BWA	<i>Broadband Wireless Access</i>
CAC	<i>Call Admission Control</i>
CBR	<i>Constant Bit Rate</i>

CI	<i>CRC Indicator</i>
CID	<i>Connection Identifier</i>
CPS	<i>Common Part Sublayer</i>
CRC	<i>Cyclic Redundancy Check</i>
CS	<i>Convergence Sublayer</i>
DCD	<i>Downlink Channel Descriptor</i>
DFS	<i>Dynamic Frequency Selection</i>
DHCP	<i>Dynamic Host Configuration</i>
DL-MAP	<i>Downlink Map</i>
DOCSIS	<i>Data Over Cable Service Interface Specification</i>
DRR	<i>Deficit Round Robin</i>
DSA	<i>Dynamic Service Addition</i>
DSA-REQ	<i>Dynamic Service Addition Request</i>
DSC	<i>Dynamic Service Change</i>
DSD	<i>Dynamic Service Delete</i>
E1	<i>Tecnologia Europeia de Transmissão</i>
EC	<i>Encryption Control</i>

EDF	<i>Earliest Deadline First</i>
EKS	<i>Encryption Key Sequence</i>
e-mail	<i>Electronic Mail</i>
ertPS	<i>Extend Real Time Polling Service</i>
FDD	<i>Frequency Division Duplexing</i>
FEC	<i>Forward Error Correction</i>
FIFO	<i>First In First Out</i>
FTP	<i>File Transfer Protocol</i>
Gbps	<i>Gigabits Por Segundo</i>
GPC	<i>Grant Per Connection</i>
GPS	<i>Generalized Processor Sharing</i>
GPSS	<i>Grant Per Subscriber Station</i>
HCS	<i>Header Check Sequence</i>
HT	<i>Header Type</i>
HTTP	<i>HyperText Transfer Protocol</i>
IE	<i>Information Element</i>
IEEE	<i>Institute of Electrical and Eletronics Engineers</i>

IP	<i>Internet Protocol</i>
LoS	<i>Line of Sight</i>
LSB	<i>Least Significant Bit</i>
Kbps	<i>Kilobits Por Segundo</i>
MAC	<i>Medium Access Control</i>
Mbps	<i>Megabits Por Segundo</i>
MIB	<i>Management Information Base</i>
MPEG	<i>Motion Picture Expert Group</i>
MRTR	<i>Minimum Reserved Traffic Rate</i>
MSB	<i>Most Significant Bit</i>
MSTR	<i>Maximum Sustained Traffic Rate</i>
NLoS	<i>Non Line of Sight</i>
nRT	<i>Non Real Time</i>
nrtPS	<i>Non Real Time Polling Service</i>
NS-2	<i>Network Simulator – 2</i>
OFDM	<i>Orthogonal Frequency Division Multiplexing</i>
OFDMA	<i>Orthogonal Frequency Division Multiplexing Access</i>

PDU	<i>Protocol Data Unit</i>
PHS	<i>Payload Header Suppression</i>
PHY	<i>Physical Layer</i>
PMP	<i>Point-to-Multipoint</i>
PPP	<i>Point-to-Point</i>
PQ	<i>Priority Queuing</i>
QAM	<i>Quadrature Amplitude Modulation</i>
QoS	<i>Quality of Service</i>
QPSK	<i>Quadrature Phase Shift Keying</i>
RLC	<i>Radio Link Control</i>
RR	<i>Round Robin</i>
RT	<i>Real Time</i>
rtPS	<i>Real Time Polling Service</i>
SAP	<i>Service Access Point</i>
SC	<i>Single Carrier</i>
SCa	<i>Single Carrier Access</i>
SDU	<i>Service Data Unit</i>

SFID	<i>Service Flow Identifier</i>
SLA	<i>Service Level Agreement</i>
SNMP	<i>Simple Network Management Protocol</i>
SNR	<i>Signal to Noise Ratio</i>
SS	<i>Subscriber Station</i>
T1	Tecnologia Americana de Transmissão
TDD	<i>Time Division Duplexing</i>
TDM	<i>Time Division Multiplexing</i>
TDMA	<i>Time Division Multiplexing Access</i>
TFTP	<i>Trivial File Transfer Protocol</i>
TLV	<i>Type/Length/Value</i>
UCD	<i>Uplink Channel Descriptor</i>
UGS	<i>Unsolicited Grant Service</i>
UL-MAP	<i>Uplink Map</i>
UNI	<i>User Network Interface</i>
VC	<i>Virtual Channel</i>
VCI	<i>Virtual Channel Identifier</i>

VoIP	<i>Voice Over Internet Protocol</i>
VP	<i>Virtual Path</i>
VPI	<i>Virtual Path Identifier</i>
WF²Q	<i>Worst-case Fair Weighted Fair Queuing</i>
WFQ	<i>Weighted Fair Queuing</i>
Wi-Fi	<i>Wireless Fidelity</i>
WiMAX	<i>Worldwide Interoperability for Microwave Access</i>
WMAN	<i>Wireless Metropolitan Area Network</i>
WRR	<i>Weighted Round Robin</i>
xDSL	<i>Various Digital Subscriber Line</i>

Capítulo 1

INTRODUÇÃO

O crescimento da demanda por acesso banda larga à Internet impulsionou a pesquisa e o desenvolvimento de diversas tecnologias, das quais, as redes de acesso banda larga sem fio (BWA – *Broadband Wireless Access*) vêm se consolidando em um cenário inicialmente dominado pelas redes cabeadas de acesso à última milha como, por exemplo, xDSL (*Various Digital Subscriber Line Technologies*), *Cable Modem* e ISDN (*Integrated Services Digital Network*), em função da sua flexibilidade, facilidade de implantação e manutenção, custo/benefício adequado e mobilidade. Assim, áreas rurais de baixa densidade populacional ou de difícil acesso físico, além de regiões metropolitanas com alta densidade demográfica, porém com pouco espaço para a passagem de fios, podem se beneficiar do acesso banda larga à Internet através das redes BWA.

A padronização favorece a interoperabilidade dos equipamentos de rede desenvolvidos por fabricantes diferentes, de modo que o IEEE (*Institute of Electrical and Electronic Engineers*) definiu uma rede BWA através do padrão IEEE 802.16 [1] [2], comercialmente conhecida como WiMAX (*Worldwide Interoperability for Microwave Access*), que é o nome do consórcio de empresas responsável pela certificação dos equipamentos produzidos para as redes desse padrão.

Diferentemente das redes de telefonia celular, o padrão IEEE 802.16 foi projetado inicialmente para a transmissão de dados e é capaz de fornecer taxas de transmissão superiores às taxas oferecidas por aquelas redes, que foram inicialmente projetadas para o tráfego de voz. As taxas de transmissão de dados oferecidas pelas redes WiMAX podem chegar a centenas de *Megabits* por segundo (Mbps), com a perspectiva de evolução para alcançar taxas da ordem de *Gigabits* por segundo (Gbps).

As redes IEEE 802.16 abrangem regiões muito amplas, com uma área de cobertura metropolitana, ao invés de cobrirem apenas uma região local de área limitada, como as redes IEEE 802.11 (*Wi-Fi – Wireless Fidelity*).

A principal vantagem das redes WiMAX é a presença inerente de mecanismos de QoS (*Quality of Service*) em sua arquitetura, característica fundamental para consolidar a chamada convergência que as futuras redes de comunicação de quarta geração (4G) prometem oferecer, cuja demanda por serviços é chamada de *quadruple-play*, termo que se refere ao tráfego de dados, vídeo e voz com mobilidade. O tráfego heterogêneo das redes 4G é gerado por aplicações como, por exemplo, *streaming* de áudio e vídeo, aplicações de transferência remota de arquivos através do protocolo FTP (*File Transfer Protocol*), aplicações de missão crítica em tempo real, IPTV (*Internet Protocol Television*), VoIP (*Voice over Internet Protocol*), *web browsing*, etc. Cada tipo de tráfego possui necessidades específicas, que devem ser atendidas de maneira diferenciada, para que a percepção da qualidade de serviço de cada usuário seja satisfatória.

O padrão IEEE 802.16 especifica duas camadas, a camada de controle de acesso ao meio (MAC – *Medium Access Control*), e a camada física (PHY - *Physical*). A camada MAC é orientada à conexão, o que significa que para uma estação de usuário (SS – *Subscriber Station*) poder transmitir seus dados, é preciso antes trocar mensagens de sinalização para estabelecer um fluxo de serviço com a estação base (BS – *Base Station*). Então, o protocolo da camada MAC efetua a associação individual de um identificador de conexão (CID – *Connection Identifier*) com um identificador de fluxo de serviço (SFID – *Service Flow Identifier*). Um fluxo de serviço possui um conjunto de parâmetros de QoS como, por exemplo, a taxa mínima requisitada e a latência máxima, e um sentido, *downlink* (DL) para o tráfego que é transmitido da estação base para as estações de usuário, ou *uplink* (UL) que é o sentido do tráfego das SSs para a BS. Os fluxos de serviços são mapeados em classes de serviços para a provisão de QoS de acordo com os seus requisitos [1] [2].

A arquitetura de QoS definida pelo padrão IEEE 802.16 é composta por mecanismos como, por exemplo, o Controle de Admissão de Conexões (CAC), o mecanismo de policiamento de tráfego e o mecanismo de escalonamento. O principal mecanismo dessa arquitetura é o escalonador, que é responsável por definir a ordem de envio dos pacotes armazenados nas filas das estações da rede, BS e SS.

O padrão especifica três escalonadores: o escalonador da BS de *downlink* (BS-DL) que seleciona diretamente os pacotes armazenados em suas filas para transmiti-los para as SSs; o escalonador da SS de *uplink* (SS-UL) que seleciona diretamente os pacotes armazenados em suas filas para transmiti-los para a BS; e o mais complexo deles, o escalonador da BS de *uplink* (BS-UL), que não tem acesso direto às filas nas SSs e depende das mensagens de requisição enviadas por elas para efetuar a alocação de largura de banda no *uplink*. A

eficiência no atendimento da demanda de tráfego *uplink* produzido pelas SSs depende do sincronismo da troca de mensagens entre a BS e cada SS para requisição e alocação de largura de banda, e também da disponibilidade de largura de banda, que é um recurso compartilhado por inúmeros usuários e cada vez mais escasso.

O padrão IEEE 802.16 especifica a necessidade do emprego de mecanismos para a provisão de QoS, porém não define as políticas de operação, outorgando essa tarefa aos desenvolvedores e fabricantes de equipamentos de rede, que assim podem diferenciar seus produtos.

A política de envio de pacotes do mecanismo de escalonamento é implementada através de um algoritmo, que pode ser baseado em disciplinas de escalonamento. De acordo com o estudo a respeito do estado da arte em escalonamento em redes IEEE 802.16, apresentado no capítulo 3, uma única disciplina de escalonamento não é capaz de garantir todos os requisitos de QoS. Portanto, a escolha de uma política eficiente depende do perfil de tráfego da rede. Por exemplo, o escalonador apresentado em [3], cujo algoritmo é implementado em um módulo WiMAX com código aberto disponível para o acesso de pesquisadores e desenvolvedores, é capaz de fornecer suporte para a provisão de QoS para todas as classes definidas pelo padrão IEEE 802.16d utilizando a disciplina de escalonamento PQ (*Priority Queing*) e, funções específicas para o atendimento das requisições das diferentes classes armazenadas em filas dedicadas, oferecendo tratamento diferenciado com priorização, adequado para atender o tráfego heterogêneo das redes WiMAX.

No entanto, no mecanismo de escalonamento da BS para o tráfego *uplink* apresentado em [3], não oferece um isolamento de tráfego completo porque as requisições das classes de serviço rtPS e nrtPS são atendidas na mesma fila de prioridade, e o comportamento do tráfego não tempo real, que é tolerante ao atraso, pode prejudicar o comportamento do tráfego tempo real, que não é tolerante ao atraso. Além disso, a eficiência desse escalonador é prejudicada pelo processo de requisição e alocação de largura de banda, em função da maneira como são atualizadas as informações na BS a respeito do tamanho das filas nas SSs, e pelo fato de que ele não aproveita os recursos que restam, após atender todas as filas, diretamente para a transmissão de dados, para maximizar a utilização de recursos no enlace *uplink*.

Assim, neste trabalho é proposta uma extensão do mecanismo de escalonamento BS-UL para redes IEEE 802.16, apresentado em [3], com o objetivo de fornecer tratamento diferenciado com priorização, separando as filas de atendimento dos tipos de tráfego de tempo real e não tempo real para aumentar o isolamento entre eles, e aumentar a eficiência de

utilização dos escassos recursos no enlace *uplink* pré-alocando os recursos restantes, após o escalonador atender as filas, diretamente para a transmissão de dados.

A análise de desempenho do mecanismo de escalonamento proposto é feita por meio de modelagem e simulação, utilizando o módulo WiMAX apresentado em [4], basicamente comparando o desempenho desse mecanismo com o desempenho do mecanismo apresentado em [3]. Os parâmetros de QoS analisados foram a vazão média, o atraso médio, a utilização percentual de recursos e o percentual de pacotes descartados.

Este trabalho está organizado da seguinte maneira:

O Capítulo 2 apresenta as extensões que compõem a família do padrão IEEE 802.16, o modelo de referência, e a estrutura das camadas MAC e PHY, abordando detalhes da estrutura do quadro da camada MAC como, por exemplo, o seu cabeçalho; as subcamadas da camada MAC, e as especificações operacionais da camada PHY. Após apresentar as principais características estruturais do padrão, são apresentados os principais conceitos sobre a provisão de QoS nas redes WiMAX, tais como classes de serviço, fluxos de serviço e os principais mecanismos que compõem a arquitetura de QoS dessas redes.

O Capítulo 3 apresenta os fundamentos básicos a respeito de escalonamento, disciplinas de escalonamento e o estado da arte da pesquisa em escalonamento em redes IEEE 802.16, descrevendo as características dos mecanismos de escalonamento, a classificação dos algoritmos e as principais propostas publicadas sobre o tema.

O Capítulo 4 apresenta a proposta de mecanismo de escalonamento da BS para o tráfego *uplink* baseado na disciplina PQ, que é uma extensão do mecanismo proposto em [3], descrevendo os problemas que motivaram o desenvolvimento dessa proposta e as soluções para oferecer tratamento diferenciado com priorização, e ainda aumentar a eficiência de utilização dos escassos recursos no enlace *uplink*. Por último, são apresentados os trabalhos relacionados.

O Capítulo 5 aborda a análise de desempenho da proposta por meio de modelagem e simulação. Os parâmetros apresentados para a avaliação de desempenho do mecanismo de escalonamento proposto são os principais parâmetros de QoS descritos pela literatura especializada como, por exemplo, vazão média, atraso médio e utilização de recursos.

Finalmente, o Capítulo 6, apresenta as conclusões gerais a respeito deste trabalho, e as sugestões para trabalhos futuros.

Capítulo 2

REDES DE ACESSO BANDA LARGA SEM FIO UTILIZANDO O PADRÃO IEEE 802.16

2.1. Introdução

O padrão IEEE 802.16 é uma alternativa de rede de acesso banda larga sem fio (BWA - *Broadband Wireless Access*), que define a camada de controle de acesso ao meio, MAC (*Medium Access Layer*), e a camada física, PHY (*Physical layer*), respectivamente camada de enlace e camada física do modelo de rede TCP/IP da Internet.

A camada MAC do padrão IEEE 802.16 é orientada à conexão, que determina que qualquer fluxo estabelecido deve ser associado a algum tipo de conexão. A associação de um fluxo de serviço com uma conexão de transporte é fundamental para a operação do protocolo da camada MAC. Os fluxos são mapeados em classes de serviços para a provisão de QoS.

As redes IEEE 802.16 foram projetadas para prover garantias de QoS para o tráfego de fluxos heterogêneos. O padrão define uma arquitetura de QoS composta, por exemplo, de mecanismos de Controle de Admissão de Conexões (CAC – *Call Admission Control*), policiamento de tráfego e escalonamento.

Este Capítulo está estruturado da seguinte maneira: A Seção 2.2 apresenta o padrão IEEE 802.16, a Seção 2.3 descreve a sua estrutura de camadas. Na Seção 2.4 é abordado o assunto sobre QoS em redes sem fio, na Seção 2.5 são apresentados conceitos importantes sobre classes de serviço, fluxo de serviço e os principais mecanismos de QoS definidas pelo padrão IEEE 802.16. Na Seção 2.6 são feitas as considerações finais sobre este Capítulo.

2.2. Padrão IEEE 802.16

A padronização de uma tecnologia é importante para promover a compatibilidade e a interoperabilidade dos equipamentos desenvolvidos por diferentes fabricantes, facilitando sua

difusão comercial. Em julho de 1999, o IEEE criou o grupo de trabalho 802.16 (IEEE 802.16 *Working Group*), com a missão de desenvolver um padrão para redes BWA, e durante o ano de 2001, foi criado o WiMAX *Forum*, uma espécie de consórcio de empresas como a AT&T, Intel, Siemens, e diversas outras, que é responsável pela certificação dos equipamentos desenvolvidos com base no padrão IEEE 802.16, cujo *web site* é apresentado em [5].

Desde 2001, o IEEE 802.16 *Working Group* têm definido novas recomendações e vêm concretizando essas mudanças através da composição de revisões do padrão original, também chamadas de extensões, apresentadas na Tabela 2.1:

Tabela 2.1: Família IEEE 802.16.

Padrão	Descrição	Ano
IEEE 802.16	Padrão BWA operando em frequências entre 10 e 66 GHz com linha de visada (LoS - <i>Line of Sight</i>)	2001
IEEE 802.16a	Padrão BWA operando em frequências entre 2 e 11 GHz sem necessidade de linha de visada (NLoS - <i>Non Line of Sight</i>) e alcance de até 50 Km	2003
IEEE 802.16b	Padrão permitindo o uso de frequências de 5 a 6 GHz não licenciadas	2003
IEEE 802.16c	Padrão permitindo a interoperabilidade entre frequências até 66 GHz com linha de visada (LOS)	2002
IEEE 802.16d	Aprimoramento dos padrões IEEE 802.16, IEEE 802.16a e IEEE 802.16c	2004
IEEE 802.16e	Padrão que especifica mobilidade com LOS para frequências entre 10 e 66 GHz e para NLOS frequências entre 2 e 11 GHz	2005
IEEE 802.16f	Padrão que introduz o conceito de redes em malha (<i>Mesh</i>)	2005
IEEE 802.16g	Padrão de suporte a mobilidade	2007
IEEE 802.16h	Padrão que permite operação em bandas ISM (<i>Industrial Scientific and Medical</i>) 2,4 GHz e 5,8 GHz	2005
IEEE 802.16i	Inclui o conceito de MIB (<i>Management Information Base</i>), que especifica quais variáveis são mantidas pelos elementos de rede	2005
IEEE 802.16j	Especifica operações em saltos múltiplos com retransmissões e interoperabilidade entre as estações retransmissoras e BSs	2009
IEEE 802.16k	Propõe a união do Padrão IEEE 802.1d (<i>bridge</i> transparente) no reconhecimento da MAC do Padrão IEEE 802.16	2007
IEEE 802.16m	Interface aérea avançada para a compatibilidade do Padrão IEEE 802.16d (nômade) e futuras redes 4G com taxas de até 1 Gbps	2009

O padrão IEEE 802.16 define uma rede BWA capaz de oferecer altas taxas de transmissão de dados, da ordem de centenas de *Megabits* por segundo (Mbps) podendo chegar futuramente a *Gigabits* por segundo (Gbps), e acesso a usuários finais distribuídos em uma ampla faixa de cobertura do sinal de rádio frequência, com raio da ordem de quilômetros (Km), como as regiões metropolitanas, e sendo por isso também chamado de *WirelessMAN* (*Wireless Metropolitan Area Network*). Além disso, permite a mobilidade dos seus nós, apresenta fácil implantação e manutenção, e dispensa a passagem de fios, característica das redes tradicionais banda larga ADSL e *Cable Modem*, que exige uma infraestrutura complexa e onerosa, principalmente em regiões suburbanas e rurais, de difícil acesso.

As redes WiMAX podem ser utilizadas como *backhaul* ou *backbone* para interligar diferentes tipos de redes públicas locais, fornecendo acesso sem fio em locais com a presença de *hot-spots* para redes *Wi-Fi*, ou fornecer acesso de última milha a usuários finais dedicados, como os usuários de Internet domésticos ou corporativos. A Figura 2.1 apresenta a ampla possibilidade de emprego das redes IEEE 802.16.

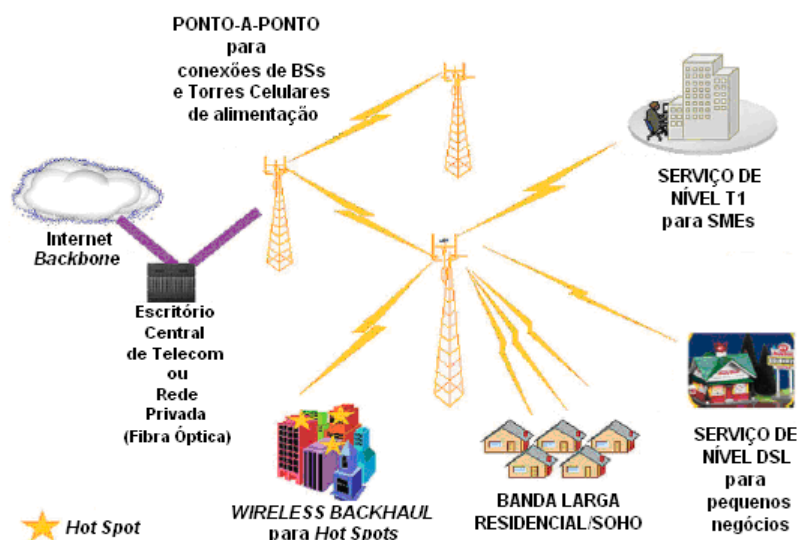


Figura 2.1: Emprego das Redes BWA do Padrão IEEE 802.16 [6].

A provisão de acesso à última milha depende da topologia de rede empregada:

PMP (*Point to Multipoint*): A estação base (BS – *Base Station*) controla e gerencia todo o tráfego entre as estações de usuários (SS – *Subscriber Station*) dentro de sua área de cobertura, ou célula, por isso é caracterizada como uma rede *single-hop* [7]. As SSs não se comunicam diretamente. Portanto, para uma SS se comunicar com outra é necessário que a SS

remetente primeiro se comunique diretamente com a BS, que por sua vez transmite a mensagem para a SS receptora. A comunicação no sentido da BS para a SS é chamada de *downlink* (DL), e a comunicação no sentido da SS para a BS é chamada de *uplink* (UL). No *downlink* as mensagens são transmitidas apenas pela BS para todas as SSs via *broadcast* e cada SS fica apenas com a mensagem destinada a ela. No *uplink*, as SSs compartilham o direito de transmissão de mensagens para a BS de acordo com a sua demanda, comunicando-se em oportunidades individuais concedidas pela BS [7]. A arquitetura de rede PMP é mostrada na Figura 2.2.

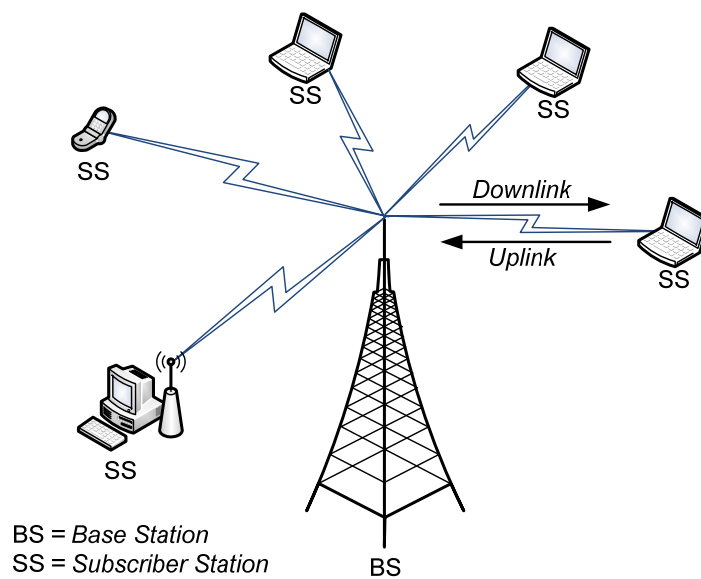


Figura 2.2: Topologia PMP.

Malha (*Mesh*): As SSs podem se comunicar entre si, sem a necessidade de um elemento centralizador como a BS, o que permite uma variabilidade de rotas entre o núcleo da rede e as diversas estações repetidoras de assinantes. Assim, é caracterizada como uma rede *multi-hop* [7]. Apresenta a vantagem de que em caso de falha de um nó, os nós vizinhos poderão se conectar com outros nós e a rede apresentará um único ponto de falha, apresentando maior confiabilidade e redundância em relação à topologia PMP. A sua desvantagem em relação à arquitetura PMP é que em cada nó repetidor (SS) é necessário implementar um mecanismo de roteamento de pacotes, aumentando seu custo de implementação. Essa arquitetura de rede é mostrada na Figura 2.3.

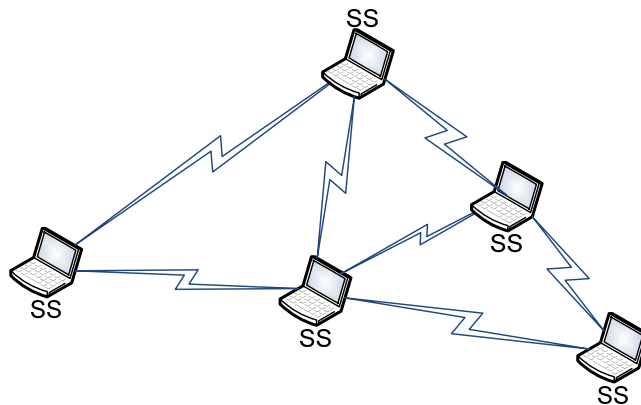


Figura 2.3: Topologia *Mesh*.

2.3. Modelo de Referência do Padrão IEEE 802.16

As propriedades do padrão IEEE 802.16 são distribuídas em três planos, caracterizadas pelo chamado modelo de referência, o qual é uma representação lógica da arquitetura da rede. A Figura 2.4 apresenta o modelo de referência do padrão IEEE 802.16.

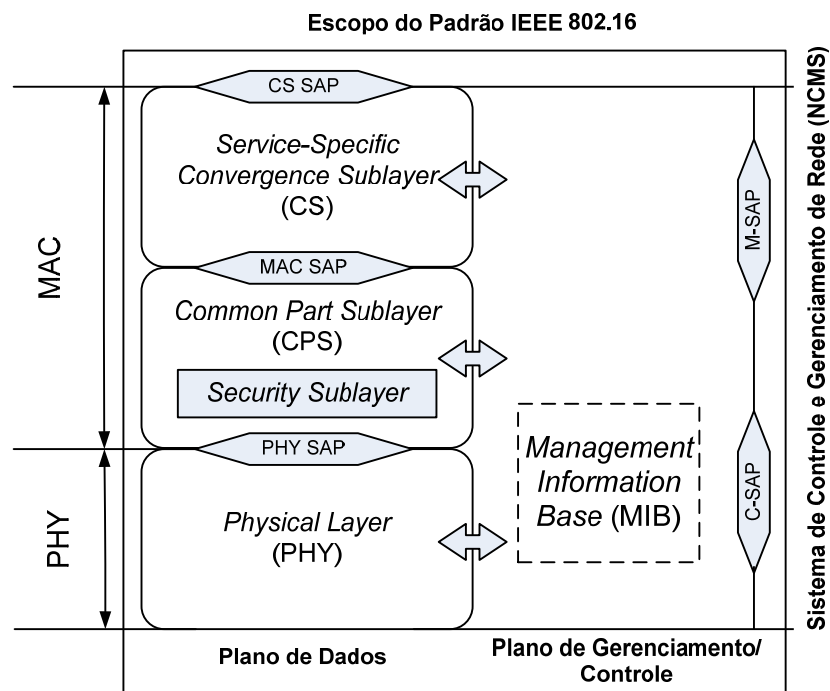


Figura 2.4: Modelo de Referência do Padrão IEEE 802.16 [2].

Os três planos definidos no modelo de referência do padrão IEEE 802.16 são: o plano de dados do usuário, o plano de controle e o plano de gerenciamento [1] [2].

O padrão define a interface aérea, a camada física (PHY) e a camada de controle de acesso ao meio (MAC). A estrutura da camada MAC permite que ela suporte múltiplas especificações de camada física, de acordo com o ambiente operacional específico. A camada MAC prove inteligência para a camada física.

As transferências de dados, mensagens de controle e estatísticas são trocadas entre as camadas através dos SAP (*Service Access Point*).

A MIB (*Management Information Base*) é uma base de informações de gerenciamento. Informações para a provisão dos fluxos de serviços são armazenadas em tabelas [8].

O sistema de controle e gerenciamento de rede (NCMS – *Network Control and Management System*) é responsável por gerenciar a interface de troca de mensagens de controle e gerenciamento entre os elementos do sistema (BS e SS). Dentro de grandes redes, é possível atribuir uma arquitetura mais flexível para as camadas MAC e PHY definidas pelo padrão IEEE 802.16 em relação às camadas de transporte e outros protocolos, através da entidade de controle e gerenciamento independente NCMS. Assim, entidades de sistemas de gerenciamento de rede gerais podem executar funções através da NCMS e protocolos de gerenciamento padrão podem ser implementados na NCMS como o protocolo SNMP (*Simple Network Management Protocol*) [2].

2.3.1. A Camada MAC

De acordo com o modelo de referência apresentado na Figura 2.4, a camada MAC é dividida em três subcamadas: CS (*Service-Specific Convergence Sublayer*), CPS (*Common Part Sublayer*) e Subcamada de Segurança (*Security Sublayer*) [1] [2].

De uma maneira geral, a camada MAC é responsável pelo controle de acesso ao meio, pela adaptação do tráfego proveniente de outras técnicas de transmissão, pelo estabelecimento e manutenção de conexões, pelo ajuste adaptativo das técnicas de transmissão e, pela provisão de QoS através de mecanismos como o controle de admissão, o mecanismo de requisição de largura de banda, o mecanismo de escalonamento, dentre outras funcionalidades [1] [2].

A camada MAC é orientada à conexão. Na arquitetura PMP, uma conexão é um enlace lógico unidirecional entre a camada MAC da BS e a camada MAC da SS [9]. As

conexões estabelecidas podem ser conexões de controle e gerenciamento, cujas mensagens são de controle e gerenciamento, ou conexões de dados, cujas mensagens transmitidas são os dados das aplicações.

Antes de serem transmitidas, as mensagens são encapsuladas dentro de um quadro da camada MAC definido pelo padrão IEEE 802.16. O tipo de mensagem a ser transmitida é determinado pelos valores dos campos presentes no cabeçalho do quadro MAC, apresentados na próxima Subseção.

2.3.1.1. MAC PDU: O Quadro da Camada MAC

Para compreender as principais propriedades da camada MAC, definida pelo padrão IEEE 802.16, é importante conhecer a estrutura de seu quadro. Portanto, nesta Seção são apresentados os principais campos e seus valores que compõem o quadro da camada MAC.

O quadro ou *frame* da camada MAC é chamado de MAC PDU (*MAC Protocol Data Unit*), mostrado na Figura 2.5, e é responsável pela troca de dados entre as camadas MAC da BS e da SS. Pode atingir um tamanho máximo de 2048 *bytes* e é constituído de um cabeçalho de tamanho fixo de 6 *bytes*, um *payload* que é opcional e apresenta tamanho variável e ainda um código de redundância cíclica (CRC – *Cyclic Redundancy Check*) opcional cujo tamanho é de 4 *bytes*. Se o *payload* estiver presente, ele pode apresentar subcabeçalhos, além de MAC SDUs (*MAC Service Data Units*) ou fragmentos de MAC SDUs.

Cabeçalho (6 Bytes)	Payload (Opcional) (Tamanho variável)	CRC (4 Bytes)
------------------------	--	------------------

Figura 2.5: Formato da MAC PDU [1] [2].

O padrão IEEE 802.16 define dois tipos de cabeçalhos da MAC PDU diferenciados de acordo com o valor do campo HT (*Header Type*) [1] [2]:

Cabeçalho Genérico (HT = 0): O quadro é composto de *payload*, que pode conter dados, ou mensagens de gerenciamento da camada MAC;

Cabeçalho de Requisição de Banda (HT = 1): O quadro não possui *payload* e é constituído apenas pelo próprio cabeçalho.

A MAC PDU das mensagens transmitidas através do canal de *downlink* (DL) utiliza apenas o primeiro tipo de cabeçalho, enquanto que a MAC PDU das mensagens transmitidas através do canal de *uplink* (UL) pode utilizar ambos os tipos de cabeçalho.

2.3.1.1.1. Cabeçalho Genérico da MAC PDU

A Figura 2.6 apresenta os campos que constituem o cabeçalho genérico da MAC PDU.

HT = 0 (1)	EC (1)	Type (6)	RSV (1)	CI (1)	EKS (2)	RSV (1)	LEN MSB (3)
LEN LSB (8)			CID MSB (8)				
CID LSB (8)			HCS (8)				

Figura 2.6: Composição do Cabeçalho Genérico da MAC PDU [1].

O campo HT determina o tipo de cabeçalho. O campo EC (*Encryption Control*) é composto por 1 *bit*, se seu valor for igual a “um” (EC = 1), isto indica que o *payload* será criptografado, e se o valor for igual a “zero” (EC = 0), ele não será. O valor do campo *Type* determina o tipo de carga presente no *payload*, adicionando características especiais de acordo com os subcabeçalhos que são inseridos imediatamente após o cabeçalho MAC genérico, como apresentado na Figura 2.7.

As características adicionadas ao quadro MAC de acordo com o valor de *Type* são [1]:

- Se *Type* = 0, o *payload* é composto por um subcabeçalho de requisição de largura de banda do tipo *piggybacking*, que consiste de uma forma opcional de solicitar banda juntamente com os pacotes de dados, evitando a necessidade de transmissão de um quadro completo para solicitação de largura de banda;
- Se *Type* = 1, o *payload* possui um subcabeçalho de empacotamento para reunir várias MAC SDUs de uma mesma conexão em uma única MAC PDU, ou seja, trata-se de uma agregação de pacotes no nível MAC;

- Se *Type* = 2, o *payload* possui um subcabeçalho de fragmentação utilizado para controlar o processo de fragmentação, onde cada MAC SDU pode ser fragmentada e transmitida de maneira independente, como mostrado na Figura 2.7, que ilustra a disposição de MAC SDUs fragmentadas compondo o *payload* de uma MAC PDU;
- O valor de *Type* = 3 é utilizado para indicar uma expansão do subcabeçalho de empacotamento ou do subcabeçalho de fragmentação no *payload*;
- O valor de *Type* = 4 é utilizado para requisitar a retransmissão de um ou mais MAC SDUs, que sofreram danos em virtude de algum erro de transmissão, através da técnica ARQ (*Automatic Repeat reQuest*);
- O valor de *Type* = 5 indica a existência de um subcabeçalho da topologia *Mesh* no *payload*.

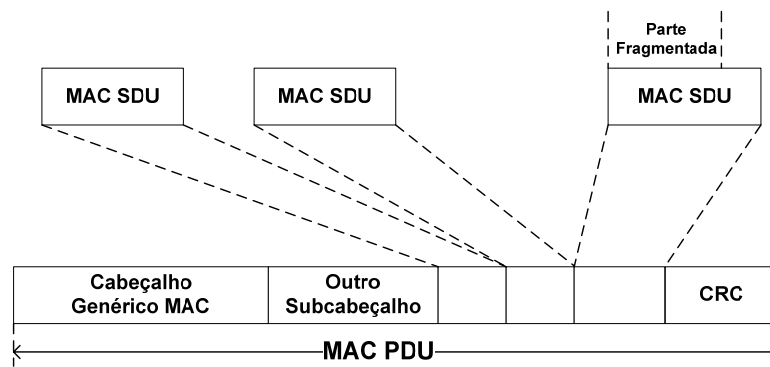


Figura 2.7: Composição Eventual da MAC PDU [10].

O campo RSV (*Reserved*) composto por 1 *bit* é reservado para uso futuro. O campo CI (CRC *Indicator*) informa se existe ou não um código CRC no final do quadro, se CI é igual a um, significa que existe, e se for nulo, não existe. O campo EKS (*Encryption Key Sequence*) de 2 *bits* indica qual chave foi utilizada na criptografia. O campo LEN (*Lenght*) determina o tamanho total da MAC PDU, incluindo o cabeçalho e o CRC. O campo CID (*Connection Identifier*) de 16 *bits* é o identificador único de cada conexão atribuído pela BS, e, finalmente, o campo HCS (*Header Check Sequence*) de 8 *bits* é utilizado para detecção de erros no cabeçalho [1].

O quadro de dados que utiliza cabeçalho genérico (HT = 0) pode conter subcabeçalhos agregados ao *payload* que são transmitidos na conexão de dados. O *payload* do quadro de gerenciamento, que também utiliza o cabeçalho do tipo genérico (HT = 0), contém mensagens

de gerenciamento MAC transmitidas nas conexões de gerenciamento utilizando codificação TLV (*Type/Lenght/Value*), que são especificadas em dezenas de tipos, de acordo com as recomendações do padrão específico.

Como as mensagens de controle e gerenciamento MAC não são enviadas através de conexões de transporte, não é possível haver dúvida em relação ao conteúdo do *payload* de um quadro MAC com cabeçalho genérico ($HT = 0$). Pois, se for recebido através de uma conexão de controle e gerenciamento, o *payload* será composto por mensagens de controle e gerenciamento e, se for recebido através de uma conexão de transporte, o *payload* será composto por dados do usuário [11].

2.3.1.1.2. Cabeçalho de Requisição de Banda da MAC PDU

As MAC PDUs que utilizam este tipo de cabeçalho servem apenas para requisitar largura de banda para uma determinada conexão e por isso não carregam *payload*. A Figura 2.8 apresenta os campos que constituem o cabeçalho de requisição de banda da MAC PDU.

HT = 1 (1)	EC = 0 (1)	Type (3)	BR MSB (11)	
BR LSB (8)			CID MSB (8)	
CID LSB (8)			HCS (8)	

Figura 2.8: Composição do Cabeçalho de Requisição de Banda da MAC PDU [1].

O valor igual a um do campo HT caracteriza o tipo de cabeçalho de requisição de banda, e devido ao fato de não carregar *payload* em seu quadro, o campo EC sempre apresentará valor nulo, uma vez que não há necessidade de criptografia de dados. O campo *Type*, composto por 3 *bits* determina o tipo de requisição de largura de banda. Se *Type* é igual a zero, significa que a requisição de banda será incremental. Se *Type* é igual a um, a requisição será agregada. O campo BR (*Bandwidth Request*) composto por 19 *bits* indica a

quantidade de *bytes* requisitados pela SS. O campo CID de 16 *bits* é o identificador único de conexão atribuído pela BS, que identifica a que conexão pertence a MAC PDU, e o campo HCS de 8 *bits* é utilizado para detecção de erros no cabeçalho.

2.3.1.2. Subcamadas da Camada MAC

A camada MAC é dividida em três subcamadas: Subcamada de Convergência Específica (CS) e Subcamada da Parte Comum (CPS), que serão descritas nas próximas Subseções, e a Subcamada de Segurança (*Security Sublayer*) [1] [2].

2.3.1.2.1. Subcamada de Convergência Específica (CS)

A Subcamada de Convergência Específica ou CS (*Service-Specific Convergence Sublayer*), por fazer fronteira diretamente com a camada superior à camada MAC é responsável por receber as PDUs das camadas acima dela e efetuar a adaptação do tráfego proveniente de outras técnicas de transmissão.

As PDUs recebidas através do CS SAP da subcamada CS são classificadas ou mapeadas para uma conexão, encapsulados em MAC SDUs e repassados a subcamada CPS através do MAC SAP e, posteriormente são entregues à entidade par conforme os requisitos de QoS associados com as características do fluxo de serviço de uma conexão particular. A subcamada CS também é responsável por opcionalmente executar a função de supressão de cabeçalho do *payload* (PHS – *Payload Header Suppression*), eliminando as partes repetidas do cabeçalho na origem e recuperando-as no destino, reduzindo o percentual de *overhead* e, consequentemente aumentando a eficiência na transmissão de dados [1] [2].

O classificador de pacotes da subcamada CS utiliza um critério de associação para mapear os pacotes que entram nas redes IEEE 802.16 que se baseia em atributos de protocolo como, por exemplo, o endereço IP de origem, o endereço IP de destino e o tipo de tráfego definido para a aplicação (CBR, VBR, etc); além de uma prioridade definida pelo classificador e a referência a um CID. Aqueles pacotes que estiverem dentro dos critérios de mapeamento especificados serão encaminhados na conexão definida pelo CID. Se os atributos de algum pacote não atenderem aos critérios definidos nos classificadores, o pacote será rotulado com um CID de valor padrão e encaminhado em uma conexão padrão, ou ainda poderá ser descartado.

Como as camadas superiores a MAC podem apresentar tecnologias específicas, por exemplo, baseada em células ATM (*Asynchronous Transfer Mode*) ou baseada em pacotes, a função de convergência da CS precisa ser específica de acordo com a técnica da camada superior. Portanto, duas especificações foram definidas pelo padrão IEEE 802.16: a Convergência de Pacotes (*Packet CS*) e a Convergência ATM (ATM CS) [1] [2] [12].

A Convergência de Pacotes oferece suporte aos protocolos baseados em pacotes como, por exemplo, IP (*Internet Protocol*) e PPP (*Point-to-Point Protocol*). As PDUs recebidas da camada superior são mapeados ou classificadas para uma conexão apropriada, encapsuladas em MAC SDUs e, então, associadas com um fluxo de serviço, bem como a um CID e a um SFID (*Service Flow Identifier*), facilitando a entrega das MAC SDUs com as características adequadas de QoS para o MAC SAP da entidade par.

A Convergência ATM associa diferentes serviços ATM com o MAC SAP, ou seja, suporta a convergência das PDUs produzidas pela entidade da camada ATM. As células ATM são mapeadas para uma determinada conexão associada a um fluxo de serviço, de acordo com os seus valores de VPI (*Virtual Path Identifier*) e do mecanismo de comutação por caminho virtual VP (*Virtual Path*), ou de acordo com os seus valores de VPI e VCI (*Virtual Channel Identifier*) e do mecanismo de comutação por canal virtual VC (*Virtual Channel*). O classificador da ATM CS possui um conjunto de critérios de casamento como, por exemplo, do par VPI e VCI com um CID de referência, e quando uma célula atende a esses critérios, as MAC SDUs podem ser repassadas para a CPS através do MAC SAP, e posteriormente são entregues para a conexão com o CID correspondente.

2.3.1.2.2. Subcamada da Parte Comum (CPS)

A subcamada da Parte Comum, CPS, é a principal subcamada da MAC em virtude das funções executadas como o controle de acesso ao sistema, o gerenciamento das conexões (estabelecimento e manutenção), a construção da MAC PDU, o suporte à camada física através do ajuste adaptativo das técnicas de transmissão, pela multiplexação dos fluxos de tráfego em conexões e pela provisão de QoS através do mecanismo de requisição de largura de banda, mecanismo de Controle de Admissão de Conexões (CAC), mecanismo de escalonamento e alocação dinâmica de recursos.

A camada MAC é orientada a conexão, o que significa que, para um usuário entrar na rede e iniciar a transmissão de dados, é necessário que a SS e a BS troquem informações de

gerenciamento e controle, garantindo que todos os parâmetros para gerenciar a conexão de forma adequada sejam informados, sendo possível mapear todos os serviços para uma conexão. Cada conexão é caracterizada por um identificador CID de 16 *bits* e cada uma das SSs possui um endereço MAC único de 48 *bits* que a identifica e é utilizado para registrá-la e autenticá-la na rede. Todo o tráfego, incluindo o tráfego não orientado a conexão, é mapeado para uma conexão [1] [2].

No momento em que uma SS entra na rede, três pares de conexões de gerenciamento, um em cada sentido (*downlink* e *uplink*), são estabelecidas [1] [2]:

Conexão de gerenciamento básica: é utilizada para transferência de mensagens MAC de tempo crítico curta e para mensagens de controle de enlace de rádio (RLC – *Radio Link Control*), que são estabelecidas pelas estações durante o período de tempo definido para o campo *Initial Ranging*, no início do *uplink subframe*;

Conexão de gerenciamento primária: é utilizada para transferência de mensagens mais longas e mais tolerantes ao atraso como, por exemplo, as mensagens de configuração e de autenticação da conexão, que são estabelecidas pelas estações durante o período de tempo definido para o campo *Initial Ranging*, no início do *uplink subframe*;

Conexão de gerenciamento secundária: é utilizada para a transferência de mensagens de gerenciamento tolerantes ao atraso baseadas em protocolos (DHCP – *Dynamic Host Configuration*, TFTP – *Trivial File Transfer Protocol* e SNMP) que são estabelecidas pelas SSs durante o período de registro de uma estação na rede.

O acesso de uma SS ao sistema é gerenciado pela camada MAC, especificamente pela subcamada CPS. O processo de inicialização ou entrada de uma estação na rede de topologia PMP pode ser basicamente descrito em três etapas [1] [2] [7] [13] [14]:

Aquisição de Canal (*Scanning*): Processo de inicialização ou de *setup* executado pela camada MAC que permite à SS rastrear uma lista de frequências de canais de operação. A busca da frequência termina assim que o quadro PHY é identificado pela SS. A sincronização é completada quando a SS recebe uma mensagem de gerenciamento DL-MAP (*Downlink Map*), que determina o mapa da disposição dos *slots* no *downlink subframe* correspondentes às respectivas SSs, e o DCD (*Downlink Channel Descriptor*), que descreve os esquemas de modulação e codificação que podem ser utilizados pela SS no canal físico. A SS busca essas informações periodicamente, permanecendo sincronizada enquanto as estiver recebendo;

***Initial Ranging*:** Após a sincronização, inicia-se a fase chamada de *Initial Ranging*. A SS espera por uma mensagem de gerenciamento UCD (*Uplink Channel Descriptor*), que descreve os parâmetros de transmissão de todos os canais de *uplink* disponíveis. Nessa fase, a

SS envia uma rajada de potência mínima, e a incrementa até obter uma resposta, determinando a potência de transmissão e corrigindo os indicadores de tempo, dia e hora. Durante esse ajuste do alcance inicial, onde a BS consegue estimar a sua distância à SS, a SS procura o *Initial Ranging Interval* dentro do UL-MAP, que determina o período cujas SSs disputam o direito de enviar mensagens de requisição para entrada na rede (RNG-REQ). Caso a requisição enviada pela SS seja aceita pela BS, a SS recebe a mensagem de confirmação (RNG-RSP) durante o mesmo período. Para completar essa etapa, acontece a negociação das capacidades, onde a SS finalmente encaminha uma mensagem de requisição de capacidades que determina os níveis de modulação, esquemas de codificação e taxas suportadas; que então são analisados pela BS e podem ser aceitos ou rejeitados;

Autenticação e Registro da SS: Após o processo de negociação das capacidades ocorre a autenticação (autorização) e o registro da SS. Durante a fase de autenticação ocorre a troca de chaves de segurança, certificação e suporte à criptografia entre a SS e a BS. Após a fase de autenticação a SS envia uma mensagem de requisição de registro da conexão para a BS. A BS verifica a identidade da SS, e caso a SS seja admitida na rede, a BS responderá à requisição de registro através de uma mensagem *Authorization Reply* (Resposta de Autorização) que contém uma chave de autorização (AK – *Authorization Key*) codificada com a chave pública da SS. Assim, é efetuado o registro da SS na rede.

Após a etapa de autenticação e registro, a SS inicializa o serviço DHCP para obter um endereço IP e os demais parâmetros que a permita o estabelecimento da conectividade IP com a BS. Os parâmetros operacionais são obtidos via protocolo TFTP. Posteriormente, ocorre a sincronização temporal, onde a SS e a BS ajustam em comum a data e a hora atuais, utilizando o protocolo especificado pelo RFC 868. Após a transferência dos parâmetros operacionais e todas as etapas anteriormente citadas, finalmente podem ser criadas as conexões de transporte.

As transferências de dados das aplicações entre as entidades da rede são efetuadas através das conexões de transporte. As conexões de transporte são unidirecionais, permitindo uma melhor separação entre os parâmetros de QoS dos tráfegos *downlink* e *uplink* [1] [2]. O encerramento das conexões pode acontecer através de solicitação da rede ou das estações, em virtude da alteração de parâmetros de QoS ou devido ao encerramento do tráfego de pacotes.

2.3.2. A Camada Física

As principais funções executadas pela camada física (PHY) são: o transporte físico dos dados (MAC PDUs) ao longo de quadros (*frames*) sucessivos; a definição de espectro e do método de duplexação; a definição das técnicas de transmissão digital (modulação e codificação); a correção de erros e a construção dos *frames*.

A faixa de frequências de operação da camada física do padrão IEEE 802.16 é de 2 a 11 GHz sem linha de visada (NLOS – *Non Line of Sight*), e de 10 a 66 GHz com linha de visada (LOS – *Line of Sight*). Para garantir a interoperabilidade, a camada física apresenta várias especificações de interface aérea, mostradas na Tabela 2.2.

Tabela 2.2: Interfaces Aéreas Definidas para a Camada Física [1].

Designação	Aplicabilidade	Duplexação
WirelessMAN-SC	10-66GHz	TDD/FDD
WirelessMAN-SCa	Bandas Licenciadas < 11 GHz	TDD/FDD
WirelessMAN-OFDM	Bandas Licenciadas < 11 GHz	TDD/FDD
WirelessMAN-OFDMA	Bandas Licenciadas < 11 GHz	TDD/FDD
WirelessHUMAN	Bandas Não-Licenciadas < 11 GHz	TDD

As especificações baseadas na tecnologia SC (*Single Carrier*) utilizam uma única portadora, enquanto que as tecnologias baseadas em OFDM (*Orthogonal Frequency Division Multiplexing*) utilizam várias portadoras ortogonais entre si. A especificação baseada na tecnologia HUMAN (*High-speed Unlicensed Metropolitan Area Network*) utiliza um esquema de seleção de frequências dinâmico (DFS – *Dynamic Frequency Selection*), que detecta e evita interferências.

A camada física opera em formato de *frames* de tamanho fixo, subdivididos em intervalos de tempo chamados de *slots* físicos (PS – *Physical Slots*). A estrutura do quadro da camada física do padrão IEEE 802.16 é composta de um *downlink subframe* e de um *uplink subframe*. No *uplink subframe* as subdivisões em intervalos de tempo recebem a denominação particular de *mini-slots*, que são agrupamentos de PSs. Embora o tamanho do quadro seja fixo, os *subframes* podem apresentar tamanhos variáveis e adaptativos.

O padrão define dois métodos de duplexação: o TDD (*Time Division Duplexing*) cujos *subframes* ocorrem em instantes de tempo diferentes, não sendo possível realizar transmissões simultâneas nos dois sentidos, e geralmente compartilham a mesma frequência de portadora; e o FDD (*Frequency Division Duplexing*) cujos *subframes* ocorrem simultaneamente no tempo,

mas em frequências de portadoras e canais diferentes. Portanto, o método de duplexação determina a maneira como a divisão do quadro é executada [1] [2] [11] [15].

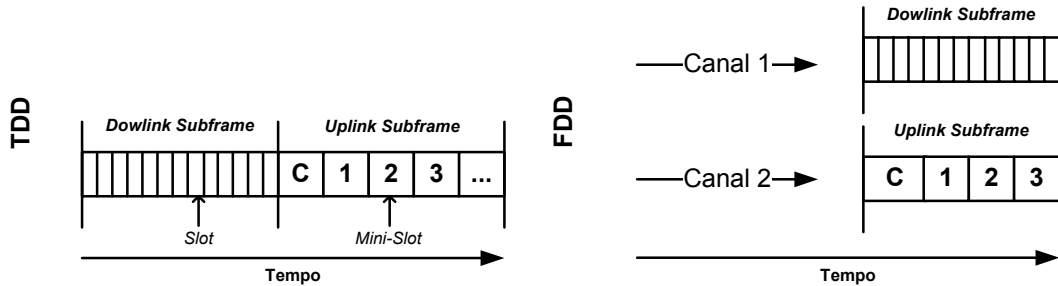


Figura 2.9: Esquema de Transmissão TDD e FDD [1] [2].

No método de duplexação TDD, o canal de *downlink* é um enlace de *broadcast* onde apenas a BS transmite a informação para as SSs do mesmo setor e utiliza basicamente o método TDM (*Time Division Multiplexing*), e o canal de *uplink* é um enlace compartilhado por múltiplas SSs que utilizam o método TDMA (*Time Division Multiple Access*) [16].

2.3.2.1. Quadro de Operação da Camada Física

A Figura 2.10 apresenta a estrutura do quadro TDD, que se trata de um esquema mais apropriado para transmissão em redes definidas pelo padrão IEEE 802.16, cuja característica em relação à utilização de largura de banda entre os canais *downlink* (DL) e *uplink* (UL) é assimétrica. Assim, a técnica TDD permite definir a taxa de transmissão dos canais de *uplink* e *downlink* mais adequadas às necessidades e expectativas de seus usuários.

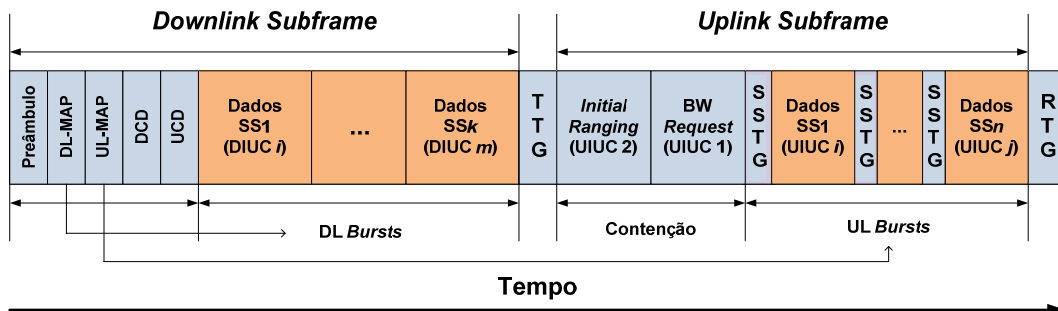


Figura 2.10: Estrutura do Quadro TDD [10].

A primeira parte do quadro TDD, utilizando a especificação OFDM, é representada pelo *downlink subframe*, e a segunda parte é representada pelo *uplink subframe*.

O primeiro campo do *downlink subframe* é o chamado Preâmbulo (FSP – *Frame Start Preamble*) que serve para a sincronização e ajuste do nível de potência (equalização) [17].

Os próximos campos chamados de DL-MAP (*Downlink Map*), ou mapa de *downlink*, e o UL-MAP (*Uplink Map*), ou mapa de *uplink*, definem os apontadores de início das sessões de dados respectivamente no *downlink* e no *uplink*. Portanto, indicam o início do *slot* de tempo de cada rajada de dados.

Os campos DL-MAP e UL-MAP contêm o identificador de canal (CID) e o mapa dos elementos de informação (IE – *Information Elements*) que descrevem o mapeamento das especificações físicas no *downlink* e no *uplink*. Os IEs definem o tempo inicial, o tempo final, o tipo de modulação e a técnica FEC (*Forward Error Correction*) da rajada ou *burst*. Os IEs são baseados em diferentes especificações de interface aérea da camada física (SC, OFDM e OFDMA), que determinam a disposição dos campos DL *Bursts* ao longo dos *subframes*. Por exemplo, na Figura 2.10, foi representada a disposição específica dos campos DL *Bursts* do quadro TDD para a especificação de interface aérea *WirelessMAN* – OFDM operando na arquitetura de rede PMP.

O esquema de mapeamento definido pelos campos de controle DL-MAP e UL-MAP determinam a largura de banda alocada para cada SS através do número de *slots* disponibilizados e indicam o perfil de transmissão a ser utilizado, *Burst Profiles*, que são definidos nos campos subsequentes, o DCD (*Downlink Channel Descriptor*) e o UCD (*Uplink Channel Descriptor*). Portanto, os campos DL-MAP e UL-MAP indicam qual *Burst Profile* uma SS utilizará no *downlink* e no *uplink*, e os campos DCD e UCD trazem as definições das características dos *Burst Profiles* disponíveis.

Após os campos de controle que compõem o *downlink subframe*, segue o campo que apresenta as sessões de dados, transmitidas na forma de rajadas (DL *Bursts*), que podem ter tamanho variável para facilitar o mapeamento dos pacotes de dados. Os DL *Bursts* são transmitidos de acordo com os respectivos perfis de modulação e codificação (*Burst Profiles*), que são diferenciados pelo DIUC (*Downlink Interval Usage Code*) associado ao *downlink* e o UIUC (*Uplink Interval Usage Code*) associado ao *uplink* (para os UL *Bursts*), e podem apresentar diferentes níveis de eficiência e robustez de transmissão.

No final do *downlink subframe* existe um campo chamado TTG (*Transmit/Receive Transition Gap*), que separa o *downlink subframe* do *uplink subframe*, e define um intervalo

de tempo de transição para a BS mudar seu modo como transmissor (*Tx*) para o modo como receptor (*Rx*), e para as SSs mudarem seu modo definido como receptor (*Rx*) para o modo como transmissor (*Tx*) [1] [2].

A segunda parte do quadro TDD é representada pelo *uplink subframe*. O primeiro campo é o período de contenção para permitir oportunidades de acesso às estações de assinantes, chamado de *Initial Ranging*.

O segundo campo do *uplink subframe* também é uma alocação baseada em contenção, mas nesse caso para requisição de largura de banda através de conexões *multicast* ou *broadcast*. O método *truncated binary exponential backoff* é utilizado para controlar as colisões durante o período de contenção.

O último campo do *uplink subframe*, *UL Bursts*, é destinado para que as SSs transmitam seus dados de acordo com os respectivos *Burst Profiles*. Os chamados *UL Bursts*, são separados pelos campos *SSTG* (*Subscriber Station Transition Gap*), que servem para redução de atividade do *burst* prévio, e então sincronizar a BS com a próxima SS que deverá transmitir os dados no *uplink*, dentro do *UL Burst* seguinte.

No final do *uplink subframe* existe o campo *RTG* (*Receive/Transmit Transition Gap*), que define um intervalo de tempo de transição para a BS mudar seu modo como receptor (*Rx*) para o modo como transmissor (*Tx*) e para as SSs mudarem seu modo definido como transmissor (*Tx*) para o modo como receptor (*Rx*) [1] [2].

2.3.2.2. Função de Adaptação do Enlace

A instabilidade das condições do meio sem fio, causada principalmente pelos múltiplos caminhos (*multipath*), pelo rápido desvanecimento (*fading*) e pela refração sofridos pelo sinal, podem prejudicar a qualidade do sinal no enlace, e consequentemente comprometer a conectividade dos nós das redes WiMAX.

Assim, para manter a integridade da comunicação entre a BS e a SS, a camada MAC deve fornecer suporte inteligente à camada física, através do método de adaptação do enlace (*Link Adaptation*), que é baseado no processo de alteração dinâmica do esquema de codificação e modulação (MCS – *Modulation Coding Scheme*), de acordo com a qualidade do sinal no enlace, que é indicada pela relação sinal/ruído (SNR – *Signal-to-Noise Ratio*).

A Tabela 2.3 apresenta os MCS utilizados de acordo com a SNR, e a quantidade de *bytes* não codificados transmitidos por *slot* [2] [12] [18] [19].

Tabela 2.3: MCS Relacionada ao Patamar SNR.

Modulação	Codificação	SNR (dB)	Slot(Bytes)
BPSK	1/2	3	12
QPSK	1/2	6	24
QPSK	3/4	8,5	36
16-QAM	1/2	11,5	48
16-QAM	3/4	15	72
64-QAM	2/3	19	96
64-QAM	3/4	21	108

As técnicas de modulação digital listadas na Tabela 2.3 apresentam as seguintes características [14] [15] [20] [21]:

64-QAM (64 – Quadrature Amplitude Modulation): Codifica seis *bits* por símbolo (6 *bits/ baud*), aplicada para as estações mais próximas da BS.

16-QAM (16 – Quadrature Amplitude Modulation): Codifica quatro *bits* por símbolo (4 *bits/ baud*), aplicada para estações a uma distância média da BS.

QPSK (Quadrature Phase Shift Keying): Codifica dois *bits* por símbolo (2 *bits/ baud*), aplicada para estações distantes da BS.

BPSK (Binary Phase Shift Keying): Codifica um *bit* por símbolo (1 *bit/ baud*), aplicada para aquelas estações mais longínquas em relação a BS.

A taxa de transmissão teórica nas redes IEEE 802.16 depende da largura de banda do canal e do MCS utilizado, pois para cada esquema de modulação existe uma relação diferente entre a eficiência de símbolo, *bits/símbolo*, e símbolos/*slot*, que influencia na quantidade necessária de recursos reservados para a transmissão de dados. Por exemplo, a transmissão utilizando um esquema mais robusto com modulação QPSK com taxa de codificação $\frac{1}{2}$ apresenta a eficiência de 1 *bit/símbolo*, enquanto a transmissão utilizando um esquema mais eficiente com modulação 16-QAM com taxa de codificação $\frac{3}{4}$ apresenta a eficiência equivalente a 3 *bits/símbolo*. Portanto, o esquema mais eficiente permite utilizar menos *slots* do que o esquema mais robusto para transmitir a mesma quantidade de dados.

Os valores da SNR são menores à medida que as distâncias entre a BS e as SSs são maiores, e conseqüentemente o MCS utilizado deve ser mais robusto e menos eficiente, como mostrado na Figura 2.11.

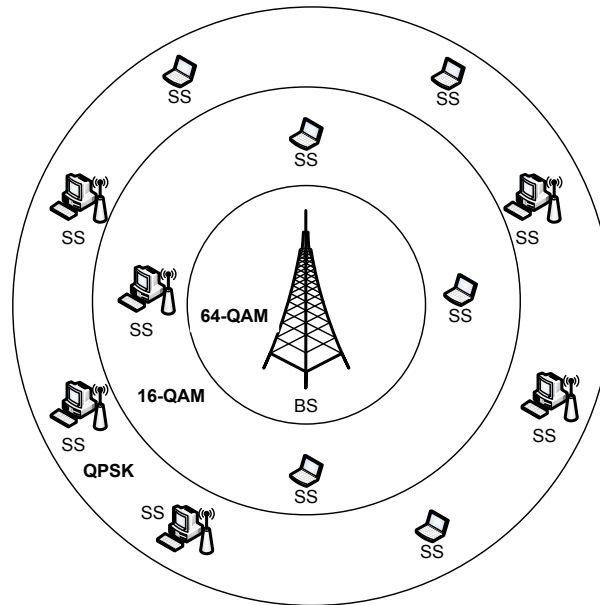


Figura 2.11: Utilização do MCS de Acordo com a Distância da SS à BS [14].

2.4. Provisão de Qualidade de Serviço (QoS) em Redes Sem Fio

Qualidade de serviço, ou simplesmente QoS (*Quality of Service*), se refere a um conjunto de parâmetros estabelecidos nos contratos de nível de serviço (SLA – *Service Level Agreement*) que expressam as expectativas dos usuários [22].

Para que os usuários possam utilizar os serviços da rede é necessário oferecer uma determinada largura de banda. A maneira como a largura de banda é distribuída determina a oferta de QoS da rede.

As redes de acesso banda larga sem fio caracterizam-se pela presença de tráfego heterogêneo e pela limitação da largura de banda, que é compartilhada. Assim, a provisão de QoS nessas redes deve ser garantida através mecanismos inteligentes, capazes de atender os diversos fluxos de tráfego produzidos pelas aplicações de rede, de acordo com suas necessidades ou requisitos.

O nível de satisfação dos usuários depende da eficiência dos mecanismos de QoS desenvolvidos. A avaliação de desempenho desses mecanismos é efetuada em relação a um conjunto de parâmetros de QoS, dentre os quais os mais comuns são [23]:

Taxa de Transmissão: quantidade de dados que podem ser transmitidos por unidade de tempo definida em bps (*bits* por segundo) ou seus múltiplos Kbps e Mbps.

Vazão: quantidade de dados transmitidos com sucesso por unidade de tempo, também definida em bps (*bits* por segundo) ou pelos seus múltiplos Kbps e Mbps.

Atraso: período de tempo transcorrido desde o momento da transmissão da mensagem pelo usuário de origem até o momento em que a mensagem é recebida pelo usuário de destino.

Variação do Atraso (*Jitter*): variação no conjunto de valores de atraso.

Taxa de Erro de Pacotes: relação entre a quantidade de pacotes perdidos (descartados, corrompidos) e a quantidade de pacotes que foi transmitida.

2.5. Provisão de Qualidade de Serviço (QoS) no Padrão IEEE 802.16

O padrão IEEE 802.16 define uma rede BWA com QoS nativa. A provisão de QoS depende da classificação dos pacotes gerados pelas aplicações que passam pela camada MAC e da associação dos fluxos de serviço a uma determinada conexão. Os fluxos são mapeados para as classes de serviço UGS (*Unsolicited Grant Service*), rtPS (*Real Time Polling Service*), nrtPS (*non Real Time Polling Service*) e BE (*Best Effort*), para a provisão de QoS.

A arquitetura do padrão IEEE 802.16, que é responsável pela provisão de QoS, é composta por um conjunto de mecanismos, representados na Figura 2.12.

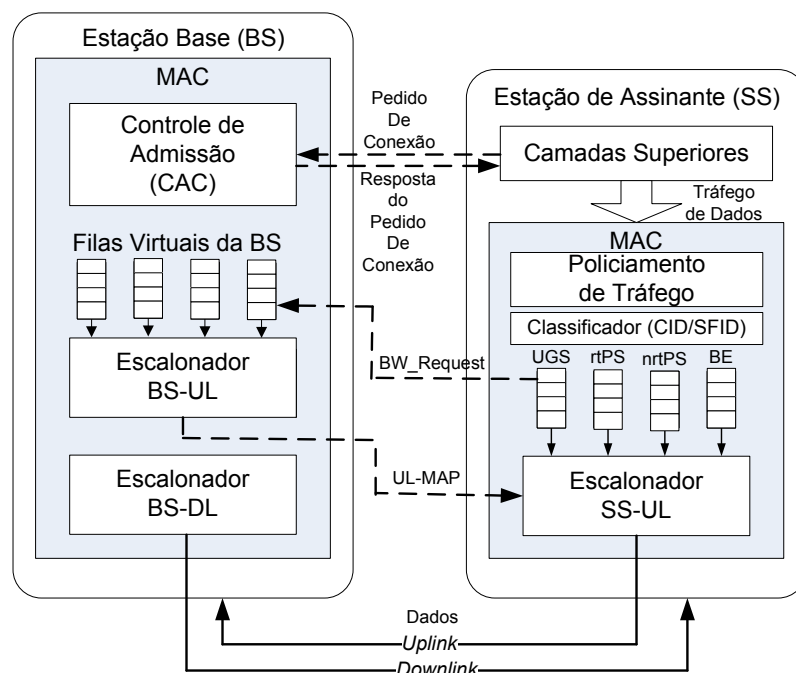


Figura 2.12: Arquitetura de QoS no padrão IEEE 802.16d [24].

Os principais mecanismos da arquitetura de QoS das redes WiMAX são: o mecanismo de policiamento de tráfego na SS, o mecanismo de Controle de Admissão de Conexões (CAC) na BS, o mecanismo de requisição de largura de banda, o mecanismo de alocação de largura de banda e os demais mecanismos de escalonamento presentes na SS e na BS.

Os elementos fundamentais do padrão como, por exemplo, o escalonador, o CAC, a BS, a SS, as classes de serviço, os fluxos de serviço, e os parâmetros de QoS, podem ser representados como objetos na Teoria do Modelo de Objeto, que permite construir um sistema modular implementável por meio de *software*.

Os objetos são representados através de retângulos, mostrados na Figura 2.13, e os atributos, que são sublinhados, identificam de maneira única os objetos aos quais pertencem. Os atributos opcionais estão entre colchetes. A relação entre o número de objetos é indicada no final de cada linha associada ao atributo.

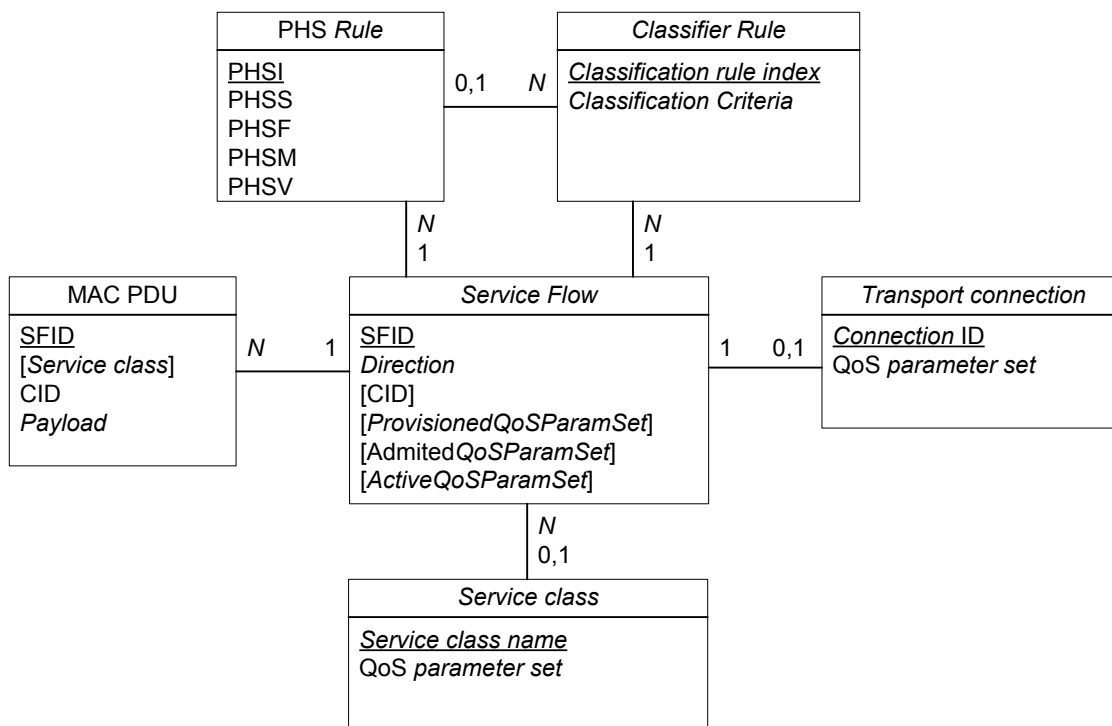


Figura 2.13: Teoria do Modelo de Objetos de Operação [2].

2.5.1. Classes de Serviço

O padrão IEEE 802.16d define quatro classes de serviço para a provisão de QoS [1]:

UGS: suporta aplicações em tempo real com pacotes de tamanho fixo gerados periodicamente, característica do tráfego CBR. Nunca solicita largura de banda periodicamente, pois define sua necessidade durante a fase de *setup*. Exemplo de aplicações: emulação de circuito E1/T1 e VoIP sem supressão de silêncio;

rtPS: suporta aplicações em tempo real para pacotes de tamanhos variáveis gerados periodicamente. Solicita largura de banda respondendo às consultas *unicast* periodicamente efetuadas pela BS. Exemplo de aplicações: *streaming* de áudio e vídeo com codificação MPEG (*Moving Pictures Expert Group*);

nrtPS: suporta aplicações para tráfego armazenado tolerante ao atraso, com pacotes de tamanhos variáveis e requisição de banda periódica. Pode utilizar oportunidades de requisição de banda *unicast*, baseada em contenção e *piggybacking*. A consulta do tipo *unicast* ocorre com menor frequência do que o rtPS, que permite receber oportunidades de requisição de banda mesmo quando a rede estiver congestionada. Exemplo de aplicações: transferência remota de arquivos (FTP);

BE: desenvolvida para suportar o tráfego de melhor esforço com taxa variável. Pode utilizar qualquer mecanismo de requisição de banda. Exemplo de aplicações: *web browsing*.

A partir da extensão IEEE 802.16e foi criada uma quinta classe de serviço [2]:

ertPS (*Extended Real Time Polling Service*): suporta aplicações em tempo real que geram pacotes de tamanhos variáveis periodicamente. A BS pode prover concessão para envio de dados de maneira periódica e não solicitada como na classe UGS, porém as alocações UGS são estáticas e as alocações ertPS são dinâmicas, podendo enviar requisições de banda quando o tamanho do pacote variar. Assim, a estação de usuário pode solicitar a mudança do tamanho da alocação de banda no *uplink*. Pode utilizar consulta do tipo *unicast*, baseada em contenção e *piggybacking* para efetuar requisição de largura de banda. Exemplo de aplicações: VoIP com supressão de silêncio.

Os parâmetros de QoS dessas classes são apresentados na Tabela 2.4.

Tabela 2.4: Parâmetros de QoS das Classes de Serviço [1].

Classes	Parâmetros de QoS
UGS	MRTR=MSTR - <i>Request/Transmission Policy - Tolerated Jitter - Maximum Latency</i>
rtPS	MRTR - MSTR - <i>Request/Transmission Policy - Maximum Latency</i>
nrtPS	MRTR - MSTR - <i>Request/Transmission Policy -Traffic Priority</i>
BE	MSTR - <i>Request/Transmission Policy - Traffic Priority</i>

Os parâmetros de QoS apresentados na Tabela 2.4 são definidos como [1] [4] [15]:

Minimum Reserved Traffic Rate (MRTR): taxa mínima reservada para um serviço, definida em bps;

Maximum Sustained Traffic Rate (MSTR): taxa de pico de informação do serviço, expressa em bps, sem considerar o *overhead* como, por exemplo, do cabeçalho MAC ou do campo CRC;

Maximum Latency: intervalo de tempo máximo entre a recepção de um pacote na subcamada CS da BS ou da SS e o encaminhamento da SDU para a interface aérea;

Tolerated Jitter: define a máxima variação do atraso (*jitter*) para a conexão;

Traffic Priority: especifica a prioridade dos fluxos de serviço que possuem os mesmos parâmetros de QoS e que pertencem à mesma classe de serviço. Aquele que apresentar maior prioridade será contemplado com um menor atraso e preferência na fila do *buffer* do escalonador;

Request/Transmission Policy: prove a capacidade de especificar certos atributos para o fluxo de serviço associado, como a formação da PDU e a restrição na requisição de banda. Por exemplo, se o valor de ***Request/Transmission Policy*** de um fluxo nrtPS for igual a um e relacionado com a consulta baseada em contenção, significa que o fluxo poderá efetuar requisição de largura de banda baseada em contenção.

2.5.2. Fluxos de Serviço

Um fluxo de serviço (*service flow*) consiste de um serviço de transporte unidirecional da camada MAC, caracterizado por um sentido, *downlink* ou *uplink*, e um conjunto de parâmetros que é definido com o objetivo de padronizar a operação entre a BS e a SS.

Assim, quando a BS ou uma SS criam uma conexão, devem associá-la a um fluxo de serviço. Todos os pacotes transmitidos nessa conexão são rotulados por um CID e um SFID univocamente relacionados. Como cada fluxo de serviço é associado a um SFID e um sentido exclusivo, é possível garantir a provisão de QoS nos sentidos *downlink* e *uplink*.

Um fluxo de serviço é caracterizado pelo conjunto de atributos [1] [2] [14] [24]:

SFID (Service Flow Identifier): identificador de fluxo de serviço de 32 *bits*. Todo fluxo de serviço deve possuir pelo menos um identificador SFID e um sentido associado;

CID (Connection Identifier): identificador de 16 *bits* da conexão associada ao fluxo de serviço. Um fluxo de serviço é mapeado para uma conexão de transporte identificada pelo

CID, cuja relação sempre é de um para um, e esse mapeamento ocorre somente quando a conexão tem um fluxo de serviço admitido;

ProvisionedQoSParamSet: conjunto de parâmetros de QoS aprovacionados através de meios externos àqueles definidos no padrão, tal como o sistema de gerenciamento de rede. Indica o limite máximo de recursos que o fluxo de serviço tem à sua disposição [11];

AdmittedQoSParamSet (AQPS): conjunto de parâmetros de QoS para os quais a BS e a SS reservam recursos, dentre eles a largura de banda e o espaço em *buffer* (memória) para viabilizar a ativação de um fluxo;

ActiveQoSParamSet (ACQPS): conjunto de parâmetros de QoS que define o serviço sendo atualmente provido para o fluxo de serviço, ou seja, define exatamente quais os recursos estão sendo utilizados pelo fluxo. A transmissão de pacotes ocorre apenas quando um fluxo de serviço está ativo, portanto, apenas quando o *ActiveQoSParamSet* está definido nas propriedades de um fluxo de serviço;

Authorization Module (Módulo de Autorização): função lógica dentro da BS que deve aprovar ou rejeitar qualquer tipo de mudança dos parâmetros de QoS e classificadores associados a um fluxo de serviço. Ele define um “envelope” que limita os possíveis valores dos conjuntos *AdmittedQoSParamSet* e *ActiveQoSParamSet*.

Existem três tipos de fluxos de serviço [1] [2] [13] [14]:

1. **Provisionado**: fluxo de serviço conhecido pela provisão do sistema de gerenciamento de rede, mas não é imediatamente ativado. O conjunto *ProvisionedQoSParamSet* indica os recursos que podem ser alocados ao fluxo de serviço. Ambos os conjuntos de parâmetros AQPS e ACQPS são nulos.
2. **Admitido**: fluxo de serviço que possui recursos reservados pela BS para o seu conjunto de parâmetros *AdmittedQoSParamSet*, porém esses parâmetros não estão ativos, ou seja, o *ActiveQoSParamSet* é nulo. Possivelmente, os fluxos de serviço admitidos foram provisionados ou sinalizados por algum outro mecanismo.
3. **Ativo**: fluxo de serviço que apresenta recursos reservados pela BS para o seu conjunto de parâmetros *ActiveQoSParamSet*, que não é nulo.

2.5.3. Mecanismos de Provisão de QoS no Padrão IEEE 802.16

A provisão de QoS nas redes WiMAX com base nos requisitos informados pelos fluxos de serviço é função dos mecanismos da arquitetura de QoS do padrão IEEE 802.16.

Os principais mecanismos são: o escalonador, o CAC e o policiamento de tráfego.

O policiamento de tráfego é útil para proteger a rede contra tráfegos mal comportados que violam os parâmetros negociados durante a fase de estabelecimento da conexão (SLA).

O mecanismo de controle de admissão (CAC) restringe o número de usuários simultaneamente ativos na rede, com objetivo de evitar a saturação dos recursos, e consequentemente a degradação do desempenho dos parâmetros de QoS dos fluxos ativos.

O padrão IEEE 802.16 especifica a necessidade de um processo de requisição e concessão (alocação) de largura de banda que consiste em definir e reservar a quantidade adequada de *slots* ou *bytes* para cada conexão transmitir seus dados.

O mecanismo de requisição de largura de banda é caracterizado pelo processo de *polling* ou consulta efetuada pela BS às SSs. A consulta pode ser efetuada a uma SS ou a um grupo de SSs, via *multicast* ou *broadcast*, e visa oferecer largura de banda para que uma SS possa negociar banda para suas conexões. O *polling* é efetuado com base na SS, enquanto que a requisição de largura de banda é efetuada por uma conexão com base em um CID, e a alocação de banda é feita com base na SS. Os tipos de *polling* são [1] [2] [4] [16] [25]:

Polling Unicast: A BS consulta cada SS de maneira exclusiva para identificar se ela necessita utilizar o meio de transmissão e, em caso afirmativo, a BS aloca largura de banda para a SS enviar sua mensagem de requisição (*BW_Request*), como mostrado na Figura 2.14. A desvantagem desse tipo de *polling* é que se a SS não tem dados aguardando para serem transmitidos (*backlog*), ela não responderá a consulta, acarretando em desperdício de banda.

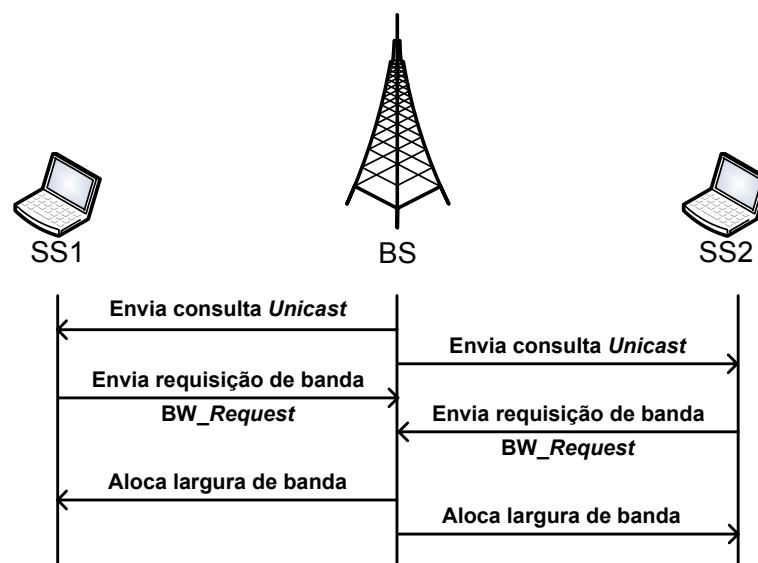


Figura 2.14: *Polling Unicast* [25].

Polling Baseado em Contenção: A BS efetua uma consulta via *multicast* ou *broadcast*. Essa técnica é empregada quando a BS não dispõe de recursos suficientes para consultar cada SS individualmente, e então disponibiliza uma largura de banda onde várias SSs disputam a oportunidade de enviar a sua requisição de largura de banda, como mostrado pela Figura 2.15. Para evitar colisões entre os pedidos simultâneos de várias SSs é utilizado o algoritmo *truncated binary exponential backoff*.

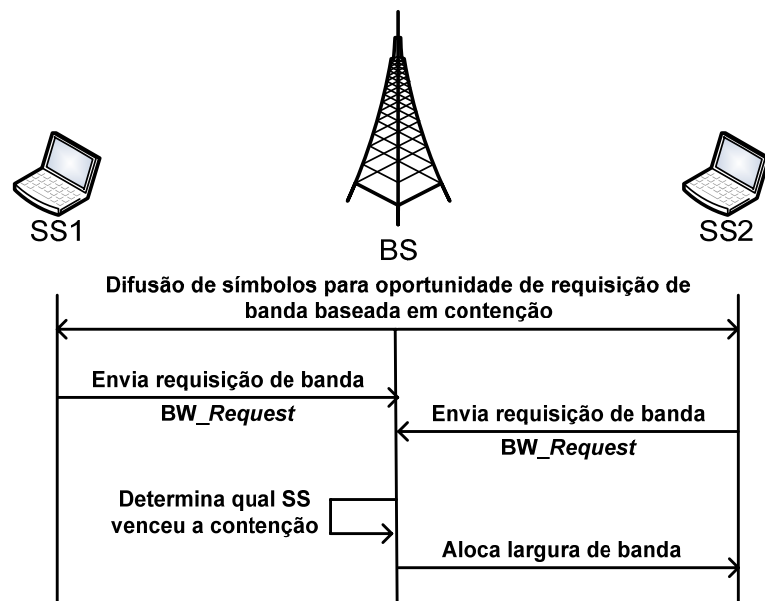


Figura 2.15: *Polling* Baseado em Contenção [25].

A requisição de banda também pode ser enviada pela SS juntamente com um pacote de dados, mas para isso o campo *Type* do cabeçalho genérico da MAC PDU deve ser definido com um valor igual à zero. O valor do campo *Type* igual a zero indica que o *payload* da MAC PDU é composto por um subcabeçalho de requisição de largura de banda do tipo *piggybacking*, chamado de GMSH (*Grant Management Sub Header*) [2].

As requisições de largura de banda podem ser do tipo agregada ou do tipo incremental. A requisição de banda agregada representa um pedido de uma quantidade total de banda necessária para uma conexão, enquanto que a requisição de banda incremental representa uma quantidade de largura de banda adicional [4]. As requisições de largura de banda através de *polling* baseado em contenção devem ser sempre do tipo agregadas devido à possibilidade de colisões e possível perda do pedido de requisição. As requisições de banda do tipo *piggybacking* são sempre incrementais.

A alocação de largura de banda é realizada de acordo com a necessidade das conexões, ou seja, de acordo com a demanda das SSs. Isso significa que, quando uma determinada SS precisa de largura de banda para transmissão de dados, deve encaminhar uma mensagem de requisição de largura de banda, convencionalmente chamada de *BW_Request* ou BR (*Bandwidth Request*), informando para a BS o tamanho da concessão (*grant*) em *bytes* necessária. Posteriormente, a BS recebe a requisição, analisa o pedido e aloca a quantidade de largura de banda possível.

O processo de alocação ou concessão de largura de banda pode ser classificado em GPSS (*Grant Per Subscriber Station*) ou em GPC (*Grant Per Connection*) [1] [14] [24] [26] [27]. No entanto, a revisão mais atual do padrão considera apenas o processo GPSS [2]. No processo GPSS, a BS concede largura de banda para uma SS, considerando a quantidade total de banda requerida por todas as suas conexões, e dessa forma a SS é responsável por redistribuir entre as suas conexões essa banda total que lhe fora concedida. No processo GPC, a BS concede largura de banda para cada conexão de uma SS, o que implica em mais processamento na BS e menos inteligência para as SSs, pois a ordem de entrega de pacotes para cada conexão já é definida pela BS. Portanto, no processo GPC a SS não precisa efetuar o processo de escalonamento localmente, mas quanto mais conexões existirem por SS, maior será o *overhead* nas transmissões em relação ao processo GPSS.

O processo GPSS apresenta uma provisão de QoS mais sofisticada porque a reserva de recursos ou a alocação é realizada para o CID da conexão básica da SS, de registro na rede, e não para o CID de cada conexão presente na SS. Nesse caso, como a SS não sabe exatamente para qual conexão os *mini-slots* foram disponibilizados, o mecanismo de escalonamento implementado na SS deve redistribuir os recursos entre as conexões, a partir de informações mais atualizadas sobre as suas filas, possibilitando atender melhor os requisitos de QoS dos diferentes tráfegos de serviço presentes na SS.

Portanto, o processo GPC é mais vantajoso para redes com um pequeno número de usuários e o processo GPSS é mais adequado para redes com muitos usuários.

Dentre todos os mecanismos para a provisão de QoS na arquitetura IEEE 802.16 implementados na camada MAC, o principal é o escalonador.

O escalonamento do tráfego *downlink* requer um único escalonador na estação base (BS-DL), enquanto o escalonamento *uplink* precisa de dois componentes, um na estação base (BS-UL) e outro na estação de usuário (SS-UL), que são representados na Figura 2.16.

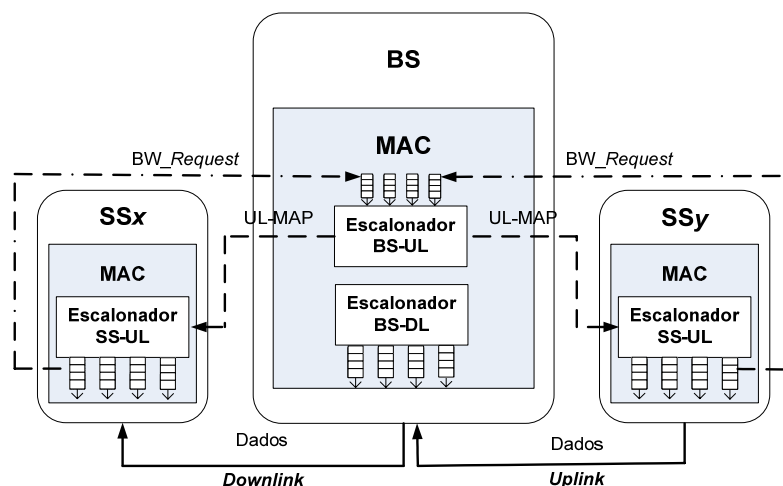


Figura 2.16: Mecanismos de Escalonamento da Rede IEEE 802.16.

O escalonamento no sentido *downlink* é relativamente mais simples do que no sentido *uplink* porque durante o *downlink subframe* a BS é a única a transmitir. Além disso, o escalonador BS-DL tem acesso direto às filas, montando e enfileirando as rajadas de acordo com os parâmetros de QoS dos fluxos, que são transmitidas através de conexões *broadcast* para todas as SSs, que utilizam apenas aqueles pacotes destinados a ela.

O escalonamento no sentido *uplink* utiliza um esquema mais complexo de consulta e concessão, que exige a coordenação da troca de mensagens entre a BS e cada SS individualmente. Além disso, o escalonador BS-UL não tem acesso direto ou conhecimento preciso do tamanho das filas nas SSs, e por isso depende das mensagens de requisição de banda (*BW_Request* ou BR) enviadas pelas SSs, para estimar a necessidade de banda e então efetuar a alocação de largura de banda no *uplink subframe*.

No próximo Capítulo é apresentada uma abordagem mais ampla a respeito do tema escalonamento em redes IEEE 802.16.

2.6. Considerações Finais

Neste capítulo foi apresentada uma alternativa de rede de acesso banda larga sem fio (BWA), comercialmente conhecida por WiMAX, definida pelo IEEE através do padrão IEEE 802.16. Inicialmente foram apresentadas as revisões ou extensões que compõem a chamada família do padrão IEEE 802.16. Também foram apresentados os dois tipos de topologia ou arquitetura de rede de acesso: PMP e *Mesh*.

As camadas do modelo de referência do padrão IEEE 802.16 foram descritas, destacando a estrutura do quadro da camada MAC, os tipos de cabeçalho da MAC PDU, as principais funções das subcamadas de Convergência Específica (CS) e da Subcamada da Parte Comum (CPS). Posteriormente, foram apresentadas as principais características da camada física (PHY), as especificações das interfaces aéreas, as técnicas de modulação digital, a estrutura do quadro utilizada pela camada física e a importância da técnica de adaptação do enlace.

Neste capítulo também foi apresentada a definição de Qualidade de Serviço (QoS) e a sua necessidade de provisão no contexto de redes BWA. Foram apresentados os conceitos básicos a respeito das classes de serviço e de fluxo de serviço, além dos principais mecanismos da arquitetura de QoS do padrão IEEE 802.16.

Capítulo 3

ESCALONAMENTO EM REDES IEEE 802.16

3.1. Introdução

A finalidade de um escalonador é definir a ordem de envio de pacotes presentes nas filas de armazenamento (*buffers*) dos elementos da rede. Em redes WiMAX, os mecanismos de escalonamento são implementados na camada MAC da BS e da SS, e sua lógica de operação é definida através de um algoritmo que pode ser baseado em alguma ou algumas disciplinas de escalonamento.

As primeiras políticas de escalonamento propostas para as redes IEEE 802.16 foram adaptadas das redes cabeadas. Com o aumento da necessidade por acesso banda larga em grande escala e do interesse pelas novas tecnologias BWA, surgiram diversas propostas de mecanismos de escalonamento para as redes IEEE 802.16, baseadas em técnicas mais sofisticadas e eficientes, cujo estado da arte é abordado neste Capítulo.

A estrutura deste Capítulo é composta por: Seção 3.2 apresenta algumas das principais disciplinas de escalonamento encontradas na literatura, Seção 3.3 apresenta o estado da arte acerca dos mecanismos de escalonamento empregados em redes do padrão IEEE 802.16 e, Seção 3.4 onde são apresentadas as considerações gerais sobre o Capítulo.

3.2. Disciplinas de Escalonamento

As primeiras propostas de escalonamento em redes IEEE 802.16 foram baseadas em disciplinas de escalonamento adaptadas das redes cabeadas como, por exemplo, WRR (*Weighted Round Robin*), EDF (*Earliest Deadline First*), etc. Por isso, nesta subseção são abordadas algumas das principais disciplinas de escalonamento encontradas na literatura.

A disciplina GPS (*Generalized Processor Sharing*) pode ser utilizada como uma referência para comparar a eficiência entre as disciplinas de escalonamento práticas [24]. É

uma disciplina ideal não implementável, que serve uma quantidade infinitesimal de dados (*bits*) de cada fila não vazia (*work-conserving*) em um período de tempo finito. Assim, pode alcançar uma justiça perfeita e uma utilização de *frame* completa, porém a fragmentação de pacotes pode resultar na redução do *goodput*, que corresponde aos dados transmitidos sem incluir o *overhead* e a perda de pacotes [28].

A disciplina de escalonamento FIFO (*First In First Out*) é considerada a mais fácil de ser implementada na prática, pois todos os pacotes que chegam são armazenados em uma fila comum cuja ordem de entrega é baseada na ordem de chegada, como mostrado na Figura 3.1. Caso a fila esteja cheia, a chegada de novos pacotes resultará em descartes. Seu comportamento é previsível, pois o atraso é determinado pelo tamanho da fila. Como não é capaz de proteger a rede contra fontes mal comportadas e não toma decisão sobre a prioridade dos fluxos, não é capaz de atender os requisitos de QoS dos diferentes tipos de tráfego [29].

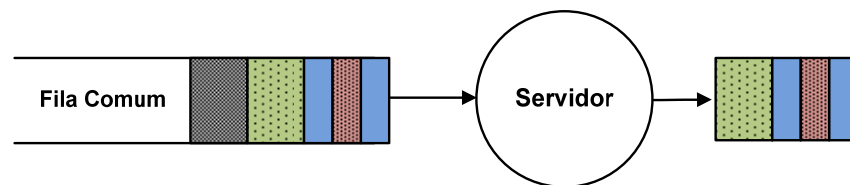


Figura 3.1: Encaminhamento de Pacotes de Acordo com o Escalonador FIFO.

A disciplina PQ (*Priority Queuing*) classifica e armazena os pacotes dos tráfegos de serviço em suas filas, permitindo que sejam atribuídos diferentes níveis de prioridade para atendimento como, por exemplo, alto, médio, baixo e normal, mostrados na Figura 3.2. Pode operar de duas formas: *strict priority* onde os pacotes das filas de menor prioridade só serão transmitidos quando as filas de maior prioridade estiverem totalmente vazias, ou *rate-controlled* onde todos os pacotes da fila de maior prioridade só serão entregues antes das demais filas de menor prioridade se a quantidade de tráfego na primeira estiver abaixo de um patamar pré-determinado [29]. A disciplina PQ é útil quando se deseja fornecer uma garantia de atraso máximo fim-a-fim para um tipo específico de tráfego, e também para o escalonamento entre classes, promovendo tratamento diferenciado. Porém, pode provocar a inanição do tráfego de menor prioridade, e até mesmo nas filas de maior prioridade, determinados fluxos podem sofrer com atrasos inadequados caso um dos fluxos armazenados nessa fila se comporte mal [30].

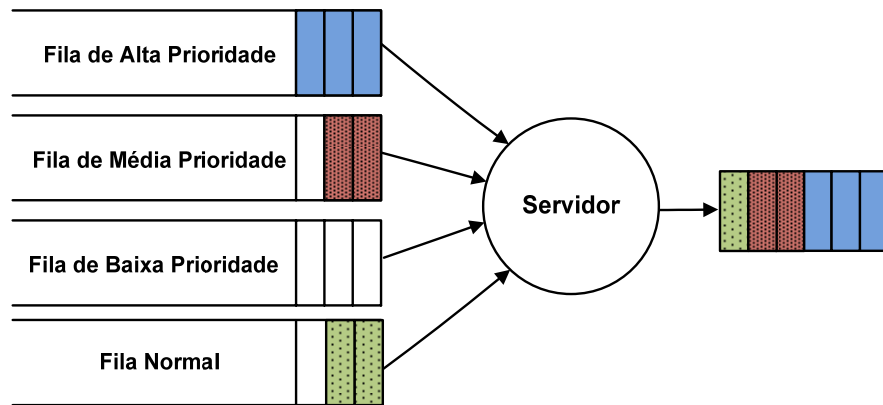


Figura 3.2: Encaminhamento de Pacotes de Acordo com o Escalonador PQ.

O objetivo da disciplina de escalonamento FQ (*Fair Queuing*) é fornecer a mesma quantidade de largura de banda para cada fluxo em um dado período de tempo. Ela classifica os pacotes recebidos pela rede em fluxos específicos e os separa em filas dedicadas. A banda é dividida igualmente em N filas [29]. Isolando cada fluxo em uma fila dedicada, previne-se que um fluxo mal comportado degrade o desempenho dos outros. Porém, não dá suporte para os fluxos com necessidades diferentes que existem nas redes WiMAX.

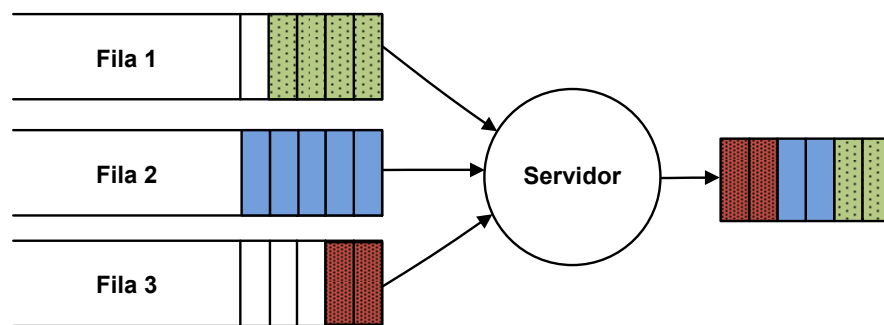


Figura 3.3: Encaminhamento de Pacotes de Acordo com o Escalonador FQ.

A disciplina de escalonamento RR (*Round Robin*) efetua uma varredura cíclica, atendendo um pacote de cada fila a cada rodada [12] [21] [31]. Sua grande vantagem é a simplicidade de implementação computacional. Porém, não é capaz de garantir os requisitos de QoS dos fluxos com necessidades diferentes, e apresenta um comportamento *nonwork-conserving*, pois pode alocar recursos para conexões que não tem nada a transmitir.

A disciplina de escalonamento WRR (*Weighted Round Robin*), também chamada de CBQ (*Class-Based Queuing*), serve rotativamente todas as filas não vazias (*work-conserving*).

Cada fila pode representar um tipo específico de tráfego [7] [32]. A cada ciclo ou rodada, uma quantidade de pacotes de cada fila, proporcional ao valor de seu peso, é selecionada para ser transmitida. O peso pode ser definido em função do comprimento das filas, do atraso dos pacotes ou ainda pode ser associado com o parâmetro de QoS de fluxo de serviço MRTR (*Minimum Reserved Traffic Rate*). Assim, o percentual destinado a cada tráfego pode ser definido como [30]:

$$P_i = \frac{W_i \cdot C}{S} \quad (3.1)$$

Onde W_i é o peso do fluxo i , S é a soma de todos os pesos e C a capacidade do canal.

Portanto, a disciplina WRR pode ser utilizada para prover garantias de largura de banda diferenciada para fluxos com necessidades diferentes, baseado em seus pesos. Porém, é eficiente somente quando os pacotes produzidos por esses fluxos têm tamanhos iguais, pois foi inicialmente desenvolvida para tratar os pacotes de tamanhos fixos das redes ATM [33]. Quando são gerados pacotes de tamanhos diferentes a proporção da divisão do tráfego pode ser imperfeita em relação ao peso definido [34] [35], como mostrado na Figura 3.4.

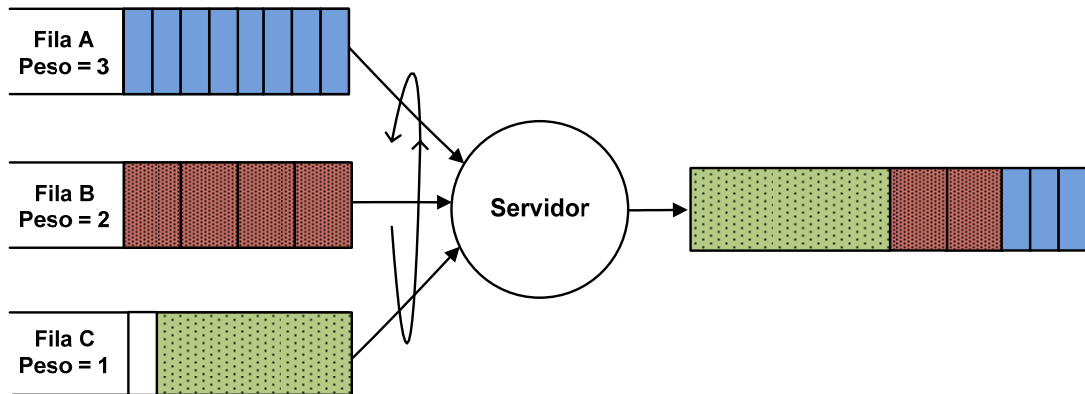


Figura 3.4: Comportamento Inadequado do Escalonamento WRR [35].

O algoritmo da disciplina DRR (*Deficit Round Robin*) associa o valor de uma variável chamada *Quantum* e o valor de outra variável chamada *Deficit Counter* com cada tipo de fluxo i armazenados em filas dedicadas. Então, serve rotativamente todas as filas não vazias (*work-conserving*) de maneira cíclica e o valor do *Quantum* remanescente da rodada anterior é adicionado ao *Deficit Counter* utilizado na próxima rodada. Assim, as filas que não foram

atendidas completamente nas rodadas anteriores são recompensadas nas próximas, de acordo com os valores desses parâmetros [36]. O tamanho do *Quantum* pode ser definido por [28]:

$$Q_i = \text{Capacidade_do_Frame} / \text{Número_de_Fluxos} \quad (3.2)$$

Onde, por exemplo, se a capacidade do *frame* for de 300 *bytes* e o número de fluxos = 4, então $Q_i = 300/4 = 75$. Porém, seu algoritmo é flexível em relação ao valor do *Quantum*, que pode apresentar valores diferentes de acordo com os requisitos de QoS dos fluxos.

A disciplina DRR é adequada para redes com pacotes de tamanhos variáveis porque considera o tamanho dos mesmos. Porém, a latência é influenciada pelo número de fluxos ativos e pelo tamanho dos pacotes. Portanto, não é adequada para escalonar o tráfego das classes de tempo real [24]. Além disso, seu algoritmo precisa da informação a respeito do tamanho exato dos pacotes armazenados nas filas, que o torna inadequado para o escalonador BS-UL das redes IEEE 802.16 [7]. Para minimizar a baixa utilização de *frame* pode ser adaptada para a disciplina DRRF (*Deficit Round Robin with Fragmentation*) [28].

A disciplina WFQ (*Weighted Fair Queuing*) é uma aproximação baseada em pacotes do algoritmo GPS [24] [29]. Os pacotes de fluxos diferentes são armazenados em filas dedicadas, atendidas de maneira cíclica. Para cada fluxo é atribuído um peso, que é utilizado para calcular o tempo associado a cada pacote, o *Finish Time*, que corresponde ao tempo em que um pacote deveria abandonar a fila, segundo o estabelecido por uma disciplina GPS. O *Finish Time* é calculado através das equações a seguir [7]:

$$S_k^i = \max\{F_i^{k-1}, V(a_i^k)\} \quad (3.3)$$

$$\Phi_i = C * W_i \quad (3.4)$$

$$F_i^k = S_k^i + L_k^i / \Phi_i \quad (3.5)$$

Onde S_k^i é o *Start Time* do pacote de índice k da SS de índice i , F_i^{k-1} é o *Finish Time* do pacote $k-1$ da SS i , $V(a_i^k)$ é o tempo virtual, a_i^k é o tempo de chegada do pacote k da SS i , Φ_i é a taxa reservada da SS i , C é a capacidade do canal, W_i é o peso da SS i , F_i^k é o *Finish Time* do pacote k da SS i , e L_k^i é o comprimento do pacote k da SS i .

Portanto, a disciplina WFQ suporta fluxos com requisitos de largura de banda diferentes e, diferente da WRR, pacotes de tamanhos variáveis porque considera o tamanho do pacote e capacidade do canal para distribuir os recursos de maneira justa. Sua grande desvantagem é a complexidade computacional em comparação à de outras disciplinas.

A disciplina de escalonamento EDF (*Earliest Deadline First*) aloca largura de banda de acordo com os requisitos de atraso e por isso é adequada para escalonar o tráfego de tempo real que é mapeado para as classes de serviço UGS e rtPS do padrão IEEE 802.16. O seu algoritmo define um tempo de vida para cada pacote, o *deadline*. Os pacotes que possuem o menor *deadline* são atendidos primeiro. O *deadline* é calculado de acordo com [7] [24] [32]:

$$d_k^i = t_k + D_i \quad (3.6)$$

Onde d_k^i é o *deadline* do pacote de índice k da sessão i , t_k é o tempo de chegada do pacote de índice k , e D_i é o atraso máximo limitado definido para a sessão i .

3.3. Estado da Arte dos Mecanismos de Escalonamento em Redes IEEE 802.16

Na arquitetura de QoS do padrão IEEE 802.16 são definidos três mecanismos de escalonamento: o escalonador da BS de *downlink* (BS-DL), que seleciona diretamente os pacotes armazenados em suas filas para serem transmitidos no enlace de saída da BS para as SSs; o escalonador da SS de *uplink* (SS-UL), que seleciona diretamente os pacotes armazenados em suas filas para serem transmitidos no enlace de saída da SS para a BS; e o escalonador da BS de *uplink* (BS-UL) que, é responsável pela alocação de largura de banda para as SSs no *uplink*. Esses mecanismos são definidos na camada MAC do padrão IEEE 802.16, e sua política de operação é implementada através de um algoritmo de escalonamento.

Para que uma proposta de escalonamento para as redes IEEE 802.16 seja satisfatória, é importante que o seu algoritmo desenvolvido apresente determinadas características como, por exemplo [24] [37] [38] [39] [40] [41] [42]:

Baixa complexidade computacional: Utilizar um algoritmo eficiente que consuma uma pequena parcela de tempo de execução, pois o escalonamento em redes WiMAX é executado a cada *frame*, e podem ser transmitidos centenas de *frames* por segundo;

Garantir o fornecimento da quantidade mínima de recursos: Fornecer pelo menos a quantidade mínima de recursos necessários para todos os usuários, de acordo com os requisitos de cada aplicação, para que a percepção da qualidade de serviço seja adequada;

Utilização eficiente do enlace de comunicação: Evitar o desperdício de recursos, não alocando banda para conexões que não possuem dados para serem transmitidos e nem para aquelas SSs com alta possibilidade de perda de pacotes durante a transmissão;

Distribuir de maneira justa os recursos da rede: Distribuição proporcional dos recursos disponíveis na rede entre todos os fluxos ativos, de acordo com os parâmetros requisitados por cada fluxo;

Isolamento entre os tráfegos diferentes (Proteção): Separação dos tipos de tráfego em filas dedicadas para evitar que fontes mal comportadas de um determinado fluxo afetem o serviço de outros com requisitos críticos;

Flexibilidade e Escalabilidade: Capacidade para atender a qualquer tipo de tráfego, principalmente o tráfego multimídia, além de ser capaz de suportar de maneira eficiente a variação do número de usuários presentes na rede ao longo do tempo.

3.3.1. Classificação dos Algoritmos de Escalonamento Utilizados em Redes IEEE 802.16

O algoritmo de um mecanismo de escalonamento IEEE 802.16 pode ser classificado de acordo com determinadas características estruturais e funcionais [7] [18] [24] [32] [42] em:

Homogêneo: O escalonador utiliza apenas um tipo de disciplina de escalonamento (WRR, DRR, WFQ, etc) para executar sua tarefa;

Híbrido: O escalonador utiliza uma combinação de disciplinas de escalonamento (PQ+EDF+WFQ, etc) com o propósito de executar determinadas tarefas. Pode ser classificado como hierárquico quando apresentar estágios com políticas distintas;

Work-Conserving: O escalonador nunca fica ocioso caso existam recursos disponíveis no canal e pelo menos um pacote em qualquer uma de suas filas. Portanto, os pacotes presentes em alguma das filas do escalonador sempre são transmitidos;

Nonwork-Conserving: O escalonador pode ficar ocioso mesmo com a presença de pacotes aguardando em alguma fila de espera, que só podem ser transmitidos quando o fluxo associado a esses pacotes se tornar elegível, após satisfazer uma determinada condição estabelecida pela lógica do algoritmo;

Channel Aware: Também classificado como oportunista ou *cross-layer*, o escalonador considera as condições do canal como, por exemplo, o nível de potência do sinal, as taxas de erros e de perdas de pacotes, para definir a ordem de transmissão dos pacotes armazenados;

Channel Unaware: O escalonador não considera as condições do canal.

Intra-Class: Algoritmo de escalonamento adequado para reservar recursos entre os fluxos de serviço da mesma classe, de acordo com os requisitos de QoS;

Inter-Class: O escalonador considera o tráfego de cada classe separadamente, adotando políticas de atendimento específicas para satisfazer os parâmetros de cada classe;

Baseado em RR: O escalonador aloca a banda de maneira comum para todos os fluxos a cada varredura cíclica efetuada ao longo de suas filas de espera;

Deadline-Based ou Time-Stamped: Escalonador projetado para oferecer melhor propriedade de atraso, marcando todos os pacotes com uma variável que representa o seu tempo de vida. Apresentam maior complexidade computacional que aqueles baseados em RR;

Rate-Latency: O escalonador é projetado para atender os parâmetros de taxa de serviço ou atraso.

As características supracitadas não são todas auto-exclusivas, pois o algoritmo de um mecanismo de escalonamento baseado, por exemplo, na disciplina WRR, pode ser classificado como: Homogêneo, *Work-Conserving*, Baseado em RR e *Rate-Latency*. Portanto, um determinado algoritmo de escalonamento pode agregar várias dessas características.

3.3.2. Propostas de Mecanismos de Escalonamento para as Redes IEEE 802.16

O padrão IEEE 802.16 define três mecanismos de escalonamento para a provisão de QoS na rede. Muitas das vezes, o objetivo em termos de desempenho desses mecanismos é garantir o atraso máximo limitado, ou a taxa mínima de serviço, ou até mesmo garantir vários desses parâmetros simultaneamente. De qualquer maneira, sua eficiência depende do tipo de algoritmo e do perfil de tráfego na rede onde será empregado, que define quais parâmetros são mais ou menos críticos para a satisfação de seus usuários.

Dentre os três escalonadores definidos pelo padrão, o escalonador BS-UL tem a tarefa mais complicada, pois a implementação de seu algoritmo, além das características apresentadas no início da Seção 3.3, deve ainda considerar os seguintes aspectos [43]:

- A distribuição dos recursos deve ser baseada nas requisições de largura de banda enviadas pelas SSs e nos parâmetros de QoS de cada conexão;
- Além da largura de banda alocada para a transmissão dos dados, deve ser alocada largura de banda para o envio de requisições de largura de banda;
- Deve ser capaz de garantir todos os parâmetros de QoS definidos pelo padrão.

Além disso, o desempenho desse mecanismo depende de diversos parâmetros das redes WiMAX como, por exemplo, o *overhead* do cabeçalho MAC, o efeito do preâmbulo da rajada de *uplink* (*uplink burst preamble*) de acordo com o número de SSs, o comprimento do *frame*, o mecanismo de requisição de banda utilizado [7], a quantidade de *slots* para a contenção utilizados no *uplink subframe* [44], o tipo de camada física utilizada, e a utilização de funções especiais como a fragmentação, o empacotamento, a concatenação e a PHS [45].

Assim, nesta Seção, são apresentadas propostas de mecanismos de escalonamento para as redes IEEE 802.16 que consideram muitos dos aspectos até aqui apresentados.

Em [3] é proposto um mecanismo de escalonamento BS-UL baseado na disciplina PQ (*Priority Queuing*), para uma rede IEEE 802.16 PMP TDD. São utilizadas três filas de prioridade, mostradas na Figura 3.5.

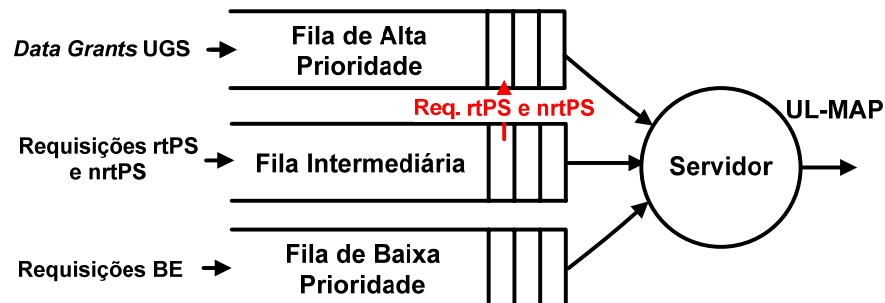


Figura 3.5: Arquitetura do Escalonador Proposto em [3].

A fila de alta prioridade armazena os *data grants* UGS e as requisições rtPS e nrtPS de maior prioridade. A fila intermediária é composta de requisições rtPS e nrtPS, armazenadas em ordem decrescente de prioridade, atribuída de forma dinâmica de acordo com o *deadline* das requisições rtPS, e com a diferença entre a taxa atual fornecida e a taxa mínima requisitada para os fluxos das requisições nrtPS e rtPS, que podem migrar para a fila de alta prioridade para garantir a provisão de seus requisitos de QoS. A fila de baixa prioridade é composta de requisições BE e somente é atendida quando as filas de maior prioridade já estiverem vazias.

O trabalho proposto por Sayenko, A. et al. [46] em relação ao mecanismo de escalonamento BS-UL, preocupa-se em alocar largura de banda de acordo com os requisitos de QoS, com o tamanho das requisições de banda das conexões, e com os parâmetros das redes WiMAX. O mecanismo de alocação de banda é modelado a partir de uma adaptação da disciplina de escalonamento WRR. O esquema proposto define basicamente três estágios:

1. Alocação de um número mínimo de *slots* para cada conexão a fim de garantir seus requisitos básicos de QoS como, por exemplo, a banda mínima requisitada;
2. Alocação de *slots* não utilizados no *frame* para o excedente de banda requisitada pelas conexões, com o objetivo de garantir um comportamento *work-conserving*.
3. Reordenação (intercalação) dos *slots* a serem transmitidos para minimizar o valor máximo do *jitter* e do atraso máximo, sem afetar a taxa de transmissão.

O *overhead* do cabeçalho e dos subcabeçalhos MAC é calculado para permitir uma alocação mais adequada da quantidade de *slots* para cada conexão. Para testar a eficiência desse mecanismo foram modelados três cenários específicos. No primeiro cenário é verificado se a distribuição dos recursos na rede para as conexões da mesma classe de serviço é justa, no segundo cenário é verificado como o escalonador aloca os recursos da rede entre as conexões de classes diferentes, e por fim, no terceiro cenário são analisados os parâmetros de QoS para SSs que utilizam esquemas com modulação diferentes.

O trabalho apresentado por Rodrigues, C. [12] foi baseado na adaptação do mecanismo de alocação de largura de banda proposto por Sayenko, A. et al. [46], para operar com um modelo de camada física OFDMA-PUSC (PUSC - *Partial Usage of Subchannels*).

O trabalho apresentado por Wongthavarawat, K. & Ganz, A. [47] é baseado na implementação de uma política de controle de admissão (CAC) e de um escalonador BS-UL nomeado de UPS (*Uplink Packet Scheduling*). O UPS utiliza uma combinação de disciplinas de escalonamento PQ *strict priority*, EDF e WFQ cuja hierarquia é apresentada na Figura 3.6. A disciplina PQ garante prioridade no atendimento das classes tempo real e as políticas específicas em cada fila dedicada oferecem tratamento diferenciado de acordo com os requisitos de cada classe. A operação da arquitetura de QoS implementada nesse trabalho pode ser descrita pelos seguintes passos:

1. Uma SS tenta estabelecer uma conexão com a BS através da troca de mensagens de sinalização, e a aplicação na SS encaminha seus parâmetros de QoS necessários;
2. O CAC avalia o pedido de conexão baseado nos parâmetros de QoS informados na etapa anterior e permite ou não o estabelecimento da conexão entre a SS e a BS;

3. Uma vez que a conexão é admitida pelo CAC, esse mecanismo interage com o UPS da BS e informa os parâmetros de *token bucket* para o mecanismo de policiamento da SS;
4. O mecanismo de policiamento de tráfego monitora o tráfego gerado pela SS, baseado nos parâmetros definidos no SLA daquela conexão informados nas etapas anteriores;
5. A cada *frame* o Módulo de Informação do UPS coleta a informação a respeito do tamanho das filas presentes nas SSs, de acordo com as mensagens *BW_Request* recebidas no *frame* anterior, e então processa essa informação, atualizando o Módulo de Base de Dados. Também calcula o *deadline* dos pacotes dos fluxos rtPS, que é utilizado como parâmetro de prioridade para garantir o requisito de latência máxima;
6. O Módulo de Atribuição de Serviço recupera a informação presente no Módulo de Base de Dados da BS, e aloca banda em termos de unidades de *slots* usados nos IE (*Information Elements*) do UL-MAP. Assim, o UL-MAP gerado pela BS é entregue para todas as SSs via *broadcast* no *downlink subframe*;
7. As SSs decodificam o UL-MAP para encontrar a quantidade de *slots* alocados para a transmissão dos pacotes de suas conexões. Após decodificar o UL-MAP, o escalonador da SS de *uplink* transmite os pacotes nos *slots* de tempo predefinidos.

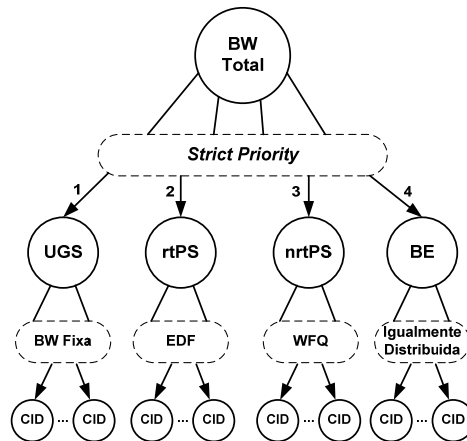


Figura 3.6: Estrutura Hierárquica do UPS Proposto por [47].

O mecanismo de escalonamento BS-DL proposto por Lera, A. et al. [39], para uma rede do padrão IEEE 802.16 PMP, utiliza um algoritmo do tipo *channel aware* e a disciplina de escalonamento WF²Q+ (*Worst-case Fair Weighted Fair Queueing* +). Esse esquema determina que apenas um grupo de fluxos classificados como elegíveis, cuja qualidade do sinal entre a BS e a SS é adequada, tenham permissão para transmitir seus pacotes. A

quantidade de SSs desse grupo varia com o tempo, de acordo com a qualidade do sinal monitorado no canal. Seus objetivos são oferecer a diferenciação de serviços de acordo com a QoS e justiça para o tráfego *downlink*. Porém, apresenta alta complexidade computacional.

Bai, X. et al. [48] propõem um mecanismo de alocação de largura de banda *cross-layer* PHY-MAC para redes IEEE 802.16 SC PMP, cujo objetivo é prover garantias de QoS a nível de conexão, otimizar o tempo de percepção pela BS da necessidade de banda para cada conexão (*refresh*) e minimizar o *overhead* das requisições de banda. A estrutura proposta opera via colaboração de dois blocos funcionais, representados na Figura 3.7, chamados de URMA (*Uplink Request Management Agent*) na SS e FSU (*Frame Scheduling Uplink*) na BS.

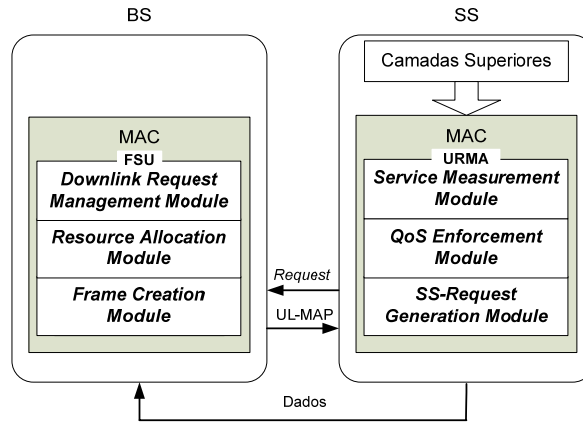


Figura 3.7: Estrutura Modular do Mecanismo de Escalonamento Proposto por [48].

Através do *SS-Request Generation Module* no URMA, primeiro são processadas as requisições de banda classe rtPS, depois da classe nrtPS e por último da classe BE, que podem ser encaminhadas para a BS a qualquer momento (qualquer *frame* de operação) para diminuir o *refresh*, e em um único pacote para minimizar o *overhead*. Na FSU, o *Resource Allocation Module* (RAM) emprega essa mesma ordem de prioridade, porém para alocar largura de banda, e seu cálculo considera a eficiência de símbolo (*bits/símbolo*). Assim, o número de símbolos necessários para atender à requisição de largura de banda é definido por:

$$rs_j^{Pi} = \frac{8r_j^{Pi}}{\eta_j} \quad (3.7)$$

Onde rs_j^{Pi} é o número de símbolos, r_j^{Pi} é o número de *bytes* requisitados, η_j é a eficiência de símbolo definida de acordo com a escolha do MCS e Pi é o índice de prioridade.

No trabalho de Sun, J. et al. [49] são propostas estratégias de escalonamento SS-UL e BS-UL para uma rede IEEE 802.16 PMP TDD, representadas na Figura 3.8, cujos objetivos são prover QoS diferenciada e justiça na distribuição dos recursos entre as classes. O escalonador SS-UL seleciona os pacotes em filas não vazias de acordo com a sua prioridade. Existem filas dedicadas para cada classe de serviço, com políticas de escalonamento específicas. Os índices de prioridades são definidos de acordo com a classe de serviço: 4-UGS, 3-rtPS, 2-nrtPS e 1-BE. O escalonamento BS-UL é feito com base em duas filas: Tipo 1 que escalona os *data grants* UGS e consultas periódicas das classes rtPS e nrtPS, e Tipo 2 que escalona as requisições rtPS, nrtPS e BE com o objetivo de garantir uma largura de banda mínima, de acordo com a equação 3.8, e depois distribui a largura de banda residual, de acordo com as equações 3.9 e 3.10.

$$b_i^{MIN} = \min\{B_i^{MIN}, BR_i\} \quad (3.8)$$

$$B^{EX} = B - \sum_i b_i^{MIN} \quad (3.9)$$

$$b_i^{EX} = B^{EX} \cdot w_i / \sum_K w_K \quad (3.10)$$

Onde b_i^{MIN} é a largura de banda alocada para a conexão i , B_i^{MIN} representa a banda mínima reservada, e BR_i representa a banda atual solicitada, B^{EX} é o total de banda residual, B é largura de banda disponível, b_i^{EX} é a banda excedente alocada para a fila da conexão i , e w_i é o peso predefinido para a conexão i .

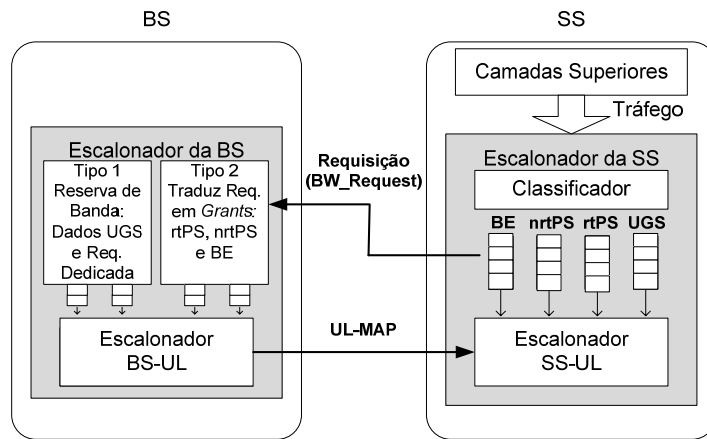


Figura 3.8: Arquitetura de Escalonamento Proposta por [49].

Ball, C. et al. [50] propõem um mecanismo de escalonamento oportunista para redes IEEE 802.16 PMP chamado de TRS (*Temporary Removal Scheduler*). As SSs cujas condições de potência do sinal são baixas em relação a um determinado patamar devem ser temporariamente removidas da lista do escalonador durante um período ajustável Tr . Quando Tr expira, novamente são verificadas as condições do enlace daqueles pacotes retirados da fila do escalonador e, se essas condições ainda forem ruins, os pacotes são novamente removidos por um período Tr . Esse processo pode ser repetido L vezes, e após um período $L.Tr$, mesmo que as condições de potência ainda não sejam favoráveis, o pacote é incluído na lista do escalonador. Se as condições do enlace melhoram, o pacote pode ser colocado novamente na lista de transmissão do escalonador. No pior dos casos o TRS produz um atraso adicional de $L.Tr$. Esse processo é representado na Figura 3.9.

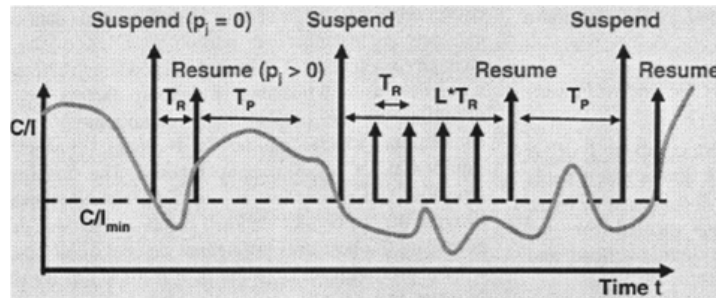


Figura 3.9: Dinâmica da Elegibilidade das SSs para Escalonamento em [50].

Tsai, T. et al. [51] propõem um escalonador da BS de *uplink* e CAC. O funcionamento do escalonador é definido em cinco etapas:

1. Todas as conexões UGS recebem concessões periódicas para a transmissão de dados;
2. Quando uma SS com os fluxos das classes rtPS, nrtPS ou BE, precisa estabelecer uma conexão com a BS, são enviados dois parâmetros, r e b , para serem avaliados pelo CAC na BS, que então decide se aceita ou não o pedido de conexão. O fluxo rtPS envia um parâmetros extra d , que especifica a latência máxima. O CAC é baseado no princípio do *token bucket*, representado na Figura 3.10. Nesse mecanismo, um pacote só poderá ser transmitido quando um *token* for disponibilizado, e os *tokens* são gerados a uma taxa r . Portanto, a quantidade máxima de dados enviados durante um período de tempo t será:

$$r \cdot t + b \quad (3.11)$$

3. Depois, com base na disciplina EDF (*Earliest Deadline First*) são determinados os *deadlines* dos pacotes da classe rtPS para garantir a latência máxima dos fluxos mapeados para essa classe;
4. Após servir os fluxos UGS e rtPS, caso existam recursos disponíveis, o mecanismo atende os fluxos nrtPS e BE de acordo com os patamares de banda mínima T_{nrtps} ou T_{be} : se a banda atual fornecida estiver acima desses patamares a prioridade do fluxo nrtPS ou BE é reduzida e se estiver abaixo a prioridade desses fluxos é aumentada. Assim, estabelece um controle que ajuda a evitar a inanição de algum tipo de tráfego;
5. Se ainda sobraem recursos, eles serão destinados para atender ao excedente de largura de banda solicitada pelas conexões nrtPS e BE.

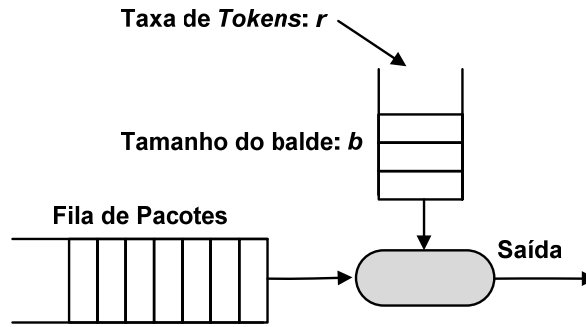


Figura 3.10: Mecanismo *Token Bucket* [51].

Gidlund, M. & Wang, G [52] propõem dois tipos de algoritmos para escalonamento no enlace *uplink*. Um algoritmo, primeiro aloca largura de banda para as classes do tráfego de tempo real (RT) e não tempo real (nRT) com base na quantidade de SSs com fluxos ativos e na MRTR, de acordo com as equações 3.12 e 3.13, com exceção da classe UGS que possui reserva de banda fixa. Depois, de acordo com a proporção de banda que foi alocada para cada classe, escalona os pacotes das classes do tráfego RT utilizando a disciplina EDF, e os pacotes das classes do tráfego nRT utilizando a disciplina WFQ, como mostrado na Figura 3.11.

$$BW_{ertPS,rtPS} = C \frac{\sum_i MRTR_i}{\sum_{j=1}^n MRTR_j} \quad (3.12)$$

$$BW_{nrtPS,BE} = C \frac{\sum_i MRTR_i}{\sum_{j=1}^n MRTR_j} \quad (3.13)$$

Onde, na equação 3.12, i representa o índice das conexões pertencentes às classes ertPS e rtPS, e na equação 3.13, i representa o índice das conexões pertencentes às classes nrtPS e BE. Em ambas as equações, C é a capacidade do canal de *uplink* e j é o índice das conexões de todas as classes.

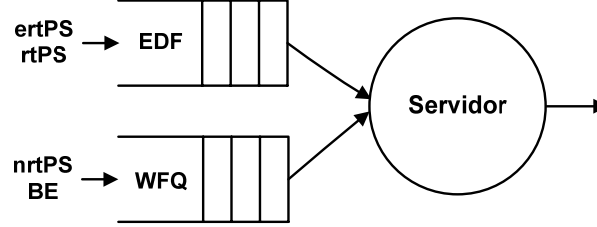


Figura 3.11: Escalonador Apresentado em [52].

O outro algoritmo de escalonamento apresentado por Gidlund, M. & Wang, G [52] é do tipo *cross-layer*, cujo principal objetivo é satisfazer todos os requisitos de QoS e o oferecer um desempenho com balanceamento ótimo entre vazão e justiça. As políticas de escalonamento do tipo *cross-layer*, em geral, melhoram a eficiência de transmissão em ambientes onde a qualidade do sinal das SSs sofre alguma degradação. No entanto, elas costumam atender apenas uma SS a cada *frame* de operação, que apresenta maior prioridade em função da qualidade do sinal e requisitos da classe. O algoritmo desse trabalho escala todas as classes seguindo a ordem de prioridade: UGS>rtPS>nrtPS>BE. Assim, de acordo com a função de prioridade, o escalonador aloca *slots* por *frame* para a conexão i :

$$i^* = \arg \max_i \phi_i(t) \quad (3.14)$$

$$\phi_i(t) = \frac{R_i(t)}{f_i(t) + \varepsilon'} \quad (3.15)$$

Onde $\phi_i(t)$ é a função de prioridade para a conexão i no instante de tempo t , $f_i(t)$ é o indicador de justiça e $R_i(t)$ é quantidade de dados que podem ser transmitidos de acordo com o esquema AMC escolhido.

Para garantir um valor predeterminado de PER (*Packet Error Rate*), cada conexão utiliza o esquema de Modulação e Codificação Adaptativa (AMC) mais adequada, que é escolhido de acordo com a qualidade do sinal entre a SS e a BS, em um ambiente de simulação cujo canal segue o modelo FSMC (*Finite State Markov Channel*).

O trabalho de Chiang, C. et al. [53] propõe um algoritmo de escalonamento integrado do tipo *cross-layer* (Transporte/MAC) que ajusta dinamicamente a taxa dos *subframes* *downlink* e *uplink* (*framing*) de acordo com o perfil assimétrico do tráfego das redes multimídia, justificando porque a divisão estática DL/UL = 1:1 pode levar a uma utilização de largura de banda ineficiente. Esse algoritmo é chamado de ABAS (*Adaptive Bandwidth Allocation Scheme*), e seu objetivo é melhorar o desempenho de vazão e utilização para o tráfego não tempo real (BE) que utiliza protocolo TCP. Ele determina o número de *slots* alocados para o *downlink subframe* (DL) e para o *uplink subframe* (UL), ajusta a divisão entre as duas partes do *frame* em uma proporção ótima de $X:Y$, como mostrada na Figura 3.12; e posteriormente informa as SSs através das mensagens de gerenciamento da MAC, DL-MAP e UL-MAP.

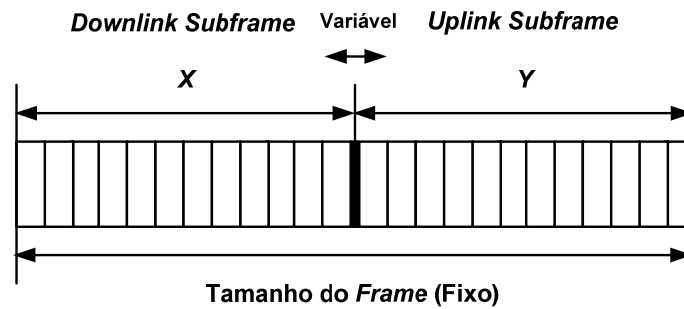


Figura 3.12: Ajuste Dinâmico do Tamanho dos *Subframes* [53].

A avaliação de desempenho da proposta de Chiang, C. et al. [53] foi efetuada por meio de modelagem e simulação, e segundo os autores, mesmo quando comparado com aproximações estáticas, ou seja, que utilizam uma proporção de banda DL/UL fixa de 1:1, o ABAS apresentou um desempenho superior em relação à vazão e a utilização de banda.

Na literatura atual, é possível encontrar diversos trabalhos que abordam o tema escalonamento *cross-layer* como, por exemplo, em [19] [32] [37] [39] [48] [51] [52] [54] [55] [56]. Basicamente, como na equação 3.14, aquela conexão i , cuja função de prioridade (PRF) apresenta maior valor em um dado instante é atendida pelo escalonador, e sua taxa de transmissão nesse instante depende do MCS utilizado, que é definido pelo mecanismo de Modulação e Codificação Adaptativa (AMC) para manter a taxa de erro de pacote (PER) abaixo de um patamar predeterminado aceitável (P_0). A PRF pode ser função do índice de prioridade da classe (UGS>rtPS>nrtPS>BE), da diferença entre a taxa atual fornecida com a

taxa mínima requisitada, da qualidade do sinal no canal, do índice de justiça, do indicador de satisfação de QoS que é definido de acordo com os requisitos da classe, dentre outros.

O trabalho de Bhandari, B. et al. [44] mostra que o aumento do tamanho do pacote pode melhorar a utilização de largura de banda no *uplink* e também aumentar a vazão em função da redução do percentual de *overhead*, mas pode impactar negativamente no atraso. Além disso, esse trabalho revelou que o aumento do número de estações ativas na rede pode comprometer a vazão no *uplink* em função do aumento do número de colisões durante o período de contenção para requisição de largura de banda.

No trabalho de Gidlund, M. & Wang, G. [19] é apresentada uma análise de desempenho dos tipos de algoritmos de escalonamento em relação aos parâmetros: utilização de *frame*, vazão média, atraso médio e justiça. A análise efetuada considera a influência do *uplink burst preamble* e do comprimento do *frame*, e concluiu que a vazão média decresce com o aumento do número de SSs em função do decréscimo da carga disponível que pode ser distribuída para cada SS e do desperdício causado pelo aumento do *overhead (uplink burst preamble)*. Os demais parâmetros como atraso, justiça e utilização de *frame* também são prejudicados quando o número de SSs aumenta. O efeito do aumento do comprimento do *frame* em relação ao atraso médio é prejudicial, pois os pacotes deverão esperar mais tempo nas filas, mas ajuda a melhorar o índice de justiça quando é utilizado um mecanismo de escalonamento WRR. No entanto, se o mecanismo utilizado é do tipo *cross-layer* o índice de justiça piora, pois seleciona sempre uma SS em cada *frame*, enquanto o WRR tenta selecionar um número máximo de SSs a cada *frame*.

No trabalho de Ali, N. A. et al. [32] é apresentada uma análise de desempenho de diversos parâmetros de QoS de acordo com a utilização do mecanismo de requisição de banda baseado em contenção e *piggyback*. Com exceção do mecanismo de escalonamento com algoritmo do tipo *cross-layer*, todos os algoritmos, EDF, WFQ, WRR e seus híbridos, apresentam um melhor desempenho utilizando o mecanismo de contenção sob carga suave da rede. Quando a carga da rede torna-se pesada, o melhor desempenho é alcançado com a utilização do mecanismo de *piggyback*, em função do grande *backlog* de dados presentes nas filas das conexões das SSs. Porém, o mecanismo com algoritmo do tipo *cross-layer* não apresenta esse mesmo comportamento porque seleciona apenas uma SS a cada *frame*, e geralmente descarrega de uma vez todos os seus pacotes, não formando *backlog* para requisição de banda através de *piggyback*.

No trabalho apresentado por Mukul, R. et al. [57] é proposto um mecanismo de escalonamento definido como *Adaptive rtPS*, cujo objetivo é reduzir o atraso da classe *rtPS*.

Na maioria dos mecanismos de escalonamento se a requisição de banda é feita no instante t_0 , a concessão ocorrerá no instante t_1 ($t_1 > t_0$), resultando em um atraso relativo. Com o objetivo de otimizar esse processo, foi desenvolvido um esquema baseado na predição da taxa de chegada de pacotes do tráfego de tempo real (rtPS) nas filas da SS, permitindo estimar a necessidade de largura de banda, e então encaminhar, de maneira avançada, uma requisição antes mesmo da SS receber os dados previstos. A análise dos resultados obtidos através de modelagem e simulação concluiu que o esse mecanismo, comparado a outros, apresentou melhor desempenho em relação aos parâmetros de atraso e tamanho do *backlog* das filas das SSs com conexões rtPS, que permite utilizar um *buffer* menor. De certa maneira, o *Adaptive* rtPS otimiza o processo, mas não evita a necessidade do envio de mensagens para requisição de largura de banda, e portanto, não reduz a carga desse tipo de mensagem na rede.

Maheshwari, S. et all [58] propõe um escalonador BS-UL que utiliza a estratégia *max-min fair allocation*, cujos objetivos são garantir os requisitos de largura de banda mínima, manter a justiça na distribuição dos recursos alcançando a máxima utilização do canal e garantir o atraso máximo limitado. Esse processo de alocação de banda para as SSs é executado em dois estágios. No primeiro estágio, a largura de banda no *uplink* é distribuída entre os fluxos das quatro classes do padrão IEEE 802.16d, de acordo com a equação 3.16, garantindo uma quantidade de banda mínima para cada classe. No segundo estágio, a banda alocada para cada classe é distribuída entre todas as SSs dessa classe, de acordo com a equação 3.17. E por fim, a ordem de transmissão entre as SSs é redefinida de acordo com o *deadline* dos pacotes UGS transmitidos por cada SS, para garantir o atraso máximo limitado.

$$flow_Alloc[i] = \left(\frac{flow_Weight[i]}{\sum_{i=1}^4 flow_Weight[i]} \right) \cdot totalUplinkBytes \quad (3.16)$$

$$flow_Alloc_SS = \frac{flow_Alloc[i]}{TotalSS} \quad (3.17)$$

Onde, na Equação 3.16, $flow_Alloc[i]$ é o total de *bytes* alocados para os fluxos da classe de índice i , $flow_Weight[i]$ é o peso definido para os fluxos ativos da classe i e $totalUplinkBytes$ é o total de *bytes* disponíveis no *uplink*. Na Equação 3.17, $flow_Alloc_SS$ é o número de *bytes* alocados para cada SS com fluxos ativos da classe i , e $TotalSS$ é o número total de SSs com fluxos ativos da classe i . Em ambas as equações i é o índice da classe: 1-UGS, 2-rtPS, 3-nrtPS e 4-BE.

O trabalho de Chen, J. et al. [59] propõe um módulo para redes IEEE 802.16 PMP, composto por mecanismos de escalonamento e CAC. Os pacotes movidos para a camada MAC recebem um atributo que indica sua prioridade, de acordo com a classe de serviço, UGS ($q5$), rtPS ($q4$), ertPS ($q3$), nrtPS ($q2$) e BE ($q1$), e então são armazenados em filas dedicadas. Na primeira rodada, o escalonador seleciona os pacotes em cada fila de acordo com a função:

$$B_{Type}^T = \min(R_{Type}, B_{Total} * qi) \quad (3.18)$$

Onde B_{Type}^T representa a largura de banda alocada para uma dada classe, R_{Type} representa a quantidade de banda solicitada pela classe, B_{Total} representa a largura de banda total disponível no sistema e qi representa a proporção de banda destinada para a classe. Na segunda rodada, a largura de banda que resta é distribuída proporcionalmente entre as classes, de acordo com os parâmetros $\{q5, q4, q3, q2, q1\}$.

O trabalho de Dutra, L. [60] apresentou uma solução prática para o emprego de redes WiMAX interligando redes MetroEthernet. No que diz respeito ao padrão IEEE 802.16, foram implementadas políticas de escalonamento *strict priority* e WRR para oferecer tratamento diferenciado com priorização, definindo que primeiro devem ser atendidos os fluxos do tráfego de tempo real, e caso ainda restem recursos, devem ser atendidos os fluxos dos demais tráfegos, na proporção do peso atribuído a cada um. A avaliação dos resultados foi efetuada por meio da implementação de uma rede com componentes físicos configurados adequadamente em um laboratório de testes.

3.4. Considerações Finais

Neste capítulo foram apresentados os principais conceitos a respeito de escalonamento em redes IEEE 802.16. Foi definido o conceito de escalonador e as características básicas a respeito dos tipos de disciplinas de escalonamento: GPS, FIFO, PQ, FQ, RR, WRR, DRR, WFQ e EDF.

Também foi apresentado o estado da arte do estudo em escalonamento em redes IEEE 802.16, destacando os principais requisitos para operarem de maneira adequada e eficiente, a classificação dos algoritmos de escalonamento de acordo com suas características estruturais e funcionais, e as principais propostas de mecanismos de escalonamento para as redes IEEE 802.16.

Capítulo 4

UMA PROPOSTA DE ESCALONADOR BASEADO NA DISCIPLINA *PRIORITY QUEUING* (PQ) PARA REDES IEEE 802.16

4.1. Introdução

A crescente demanda por serviços banda larga e as vantagens oferecidas pelo acesso através das redes BWA fazem com que o padrão IEEE 802.16 venha atraindo o interesse de muitos pesquisadores e fabricantes de equipamentos para o desenvolvimento de soluções que atendam a essa demanda.

O padrão IEEE 802.16 apresenta diversas características importantes como a presença inerente de mecanismos para a provisão de QoS. Todavia, ele não especifica as políticas a serem adotadas por estes mecanismos e, como as redes WiMAX agregam diversos tipos de tráfego em um sistema com recursos limitados, eles devem ser capazes de prover garantias de QoS para classes heterogêneas de tráfego que possuem requisitos diferentes, bem como maximizar a eficiência na utilização dos recursos no enlace *uplink*.

O escalonador da BS para o tráfego *uplink* (BS-UL) é responsável por alocar banda de maneira dinâmica para que as SSs possam transmitir seus dados, de acordo com os requisitos de QoS definidos pelo padrão IEEE 802.16 e, com as mensagens de requisição de banda enviadas pelas SSs. O mecanismo BS-UL é o mais complexo dos três escalonadores WiMAX porque não tem acesso direto as filas de pacotes nas SSs e depende do processo de requisição e alocação de largura de banda para efetuar a concessão de recursos às SSs, sob demanda.

Assim, neste capítulo é apresentada uma proposta de mecanismo de escalonamento BS-UL baseado na disciplina de escalonamento PQ (*Priority Queuing*), que é uma extensão do trabalho apresentado em [3]. Os principais objetivos desta proposta são oferecer tratamento diferenciado com priorização de tráfego, aumentar o isolamento entre os tipos de tráfego e a eficiência de utilização de recursos sem degradar os níveis de QoS oferecidos aos usuários.

Este capítulo está organizado da seguinte forma: na Seção 4.2 são descritos os problemas que motivaram o desenvolvimento da proposta deste trabalho, na Seção 4.3 é apresentada a solução proposta, na Seção 4.4 são apresentados os trabalhos relacionados, e na Seção 4.5 são feitas as considerações finais sobre o capítulo.

4.2. Descrição do Problema

O padrão IEEE 802.16 define uma rede BWA sem fio com a presença de mecanismos de provisão de QoS inerentes à sua estrutura, cuja necessidade é definida, mas a política de como devem operar não, ficando a cargo de desenvolvedores e fabricantes de equipamentos de rede para que eles possam diferenciar seus produtos. O principal mecanismo da arquitetura de QoS deste padrão é o escalonador.

Dos três mecanismos de escalonamento empregados nas redes WiMAX, o escalonador BS-UL tem a tarefa mais complicada, pois não tem acesso direto às filas nas SSs e tem de alocar os *mini-slots* no *uplink subframe* baseado em filas virtuais na BS, que são definidas através das mensagens de requisição de largura de banda (*BW_Request*) enviadas pelas SSs à BS [10]. Dessa maneira, a BS precisa estar sincronizada com várias SSs para gerenciar adequadamente as operações de requisição e alocação de largura de banda.

Assim, quando uma SS deseja transmitir seus pacotes, precisa antes obter uma oportunidade para o envio de requisição de largura de banda para a BS, informando a quantidade de *mini-slots (bytes)* necessários para transmitir seus dados. Depois que a BS recebe a requisição de banda enviada pela SS, o mecanismo de escalonamento BS-UL deve processá-la juntamente com as requisições enviadas por outras SSs, e de acordo com a capacidade do sistema e com os requisitos de QoS de cada fluxo, em seguida retornar via UL-MAP a informação sobre a quantidade e a disposição dos *mini-slots* onde cada SS poderá transmitir seus pacotes no *uplink subframe*. Posteriormente, o escalonador SS-UL em cada SS deve encaminhar os pacotes de suas filas nos *grants* concedidos pelo escalonador BS-UL.

O processo descrito anteriormente é executado ao longo de mais de um *frame*, e acarreta um atraso, chamado de atraso de acesso da rede, entre o momento da necessidade imediata de banda de uma SS e o momento em que é disponibilizada a oportunidade para a SS efetivamente poder transmitir os pacotes armazenados em suas filas [57]. Além disso, no momento em que a requisição é atendida a necessidade de banda da SS já pode ter sido alterada, uma vez que novos pacotes podem ter sido gerados e armazenados em suas filas. A

diferença entre a necessidade de banda percebida pela BS e a necessidade atual da SS tem um grande impacto no desempenho do mecanismo de escalonamento BS-UL [48] [61].

As redes IEEE 802.16 foram projetadas para prover garantias de QoS a diversos tipos de tráfego com requisitos específicos. Como nem sempre é possível oferecer toda a largura de banda solicitada pelas aplicações, os mecanismos de QoS precisam ter a capacidade de oferecer tratamento diferenciado com priorização de serviços. Na prática, isso significa que, por exemplo, pacotes transmitindo vídeo e voz de uma aplicação de vídeo-conferência em tempo real devem ter prioridade de tráfego sobre pacotes transportando dados de uma aplicação de correio eletrônico (*e-mail*) ou de transferência remota de arquivos através do protocolo FTP (*File Transfer Protocol*), caso contrário os usuários de vídeo-conferência podem experimentar cortes e falta de sincronismo, tornando o serviço inviável.

De acordo com o estado da arte da pesquisa em escalonamento em redes IEEE 802.16, apresentada no Capítulo 3, a eficiência de um mecanismo de escalonamento depende do tipo de seu algoritmo, cuja implementação é baseada em disciplinas de escalonamento, do perfil de tráfego da rede, pois é preciso considerar quais são os parâmetros mais críticos para o tipo de rede onde será empregado, e de diversos parâmetros como, por exemplo, o comprimento do *frame* e a maneira como é requisitada a largura de banda [7].

O emprego de uma única disciplina não atende a todos os requisitos necessários para a provisão de QoS em redes IEEE 802.16 [32]. Quando é possível garantir, por exemplo, o parâmetro de vazão, o parâmetro de justiça pode não ser satisfeito, onde o ideal seria um desempenho balanceado entre ambos.

Em função disso, surgiram muitos trabalhos com novas propostas de escalonamento baseados em políticas mais sofisticadas e complexas [47] [48] [49] [52], com objetivos específicos e eficiência relativa. Também é importante destacar, que redes WiMAX onde não foram implementados mecanismos de CAC e policiamento de tráfego, como é o caso da proposta apresentada em [3], são totalmente dependentes de um mecanismo de escalonamento ainda mais eficiente para garantir a provisão de QoS com flexibilidade e escalabilidade.

Um dos principais problemas que os desenvolvedores de mecanismos de QoS para as redes IEEE 802.16 devem considerar no seu projeto é que a capacidade de transmissão no *uplink* é limitada pela largura de banda, que é compartilhada entre vários usuários, e geralmente é escassa em relação a essa demanda. A tendência futura é de que se torne cada vez mais escassa devido ao aumento da necessidade por acesso banda larga.

Além disso, as mensagens de gerenciamento e controle transmitidas através das conexões de gerenciamento são *inband*, isso significa que elas são transmitidas no mesmo

canal onde também são transmitidos os dados, e assim, podem competir pela largura de banda no *uplink* com as mensagens de dados transmitidas através das conexões de transporte. Também, é importante destacar que a utilização dos *frames* nem sempre é adequada, pois são compostos por um número inteiro fixo de *slots* físicos e podem ser transmitidos com *slots* vazios, ou seja, que não foram alocados para nenhuma SS.

Portanto, desenvolver um mecanismo de escalonamento para redes IEEE 802.16 eficiente e capaz de atender todos os requisitos de QoS é um desafio bastante complexo.

4.3. Solução Proposta

Nesta Seção é apresentada uma proposta de mecanismo de escalonamento para alocação de largura de banda no *uplink* para redes IEEE 802.16, cujos principais objetivos são: oferecer tratamento diferenciado com priorização para o tráfego heterogêneo das redes WiMAX; aumentar o nível de isolamento e proteção entre as classes de tráfego tempo real e não tempo real; aumentar a eficiência de utilização de recursos da rede sem degradar os níveis de QoS oferecidos aos usuários; e consequentemente melhorar o desempenho dos parâmetros de QoS de vazão média, utilização percentual de recursos e atraso médio.

A política da solução escolhida é baseada na disciplina de escalonamento PQ (*Priority Queuing*), que armazena os pacotes dos tráfegos de serviço em filas dedicadas, permitindo que sejam atribuídos diferentes níveis de prioridade para seu atendimento como, por exemplo, alto, médio, baixo e normal. Portanto, é adequada para oferecer tratamento diferenciado com priorização para o tráfego tempo real que deve ser armazenado nas filas de maior prioridade para que seja atendido primeiro do que o tráfego não tempo real, que é tolerante ao atraso.

De acordo com o estado da arte da pesquisa em escalonamento em redes IEEE 802.16, nenhuma política de escalonamento homogênea é capaz de atender todos os requisitos de QoS. Por isso, surgiram propostas que utilizam políticas híbridas, como a apresentada em [3], que é implementada em um módulo com código aberto disponível para o acesso de pesquisadores e desenvolvedores, é baseada na disciplina PQ híbrida, com funções específicas para atender as diferentes classes de tráfego armazenadas em suas filas, e por isso é capaz de fornecer suporte para a provisão de QoS para todas as classes do padrão IEEE 802.16d GPSS.

No entanto, o mecanismo de escalonamento apresentado em [3], possui três filas de prioridades, mostradas na Figura 3.5, onde as requisições das classes rtPS e nrtPS são armazenadas na mesma fila, e por isso não oferece um isolamento de tráfego completo porque

o comportamento do tráfego da classe nrtPS pode prejudicar o comportamento do tráfego da classe rtPS. A eficiência desse escalonador é prejudicada pelo processo de requisição e alocação de largura de banda, em função da maneira como são atualizadas as informações na BS a respeito do tamanho das filas nas SS. Além disso, como ele não aproveita os *mini-slots* que restam, após atender todas as filas, diretamente para a transmissão de dados, não é capaz de maximizar a eficiência de utilização dos recursos no enlace *uplink*.

Assim, o mecanismo proposto neste trabalho é uma extensão do trabalho apresentado em [3], e apresenta uma estrutura com quatro filas, representadas na Figura 4.1, com uma separação entre as requisições de cada tipo de tráfego, tempo real (RT - *Real Time*) e não tempo real (nRT - *non Real Time*), pois possuem requisitos de QoS diferentes.

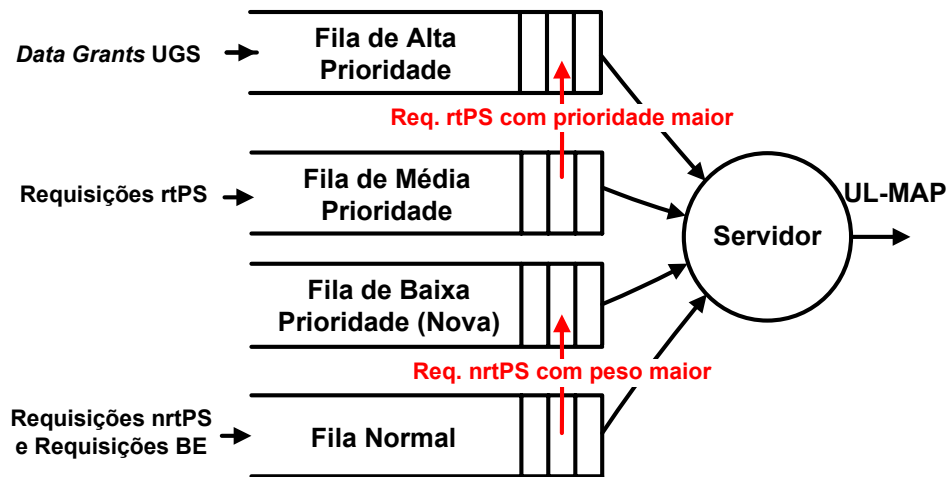


Figura 4.1: Estrutura do Mecanismo de Escalonamento Proposto.

A disciplina PQ utilizada opera na forma *strict priority*. Assim, o escalonador atende completamente a fila de alta prioridade primeiro, e após, caso ainda existam recursos, atende completamente a fila de média prioridade, e na sequência, as filas de baixa prioridade e normal.

A fila de alta prioridade é composta por *data grants* da classe UGS, que apresenta largura de banda fixa reservada, e requisições rtPS com prioridade maior. A fila de média prioridade, que corresponde à fila intermediária do mecanismo base apresentado em [3], é composta de requisições rtPS, e caso existam recursos disponíveis, podem migrar para a fila de alta prioridade para satisfazer seus requisitos de QoS. A fila de baixa prioridade (nova), armazena as requisições nrtPS com peso maior. A fila normal, além das requisições BE,

armazena também as requisições nrtPS, e caso existam recursos disponíveis, as requisições nrtPS podem migrar para a fila de baixa prioridade, para satisfazer seus requisitos de QoS.

O objetivo da criação da nova fila foi separar as requisições das classes rtPS e nrtPS, para isolar e proteger o tráfego de tempo real. Assim, as requisições rtPS, que possuem o requisito de atraso máximo limitado, são atendidas primeiro em relação às da classe nrtPS. O tráfego da classe nrtPS não tem seu desempenho prejudicado pelo fato de suas requisições não migrarem para a fila de alta prioridade, como no mecanismo base apresentado em [3], porque é compensado pelo esquema de pré-alocação, que é descrito mais adiante neste Capítulo.

Antes de atender a cada fila de acordo com a política de escalonamento PQ, são executadas funções específicas que definem a ordem de armazenamento, e consequentemente de atendimento das requisições de cada classe de serviço dentro dessas filas. A política de atendimento das requisições das classes UGS e BE é do tipo FIFO, ou seja, são atendidas na ordem em que chegarem. As requisições da classe rtPS são ordenadas e atendidas em ordem decrescente de prioridade, atribuída de forma dinâmica de acordo com o *deadline* e com a diferença entre a taxa atual fornecida e a taxa mínima requisitada para os fluxos dessas requisições, de acordo com as funções definidas em [3].

A política de atendimento das requisições nrtPS armazenadas na fila de baixa prioridade, mostrada no pseudocódigo Escalonamento_nrtPS, define que as requisições cujos fluxos possuem um valor de peso maior devem ser atendidas primeiro. O peso é proporcional à MRTR, que de acordo com o padrão IEEE 802.16 é um parâmetro de QoS mandatório para a classe nrtPS. Pois, mesmo conexões diferentes que utilizam um mesmo tipo de serviço, como pode ser o caso de aplicações diferentes, podem apresentar necessidades de taxas mínimas de serviço diferentes.

Após o escalonador atender todas as suas filas é montada a mensagem de controle UL-MAP, que define a disposição e a quantidade de *mini-slots* que cada SS poderá utilizar para transmitir seus dados no *uplink subframe*. Eventualmente, após o escalonador atender a todas as filas, ainda podem restar *mini-slots* que não foram alocados para nenhuma SS, uma vez que o *uplink subframe* é dividido em um número inteiro fixo de *mini-slots*.

A preocupação de como efetuar o aproveitamento dos *mini-slots* não utilizados no *uplink subframe* é uma das premissas para a melhoria de desempenho dos mecanismos de escalonamento defendida por Sayenko, A. et al. [46] e, portanto, sua implementação é também considerada relevante para alcançar o objetivo da proposta deste trabalho. Assim, após o escalonador atender todas as suas filas, aqueles *mini-slots* que ainda não foram

definidos para nenhuma SS dentro da mensagem de gerenciamento UL-MAP, precisam ser alocados diretamente para a transmissão dos dados das SSs.

Assim, foi desenvolvido o esquema de pré-alocação, apresentado no pseudocódigo Pré_Alocação_nrtPS, para aproveitar os *mini-slots* não utilizados no *uplink subframe*. Após o escalonador atender todas as filas, aqueles *mini-slots* que restam são alocados de maneira antecipada para uma SS transmitir seus dados, ou seja, antes do envio de um pedido de requisição de largura de banda da SS a BS, evitando a necessidade do envio desse tipo de mensagem de gerenciamento e, conseqüentemente reduzindo sua carga na rede.

O esquema de pré-alocação proposto pode atender a um fluxo nrtPS de uma SS cuja fila de armazenamento acabara de receber pacotes de dados, sem precisar esperar por uma oportunidade (*polling*) para o envio de uma mensagem de requisição de largura de banda, além de todo o tempo decorrente do processo de requisição e alocação de largura de banda (no frame n a SS encaminha a requisição de banda, indicando a quantidade de *bytes* que precisa para transmitir seus dados de acordo com o tamanho de suas filas de pacotes; depois, no frame $n+1$, a requisição é processada pelo escalonador BS-UL, que gera o UL-MAP; e então no frame $n+2$ a SS transmite seus pacotes de acordo com o UL-MAP recebido).

Além disso, como podem chegar novos pacotes de dados nas filas das SSs durante o intervalo de tempo decorrente do processo de requisição e alocação de largura de banda, esse esquema pode atender mais precisamente em relação a outros mecanismos a nova quantidade de largura de banda necessitada pelas SSs, sem a necessidade do envio de uma requisição de largura de banda atualizada, como mostrado na Figura 4.2.

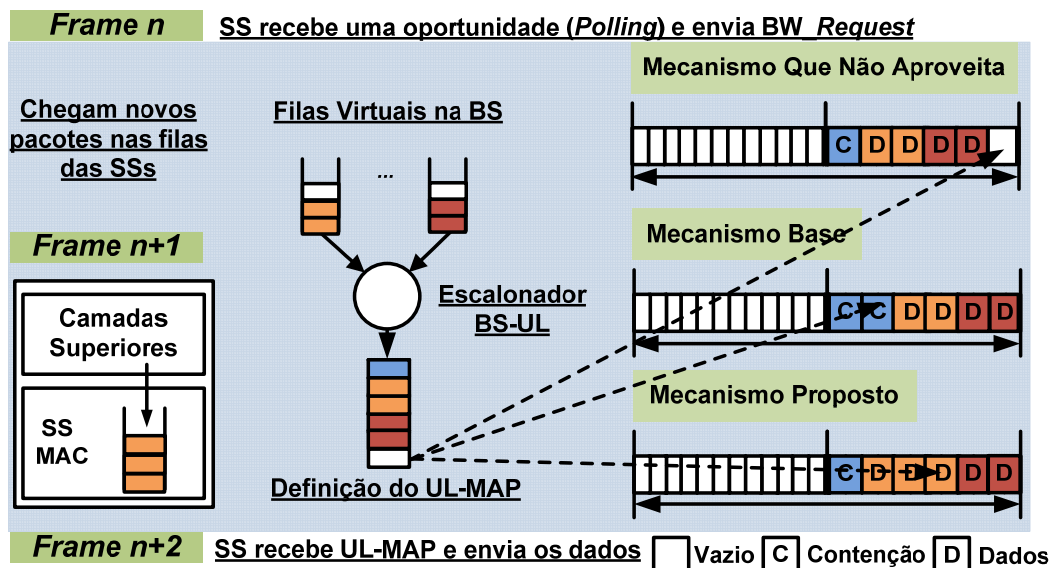


Figura 4.2: Vantagem do Esquema de Pré-Alocação.

Portanto, através da pré-alocação pretende-se reduzir o atraso de acesso da rede, minimizar o problema da percepção desatualizada da necessidade de largura de banda das SSs pela BS, que são resultantes dos processos descritos anteriormente, e ainda diminuir a carga de mensagens de gerenciamento, ao evitar a necessidade do envio de uma mensagem para a requisição de largura de banda, aproveitando prontamente os recursos disponíveis diretamente para a transmissão de dados; e consequentemente aumentar a eficiência de utilização dos recursos no enlace *uplink*.

A seguir é apresentado o pseudocódigo proposto.

Escalonamento_nrtPS(N , Vetor_Req_nrtPS, Numero_Req_nrtPS)

entrada: N , Vetor_Req_nrtPS, Numero_Req_nrtPS.

```

1. início
2.   para ( $i = 0$  até  $N/n-1$ ) faça
3.      $PF_i = MRTR_i$ ;
4.   fim para
5.   cria novo Vetor_Req_nrtPS[Numero_Req_nrtPS];
6.   enquanto (Existe Req)
7.     se (Req.Tipo = Req_nrtPS) faça
8.       Vetor_Req_nrtPS[ $n$ ] = Req_nrtPS;
9.     fim se
10.    Req.Apontador = Req->proximo;
11.  fim enquanto
12.  se ( $PF_i \neq s$ ) faça
13.    Ordena_Req_nrtPS_Acordo_com_Peso(Vetor_Req_nrtPS, Numero_Req_nrtPS);
14.  fim se
15.  se ((Existe Req) e (Req.Tipo = Req_nrtPS)) faça
16.    fluxo_PreAlocado = Vetor_Req_nrtPS[0].SFID; //1ª Req_nrtPS a ser atendida
17.  fim se
18.  enquanto (( $n < \text{Numero\_Req\_nrtPS}$ ) e ( $N > 0$ )) faça //  $n$  tem valor inicial nulo
19.    Vetor_Req_nrtPS[ $n$ ].Flag = 1;
20.    SlotsAlocados = Vetor_Req_nrtPS[ $n$ ].NumSlots;
21.    se (Vetor_Req_nrtPS[ $n$ ].NumSlots >  $N$ ) faça
22.      SlotsAlocados =  $N$ ;
23.      Vetor_Req_nrtPS[ $n$ ].NumSlots -= (SlotsAlocados - 1);
24.      Vetor_Req_nrtPS[ $n$ ].Flag = 0;
25.    fim se
26.    Insere_Fila_Baixa_Prioridade(SlotsAlocados*);
27.     $N = N - \text{SlotsAlocados}$ ;
28.     $n++$ ;
29.  fim enquanto
30. fim

```

*Insere Req_nrtPS com Req_nrtPS. NumSlots = SlotsAlocados.

Observação: Fila_Normal e Fila_Baixa_Prioridade são listas encadeadas com alocação dinâmica de memória.

Pré_Alocação_nrtPS(N)

entrada: N .

//Após atender todas as filas pré-aloca os *mini-slots* restantes no UL para fluxo_PreAlocado

1. **início**

2. **se** ($N > 0$) **faça**

3. SlotsAlocados = N ; //mini-slots pré-alocados para fluxo_PreAlocado

4. $N = 0$;

5. **fim se**

6. **fim**

Lista de Parâmetros

N = Número de *mini-slots* disponíveis no *uplink*;

N_{fn} = Número de fluxos nrtPS ativos;

Numero_Req_nrtPS = Número de requisições nrtPS na fila normal;

MRTR i = Taxa mínima requisitada por um fluxo nrtPS;

PF i = Peso atribuído ao fluxo nrtPS;

Req_nrtPS = Estrutura de uma requisição do fluxo nrtPS na fila normal;

Vetor_Req_nrtPS = Vetor de referências para as estruturas de requisições nrtPS Req_nrtPS;

Req_nrtPS.NumSlots = Número de *mini-slots* requisitados;

Req_nrtPS.Peso = Peso do fluxo da requisição nrtPS (PF i);

Req_nrtPS.SFID = Identificador de fluxo da requisição nrtPS;

Req.Tipo = Classe de serviço da requisição;

Req.Flag = *Flag* da requisição na fila normal (0: manter na fila, 1: excluir da fila);

Req.Apontador = Indica a posição da requisição na fila normal;

SlotsAlocados = Número de *mini-slots* alocados;

fluxo_PreAlocado = SFID do fluxo pré-alocado;

Descrição do pseudocódigo Escalonamento_nrtPS(N , Vetor_Req_nrtPS, Numero_Req_nrtPS):

Linhas 2 a 4 - Inicialmente, a variável PF i é atualizada com o valor do peso de cada fluxo, de acordo com o valor MRTR atribuído, obtido a partir da tabela de fluxos ativos;

Linha 5 - Criação de uma nova instância para o vetor Vetor_Req_nrtPS, que armazena os elementos do tipo de estrutura Req_nrtPS;

Linhas 6 a 11 - Enquanto existirem requisições do tipo nrtPS na fila normal, as referências para cada requisição nrtPS são armazenadas no vetor Vetor_Req_nrtPS;

Linhas 12 a 14 - Se existirem pesos diferentes, as referências para cada requisição nrtPS armazenadas em Vetor_Req_nrtPS são reordenadas em ordem decrescente de peso;

Linhas 15 a 17 - Se na rodada de escalonamento atual existir alguma requisição do tipo nrtPS, a variável fluxo_PreAlocado é atualizada com o valor do identificador de fluxo (SFID) da primeira referência para uma requisição nrtPS na fila normal, senão continua com o valor da rodada anterior;

Linhas 18 a 29 - Na sequência, enquanto não for alcançado o último elemento de $\text{Vetor_Req_nrtPS}[n]$ e existirem *mini-slots* disponíveis no *uplink subframe* ($N > 0$), a referência para a requisição do tipo nrtPS é atendida:

- **Linha 19** - Inicialmente, a variável *Flag* é “setada”, definindo que essa requisição deve ser excluída da fila normal após ser atendida;
- **Linha 20** - Depois, a variável *SlotsAlocados*, que define a quantidade de *mini-slots* que serão alocados, é atualizada pela quantidade de *mini-slots* requisitados ($\text{Vetor_Req_nrtPS}[n].\text{NumSlots}$).
- **Linhas 21 a 26** - Se a quantidade de *mini-slots* requisitados está acima da capacidade disponível do *uplink subframe* ($\text{Vetor_Req_nrtPS}[n].\text{NumSlots} > N$), a requisição é inserida ou migra para a fila de baixa prioridade com uma quantidade de *mini-slots* a serem alocados igual à capacidade disponível N ($\text{SlotsAlocados} = N$), e o restante ($\text{Vetor_Req_nrtPS}[n].\text{NumSlots} - \text{SlotsAlocados}$), ou seja, a quantidade de *mini-slots* requisitados que não puderam ser alocados mais um (espaço extra para o controle do processo de fragmentação), permanece na fila normal (que não tem nenhuma prioridade) para ser atendida quando existirem recursos disponíveis e, para isso o valor de *Flag* é desabilitado.
- **Linhas 18,19, 20 e 26** - Caso contrário ($\text{Vetor_Req_nrtPS}[n].\text{NumSlots} \leq N$), a requisição é inserida na a fila de baixa prioridade com uma quantidade de *mini-slots* a serem alocados igual à quantidade requisitada ($\text{SlotsAlocados} = \text{Vetor_Req_nrtPS}[n].\text{NumSlots}$).
- **Linha 27** - A variável N que indica a quantidade de *mini-slots* disponíveis no *uplink subframe* é decrementada pela quantidade de *mini-slots* alocados para a requisição ($N = N - \text{SlotsAlocados}$).
- **Linha 28** - A variável n é incrementada para que o próximo elemento de $\text{Vetor_Req_nrtPS}[n]$ seja atendido.

Todos os passos descritos da linha 18 até a linha 29 são repetidos até que não existam mais *mini-slots* disponíveis no *uplink subframe* ($N = 0$) ou todas as requisições nrtPS na fila normal tenham sido atendidas ($n = \text{Numero_Req_nrtPS}$).

Descrição do pseudocódigo $\text{Pré_Alocação_nrtPS}(N)$:

Linhas 1 a 6 - Após as quatro filas serem atendidas, este método verifica se restaram *mini-slots* no *uplink subframe* ($N > 0$) e, em caso positivo, aloca os mesmos (N) para o fluxo cujo SFID corresponde ao valor da variável fluxo_PreAlocado , atualizada em

Escalonamento_nrtPS. A cada *frame*, a quantidade de *mini-slots* disponíveis ou livres no *uplink subframe* é atualizada em uma variável N , que é decrementada pela quantidade de *mini-slots* alocados para cada requisição enquanto o escalonador atende suas filas. Após atender todas as filas, aqueles *mini-slots* que não foram alocados para nenhuma SS, são destinados às conexões das SSs com fluxos nrtPS. Portanto, após o escalonador atender todas as suas filas, uma quantidade de *mini-slots* igual ao valor da variável N é pré-alocada ao fluxo nrtPS associado à primeira requisição da ordem de atendimento no *frame* mais atual onde existiram requisições nrtPS.

Diferentemente da disciplina WRR, que tenta selecionar um número máximo de SSs para alocação a cada *frame*, o algoritmo proposto tenta atender ao pedido (requisição) completo, e caso a capacidade do *subframe* seja superior à quantidade de recursos solicitados pela primeira requisição, a próxima requisição é atendida completamente ou parcialmente de acordo com a capacidade do *subframe*, e assim sucessivamente. De acordo com essa política, o nível de segmentação do *uplink subframe* pode ser reduzido, atenuando o efeito negativo do *uplink burst preamble*, podendo produzir um desempenho melhor em relação aos parâmetros de QoS como, por exemplo, a vazão.

Em um ambiente com restrições acentuadas de recursos como o das redes sem fio, o ideal é utilizar uma quantidade mínima possível de recursos para as mensagens de sinalização ou *overhead*, e utilizar mais recursos para a transmissão de dados das aplicações. Pois, a distribuição da capacidade de transmissão disponível afeta a qualidade de serviço percebida pelos usuários das aplicações. Portanto, na prática, um esquema capaz de reduzir o número de mensagens de gerenciamento como o preâmbulo de sincronização (*uplink burst preamble*), o SSTG (*Subscriber Station Transition Gap*), que separa as rajadas de dados das SSs no *uplink subframe*, mostrados na Figura 2.10, e a quantidade de *slots* disponibilizados para o envio de requisições de largura de banda, pode produzir um desempenho melhor em relação aos parâmetros de QoS como, por exemplo, a vazão.

O método de pré-alocação foi designado para atender somente os fluxos da classe nrtPS porque, a partir da criação da nova fila (baixa prioridade), deixaram de ter o mesmo nível de prioridade que os fluxos da classe rtPS, e assim, foram compensados para não ter a sua eficiência comprometida. No entanto, esse método pode ser facilmente adaptado para atender a outras classes como rtPS e BE, de acordo com as necessidades do perfil de tráfego de outros ambientes de rede WiMAX.

No Capítulo subsequente são apresentados os resultados da avaliação de desempenho do mecanismo de escalonamento proposto para o tráfego *uplink*, mensurado através de

modelagem e simulação dos parâmetros de QoS, como a vazão, o atraso, a utilização percentual de recursos e o descarte de pacotes nos *buffers* dos elementos da rede, para as classes definidas pelo padrão IEEE 802.16d.

4.4. Trabalhos Relacionados

A proposta de mecanismo de escalonamento da BS para o tráfego *uplink* apresentada em [3] é estruturada com três filas de prioridade, baseado na disciplina PQ que opera na forma *strict priority*, e as requisições das classes rtPS e nrtPS são armazenadas na mesma fila, diferente da proposta deste trabalho, que define uma estrutura com quatro filas para separar e proteger o tráfego da classe rtPS do tráfego da classe nrtPS, que é tolerante ao atraso.

Em [31] é apresentada uma técnica de escalonamento baseada em PQ com duas filas, uma de alta e outra de baixa prioridade, utilizando MDRR (*Modified Deficit Round Robin*) e, AMC (*Adaptive Modulation and Coding*) que considera a qualidade do sinal para definir o MCS, e consequentemente a taxa de transmissão. Não são definidas claramente as principais vantagens dessa combinação de técnicas diante do desempenho dos requisitos de QoS das classes do padrão IEEE 802.16, além da conclusão intuitiva de ser adequada para diferenciação dos níveis de QoS entre as classes; e nem tampouco resultados práticos.

Em [45] não é apresentada a comparação de desempenho do mecanismo de escalonamento desenvolvido com nenhum outro mecanismo de escalonamento. A comparação de desempenho é executada em relação à eficiência do mecanismo proposto utilizando ou não determinadas funções especiais que são definidas no padrão IEEE 802.16 como, por exemplo, a fragmentação, o empacotamento e a PHS. Sua avaliação foi efetuada por meio de modelagem e simulação utilizando um *software* de simulação proprietário.

Em trabalhos como o apresentado em [47], o escalonador BS-UL é baseado na disciplina PQ que opera na forma *strict priority* com quatro filas, como na proposta deste trabalho. No entanto, não apresenta uma política efetiva para aproveitar os *mini-slots* restantes no *uplink subframe*, após o escalonador atender todas as suas filas, utilizando uma técnica que aloca a largura de banda de maneira antecipada antes de receber a requisição de banda (pré-alocação), como é apresentada neste trabalho.

Em [49] é proposto um mecanismo de escalonamento BS-UL com apenas duas filas de prioridades. A fila de maior prioridade armazena os *data grants* UGS e os *grants* periódicos para requisição de banda das conexões rtPS e nrtPS. A fila de menor prioridade escalona os

grants de dados das conexões rtPS, nrtPS e BE. O escalonamento da fila de menor prioridade acontece em duas etapas, na primeira etapa é alocada a banda mínima para cada conexão, e na segunda etapa, caso exista banda disponível, o excedente de banda solicitada é distribuído de acordo com o peso de cada conexão.

No trabalho apresentado em [62], é proposto um mecanismo baseado em uma política de escalonamento híbrida DFPQ (*Deficit Fair Priority Queue*) que aloca largura de banda de forma integrada para o *downlink subframe* e *uplink subframe*, diferente deste trabalho cuja proposta de alocação de banda para o *uplink subframe* é efetuada de forma independente. Além disso, sua avaliação de desempenho não apresenta resultados em relação ao parâmetro de atraso, que é essencial para verificar o nível de QoS oferecido para as classes de tempo real.

Outra proposta que utiliza *software* de simulação de redes de computadores proprietário é apresentada em [63]. O mecanismo de escalonamento *uplink* da SS é baseado na disciplina PQ híbrida. O mecanismo de escalonamento *uplink* da BS é baseado na política *max-min fairness*. O trabalho não apresenta uma descrição substancial a respeito da operação lógica e da implementação das políticas de escalonamento utilizadas. Os resultados analisados não consideram a classe nrtPS.

Em [64] é proposto um mecanismo de escalonamento *uplink* da BS cuja política de atendimento é baseada em prioridades, porém não das filas, mas diretamente das SSs. A prioridade de cada SS é indicada por um parâmetro dinâmico definido pelo algoritmo, calculado de acordo com a largura de banda total requisitada e o tempo em espera para uma SS ser atendida. O algoritmo define uma maneira de utilizar os *mini-slots* que restam no *uplink subframe*, mas que são utilizados apenas se existirem SSs com *backlog*, diferente da proposta deste trabalho que aloca os *mini-slots* restantes mesmo que não exista *backlog*, por isso são pré-alocados. Para melhorar a eficiência da proposta de escalonamento, é integrado um mecanismo de ajuste de *polling*. O principal objetivo da proposta é melhorar o desempenho do tráfego da classe nrtPS sem degradar os níveis de QoS do tráfego da classe rtPS. Não é apresentada uma análise de resultados para as classes UGS e BE.

No trabalho de [65] é proposta uma arquitetura de QoS para o protocolo MAC IEEE 802.16. Porém, não é apresentada uma descrição substancial a respeito da política utilizada no mecanismo de escalonamento da BS para o tráfego *uplink* (BS-UL). No caso do mecanismo de escalonamento SS-UL é proposta uma política baseada em PQ híbrida, com cinco filas de prioridades, que utilizam políticas específicas como, por exemplo, WFQ e WRR, de acordo com a classe de serviço armazenada. Essas políticas específicas adotadas para atender as

diferentes classes em filas dedicadas apresentam uma implementação mais complexa do que as adotadas na proposta deste trabalho. Não são apresentados resultados e nem tampouco uma análise de desempenho, que são propostas para trabalhos futuros.

Diversas propostas baseadas em PQ como, por exemplo, [3] [45] [47] [62] [63] [66] [67], empregam disciplinas como, por exemplo, EDF, WRR, WFQ, ou políticas semelhantes para atender as diferentes classes armazenadas em filas dedicadas. No entanto, nenhum desses trabalhos apresenta uma política efetiva para aproveitar os *mini-slots* restantes no *uplink subframe*, após o escalonador atender todas as suas filas, com o objetivo de aumentar a eficiência de utilização dos recursos no enlace *uplink* através de uma técnica que aloca a largura de banda para as conexões da rede de maneira antecipada (pré-aloca) antes de receber a requisição de banda, como é apresentada neste trabalho. Além disso, o mecanismo proposto possui uma implementação relativamente mais simples do que a maioria dos trabalhos relacionados, que é uma característica apresentada em propostas recentes, como no trabalho [68].

4.5. Considerações Finais

Neste Capítulo, inicialmente foram apresentados os principais desafios para o projeto de um mecanismo de escalonamento da BS para o tráfego *uplink* das redes WiMAX e, na sequência a solução proposta.

A natureza operacional do escalonador BS-UL é mais complexa do que a dos demais escalonadores das redes IEEE 802.16 porque a BS depende das mensagens de requisição de largura de banda enviadas pelas SSs para efetuar a alocação de largura de banda no *uplink*. Além disso, os recursos no enlace *uplink* são escassos e são compartilhados por vários usuários que possuem uma demanda com grande volume e diversidade de tráfego.

Como o padrão deixa em aberto a política de escalonamento a ser utilizada, foi proposta neste Capítulo uma extensão do mecanismo de escalonamento BS-UL apresentado em [3], que é baseado na disciplina de escalonamento PQ (*Priority Queuing*).

A proposta deste trabalho apresenta uma estrutura com quatro filas de prioridades, com separação das requisições dos tipos de tráfego nRT e RT, diferente do mecanismo base que possui três filas e as requisições nrtPS e rtPS são armazenadas nas mesmas filas. O mecanismo proposto ainda utiliza um esquema que aproveita os *mini-slots* restantes no *uplink subframe*, após o escalonador atender todas as suas filas, diretamente para a transmissão de

dados das SSs, pré-alocando esses *mini-slots* livres para os fluxos da classe nrtPS, que é compensada pelo fato de suas requisições não migrarem para a fila de maior prioridade como no mecanismo base.

O objetivo da criação da nova fila foi isolar e proteger o tráfego da classe rtPS, que não é tolerante ao atraso, do tráfego da classe nrtPS. O objetivo do esquema de pré-alocação definido para a classe nrtPS foi aumentar a eficiência de utilização dos recursos no *uplink*.

Através da análise dos resultados, que serão apresentados no próximo Capítulo, será possível comprovar a eficiência da proposta apresentada neste Capítulo.

Capítulo 5

ANÁLISE DE DESEMPENHO DO ESCALONADOR PROPOSTO

5.1. Introdução

Neste Capítulo é avaliada a proposta do mecanismo de escalonamento descrita no Capítulo 4 por meio de modelagem e simulação. Os resultados obtidos a partir do mecanismo proposto são comparados com os resultados obtidos pelo mecanismo base proposto em [3], em relação aos parâmetros de vazão média, atraso médio, utilização percentual de recursos e percentual de descarte de pacotes, nos experimentos de simulação realizados para diferentes cenários.

Este Capítulo está estruturado da seguinte maneira: na Seção 5.2 são apresentadas as características básicas da ferramenta utilizada para a modelagem e simulação, na Seção 5.3 é descrito o ambiente de rede, os parâmetros de simulação utilizados e as características das fontes de tráfego; na Seção 5.4 são apresentados e comparados os resultados obtidos entre o mecanismo proposto e o mecanismo base, em cenários com características distintas. Finalmente, na Seção 5.5 são feitas as considerações gerais sobre o Capítulo.

5.2. O Simulador para o Ambiente de Rede IEEE 802.16 (WiMAX)

O mecanismo de escalonamento proposto é modelado e avaliado através do *software* de simulação de redes de computadores NS-2 (*Network Simulator-2*) [69], com o módulo simulador WiMAX proposto em [4], que foi baseado no módulo DOCSIS apresentado em [70]. Nessa versão do módulo WiMAX não são implementados os mecanismos de controle de admissão de conexões (CAC) e de policiamento de tráfego. Portanto, os mecanismos de escalonamento implementados são os principais responsáveis pela provisão de QoS para o tráfego modelado nos ambientes de simulação.

A escolha de uma ferramenta computacional para a simulação de redes de computadores foi baseada na conveniência de sua implementação, que não apresenta a necessidade de montagem e adaptação dos componentes físicos que poderiam despendar mais tempo, espaço e recursos do que através do modelo lógico oferecido pelo *software*. Além disso, o *software* de simulação de redes de computadores escolhido, NS-2, é amplamente utilizado pela comunidade acadêmica e por desenvolvedores de soluções para redes porque apresenta um código aberto com estrutura modular, que pode ser alterada através da modificação de seus módulos existentes ou da inclusão de novos módulos, oferece suporte para diversas tecnologias de redes, e é suportado por diversos sistemas operacionais como, por exemplo, o Linux.

O NS-2 é uma ferramenta baseada em eventos, cujo algoritmo principal é uma estrutura hierárquica de classes definidas em linguagem de programação C++ orientada a objetos, que é compilada e eficiente para se trabalhar com grandes fluxos de dados; e utiliza a linguagem OTcl (*Object Tool Command Language*), que é interpretada e interativa, como a interface para geração de *scripts* de simulação onde são comuns as alterações dos parâmetros do cenário modelado.

A manipulação dos dados obtidos através da simulação (*traces*) foi executada a partir de *scripts* AWK (Aho, Weinberger and Kernighan), que permitem formatar os resultados de maneira mais consistente para efetuar a sua análise.

5.3. Ambiente de Rede e Parâmetros de Simulação

Para fins de análise da provisão de QoS em redes IEEE 802.16 com arquitetura de rede PMP e modo TDD, foram definidos cenários de simulação com diferentes condições para o tráfego *uplink*.

Em todos os cenários, foram utilizados modelos de rede com as mesmas características apresentadas na Tabela 5.1, diferenciando-se em relação à quantidade de SSs, todas elas uniformemente distribuídas em relação à BS, como mostrado na Figura 5.1, aos tipos de tráfego de serviço produzidos pelas SSs, e aos parâmetros como, por exemplo, a MRTR e a capacidade dos *buffers*. Na modelagem realizada, o canal de comunicações foi considerado ideal, desprezando-se as características do meio físico, pois o foco desse trabalho é a camada MAC.

Tabela 5.1: Parâmetros de Configuração da Rede.

Parâmetros	Valor
Arquitetura da Rede	PMP
Duplexação	TDD
Tamanho do <i>Frame</i>	5ms
Capacidade do Canal	40Mbps
Nº <i>Mini-Slots</i> de Contenção	3
Tempo de Simulação	1000s

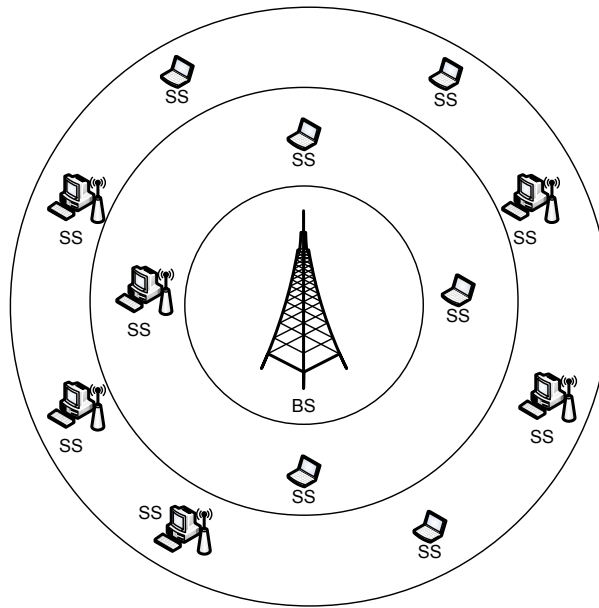


Figura 5.1: Cenário de Simulação de uma Rede IEEE 802.16 PMP TDD.

Cada SS possui apenas uma aplicação. As aplicações são representadas pelas fontes de tráfego de voz, vídeo, transferência de arquivos via protocolo FTP e *Web* que são servidas, respectivamente pelos serviços UGS, rtPS, nrtPS e BE. Cada serviço adotou um Agente Gerador de Tráfego (AGT) com características específicas. O tráfego de voz é modelado por uma fonte *on/off* cuja duração segue uma distribuição exponencial com médias iguais a 1,2 e 1,8 segundos, respectivamente, com pacotes de 66 *bytes* gerados a cada 20 ms durante o período *on* [71]. O tráfego de vídeo foi obtido a partir de *traces* de vídeos reais cujas características são apresentadas na Tabela 5.2 [72]. O tráfego *Web* é modelado por um AGT que utiliza uma distribuição híbrida Lognormal/Pareto, cujo corpo corresponde a uma área de 0,88 e é modelado por uma distribuição Lognormal com média igual a 7247 *bytes*, enquanto que a cauda é modelada por uma distribuição Pareto com média de 10558 *bytes* [73]. O tráfego de transferência arquivos via FTP é modelado por um AGT que utiliza uma

distribuição exponencial com média de 512 *Kbytes*, cujo valor da taxa de serviço é definido em cada cenário.

Tabela 5.2: Características dos *Traces* de Vídeos Utilizados nas Simulações [3].

Vídeo	Taxa de Pico (Mbps)	Variância dos <i>Frames</i>	Hurst
<i>Baseball</i>	1,859760	38430999,822669	0,684708
<i>Lecture Gupta</i>	2,218080	71620243,645425	0,777175
<i>Friends</i>	1,175520	71620243,645425	0,777175
<i>Lady Tramp</i>	0,910080	82822030,584567	0,813108
<i>Ice Age</i>	1,058400	11347671,977999	0,834963
<i>Dreamcatcher</i>	0,932400	8685118,764458	0,700975

Para os fluxos de serviço da classe UGS o intervalo entre os *grants* de dados periódicos foi definido em 20 ms. O intervalo destinado para *polling unicast* (*grants* periódicos) para os fluxos da classe rtPS é igual a 20 ms e para os fluxos nrtPS é igual a 1 s. O requisito de atraso máximo limitado para os fluxos da classe rtPS é igual a 100ms. O requisito de banda mínima (MRTR) para os fluxos rtPS e nrtPS são especificados em cada cenário simulado.

A capacidade dos *buffers* utilizados nos elementos da rede é definida em cada cenário. Cada *buffer* possui um processo de alocação dinâmica de memória para armazenar um número inteiro de pacotes de tamanhos variáveis. A política de descarte aplicada em cada fila foi *DropTail*, ou seja, na falta de espaço de armazenamento o último pacote é descartado.

Funções especiais como a fragmentação, a concatenação e *piggybacking*, foram habilitadas, de acordo com as recomendações do padrão IEEE 802.16 para cada classe.

5.4. Apresentação e Análise dos Resultados

Para verificar a eficiência do mecanismo proposto foram modelados quatro cenários, basicamente comparando o desempenho da proposta deste trabalho com o desempenho da proposta base apresentada em [3].

Particularmente, o objetivo da modelagem do primeiro cenário foi analisar se o compartilhamento da largura de banda entre os fluxos nrtPS é proporcional ao seu peso, e a justiça na distribuição dos recursos da rede entre os usuários. Na modelagem do segundo cenário, o objetivo foi analisar a capacidade do mecanismo de escalonamento oferecer

tratamento diferenciado, de acordo com a prioridade e os requisitos de QoS de cada tipo de tráfego, para as classes UGS, rtPS, nrtPS e BE.

Nos dois primeiros cenários, o tamanho dos *buffers* é superdimensionado (1000 pacotes), pois o objetivo é efetuar uma comparação de desempenho entre a proposta deste trabalho e o mecanismo base mais fidedigna possível em relação aos parâmetros utilizados em [3]. Porém, no terceiro cenário, cujo objetivo é analisar o efeito do dimensionamento de um *buffer* de tamanho mais realístico, foram definidas novas capacidades.

Na modelagem do quarto cenário, o objetivo foi particularmente analisar o nível de proteção oferecida ao tráfego rtPS diante do aumento da carga de tráfego na rede produzida pelo tráfego nrtPS.

Portanto, a avaliação de desempenho é efetuada através da comparação dos resultados apresentados nas figuras 5.2 até 5.13 das próximas subseções, entre o mecanismo base proposto em [3], cuja legenda é definida por PQ-3F, pois se trata de um mecanismo de escalonamento PQ com três filas, e a sua extensão proposta nesse trabalho, que é definida pela legenda PQ-4F, pois é composto por quatro filas em sua estrutura.

5.4.1. Cenário I

O primeiro cenário de rede WiMAX modelado, consiste de uma BS associada a um número inicial de 10 SSs, variando até um número de 30 SSs. Cada SS possui um único fluxo de serviço nrtPS associado a uma aplicação modelada por um AGT que fornece uma taxa em torno de 600 Kbps. Exceto os fluxos das estações SS5 e SS10 que apresentam uma MRTR igual a 500 Kbps, todos os fluxos de serviço das demais SSs apresentam uma MRTR com valor igual a 200 Kbps. Isto foi feito para verificar se o compartilhamento da largura de banda entre esses fluxos é proporcional ao seu peso, que é diretamente proporcional ao parâmetro MRTR.

Os resultados apresentados na Figura 5.2 revelam que, entre uma quantidade de 20 SSs até 30 SSs presentes na rede, o desempenho de vazão média apresentado pelo mecanismo PQ-4F (PQ-4F Média) foi sempre superior ao desempenho apresentado pelo mecanismo PQ-3F (PQ-3F Média).

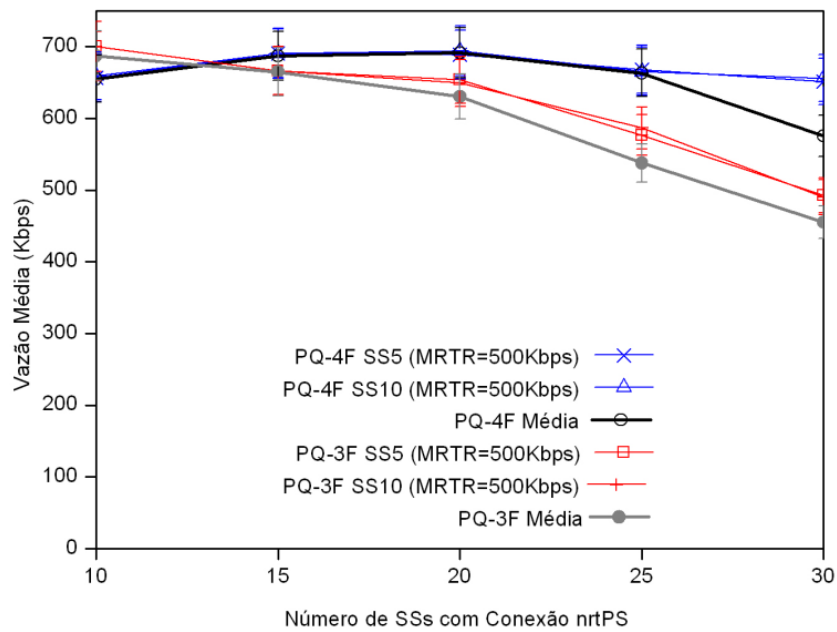


Figura 5.2: Vazão das SSs nrtPS (PQ-4F vs. PQ-3F).

Comparando-se a vazão obtida pelas SSs com maior MRTR (SS5 e SS10) em relação à vazão média de todas SSs (PQ-4F Média e PQ-3F Média), concluiu-se que ambos os mecanismos são capazes de priorizar o atendimento daquelas SSs cujos fluxos apresentam maior peso, garantindo a elas maior vazão quando os recursos na rede tornaram-se escassos.

A Figura 5.3 apresenta a vazão de dez conexões concorrentes nrtPS, a fim de verificar se o mecanismo PQ-4F distribui a largura de banda de maneira justa entre as conexões nrtPS que possuem os mesmos parâmetros de QoS (MRTR). Os valores obtidos para a maioria das SSs apresentam pouca diferença uns em relação aos outros, com uma diferença mais acentuada para aquelas SSs cujo valor do parâmetro MRTR é igual a 500 Kbps em relação às demais SSs que apresentam um valor igual a 200 Kbps, a partir de 25 SSs presentes na rede.

A partir de 25 SSs, o congestionamento da rede não permitiu que o tráfego gerado por todas as SSs fosse atendido por completo, e a vazão obtida pelas SSs cujo valor do parâmetro MRTR é igual a 500 Kbps foi superior ao das demais SSs cujo MRTR é igual a 200 Kbps, porque o algoritmo de escalonamento implementado para atender às requisições da classe nrtPS, definido no Capítulo 4, determina a prioridade de atendimento de acordo com esse parâmetro. Assim, conclui-se que o algoritmo proposto é eficiente em relação à priorização de atendimento dos fluxos nrtPS que apresentam necessidades de taxas mínimas de serviço diferentes, e é justo em relação ao compartilhamento da largura de banda entre os fluxos que apresentam necessidades semelhantes.

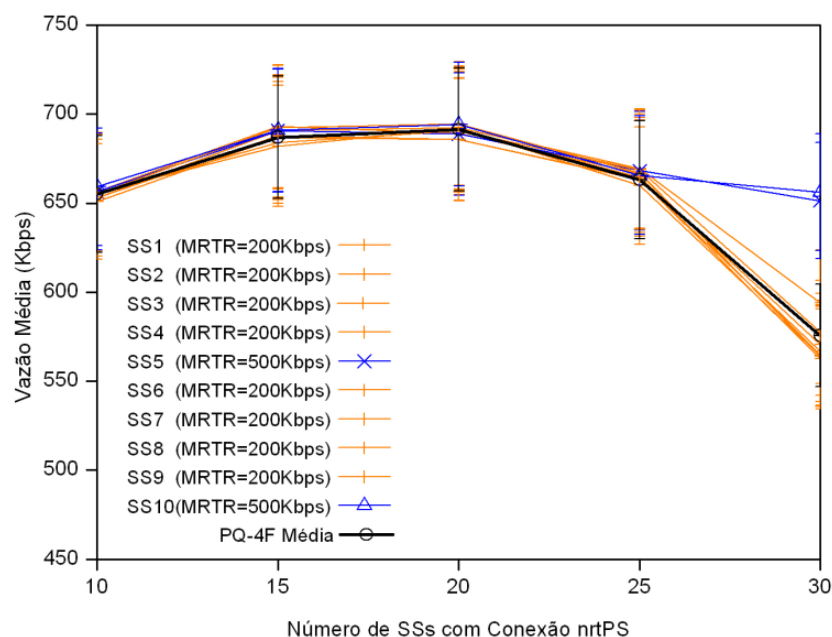


Figura 5.3: Vazão de Dez Conexões Concorrentes nrtPS (PQ-4F).

Todos os resultados obtidos foram gerados utilizando o método de replicação, executando cada simulação dez vezes com sementes diferentes. Nas Tabelas 5.3 e 5.4 são mostrados os intervalos de confiança (95%) da vazão apresentada nas Figuras 5.2 e 5.3.

Tabela 5.3: Intervalos de Confiança de Vazão (Figura 5.2).

SS	IC(95%)
PQ-4F SS5(MRTR=500Kbps)	671,059±22,445
PQ-4F SS10(MRTR=500Kbps)	673,258±22,451
PQ-4F Média (Todas as SSs)	654,495±57,943
PQ-3F SS5(MRTR=500Kbps)	617,979±103,540
PQ-3F SS10(MRTR=500Kbps)	615,717±98,757
PQ-3F Média (Todas as SSs)	595,041±120,271

Tabela 5.4: Intervalos de Confiança de Vazão (Figura 5.3).

SS	IC(95%)
SS1(MRTR=200Kbps)	649,935±63,556
SS2(MRTR=200Kbps)	650,836±62,837
SS3(MRTR=200Kbps)	659,928±50,003
SS4(MRTR=200Kbps)	654,104±64,797
SS6(MRTR=200Kbps)	655,393±56,528
SS7(MRTR=200Kbps)	653,715±60,405
SS8(MRTR=200Kbps)	651,329±61,332
SS9(MRTR=200Kbps)	654,649±66,344

A Figura 5.4 apresenta os resultados referentes ao percentual de utilização da capacidade máxima do canal de *uplink*, pelos mecanismos de escalonamento PQ-3F e PQ-4F, com a variação da carga na rede. O percentual de utilização foi determinado através de equação 5.1, que depende da razão entre a quantidade total de dados transmitidos com sucesso pelo mecanismo de escalonamento ao longo de seu tempo de operação, pela taxa de transmissão do canal *uplink* (20 Mbps).

$$Utilização = \frac{D_{UL} / \Delta t}{R_{UL}} \cdot 100\% \quad (5.1)$$

Onde: *Utilização* representa o percentual de utilização, D_{UL} é total de dados transmitidos no *uplink* (bits), Δt é o tempo total (segundos) e R_{UL} é a taxa de transmissão do *uplink* (bps).

Na Figura 5.4, verifica-se que, com a variação do número de SSs presentes na rede, a partir de um número equivalente a 20 SSs, o desempenho do mecanismo PQ-4F foi muito superior ao desempenho do mecanismo PQ-3F.

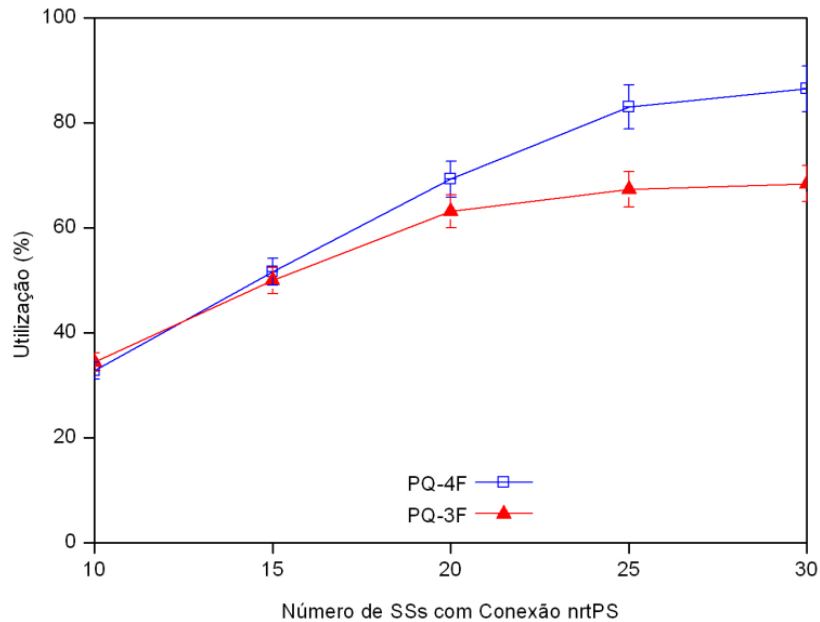


Figura 5.4: Percentual de Utilização (PQ-4F vs. PQ-3F).

Após o escalonador atender suas filas, o mecanismo PQ-4F pré-aloca os recursos que restam diretamente para a transmissão de dados das SSs, antes de receber o pedido de requisição de largura de banda, evitando a necessidade do envio desse tipo de mensagem.

Assim, o mecanismo PQ-4F é capaz de aumentar a eficiência de utilização dos recursos no enlace *uplink* e utilizar menos recursos para as operações de controle e gerenciamento, como é mostrado na Figura 5.5, que representa o percentual de pacotes de gerenciamento (MAC) em relação ao total de pacotes no *uplink*. Portanto, o PQ-4F oferece uma utilização de recursos mais eficiente para a transmissão de dados do que o PQ-3F.

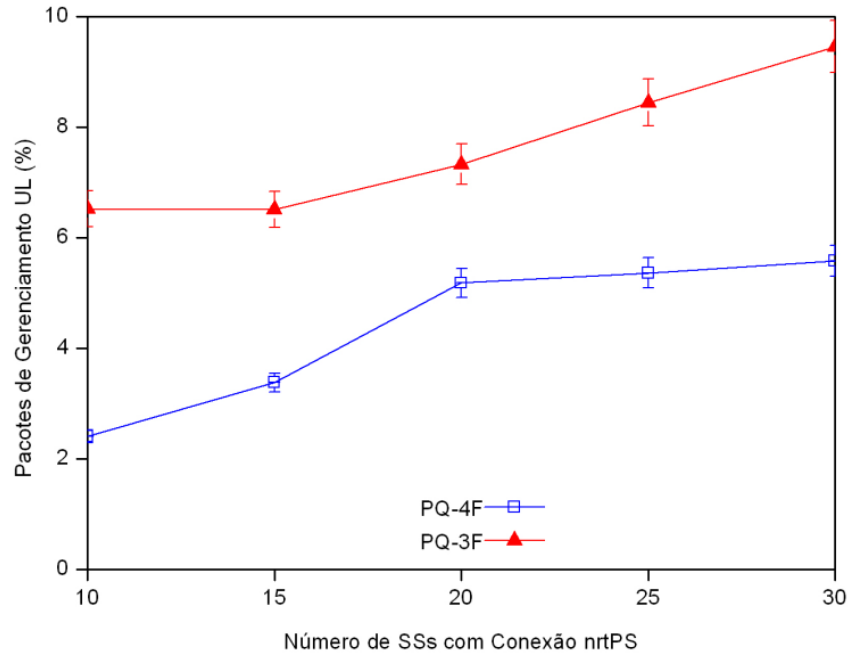


Figura 5.5: Percentual de Pacotes de Gerenciamento no *Uplink* (PQ-4F vs. PQ-3F).

5.4.2. Cenário II

O segundo cenário modelado, apresenta inicialmente 20 conexões BE, 20 conexões nrtPS, 6 conexões rtPS, e de 15 à 35 conexões UGS. A taxa média produzida pelos AGTs dos fluxos BE e nrtPS foi em torno de 200 Kbps para cada SS. O valor do parâmetro MRTR para todas as SSs com o fluxo nrtPS é igual a 200 Kbps e, para as 6 SSs rtPS foram definidos os seguintes valores de MRTR: 900, 900, 500, 200, 200 e 550 Kbps, respectivamente.

Na Figura 5.6, verifica-se que a vazão obtida pelas SSs da classe nrtPS, através do mecanismo PQ-4F, apresenta um desempenho levemente superior à vazão obtida através do PQ-3F, mas em relação às SSs da classe BE a melhoria de desempenho foi mais significativa. Como o PQ-4F atende a toda largura de banda solicitada pelo fluxo nrtPS através de um esquema de utilização mais eficiente, restam mais recursos para serem aproveitados pelos fluxos BE no *uplink*.

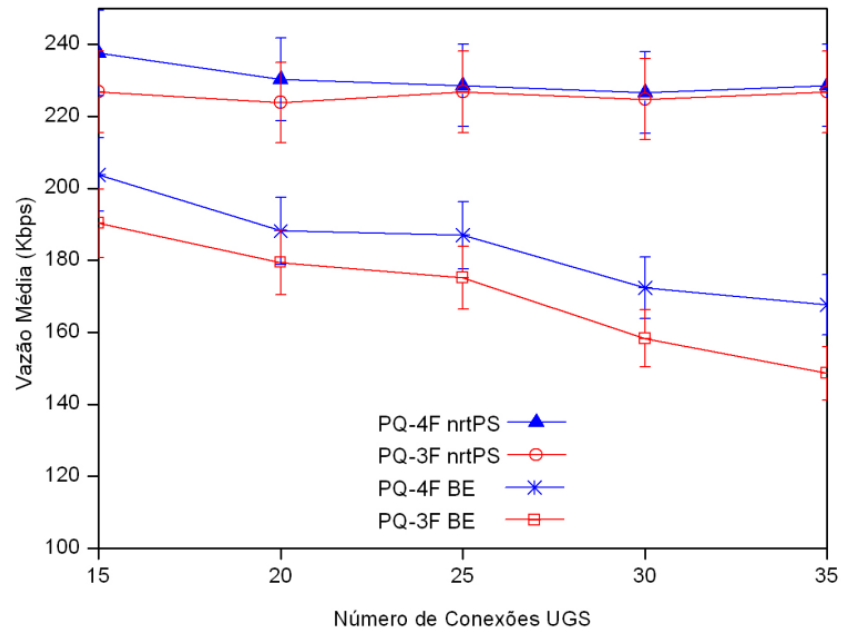


Figura 5.6: Vazão Média das Conexões nrtPS e BE (PQ-4F vs. PQ-3F).

A melhoria de desempenho em relação à vazão alcançada para as classes nRT, a partir do mecanismo de escalonamento PQ-4F, foi efetiva e não prejudicou os outros tráfegos RT. Pelo contrário, o fluxo rtPS apresentou uma melhoria de desempenho em relação ao atraso médio, como é mostrado na Figura 5.7.

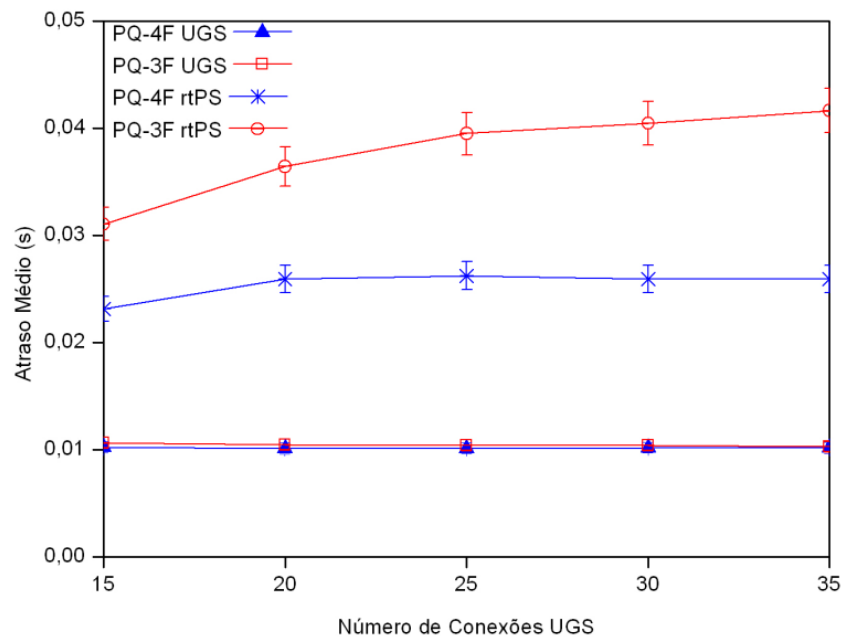


Figura 5.7: Atraso Médio das SSs UGS e rtPS (PQ-4F vs. PQ-3F).

O desempenho de atraso médio do mecanismo PQ-4F foi melhor do que o desempenho do mecanismo PQ-3F porque no PQ-3F as requisições rtPS e nrtPS, que são armazenadas na mesma fila, são atendidas ao mesmo tempo e; no mecanismo PQ-4F as requisições rtPS, que são armazenadas na fila de alta prioridade, são atendidas primeiro do que as requisições nrtPS, que são armazenados na fila de baixa prioridade e não migram para a fila de alta prioridade como no PQ-3F, fato que não compromete o desempenho de vazão da classe nrtPS, como apresentado na Figura 5.6, porque o mecanismo PQ-4F oferece para essa classe a alternativa de pré-alocação.

A quantidade máxima de SSs presentes no Cenário *I* é igual a 30, enquanto que no Cenário *II* essa quantidade é igual a 81. À medida que mais usuários entram na rede, aumenta o número de mensagens de gerenciamento produzidas, e inevitavelmente a competição pela largura de banda, fazendo-se necessário um esquema eficiente de utilização de recursos no enlace *uplink*, como o esquema proposto de pré-alocação, para atender a essa demanda com níveis satisfatórios de QoS.

No Cenário *II*, além do número de SSs presentes na rede ser superior em relação ao número do Cenário *I*, novos tipos de tráfego foram gerados, e classes de maior prioridade de atendimento como UGS e rtPS passaram a competir pela largura de banda disponível. Mesmo assim, a vazão média da classe nrtPS foi adequada, pois a taxa de dados gerada pelos AGTs, que foi definida em torno de 200 Kbps para cada conexão, foi atendida.

5.4.3. Cenário *III*

O terceiro cenário de rede modelado consiste de uma BS associada a um número inicial de 20 SSs, variando até um número igual a 60 SSs com fluxo nrtPS. Cada SS é associada a um AGT que fornece uma taxa em torno de 600 Kbps. A MRTR de todos os fluxos é igual a 200 Kbps.

A modelagem deste cenário tem o objetivo de analisar o desempenho do mecanismo de escalonamento PQ-4F em relação à variação dos valores de parâmetros estruturais dos elementos da rede como, por exemplo, o tamanho dos *buffers*. Foram definidos quatro valores diferentes para a capacidade dos *buffers*: 100, 50, 25 e 10 pacotes.

Na Figura 5.8 é apresentado o percentual de pacotes descartados nos *buffers*, que é a relação entre o número de pacotes descartados com o total de pacotes gerados. Verificando a Figura 5.8, conclui-se que, além do aumento do número de SSs, a redução do tamanho dos

buffers contribui para o aumento do percentual de descarte, e consequentemente compromete o desempenho da vazão média das SSs presentes na rede, como apresentado na Figura 5.9.

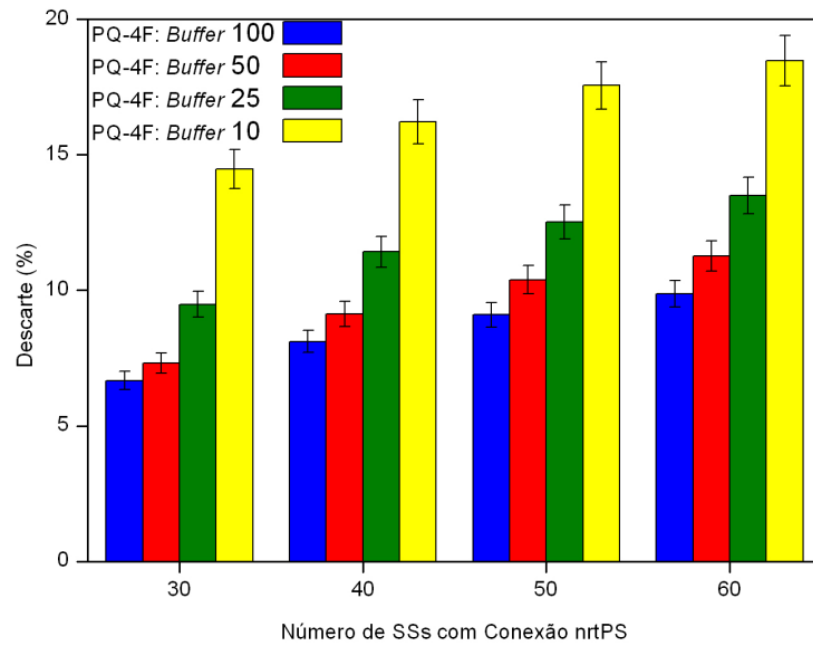


Figura 5.8: Percentual de Pacotes Descartados nos *Buffers* (PQ-4F).

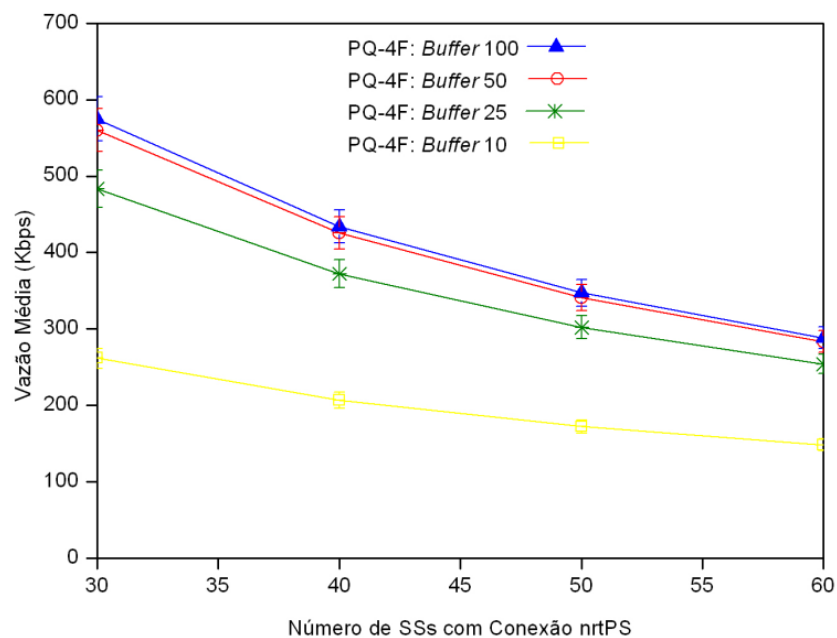


Figura 5.9: Vazão Média das Conexões nrtPS (PQ-4F).

Na Figura 5.10 é apresentada a comparação do percentual de pacotes descartados, e na Figura 5.11 é apresentada a comparação do parâmetro de vazão média, entre os mecanismos

PQ-3F e PQ-4F, em relação à capacidade dos *buffers*, para uma quantidade de 60 SSs presentes na rede.

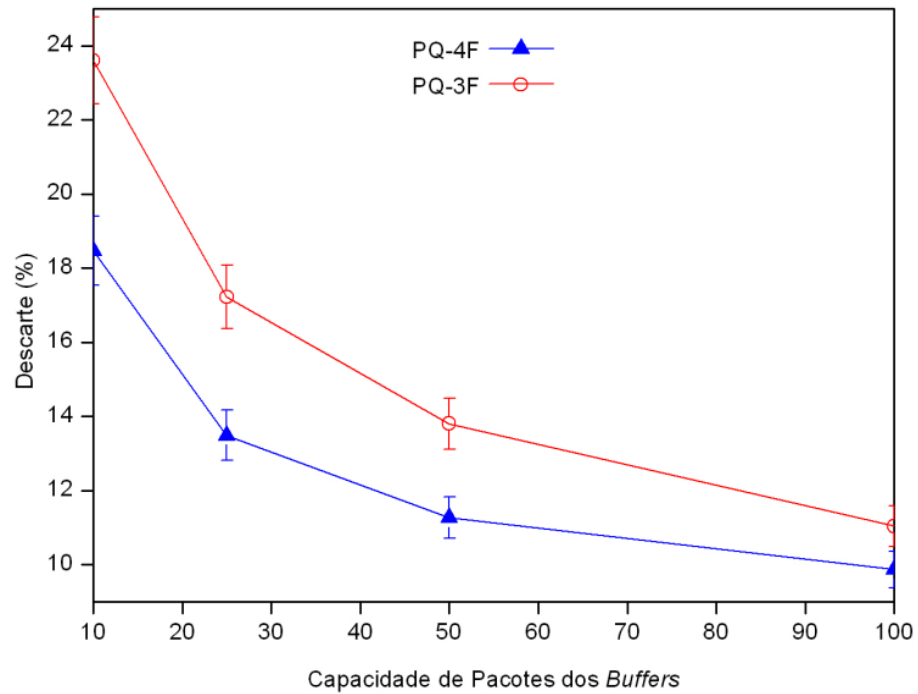


Figura 5.10: Percentual de Pacotes Descartados nos *Buffers* (PQ-4F vs. PQ-3F).

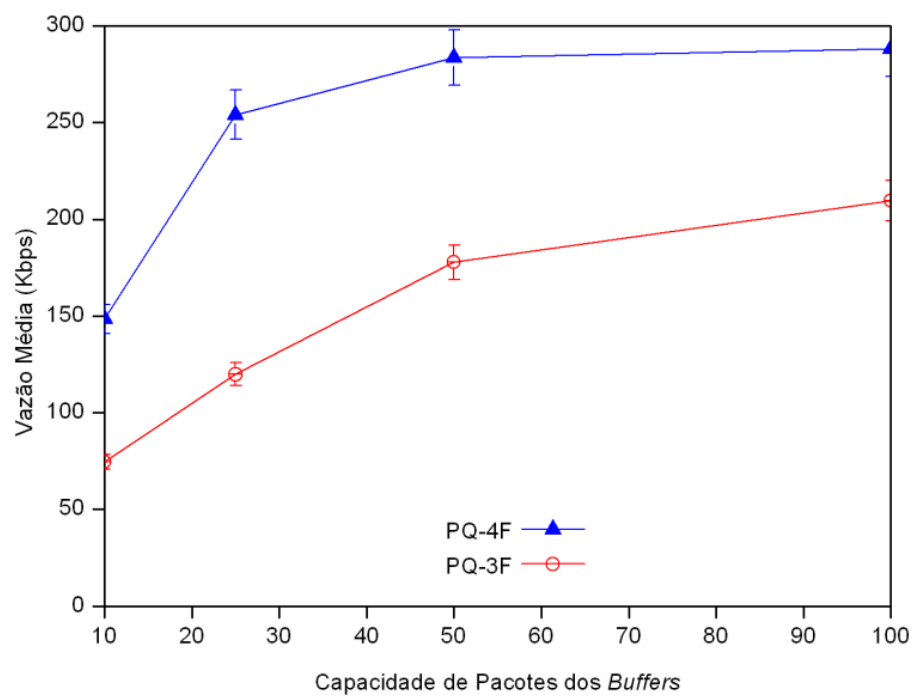


Figura 5.11: Vazão Média em Relação à Capacidade dos *Buffers* (PQ-4F vs. PQ-3F).

A partir da análise das Figuras 5.10 e 5.11, verifica-se que com o aumento da capacidade dos *buffers* dos elementos da rede, o percentual de pacotes descartados diminui, e a vazão média das 60 SSs presentes na rede aumenta. Em ambas as figuras, o desempenho do PQ-4F foi melhor do que o desempenho apresentado pelo PQ-3F. Essa diferença de desempenho ocorre porque o mecanismo PQ-4F aproveita os recursos que restam diretamente para a transmissão de dados, como já foi justificado na subseção 5.4.1.

5.4.4. Cenário IV

O objetivo da modelagem deste cenário é avaliar o nível de proteção oferecido ao tráfego rtPS pelos mecanismos de escalonamento comparados. Esse cenário consiste de uma BS associada a um número de 6 SSs com conexões rtPS, com as mesmas características daquelas apresentadas no Cenário II, e um número inicial de 10 SSs, variando até um número de 30 SSs com conexões nrtPS, cujo tráfego é produzido por um AGT que fornece uma taxa em torno de 800 Kbps. Todos os fluxos nrtPS apresentam uma MRTR igual a 200 Kbps.

Na Figura 5.12 é mostrado que o desempenho do parâmetro de atraso médio do mecanismo PQ-4F manteve-se mais estável que o do mecanismo PQ-3F, pois apenas as requisições da classe rtPS podem migrar para fila de maior prioridade do PQ-4F, e portanto nesse nível não sofrem a concorrência das requisições nrtPS, como no mecanismo PQ-3F.

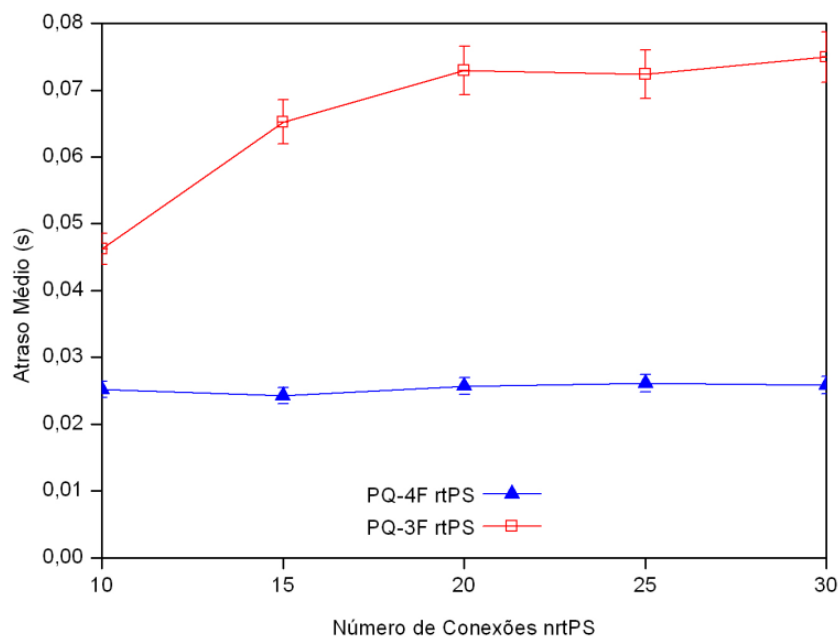


Figura 5.12: Atraso Médio das Conexões rtPS (PQ-4F vs. PQ-3F).

Tanto o mecanismo PQ-4F quanto o mecanismo PQ-3F conseguem atender a taxa de serviço requisitada pela classe rtPS, e por isso a vazão média produzida por ambos é semelhante, como é apresentado na Figura 5.13. No entanto, o desempenho da vazão média da classe nrtPS produzido pelo mecanismo PQ-4F é melhor do que o desempenho do mecanismo base PQ-3F porque possui um esquema mais eficiente de utilização de recursos no *uplink*, capaz de compensar a desvantagem das requisições nrtPS não poderem migrar para a fila de maior prioridade com as requisições rtPS como no mecanismo base.

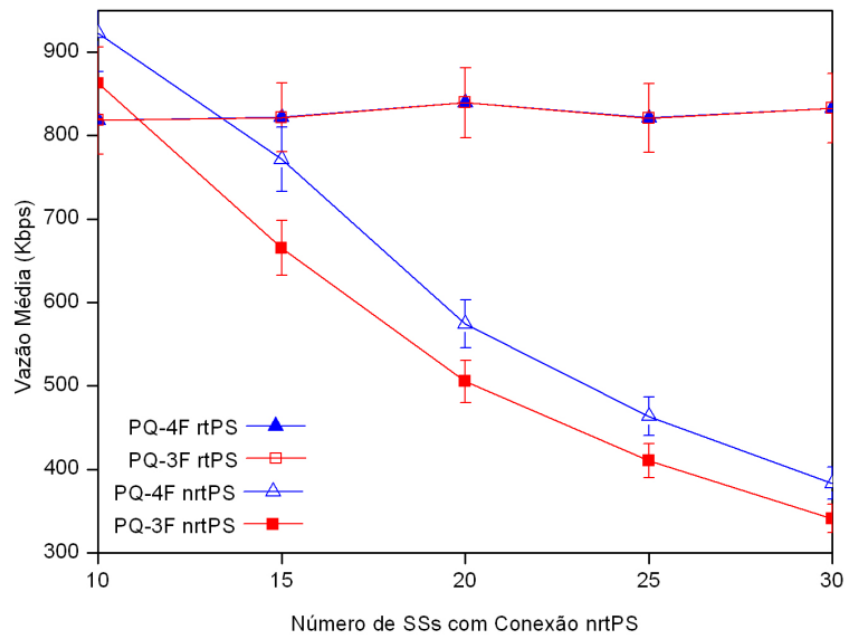


Figura 5.13: Vazão Média das Conexões rtPS e nrtPS (PQ-4F vs. PQ-3F).

O período definido para o *polling unicast* da classe rtPS foi cinquenta vezes menor do que o período definido para o *polling unicast* da classe nrtPS, cuja vazão média diminuiu à medida que a carga de tráfego na rede aumentou.

De acordo com a análise de resultados apresentada nesse cenário, o mecanismo proposto melhorou o desempenho do atraso médio da classe rtPS, evitando que ele fosse degradado pelo aumento da carga do tráfego não tempo real, e melhorou o desempenho da vazão da classe nrtPS.

5.5. Considerações Finais

Neste Capítulo realizou-se a apresentação e análise de resultados obtidos a partir da modelagem e simulação do mecanismo de escalonamento da BS para o tráfego *uplink* proposto no Capítulo 4, para diferentes cenários com características de tráfego e carga diversas.

O desempenho apresentado pelo mecanismo de escalonamento proposto, PQ-4F, foi superior ao desempenho apresentado pelo mecanismo base, PQ-3F, em relação aos parâmetros de vazão média, atraso médio, utilização percentual de recursos, carga de pacotes de gerenciamento e descarte de pacotes nos *buffers*, pois prioriza o atendimento das requisições do tráfego tempo real e pré-aloca os recursos livres diretamente para a transmissão de dados; que permite diminuir o atraso de acesso da rede, minimizar o problema da diferença entre a necessidade de largura de banda percebida pela BS e a necessidade atual da SS, reduzir a carga de mensagens de gerenciamento, o *backlog* e o descarte de pacotes nos *buffers* das estações da rede e, sobretudo, aumentar a eficiência de utilização de recursos no enlace *uplink*.

A análise dos resultados obtidos a partir do mecanismo de escalonamento baseado no esquema de pré-alocação desenvolvido neste trabalho consolidou a idéia de que os mecanismos que aproveitam os recursos restantes no *uplink* diretamente para a transmissão de dados das SSs, como, também é o caso dos mecanismos propostos nos trabalhos apresentados em [74] [75], são capazes de melhorar a eficiência de utilização de recursos no enlace *uplink*.

Portanto, o mecanismo de escalonamento BS-UL proposto atingiu seus principais objetivos: oferecer tratamento diferenciado com priorização de tráfego e aumentar a eficiência de utilização dos recursos no enlace *uplink* sem degradar os níveis de QoS oferecidos aos usuários.

Capítulo 6

CONCLUSÕES GERAIS

Este trabalho abordou uma alternativa de rede de acesso banda larga sem fio (BWA) em franca expansão ao nível de pesquisa, desenvolvimento de soluções e implementação, comercialmente conhecida por WiMAX, e padronizada como IEEE 802.16.

Inicialmente, foram abordados os principais fatores que determinam a preferência do emprego das redes sem fio WiMAX para provisão de acesso banda larga frente às tecnologias tradicionais como, por exemplo, a mobilidade, o seu custo-benefício e facilidade de implantação e manutenção em relação às redes cabeadas tradicionais xDSL e *Cable Modem*, além da maior capacidade de abrangência em relação a das redes sem fio *Wi-fi*, a capacidade de atingir altas taxas de transmissão de dados superiores à das redes de telefonia celular, e principalmente a presença inerente de uma arquitetura de QoS. Portanto, são essas as razões pelas quais as redes WiMAX são consideradas como uma das tecnologias mais promissoras para atender o tráfego heterogêneo das redes de comunicação 4G.

Também foi mostrado que o padrão IEEE 802.16 especifica a camada física (PHY) e a camada de controle de acesso ao meio (MAC), é orientado à conexão, e que a comunicação entre os elementos da rede, estação base (BS) e estação de assinante (SS), é gerenciada através de fluxos de serviços, que são mapeados e atendidos de acordo com os requisitos de QoS das classes UGS, rtPS, nrtPS, BE, e uma quinta classe, ertPS, que foi definida a partir do padrão IEEE 802.16e. Portanto, o padrão é capaz de prover QoS para classes heterogêneas com requisitos diferentes. O tráfego de tempo real (RT) é mapeado para as classes UGS, ertPS ou rtPS, e o tráfego não tempo real (nRT) para as classes nrtPS ou BE. O tráfego de tempo real possui requisitos de QoS específicos e mais críticos como o atraso máximo limitado, e por isso precisa de tratamento diferenciado com priorização.

Foram apresentados os principais mecanismos que compõem a arquitetura de QoS definida na camada MAC do padrão IEEE 802.16 como, por exemplo, o mecanismo de controle de admissão de conexões (CAC), que limita o número de conexões na rede para evitar que a QoS exigida pelos fluxos das conexões ativas seja deteriorada, o mecanismo de

policiamento de tráfego que estabelece que a utilização de recursos por parte das aplicações seja mantida dentro dos níveis estabelecidos em contratos de SLA, e também, o mais importante, o mecanismo de escalonamento. O mecanismo de escalonamento, que é responsável por definir a ordem de envio dos pacotes armazenados nas filas da BS e das SSs, é implementado através de um algoritmo, que pode ser baseado em diversas disciplinas de escalonamento. Embora o padrão determine a necessidade estrutural de mecanismos de QoS, a escolha da política de operação é deixada a cargo dos desenvolvedores e fabricantes de equipamentos de rede a fim de que estabeleçam sua eficiência desejada.

Assim, considerou-se importante apresentar a teoria básica a respeito de algumas das principais disciplinas de escalonamento e o estado da arte da pesquisa em escalonamento em redes IEEE 802.16. Esse levantamento mostrou que uma única disciplina de escalonamento não é capaz de atender todos os requisitos de QoS dessas redes, e conseqüentemente o desenvolvimento de um mecanismo de escalonamento é uma tarefa complexa, cuja escolha da política mais adequada deve considerar o perfil de tráfego da rede, identificando quais são os requisitos mais ou menos críticos para seus usuários.

O mecanismo de escalonamento da BS para o tráfego *uplink* (BS-UL), proposto em [3], que foi implementado em um módulo WiMAX com código aberto disponível para a pesquisadores e desenvolvedores, tem a vantagem de oferecer suporte para todas as classes definidas pelo padrão IEEE 802.16d porque utiliza a disciplina de escalonamento PQ (*Priority Queuing*) híbrida, que é capaz de fornecer tratamento diferenciado com priorização, e políticas específicas para atender os diferentes tráfegos de acordo com seus requisitos de QoS. No entanto, ele não oferece um isolamento de tráfego completo porque as requisições das classes rtPS e nrtPS são armazenadas nas mesmas filas, e a maneira como são aproveitados os *mini-slots* que restam após o escalonador atender suas filas não é capaz de maximizar a eficiência de utilização dos recursos no enlace *uplink*.

A provisão de QoS em redes WiMAX é uma tarefa difícil, uma vez que sua largura de banda é compartilhada e limitada, e torna-se cada vez mais escassa em função do aumento da necessidade por acesso banda larga à Internet. Além disso, as mensagens de dados das aplicações de usuários transmitidas através das conexões de transporte competem pela largura de banda com as mensagens de controle e gerenciamento, transmitidas no mesmo canal *uplink* (*inband*) através das conexões de gerenciamento. Portanto, o desenvolvimento de soluções eficientes é essencial para garantir a escalabilidade e flexibilidade dessas redes.

Assim, neste trabalho foi apresentada uma proposta de extensão do mecanismo de alocação de largura de banda, escalonador BS-UL, apresentado em [3], com o objetivo de

oferecer tratamento diferenciado com priorização, isolando completamente os diferentes tráfegos através da separação das requisições das classes de serviço em filas dedicadas, bem como aumentar a eficiência de utilização dos recursos no enlace *uplink*, utilizando um novo esquema que pré-aloca os recursos que restam, após o escalonador atender as filas, diretamente para a transmissão de dados.

Diferentemente do mecanismo base apresentado em [3] que utiliza três filas de escalonamento, a proposta deste trabalho foi desenvolvida com uma estrutura com quatro filas, separando as requisições rtPS, atendidas primeiro na fila de alta prioridade, das requisições nrtPS, atendidas posteriormente na nova fila (baixa prioridade), porque a classe rtPS possui o requisito de QoS de atraso máximo limitado. Também foram implementadas políticas específicas para atender as diferentes classes armazenadas em filas dedicadas, de acordo com os requisitos de QoS definidos pelo padrão IEEE 802.16d. A política de atendimento desenvolvida para a nova fila determina que os fluxos com maior peso, proporcional à sua MRTR, recebam prioridade de atendimento, permitindo atender melhor a demanda diferenciada de cada aplicação que utiliza esse serviço.

Um novo esquema, chamado de pré-alocação, foi desenvolvido para compensar a classe nrtPS, cujas requisições não migram para a fila de alta prioridade como no mecanismo base apresentado em [3], e também para maximizar a eficiência de utilização dos recursos no enlace *uplink*, que é afetada pelo atraso de acesso da rede e pela diferença entre a necessidade de banda percebida pela BS e a necessidade imediata de banda das SSs, resultantes da dinâmica de operação dos processos de requisição e alocação de largura de banda.

A pré-alocação concede banda de maneira antecipada, evitando a necessidade do envio de mensagens de requisição de banda, e assim, diminuiu a quantidade produzida dessas mensagens na rede, como mostrado na Figura 5.5, bem como o atraso entre o momento que uma SS necessita transmitir seus dados e o momento que ela realmente pode transmiti-los, além de tornar mais preciso o processo de alocação de largura de banda, à medida que os novos pacotes que chegam às filas das SSs podem ser prontamente atendidos.

A apresentação e a análise dos resultados obtidos a partir do mecanismo de escalonamento proposto neste trabalho foram efetuadas por meio de modelagem e simulação, utilizando o *software* para simulação de redes de computadores NS-2 (*Network Simulator - 2*) com módulo WiMAX proposto em [4]. Foram modelados quatro cenários com fontes geradoras de tráfego com distribuição estocástica, baseadas em trabalhos sobre o estudo de modelagem de tráfego, que representem de maneira mais fidedigna as condições do tráfego das redes reais.

No cenário de rede modelado com a presença de tráfego homogêneo, representado pelas fontes que utilizam o serviço nrtPS, o mecanismo proposto neste trabalho conseguiu atender a banda mínima requisitada por todos os usuários, distribuir de maneira justa a largura de banda entre aqueles fluxos que possuem a mesma MRTR, bem como fornecer maior largura de banda para aqueles fluxos que possuem maior MRTR (maior peso). Além disso, o mecanismo proposto apresentou melhor desempenho de vazão média do que o da proposta base porque utiliza um esquema que pré-aloca os *mini-slots* restantes, após o escalonador atender as filas, diretamente para a transmissão dos dados das SSs com fluxos nrtPS.

No cenário de rede modelado com a presença de tráfego heterogêneo, o mecanismo proposto neste trabalho mostrou-se capaz de oferecer tratamento diferenciado, priorizando o atendimento das classes de tempo real, porque atende primeiramente as classes UGS e rtPS garantindo que os valores de atraso médio obtidos ficassem abaixo do valor definido pelo parâmetro de atraso máximo limitado, e ao mesmo tempo menor do que o valor do atraso médio obtido através do mecanismo base, para o caso da classe rtPS. Além disso, o mecanismo proposto ainda conseguiu atender a taxa de transmissão dos fluxos da classe nrtPS e apresentou desempenho de vazão média um pouco superior em relação ao desempenho do mecanismo base para a classe BE porque emprega um esquema de utilização de recursos no *uplink* mais eficiente, permitindo que sobrassem mais recursos para serem aproveitados pela classe BE.

Portanto, os resultados obtidos mostraram principalmente que o mecanismo proposto produziu um atraso médio menor para a classe rtPS, uma vazão média maior para as classes nrtPS e BE, sem prejudicar os níveis de QoS da classe UGS e; considerando o Cenário 1 de simulação com a presença de tráfego homogêneo nrtPS, aumentou a eficiência de utilização dos recursos no enlace *uplink* em cerca de 18%, como mostrado na Figura 5.4, em comparação com resultados obtidos através do mecanismo base proposto em [3].

Os principais desafios encontrados no processo de modelagem e simulação foram o processo de instalação e configuração do *software* para simulação de redes de computadores e seus componentes, tais como bibliotecas e classes adicionais para suportar o módulo WiMAX apresentado em [4]. Posteriormente, foi necessário pesquisar minuciosamente os valores dos parâmetros utilizados nos cenários de rede modelados no trabalho apresentado em [3], para que fossem reproduzidas as mesmas condições e, consequentemente garantir com segurança que os resultados obtidos e sua análise fossem fidedignos. Nas primeiras simulações ocorreram problemas devido às restrições da capacidade de processamento das máquinas utilizadas para executarem o *software* de simulação de redes de computadores. Assim, foi

necessário aumentar a capacidade de memória para a execução de programas previamente configurada nas máquinas, pois antes de efetuar esse *upgrade* diversas foram as vezes que se fez necessário refazer a simulação a partir do começo, pelo fato de ocorrerem erros por falta de capacidade de memória, em virtude do modo como foram modelados os *buffers* (alocação dinâmica de memória) utilizados para armazenamento de pacotes nos elementos da rede WiMAX, diante da carga de tráfego proposta pelo trabalho apresentado em [3].

Como trabalhos futuros, sugere-se associar ao mecanismo de escalonamento uma política que considera as condições de qualidade do sinal das SS no meio físico para tomada de decisão de escalonamento (*cross-layer*), bem como associar ao esquema de pré-alocação um método preditivo para tornar o processo mais flexível e eficiente, sendo capaz de atender outros tipos de tráfego em outros ambientes de rede, e assim, atender a maior variedade de perfis de tráfego possíveis das redes WiMAX. Também é sugerido mudar a maneira como são definidos os pesos associados aos fluxos nrtPS, através de parâmetros dinâmicos, para que possam refletir de maneira mais realística, o comportamento variável do tráfego produzido nessas redes.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] IEEE 802.16 Working Group. IEEE Standard for Local and Metropolitan Area Networks – Part. 16: Air Interface for Fixed Broadband Wireless Access Systems. Outubro 2004.
- [2] IEEE Std 802.16-2009, IEEE Standard for Local and Metropolitan Area Networks – Part 16: Air Interface for Broadband Wireless Access Systems, Maio 2009.
- [3] FREITAG, J. & DA FONSECA, N., Escalonamento com Qualidade de Serviço em redes IEEE 802.16, Anais do 25º Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC), pp. 723-736, Maio 2007.
- [4] FREITAG, J. & DA FONSECA, N., Um Módulo para Simulação de Redes WiMAX no Simulador NS-2, Anais do XXVIII Congresso da SBC, Julho 2008.
- [5] WiMAX Forum, Disponível em: <http://www.wimaxforum.org/>, Acessado em 2009.
- [6] WiMAX Forum, IEEE 802.16a Standard and WiMAX Igniting Broadband Wireless Access, White Paper, pp. 1-7, 2003.
- [7] DHRONA, P., A Performance Study of Uplink Scheduling Algorithms in Point to Multipoint WiMAX Networks, Master's Thesis, Queen's University, Kingston, Canada, Dezembro 2007.
- [8] CHOU, J. et all, MAC and PHY MIB for WirelessMAN and WirelessHUMAN BS and SS, IEEE S802. 16mgt-04, vol. 4, Maio 2004.
- [9] TALWALKAR, R., Analysis of Quality of Service (QoS) in WiMAX Networks, Master's Thesis, Florida Atlantic University, Boca Raton, Florida, Maio 2008.
- [10] ŞEKERCİOĞLU, Y. et all, A Survey of MAC Based QoS Implementations for WiMAX Networks, Elsevier, Computer Networks, vol. 53, no. 14, pp. 2517-2536, 2009.

- [11] MALINI, R., Uma Arquitetura de QoS e Modelo de Simulação para o Padrão IEEE 802.16d como Suporte a Serviços Multimídia distribuídos, Dissertação de Mestrado em Informática, Universidade Federal do Espírito Santo, Vitória, Maio 2007.
- [12] RODRIGUES, C., Escalonamento de Tráfego em Redes Wimax no Modo PMP, Dissertação de Mestrado, UnB, Universidade de Brasília, Faculdade de Tecnologia, Departamento de Engenharia Elétrica, Julho 2009.
- [13] PRASAD, R. & VELEZ, F. J., WiMAX Networks Techno-Economic Vision and Challenges, Springer, 2010.
- [14] LIMA, L. et al, WiMAX: Padrão IEEE 802.16 para Banda Larga Sem Fio, Pontifícia Universidade Católica do Rio de Janeiro, Setembro 2004.
- [15] XIAO, Y. & PAN, Y., Emerging Wireless LANs, Wireless PANs, and Wireless MANs: IEEE 802.11, IEEE 802.15, IEEE 802.16 Wireless Standard Family, ed. John Wiley & Sons, 2009.
- [16] CICCONETTI, C. et al, Quality of Service Support in IEEE 802.16 Networks, IEEE Network, vol. 20, issue 2, pp. 50-55, Abril 2006.
- [17] FIGUEIREDO, F. & PEREIRA, L., Tecnologia WIMAX: Uma Visão Geral, Cad. CPqD Tecnologia, Campinas, vol. 4, no. 2, pp. 7-26, 2008.
- [18] MA, M. et al, Current Technology Developments of WiMAX Systems, Springer Science + Business Media B.V., 2009.
- [19] GIDLUND, M. & WANG, G., Uplink Scheduling Algorithms for QoS Support in Broadband Wireless Access Networks, Journal of Communications, vol. 4, no. 2, pp. 133-142, 2009.
- [20] NUAYMI, L., WiMAX: Technology for Broadband Wireless Access, ed. John Wiley & Sons, 2007.

- [21] SOARES, C., Proposta de Algoritmo de Controle de Admissão de Conexões Baseado em Threshold para as Redes IEEE 802.16, Dissertação de Mestrado, Universidade Federal de Uberlândia, Maio 2009.
- [22] PARK, K., QoS in Packet Networks, Springer, 2005.
- [23] KUROSE, J. & ROSS, K., Redes de Computadores e a Internet: Uma Nova Abordagem Top-down, Addison Wesley, 3ª. Ed., São Paulo 2006.
- [24] ZHANG, Y. et all, WIMAX Network Planning and Optimization, ed. Auerbach Publications, 2008.
- [25] BELGHITH, A. & NUAYMI, L., Design and Implementation of a QoS-included WiMAX Module for NS-2 Simulator, First International Conference on Simulation Tools and Techniques for Communications, Networks and Systems, SimuTools 2008, Marseille, France, pp. 1-8, Março 2008.
- [26] EKLUND, C. et all, IEEE Standard 802.16: A Technical Overview of the WirelessMAN™ Air Interface for Broadband Wireless Access, IEEE Communications Magazine, vol. 40, no. 6, pp. 98-107, Julho 2002.
- [27] OROPEZA, A., Análise e Simulação do Sistema de Comunicações WIMAX (IEEE 802.16 - 2004) com ADS (Advanced Design System), Dissertação de Mestrado, Escola Politécnica da Universidade de São Paulo, 2008.
- [28] SO-IN, C. et all, A Deficit Round Robin with Fragmentation Scheduler for IEEE 802.16e Mobile WiMAX, IEEE SARNOFF '09, Sarnoff Symposium, pp. 1-7, Maio 2009.
- [29] SEMERIA, C., Supporting Differentiated Service Classes: Queue Scheduling Disciplines, Juniper Networks Inc., Part Number:: 200020-001 12/01, USA, 2001.
- [30] ALECRIM, P., Uma Proposta para a Provisão de QoS em Redes Backbone UMTS, Dissertação de Mestrado, Universidade Federal de Uberlândia, Abril 2006.

- [31] RAVICHANDIRAN, C., Analysis and Modification of Scheduling Algorithm for IEEE 802.16e (Mobile WiMAX), 2010 Second International Conference on Communication Software and Networks, pp. 434-438, 2010.
- [32] ALI, N. A. et al, A Performance Study of Uplink Scheduling Algorithms in Point-to-Multipoint WiMAX Networks, Elsevier, Computer Communications, vol. 32, pp. 511-521, 2009.
- [33] KATEVENIS, M. et al, Weighted Round-Robin Cell Multiplexing in a General-Purpose ATM Switch Chip, IEEE Journal on Selected Areas in Communications, vol. 9, pp. 1265-1279, Outubro 1991.
- [34] SUBRAMANYAM, Y. & VENKATESWARLU, Y., WiMAX Base Station Scheduling Algorithms, Tata Consultancy Services, From site : <http://whitepapers.techrepublic.com.com/abstract.aspx?docid=1183797>, Julho 2009.
- [35] ROESLER, V., Controle de Congestionamento em Redes de Computadores, Relatório Técnico: RT 05/2003, Unisinos, Centro de Ciências Exatas e Tecnológicas, PRAV: Pesquisa em Redes de Alta Velocidade, 2003.
- [36] SHREEDHAR, M. & VARGHESE, G., Efficient Fair Queuing Using Deficit Round-Robin, IEEE/ACM Transactions on Networking, vol. 4, no. 3, pp. 375-385, Junho 1996.
- [37] SO-IN, C. et al, Scheduling in IEEE 802.16e Mobile WiMAX Networks: Key Issues and a Survey, IEEE Journal on Selected Areas in Communications, vol. 27, no. 2, pp. 156-171, 2009.
- [38] FATTAH, H. & LEUNG, C., An Overview of Scheduling Algorithms in Wireless Multimedia Networks, IEEE Wireless Communications, vol. 9, no. 5, pp. 76-83, 2002.
- [39] LERA, A. et al, Channel-Aware Scheduling for QoS and Fairness Provisioning in IEEE 802.16/WiMAX Broadband Wireless Access Systems, IEEE Network, vol. 21, issue 5, pp. 34-41, Setembro 2007.

- [40] FREITAG, J. & DA FONSECA, Algoritmos para Controle de Admissão e Escalonamento do Tráfego Uplink em Redes IEEE 802.16, Anais do VIII WPERFORMANCE, 2009.
- [41] YEO, K. et al, A Packet Scheduler with Forward Equilibrium and Adaptive Fair Selection Method in Wireless Multimedia Traffic Environment, Proceedings of ICCT2003, vol.2, pp. 1328-1331, Julho 2003.
- [42] MSADAA I. et al, Scheduling and CAC in IEEE 802.16 Fixed BWNs: A Comprehensive Survey and Taxonomy, IEEE Communications Surveys & Tutorials, Fourth Quarter 2010, vol. 12, pp. 459-487, 2010.
- [43] FREITAG, J., Mecanismos para a Provisão de Qualidade de Serviço em Redes IEEE 802.16, Tese de Doutorado, Unicamp, Janeiro 2010.
- [44] BHANDARI, B. et al, Uplink Performance of the IEEE 802.16 Medium Access Control (MAC) Layer Protocol, ICPWC 2005, Personal Wireless Communications, pp. 5-8, 2005.
- [45] SETTEMBRE, M. et al, Performance Analysis of an Efficient Packet-Based IEEE 802.16 MAC Supporting Adaptive Modulation and Coding, Proceedings of International Symposium on Computer Networks, pp.11-16, Junho 2006.
- [46] SAYENKO, A. et al, Ensuring the QoS Requirements in 802.16 Scheduling, Proceedings of the 9th ACM International Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems, pp. 108-117, Outubro 2006.
- [47] WONGTHAVARAWAT, K. & GANZ, A., Packet Scheduling for QoS Support in IEEE 802.16 Broadband Wireless Access Systems, International Journal of Communications Systems, vol. 16, no. 1, pp. 81-96, Fevereiro 2003.
- [48] BAI, X. et al, Robust QoS Control for Single Carrier PMP Mode IEEE 802.16 Systems, IEEE Transactions on Mobile Computing, vol. 7, pp. 416-429, 2007.

- [49] SUN, J. et al, Quality of Service Scheduling for 802.16 Broadband Wireless Access Systems, IEEE 63rd Vehicular Technology Conference, vol. 3, pp. 1221-1225, Maio 2006.
- [50] BALL, C. et al, Performance Analysis of Temporary Removal Scheduling Applied to Mobile WiMax Scenarios in Tight Frequency Reuse, The 16th Annual IEEE International Symposium on Personal Indoor and Mobile Radio Communications, PIMRC'2005, Berlin, pp. 11–14, Setembro 2005.
- [51] TSAI, T. et al, CAC and Packet scheduling Using Token Bucket for IEEE 802.16 Networks, Journal of Communications, vol. 1, no. 2, pp. 30, Maio 2006.
- [52] GIDLUND, M. & WANG, G., Design of Uplink Schedulers for Broadband Wireless Access Networks, WAMICON '09, IEEE 10th Annual, Wireless and Microwave Technology Conference, pp. 1-2, 2009.
- [53] CHIANG, C. et al, Adaptive Downlink/Uplink Bandwidth Allocation in IEEE 802.16 (WiMAX) Wireless Networks: A Cross-Layer Approach, GLOBECOM 07, Global Telecommunications Conference, Washington, USA, pp. 4775-4779, Novembro 2007.
- [54] LIU, Q. et al, Cross-Layer Scheduler Design with QoS Support for Wireless Access Networks, Proceedings of the 2nd Int'l Conf. on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine '05), pp. 8-21, 2005.
- [55] NASCIMENTO, A. & RODRIGUES, J., A New Cross-Layer Based Dynamic Resource Allocator for IEEE 802.16e Networks, IWCLD '09, Second International Workshop on Palma de Mallorca, Cross Layer Design, pp. 1-6, Junho 2009.
- [56] PITIC, R. et al, Performance Evaluation of Utility-Based Scheduling Schemes with QoS Guarantees in IEEE 802.16/WiMAX Systems, Wireless Communications and Mobile Computing, Published online in Wiley InterScience, (www.interscience.wiley.com) DOI: 10.1002/wcm.802, 2009.

- [57] MUKUL, R. et al, An Adaptive Bandwidth Request Mechanism for QoS Enhancement in WiMax Real Time Communication, Wireless and Optical Communications Networks, 2006 IFIP International Conference on, Bangalore, India, pp. 1-5, April 2006.
- [58] MAHESHWARI, S. et al, An Efficient QoS Scheduling Architecture for IEEE 802.16 Wireless MANs, Master's Thesis, Indian Institute of Technology, Mumbai, Indian, 2005.
- [59] CHEN, J. et al, The Design and Implementation of WiMAX Module for ns-2 Simulator, ACM, Proceeding from the 2006 Workshop on ns-2: The IP Network Simulator, pp. 1-9, 2006.
- [60] DUTRA, L., Solução de Qualidade de Serviço Fim-a-Fim em Redes Metropolitanas Heterogêneas, Dissertação de Mestrado, UnB, Universidade de Brasília, Faculdade de Tecnologia, Departamento de Engenharia Elétrica, Fevereiro de 2008.
- [61] ARCIA, A. et al, Improving Uplink Bandwidth Management in 802.16 Networks, IEEE 34th Conference on Local Computer Networks, pp. 233-236, 2009.
- [62] CHEN, J. et al, A Service Flow Management Strategy for IEEE 802.16 Broadband Wireless Access Systems in TDD Mode, In Proceedings of the IEEE International Conference on Communications, pp. 3422-3426, 2005.
- [63] ZHANG, G. et al, Quality of Service Scheduling Based on GPSS in IEEE 802.16 WiMAX Networks, 4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008 (WiCOM 08), pp. 1-4, Outubro 2008.
- [64] LIANG, J. et al, Performance Evaluation of an Integrated and Efficient Uplink Scheduler for WiMAX Network, IEEE International Conference on Communications Technology and Applications (ICCTA '09), pp. 326-330, 2009.
- [65] CHU, G., WANG, D., MEI, S., A QoS Architecture for the MAC Protocol of IEEE 802.16 BWA System, IEEE Communications, Circuits and Systems and West Sino Expositions, pp. 435-439, 2002.

- [66] TARCHI, D. et al, Quality of Service Management in IEEE 802.16 Wireless Metropolitan Area Networks, Proceedings of the IEEE International Conference on Communications, pp. 1789-1794, 2006.
- [67] MAHMOOD, R. et al, A Novel Parameterized QoS Based Uplink and Downlink Scheduler for Bandwidth/Data Management over IEEE 802.16d Network, International Journal of Recent Trends in Engineering, vol. 2, no. 1, pp. 42-46, Novembro 2009.
- [68] SAYENKO, A. et al, Scheduling Solution for the IEEE 802.16 Base Station, Elsevier, Proceedings of the Computer Networks, vol. 52, n. 1, pp. 96-115, 2008.
- [69] THE NETWORK SIMULATOR – NS-2, Disponível em: <http://www.isi.edu/nsnam/ns/>, Acessado em: Agosto de 2008.
- [70] SHRIVASTAV, N., A Network Simulator Model of the DOCSIS Protocol and a Solution to the Bandwidth-Hog Problem in the Cable Networks, Master's Thesis, Faculty of North Carolina State University, 2003.
- [71] BRADY, P., A Model for Generating On-Off Speech Patterns in Two-Way Conversations, Bell System Technical Journal, vol. 48, pp. 2445-2472, 1969.
- [72] SEELING, P. et al, Network Performance Evaluation Using Frame Size and Quality Traces of Single-Layer and Two-Layer Video: A Tutorial. IEEE Communications Surveys and Tutorials, vol. 6, no. 2, pp. 58-78, 2004.
- [73] BARFORD, P. et al, Changes in Web Client Access Patterns: Characteristics and Caching Implications, Technical Report 1998-023, Boston University, 1998.
- [74] FERREIRA, F. A. & GUARDIEIRO, P. R., Uma Proposta de Algoritmo de Escalonamento Baseado na Disciplina *Priority Queuing* (PQ) para Redes IEEE 802.16, 9th International Information and Telecommunication Technologies Symposium (I2TS), Rio de Janeiro – RJ, Dezembro 2010.

- [75] FERREIRA, F. A. & GUARDIEIRO, P. R., Uma Proposta de Escalonador Baseado na Disciplina *Priority Queuing* (PQ) para Redes IEEE 802.16, VIII Conferência de Estudos em Engenharia Elétrica (CEEL), Uberlândia – MG, Outubro 2010.