

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



PROPOSTA DE UM ESQUEMA PARA A
PROVISÃO DE QoS NO PADRÃO IEEE 802.16

Eduardo Castilho Rosa

Eduardo Castilho Rosa

**PROPOSTA DE UM ESQUEMA PARA A PROVISÃO DE QoS NO
PADRÃO IEEE 802.16**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal de Uberlândia, como parte dos requisitos para obtenção do grau de Mestre em Ciências, aprovada em 7 de julho de 2011 pela banca examinadora:

Paulo Roberto Guardieiro, Dr. - Orientador (UFU)

Anilton Salles Garcia, Dr. (UFES)

Luis Fernando Faina, Dr. (UFU)

Eduardo Castilho Rosa

**PROPOSTA DE UM ESQUEMA PARA A PROVISÃO DE QoS NO
PADRÃO IEEE 802.16**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal de Uberlândia, como parte dos requisitos para obtenção do grau de Mestre em Ciências.

Prof. Paulo Roberto Guardieiro, Dr.
Orientador

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

Dedico esse trabalho aos meus pais Alice Castilho e José Marcos.

Agradecimentos

Agradeço primeiramente à Deus por ter me sustentado com vida e saúde até a conclusão desse trabalho.

Ao meu orientador Prof. Paulo Roberto Guardieiro pela paciência e pelo valioso aprendizado que me proporcionou ao longo desse trabalho.

À minha mãe Alice Castilho e a meu pai José Marcos, sem o apoio dos quais não seria possível a realização desse trabalho.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo incentivo financeiro na forma de bolsa de mestrado, para que a pesquisa pudesse ser viabilizada.

À Faculdade de Engenharia Elétrica da Universidade Federal de Uberlândia pelos recursos oferecidos para a execução desse trabalho.

Aos meus amigos de laboratório, em especial, ao Antônio, Flávio e Éderson, pela ajuda e ensinamentos.

Enfim, à todos aqueles que contribuíram direta ou indiretamente para a realização desse trabalho.

“E conhecereis a verdade, e a verdade vos libertará”

João 8:32

Resumo

Rosa, E. C., *Proposta de Um Esquema Para a Provisão de QoS no Padrão IEEE 802.16*, UFU, Uberlândia, Brasil, 2011, 114p.

O padrão IEEE 802.16 é designado para prover conectividade em banda larga sem fio para usuários fixos e móveis em uma ampla área de cobertura com altas taxas de transferência de dados. A principal característica incorporada por esse padrão é a provisão de qualidade de serviço (QoS) aos usuários finais, garantida por meio de algoritmos para escalonamento de pacotes tanto na estação base como nas estações de assinantes e também políticas para controle de admissão de conexões (CAC). Pelo fato do padrão não especificar políticas para a implementação desses mecanismos, propõe-se nesse trabalho um novo algoritmo de escalonamento *uplink* na estação base que é capaz de atender as principais classes de serviços definidas pelo padrão e também uma política de CAC dinâmica baseada em predição de atraso. O mecanismo de escalonamento proposto atende as filas de requisições de forma cíclica e baseia-se em um esquema de reserva estática de largura de banda para separar os fluxos de tempo real, não tempo real com o requisito de taxa mínima e não tempo real sem requisitos de vazão (tráfego de melhor esforço). Já a política de CAC foi proposta para a classe de serviço de tempo real com taxa variável e baseia-se na predição do atraso médio calculado de acordo com o estado de ocupação dos *buffers* nas estações assinantes. Avaliou-se, por meio de modelagem e simulação, vários cenários de redes com o intuito de testar a eficiência do mecanismo de escalonamento e CAC. Pelos resultados obtidos, verificou-se que o mecanismo proposto é capaz de atender os requisitos de atraso máximo limitado para fluxos de tempo real e vazão mínima para fluxos de não tempo real. Além disso, constatou-se que o mecanismo proposto é capaz de evitar o *bandwidth starvation* dos fluxos de baixa prioridade.

Palavras-chave: WiMAX, IEEE 802.16, QoS, Escalonamento, CAC.

Abstract

Rosa, E. C., *Proposal of Framework for QoS Provisioning in IEEE 802.16 Networks*, UFU, Uberlândia, Brazil, 2011, 114p.

The IEEE 802.16 standard is designed to provide wireless broadband connectivity for both fixed and mobile users in a wide coverage area with high rates of data transfer. The main feature incorporated by the IEEE 802.16 standard is the quality of service (QoS) provisioning by means of packets scheduling, traffic policing and connection admission control (CAC). Since the IEEE 802.16 standard does not define the policies for admission control and packet scheduling, in this paper we propose a QoS provisioning mechanism which consists of a novel scheduling algorithm for uplink traffic and a dynamic CAC scheme for IEEE 802.16 networks. The proposed scheduling mechanism serves the requests queues in a cyclical mode and is based on a static bandwidth reservation mechanism in order to separate the real-time and non real time flows. The CAC policy is proposed for the real-time service class with variable rate and is based on a delay prediction scheme which uses the buffer size information in SSs. Through modeling and simulation, we evaluated various network scenarios in order to test the efficiency of CAC and scheduling mechanism. From the results, it was found that the proposed mechanism is able to meet the maximum delay requirements of the maximum delay bound for real time flows and minimum throughput for streams of non-real time. Moreover, it was found that the proposed mechanism is able to avoid the bandwidth starvation of low priority flows.

Index terms: WiMAX, IEEE 802.16, QoS, Scheduling, CAC.

Sumário

Lista de Figuras	xi
Lista de Tabelas	xiii
Lista de Abreviaturas e Siglas	xiv
1 Introdução	20
2 Padrão IEEE 802.16	25
2.1 Introdução	25
2.2 Histórico de Evolução	26
2.3 Arquitetura do Padrão IEEE 802.16	28
2.3.1 Modo de Operação	30
2.3.1.1 Ponto a multiponto	30
2.3.1.2 Malha	31
2.3.2 Modelo de Referência	32
2.3.2.1 Camada MAC	33
2.3.2.1.1 Subcamada de Convergência Específica	33
2.3.2.1.2 Subcamada da Parte Comum	34
2.3.2.1.3 Subcamada de Segurança	36
2.3.2.2 Camada Física	38
2.3.2.2.1 Técnicas de Duplexação	38
2.3.2.2.2 Estrutura do Quadro TDD	38
2.3.2.2.3 Especificação WirelessMAN-SC PHY	40
2.3.2.2.4 Especificação WirelessMAN-SCa PHY	41
2.3.2.2.5 Especificação WirelessMAN-OFDM PHY	41

2.3.2.2.6	Especificação WirelessMAN-OFDMA PHY	42
2.4	Considerações Finais	43
3	Provisão de QoS em Redes IEEE 802.16	45
3.1	Introdução	45
3.2	QoS no Padrão IEEE 802.16	46
3.3	Mecanismos para a Provisão de QoS no Padrão IEEE 802.16	50
3.3.1	Escalonamento de Pacotes	51
3.3.1.1	Estratégias de Escalonamento Homogêneo	52
3.3.1.2	Estratégias de Escalonamento Híbridas	57
3.3.1.3	Escalonamento com Abordagem <i>Cross-Layer</i>	58
3.3.2	Controle de Admissão de Conexões	59
3.3.2.1	Mecanismos de CAC Pró-ativos	62
3.3.2.2	Mecanismos de CAC Reativos	65
3.4	Considerações Finais	67
4	Proposta de Um Esquema Para a Provisão de QoS no Padrão IEEE 802.16	69
4.1	Introdução	69
4.2	Descrição do Problema	70
4.3	Solução Proposta	70
4.3.1	Mecanismo de Escalonamento <i>Uplink</i> Proposto	71
4.3.2	Mecanismo de CAC Dinâmico Proposto	73
4.4	Trabalhos Relacionados	77
4.5	Considerações Finais	79
5	Avaliação dos Mecanismos de Escalonamento e CAC Propostos	81
5.1	Introdução	81
5.2	Modelagem e Simulação	81
5.2.1	Ferramentas de Simulação	81
5.2.2	Ambiente e Parâmetros de Simulação	83
5.3	Resultados Experimentais	83
5.3.1	Avaliação do Escalonador Proposto para rtPS	84

5.3.2	Avaliação do Escalonador Proposto Considerando as Classes UGS e rtPS	87
5.3.3	Avaliação do Escalonador Proposto para Fluxos nrtPS e BE	89
5.3.4	Avaliação do Escalonador Proposto em Relação ao Mecanismo de Reserva	91
5.3.5	Avaliação do Mecanismo de CAC Proposto para Fluxos rtPS	93
5.3.6	Avaliação do CAC para Fluxos Não Tempo Real	94
5.3.7	Validação do Mecanismo de CAC Proposto	96
5.4	Considerações Finais	98
6	Conclusões Gerais	99
	Referências Bibliográficas	103
	Apêndice A	111

Lista de Figuras

2.1	Infraestrutura de operação no padrão IEEE 802.16.	28
2.2	Modo de operação PMP.	30
2.3	Modo de operação <i>mesh</i>	31
2.4	Modelo de referência do padrão IEEE 802.16 (adaptado de [1]).	32
2.5	Subcamadas no padrão IEEE 802.16 (adaptado de [1]).	33
2.6	Formato genérico da MAC PDU no padrão IEEE 802.16 [2].	35
2.7	Funcionalidades da subcamada CPS (adaptado de [1]).	36
2.8	Modo de duplexação TDD e FDD (adaptado de [3]).	39
2.9	Estrutura do quadro TDD.	40
2.10	Eficiência espectral da técnica OFMD [4].	42
2.11	Comparação entre a técnica OFDM e OFDMA (adaptado de [4]).	43
3.1	Estabelecimento de um fluxo de serviço ou conexão.	48
3.2	Arquitetura base do padrão IEEE 802.16 [5].	51
3.3	<i>Framework</i> para o escalonamento na BS e SS [6].	52
3.4	Hierarquia dos algoritmos de escalonamento para o padrão IEEE 802.16 [6].	52
3.5	Comparação entre o resultado de um ciclo de escalonamento utilizando os mecanismos RR e WRR (adaptado de [6]).	54
3.6	Classificação dos mecanismos de CAC.	62
3.7	Mecanismo de CAC baseado em otimização.	64
3.8	Funcionamento de um esquema de CAC reativo.	65
4.1	(a) Arquitetura original do padrão IEEE 802.16 e (b) Mecanismos propostos.	71
4.2	Mecanismo de reserva de largura de banda proposto para o escalonamento.	72
5.1	Aplicativo desenvolvido para analisar o trace gerado pelo NS-2.	82
5.2	Cenário de rede considerado na simulação.	83

5.3	Desempenho do escalonador rtPS proposto em termos de atraso médio. . .	86
5.4	Desempenho do escalonador rtPS proposto em termos de vazão média. . .	86
5.5	Desempenho do escalonador proposto para fluxos tempo real em termos de atraso médio.	88
5.6	Desempenho do escalonador proposto em termos de vazão média para fluxos rtPS (A) e fluxos UGS (B).	88
5.7	Desempenho do escalonador proposto para fluxos rtPS em termos de atraso (A) e vazão (B), considerando estações nrtPS e BE presentes na rede. . . .	90
5.8	Desempenho do escalonador proposto em termos de vazão média para fluxos não tempo real.	90
5.9	Análise da variação da reserva W para fluxos rtPS em termos de atraso médio.	92
5.10	Análise da variação da reserva W para fluxos nrtPS em termos de vazão média	92
5.11	Comparação entre o atraso médio predito na rede em relação ao atraso medido.	94
5.12	Desempenho do CAC preditivo proposto para fluxos rtPS.	94
5.13	Desempenho do CAC preditivo para fluxos não tempo real).	95
5.14	Desempenho do CAC preditivo considerando o atraso médio para fluxos UGS.	96
5.15	Desempenho do CAC preditivo em relação ao mecanismo de CAC baseado em taxa mínima	97

Lista de Tabelas

2.1	Evolução do padrão IEEE 802.16	29
2.2	Características das especificações PHY [7].	43
3.1	Parâmetros de QoS no fluxo de serviço [8].	47
3.2	Comparação entre as disciplinas de escalonamento homogêneas.	56
3.3	Comparação entre as estratégias de escalonamento híbridas.	58
3.4	Comparação entre as estratégias de escalonamento que utilizam a abordagem <i>Cross-Layer</i>	60
3.5	Critérios para a classificação dos mecanismos de CAC [9].	61
3.6	Comparação entre as propostas de CAC pró-ativos.	66
3.7	Comparação entre as propostas de CAC's reativas.	67
4.1	Características dos trabalhos relacionados ao mecanismo proposto.	80
5.1	Parâmetros de simulação referentes às camadas MAC e física.	84
5.2	Eventos ocorridos na simulação em função do tempo.	87
5.3	Eventos ocorridos na simulação em função do tempo.	88
5.4	Eventos ocorridos na simulação em função do tempo.	91
5.5	Eventos ocorridos na simulação em função do tempo.	93
5.6	Eventos ocorridos na simulação em função do tempo.	94
5.7	Eventos ocorridos na simulação em função do tempo.	96
5.8	Eventos ocorridos na simulação em função do tempo.	97

Lista de Abreviaturas e Siglas

3GPP	<i>3th Generation Partnership Project</i>
4G	<i>4th Generation</i>
AES	<i>Advanced Encryption Standard</i>
ATM	<i>Asynchronous Transfer Mode</i>
AWK	<i>Aho, Weinberger and Kernighan</i>
BE	<i>Best Effort</i>
BPSK	<i>Binary Phase Shift Keying</i>
BRM	<i>Bandwidth Reservation Mechanism</i>
BS	<i>Base Station</i>
BWA	<i>Bandwidth Wireless Access</i>
CAC	<i>Connection Admission Control</i>
CBR	<i>Constant Bit Rate</i>
CI	<i>CRC Indicator</i>
CID	<i>Connection Identifier</i>
CMAC	<i>Cipher-based Message Authentication Code</i>
CPS	<i>Common Part Sublayer</i>
CRC	<i>Cyclic Reduncancy Check</i>
CS	<i>Service-Specific Convergence</i>

DAMA	<i>Demand Assigned Multiple Access</i>
DCD	<i>Downlink Channel Descriptor</i>
DES	<i>Data Encryption Standard</i>
DiffServ	<i>Differentiated Services</i>
DL-MAP	<i>Downlink Map</i>
DRR	<i>Deficit Round Robin</i>
DSA-ACK	<i>Dynamic Service Addition Acknowledgment</i>
DSA-REQ	<i>Dynamic Service Addition Request</i>
DSA-RSP	<i>Dynamic Service Addition Response</i>
DSL	<i>Digital Subscriber Line</i>
DVB	<i>Digital Video Broadcasting</i>
EC	<i>Encryption Control</i>
EDD	<i>Earliest Due Date</i>
EDF	<i>Earliest Deadline First</i>
EKS	<i>Encryption Key Sequence</i>
ertPS	<i>Extended Real-Time Polling Service</i>
FDD	<i>Frequency Division Duplexing</i>
FDM	<i>Frequency Division Multiplexing</i>
FEC	<i>Forward Error Correction</i>
FIFO	<i>First in first out</i>
FTP	<i>File Transfer Protocol</i>
GPC	<i>Grant per Connection</i>
GPSS	<i>Grant per Subscriber Station</i>

HCS	<i>Header Check Sequence</i>
HMAC	<i>Hashed Message Authentication Code</i>
HT	<i>Header Type</i>
HTTP	<i>Hyper Text Transfer Protocol</i>
HUMAN	<i>High-speed Unlicensed MAN</i>
IEEE	<i>Institute of Electric and Electronic Engineers</i>
IMT	<i>International Mobile Telecommunications</i>
IntService	<i>Integrated Services</i>
IPv4	<i>Internet Protocol - 4</i>
IPv6	<i>Internet Protocol - 6</i>
ISO	<i>International Organization for Stadardization</i>
ISP	<i>Internet Service Provider</i>
ITU	<i>International Telecommunications Union</i>
ITU-R	<i>International Telecommunications Union - Radiocommunication Sector</i>
LEN-MSB	<i>Length of Most Significant Byte</i>
LOS	<i>Line of Sight</i>
LSB	<i>Least Significant Byte</i>
LTE	<i>Long Term Evolution</i>
MAC	<i>Media Access Control</i>
MAN	<i>Metropolitan Access Network</i>
MCS	<i>Modulation Coding Scheme</i>
MIB	<i>Management Information Base</i>
MIMO	<i>Multiple Input Multiple Output</i>

MPEG	<i>Moving Pictures Experts Group</i>
MRTR	<i>Minimum Reserved Traffic Rate</i>
mSIR	<i>Maximum Signal-to-interference</i>
NIST	<i>National Institute of Standards and Technology</i>
NLOS	<i>No Line of Sight</i>
nrtPS	<i>Non Real-Time Polling Service</i>
NS-2	<i>Network Simulator - 2</i>
O-DDR	<i>Opportunistic Deficit Round Robin</i>
OFDM	<i>Orthogonal Frequency Division Multiplexing</i>
OFDMA	<i>Orthogonal Frequency-Division Multiple Access</i>
OSI	<i>Open Systems Interconnection</i>
PDU	<i>Packet Data Unit</i>
PHS	<i>Packet Header Supression</i>
PHY	<i>Physical Layer</i>
PMP	<i>Point-to-multipoint</i>
PQ	<i>Priority Queuing</i>
PS	<i>Privacy Sub-layer</i>
QAM	<i>Quadrature Amplitude Modulation</i>
QoS	<i>Quality of Service</i>
QPSK	<i>Quadrature Phase Shift Keying</i>
RR	<i>Round Robin</i>
RSA	<i>Rivest Shamir Adleman</i>
RTG	<i>Receive Transition Gap</i>

rtPS	<i>Real-Time Polling Service</i>
SAP	<i>Service Access Point</i>
SC	<i>Single Carrier</i>
SCa	<i>Single Carrier Adaptive</i>
SDU	<i>Service Data Unit</i>
SFID	<i>Service Flow Identifier</i>
SINR	<i>Signal-to-interference-plus-noise Ratio</i>
SIR	<i>Signal-to-interference</i>
SNR	<i>Signal to Noise Ratio</i>
SQL	<i>Structured Query Language</i>
SS	<i>Subscriber Station</i>
TDD	<i>Time Division Duplexing</i>
TDM	<i>Time Division Multiplexing</i>
TDMA	<i>Time Division Multiple Access</i>
TR	Taxa Reservada
TRS	<i>Temporary Removal Scheduling</i>
TTG	<i>Transmit Transition Gap</i>
UCD	<i>Uplink Channel Descriptor</i>
UGS	<i>Unsolicited Grant Service</i>
UL-MAP	<i>Uplink Map</i>
VCI	<i>Virtual Channel Identifier</i>
VoIP	<i>Voice Over Internet Protocol</i>
VPI	<i>Virtual Path Identifier</i>

WFQ	<i>Weighted Fair Queuing</i>
WiFi	<i>Wireless Fidelity</i>
WiMAX	<i>Worldwide Interoperability for Microwave Access</i>
WiMAX-2	<i>Worldwide Interoperability for Microwave Access Advanced</i>
WLAN	<i>Wireless Local Area Network</i>
WRR	<i>Weighted Round Robin</i>

Capítulo 1

Introdução

Nos últimos anos, a crescente demanda por acesso banda larga sem fio vêm possibilitando o surgimento de várias tecnologias de redes capazes de atender as expectativas dos usuários tanto em termos de taxas de transmissão de dados elevadas como também em relação ao suporte à mobilidade. Desde o aparecimento das primeiras tecnologias para a transmissão de dados sem fio, através de satélites de comunicações, por exemplo, a tecnologia *wireless*, como também é chamada, vem sendo desenvolvida para fazer parte não somente das comunicações de longas distâncias mas também para cobrir regiões locais e metropolitanas e fornecer acesso à Internet.

Um elemento chave nesse processo e que pode ser considerado a base para o desenvolvimento das redes de acesso sem fio atuais é o surgimento da tecnologia WiFi (*Wireless Fidelity*). Também conhecido como padrão IEEE 802.11 [10], essa tecnologia de WLAN (*Wireless Local Area Network*) se tornou muito popular ao longo dos últimos anos e amplamente difundida em razão dos baixos custos de implantação em termos de infraestrutura. Entretanto, apesar desse padrão vir sofrendo várias adaptações e melhoramentos (IEEE 802.11a, IEEE 802.11b, IEEE 802.11g, IEEE 802.11n, etc), ele não foi criado com o intuito de cobrir regiões metropolitanas e de longas distâncias nem de oferecer suporte à qualidade de serviço (*QoS - Quality of Service*), requisitos fundamentais para atender as reais necessidades e expectativas dos usuários modernos.

Em razão disso, nos últimos anos um considerável investimento por parte do mercado das telecomunicações sem fio, aliado à necessidade em se ter padrões abertos para prover interoperabilidade de equipamentos, institutos como o IEEE (*Institute of Electric and Electronic Engineers*) e o 3GPP (*3th Generation Partnership Project*), por exemplo, em

seus respectivos padrões IEEE 802.16 e LTE (*Long Term Evolution*), vêm desenvolvendo especificações para o acesso banda larga sem fio que cobrem regiões metropolitanas e são capazes de oferecer suporte nativo à QoS com elevadas taxas de transmissão de dados.

Nesse processo, o padrão IEEE 802.16, também conhecido como WiMAX (*Worldwide Interoperability for Microwave Access*), tem recebido uma atenção especial tanto da comunidade científica internacional como também das principais empresas de tecnologias sem fio espalhadas pelo mundo. Por meio do consórcio WiMAX Forum, firmado por volta de 2001 com o objetivo de certificar os produtos baseados no padrão IEEE 802.16, grande parte dos investimentos nessa tecnologia vem sendo empregados na implantação de redes de pequeno porte para realização de testes operacionais e no desenvolvimento de equipamentos certificados e prontos para serem utilizados ao redor do mundo. Dessa maneira, o padrão IEEE 802.16 vem se despontando como uma tecnologia promissora capaz de implementar os futuros sistemas de quarta geração (4G).

Basicamente, o padrão IEEE 802.16 define a especificação para a interface aérea e o protocolo para o acesso ao meio em redes metropolitanas sem fio (MAN - *Metropolitan Access Network*), sendo capaz de fornecer altas taxas de transmissão para o acesso comercial e residencial à Internet. A estrutura em camadas do padrão IEEE 802.16 permite a interoperabilidade do WiMAX com diferentes tecnologias como, por exemplo, as redes baseadas na arquitetura TCP/IP e redes ATM (*Asynchronous Transfer Mode*). Além disso, o padrão permite um eficiente e flexível modo de transferência de dados através da especificação de múltiplas camadas físicas como as baseadas na multiplexação por divisão de frequências em uma única portadora e multiplexação por divisão de frequências com múltiplas portadoras ortogonais.

Na camada física o padrão IEEE 802.16 permite as técnicas de duplexação do meio físico baseadas na divisão do tempo e na divisão da frequência. Versões mais recentes do padrão permitem ainda a transmissão de dados utilizando múltiplas camadas físicas que proporcionam uma alta eficiência espectral.

Já na camada de controle de acesso ao meio, o padrão fornece tratamento diferenciado ao tráfego gerado pelas aplicações, por meio da especificação de fluxos de serviços específicos como voz, vídeo e dados. Esses fluxos são destinados a atender os principais requisitos de qualidade de serviço das aplicações tais como atraso máximo limitado, vazão mínima sustentável, taxa de perdas mínima e variação de atraso máxima limitada.

Como complemento para a provisão de qualidade de serviço, o padrão IEEE 802.16 define também uma arquitetura base que consiste em mecanismos para escalonamento de pacotes e controle de admissão de conexões (CAC - *Connection Admission Control*). Entretanto, apesar da adoção desses mecanismos ser obrigatória para se ter qualidade de serviço, não são definidos políticas ou algoritmos específicos para escalonamento e CAC, ou seja, o padrão deixa em aberto a implementação dessas políticas, como forma dos fabricantes de equipamentos de redes diferenciarem seus produtos.

Fundamentado no fato do padrão IEEE 802.16 não definir mecanismos específicos para a provisão de qualidade de serviço, neste trabalho propõe-se um esquema para o fornecimento de qualidade de serviço às aplicações que consiste na implementação de um algoritmo para o escalonamento *uplink* na estação base e uma política para CAC. O mecanismo de escalonamento proposto é capaz de atender as principais classes de serviços definidas pelo padrão e baseia-se num esquema de reserva estática de largura de banda. Já a política de CAC é destinada a uma classe de serviço específica de tempo real e é desempenhada de maneira dinâmica com base na predição do atraso na rede, a qual é calculada com base no comprimento atual das filas nas estações assinantes.

O mecanismo de escalonamento proposto prioriza os fluxos tempo real em detrimento aos fluxos não tempo real no sentido de prover garantias estritas quanto ao atraso máximo limitado e vazão mínima sustentável. São especificadas quatro filas na estação base para armazenar os *grants* necessários ao envio dos dados e as requisições de largura de banda provenientes dos fluxos com taxa variável. A disciplina de atendimento às filas de armazenamento baseia-se na estratégia de enfileiramento prioritário, porém são realizadas preempções nas filas de mais alta prioridade no sentido de prover garantias de vazão para os fluxos de mais baixa prioridade.

O CAC proposto limita a quantidade de fluxos simultâneos na rede no sentido de controlar o atraso dos fluxos de tempo real. Esse controle é desempenhado através da estimativa do atraso médio global que uma potencial conexão sofrerá na rede se for admitida e utiliza informações sobre o estado de ocupação dos *buffers* nas estações de assinantes para realizar a predição.

Como principais contribuições deste trabalho ([61], [62], [63], [64], [65] e [66]) destacam-se um levantamento bibliográfico e análise comparativa do estado da arte no que se refere a mecanismos para a provisão de qualidade de serviço e a proposta de uma política para

escalonamento *uplink* e controle de admissão na estação base. Esses mecanismos estão de acordo com o padrão IEEE 802.16 e são capazes de prover qualidade de serviço às aplicações, conforme demonstrado nos estudos de avaliação de desempenho da proposta. Essas contribuições, segundo a pesquisa bibliográfica realizada, apresentam aspectos inovadores.

O restante desta dissertação está organizado como descrito a seguir.

No Capítulo 2 são apresentadas as principais características do padrão IEEE 802.16 no que diz respeito a arquitetura, modos de operação e modelo de referência. Inicialmente, apresenta-se um histórico sobre a evolução da família de padrões IEEE 802.16, desde o lançamento de sua primeira versão até os dias atuais. Em seguida, as arquiteturas ponto-multiponto e em malha são descritas de maneira sucinta, apresentando as vantagens, desvantagens e aplicabilidade das mesmas. A camada de controle de acesso ao meio é também abordada, apresentando-se a estrutura dos quadros de operação e formato da unidade de dados da MAC (*Media Access Control*). Na camada física são apresentadas as mais importantes especificações para a interface aérea existentes no padrão IEEE 802.16, como, por exemplo, as baseadas em uma única ou múltiplas portadoras.

O Capítulo 3 apresenta os principais aspectos referentes à provisão de qualidade de serviço nas redes WiMAX. Inicialmente são descritas as cinco classes de serviço especificadas para o padrão IEEE 802.16 bem como a sinalização necessária para o estabelecimento de novas conexões. Apresenta-se também o estado da arte em relação aos mecanismos de escalonamento de pacotes e políticas de CAC existentes na literatura. Os mecanismos de escalonamento são divididos em disciplinas homogêneas, híbridas e *cross-layer* e as políticas de CAC são abordadas segundo os esquemas pró-ativos e reativos.

No Capítulo 4 descreve-se o problema a ser abordado e apresenta-se a solução proposta a qual consiste em um algoritmo para escalonamento *uplink* na estação base e um mecanismo para controle de admissão de conexões com atraso preditivo. Além disso, são apresentados também alguns trabalhos relacionados à proposta que são considerados relevantes para a pesquisa, seguido por uma análise comparativa desses trabalhos em relação ao mecanismo proposto.

O Capítulo 5 traz os detalhes dos procedimentos utilizados na avaliação da proposta apresentada. As ferramentas para simulação e análise dos resultados são descritas juntamente com os principais parâmetros envolvidos nos experimentos. Em seguida, são apresentados e analisados os resultados obtidos por meio de modelagem e simulação, os

quais consideram vários cenários.

Finalmente, no Capítulo 6 são apresentadas as conclusões gerais relacionadas a este trabalho, bem como algumas sugestões para trabalhos futuros.

Capítulo 2

Padrão IEEE 802.16

2.1 Introdução

O padrão IEEE 802.16, impulsionado pelo surgimento e crescimento cada vez mais acentuado de aplicações com tráfegos heterogêneos (voz, video e dados), vem se tornando uma opção atrativa de acesso banda larga sem fio para a última milha. Isso se deve principalmente em razão dessas redes oferecerem uma boa relação custo-benefício ao usuário final, ou seja, serem redes com alta capacidade de transmissão de dados a um custo de implantação relativamente baixo.

A grande vantagem do padrão IEEE 802.16 quando comparado às tecnologias de acesso cabeadas é a rápida disponibilização de serviços em áreas de difícil acesso. Dessa forma, o padrão IEEE 802.16 permitirá acelerar a introdução da tecnologia de banda larga sem fio no mercado, bem como aumentar o desempenho e confiabilidade dos serviços oferecidos pelos provedores de acesso [8].

Nesse sentido, o objetivo deste capítulo é apresentar a estrutura geral do padrão IEEE 802.16. Para isso, inicialmente é apresentado na Seção 2.2 um histórico sobre a evolução do padrão IEEE 802.16, desde o lançamento da sua primeira versão, em 2002, até a versão atual, o IEEE 802.16m. Em seguida, na Seção 2.3, é descrita minuciosamente a especificação do padrão IEEE 802.16, em relação à infraestrutura de operação e às camadas que o compõe seguindo o modelo de referência. Finalmente, encerra-se este capítulo na Seção 2.4 com a apresentação das considerações finais.

2.2 Histórico de Evolução

A primeira versão do padrão IEEE 802.16 ficou pronta em outubro de 2001 e tornou-se pública em abril de 2002. Essa versão inicial simplesmente definia uma especificação para a interface aérea de redes metropolitanas (MANs). A intenção por trás dessa primeira versão do padrão era definir uma tecnologia de acesso banda larga sem fio (BWA - *Bandwidth Wireless Access*) para usuários fixos, como uma alternativa para as redes de acesso cabeadas, tais como as redes DSL (*Digital Subscriber Line*). Por essa razão, o padrão IEEE 802.16 original define uma arquitetura de rede onde os recursos são compartilhados por um nó central chamado de estação base (BS - *Base Station*) e um conjunto de estações-assinantes (SS - *Subscriber Station*). Esse modo de operação, que é investigado mais detalhadamente na Seção 2.3.1, se encaixa perfeitamente em um cenário típico de BWA, onde múltiplas estações são servidas por um provedor de serviços de Internet ou serviços de distribuição em *broadcasting* de vídeos digitais (DVB - *Digital Video Broadcasting*), por exemplo [11].

Na versão preliminar do padrão IEEE 802.16, a camada de acesso ao meio (MAC) foi projetada para ser orientada à conexão com suporte à QoS. Entretanto, apesar do padrão ter sido modelado para suportar um conjunto de interfaces aéreas baseadas num protocolo MAC comum, as especificações para a camada física dependiam de regulamentações para o uso. Inicialmente, a faixa de frequências de operação do padrão IEEE 802.16 foi definida para o intervalo 10-66 GHz, operando somente em ambientes com linha de visada (LOS - *Line of Sight*) e usando uma única portadora de transmissão [11].

Em janeiro de 2003 foi ratificada uma nova versão do padrão, intitulada IEEE 802.16a-2003 [12], com o objetivo de prover acesso em banda larga para usuários fixos na última milha. Como um melhoramento do padrão até então em vigor, o padrão IEEE 802.16a incorporou uma faixa de frequência não licenciada de 2-11 GHz. Com esse espectro de frequência mais baixo, viabilizou-se a operação em ambientes sem linha de visada (NLOS - *No Line of Sight*).

O padrão IEEE 802.16c (o projeto do padrão IEEE 802.16b foi abandonado pelo IEEE) pouco acrescentou em relação ao 802.16a. O principal melhoramento foi tornar o espectro de frequências de 2-66 GHz em conformidade para ambientes com e sem linha de visada e a adição de alguns aspectos concernentes à QoS.

Seguindo a linha de evolução, em setembro de 2003 o grupo de trabalho ‘d’ do IEEE

iniciou um processo de revisão dos padrões 802.16, 802.16a e 802.16c, o que acabou instituindo o padrão IEEE 802.16d. Em junho de 2004, o projeto IEEE 802.16d foi finalizado e aprovado pelo IEEE, como substituto dos padrões até então existentes, e passou a ser denominado IEEE 802.16-2004 [13]. Esse novo padrão passou a suportar tipos de camadas físicas adicionais: multiplexação por divisão de frequência ortogonal (OFDM - *Orthogonal Frequency Division Multiplexing*) e acesso múltiplo por divisão de frequência ortogonal (OFDMA - *Orthogonal Frequency-Division Multiple Access*). Uma outra característica também importante acrescentada no padrão IEEE 802.16-2004 foi o suporte a antenas MIMO (*Multiple Input Multiple Output*), que de certa maneira aumentava o alcance do sinal com multipercursos e facilitava as instalações através de antenas *indoor* [11].

Em 2005, uma nova versão do padrão foi lançada para permitir operações combinadas entre estações fixas e móveis usando faixas de frequências licenciadas. Essa nova versão, denominada IEEE 802.16e-2005 [14], foi definida como um melhoramento do padrão IEEE 802.16-2004 e adicionou várias características em relação à mobilidade, incluindo economia de energia, operações de *handover* e uma camada física OFDMA melhorada. Depois do lançamento dessa versão em 2005, deu-se continuidade ao projeto do padrão com a definição de uma base de informação de gerenciamento (MIB - *Management Information Base*) para a camada MAC e física (IEEE 802.16f [15]), um plano de gerenciamento e procedimentos (IEEE 802.16g [16]), o aperfeiçoamento das operações em faixas de frequências não licenciadas (IEEE 802.16h [17]), a introdução de estações com a capacidade de retransmissão (IEEE 802.16j [18]) e o refinamento das operações envolvendo mobilidade na camada física e MAC (IEEE 802.16-2009 [19]). Essa última atualização traz ainda a possibilidade das estações operarem em modo *half-duplex* e a duplexação por divisão de frequências (FDD) em OFDMA [11].

No início de 2007, o grupo de trabalho do IEEE 802.16 iniciou um novo projeto chamado IEEE 802.16m [20] (esse padrão é conhecido também como *WirelessMAN-Advanced* ou WiMAX-2), com o objetivo de agregar características ao padrão IEEE 802.16j a fim de atender os rigorosos requisitos do ITU-R (*International Telecommunications Union - Radiocommunication Sector*), através do IMT-Advanced (*International Mobile Telecommunications*), e representar a tecnologia de quarta geração (4G). Entre as principais melhorias agregadas pelo padrão IEEE 802.16m ao longo desses últimos anos destacam-se as novas funcionalidades na camada física que inclui o suporte para operar com múltiplas

portadoras simultaneamente, o que possibilita taxas de transmissão que em teoria podem exceder 1 Gbps [21]. Após quatro anos de desenvolvimento e inúmeros testes operacionais, em abril de 2011 o IEEE aprovou definitivamente o padrão IEEE 802.16m.

A Tabela 2.1 traz um resumo do processo de evolução do padrão IEEE 802.16.

2.3 Arquitetura do Padrão IEEE 802.16

A arquitetura geral do padrão IEEE 802.16 consiste basicamente de uma estação central, BS, e várias estações assinantes, SSs, as quais se comunicam utilizando o espectro de radiofrequência. A BS está ligada à infraestrutura de rede cabeada e provê acesso à Internet para as SSs que estão sob seu raio de cobertura.

O padrão IEEE 802.16 foi projetado para servir de rede de acesso e também um meio para a interligação de redes heterôgeneas. Em razão disso, as SSs podem representar tanto dispositivos móveis quanto computadores pessoais e até mesmo redes locais, como ilustrado na Figura 2.1. O padrão, no entanto, não especifica como deve ocorrer a comunicação dentro das subredes formadas pelas SSs, delegando essa tarefa a protocolos específicos de redes como o IEEE 802.3 (Ethernet) ou IEEE 802.11 (padrão WiFi).

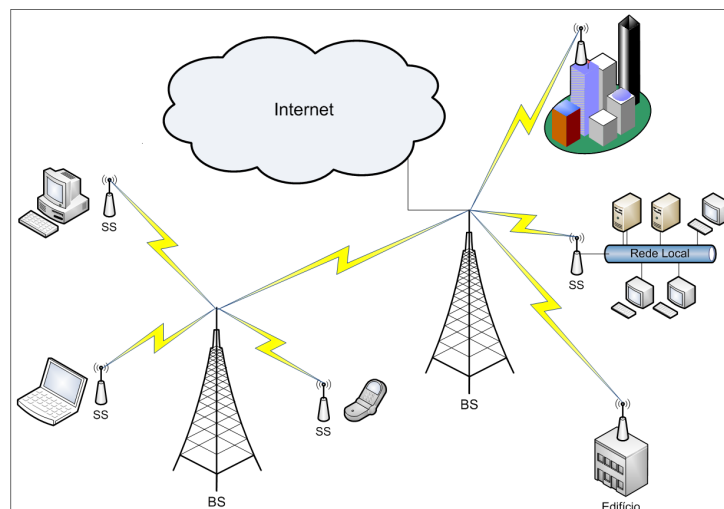


Figura 2.1: Infraestrutura de operação no padrão IEEE 802.16.

O padrão IEEE 802.16 especifica duas topologias que determinam a maneira como as SSs se comunicam na rede. O modo tradicional de operação é o PMP (*Point-to-multipoint*), no qual os recursos são compartilhados entre as SSs, e o segundo é o modo de operação em malha ou *mesh*, em que não há compartilhamento. Nas Seções seguintes

Tabela 2.1: Evolução do padrão IEEE 802.16

Versão	Data	Características
802.16	2001	- Definição da MAC e PHY (10-63 GHz). - Operação no modo PMP. - Linha de visada (LOS).
802.16a	2003	- Alteração da MAC e PHY para operar em 2-11 GHz. - Operação sem linha de visada (NLOS).
802.16b	abandonado	abandonado
802.16c	2002	Adequação do espectro 2-66 GHz para LOS e NLOS.
802.16d	2003	Revisão das versões 802.16, 802.16a e 802.16c.
802.16-2004	2004	- Incorporação das camadas físicas OFDM e OFDMA. - Suporte à antenas MIMO.
802.16e	2005	- Suporte a mobilidade. - Operação combinada entre estações móveis e fixas.
802.16e-2005	2005	- Melhoramento dos aspectos relacionados à mobilidade. - Incorporação de novas operações de handover. - Melhoramento da camada OFDMA.
802.16f	2005	Definição da MIB para a camada MAC e PHY.
802.16g	2007	Definição de um plano de gerenciamento e procedimentos.
802.16h	2010	Aperfeiçoamento das operações em faixas de frequências não licenciadas.
802.16j	2009	Estações com capacidade de retransmissão.
802.16-2009		- Refinamento das operações envolvendo mobilidade. - Possibilidade das SSs operarem no modo <i>half-duplex</i>. - FDD em OFDMA.
802.16m	2007	- Introdução de um superquadro que permite um rápido acesso ao enlace. - Novos esquemas de subcanalização (<i>overhead</i> menor). - Uso de OFDMA em ambos sentidos <i>downlink</i> e <i>uplink</i>.

serão apresentados detalhadamente esses dois modos de operação.

2.3.1 Modo de Operação

2.3.1.1 Ponto a multiponto

A primeira versão do padrão IEEE 802.16 foi projetada para ser uma opção de rede de acesso mais viável em relação às tecnologias cabeadas (xDSL, *cable modem*, etc) em regiões desprovidas de infraestrutura (zonas rurais, por exemplo). Para isso, essa versão foi desenvolvida para permitir que a estação base, em uma posição estratégica, alcançasse vários usuários simultaneamente. Definiu-se então o modo de operação PMP, que consiste de uma BS ligada a várias SSs em topologia estrela, conforme mostrado na Figura 2.2.

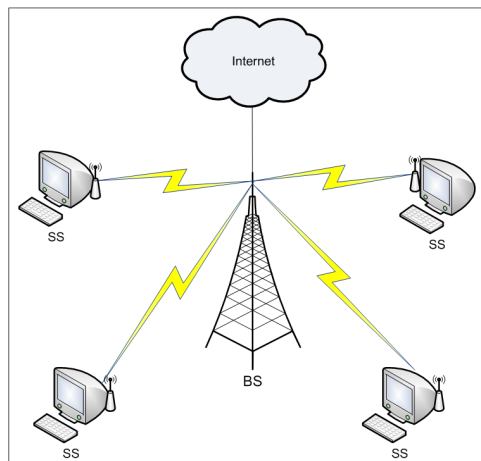


Figura 2.2: Modo de operação PMP.

No modo PMP, as SSs dentro de um determinado setor recebem o mesmo sinal enviado em *broadcast* pela BS geralmente vindo do provedor de serviços de Internet (ISP - *Internet Service Provider*). Cada SS precisa então capturar e processar somente o tráfego endereçado à ela própria, fazendo com que sua arquitetura, em termos de *software*, seja leve [22].

Apesar de simples, a arquitetura PMP está longe de ser uma abordagem adequada para representar o atual cenário das redes de acesso de banda larga sem fio. Essa topologia de rede tem sido usada com sucesso há várias décadas em sistemas telefônicos, redes de TV a cabo e várias outras aplicações. No entanto, para sistemas sem fio, essa arquitetura requer um considerável investimento, dependendo do tamanho da rede, em infraestrutura. Isso ocorre devido a necessidade da instalação de repetidores ou terminais remotos em

locais estratégicos, a fim de que o sinal transmitido chegue com qualidade em toda a área de cobertura [1].

2.3.1.2 Malha

Partindo das restrições impostas pelo modo de operação PMP, o padrão IEEE incorporou em várias de suas versões um modo de operação adicional conhecido como modo de operação em malha ou *mesh*, como também é conhecido. Em contraste com o modo PMP, no modo *mesh* o tráfego não está restrito a fluir somente no sentido BS-SS e SS-BS mas também pode ocorrer diretamente entre as SSs. Na topologia *mesh*, portanto, não é necessário investir em infraestrutura extra para prover escalabilidade, pois como existem enlaces redundantes entre cada estação, a área de cobertura é automaticamente ampliada. Dessa maneira, as estações são dinamicamente auto-configuradas e precisam adotar a arquitetura *ad hoc*. Para que haja sucesso na comunicação, no entanto, as SSs precisam atuar como roteadores *mesh*, e isso torna o processo de comunicação um pouco mais complexo. A Figura 2.3 ilustra a arquitetura típica no modo de operação *mesh*.

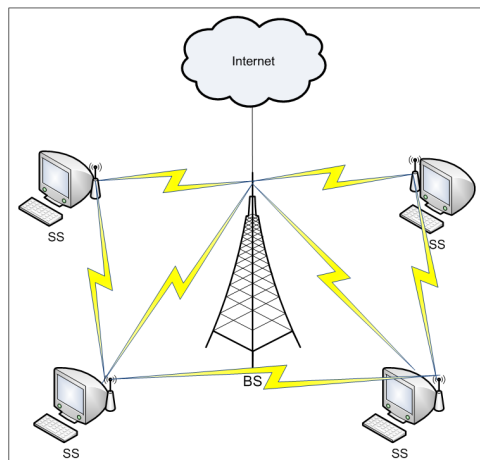


Figura 2.3: Modo de operação *mesh*.

Apesar de mais complexa, a topologia *mesh* se mostra mais adequada para os atuais sistemas de telecomunicações sem fio, uma vez que não impõe limitações quanto à posição das estações nem tampoco requer equipamentos adicionais para um bom funcionamento em qualquer ambiente.

Uma vez identificadas as características principais de cada uma das versões do padrão IEEE 802.16, em seu processo de evolução, e ainda a maneira como uma rede WiMAX funciona, em relação aos modos de operação, é apresentado nas seções seguintes o modelo

de referência do padrão IEEE 802.16, que torna possível toda a comunicação entre as estações.

2.3.2 Modelo de Referência

O padrão IEEE 802.16 está baseado no Modelo OSI (*Open Systems Interconnection*) da ISO (*International Organization for Standardization*) e, por esse motivo, ele especifica uma camada MAC e uma camada física (PHY - *Physical Layer*) para possibilitar o acesso à Internet em banda larga sem fio e interoperar com outros padrões de redes. O modelo de referência determina três planos distintos para organizar as funções dentro dessas duas camadas: plano de dados, plano de controle e plano de gerenciamento. O plano de dados define a maneira como as informações dos usuários devem ser transportadas pela rede. Esse plano trabalha em conjunto com o plano de controle, que tem como objetivo a execução de funções como balanceamento de carga e controle de congestionamento, por exemplo. Tanto o plano de dados como o de controle são desempenhados automaticamente pela rede. Ao contrário, o plano de gerenciamento é operado pelos administradores da rede em funções como engenharia de tráfego e monitoramento dos parâmetros da rede. Apesar disso, o padrão define que as ações em cada um desses planos devem ser coordenadas para que a rede possa ter um bom desempenho [1]. A Figura 2.4 a seguir ilustra o modelo de referência do padrão IEEE 802.16 e seus três planos.

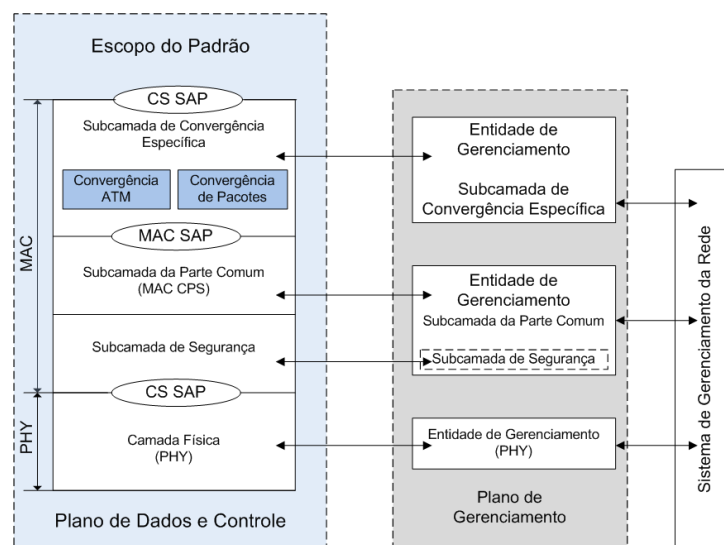


Figura 2.4: Modelo de referência do padrão IEEE 802.16 (adaptado de [1]).

2.3.2.1 Camada MAC

A camada MAC está situada logo acima da camada física no modelo de referência do padrão IEEE 802.16. Sua principal tarefa é prover uma interface entre as camadas superiores e a camada física a fim de possibilitar a transferência dos dados. Os protocolos que operam dentro da camada MAC são responsáveis pela execução das principais funções do padrão IEEE 802.16, entre elas os mecanismos para a provisão de QoS e gerenciamento de mobilidade.

No padrão IEEE 802.16, o protocolo MAC foi inicialmente projetado para a arquitetura PMP. Entretanto, o modo de operação *mesh* tornou necessário o suporte a várias camadas físicas diferentes operando simultaneamente. A estrutura da camada MAC apresentada nesta seção considera apenas o modo de operação PMP.

Conforme pode ser visto na Figura 2.5, no modelo de referência do padrão IEEE 802.16 a camada MAC é dividida em subcamadas com funções específicas. As funções desempenhadas dentro dessas subcamadas são coordenadas de maneira a permitir a transmissão correta dos pacotes na rede IEEE 802.16. A Figura 2.5 ilustra as subcamadas e suas principais funções.

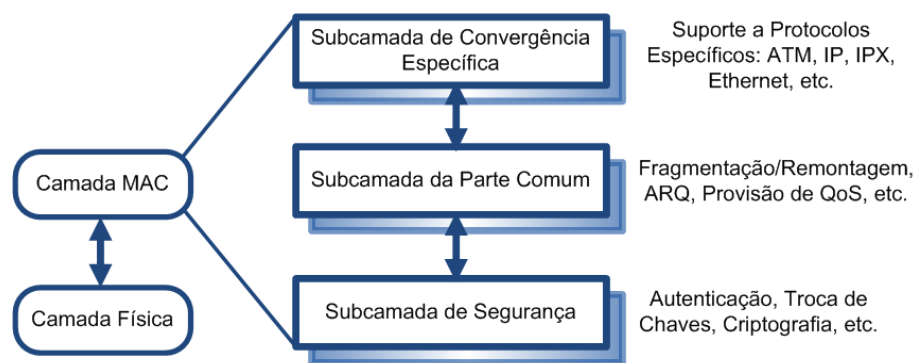


Figura 2.5: Subcamadas no padrão IEEE 802.16 (adaptado de [1]).

2.3.2.1.1 Subcamada de Convergência Específica

A subcamada de convergência específica (CS - *Service-Specific Convergence*) tem por função principal prover o mapeamento dos pacotes que vêm da rede externa, recebidos através do CS-SAP (*Service Access Point*) em SDUs (*Service Data Units*) específicas para o padrão IEEE 802.16, que serão recebidas na subcamada da parte comum (CPS - *Common Part Sublayer*) através do MAC-SAP [1]. Ao receber um pacote, a camada CS

inicialmente remove as informações redundantes contidas nos cabeçalhos. Esse processo é denominado supressão de cabeçalho (PHS - *Packet Header Supression*). Depois disso, o pacote é classificado em um fluxo de serviço específico (SFID - *Service Flow Identifier*), de acordo com sua classe (dados, voz, vídeo), para facilitar o processo de provisão de QoS posteriormente.

Para prover esse mapeamento, a CS necessita de múltiplas especificações para tratar de maneira diferenciada os pacotes de acordo com o protocolo utilizado, no sentido de adequar todos em um só formato. A fim de atender essa necessidade, o padrão IEEE 802.16 define duas subcamadas de convergência específicas: convergência de pacotes e convergência ATM.

A subcamada de convergência de pacotes é definida para mapear as unidades de dados (PDUs - *Packet Data Units*) provenientes de protocolos como o IPv4/IPv6 e Ethernet, por exemplo, em SDUs reconhecíveis para o padrão IEEE 802.16. Particularmente, essa camada é responsável pela interoperabilidade da rede WiMAX com qualquer outra rede baseada em pacotes, uma vez que possibilita a troca de dados independentemente do protocolo utilizado.

A subcamada de convergência ATM, por outro lado, tem por função fazer a associação das células que entram no domínio WiMAX, juntamente com seus parâmetros de conexão como VPI (*Virtual Path Identifier*) e VCI (*Virtual Channel Identifier*), para um determinado fluxo de serviço a fim de que os requisitos de QoS da conexão ATM sejam mantidos.

2.3.2.1.2 Subcamada da Parte Comum

A subcamada da parte comum (CPS) está localizada entre as subcamadas de serviço específico e de segurança e engloba todas as funcionalidades de maior relevância da camada MAC, como por exemplo, a fragmentação e remontagem dos pacotes e os mecanismos para a provisão de QoS. Além disso, essa subcamada é responsável pelo estabelecimento e manutenção das conexões criadas para o transporte dos dados, uma vez que o padrão IEEE 802.16 é orientado à conexão.

A CPS é responsável pela criação da PDU que será transmitida para sua entidade par, através de uma conexão específica. Essa PDU, cujo tamanho máximo é 2048 *bytes*, é composta por um cabeçalho inicial de tamanho fixo (6 *bytes*), um *payload* opcional de tamanho variável e um código de redundância cíclica (CRC - *Cyclic Redundancy Check*)

também opcional de 4 *bytes*, conforme pode ser visto na Figura 2.6.

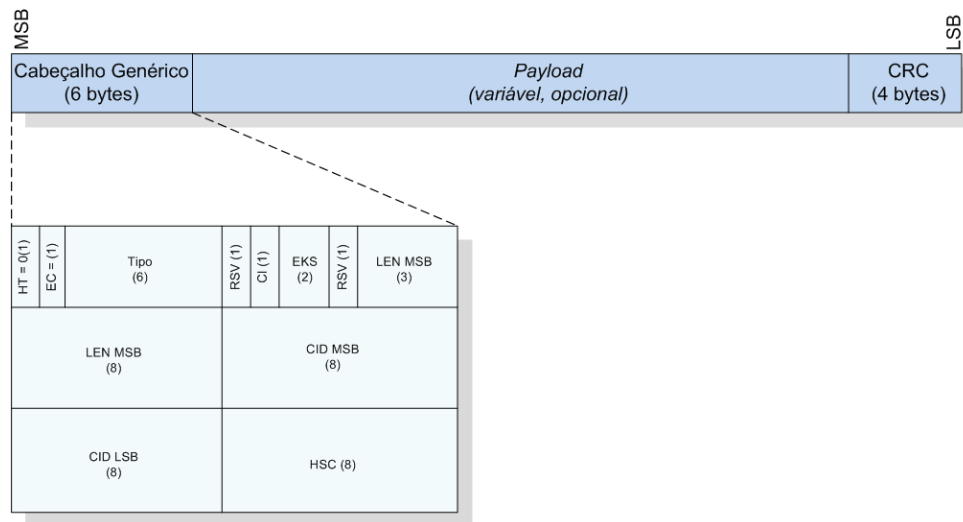


Figura 2.6: Formato genérico da MAC PDU no padrão IEEE 802.16 [2].

O campo HT (*Header Type*) indica o tipo do cabeçalho presente na MAC PDU podendo este ser um tipo genérico (HT = 0), indicando a presença de um *payload* qualquer na PDU ou um cabeçalho específico para requisição de largura de banda (HT = 1) [2]. Em seguida, o padrão especifica um campo EC (*Encryption Control*) de um *bit* indicando se o *payload* está ou não criptografado. Um campo de 6 *bits* é utilizado em seguida para especificar o tipo de carga útil que está sendo transportada pela PDU. Esse campo possibilita a indentificação de até 64 tipos diferentes de carga útil (2^6). Dois campos de um *bit* cada são reservados para uso futuro. O campo CI (*CRC Indicator*) informa se há (CI = 1) ou não (CI = 0) um código CRC no final da MAC PDU. O campo EKS (*Encryption Key Sequence*) de 2 *bits* indica qual chave foi utilizada na criptografia. Sua ausência é denotada pelo campo EC = 0. O campo LEN MSB (*Length of Most Significant Byte*) de 11 *bits* informa o tamanho total da MAC PDU (comprimento do quadro, incluindo CRC). O campo CID MSB e CID LSB (*Connection Identifier - Least Significant Byte*) possui 16 *bits* e informa qual conexão a PDU pertence. Por último, o padrão especifica o campo HCS (*Header Check Sequence*) de 8 *bits* para detectar erros presentes no cabeçalho [2].

Tendo em vista que o padrão IEEE 802.16 possibilita a interoperabilidade com redes heterogêneas e possui ainda uma limitação quanto ao tamanho máximo da carga útil a ser transportada (2048 *bytes*), são definidas na subcamada CPS as funções para a fragmentação e remontagem de MAC PDUs, ou seja, funções para dividir a MAC PDU em partes menores e juntar esses fragmentos em uma MAC PDU grande. A fragmentação,

além de possibilitar a adequação quanto ao tamanho do *payload* a ser transportado ainda tem a vantagem de amenizar o efeito das perdas. Por outro lado, ela contribui para uma diminuição sutil da vazão total da rede, pois requer *overheads* adicionais.

Na subcamada CPS também são implementados os mecanismos para a requisição de largura de banda especificados pelo padrão IEEE 802.16. O padrão define intervalos de *polling* periódicos em que as SSs podem solicitar banda individualmente através de mensagens de controle identificadas pelo campo *type*. Além disso, o padrão possibilita o envio dessas mensagens de requisição em *piggyback* sobre os dados transmitidos no sentido *uplink* e também define uma porção de largura de banda reservada exclusivamente para as SSs competirem pela oportunidade (mecanismo de contenção) de enviar a requisição. A partir dessas requisições, a BS aloca *grants* ou autorizações para cada SS solicitante, através do modo de concessão GPSS (*Grant per Subscriber Station*), ou para cada conexão individual, através do modo de concessão GPC (*Grant per Connection*). A Figura 2.7 mostra as principais funcionalidades da CPS.

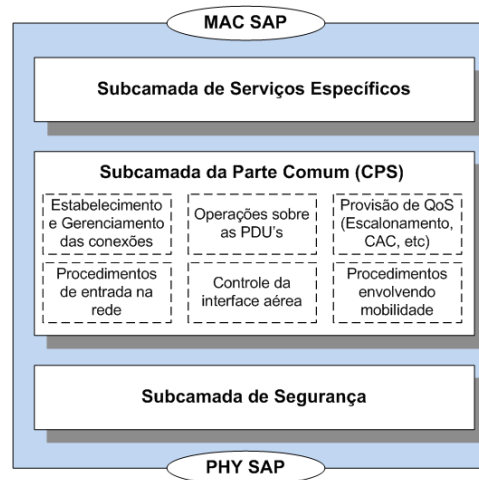


Figura 2.7: Funcionalidades da subcamada CPS (adaptado de [1]).

As funções para a provisão de QoS bem como os detalhes do estabelecimento de conexões, que formam a essência da subcamada CPS, são discutidos no Capítulo 3.

2.3.2.1.3 Subcamada de Segurança

Em razão das redes sem fio serem menos seguras, devido a própria infraestrutura de acesso, uma atenção especial deve ser dada aos aspectos concernentes à segurança [1]. Partindo desse pressuposto, o padrão IEEE 802.16 especifica uma camada exclusiva

para controle de acesso e criptografia (PS - *Privacy Sub-layer*), com o intuito de prover segurança no nível de enlace.

Para cumprir esse objetivo, a subcamada PS define um protocolo de encapsulamento e um protocolo de gerenciamento de chaves. O protocolo de encapsulamento tem por finalidade criptografar os pacotes de dados que cruzam a rede IEEE 802.16. Ele define um conjunto de premissas criptográficas e algoritmos de autenticação, além de regras para aplicar esses algoritmos na carga útil das MAC PDUs. Já o protocolo de gerenciamento de chaves é responsável pela definição de regras para a distribuição de chaves entre BS e SSs. Além disso, a BS utiliza esse protocolo para controlar o acesso aos serviços da rede, por parte das SSs [1].

O padrão IEEE 802.16 define basicamente quatro algoritmos de criptografia:

- *Rivest Shamir Adleman* (RSA) - O RSA é o algoritmo de criptografia de chave pública mais amplamente utilizado em razão da dificuldade de ser violado. No padrão IEEE 802.16 ele é utilizado para criptografar as mensagens de autorização usando a chave pública na SS. Ele é utilizado também para o transporte seguro das chaves entre a BS e a SS [23].
- *Data Encryption Standard* (DES) - Esse método de criptografia foi adotado como padrão pelo governo americano em 1976 e foi utilizado em larga escala internacionalmente. Nos últimos anos não vem sendo muito utilizado em razão do surgimento de máquinas com elevado potencial de processamento que possibilita a rápida descoberta das chaves de 56 *bits*. Apesar disso, ele é obrigatório no padrão IEEE 802.16 para a criptografia de chaves e de dados [24].
- *Advanced Encryption Standard* (AES) - O AES é um método de criptografia de chave simétrica, ou seja, possui uma única chave tanto para a BS como para a SS. Esse padrão foi uma evolução do DES e é também utilizado para a encriptação de chaves e dados [25].
- *Hashed Message Authentication Code* (HMAC) e *Cipher-based Message Authentication Code* (CMAC) - O HMAC e CMAC são códigos utilizados para proteger tanto mensagens de dados como de controle. O padrão IEEE 802.16 inclui o uso do HMAC para controle de integridade e autenticação. O CMAC é usado no IEEE 802.16e como uma alternativa ao HMAC [26].

O padrão IEEE 802.16 define que esses algoritmos de criptografia devem ser aplicados somente sobre a carga útil das MAC PDUs, ou seja, o cabeçalho genérico e as mensagens de gerenciamento sempre serão transmitidas sem criptografia. Apesar de importante para o padrão IEEE 802.16 a exploração detalhada desses algoritmos está fora do escopo deste trabalho.

2.3.2.2 Camada Física

Como o próprio nome sugere, o objetivo principal da camada física (PHY) é transportar os dados do usuário, na forma de *bits*, entre nós adjacentes da rede. Isso implica que a camada física deve conter funções para codificar os *bits* dos usuários em um formato adequado para a transmissão no meio sem fio. Além disso, em razão do enlace sem fio ser compartilhado, é necessário o emprego de alguma forma de duplexação. No padrão IEEE 802.16 são definidas duas dessas técnicas, especificadas nas seções seguintes.

2.3.2.2.1 Técnicas de Duplexação

O padrão IEEE 802.16 especifica duas técnicas de duplexação do meio físico. A primeira delas é conhecida como duplexação por divisão do tempo (TDD - *Time Division Duplexing*) e a segunda de duplexação por divisão de frequência (FDD - *Frequency Division Duplexing*). No modo de duplexação TDD, o sistema transmite e recebe *bits* dentro de um mesmo canal de radiofrequência, assumindo porções de tempos específicas para a transmissão e recepção desses *bits*. Já a técnica FDD requer duas faixas de frequências separadas, uma para cada direção, geralmente de 50 a 100 MHz, porém as transmissões são realizadas simultaneamente [7].

A técnica TDD apresenta uma vantagem sobre o FDD pois permite a alocação do espectro em blocos adjacentes, o que facilita as operações de *hardware*. Além disso, como o TDD não requer separação de banda, a utilização do espectro é feita eficientemente para ambos os sentidos, *upstream* e *downstream*. Por essa razão, a técnica mais utilizada tem sido a TDD, a qual é enfatizada nesta dissertação. A Figura 2.8 ilustra a diferença entre os modos TDD e FDD.

2.3.2.2.2 Estrutura do Quadro TDD

A transmissão de dados no modo de duplexação TDD exige que os tempos dedicados ao *downstream* e *upstream* sejam conhecidos. Isso é realizado através de um quadro de

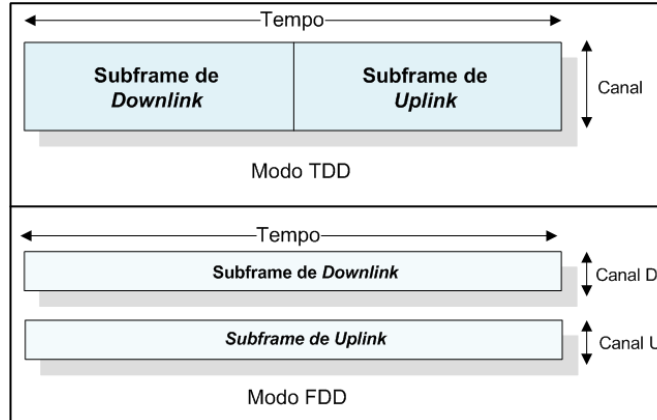


Figura 2.8: Modo de duplexação TDD e FDD (adaptado de [3]).

operação na camada física que especifica com exatidão os instantes de tempos em que cada evento deve acontecer, tais como: transmissão de uma rajada de dados, recebimento do preâmbulo e transição do modo *downlink* para o modo *uplink* e vice-versa.

No modo TDD, a BS basicamente transmite um sinal TDM (*Time Division Multiplexing*). Esse sinal é composto por uma série de porções de tempo chamadas de *time slots*, uma para cada SS na rede. Conforme mostrado na Figura 2.9, o subframe de *downlink* inicia-se com um preâmbulo, que é utilizado para a sincronização entre a BS e as SSs. Esse preâmbulo, sendo informação de controle, evidentemente não passa pelo processo de criptografia como já mencionado [1].

Depois do preâmbulo, o padrão define uma seção de controle, enviada em *broadcast*, chamada de mapa de *uplink* (UL-MAP) e mapa de *downlink* (DL-MAP). Essa seção de controle especifica quando as transmissões na camada física ocorrem dentro do quadro TDD. A porção DL-MAP da seção de controle permite que as SSs identifiquem as características do canal de *downlink*. Assim sendo, as SSs têm condições de acessarem corretamente os dados que a elas são destinados. Já o UL-MAP é utilizado para definir o instante de tempo em que cada SS acessará o canal de *uplink*, para transmitir dados ou informações de controle [1].

Em seguida, o padrão especifica um descritor de canal de *downlink* (DCD - *Downlink Channel Descriptor*) e de *uplink* (UCD - *Uplink Channel Descriptor*) necessário para que as SSs conheçam os perfis das rajadas de *downlink* e *uplink*, respectivamente, incluindo as modulações e esquemas de correção de erros (FEC - *Forward Error Correction*).

Após o DCD e UCD, é definido um conjunto específico de *time slots* que comporá o

subquadro de *downlink*, transmitido em *broadcast* pela BS. Esse conjunto é subdividido em porções unitárias, uma para cada SS de destino, ou seja, cada SS que receberá os dados. Depois disso, encerra-se o subquadro de *downlink* com uma seção chamada TTG (*Transmit Transition Gap*), que representa uma porção pequena de tempo necessária para as SSs mudarem do modo de recepção para transmissão [1].

Terminado o subquadro de *downlink*, o padrão especifica um conjunto de *time slots* destinados às SSs que queiram entrar na rede. Esse processo é denominado *initial ranging* e compõe uma das etapas que as SSs devem passar para entrar na rede.

Como as SSs já conhecem os instantes de tempo em que cada uma transmitirá (através do UL-MAP), os *time slots* seguintes serão todos utilizados para a transmissão *uplink* das SSs tanto para dados como para informações de controle. Ao final do quadro, o padrão define uma seção chamada RTG (*Receive Transition Gap*), necessária para separar os quadros subseqüentes e permitir a mudança do *status* das estações de “transmitindo” para “recebendo”. A Figura 2.9 mostra o formato detalhado do quadro TDD.

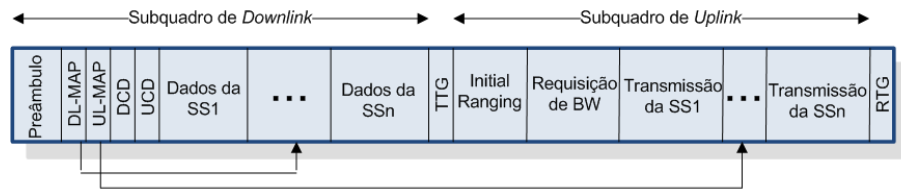


Figura 2.9: Estrutura do quadro TDD.

Para a transmissão de dados no meio físico, o padrão IEEE 802.16 define cinco especificações básicas para a camada física no que diz respeito à utilização do espectro de frequências. Essas especificações serão detalhadas nas seções seguintes.

2.3.2.2.3 Especificação WirelessMAN-SC PHY

Nessa especificação, a camada física utiliza uma única portadora, entre 10 e 66 GHz, para a transmissão dos *bits* e foi projetada para a arquitetura PMP em ambientes LOS. No sentido *uplink*, o compartilhamento do enlace sem fio é realizado através do acesso múltiplo por divisão de tempo (TDMA - *Time Division Multiple Access*) e acesso múltiplo sob demanda (DAMA - *Demand Assigned Multiple Access*) [7].

No modo TDMA, as estações acessam o enlace em períodos de tempo pré-fixados, para evitar que haja colisão. Esses tempos são definidos pelo escalonador na camada MAC e são repassados às estações por meio da mensagem UL-MAP, conforme já explanado. No modo

DAMA, no entanto, não existem tempos pré-estabelecidos para as estações acessarem o meio. Ao invés disso, esse esquema permite que as estações disputem, sob demanda, o acesso ao enlace sem fio e isso pode fazer com que ocorram colisões.

Essa especificação utiliza quadros de operação com 0,5, 1 ou 2 milissegundos. Esse quadro, por sua vez, é dividido em *slots* físicos com o propósito de permitir a alocação de largura de banda na camada MAC e identificar as transições da camada física [7].

2.3.2.2.4 Especificação WirelessMAN-SCa PHY

Essa variante usa um esquema de modulação em uma única portadora, dentro do espectro de frequências de 2 a 11 GHz, e foi modelada para operações em ambientes NLOS. Ela suporta pelo menos um tipo de duplexação, FDD ou TDD e, assim como na especificação WirelessMAN-SC PHY, utiliza-se o modo de acesso TDMA no sentido *uplink* além do modo TDM no sentido *downlink*. Essa especificação define ainda estruturas de quadros que permitem melhor equalização e estimação de performance em ambientes com atrasos de espalhamento elevados. As modulações utilizadas nessa especificação são: BPSK (*Binary Phase Shift Keying*), QPSK (*Quadrature Phase Shift Keying*), 16-QAM (*QAM - Quadrature Amplitude Modulation*), 64-QAM e 256-QAM, sendo essa última opcional [27].

2.3.2.2.5 Especificação WirelessMAN-OFDM PHY

Essa especificação foi projetada para permitir a operação em ambientes NLOS através da modulação OFDM. A técnica OFDM é uma forma avançada de FDM (*Frequency Division Multiplexing*) onde as frequências são ortogonais entre si, ou seja, são cuidadosamente sobrepostas de maneira a permitir uma utilização mais eficiente da largura de banda. Nessa especificação, o padrão define um conjunto de 255 subportadoras, sendo 192 destinadas à transmissão de dados, 8 para informações de controle e 55 bandas de guarda. Cada canal pode conter um subconjunto de subportadoras e elas não necessariamente precisam ser adjacentes como no FDM. A Figura 2.10 mostra a eficiência espectral do OFDM quando comparado ao FDM.

A grande vantagem da especificação OFDM, quando comparado ao FDM tradicional e as especificações SC (*Single Carrier*) e SCa (*Single Carrier Adaptive*), é que ele permite a transmissão de dados em mais de uma portadora no mesmo *time slot*, que nesse contexto é conhecido como símbolo. Isso permite que o padrão IEEE 802.16 apresente um alto

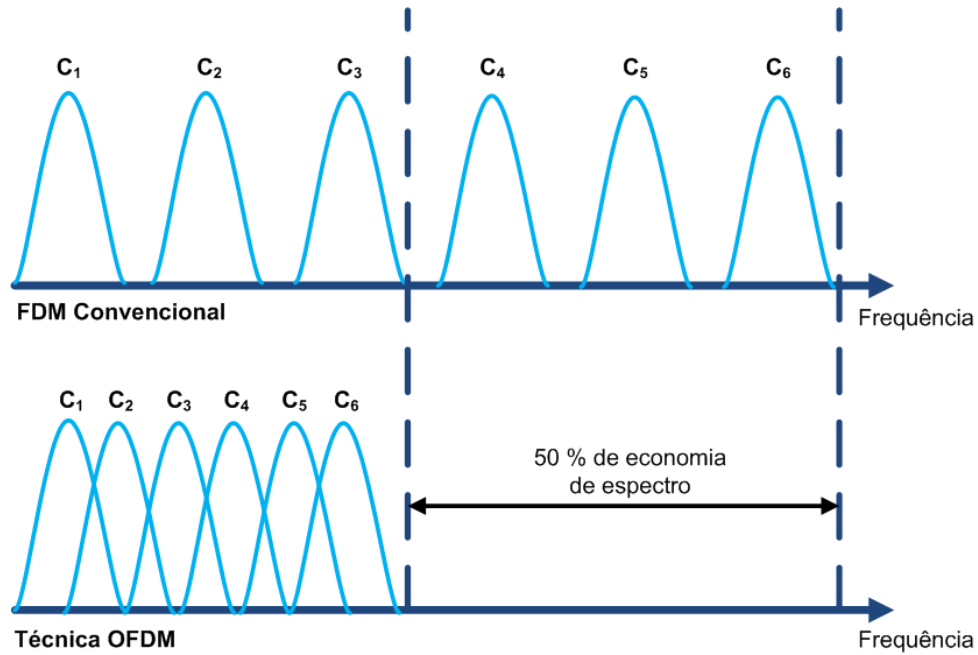


Figura 2.10: Eficiência espectral da técnica OFMD [4].

ganho em vazão. Apesar disso, as subportadoras em cada símbolo são mapeadas para um único usuário, o que de certa maneira limita a vazão agregada final.

O padrão IEEE 802.16 especifica também uma interface aérea chamada WirelessMAN-HUMAN (HUMAN - *High-speed Unlicensed MAN*), semelhante ao WirelessMAN-OFDM, designada para operar em bandas não licenciadas e modo de operação exclusivo TDD.

2.3.2.2.6 Especificação WirelessMAN-OFDMA PHY

A WirelessMAN-OFDMA PHY está entre as três especificações definidas no padrão IEEE 802.16 para aplicações abaixo de 11 GHz. No processo de evolução do padrão, conforme visto na Seção 2.2, essa especificação foi acrescentada na versão IEEE 802.16e-2005 com o intuito de prover maior flexibilidade de operação em diferentes cenários [1].

De forma semelhante ao OFDM, a tecnologia OFDMA emprega múltiplas subportadoras sobrepostas (a especificação trabalha com até 2048 subportadoras). A principal diferença está na subdivisão dessas subportadoras em grupos. Cada grupo é chamado de subcanal. As subportadoras que formam um subcanal não precisam ser adjacentes. No fluxo de transmissão *downlink*, esses subcanais podem ser requisitados por diferentes receptores. Já no fluxo de transmissão *uplink*, um transmissor, nesse caso a própria SS, pode ser associado a um ou mais subcanais (subcanalização) o que permite um elevado ganho em vazão. Usando esse conceito de subcanalização, uma estação base pode, num

mesmo *time slot*, alocar mais capacidade de transmissão para as SSs com uma menor SNR (*SNR* - *Signal to Noise Ratio*) e menor potência para as SSs com grande relação sinal-ruído [4]. A Figura 2.11 ilustra a diferença entre os sistemas OFDM e OFDMA.

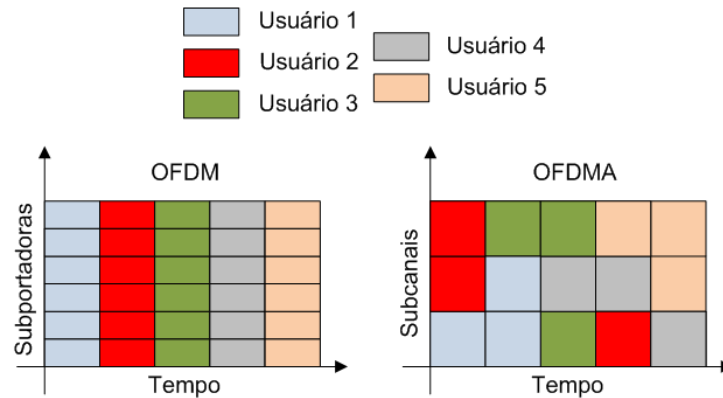


Figura 2.11: Comparação entre a técnica OFDM e OFDMA (adaptado de [4]).

A subcanalização do tráfego *uplink* pode conservar a capacidade de transmissão das SSs, uma vez que ela pode concentrar a potência somente para os subcanais a ela alocados. Essa característica de conservação de energia é particularmente útil uma vez que é uma questão crítica no projeto de sistemas *wireless* [4].

A Tabela 2.2 traz um resumo das principais características sobre as especificações para a camada física existentes no padrão IEEE 802.16.

Tabela 2.2: Características das especificações PHY [7].

Especificação	Operação	Frequência	Duplexação
WirelessMAN-SC	LOS	10-66 GHz	TDD, FDD
WirelessMAN-SCa	LOS e NLOS	2-11 GHz	TDD, FDD
WirelessMAN-OFDM	LOS e NLOS	2-11 GHz	TDD, FDD
WirelessMAN-OFDMA	LOS e NLOS	2-11 GHz	TDD, FDD
WirelessMAN-HUMAN	LOS e NLOS	2-11 GHz	TDD

2.4 Considerações Finais

Neste capítulo, identificou-se as principais características do padrão IEEE 802.16 como rede de acesso banda larga sem fio. Inicialmente foi apresentado o histórico de evolução

da família de padrões IEEE 802.16, desde a sua primeira versão que surgiu em 2001 até os dias atuais. Também foram apresentadas as arquiteturas PMP e *mesh*, fornecendo uma visão geral de como ocorre a comunicação entre as estações de assinantes e a estação base. Uma atenção especial foi dada ao modelo de referência do padrão, o qual consiste na especificação detalhada das camadas MAC e Física. As três subcamadas que compõem a MAC foram descritas e as funcionalidades de cada uma delas apresentadas, incluindo a interoperabilidade da rede WiMAX com outras redes de comutação de pacotes e os aspectos concernentes à segurança. Por fim, foram apresentadas as principais especificações para a camada física existentes no padrão IEEE 802.16.

Capítulo 3

Provisão de QoS em Redes IEEE 802.16

3.1 Introdução

O termo Qualidade de Serviço (QoS) surgiu originalmente na área de comunicações para descrever um certo conjunto de características técnicas que se observava nas transmissões de dados [28]. O modelo de referência OSI, por exemplo, que surgiu no final dos anos 70, já descrevia alguns desses parâmetros (taxa da transmissão, vazão, latência, taxa de erros, taxa de perdas) e recomendava que eles deveriam ser monitorados a fim de garantir um certo nível de qualidade na transmissão. Em agosto de 1994 foi publicada a recomendação E.800 do ITU (*International Telecommunications Union*) [29] o qual definiu a QoS como sendo o efeito coletivo do desempenho do serviço que determina o grau de satisfação de um usuário desse serviço. Essa definição sugere que a QoS está associada às características de rede que as aplicações necessitam para funcionarem corretamente, ou seja, aplicações com requisitos estritos de atraso, por exemplo, devem rodar somente em redes que fornecem QoS, a menos que essas aplicações sejam do tipo elásticas.

Com o surgimento e a rápida expansão das aplicações multimídia, a preocupação em prover QoS aos usuários finais passou a ser uma questão primordial. Várias tentativas de prover QoS sobre a arquitetura TCP/IP existente foram realizadas, como, por exemplo, a criação dos Serviços Diferenciados (*DiffServ*) e Serviços Integrados (*IntServ*), alguns dos quais até aparecendo como uma alternativa promissora. Entretanto, uma série de desafios surgem quando se trata de prover QoS em redes de acesso sem fio, incluindo o próprio padrão IEEE 802.16, principalmente em razão do modo peculiar como essas redes se organizam e da instabilidade inerente ao meio físico, o qual é baseado em radiofrequência,

que deve ser considerada.

Dessa maneira, o objetivo principal deste capítulo é descrever os principais mecanismos adotados pelo padrão IEEE 802.16 no que se refere à provisão de QoS. Para isso, na Seção 3.2 é introduzida a arquitetura para a provisão de QoS, consistindo na especificação dos fluxos e classes de serviço definidos para o padrão. Na Seção 3.3 são apresentados os principais mecanismos para a provisão de QoS, incluindo as técnicas para escalonamento de pacotes (Seção 3.3.1) e controle de admissão de conexões (Seção 3.3.2). Esses mecanismos são descritos de acordo com o estado da arte das pesquisas relacionadas ao tema em questão. Finalmente, na Seção 3.4 as considerações finais são apresentadas.

3.2 QoS no Padrão IEEE 802.16

Segundo M. Ma *et al.* [30], o termo QoS pode ser interpretado sob duas perspectivas diferentes: a do usuário e a da rede. Da perspectiva do usuário, a QoS refere-se à qualidade da aplicação sob o ponto de vista do usuário final. Já na perspectiva da rede, a QoS refere-se à qualidade do serviço que a rede oferece para as aplicações (ou usuários) em termos de parâmetros como latência máxima, vazão mínima e taxa de perdas. Partindo dessas premissas, o objetivo geral da rede, no que diz respeito a provisão de QoS, é fornecer um serviço que atenda adequadamente as necessidades dos usuários ao passo que maximiza a utilização de recursos da rede.

Conforme apresentado no Capítulo 2, a garantia de QoS no padrão IEEE 802.16 é realizada na camada MAC, especificamente na subcamada CPS. Uma das características fundamentais existentes na CPS para a provisão de QoS é a associação dos pacotes que passam por ela a um fluxo de serviço, com o objetivo de fornecer tratamento diferenciado ao tráfego gerado pelas aplicações. O fluxo de serviço é identificado por um valor numérico exclusivo de 32 *bits* (SFID) e pode ser visto como um meio para o transporte unidirecional de SDUs. De acordo com o padrão, cada fluxo de serviço é caracterizado por um conjunto específico de parâmetros de QoS conforme mostrado na Tabela 3.1.

O padrão IEEE 802.16 foi projetado para ser orientado à conexão ao nível da camada MAC. Nesse caso, cada conexão é identificada por um inteiro exclusivo de 16 *bits* em cada direção (CID), o qual é associado a um fluxo de serviço no momento do estabelecimento da conexão. O procedimento para se estabelecer uma conexão ocorre da seguinte maneira:

Tabela 3.1: Parâmetros de QoS no fluxo de serviço [8].

Parâmetro	Características
Taxa mínima reservada	Especifica a taxa de transferência de dados mínima para o fluxo de serviço, necessária para garantir uma vazão final mínima para as aplicações.
Taxa máxima sustentável	Define a taxa de pico para o serviço, tanto no sentido <i>down-link</i> como <i>uplink</i> . O algoritmo para o controle desse parâmetro não é definido pelo padrão.
Latência máxima	Especifica a latência máxima entre a transmissão do pacote para a interface de radiofrequência e a recepção do pacote na interface de rede da BS ou da SS.
<i>Jitter</i> tolerável	Define a variação máxima permitida para a latência da conexão.
Rajada de tráfego máxima	Define o tamanho máximo da rajada que deve ser disponibilizado para o serviço.
Tipo do fluxo de serviço	Determina o tipo de serviço ao qual o fluxo deve ser associado. Se o parâmetro é omitido, usa-se o serviço de melhor esforço.
Intervalo de <i>grants</i>	Intervalos entre concessões sucessivas de largura de banda (<i>grants</i>) para um fluxo de serviço tempo real com taxa constante.
Intervalo de <i>polling</i>	Intervalo máximo entre <i>grants</i> sucessivos alocados para um fluxo de serviço em tempo real enviar requisição de banda.
Prioridade do tráfego	O valor desse parâmetro especifica a prioridade atribuída ao tráfego. Dados dois fluxos de serviço idênticos em todos os parâmetros de QoS exceto na prioridade, o fluxo de serviço com maior prioridade deve ter latência menor e prioridade maior no processo de armazenamento no <i>buffer</i> .

A CPS, ao receber uma PDU proveniente da CS, verifica se já existe uma conexão para o destino indicado no cabeçalho da PDU. Se houver, a PDU é mapeada para essa conexão e para o correspondente fluxo de serviço. No entanto, se não houver conexão para o destino especificado, a camada CPS inicializa o processo de estabelecimento de conexão. Para

isso, ela envia uma mensagem chamada DSA-REQ (*Dynamic Service Addition Request*) para a BS contendo o conjunto de parâmetros de QoS referente a essa conexão. O CAC, de acordo com o fluxo de serviço contido na mensagem DSA-REQ e das conexões presentes na rede, decidirá se aceita ou não o pedido da conexão. Essa resposta é inserida na mensagem DSA-RSP (*Dynamic Service Addition Response*) e enviada de volta à SS. A estação completa o processo enviando uma mensagem de confirmação chamada DSA-ACK (*Dynamic Service Addition Acknowledgment*). Esse procedimento pode também ser utilizado para a alteração dinâmica dos parâmetros de QoS referentes a um fluxo de serviço já estabelecido. A Figura 3.1 mostra como esse procedimento é realizado.

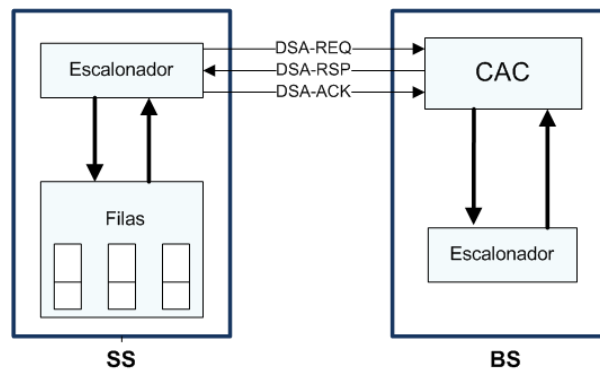


Figura 3.1: Estabelecimento de um fluxo de serviço ou conexão.

Além disso, para suportar uma variedade de aplicações, o padrão IEEE 802.16 especifica cinco classes de serviço que, associadas aos fluxos de serviço (SFID) e as conexões de dados previamente criadas (CID), permitirá ao escalonador de pacotes tratar os dados de acordo com prioridades diferentes. Essas classes de serviço são especificadas como segue:

- **UGS** - A classe de serviço UGS (*Unsolicited Grant Service*) é destinada a aplicações tempo real que geram pacotes de dados de tamanho fixo em períodos regulares, como ocorrem na emulação de circuitos E1/T1 e voz sobre o protocolo IP (VoIP - *Voice Over Internet Protocol*) sem supressão de silêncio. O escalonador na BS aloca *grants* periódicos para as SSs transmitirem os dados a fim de atender o requisito de taxa constante, e por esse motivo as conexões UGS nunca requisitam largura de banda. A quantidade de largura de banda alocada para as conexões UGS é calculada de acordo com a taxa de tráfego mínima reservada definida pelo fluxo de serviço de cada conexão [22].
- **rtPS** - A classe de serviço rtPS (*Real-Time Polling Service*) é destinada a aplicações

tempo real que geram pacotes de dados de tamanho variável em períodos regulares, como ocorrem na transmissão de vídeo *streaming* com codificação MPEG (*Moving Pictures Experts Group*). Diferentemente do que ocorre com as conexões UGS, as conexões rtPS devem informar a BS suas necessidades de largura de banda. Desse modo, a BS deve periodicamente alocar largura de banda para as conexões rtPS solicitarem banda de acordo com seus requisitos. Isso corresponde ao mecanismo de *polling* para requisição de largura de banda. O padrão IEEE 802.16 especifica três variantes de mecanismos de *polling*: *unicast polling*, *multicast polling* e *broadcast polling*. No entanto, somente *unicast polling* pode ser usado para conexões rtPS [22].

- **ertPS** - A classe de serviço rtPS extendida é um novo serviço de escalonamento introduzido pelo padrão IEEE 802.16e-2005 [14] para suportar fluxos de serviço tempo real que geram pacotes de dados de tamanho variável em períodos fixos, tais como aplicações de voz sobre o protocolo IP (VoIP) com supressão de silêncio. Da mesma forma que na classe UGS, a BS provê *grants* para a transmissão dos dados sem a necessidade de mecanismos explícitos para requisição de banda, o qual economiza largura de banda do canal e diminui a latência. Entretanto, diferentemente das alocações UGS, onde os pacotes são de tamanhos fixos, as alocações ertPS são dinâmicas como no rtPS. Por *default*, o tamanho das alocações corresponde ao valor corrente da taxa de tráfego máxima sustentável de cada conexão. A SS pode, no entanto, alterar o tamanho da alocação dinamicamente [22].
- **nrtPS** - A classe de serviço nrtPS (*Non Real-Time Polling Service*) é destinada a aplicações tolerantes a atrasos que necessitam de largura de banda mínima como, por exemplo, transferência de arquivos por FTP (*File Transfer Protocol*). O mecanismo de *polling* também pode ser aplicado para conexões nrtPS. Porém, diferentemente da classe rtPS, as conexões nrtPS não necessariamente empregam o *unicast polling*, ou seja, podem ser empregados também *multicast polling* e *broadcast polling*, e o *polling* deve ser regular, porém não necessariamente periódico. Além de solicitar banda em cada intervalo pré-definido de *polling unicast*, nessa classe a BS permite também que as SSs solicitem banda por *piggyback* e/ou mecanismos de contenção ou disputa [22].

- **BE** - A classe de melhor esforço (BE - *Best Effort*) é destinada a aplicações que não possuem requisitos de atraso nem requerem largura de banda garantida. Nessa classe é realizada solicitação de largura de banda somente por *piggyback* e/ou mecanismos de contenção. Exemplos de aplicações nessa classe incluem tráfego Web (HTTP - *Hyper Text Transfer Protocol*) e correio eletrônico [22].

Uma vez que o padrão IEEE 802.16 opera com cargas de tráfego heterogêneas, é essencial procurar soluções para a provisão de QoS que efetivamente atenda essa ampla variedade de aplicações. Sendo assim, a especificação dessas quatro classes de serviço permite um tratamento diferenciado do tráfego gerado pelas aplicações que será fundamental para atender os diferentes requisitos de QoS como atraso máximo limitado e vazão mínima tolerável.

3.3 Mecanismos para a Provisão de QoS no Padrão IEEE 802.16

A criação de classes e fluxos de serviço por si só não é suficiente para garantir QoS em uma rede IEEE 802.16. Prover QoS implica, além disso, executar uma série de mecanismos como, por exemplo, gerenciar eficientemente os recursos no enlace (incluindo técnicas para o controle de acesso ao meio), executar políticas de escalonamento e CAC tanto na BS quanto nas SSs, realizar policiamento e moldagem de tráfego também na BS e SS, além de prover técnicas para o controle de erros, no sentido de satisfazer e manter os requisitos de QoS [1]. O foco deste trabalho, no entanto, resume-se a escalonamento e CAC, os quais são abordados mais detalhadamente nas seções seguintes.

O padrão IEEE 802.16 define uma arquitetura base para a provisão de QoS, a qual está especificada na Figura 3.2. Como pode-se perceber, o padrão deixa em aberto as questões referentes a escalonamento, CAC e policiamento. Partindo desse pressuposto, é fornecido nas seções seguintes um detalhamento sistemático desses mecanismos, apresentando o estado da arte das pesquisas relacionadas ao desenvolvimento de políticas tanto para escalonamento quanto para CAC.

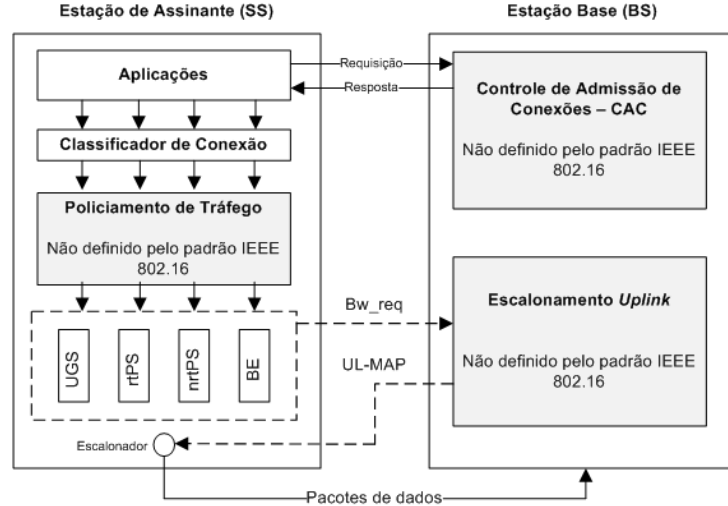


Figura 3.2: Arquitetura base do padrão IEEE 802.16 [5].

3.3.1 Escalonamento de Pacotes

Escalonamento de pacotes refere-se ao processo de decisão usado para escolher quais pacotes, armazenados nos *buffers* de transmissão, devem ser servidos ou descartados. Além disso, é tarefa do escalonamento de pacotes resolver o processo de disputa por largura de banda e determinar a alocação de banda de maneira justa entre os usuários [30].

Nos últimos anos, uma ampla variedade de algoritmos para escalonamento de pacotes tem sido propostos para redes cabeadas, no sentido de prover QoS às aplicações/usuários. Entretanto, esses algoritmos não podem ser aplicados diretamente sobre redes sem fio, incluindo o padrão IEEE 802.16, em razão da capacidade limitada do enlace sem fio e o estado do canal ser dependente da localização da SS em relação à BS [30].

O padrão IEEE 802.16 define um *framework*, conforme ilustrado na Figura 3.3, para o escalonamento tanto na BS (*downlink* e *uplink*) como nas SS (*uplink*), porém não define algoritmos ou disciplinas de escalonamento obrigatórias, conforme já mencionado. Sendo assim, nos últimos anos uma atenção especial tem sido dada à elaboração de propostas de escalonamento para o padrão IEEE 802.16 que visam o suporte aos diferentes aspectos das redes WiMAX (altas taxas de erros, largura de banda limitada, mobilidade dos usuários, capacidade do canal dependente de localização, etc).

No contexto deste trabalho, as disciplinas de escalonamento para o padrão IEEE 802.16 são divididas em categorias, de acordo com a natureza do escalonamento. Vários autores propõem classificações diferentes para os algoritmos de escalonamento, porém

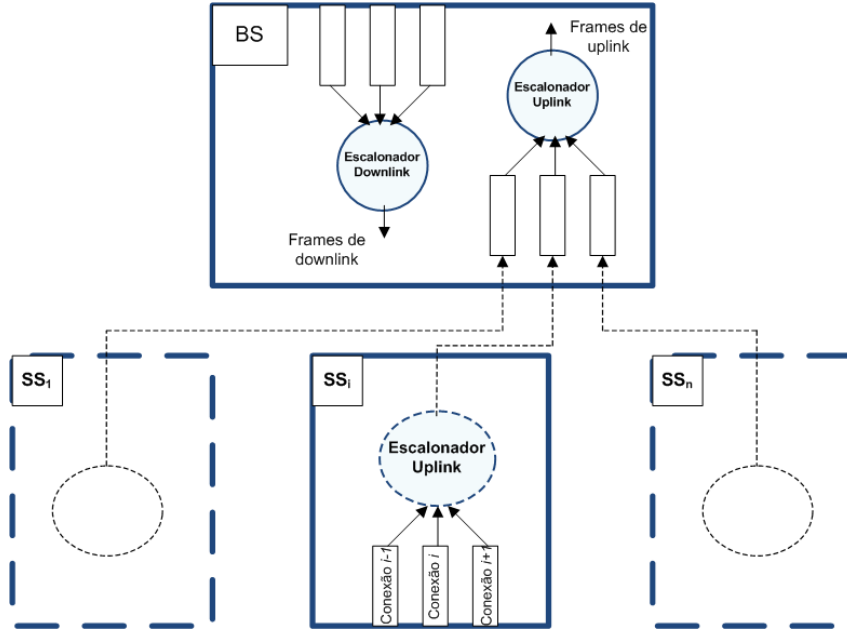


Figura 3.3: *Framework* para o escalonamento na BS e SS [6].

neste trabalho optou-se por utilizar três categorias: algoritmos homogêneos, algoritmos híbridos e algoritmos que utilizam a abordagem *cross-layer*, conforme mostra a Figura 3.4.

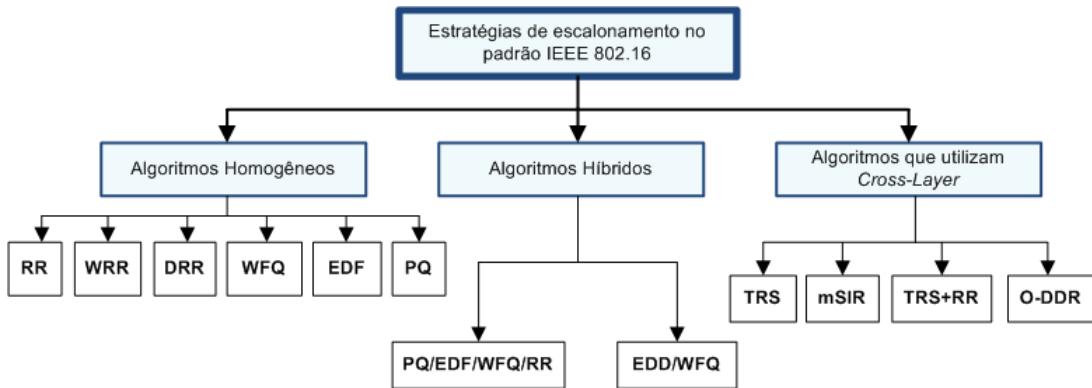


Figura 3.4: Hierarquia dos algoritmos de escalonamento para o padrão IEEE 802.16 [6].

3.3.1.1 Estratégias de Escalonamento Homogêneo

As estratégias de escalonamento homogêneo empregam uma única política ou disciplina para servir os pacotes independentemente do tipo de tráfego. Essas estratégias incluem os algoritmos desenvolvidos especificamente para o escalonamento em redes cabeadas mas que podem ser adaptados para o padrão IEEE 802.16.

O primeiro e mais simples desses algoritmos é o RR (*Round Robin*), ou escalonador cíclico como também é conhecido. Nesse tipo de escalonador, os pacotes armazenados nas filas são atendidos numa espécie de rodízio, um por vez. Desse modo, os recursos do canal são distribuídos igualmente entre as filas. Essa técnica é adequada para estações que possuem o mesmo tipo de tráfego. No entanto, apesar de algumas pesquisas utilizarem esse mecanismo para a comparação de resultados, essas condições não são geralmente aplicáveis para o padrão IEEE 802.16, principalmente em razão da existência de classes de serviço que precisam de tratamento diferenciado, o que inviabiliza o uso direto desse algoritmo.

Um melhoramento da técnica RR pode ser realizado estabelecendo-se prioridades diferentes às filas através de pesos estáticos. Essa técnica é conhecida como WRR (*Weighted Round Robin*) e foi originalmente proposta para redes ATM. Nessa estratégia, a fila que possuir pacotes referentes à classe de serviço com maior prioridade poderá ser atendida por mais tempo (determinado pelo peso) e assim garantir um nível diferenciado de serviço. Desta maneira, o uso do escalonador WRR pode ser adequado para o padrão IEEE 802.16 porque diferentes valores de pesos podem ser designados a cada uma das filas e assim fornecer a quantidade de recursos necessária para atender os requisitos de atraso e vazão de cada classe. Segundo [31], os pesos para cada uma das filas podem ser calculados de acordo com a taxa mínima reservada (MRTR - *Minimum Reserved Traffic Rate*) de cada classe: $w_i = mrtr_i / \sum_{j=1}^n mrtr_j$. A Figura 3.5 mostra o resultado de uma rodada de escalonamento usando as disciplinas RR e WRR. Como pode-se observar, o mecanismo WRR possibilita o escalonamento de uma quantidade maior de pacotes da fila de maior prioridade (Fila C, 3 pacotes) do que a disciplina RR (Fila C, 2 pacotes).

Tanto o RR quanto o WRR atende as filas ciclicamente independentemente do tamanho do pacote que está na cabeça da fila (o próximo a ser atendido). Essa maneira de atender as filas não é adequada para redes que possuem tráfego cujos pacotes variam muito em tamanho, pois as filas que armazenarem pacotes com comprimento médio menor tenderão a ter uma vazão final menor em relação as filas que tiverem pacotes com tamanho médio maior. Para resolver esse problema, um escalonador chamado DRR (*Deficit Round Robin*) pode ser utilizado. Nesse escalonador, uma porção fixa Q_i juntamente com um contador DC_i são atribuídos a cada fila, sendo $Q_i = f/nq$, segundo [32], e $DC_i = Q_i$ onde f = comprimento do *frame*, nq = quantidade de filas e $0 \leq i \leq nq$. As filas também são

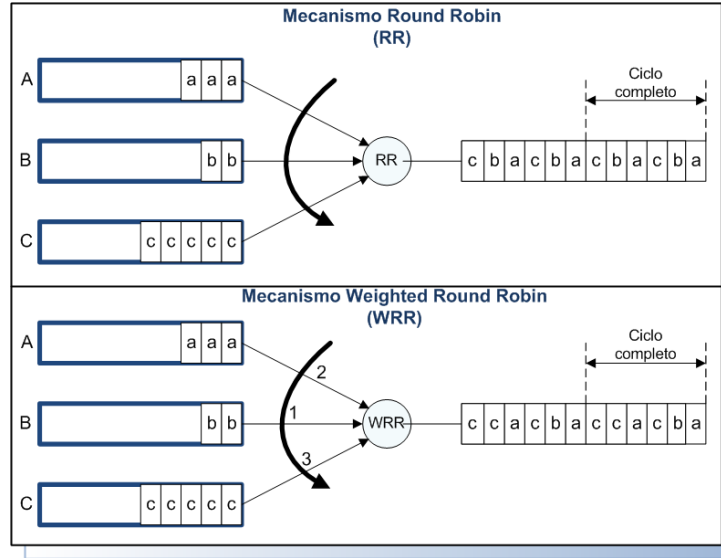


Figura 3.5: Comparação entre o resultado de um ciclo de escalonamento utilizando os mecanismos RR e WRR (adaptado de [6]).

atendidas ciclicamente, entretanto, um pacote só é transmitido se seu comprimento for menor ou igual à DC_i . Nesse caso, após a fila ser atendida, o contador DC_i é decrementado pelo comprimento do pacote transmitido. No entanto, se o comprimento do pacote for superior à DC_i , a fila não é atendida (passa-se para a próxima fila) e o contador DC_i é incrementado por Q_i . Esse processo se repete para todas as filas e só termina quando acabarem os pacotes ($DC_i = 0$) ou o frame estiver totalmente preenchido. Desde que o escalonador tenha acesso direto às filas, como é o caso do BS-DL e SS-UL no padrão IEEE 802.16, a disciplina DRR pode ser empregada facilmente. Entretanto, a desvantagem do DRR é que o tamanho máximo do pacote está limitado ao comprimento do frame (f).

Uma estratégia que também considera o tamanho do pacote e que vem sendo bastante utilizada em propostas de escalonamento para o padrão IEEE 802.16 é a disciplina WFQ (*Weighted Fair Queuing*). Da mesma forma que no WRR, a disciplina WFQ também define pesos estáticos para cada uma das filas e a política de atendimento é cíclica. No entanto, cada pacote que chegar ao *buffer* será marcado com uma estampa de tempo inicial e final, que são calculadas com base no peso da SS, no tamanho do pacote e na capacidade do canal *uplink*. O escalonamento é realizado de acordo com essas estampas de tempo e a política de atendimento às filas adotada evita o *bandwidth starvation* das classes de menor prioridade. No entanto, esse mecanismo não garante os requisitos de atraso máximo para as aplicações de tempo real.

Os escalonadores do tipo WRR, DRR e WFQ naturalmente podem ser empregados em redes IEEE 802.16, porém somente para classes de serviço de não tempo real (nrtPS e BE), pois não consideram o requisito de atraso máximo limitado. Uma disciplina mais adequada para aplicações de tempo real (rtPS e ertPS) é a chamada EDF (*Earliest Deadline First*). Esse algoritmo assume um *deadline* para cada pacote e, juntamente com o requisito de atraso máximo da classe correspondente a esse pacote, será dada maior prioridade àqueles pacotes que precisam chegar mais rapidamente ao destino. A desvantagem dessa disciplina é que ela geralmente é empregada sobre uma única fila e requer o uso de algoritmos de ordenamento, que agregará maior complexidade ao escalonamento.

Uma variante da disciplina EDF é conhecida na literatura como EDD (*Earliest Due Date*). A diferença básica é que, ao invés de se utilizar os valores absolutos de *deadlines*, o EDD atribui um prazo de validade a cada pacote. Nesse algoritmo, um pacote ao qual é atribuído um prazo mais próximo do instante de chegada sofre um atraso menor na fila de espera do que outro ao qual tenha sido atribuído um prazo mais distante. Entretanto, dependendo da carga, pode não ser possível servir todos os pacotes antes de se atingir os respectivos prazos que lhes foram atribuídos [33].

No sentido de prover garantias estritas de QoS para as aplicações de tempo real, as quais não são fornecidas na íntegra pelas disciplinas WRR, DRR, WFQ e EDF, o padrão IEEE 802.16 possibilita a adoção de uma política de escalonamento rigorosa, conhecida na literatura pelo termo “Filas de Prioridades” (PQ - *Priority Queuing*). Nessa disciplina, os pacotes são servidos ciclicamente seguindo a ordem de prioridade decrescente (alta, média e depois baixa prioridade). Em cada fila, a política usualmente adotada é a FIFO (*First in, First out*) e a fila com prioridade i só é atendida quando as filas com prioridade maior que i estiverem vazias. Apesar de garantir requisitos estritos de atraso e vazão para as classes de serviço UGS e rtPS, a disciplina PQ pode causar *bandwidth starvation* nas filas de menor prioridade, pois se não houverem políticas de policiamento de tráfego nas SSs com conexões de tempo real (alta prioridade), as filas das conexões de baixa prioridade nunca serão atendidas.

A Tabela 3.2 traz um resumo com as vantagens e desvantagens de cada disciplina homogênea discutida nessa seção.

Tabela 3.2: Comparação entre as disciplinas de escalonamento homogêneas.

Disciplina	Vantagens	Desvantagens
RR	- Baixa complexidade e de fácil implementação computacional. - Não ocorre <i>bandwidth starvation</i>, pois todas as classes tem a mesma prioridade.	- Ineficiente para tráfegos heterogêneos. - Não provê QoS para as aplicações. - Inadequada para redes com tráfego cujos pacotes variam muito em tamanho.
WRR	- Idem RR. - Possibilita tratamento diferenciado para as filas. - Garante vazão mínima.	Inadequada para redes com tráfego cujos pacotes variam muito em tamanho.
DRR	- Baixa complexidade computacional ($O(1)$) [34]. - Provê justiça [34].	- Impõe tamanho máximo ao pacote [32]. - Inadequada para classes de tráfego com requisitos estritos de atraso e vazão.
WFQ	- Justiça entre aplicações com um mesmo tipo de tráfego (ideal para a classe nrtPS). - Não provoca <i>bandwidth starvation</i> [35]. - Provê garantias de vazão para todos os fluxos [35].	- Não é escalável [35]. - Não apresenta garantias de atraso [35].
EDF	Atende o requisito de atraso para classes de tempo real.	Requer algoritmos de ordenamento (complexidade computacional de no mínimo $O(n * \log(n))$).
PQ	- Baixa complexidade computacional em relação a outras disciplinas [36]. - Diferenciação do tráfego de acordo com níveis de prioridades [36].	Provoca <i>bandwidth starvation</i> para fluxos de baixa prioridade.

3.3.1.2 Estratégias de Escalonamento Híbridas

Conforme pode ser observado na Tabela 3.2, a utilização de uma disciplina de escalonamento isolada para o padrão IEEE 802.16 pode não ser adequada, em razão de cada classe de serviço possuir requisitos de QoS diferentes. Sendo assim, existem várias propostas de escalonamento disponíveis na literatura que consideram uma combinação das disciplinas homogêneas mencionadas na Seção 3.3.1.1 no sentido de aproveitar o que cada uma tem de melhor.

No trabalho apresentado por H. Wang *et al.* [37][38], por exemplo, adota-se a política EDF para escalonar o tráfego rtPS, a disciplina WFQ para o tráfego nrtPS, a RR para classe BE e a disciplina PQ para todos os tráfegos respeitando os níveis de prioridade UGS > rtPS > nrtPS > BE. Essa estratégia possibilita atender adequadamente cada classe de serviço, considerando os requisitos de atraso máximo limitado para conexões UGS e rtPS (através da disciplina EDF), garantia de largura de banda mínima para conexões nrtPS (através da disciplina WFQ) e distribuição justa dos recursos para conexões BEs (através da disciplina RR).

Uma outra estratégia híbrida é apresentada por K. Vinay *et al.* [39]. Nesse mecanismo, um escalonador do tipo EDD é usado para serviços de tempo real enquanto o escalonador WFQ é usado para serviços de não tempo real. O mecanismo EDD prioriza as SSs com *deadline* próximos a expirar, e assim garante o requisito de atraso máximo para as aplicações de tempo real, enquanto que o escalonador WFQ provê a vazão mínima necessária para um bom funcionamento das aplicações não sensíveis a atraso. A latência no mecanismo WFQ (D_{wfq}) é calculada através da equação 3.1, sendo n o número de fluxos de serviço, w_i o peso referente à fila i e R a taxa de transmissão do enlace:

$$D_{wfq} = \sum_{i=1}^n w_i / (R * w_i) \quad (3.1)$$

Um mecanismo semelhante ao descrito em [39] é proposto por D. Tarchi *et al.* [40]. A diferença básica é que a classe de serviço nrtPS possui prioridade maior do que a classe BE. A avaliação de desempenho em ambos mostra que a estratégia híbrida (EDD/WFQ) reduz o atraso médio para as aplicações de tempo real quando comparado à disciplina EDD isolada.

A Tabela 3.3 traz uma análise comparativa em termos de vantagens e desvantagens das estratégias de escalonamento híbridas discutidas nesta seção.

Tabela 3.3: Comparação entre as estratégias de escalonamento híbridas.

Disciplina	Ref.	Vantagens	Desvantagens
EDF/WFQ/RR	[37] [38]	• Garantia de atraso limitado para aplicações tempo real. • Justiça.	• Maior complexidade.
EDD/WFQ	[39]	• Redução do atraso quando comparado à disciplina EDD isolada.	• Maior complexidade. • Fluxos BE tem mesma prioridade do que nrtPS.
EDF/WFQ	[40]	• Fluxos nrtPS possuem prioridade maior do que fluxos BE.	• Poucos recursos destinados a fluxos BE. • Perdas de <i>bytes</i> devido à política WFQ adotada.

3.3.1.3 Escalonamento com Abordagem *Cross-Layer*

A última classe de algoritmos para escalonamento de pacotes considerada nesse trabalho é baseada na abordagem *cross-layer*, na qual os algoritmos utilizam informações da camada física para tomar decisão (em relação à ordem de transmissão e alocação de recursos) na camada MAC ou vice-versa.

O primeiro e mais simples escalonador dessa natureza é conhecido na literatura como mSIR (*Maximum Signal-to-interference*) e foi proposto exclusivamente para redes sem fio [6]. No caso do padrão IEEE 802.16, esse algoritmo aloca recursos de rádio para as SSs que possuem uma alta relação sinal/interferência (SIR - *Signal-to-interference*). Dessa maneira, ele oferece uma alta eficiência espectral. Porém, a desvantagem desse mecanismo é que as SSs que possuírem uma relação SIR sempre baixa nunca serão servidas.

C. Ball *et al.* [41] apresentam um escalonador *cross-layer* exclusivo para o padrão IEEE 802.16 chamado TRS (*Temporary Removal Scheduling*). Esse escalonador funciona da seguinte maneira: A cada período de tempo ajustável T_r , o algoritmo faz uma varredura na fila de pacotes prontos para serem escalonados e remove aqueles cujas estações apresentam condições de potência de sinal baixa. Quando o valor T_r expirar, as condições do canal são novamente verificadas e se permanecerem ruins, os pacotes permanecem na fila por mais um período T_r . Esse processo se repete até que as condições do canal melhorem ou atinja um valor limite L . Esse mecanismo evita que um pacote seja transmitido em um

canal cujas condições de potência são ruins, o que resultaria em uma vazão final baixa. No pior caso, o mecanismo TRS adiciona um atraso de $L * T_r$ aos pacotes.

O mecanismo TRS pode ser combinado com a estratégia RR e assim fornecer um desempenho melhor, como a estratégia apresentada por A. Belghith [6]. Considerando que existam n pacotes na fila do escalonamento e que desses, k pacotes sejam removidos temporariamente da fila pelo mecanismo TRS, esse escalonador combinado, chamado de TRS/RR, aloca uma porção fixa de $1/(n - k)$ da capacidade do canal para cada estação.

No trabalho apresentado por H. K. Rath [42], uma extensão do mecanismo de escalonamento DRR é proposta para o tráfego *uplink* utilizando a abordagem *cross-layer* (O-DDR - *Opportunistic Deficit Round Robin*). Nesse mecanismo, a BS periodicamente consulta as SSs (a cada k frames) e determina um conjunto C de estações aptas a transmitirem dados e solicitação de largura de banda. Cada estação pertencente ao conjunto C transmitirá somente quando suas filas não estiverem vazias e o nível da relação sinal/interferência mais o nível da relação sinal/ruído (*SINR* - *Signal-to-interference-plus-noise Ratio*) estiverem acima de um *threshold*, chamado de $SINR_{th}$. O conjunto C muda dinamicamente e essas mudanças dependem do estado do enlace sem fio para cada SS. A vantagem desse mecanismo é que ele se mostra adequado para canais sujeitos a interferências e ruídos, situações típicas em redes IEEE 802.16. O problema é que esse mecanismo é mais complexo em função de exigir técnicas sofisticadas para acessar o canal e obter os valores da SINR corretos.

A Tabela 3.4 mostra resumidamente os principais mecanismos de escalonamento que utilizam a abordagem *cross-layer* discutidos nessa seção, bem como uma análise comparativa breve.

3.3.2 Controle de Admissão de Conexões

Juntamente com os mecanismos de escalonamento de pacotes, o CAC é um outro mecanismo de grande importância para a garantia de QoS no padrão IEEE 802.16. Ele consiste em um procedimento que visa decidir se uma nova solicitação de conexão deve ou não ser aceita pela rede, levando-se em conta os recursos já alocados para as conexões existentes e os requisitos de QoS para a nova conexão. Essa decisão não pode ser tomada de qualquer maneira, pois o CAC deve ser o mais eficiente possível. Caso a política do controle aceite um número excessivo de conexões, por exemplo, o sistema não terá como

Tabela 3.4: Comparação entre as estratégias de escalonamento que utilizam a abordagem *Cross-Layer*.

Disciplina	Ref.	Vantagens	Desvantagens
mSIR	[6]	• Se adapta a ambientes com canais sujeito à ruídos e interferência.	• <i>Starvation</i> das SSs que apresentaram níveis de SNR baixo.
TRS	[41]	• Não provoca <i>starvation</i> das SSs com níveis de SNR baixo.	• Atraso adicional para pacotes cujas SSs apresentam níveis de SNR baixo, não sendo adequado para aplicações de tempo real.
TRS/RR	[6]	• Provê justiça.	• Baixa utilização do canal.
O-DDR	[42]	• Adequado para canais sujeitos à interferências e ruídos.	• Implementação complexa.

garantir a QoS das conexões existentes. Entretanto, caso a política admita um número muito pequeno de conexões pode ocorrer um desperdício dos recursos da rede.

Nos últimos anos, mecanismos para CAC tem sido amplamente estudados em redes cabeadas como uma ferramenta essencial para controle de congestionamento e provisão de QoS. Entretanto, o problema do CAC em redes sem fio apresenta maior complexidade devido às particularidades existentes no meio físico que incluem interferência de acesso múltiplo, requisições de *handoff/handover* (quando uma SS móvel muda para uma outra área de cobertura) e limitação de largura de banda [9].

Sendo assim, existem várias situações desafiadoras em que o CAC deve ser empregado e que gera impacto direto na provisão de QoS. Em [9] (Tabela 3.5) identifica-se duas dessas situações:

- Controle da qualidade do sinal:** Nessa situação, o CAC é essencial para garantir a qualidade do sinal em relação ao limite de interferência nas redes sem fio, nesse caso, as redes IEEE 802.16. Por exemplo, essas redes têm uma capacidade limitada na carga de tráfego imposta na rede, de maneira que quanto maior for essa carga, maior será a deteriorização da qualidade do sinal para os usuários em termos de nível de SIR e/ou SINR. Deste modo, esquemas de CAC admitem usuários somente se a rede pode manter uma qualidade de sinal mínima para os usuários já admitidos, incluindo a nova conexão. Nesse caso, o critério de admissão pode ser o número de

usuários.

- **Controle da probabilidade de queda da conexão:** Considerando que a queda de uma conexão ativa é usualmente mais incômodo que o bloqueio de uma nova conexão, um mecanismo de CAC pode ser empregado em uma rede móvel IEEE 802.16 para controlar a probabilidade de falha no *handoff*. Isso pode ser feito reservando-se alguns recursos exclusivamente para admitir o tráfego móvel. O critério de admissão também pode ser o número de usuários ou uma estimativa da probabilidade de falha no *handoff*.

Tabela 3.5: Critérios para a classificação dos mecanismos de CAC [9].

Critério	Opção	Vantagens	Desvantagens
Centralização	Central	Mais eficiente	Complexo, impraticável
	Distribuído	Simples, Praticável	Menos eficiente
Escala	Global	Mais eficiente	Mais complexo
	Semi-local	Eficiente, menos complexo	Complexidade média
	Local	Mais simples	Menos eficiente
Otimização	Ótimo	Mais eficiente	Difícil de encontrar. Baseado em simplificações.
	Subótimo	Mais realístico. Escalável	Menos eficiente
Mobilidade	Fixo	Não necessita considerar operações de <i>handover</i>	Necessita de um modelo de ocupação da célula
	Móvel	Não necessita de um modelo de ocupação da célula	Precisa considerar operações de <i>handover</i>
Tempo de decisão	Pró-ativo	Rápido	Mais propenso a erros. Necessita de informações <i>a priori</i>
	Reativo	Menos propenso a erros. Não necessita de informações <i>a priori</i> .	Lento

Além disso, políticas de CAC também são empregadas em situações onde critérios subjetivos devem ser avaliados. Por exemplo, existe um consenso entre pesquisadores e

operadores de rede de que serviços de voz devem ter prioridade mais alta que o serviço de dados, o que é uma realidade. Entretanto, usuários de serviços de dados podem querer pagar mais para ter um serviço diferenciado e o CAC precisa levar isso em consideração no processo de admissão.

Como se pode perceber, o contexto e aplicabilidade dos esquemas de CAC são imensos, o que faz surgirem inúmeras formas de classificá-los. No estudo apresentado em [9], os esquemas de CAC são classificados por categorias de acordo com essa variedade de situações. Essa classificação também baseia-se nas informações necessárias para o CAC tomar a decisão, conforme pode ser visualizado na Tabela 3.5.

Nesta dissertação, optou-se por dividir as estratégias de CAC em dois grandes grupos, de acordo com o critério tempo de decisão apresentado na Tabela 3.5. O primeiro deles refere-se ao grupo de mecanismos de CAC reativos (CAC baseado em degradação ou “roubo” de largura de banda) nos quais os algoritmos de CAC precisam atuar sobre as conexões de dados mesmo depois de já as terem aceitado. Nesse grupo, o CAC está intimamente ligado à política de escalonamento adotada. O segundo grupo refere-se às políticas de CAC pró-ativas (CAC baseados em taxa mínima, em maximização do lucro e em heurísticas). Nesse grupo, os algoritmos de CAC processam os dados e realizam todos os cálculos antes da tomada de decisão. Uma vez aceita pela rede, o algoritmo não atua mais sobre a conexão. A Figura 3.6 ilustra esses dois grupos.

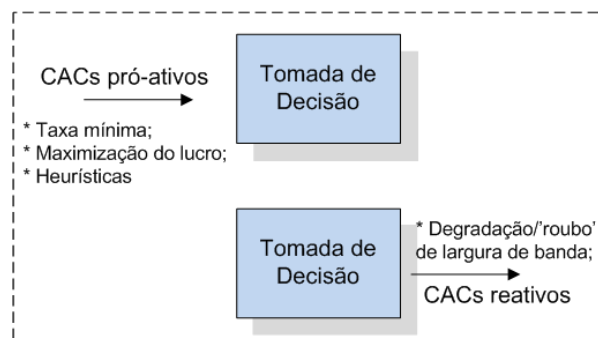


Figura 3.6: Classificação dos mecanismos de CAC.

3.3.2.1 Mecanismos de CAC Pró-ativos

O mecanismo de CAC pró-ativo mais simples e de fácil implementação que existe é o baseado em taxa mínima. Nesse esquema, as únicas informações necessárias para o CAC decidir se uma solicitação de conexão deve ou não ser aceita pela rede é o requisito de

taxa mínima da nova conexão (C_r), a quantidade de conexões já admitidas na rede (n), com suas respectivas taxas reservadas (TR_{ij}) e a capacidade do canal (C). A decisão de admitir ou não a nova conexão deve ser realizada levando-se em conta a restrição apresentada na equação 3.2.

$$C_r + \sum_{i=1}^n \sum_{j=1}^n TR_{ij} \leq C \quad (3.2)$$

K. F. Tsang *et al.* [43] propõem um mecanismo de CAC baseado em taxa mínima que segmenta a largura de banda para cada classe de serviço, incluindo uma parcela para operações de *handover*, e utiliza essas faixas para acomodar as conexões admitidas reservando-se a taxa mínima se elas pertencerem ao grupo [rtPS, nrtPS ou BE]. A vantagem desse mecanismo é a sua simplicidade de implementação na BS. Entretanto, como os *thresholds* definidos para segmentar a largura de banda são fixos, o algoritmo proposto pode desperdiçar largura de banda se houverem muitas conexões de uma mesma classe em detrimento a outras de menor prioridade, pois essas conexões não poderão utilizar a faixa de largura de banda reservada para as outras classes.

Chaudhry e Guha [44] utilizam um esquema de CAC em que a largura de banda também é segmentada entre as conexões de dados e operações de *handover*, no entanto, os *thresholds* são alterados dinamicamente e isso implica numa melhor utilização da largura de banda. Nesse esquema, cada conexão é caracterizada por dois parâmetros $[b_{min}, b_{max}]$, onde b_{min} reflete o requisito de taxa mínima e b_{max} o requisito de taxa máxima. Para as conexões de tempo real (UGS e rtPS), o mecanismo proposto reserva a largura de banda b_{max} durante todo o processo de comunicação e, se essas conexões forem de não tempo real, são reservados b_{min} para cada conexão.

J. F. Borin [8] propõe um esquema de CAC pró-ativo que garante recursos suficientes para o escalonador atender o requisito de taxa mínima das conexões UGS, rtPS, ertPS e nrtPS, bem como a latência máxima das conexões tempo real que não ultrapassa essa taxa. A principal contribuição desse trabalho no que diz respeito à política de CAC adotada é o cálculo do parâmetro TR_{ij} utilizado na equação 3.2. Nessa proposta, ao invés de considerar apenas a taxa mínima, o parâmetro TR_{ij} inclui também uma estimativa da sobrecarga gerada pelo tipo de serviço associado à conexão. Essa abordagem possibilita que se faça um cálculo mais preciso em relação à largura de banda alocada para as conexões.

Além do mecanismo de CAC baseado em taxa mínima, os mecanismos de CAC pró-ativos são compostos também pelas estratégias baseadas em otimização. O princípio de funcionamento dos mecanismos dessa natureza consiste em aplicar uma dada heurística sobre um determinado conjunto de requisições potenciais a fim de obter o maior ganho possível para a rede. A Figura 3.6 mostra como isso funciona.

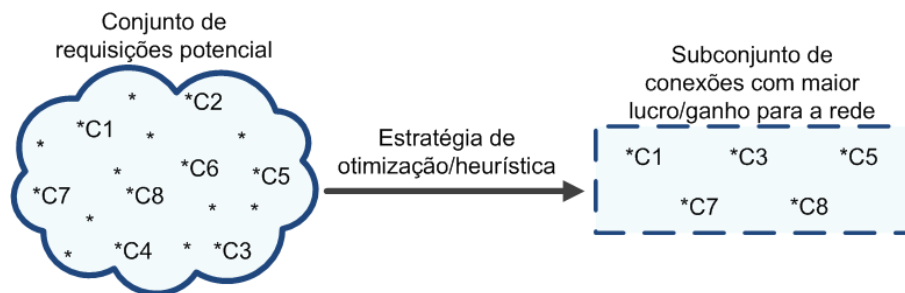


Figura 3.7: Mecanismo de CAC baseado em otimização.

Uma estratégia simples de CAC baseada em otimização é apresentado por J. F. Borin [45]. A política adotada nessa proposta visa encontrar uma combinação ótima de conexões que, a longo prazo, seja capaz de fornecer o maior lucro possível respeitando a capacidade da rede. Para isso, o problema é modelado como um clássico problema da mochila [46] no qual a capacidade da rede é o tamanho da mochila e a quantidade de banda requisitada por cada conexão representa o tamanho de cada objeto. Como todo o processamento é realizado antes das conexões serem aceitas pela rede, o tempo computacional gasto nesse processo não interfere na garantia de latência máxima para as aplicações de tempo real.

Bo Rong *et al.* [47] propõem um outro esquema de CAC pró-ativo o qual também é modelado como um problema de otimização sobre uma determinada função objetivo. Os autores consideram a função objetivo como sendo a razão SB/B , onde SB é a largura de banda estimada que as conexões podem experimentar se forem aceitas pela rede e B denota a largura de banda de acesso ao meio físico. Os resultados obtidos por meio de modelagem e simulação demonstraram que a proposta de um esquema de CAC ótimo é capaz de atender os requisitos tanto das SSs (atraso e vazão) como também dos provedores de serviço (maximização do lucro).

S. Khemiri *et al.* [48] utilizaram um esquema de CAC que também se baseia em otimização mas que está fundamentado em dois objetivos básicos: (1) Garantir QoS para as classes de serviço definidas no padrão IEEE 802.16 e (2) Maximizar o ganho médio da rede em relação aos recursos disponíveis no meio físico. O esquema proposto é

modelado através de um processo de decisão de *Markov*. Uma vez que processos de decisão de *Markov* requerem frequentemente a utilização de programação linear, que apresenta um alto custo computacional, os autores propõem uma solução baseada na técnica de relaxação lagrangeana [49].

A Tabela 3.6 mostra uma análise comparativa entre as propostas de CAC's pró-ativas, entre elas as baseadas em taxa mínima e otimização, discutidas nessa seção.

3.3.2.2 Mecanismos de CAC Reativos

Os mecanismos de CAC reativos compreendem as estratégias baseadas em degradação ou empréstimo de largura de banda. Nesses mecanismos, existe uma forte ligação entre a política de CAC adotada e o módulo de alocação de largura de banda na BS, o qual é um componente essencial para o funcionamento do escalonador. A Figura 3.8 mostra o comportamento do CAC pertencente a esse grupo. Nota-se que uma vez aceita pela rede, a conexão estará sujeita a ter seus parâmetros de QoS alterados, dependendo das requisições para estabelecimento de conexões que chegarem.

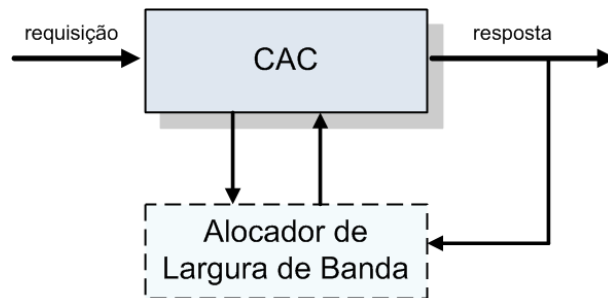


Figura 3.8: Funcionamento de um esquema de CAC reativo.

T. C. Tsai *et al.* [50][51] apresentam um esquema de CAC que utiliza o princípio de degradação de largura de banda. Ao receber a solicitação para o estabelecimento de uma conexão, o algoritmo de CAC inicialmente verifica se a largura de banda disponível é maior do que o requisito de taxa mínima dessa conexão. Em caso afirmativo, o algoritmo aceita a conexão. Caso contrário, o algoritmo verifica se as conexões das classes de menor prioridade estão utilizando largura de banda acima de seus respectivos *thresholds* (T_{nrtPS} e T_{BE}). Se estiverem, o algoritmo degradará a largura de banda alocada para cada conexão (r_i) por um fator c , fazendo $r_i * c$, tal que $0 < c < 1$. Se a largura de banda alocada para esses fluxos já estiverem no limiar, o mesmo procedimento é feito para as classes de maior prioridade.

Tabela 3.6: Comparação entre as propostas de CAC pró-ativos.

Estratégia	Ref.	Vantagens	Desvantagens
Taxa Mínima	[43]	• Simplicidade de implementação na BS.	• Causa desperdício de largura de banda. • Garante apenas a taxa mínima.
	[44]	• Utilização eficiente da largura de banda. • Baixa probabilidade de bloqueio.	• Os resultados são satisfatórios apenas quando a taxa de chegada de conexões está entre 0.2 e 0.28. • Garante apenas a taxa mínima.
	[45]	• Implementação computacional simples. • Considera sobrecarga gerada pelos <i>overheads</i>, produzindo resultados mais realísticos.	• Garante apenas a taxa mínima reservada para os fluxos rtPS e nrtPS mais a estimativa de sobrecarga.
Otimização	[45]	• Encontra a solução ótima (conjunto de conexões que produzem o maior ganho possível para a rede).	• Implementação complexa (problema da mochila).
	[47]	• Abordagem multi-objetiva. • A proposta inclui tanto algoritmos ótimos quanto aproximados.	• Implementação complexa.
	[48]	• Solução ótima, capaz de prover QoS para todas as classes do padrão IEEE 802.16.	• Modelagem analítica por cadeia de <i>Markov</i> complexa. • O problema são decompostos e solucionados usando programação linear (alta complexidade computacional).

H. Wang *et al.* e C. F. Leal [37][38][52] também utilizam o conceito de empréstimo de largura de banda. Nesse esquema, no entanto, assume-se que a largura de banda para cada classe de serviço varia entre $[b_j^{min}, b_j^{max}]$, com $j \in [\text{rtPS}, \text{nrtPS} \text{ e } \text{BE}]$ e b_{UGS} sendo a alocação para a classe UGS. Quando existe banda suficiente disponível no sistema (baixo

número de conexões ativas), as conexões podem transmitir a altas taxas até o limite de b_j^{max} , porém conforme o número de conexões no sistema aumenta, as conexões das classes rtPS e nrtPS podem dispensar certa quantidade de largura de banda com o objetivo de permitir que um número maior de novas conexões seja aceito, maximizando dessa forma a utilização da rede. Diferentemente de [50][51], nessa proposta a degradação é feita em passos, onde δ é a quantidade de largura de banda degradada para cada passo de degradação das conexões das classes rtPS e nrtPS. Considera-se que todas as conexões de uma classe devem manter o mesmo nível de degradação l_j^n , fazendo a largura de banda reservada para cada conexão ser $b_j^{max} - l_j^n * \delta$ que satisfaz $b_j^{max} - l_j^n * b_j^{min} \leq b_j^{min}$. Os resultados obtidos por meio de modelagem e simulação demonstraram que em ambas as propostas a taxa de utilização do canal permaneceu alta e a probabilidade de bloqueio de novas conexões se mostrou baixa.

Finalmente, a Tabela 3.7 resume as propostas de CAC's reativos, ou seja, os baseados em degradação de largura de banda, que foram discutidas nessa seção.

Tabela 3.7: Comparação entre as propostas de CAC's reativas.

Ref.	Vantagens	Desvantagens
[50] [51]	• Utilização eficiente da largura de banda. • Não causa <i>starvation</i>.	• Implementação e modelagem matemática Complexa
[37] [38] [52]	• Utilização eficiente da largura de banda.	• Implementação complexa. • Não provê justiça entre as classes.

3.4 Considerações Finais

Neste capítulo, apresentou-se os principais aspectos existentes no padrão IEEE 802.16 no que diz respeito à provisão de QoS. Foram abordadas as principais características dos fluxos e classes de serviço especificadas pelo padrão, bem como o processo para estabelecimento da conexão ou fluxo de serviço. Os mecanismos para escalonamento de pacotes e CAC, essenciais para a provisão de QoS e não definidos pelo padrão, também foram descritos em detalhes. O escalonamento foi abordado de acordo com as princi-

país propostas existentes na literatura, enquadrando-as em três categorias principais de algoritmos: (1) Algoritmos homogêneos, (2) Algoritmos híbridos e (3) Algoritmos que utilizam a abordagem *cross-layer*. Já os mecanismos de CAC também foram abordados de acordo com o estado da arte, enquadrando cada uma das propostas em dois grandes grupos: (1) Mecanismos de CAC pró-ativos e (2) Mecanismos de CAC reativos. Esses conceitos foram abordados de maneira a fundamentar a proposta para a provisão de QoS que é apresentada no próximo capítulo.

Capítulo 4

Proposta de Um Esquema Para a Provisão de QoS no Padrão IEEE 802.16

4.1 Introdução

O padrão IEEE 802.16 vem de encontro a uma necessidade cada vez mais crescente em fornecer acesso banda larga sem fio para longas distâncias, principalmente em regiões onde a infraestrutura é precária, como, por exemplo, zonas rurais ou áreas suburbanas onde os custos para a implantação de redes cabeadas são onerosos. Além disso, através da especificação das quatro classes de serviços básicas (UGS, rtPS, nrtPS e BE), descritas no Capítulo 3, o padrão é capaz de fornecer suporte a uma ampla variedade de aplicações e ainda prover qualidade de serviço (QoS) aos usuários.

Tendo em vista que os principais mecanismos para a provisão de QoS no padrão IEEE 802.16 estão associados às políticas de escalonamento de pacotes e CAC e os mesmos não são definidos pelo padrão, este capítulo visa descrever a proposta de um esquema básico para a provisão de QoS que consiste em um mecanismo para escalonamento *uplink* na estação base (BS) e uma política de CAC com atraso preditivo.

Dessa maneira, este capítulo estrutura-se da seguinte forma: Na Seção 4.2 é feita a descrição do problema e a motivação para a pesquisa. A Seção 4.3 traz em detalhes a solução proposta sendo composta pela Seção 4.3.1 (mecanismo de escalonamento) e Seção 4.3.2 (mecanismo de CAC). Na Seção 4.4 são apresentados os trabalhos relacionados à

solução proposta e a Seção 4.5 são feitas as considerações finais.

4.2 Descrição do Problema

Conforme foi apresentado no Capítulo 3, o padrão IEEE 802.16 deixa em aberto as questões referentes à provisão de QoS, ou seja, ele não define as políticas que devem ser usadas na implementação dos mecanismos de escalonamento e CAC. Dessa maneira, o padrão possibilita a elaboração de propostas por parte da comunidade científica ao passo que permite aos fabricantes de equipamentos de redes diferenciarem seus produtos, uma vez que existe a liberdade em adotar políticas e mecanismos para escalonamento e CAC próprios. Essa é uma das principais motivações para a elaboração da proposta apresentada nesta dissertação.

Além disso, existe uma forte carência de trabalhos na literatura que exploram algoritmos para escalonamento e CAC trabalhando conjuntamente. Na maioria dos trabalhos, mecanismos para escalonamento são propostos e avaliados considerando que já existem políticas de CAC implementadas, ou seja, a quantidade de conexões e a carga de tráfego imposta na rede estão limitadas à capacidade de transmissão no enlace, o que de certa forma dispensa ou torna implícita a adoção de um mecanismo de CAC. Por outro lado, propostas de CAC também são avaliadas na maioria dos casos sem levar em consideração mecanismos de escalonamento.

4.3 Solução Proposta

Em razão da inexistência de mecanismos obrigatórios para escalonamento e CAC no padrão IEEE 802.16 e a escassez de trabalhos que focam escalonamento e CAC juntos, principalmente em se tratando de CAC dinâmico, neste trabalho propõe-se um algoritmo de escalonamento híbrido para o tráfego *uplink* que atende as quatro principais classes de serviços especificadas no padrão (UGS, rtPS, nrtPS e BE), sendo capaz, portanto, de prover QoS às aplicações/usuários. O mecanismo de CAC deve ser pró-ativo, de forma a utilizar um esquema de predição do atraso na rede para a tomada de decisão. O esquema proposto pode ser visto na Figura 4.1.

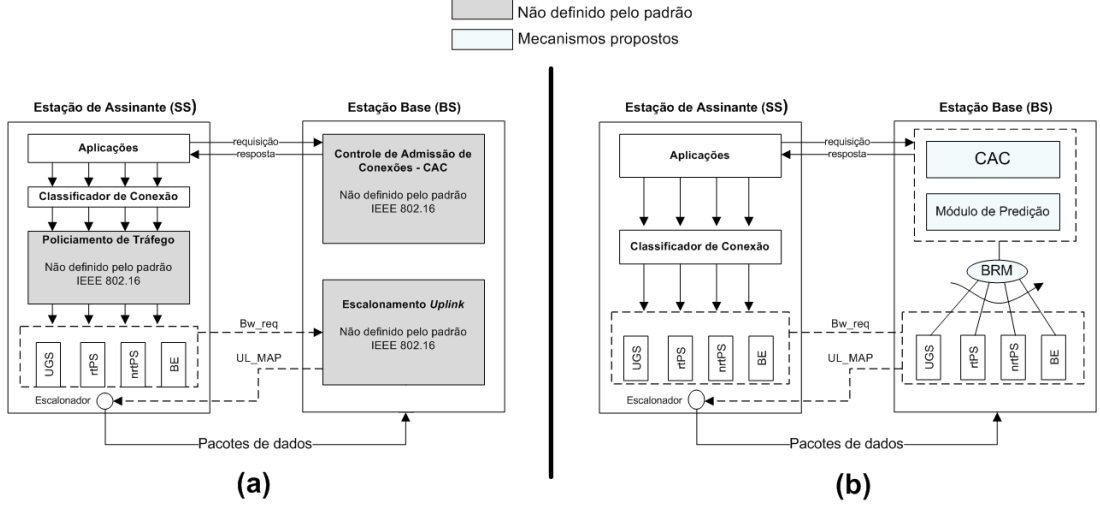


Figura 4.1: (a) Arquitetura original do padrão IEEE 802.16 e (b) Mecanismos propostos.

4.3.1 Mecanismo de Escalonamento *Uplink* Proposto

A estratégia de escalonamento proposta se baseia em um mecanismo de reserva de largura de banda com porções estáticas que visam separar logicamente os fluxos tempo real (UGS e rtPS), fluxos não tempo real com o requisito de largura de banda mínima garantida (nrtPS) e fluxos de melhor esforço (BE). Esse mecanismo de reserva foi criado com o intuito de evitar a “inanição” dos fluxos de baixa prioridade (*bandwidth starvation*), como nrtPS e BE, quando a rede estiver operando sob uma carga de tráfego elevada.

O mecanismo de reserva proposto, denominado BRM (*Bandwidth Reservation Mechanism*), aloca uma parcela fixa de largura de banda para as conexões pertencentes às classes de serviço UGS e rtPS (W bps), a qual é alocada sob demanda para os fluxos dessas classes que chegam ao longo do tempo, obedecendo a ordem de prioridade $UGS > rtPS$. A parcela W é utilizada prioritariamente para fluxos de tempo real, porém, para evitar desperdício de largura de banda, essa porção poderá ser utilizada para fluxos nrtPS e BE desde que não haja fluxos UGS e/ou rtPS presentes na rede ou a largura de banda total solicitada por essas conexões seja menor que W . Uma segunda parcela é reservada para as conexões da classe nrtPS (T bps) e, da mesma forma que para a porção W , T também poderá ser utilizada para fluxos de outras classes se não houverem conexões nrtPS ou se essas conexões não forem suficientes para ocuparem toda a reserva T . Finalmente, uma parcela relativamente pequena é destinada a atender os fluxos BE (R bps), para evitar *bandwidth starvation* nessa classe, a qual também poderá ser utilizada para outros fluxos. A Fig. 4.2 ilustra o esquema de reserva proposto.

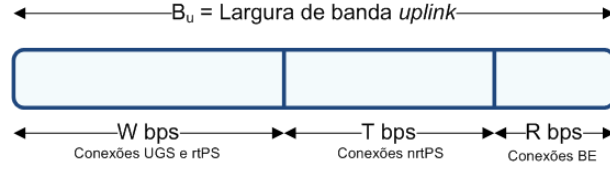


Figura 4.2: Mecanismo de reserva de largura de banda proposto para o escalonamento.

O mecanismo de escalonamento *uplink* proposto define quatro filas na BS, uma para cada classe de serviço, e estas são atendidas de acordo com os níveis de prioridades especificados para cada fluxo de serviço: $UGS > rtPS > nrtPS > BE$. A fila das conexões UGS armazena os *grants* periódicos para o envio dos dados enquanto que as filas rtPS, nrtPS e BE armazenam as requisições de largura de banda provenientes das SSs. A política de atendimento a essas filas é cíclica e se comporta de maneira semelhante a disciplina de escalonamento PQ. No entanto, essas filas serão atendidas de forma preemptiva com base nas reservas W , T e R .

Em cada ciclo de escalonamento, as SSs com conexões UGS serão atendidas em primeiro lugar em razão de requererem largura de banda constante e garantida. Nesse caso, a condição $\sum_{i=1}^{n_{ugs}} b_i \leq W$ (n_{ugs} = número de conexões UGS e b_i = taxa da i -ésima conexão UGS) deve ser respeitada, caso contrário realiza-se uma preempção e a próxima fila é atendida. Na fila de conexões UGS é aplicada a política FIFO. Em seguida, serão atendidas as SSs com conexões rtPS ordenando-as de acordo com o estado de ocupação de seus *buffers* e fornecendo toda a largura de banda solicitada por cada fluxo. Nesse caso, para não haver preempção nessa fila, a seguinte condição deve ser satisfeita: $\sum_{i=1}^{n_{ugs}} b_i + \sum_{j=1}^{n_{rtPS}} b_j^{req} \leq W$, sendo b_j^{req} o valor da requisição rtPS referente a j -ésima SS e n_{rtPS} a quantidade de fluxos rtPS presentes na rede. Em seguida a fila das conexões nrtPS é atendida distribuindo-se equitativamente a largura de banda residual $W + T - (\sum_{i=1}^{n_{ugs}} b_i + \sum_{j=1}^{n_{rtPS}} b_j^{req})$ entre as estações nrtPS presentes na rede. Dessa maneira, será fornecida para cada fluxo nrtPS a porção de largura de banda $\min(b_k^{req}, b_{med})$, onde b_k^{req} é o valor da requisição nrtPS referente à estação k e b_{med} é a largura de banda média resultante da expressão $\sum_{k=1}^{n_{nrtPS}} b_k^{req} / n_{nrtPS}$, sendo n_{nrtPS} a quantidade de estações com conexões nrtPS. Seguindo essa política de atendimento, as conexões nrtPS não sofrerão *bandwidth starvation*, pois no pior caso (situação em que W for totalmente utilizado), a parcela T será destinada exclusivamente aos fluxos nrtPS. Por último, escalonam-se as estações com fluxos BE também distribuindo-se a largura de banda residual $W + T - (\sum_{i=1}^{n_{ugs}} b_i + \sum_{j=1}^{n_{rtPS}} b_j^{req} + \sum_{k=1}^{n_{nrtPS}} \min(b_k^{req}, b_{med}))$ entre

cada fluxo BE presente na rede. No pior caso, ou seja, na situação em que as parcelas W e T estiverem 100% alocadas aos fluxos UGS, rtPS e nrtPS, restará a reserva R que será alocada exclusivamente para os fluxos BE. Dessa maneira, o mecanismo proposto evita o *bandwidth starvation* na classe BE. No entanto, o efeito negativo disso é que a porção R pode ficar subutilizada se não houverem fluxos BE na rede. O algoritmo 4.1 descreve com detalhes como o escalonamento proposto é desempenhado considerando as quatro classes de serviço.

Algoritmo 4.1 EscalonaFluxosUplink

Entrada: N_b = Número total de símbolos OFDM disponíveis para escalonar.

Saída: UL-MAP = Mapa de *uplink* (alocações de largura de banda para cada SS).

- 1: UL-MAP = CriaMensagemUL-MAP()
 - 2: N_b , UL-MAP = Escalona_Fluxos_UGS(N_b , UL-MAP)
 - 3: N_b , UL-MAP = Escalona_Fluxos_RTPS(N_b , UL-MAP)
 - 4: N_b , UL-MAP = Escalona_Fluxos_NRTPS(N_b , UL-MAP)
 - 5: N_b , UL-MAP = Escalona_Fluxos_BE(N_b , UL-MAP)
 - 6: **retorna** UL-MAP;
-

Os detalhes dos algoritmos para cada uma das classes de serviço mostrados no algoritmo 4.1 estão apresentados no Apêndice A. Como cada escalonamento, com exceção da classe de serviço rtPS que requer uma função de ordenação, possui um laço de repetição contendo n_i operações, sendo n_i a quantidade de conexões pertencentes à classe i , a complexidade computacional do algoritmo de escalonamento *uplink* proposto, que influencia fortemente em sua escalabilidade, é dado por: $\max(O(n_{ugs}), O(n_{rtps}^2), O(n_{nrtps}), O(n_{be})) = O(n_{rtps}^2)$, sendo, portanto, uma complexidade computacional relativamente baixa.

4.3.2 Mecanismo de CAC Dinâmico Proposto

O mecanismo de CAC foi proposto para a classe de serviço rtPS e funciona de maneira conjunta à política de escalonamento *uplink*. Nesse mecanismo, a BS ao receber a solicitação para o estabelecimento de uma conexão rtPS, através da mensagem DSA-REQ, realiza uma predição sobre o atraso médio global que essa potencial conexão poderá sofrer no sistema se for admitida. Se o valor predito for menor ou igual a um valor de referência, então o pedido de conexão é aceito. Caso contrário ele será rejeitado. Essa condição é

verificada para cada conexão rtPS que espera entrar na rede. Como o mecanismo de CAC é proposto para os fluxos rtPS, as conexões pertencentes às outras três classes de serviço serão aceitas automaticamente pela rede, entretanto, a quantidade de largura de banda alocada pelo escalonador *uplink* para cada fluxo ao longo do tempo deverá respeitar as parcelas W , T e R . Considerando que o atraso máximo rtPS fica limitado a um dado valor de *threshold* médio, obtém-se também uma vazão média limitada para os fluxos admitidos na rede. O algoritmo 4.2 traz em detalhes as operações realizadas pelo mecanismo de CAC proposto.

Algoritmo 4.2 ControleAdmissãoConexoesRTPS

Entrada: Mensagem de requisição DSA-REQ.

Saída: Mensagem de resposta DSA-RSP.

```

1:  $c \leftarrow \text{DSA-REQ.connection};$ 
2: se ( $c.type = rtPS$ ) então
3:    $A_t \leftarrow$  Busca o atraso predito atual;
4:   se ( $A_t < threshold$ ) então
5:     DSA-RSP.flag  $\leftarrow$  1; //Aceita  $c$ 
6:   senão
7:     DSA-RSP.flag  $\leftarrow$  0; //Rejeita  $c$ 
8:   fim se
9: senão
10:  DSA-RSP.flag  $\leftarrow$  1; //Aceita a conexão  $c$ .
11: fim se
12: retorna DSA-RSP;
```

Ao admitir uma conexão, o algoritmo de CAC proposto não define um *threshold* máximo para a alocação de largura de banda às SSs com conexões rtPS. Ao invés disso, o esquema proposto aloca a quantidade de símbolos OFDM¹ necessária para transmitir a requisição inteira (todos os *bytes* solicitados), de acordo com o esquema de modulação e codificação empregado (*MCS - Modulation Coding Scheme*). Assim, o cálculo da largura de banda alocada para todas as conexões rtPS na rede varia estocasticamente, uma vez que ela dependerá dos valores das requisições presentes na fila rtPS.

¹Símbolo OFDM é a menor unidade de alocação de largura de banda existente no *frame* TDD, considerando a modulação por divisão de frequência ortogonal (OFDM).

Considerando que haja, em um determinado instante de tempo t , no máximo K conexões rtPS que podem ser escalonadas no quadro corrente e que existem N conexões rtPS na rede, a alocação média para cada estação rtPS (b_i), cujos valores das requisições estão armazenados na fila Q_2 ($Q_{i,j}$ é o *data grant* ou requisição de largura de banda j referente à fila i , onde $i \in [1 = UGS, 2 = rtPS, 3 = nrtPS, 4 = BE]$) é calculada de acordo com as equações (4.1), (4.2) e (4.3):

$$b_i = \frac{\sum_{j=1}^K Q_{2,j}}{K} \quad (4.1)$$

sujeito à:

$$K * b_i \leq W - \sum_{j=1}^{n_{ugs}} Q_{1,j} \quad (4.2)$$

$$0 < K \leq N \quad (4.3)$$

O mecanismo de predição do atraso é desempenhado de acordo com a arquitetura mostrada na Figura 4.1 (b). Conforme apresentado na seção anterior, cada requisição R_i que chegar à BS proveniente das classes rtPS, nrtPS e BE, a qual reflete uma estimativa do comprimento das filas nas aplicações, é armazenada em filas separadamente (a fila UGS armazena *grants* a serem enviados periodicamente e não requisições). O módulo de predição desenvolvido acessará o conteúdo da fila rtPS e, a cada intervalo fixo f que corresponde à duração do *frame* TDD, realizará a predição. Quando a BS receber a mensagem DSA-REQ para criar uma nova conexão, o valor predito corrente será utilizado no processo de admissão (linha 3 do algoritmo 4.2).

A complexidade do algoritmo de CAC proposto, no pior caso, resume-se ao tempo de processamento gasto para buscar o atraso predito na rede atual ($O(1)$) na linha 3, tendo em vista que as demais operações são elementares.

Antes de calcular o atraso predito na rede é necessário fazer uma estimativa do estado de ocupação médio dos *buffers* nas SSs que possuem conexões rtPS. Para isso, a equação 4.4 descreve a dinâmica do tamanho médio das filas, em *bytes*, referentes às SSs que não puderam ter suas conexões totalmente escalonadas no quadro TDD corrente (B_t), considerando o instante de tempo t , N conexões e no máximo K conexões sendo escalonadas no *frame* corrente.

$$B_t = \begin{cases} 0, & \text{se } N=K; \\ \sum_{i=1}^{N-K} Q_{2,i} \\ \frac{1}{N-K}, & \text{se } N>K; \end{cases} \quad (4.4)$$

De acordo com a equação 4.4, pode-se ter a seguinte proposição:

Proposição 1. *Quando a rede está levemente congestionada, isto é, $N = K$, B_t é zero. Por outro lado, quanto N tende ao infinito através da adição de novas conexões, ambos B_t e $N - K$ tendem ao infinito ao passo que K tende a 1.*

Demonstração. Seja R_t um conjunto com n requisições de conexões no tempo t ($R_t = [r_1, r_2, \dots, r_n]$) e C a capacidade do canal. Então, tem-se no máximo k estações, assumindo $k = n - 1$, tal que:

$$\sum_{i=1}^k r_i \leq C \text{ e } B_t = \frac{\sum_{i=k+1}^n r_i}{n - k} \quad (4.5)$$

Agora, considerando o instante $t + f$ e $n = n + 1$ (nova conexão), como o conjunto R_{t+f} está ordenado, então deve existir um k' tal que:

$$\sum_{i=k+1}^n r_i + F(i, t + f) + \sum_{j=1}^{k'} r_j \leq C \text{ e} \quad (4.6)$$

$$B_{t+f} = \frac{\sum_{i=k'+1}^n r_i + F(nova - conexao, t + f)}{n - k'} \quad (4.7)$$

onde $F(i, j)$ é o tráfego gerado pela estação i no instante j . Para satisfazer a equação 4.6, como $F(i, j) \neq 0$, necessariamente $k' \leq k$. Supondo que a capacidade do *buffer* (B) seja infinita, $B_{t+f} \geq B_t$, pois não haverá transbordamento. Portanto, quando n tende ao infinito ($n \rightarrow \infty$) intuitivamente percebe-se que k' tenderá a 1 ($k' \rightarrow 1$), k' denotando a quantidade de estações servidas no *frame* corrente e ambos $(n - k')$ e B_{t+nf} tenderão por consequência ao infinito, ou seja, $(n - k) \rightarrow \infty$ e $B_{t+nf} \rightarrow \infty$.

□

Pela Proposição 1, nota-se que é necessário controlar a variável B_t quando a rede já está operando sob uma carga de tráfego elevada e o mecanismo de CAC proposto atua exatamente no sentido de estabelecer esse controle.

Através da estimação do comprimento médio das filas (B_t), calculado por meio da equação 4.4, a predição do atraso na rede (A_t) pode então ser calculada segundo a equação 4.8:

$$A_t = \frac{8 * B_t}{r} * T_s * f \quad (4.8)$$

onde r é a eficiência do MCS empregado (*bits*/símbolo), T_s é o tempo do símbolo OFDM (tempo/símbolo) e f é a duração do quadro TDD.

A equação 4.8 mostra que o atraso predito é diretamente proporcional a B_t , uma vez que r , T_s e f são constantes.

4.4 Trabalhos Relacionados

Existe uma ampla variedade de trabalhos na literatura que exploram mecanismos para escalonamento e CAC, muitos dos quais já abordados no Capítulo 3. Nesta Seção, no entanto, são discutidos e analisados os trabalhos relacionados diretamente ao mecanismo proposto nesta dissertação no que diz respeito a escalonamento e CAC.

M. Castrucci *et al.* [53] apresentam uma proposta de algoritmo de CAC que utiliza predição de atraso para fluxos rtPS. Diferentemente do mecanismo de CAC apresentado neste capítulo, que utiliza unicamente o critério de atraso (através de um algoritmo preditivo) para o processo de admissão, o algoritmo proposto em [53] é desempenhado considerando dois critérios para a admissão: (1) largura de banda mínima e (2) atraso máximo limitado. Quando a BS recebe a mensagem DSA-REQ solicitando a criação de um fluxo de serviço verifica-se inicialmente se o requisito de largura de banda mínima desse fluxo pode ser atendido. Se não puder, a conexão é imediatamente rejeitada. Caso contrário, como a garantia de largura de banda mínima não é suficiente para garantir os requisitos de QoS das aplicações de tempo real, o algoritmo realiza uma segunda verificação para determinar se a rede é capaz de garantir o atraso máximo limitado requerido. A proposta apresentada neste capítulo, entretanto, não necessita verificar esses dois critérios, pois, garantindo-se o atraso máximo limitado, a largura de banda é implicitamente garantida através do mecanismo de reserva de largura de banda. Tanto o mecanismo de CAC proposto neste capítulo quanto o algoritmo proposto em [53] utilizam um esquema de predição de atraso. No entanto, neste trabalho a predição é realizada considerando o

comprimento das filas virtuais nas SSs, que reflete o comportamento real da rede enquanto que em [53] a predição é desempenhada de acordo com algoritmos numéricos, mais precisamente por meio das equações de Yule-Walker e recursão de Levinson-Durban [54][55], que produzem resultados aproximados baseados em medidas.

F. Hou *et al.* [56] desenvolveram um esquema de CAC baseado em reservas que utiliza duas variáveis de controle, C_u (porção dedicada ao tráfego UGS) e C_r (porção exclusiva para UGS e rtPS), para segmentar a largura de banda, sendo a largura de banda residual $C - C_u - C_r$ destinada ao tráfego nrtPS e BE. Nesta dissertação também propõe-se um esquema de reserva de largura de banda, no entanto essa reserva é utilizada no escalonamento *uplink* e não diretamente no CAC. Outra diferença básica é que em [56] o mecanismo de CAC é reativo, ou seja, é baseado em degradação de largura de banda enquanto que o CAC proposto nesse trabalho se baseia em uma política pró-ativa. Uma característica comum entre esses trabalhos, no entanto, é que ambos utilizam uma parcela de largura de banda exclusiva para as classes nrtPS e BE. No entanto, no mecanismo proposto, a classe BE, mesmo possuindo menor prioridade, não sofre *bandwidth starvation* se houverem muitas conexões nrtPS.

H. Wang *et al.* [38] propõem um esquema de CAC que reserva largura de banda exclusiva para o tráfego UGS (U) no sentido de diminuir a probabilidade de bloqueio de novas conexões ao passo que permite aos demais fluxos de serviço compartilharem a largura de banda residual ($B - U$), sendo B a largura de banda total. Para evitar que os fluxos de serviço de baixa prioridade utilizem largura de banda além da quantidade que necessitam, os autores definem um *threshold* máximo para a classe nrtPS e utilizam o princípio de degradação de largura de banda para aumentar a taxa de aceite de novas conexões. A proposta apresentada nesta dissertação, no entanto, não define *thresholds* máximos para a alocação de largura de banda aos fluxos de serviço de baixa prioridade. Ao invés disso, o escalonador proposto aloca toda a largura de banda solicitada mesmo para as classes de menor prioridade e o mecanismo de reserva proposto se encarregará de evitar a monopolização ou superutilização da largura de banda por parte das aplicações mal comportadas.

K. Wongthavarawat e A. Ganz [5] apresentam uma arquitetura de provisão de QoS semelhante à proposta neste trabalho, que também consiste em um algoritmo de escalonamento *uplink* híbrido juntamente com uma política de CAC. No entanto, as disciplinas

homogêneas utilizadas no escalonamento dos fluxos rtPS (EDF) e nrtPS (WFQ) são diferentes das utilizadas neste trabalho, as quais são baseadas no comprimento das filas e na disciplina baseada na distribuição justa de recursos, respectivamente. Além disso, a política adotada em [5] para escalonar todos os fluxos é a PQ, enquanto que no mecanismo proposto utiliza-se o método PQ preemptivo com base nas reservas W , T e R . Em relação ao mecanismo de CAC, em [5] emprega-se um *token bucket* para controlar o atraso juntamente com algoritmo de policiamento de tráfego nas SSs. A proposta de CAC apresentada neste trabalho, por outro lado, se fundamenta unicamente na predição do atraso na rede, calculada com base nas informações reais sobre o estado de ocupação dos *buffers* nas SSs por meio do mecanismo de requisição de largura de banda.

Finalmente, um outro trabalho relacionado com essa pesquisa e que apresenta características semelhantes ao mecanismo proposto é apresentado por J. F. Borin *et al.* [57]. Nesse trabalho, os autores adotam uma política de CAC pró-ativa que também trabalha em conjunto com um mecanismo de escalonamento *uplink* na BS. Diferentemente da proposta apresentada nesta dissertação, entretanto, a qual permite aos fluxos de serviços transmitirem dados além do requisito mínimo, a política de CAC em [57] é capaz de garantir somente a taxa mínima reservada para cada fluxo de serviço juntamente com uma estimativa sobre o *overhead* gerado pelas mensagens de requisição. Além disso, em [57] não se considera um mecanismo de CAC dinâmico, com entradas periódicas de conexões na rede, e também o canal físico considerado não emprega o esquema de modulação e codificação OFDM.

A Tabela 4.1 traz um resumo sobre os trabalhos relacionados ao mecanismo proposto discutidos nessa Seção.

4.5 Considerações Finais

Neste capítulo, apresentou-se um esquema para a provisão de QoS no padrão IEEE 802.16, a qual consiste em um algoritmo de escalonamento *uplink* na BS bem como um mecanismo de CAC pró-ativo. Foi apresentada, como motivação principal para a solução proposta, o fato do padrão deixar em aberto as questões referentes à provisão de QoS e também a escassez de trabalhos na literatura que focam escalonamento e CAC juntos. O mecanismo de escalonamento proposto é designado para atender as quatro principais

Tabela 4.1: Características dos trabalhos relacionados ao mecanismo proposto.

Trabalho	Escalonamento	CAC	Características
[53]	Não detalha	Pró-ativo	• CAC realizado em dois passos. • Predição de atraso numérica. • CAC reserva apenas a taxa mínima.
[56]	Não detalha	Reativo	• Reserva de largura de banda para CAC • Fatia exclusiva de largura de banda para as classe nrtPS e BE. • Pode provocar <i>bandwidth starvation</i> para a classe BE.
[38]	Não detalha	Reativo	• Definição de um <i>threshold</i> máximo para os fluxos de baixa prioridade. • A proposta envolve modelagem analítica.
[5]	Híbrido	Pró-ativo	• Escalonamento de todas as classes baseado na disciplina <i>PQ</i>. • Mecanismo de CAC utiliza <i>token bucket</i> e policiamento.
[57]	Homogêneo	Pró-ativo	• Escalonamento baseado na disciplina <i>PQ</i>. • CAC garante somente a taxa mínima para cada conexão, mais uma estimativa sobre o <i>overhead</i>.

classes de serviços especificadas no padrão IEEE 802.16 (UGS, rtPS, nrtPS e BE) e está fundamentada na disciplina PQ com preempção baseada em reservas de largura de banda estáticas. Já a política de CAC se baseia em um esquema de predição do atraso na rede e utiliza informações sobre o estado de ocupação dos *buffers* nas SSs. Constatou-se que ambos os algoritmos apresentam complexidade computacional baixa e são de fácil implementação na BS. Por fim, foram apresentados os trabalhos relacionados de maior relevância que serviram de base para a solução proposta.

Capítulo 5

Avaliação dos Mecanismos de Escalonamento e CAC Propostos

5.1 Introdução

Neste capítulo, avalia-se por meio de modelagem e simulação o mecanismo para a provisão de QoS proposto no Capítulo 4. Para isso, tanto o algoritmo de escalonamento quanto o de CAC são avaliados de acordo com as quatro principais classes de serviço especificadas pelo padrão, considerando cenários de redes tipicamente urbanos.

A Seção 5.2 descreve as ferramentas e os ambientes de rede utilizados nas simulações, a Seção 5.3 apresenta uma análise dos resultados obtidos e a Seção 5.3 traz as considerações finais deste Capítulo.

5.2 Modelagem e Simulação

5.2.1 Ferramentas de Simulação

A análise de desempenho dos mecanismos de escalonamento e CAC propostos no Capítulo 4 foi realizada através da ferramenta de simulação de redes NS-2 (*Network Simulator - 2*) [58]. O NS-2 é um simulador de eventos discretos escrito nas linguagens C++ e oTCL que suporta a arquitetura TCP/IP e permite ainda a inclusão de módulos que simulam protocolos de redes mais específicos como, por exemplo, o padrão IEEE 802.16, IEEE 802.20 e redes LTE (*Long Term Evolution*).

Foi acrescentado ao NS-2 o módulo para a simulação de redes WiMAX desenvolvido pelo NIST (*NIST - National Institute of Standards and Technology*) [59]. Esse módulo implementa as camadas MAC e física e provê suporte para a inclusão de escalonadores UGS, rtPS, nrtPS e BE e também simula a transmissão de dados no enlace físico por meio de um esquema de modulação e codificação.

Para avaliar o mecanismo de CAC proposto, foi necessário incorporar ao NS-2 a possibilidade de adicionar conexões ou fluxos de serviços na rede dinamicamente no tempo, uma vez que as conexões de dados e de controle eram criadas somente no início da simulação, quando as SSs se registravam na rede.

Além disso, em razão dos cenários de simulação considerarem a entrada periódica de novas conexões na rede, foi desenvolvido um aplicativo para analisar o arquivo de *trace* gerado nas simulações, sendo este aplicativo uma opção mais flexível no tratamento dos dados em relação aos scripts AWK (*AWK - Aho, Weinberger and Kernighan*). O aplicativo extrai do arquivo de *trace* informações como o tempo em que o evento ocorreu, a identificação da estação que gerou ou recebeu os dados e o tamanho juntamente com a identificação do pacote, por exemplo, e armazena essas informações em um banco de dados. Isso possibilita um rápido processamento por meio de consultas SQL (*Structured Query Language*) e facilita a geração de gráficos de vazão e atraso em relação a intervalos de tempo configuráveis.

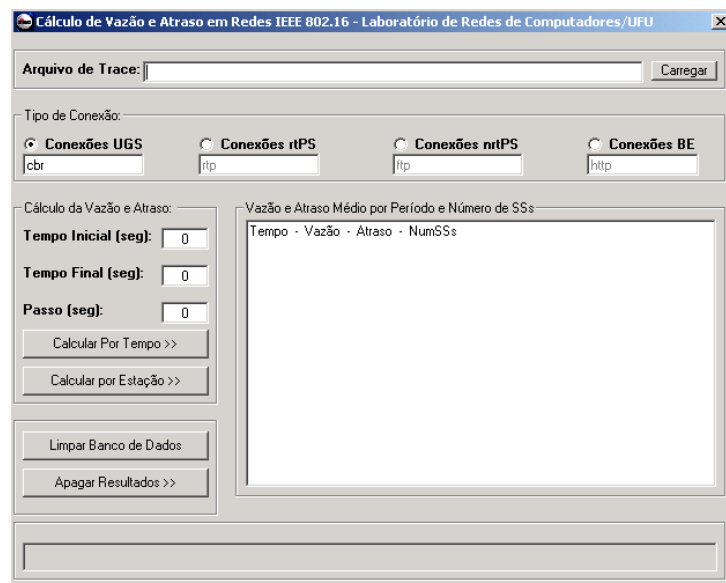


Figura 5.1: Aplicativo desenvolvido para analisar o trace gerado pelo NS-2.

5.2.2 Ambiente e Parâmetros de Simulação

Os cenários considerados nas simulações envolveram uma BS e várias SSs entrando na rede sucessivamente, em intervalos de tempo pré-fixados. A distância máxima permitida entre uma SS e a BS é de 500 metros, o que possibilita a utilização do esquema de modulação e codificação OFDM-64QAM3/4, que apresenta um alto ganho em vazão [60]. A Figura 5.2 ilustra o cenário considerado nas simulações.

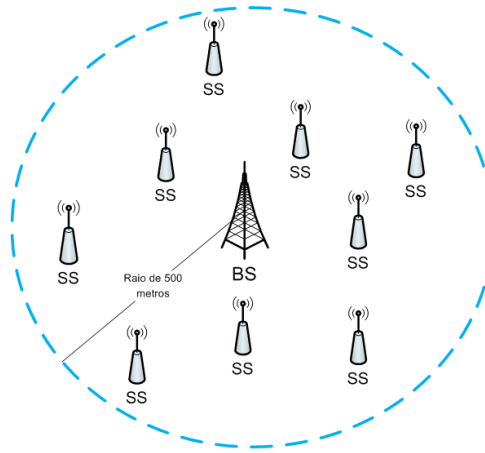


Figura 5.2: Cenário de rede considerado na simulação.

As simulações realizadas consideraram uma conexão por estação e o modo GPSS foi utilizado na concessão de largura de banda. Os principais parâmetros de simulação referentes às camadas MAC e física são apresentados na Tabela 5.1. Esses parâmetros foram escolhidos em razão de serem utilizados na maioria dos trabalhos na literatura.

O tráfego UGS foi modelado como uma fonte CBR com pacotes de 100 *bytes* gerados em períodos de 20 milisegundos. Para o tráfego rtPS, considerou-se uma fonte que gera pacotes em intervalos fixos de 10 milisegundos e pacotes variando uniformemente entre 200 e 1000 *bytes*, produzindo uma taxa média de 480 Kbps. Já para o tráfego nrtPS, utilizou-se uma fonte FTP com taxa mínima de 160 Kbps e máxima de 800 Kbps. Para o tráfego de melhor esforço, utilizou-se uma fonte que gera rajadas em intervalos fixos produzindo uma taxa média de 480 Kbps.

5.3 Resultados Experimentais

Os resultados obtidos por experimentos de simulação nesta seção referem-se a cinco execuções para cada cenário individual, com o objetivo de produzir um intervalo de confi-

Tabela 5.1: Parâmetros de simulação referentes às camadas MAC e física.

Parâmetro	Valor
Frequencia de Operação	3,5 GHz
Largura de Banda	5 MHz
Duplexação	TDD
Antena	Omnidirecional
Modelo de Propagação	Two Ray Ground
Duração do Quadro	20 ms
<i>Cyclic Prefix</i> (CP)	0,25
MCS Empregado	OFDM 64QAM3/4
Taxa de Transmissão <i>Uplink</i>	7,70 Mbps

ança de 95%. Em todos os cenários considerou-se a entrada periódica de estações rtPS na rede em intervalos fixos de 5 segundos. Após se registrar na rede, cada estação começa a transmissão de dados e essa transmissão permanece até o término da simulação. A vazão e o atraso médio são calculados em intervalos periódicos $(t, t + k)$, tomando-se $k = 5$, de acordo com as equações 5.1 e 5.2.

$$vazao_{t,t+k}^z = \frac{\sum_{i=1}^{Tot} comp_pacote_{i,z}^{t,t+k}}{Tot} \quad (5.1)$$

$$atraso_{t,t+k}^z = \frac{\sum_{i=1}^N (\sum_{j=1}^{P_{cont,i}} (Rx_{j,z}^{t,t+k} - Tx_{j,z}^{t,t+k}) / P_{cont,i})}{N} \quad (5.2)$$

onde z = classe de serviço; Tot = total de pacotes recebidos no intervalo $[t, t + k]$; $comp_pacote_{i,z}^{t,t+k}$ = comprimento do i -ésimo pacote recebido no intervalo $[t, t + k]$; $P_{cont,i}$ = quantidade de pacotes recebidos pela estação i ; $Rx_{j,z}^{t,t+k}$ e $Tx_{j,z}^{t,t+k}$ = tempo de recepção e transmissão do j -ésimo pacote, respectivamente; k = intervalo de amostra (5 segundos) e N = quantidade total de estações na rede no intervalo $[t, t + k]$.

5.3.1 Avaliação do Escalonador Proposto para rtPS

Nesta seção analisa-se o desempenho do algoritmo de escalonamento proposto considerando a classe de serviço rtPS. Neste experimento, as SSs com conexões rtPS começam a se registrar na rede a partir do instante 15 segundos, tendo no máximo 20 estações, e a

proporção $W:T:R$ considerada foi de 0,93:0,05:0,02 com tráfego de *background* referente às classes nrtPS e BE em torno de 1 Mbps. Esses valores foram escolhidos em razão de serem adequados para um cenário onde somente fluxos rtPS são avaliados (nesse caso W deve ter percentual significativo). As SSs começam a transmitir dados assim que o procedimento de entrada na rede finaliza e continuam a transmissão até o término da simulação, no instante de tempo 130 segundos. A intenção de fazer as SSs permanecerem na rede até o fim da simulação é que, dessa forma, a adição periódica de conexões, de alguma forma, irá interferir nas conexões já existentes. Considerou-se como métrica de avaliação para este cenário o atraso médio, que é calculado também em intervalos fixos de 5 segundos.

Considerou-se como métrica de comparação para este cenário um escalonador rtPS baseado na disciplina FIFO com distribuição equitativa de recursos. Nesse escalonador, a ordem de atendimento da fila rtPS respeita o instante de chegada das requisições, diferentemente do mecanismo proposto que prioriza as estações cujos comprimentos das filas são maiores. A justificativa para a escolha da disciplina FIFO é sua simplicidade de implementação.

Conforme pode ser observado pelo gráfico apresentado na Figura 5.3, o algoritmo proposto apresenta um desempenho superior ao escalonador FIFO em termos de atraso médio, quando a rede está saturada, isto é, quando a utilização da largura de banda reservada W está próxima do valor pré-fixado (93%). Essa situação acontece a partir do instante de tempo 80 segundos. Isso ocorre devido ao fato de que, nessa situação, as conexões rtPS já não podem mais ser totalmente escalonadas no mesmo quadro TDD, o que provoca um aumento nas filas das SSs que não puderam ter suas conexões escalonadas naquele instante, pois no mínimo serão escalonadas no quadro seguinte (vide Proposição 1 apresentada no Capítulo 4). A justificativa para o desempenho superior do algoritmo proposto é que ele prioriza as SSs cujas filas são maiores, o que não acontece na disciplina FIFO, e o principal fator que contribui para o atraso é exatamente o acúmulo de pacotes nos *buffers* das SSs. Quando a rede está levemente congestionada (do instante 15 segundos até 75 segundos), ou seja, quando todas as conexões podem ser escalonadas no quadro TDD corrente, não ocorre acúmulo de pacotes nas filas das SSs, e, conseqüentemente, o desempenho dos dois algoritmos são semelhantes.

Analisando-se a vazão média calculada ao longo do tempo, por meio do gráfico apre-

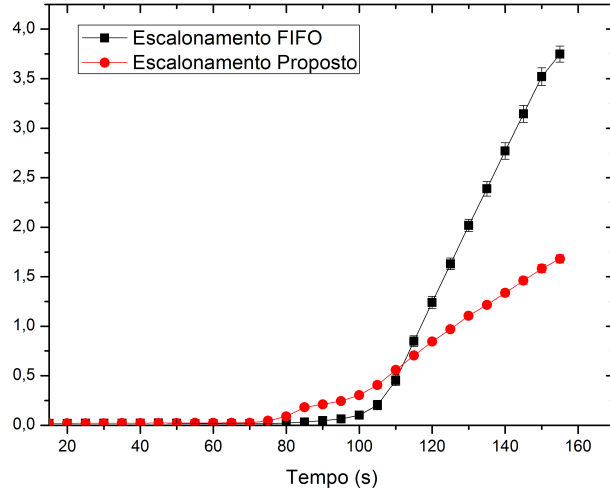


Figura 5.3: Desempenho do escalonador rtPS proposto em termos de atraso médio.

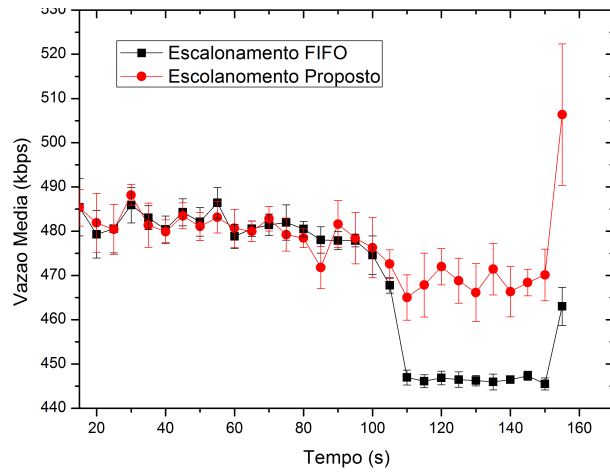


Figura 5.4: Desempenho do escalonador rtPS proposto em termos de vazão média.

sentado na Figura 5.4, percebe-se também um desempenho superior do algoritmo de escalonamento proposto em relação à política FIFO. A justificativa para tal comportamento é justamente o fato de, a partir do instante de tempo 80 segundos, a rede não ser capaz de escalonar todos os fluxos rtPS no mesmo quadro TDD, o qual provoca um atraso médio substancial das conexões rtPS presentes na rede, que por sua vez, resulta em baixa vazão final.

A Tabela 5.2 traz um resumo sobre os principais eventos ocorridos ao longo da simulação.

Tabela 5.2: Eventos ocorridos na simulação em função do tempo.

Instante (seg)	Evento
0-10	* <i>Initial ranging</i> e procedimentos de entrada na rede
10-15	* Entra uma estação nrtPS e uma estação BE na rede produzindo um tráfego de <i>background</i> em torno de 1 <i>Mbps</i>
15-115	* Entra um estação rtPS na rede a cada 5 segundos
115-160	* Não há entrada de novas estações na rede

5.3.2 Avaliação do Escalonador Proposto Considerando as Classes UGS e rtPS

Neste experimento, procurou-se avaliar o desempenho do escalonador proposto em relação aos fluxos de serviço de tempo real UGS e rtPS. Considerou-se também a proporção $W:T:R$ de 0,93:0,05:0,02, pois só são analisadas conexões tempo real. Para tal, foram adicionadas 8 estações UGS na rede no instante 15 segundos e esse número permaneceu constante ao longo do tempo. Periodicamente, em intervalos de 5 segundos, foram adicionadas SSs com conexões rtPS a fim de verificar a influência dessas estações no atraso das conexões UGS já presentes na rede. Assim como na avaliação anterior, o número máximo de estações rtPS foi fixado em 20. Conforme pode ser observado no gráfico da Figura 5.5, a adição de conexões rtPS na rede não interferiu no atraso das conexões UGS. Isso ocorre porque, como as estações UGS têm recursos garantidos, somente quando há largura de banda suficiente na rede é que as conexões das outras classes são servidas. Como conexões rtPS são adicionadas na rede periodicamente, a utilização da reserva W aumenta a cada instante. Neste experimento, quando a utilização da largura de banda reservada W aproxima-se de 0,93, ou seja, aproxima-se do patamar fixado *a priori*, as conexões rtPS não puderam mais ser escalonadas no quadro TDD corrente, o que provocou um aumento das filas, e consequentemente do atraso médio.

Em relação à vazão média para os fluxos UGS e rtPS, percebe-se também, pelo gráfico apresentado na Figura 5.6, que o mecanismo proposto de fato prioriza os fluxos UGS e, mesmo quando a porção W está totalmente em uso, as estações UGS que entraram na rede primeiro não sofreram com a entrada periódica de estações rtPS.

A Tabela 5.3 mostra os principais eventos ocorridos ao longo da simulação.

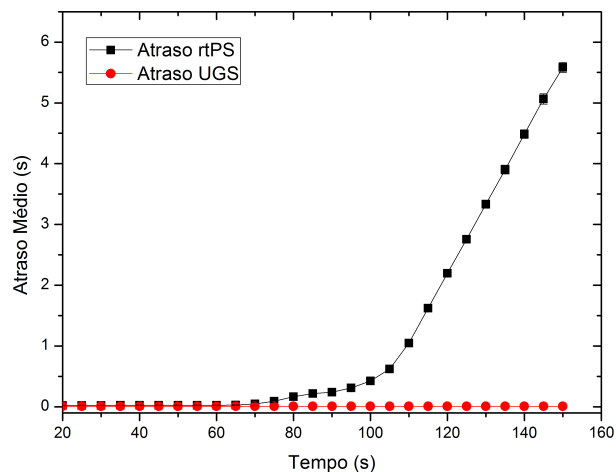


Figura 5.5: Desempenho do escalonador proposto para fluxos tempo real em termos de atraso médio.

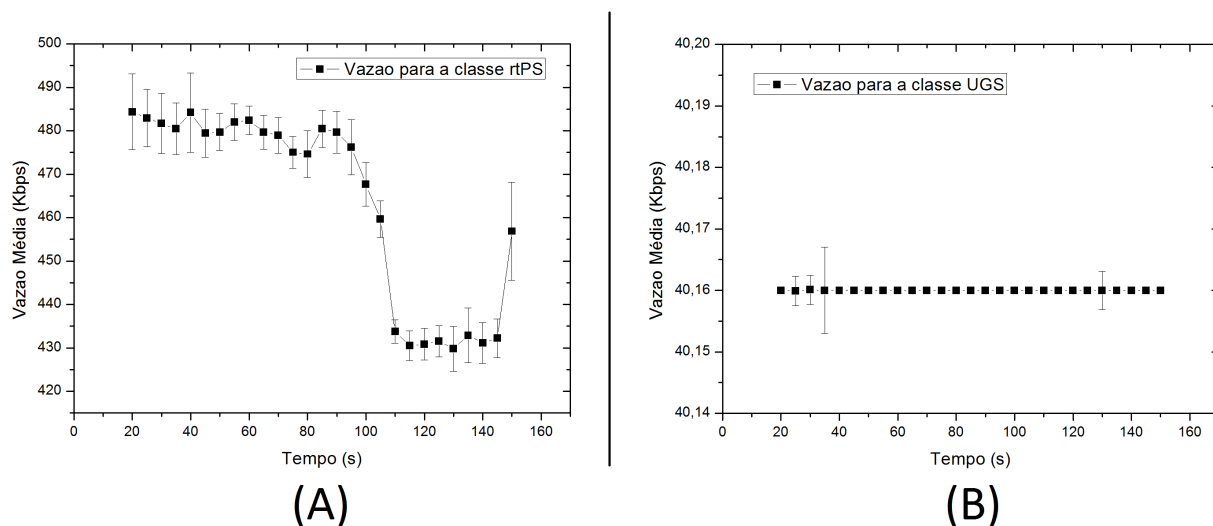


Figura 5.6: Desempenho do escalonador proposto em termos de vazão média para fluxos rtPS (A) e fluxos UGS (B).

Tabela 5.3: Eventos ocorridos na simulação em função do tempo.

Instante (seg)	Evento
0-10	* <i>Initial ranging</i> e procedimentos de entrada na rede
10-15	* 8 estações UGS são adicionadas a rede
15-115	* 1 estação rtPS é adicionada a rede a cada 5 segundos
115-160	* Não há entrada de novas estações na rede

5.3.3 Avaliação do Escalonador Proposto para Fluxos nrtPS e BE

O objetivo deste experimento é verificar a influência da entrada de novas conexões rtPS na rede em relação aos fluxos não tempo real além de verificar o comportamento do atraso e vazão dos fluxos rtPS entrantes. Para isso, foram adicionadas 5 estações nrtPS na rede entre os instantes 15 e 20 segundos (uma nova conexão a cada segundo) e 5 estações BE entre os instantes 20 e 25 segundos (também uma nova conexão a cada segundo) e, a partir do instante 25 segundos, foram adicionadas periodicamente estações rtPS na rede até um limite de 15 estações. A proporção $W:T:R$ foi agora fixada em 0,64:0,32:0,04, uma vez que pretende-se avaliar os fluxos nrtPS e BE (nesse caso T deve ter uma proporção maior) e a taxa de tráfego para cada conexão UGS foi de 32 Kbps. Analisando o atraso e vazão média para as conexões rtPS ao longo do tempo, por meio do gráfico da Figura 5.7, observa-se quatro situações distintas. A primeira delas, chamada de situação 1, acontece entre os instantes 15 e 75 segundos onde tanto a vazão quanto o atraso permanecem praticamente constantes. A segunda, chamada de situação 2, pode ser identificada entre os instantes 75 e 95 segundos, onde nota-se um aumento exponencial do atraso bem como uma diminuição também exponencial da vazão média. Uma terceira situação, chamada de situação 3, pode ser identificada entre os instantes 95 e 130 segundos, onde ocorre um aumento linear do atraso médio enquanto que a vazão média obtida na rede permanece constante. Por fim, uma quarta situação é identificada entre os instantes 130 e 160 segundos, chamado aqui de situação 4, onde ocorre o contrário, ou seja, o atraso médio diminui enquanto que a vazão aumenta. A explicação para a ocorrência dessas situações é a seguinte. A situação 1 acontece em razão de no intervalo em que ela ocorre a ocupação da reserva W está abaixo de 100%. Já a situação 2 ocorre devido à proposição 1 apresentada no Capítulo 4 a qual mostra que o atraso médio cresce muito rapidamente à medida em que novas conexões chegam à rede. A terceira situação ocorre em razão de no intervalo [95 - 130 segundos] não haver mais conexões entrando na rede e a última situação ocorre pelo fato de que o *backlog* (total de requisições não atendidas) total acumulado até então é consumido periodicamente ao longo do tempo, diminuindo assim a ocupação média da reserva W .

Analisando-se agora o comportamento da vazão média para os fluxos nrtPS e BE percebe-se, pelo gráfico apresentado na Figura 5.8, que entre os instantes 15 e 25 segundos ocorre um aumento das vazões nrtPS e BE. Isso ocorre em razão de nesse intervalo as

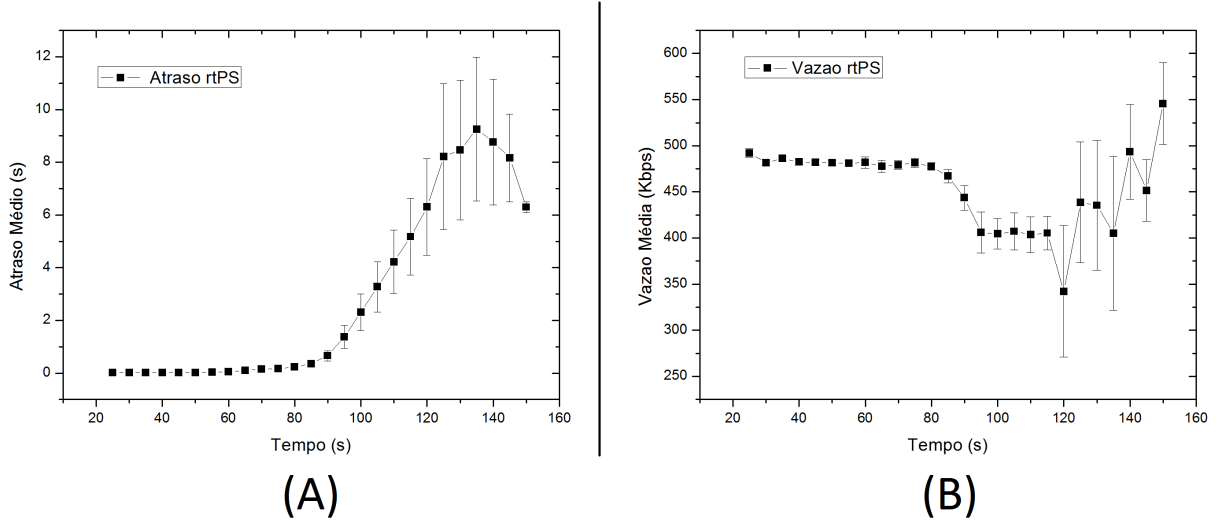


Figura 5.7: Desempenho do escalonador proposto para fluxos rtPS em termos de atraso (A) e vazão (B), considerando estações nrtPS e BE presentes na rede.

estações dessas classes estarem entrando na rede e não haver conexões UGS e/ou rtPS presentes na rede. A partir do momento em que as conexões rtPS começam a chegar e serem admitidas pela rede, do instante 25 segundos em diante, a vazão de ambas as classes nrtPS e BE caem até chegarem ao limite da reserva T e R , respectivamente. Conforme apresentado no Capítulo 4, na Figura 5.8 comprova-se que o mecanismo de escalonamento proposto evita o *bandwidth starvation* para os fluxos de baixa prioridade.

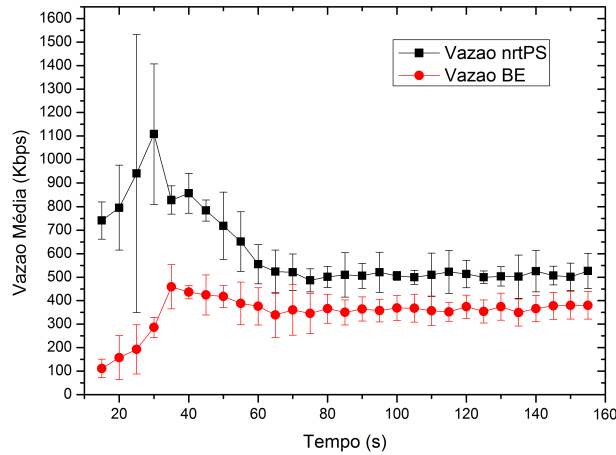


Figura 5.8: Desempenho do escalonador proposto em termos de vazão média para fluxos não tempo real.

A Tabela 5.4 traz a lista dos principais eventos ocorridos ao longo da simulação.

Tabela 5.4: Eventos ocorridos na simulação em função do tempo.

Instante (seg)	Evento
0-10	* <i>Initial ranging</i> e procedimentos de entrada na rede
10-15	* * *
15-20	* 1 estação com conexão nrtPS é adicionada a rede a cada 1 segundo (máximo de 5 estações)
20-25	* 1 estação com conexão BE é adicionada a rede a cada 1 segundo (máximo de 5 estações)
25-95	* 1 estação com conexão rtPS é adicionada a rede a cada 5 segundos (máximo de 15 estações)
95-160	* Não há entrada de novas estações na rede

5.3.4 Avaliação do Escalonador Proposto em Relação ao Mecanismo de Reserva

Neste outro experimento de simulação procurou-se verificar a influência da variação da parcela W no atraso das conexões rtPS, bem como na vazão média dos fluxos nrtPS. Para isso, considerou-se conexões nrtPS entrando na rede a cada 1 segundo a partir do instante 15 até 22 segundos. Foram consideradas também conexões rtPS entrando na rede a cada 5 segundos a partir de 25 segundos, sendo três simulações distintas com um valor de W para cada uma delas. A justificativa para a entrada periódica de fluxos nrtPS na rede entre os instantes 15 e 22 segundos é que isso permite uma análise mais realista em relação ao impacto da adição de fluxos nrtPS sobre a vazão média obtida na rede do que ao levar-se em conta a entrada simultânea dessas conexões. Conforme mostrado na Figura 5.9, o comportamento do atraso médio em cada período variou de acordo com o valor da reserva W , indicando um crescimento mais acentuado da curva de atraso quando se considera valores para W menores. Isso se justifica em razão de que quando se atribui valores menores para W , o mecanismo de escalonamento proposto realiza a preempção mais precocemente em relação ao uso de valores para W maiores. Evidentemente, a partir do momento em que ocorre a preempção, ou seja, a partir do momento em que nem todas as conexões podem mais ser escalonadas para o quadro TDD corrente, o tamanho médio das filas nas SSs aumenta muito rapidamente, conforme a Proposição 1 enunciada

no Capítulo 4. O comportamento atípico observado entre os instantes 85 e 95 segundos (redução da vazão) quando se considera $W = 0.06$ é decorrente de descartes de pacotes.

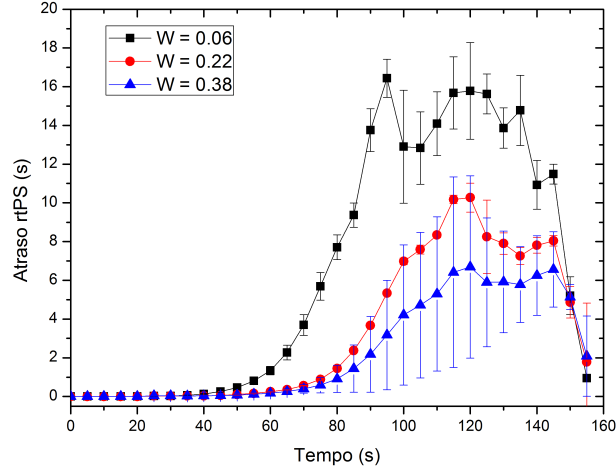


Figura 5.9: Análise da variação da reserva W para fluxos rtPS em termos de atraso médio.

Conforme identificado no gráfico da Figura 5.10, percebe-se que na medida em que o valor de W diminui, apesar disso acarretar em um aumento prematuro no atraso das conexões rtPS, a vazão nrtPS aumenta proporcionalmente. Isso se justifica pelo fato de que, como a porção $W + T$ é constante ao longo do tempo, se a reserva W diminuir por um fator k , ou seja, passar a ser $k * W$ sendo $0 < k \leq 1$, a porção T sofrerá um aumento proporcional de $1 - k$ passando a ser $T + W * (1 - k)$. Isso faz com que, independentemente do valor atribuído à W , desde que $T \neq 0$, não haverá *bandwidth starvation* para os fluxos nrtPS.

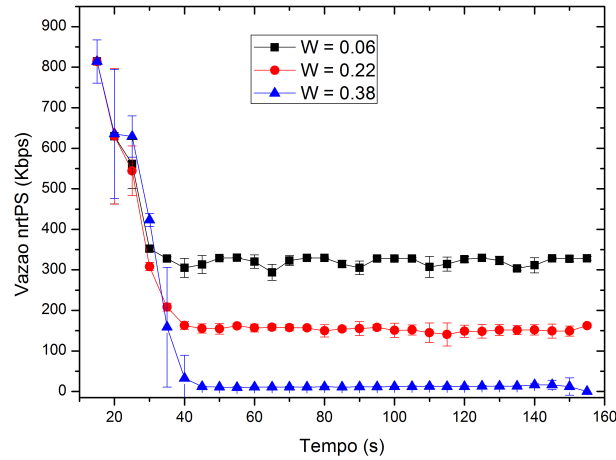


Figura 5.10: Análise da variação da reserva W para fluxos nrtPS em termos de vazão média

A Tabela 5.5 traz a lista dos principais eventos ocorridos nesse experimento.

Tabela 5.5: Eventos ocorridos na simulação em função do tempo.

Instante (seg)	Evento
0-10	* <i>Initial ranging</i> e procedimentos de entrada na rede
10-15	* * *
15-22	* 1 estação com conexão nrtPS é adicionada à rede a cada 1 segundo
22-25	* * *
25-95	* 1 estação com conexão rtPS é adicionada à rede a cada 5 segundos
95-160	* Não há entrada de novas estações na rede

5.3.5 Avaliação do Mecanismo de CAC Proposto para Fluxos rtPS

Procurou-se verificar nesta seção o desempenho do mecanismo de CAC preditivo para a classe de serviço rtPS. Para tal, foram consideradas apenas estações rtPS, entrando na rede em intervalos de 5 segundos, até um limite máximo de 20 estações.

O primeiro teste realizado neste experimento consiste em avaliar a eficiência do mecanismo de predição do atraso proposto, segundo a equação 4.8 apresentada no Capítulo 4. Conforme mostrado no gráfico da Figura 5.11, o atraso predito na rede segue a mesma tendência em relação ao atraso obtido na rede para cada intervalo fixo de 20 ms. Isso demonstra que o mecanismo de CAC proposto é capaz de prever com exatidão o valor do atraso na rede, o qual é fundamental para uma tomada de decisão mais precisa.

Neste experimento, o *threshold* do atraso rtPS utilizado pelo algoritmo foi de 200 ms. O gráfico apresentado na Figura 5.12 mostra que a utilização do esquema de CAC proposto foi capaz de bloquear a entrada de novas conexões na rede a partir do momento em que o atraso médio predito atingiu um patamar máximo. Isso fez com que o atraso médio rtPS fixasse em torno de 200 ms, ou seja, o mecanismo de CAC preditivo evitou que a entrada de novas conexões, quando a rede está operando sob uma carga de tráfego elevada, prejudicasse a QoS das conexões já existentes.

A Tabela 5.6 traz a lista dos principais eventos ocorridos nesse experimento de simulação.

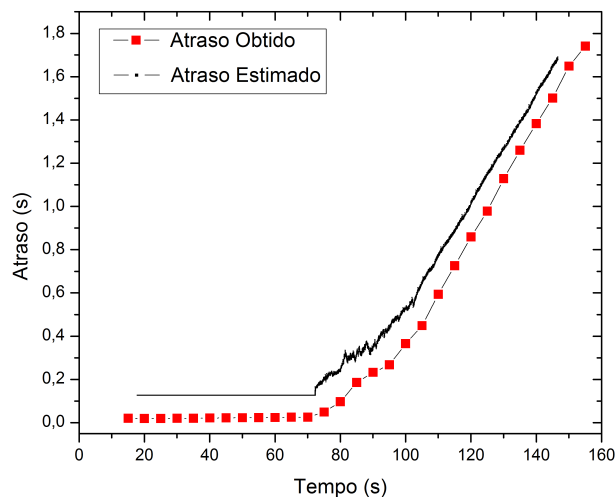


Figura 5.11: Comparação entre o atraso médio predito na rede em relação ao atraso medido.

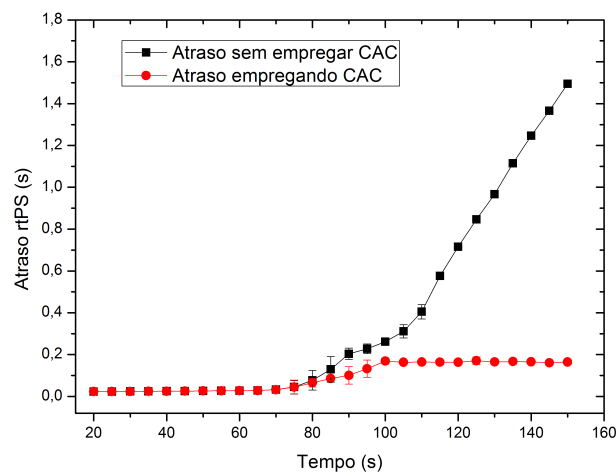


Figura 5.12: Desempenho do CAC preditivo proposto para fluxos rtPS.

Tabela 5.6: Eventos ocorridos na simulação em função do tempo.

Instante (seg)	Evento
0-10	* <i>Initial ranging</i> e procedimentos de entrada na rede
10-15	* * *
15-115	* 1 estação com conexão rtPS é adicionada à rede a cada 5 segundos
115-160	* Não há entrada de novas estações na rede

5.3.6 Avaliação do CAC para Fluxos Não Tempo Real

O objetivo desta avaliação consiste em aplicar o mecanismo de CAC sobre as conexões rtPS no sentido de avaliar seu efeito sobre a vazão das conexões não tempo real (nrtPS e

BE) e atraso das conexões UGS. Para isso, considerou-se 4 estações UGS registrando-se na rede entre 10 e 14 segundos, 4 estações nrtPS registrando-se na rede entre 15 e 19 segundos, 4 estações BE entrando na rede entre 20 e 24 segundos e, a partir do instante 25 segundos e a cada período de 5 segundos, foram adicionadas estações rtPS na rede até um limite máximo de 15 estações.

Pelo gráfico apresentado na Figura 5.13, nota-se que a vazão média das conexões nrtPS e BE segue a mesma tendência de aumento e redução, respectivamente, independentemente de aplicar ou não o mecanismo de CAC preditivo. A justificativa para tal comportamento é que, a partir do momento em que o CAC começa a bloquear a entrada de novas conexões rtPS na rede, a utilização da largura de banda W está próxima de 100% e isso faz com que, a partir desse instante, a reserva T fique alocada exclusivamente para os fluxos nrtPS e BE, fazendo com que a política de CAC não cause impacto direto nas vazões nrtPS e BE.

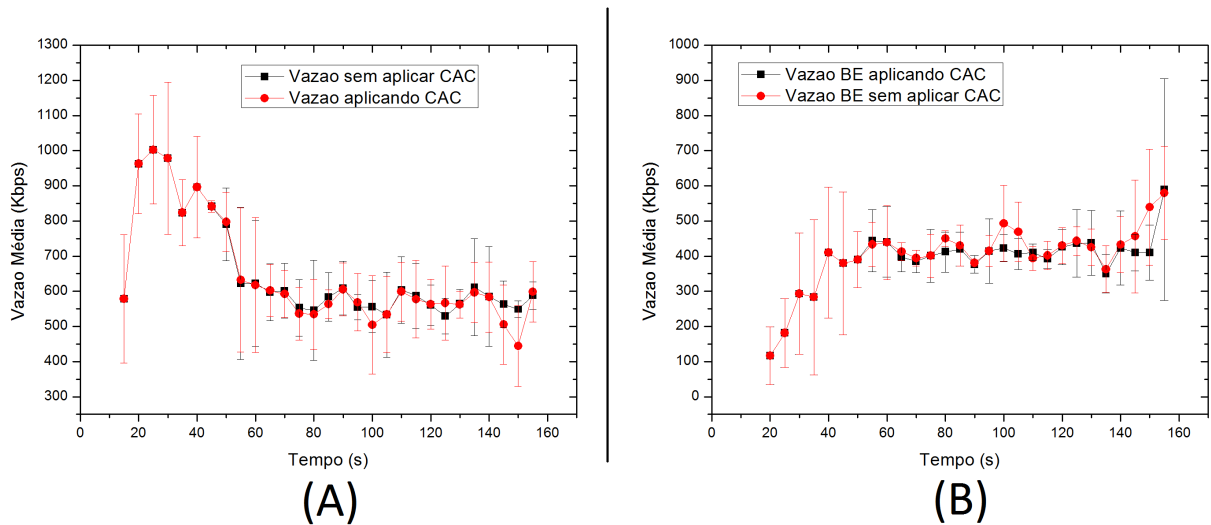


Figura 5.13: Desempenho do CAC preditivo para fluxos não tempo real).

Agora, analisando-se o atraso médio para as conexões UGS por meio da Figura 5.14, percebe-se que ele apresenta o mesmo comportamento ao longo do tempo independentemente de se aplicar ou não o mecanismo de CAC. Isso se justifica pelo fato das conexões UGS terem prioridade no escalonamento, alocando-se, dessa maneira, um percentual constante de W em cada ciclo de escalonamento, independentemente de se aplicar ou não a política de CAC. Além disso, a entrada de novas conexões rtPS na rede não influencia de maneira direta o atraso UGS, conforme já comprovado na Seção 5.3.2.

A Tabela 5.7 traz a lista dos principais eventos ocorridos nesse experimento de simu-

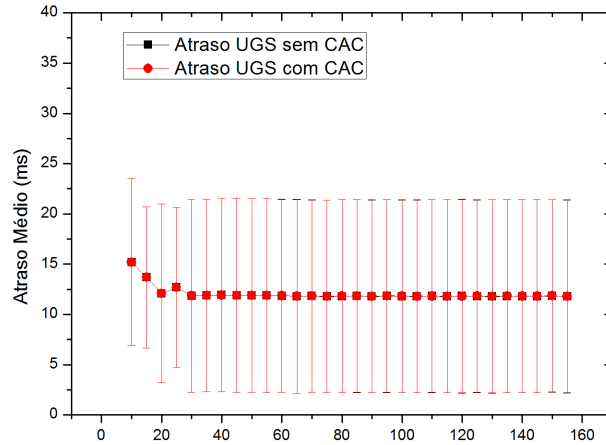


Figura 5.14: Desempenho do CAC preditivo considerando o atraso médio para fluxos UGS.

lação.

Tabela 5.7: Eventos ocorridos na simulação em função do tempo.

Instante (seg)	Evento
0-10	* <i>Initial ranging</i> e procedimentos de entrada na rede
10-14	* 1 estação com conexão UGS é adicionada à rede a cada 1 segundo
14-19	* 1 estação com conexão nrtPS é adicionada à rede a cada 1 segundo
19-115	* 1 estação com conexão rtPS é adicionada à rede a cada 5 segundos
115-160	* Não há entrada de novas estações na rede

5.3.7 Validação do Mecanismo de CAC Proposto

Finalmente, nesta seção avalia-se a eficiência do mecanismo de CAC proposto em relação à política de CAC pró-ativa baseada em taxa mínima, bastante difundida na literatura e explorada com detalhes no Capítulo 4. Para isso, considerou-se apenas fluxos rtPS adicionados à rede a partir de 15 segundos de simulação. Neste experimento de simulação, analisou-se somente o parâmetro de atraso médio até o momento em que os mecanismos de CAC iniciaram o bloqueio de novas conexões, aproximadamente no instante 70 segundos. Conforme mostrado no gráfico da Figura 5.15, embora os dois mecanismos tenham cumprido seu papel em bloquear a entrada de novas conexões quando a rede estava saturada, o mecanismo de CAC proposto foi capaz de manter o atraso

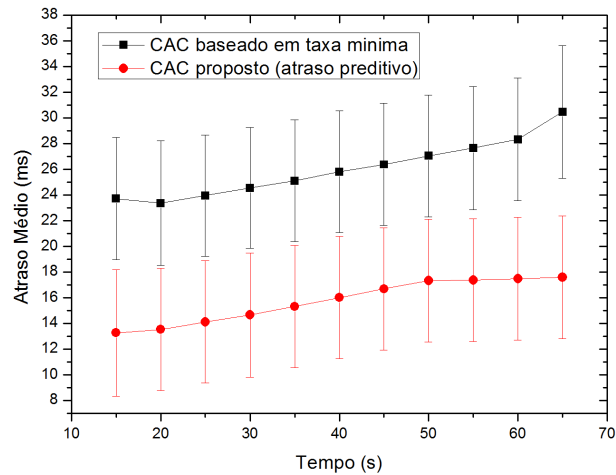


Figura 5.15: Desempenho do CAC preditivo em relação ao mecanismo de CAC baseado em taxa mínima

médio inferior ao esquema de CAC baseado em taxa mínima. A razão para isso é que o mecanismo de CAC proposto monitora e verifica o atraso médio da rede antes da admissão e isso faz com que a BS admita uma quantidade maior de conexões em relação à política de CAC baseada em taxa mínima, a qual considera apenas o número de conexões ativas na rede. Pelo gráfico apresentado na Figura 5.15 nota-se que o atraso médio quando se utiliza o mecanismo de CAC baseado em taxa mínima é superior justamente pelo fato do impacto, em termos de atraso, da admissão de uma nova conexão não ser levado consideração no processo de admissão.

A Tabela 5.8 traz a lista dos principais eventos ocorridos nesse experimento de simulação.

Tabela 5.8: Eventos ocorridos na simulação em função do tempo.

Instante (seg)	Evento
0-10	* <i>Initial ranging</i> e procedimentos de entrada na rede
10-15	* * *
15-65	* 1 estação com conexão rtPS é adicionada à rede a cada 5 segundos

5.4 Considerações Finais

Neste capítulo, avaliou-se por meio de modelagem e simulação os algoritmos para escalonamento e CAC propostos no Capítulo 4. Foi apresentada uma ferramenta alternativa desenvolvida exclusivamente para analisar o arquivo de *trace* gerado nas simulações do NS-2, que se mostrou mais adequada em relação aos tradicionais *scripts* AWK. Os ambientes de redes e de simulação também foram apresentados, bem como as características de tráfego e parâmetros de simulação. Os resultados obtidos foram analisados de acordo com as métricas de atraso e vazão e o comportamento foi analisado em função do tempo, uma vez que considerou-se a entrada periódica de conexões na rede. Nos cenários analisados avaliou-se a eficiência do mecanismo de escalonamento proposto em termos de atraso e vazão, a eficiência do mecanismo de predição de atraso, o esquema de CAC proposto tanto para fluxos tempo real quanto não tempo real e também o esquema de reserva de largura de banda, considerando a variação da parcela de largura de banda W e seu efeito sobre o atraso dos fluxos rtPS e vazão das conexões nrtPS.

Capítulo 6

Conclusões Gerais

O padrão IEEE 802.16 foi projetado para cobrir regiões metropolitanas e prover acesso banda larga sem fio em regiões desprovidas de infraestrutura, prometendo taxas de transferência de dados que podem chegar a 1 Gbps. A principal característica desse padrão é a provisão de qualidade de serviço aos usuários finais, garantida através de mecanismos para escalonamentos de pacotes e controle de admissão de conexões. Entretanto, em razão do padrão deixar em aberto a implementação desses mecanismos, essa tem sido uma ampla área de pesquisa o que tem motivado o desenvolvimento deste trabalho.

Em razão disso, o objetivo principal deste trabalho foi apresentar uma arquitetura para a provisão de qualidade de serviço no padrão IEEE 802.16 que fosse compatível com a especificação do padrão. Foi necessário então apresentar as características básicas do padrão bem como o estado da arte no que se refere a mecanismos para a provisão de QoS.

Para cumprir esse objetivo, foi apresentado no Capítulo 2 a especificação e características gerais do padrão IEEE 802.16. Foram descritas as arquiteturas PMP e *mesh* bem como o modelo de referência com ênfase nas camadas MAC e física, que formam a essência do padrão.

Em seguida, no Capítulo 3, apresentou-se os mecanismos necessários para a provisão de qualidade de serviço em redes IEEE 802.16. Inicialmente, foram apresentadas as classes de serviços definidas pelo padrão juntamente com o mecanismo necessário para o estabelecimento de fluxos de serviço. A partir disso, foram abordados os mecanismos de escalonamento de pacotes e esquemas de CAC. Procurou-se abordar esses mecanismos de acordo com algumas classificações encontradas na literatura. Em razão disso, os mecanismos de escalonamento foram apresentados segundo as disciplinas homogêneas, híbridas

e as que utilizam abordagem *cross-layer*. Já os mecanismos para controle de admissão (CAC) foram apresentados segundo as estratégias pró-ativas e reativas.

Com base nisso, no Capítulo 4 propôs-se um mecanismo para escalonamento *uplink* na estação base bem como um algoritmo para controle de admissão de conexões destinados aos fluxos de serviço rtPS. O mecanismo de escalonamento atende as quatro principais classes de serviço especificadas pelo padrão, cada uma sendo armazenada em uma fila específica na estação base, em relação a níveis diferenciados de prioridade: UGS > rtPS > nrtPS > BE. A política de atendimento às filas é baseada na disciplina PQ preemptiva com base em um esquema de reserva de largura de banda. A política de CAC trabalha de forma conjunta ao mecanismo de escalonamento e é designada para controlar a entrada de conexões rtPS na rede. Esse controle é realizado de acordo com uma predição, com base do comprimento das filas nas estações assinantes, sobre o atraso médio que uma possível conexão causará ao sistema se for admitida.

Por meio de modelagem e simulação, avaliou-se no Capítulo 5 a eficiência tanto do mecanismo de escalonamento *uplink* como também da política de CAC propostos. Todos os experimentos de simulação realizados consideraram a entrada periódica de conexões rtPS na rede e a análise foi realizada em função do tempo, tendo como métricas de avaliação o atraso e a vazão médios.

Na Seção 5.3.1 avaliou-se o comportamento do mecanismo de escalonamento proposto tanto em termos de atraso como de vazão. Para isso, um mecanismo simples para escalonar a fila de conexões rtPS foi desenvolvido, baseado na disciplina FIFO. Constatou-se um desempenho superior, a partir do instante de congestionamento, do mecanismo proposto em relação à política FIFO, justamente em razão do mecanismo proposto priorizar as estações com filas maiores e atribuir toda a largura de banda solicitada por cada fluxo. Verificou-se, por meio dessa análise comparativa, que o fato de priorizar os fluxos que apresentam mais pacotes em suas filas é uma abordagem válida e mostrou-se eficiente para a obtenção de QoS.

Nas Seções 5.3.2, 5.3.3 e 5.3.4 avaliou-se o desempenho do mecanismo de escalonamento para fluxos tempo real, fluxos não tempo real e com relação à variação da parcela de largura de banda W . Para fluxos tempo real constatou-se que a adição periódica de conexões rtPS na rede não interfere no atraso/vazão dos fluxos UGS já presentes na rede, o que comprova o fato do escalonador proposto atender as filas com prioridades diferentes

(UGS > rtPS). Esse mesmo comportamento foi observado para os fluxos não tempo real (nrtPS e BE) e verificou-se também uma diferença em termos de vazão e atraso para esses fluxos, comprovando-se o fato do escalonador priorizar os fluxos nrtPS em relação aos fluxos BE. A análise da variação da parcela W , por meio de três simulações com três diferentes valores para W , demonstrou que ela influencia fortemente no atraso rtPS e na vazão média dos fluxos nrtPS a partir do instante em que a rede está congestionada. Sendo assim, essa análise serviu de subsídio para um melhoramento dessa proposta em trabalhos futuros, os quais podem considerar a parcela W variando dinamicamente ao longo do tempo de acordo com a carga de tráfego imposta na rede.

As Seções 5.3.5 e 5.3.6 mostraram o desempenho do mecanismo de CAC proposto para a classe de serviço rtPS. O mecanismo de predição do atraso, componente chave para a tomada de decisão no algoritmo de CAC, mostrou-se eficaz na estimação do atraso médio quando comparado ao atraso obtido na rede. Essa constatação indica que o mecanismo proposto é capaz de tomar uma decisão precisa no que diz respeito a aceitar ou não uma solicitação para estabelecimento de conexão. Utilizando esse mecanismo de predição, ao aplicar o CAC sobre as conexões rtPS entrantes na rede observou-se que o mesmo é capaz de bloquear novas conexões quando o atraso médio da rede está acima de um limiar e com isso é possível manter níveis de QoS aceitáveis para os fluxos rtPS já presentes na rede. Constatou-se também que ao considerar conexões não tempo real, o mecanismo de CAC não influencia de maneira direta na vazão média dos fluxos nrtPS e BE. Uma vez que o mecanismo de CAC trabalha de forma conjunta com a política de escalonamento, a qual utiliza um esquema de reserva de largura de banda, esse comportamento se justifica.

Na Seção 5.3.7 avaliou-se a eficácia do algoritmo de CAC proposto. Para isso, foi desenvolvido um algoritmo de CAC simples baseado em taxa mínima com o objetivo de compará-lo ao CAC proposto. Verificou-se claramente um desempenho superior do mecanismo proposto em relação ao CAC baseado em taxa mínima. Isso se justifica pelo fato do CAC baseado em taxa mínima considerar apenas a quantidade de conexões presentes na rede ao passo que o mecanismo de CAC preditivo considera o estado atual da rede baseado no atraso médio rtPS.

Apesar de suficiente para validar o mecanismo proposto, várias questões relevantes e que podem trazer resultados importantes não foram tratadas nesse trabalho. A primeira delas é o fato do mecanismo de CAC ter sido proposto apenas para a classe de serviço rtPS.

Uma sugestão para trabalho futuro é a extensão do mecanismo de CAC proposto para as demais classes de serviço especificadas pelo padrão IEEE 802.16. A segunda questão é que o mecanismo de escalonamento *uplink* pode utilizar algum mecanismo dinâmico que se beneficie do empréstimo de largura de banda dos fluxos de baixa prioridade, ao invés de um esquema de reserva estática de largura de banda. Além disso, a análise dos mecanismos de escalonamento e CAC propostos, cujos resultados foram comparados com os resultados da disciplina FIFO e da política baseada em taxa mínima, respectivamente, poderá ser melhor avaliada em trabalhos futuros por meio de comparações com outras disciplinas de escalonamento e algoritmos de CAC. Por fim, a arquitetura para a provisão de QoS proposta neste trabalho pode ser complementada através do acréscimo de um módulo para policiamento do tráfego na BS.

Referências Bibliográficas

- [1] R. Prasad and F. J. Velez, *WiMAX Networks: Techno-Economic Vision and Challenges*. Nova York: Springer, 1 ed., 2010.
- [2] C. Eklund, “IEEE Standard 802.16: A Technical Overview of the WirelessMAN Air Interface for Broadband Wireless Access,” *IEEE Communications Magazine*, June, 2002.
- [3] IEEE 802.16-2001, “IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems,” tech. rep., IEEE Computer Society, April, 2002.
- [4] E. Honório, E. Augusto, and L. Cabral, “OFDMA e WCDMA,” *Faculdade Integrada do Recife*, Disponível em: <http://www.arquivos.fir.br>, October, 2006.
- [5] K. Wongthavarawat and A. Ganz, “Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems,” *International Journal of Communication Systems*, vol. 16, pp. 81–96, February, 2003.
- [6] A. Belghith and L. Nuaymi, *Scheduling Techniques for WiMAX*, pp. 61–83. Current Technology Developments of WiMAX Systems, Singapore: Springer, 2009.
- [7] F. Ohrtman, *WiMAX Handbook: Building 802.16 Wireless Networks*. Nova York: McGraw-Hill Communications, 1 ed., 2005.
- [8] J. F. Borin, *Mecanismos para Provisão de Qualidade de Serviço em Redes IEEE 802.16*. PhD thesis, Universidade Estadual de Campinas, Campinas/SP, Brazil, January, 2010.

- [9] M. H. Ahmed, "Call Admission Control in Wireless Networks: A Comprehensive Survey," *IEEE Communications Surveys*, vol. 7, pp. 50–68, January, 2005.
- [10] IEEE 802.11-2007, "IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," tech. rep., IEEE Computer Society, June, 2007.
- [11] A. Bacioccola and C. Cicconetti, "IEEE 802.16: History, Status and Future Trends," *Computer Communications*, vol. 33, pp. 113–123, November, 2009.
- [12] IEEE 802.16a-2003, "Amendment to IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Wireless Access Systems - Medium Access Control Modifications and Additional Physical Layers Specifications for 2-11 GHz," tech. rep., IEEE Computer Society, April, 2003.
- [13] IEEE 802.16-2004, "IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems," tech. rep., IEEE Computer Society, October, 2004.
- [14] IEEE 802.16e-2005, "Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Band," tech. rep., IEEE Computer Society, February, 2006.
- [15] IEEE 802.16f, "IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems - Management Information Base," tech. rep., IEEE Computer Society, December, 2005.
- [16] IEEE 802.16g, "IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems - Amendment 3: Management Plane Procedures and Services," tech. rep., IEEE Computer Society, December, 2007.
- [17] IEEE 802.16h/D9, "IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems - Improved Coexistence Mechanisms for License-Exempt Operation," tech. rep., IEEE Computer Society, March, 2009. Draft D9.

- [18] IEEE 802.16j, “IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems - Multihop Relay Specification,” tech. rep., IEEE Computer Society, May, 2009.
- [19] IEEE 802.16-2009, “IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems,” tech. rep., IEEE Computer Society, May, 2009.
- [20] IEEE 802.16m/D2, “Amendment to IEEE Std 802.16-2009, developed by Task Group m Working Document,” tech. rep., IEEE Computer Society, May, 2009. Draft D2, Disponível em: <http://wirelessman.org/pubs/802.16m.html>.
- [21] S. Ahmadi, “An Overview of Next-Generation Mobile WiMAX Technology,” *IEEE Communications Magazine*, vol. 9, pp. 84–98, June, 2009.
- [22] I. Msadaa, D. Camara, and F. Filali, “Scheduling and CAC in IEEE 802.16 Fixed BWNs: A Comprehensive Survey and Taxonomy,” *Communications Surveys Tutorials, IEEE*, vol. 12, pp. 459–487, Fourth quarter, 2010.
- [23] M. Shand and J. Vuillemin, “Fast Implementations of RSA Cryptography,” in *Computer Arithmetic, 1993. Proceedings., 11th Symposium on*, pp. 252–259, June, 1993.
- [24] S.-J. Han, H.-S. Oh, and J. Park, “The Improved Data Encryption Standard (DES) Algorithm,” in *Spread Spectrum Techniques and Applications Proceedings, 1996., IEEE 4th International Symposium on*, vol. 3, pp. 1310–1314, September, 1996.
- [25] N. Saqib, F. Rodriguez-Henriquez, and A. Diaz-Perez, “AES Algorithm Implementation - An Efficient Approach for Sequential and Pipeline Architectures,” in *Proceedings of the Fourth Mexican International Conference on*, pp. 126–130, September, 2003.
- [26] E. Khan, M. El-Kharashi, F. Gebali, and M. Abd-El-Barr, “A Reconfigurable Hardware Unit for the HMAC Algorithm,” in *Information and Communications Technology, 2005. Enabling Technologies for the New Knowledge Society: ITI 3rd International Conference on*, pp. 861–874, December, 2005.
- [27] O. C. M. B. Duarte, P. M. Esposito, and R. B. Z. Antas, “WiMAX: IEEE 802.16 Standard,” tech. rep., Universidade Federal do Rio de Janeiro (UFRJ), 2006.

- [28] A. Vogel, B. Kerherve, G. von Bochmann, and J. Gecsei, "Distributed Multimedia and QoS: a Survey," *Multimedia, IEEE*, vol. 2, no. 2, pp. 10–19, 1995.
- [29] ITU-T, *Rec. E.800. Terms and Definitions Related To Quality of Service and Network Performance Including Dependability*, August, 1994.
- [30] M. Ma and J. Lu, *QoS Provision Mechanism in WiMAX*, pp. 85–113. Current Technology Developments of WiMAX Systems, Singapore: Springer, 2009.
- [31] N. A. Ali, P. Dhrona, and H. Hassanein, "A Performance Study of Uplink Scheduling Algorithms in Point-to-multipoint WiMAX Networks," *Computer Communications*, vol. 32, pp. 511–521, September, 2008.
- [32] C. So-In, R. Jain, and A.-K. Tamimi, "A Deficit Round Robin with Fragmentation scheduler for IEEE 802.16e Mobile WiMAX," in *Sarnoff Symposium, 2009. SARNOFF '09. IEEE*, pp. 1–7, 2009.
- [33] J. A. Afonso, *Acesso Local sem Fios em Redes de Comunicação: Escalonamento de Tráfego de Tempo Real em Sistemas de Aquisição de Dados e Controle*. PhD thesis, Universidade do Minho, Braga, Portugal, November, 2004.
- [34] Y. Zhang, *WiMAX Network Planning and Optimization*. CRC Press, 2008.
- [35] B. Stewart and D. Donohue, "CCNP ONT Quick Reference Sheets." Cisco Press, 2007.
- [36] C. Semeria, "Supporting Differentiated Service Classes: Queue Scheduling Disciplines," tech. rep., Juniper Networks, 2001.
- [37] H. Wang, B. He, and D. P. Agrawal, "Above Packet Level Admission Control and Bandwidth Allocation for IEEE 802.16 Wireless MAN," *Simulation Modelling Practice and Theory*, vol. 15, no. 4, pp. 366–382, 2007.
- [38] H. Wang, W. Li, and D. P. Agrawal, "Dynamic Admission Control and QoS for 802.16 Wireless MAN," in *Wireless Telecommunications Symposium, WTS'05 Conference on*, pp. 60–66, April, 2005.
- [39] K. Vinay, N. Sreenivasulu, D. Jayaram, and D. Das, "Performance Evaluation of End-to-end Delay by Hybrid Scheduling Algorithm for QoS in IEEE 802.16 Network," in

Wireless and Optical Communications Networks, IFIP International Conference on, pp. 11–13, April, 2006.

- [40] D. Tarchi, R. Fantacci, and M. Bardazzi, “Quality of Service Management in IEEE 802.16 Wireless Metropolitan Area Networks,” in *International Conference on Communications, (ICC’06)*, pp. 11–15, 2006.
- [41] C. Ball, F. Trembl, X. Gaube, and A. Klein, “Performance Analysis of Temporary Removal Scheduling Applied to Mobile WiMax Scenarios in Tight Frequency Reuse,” in *Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005. IEEE 16th International Symposium on*, vol. 2, pp. 888–894, September, 2005.
- [42] H. K. Rath, A. Bhorkar, and V. Sharma, “An Opportunistic DRR (O-DRR) Uplink Scheduling Scheme for IEEE 802.16-based Broadband Wireless Networks,” in *IETE, International Conference on Next Generation Networks (ICNGN)*, 2006.
- [43] K. Tsang, L. Lee, H. Tung, R. Lam, Y. Sun, and K. Ko, “Admission Control Scheme for Mobile WiMAX Networks,” in *Consumer Electronics, 2007. ISCE 2007. IEEE International Symposium on*, pp. 1–5, June, 2007.
- [44] S. Chaudhry and R. Guha, “Adaptive Connection Admission Control and Packet Scheduling for QoS Provisioning in Mobile WiMAX,” in *Signal Processing and Communications, 2007. ICSPC 2007. IEEE International Conference on*, pp. 1355–1358, November, 2007.
- [45] J. Freitag and N. Fonseca, *Quality of Service Architectures for Wireless Networks: Performance Metrics and Management*, Chapter 4. Admission Control and Scheduling for QoS Provisioning in WiMAX Networks, pp. 183–202. IGI Global, 2010.
- [46] F. Prado e M. Arenales, “O Problema da Mochila Compartimentada e Aplicações,” *Pesquisa Operacional*, vol. 22, pp. 285–304, July, 2002.
- [47] B. Rong, Y. Qian, K. Lu, H.-H. Chen, and M. Guizani, “Call Admission Control Optimization in WiMAX Networks,” *Vehicular Technology, IEEE Transactions on*, vol. 57, pp. 2509–2522, July, 2008.

- [48] S. Khemiri, K. Boussetta, N. Achir, and G. Pujolle, “Optimal Call Admission Control for an IEEE 802.16 Wireless Metropolitan Area Network,” in *Network Control and Optimization* (T. Chahed and B. Tuffin, eds.), vol. 4465 of *Lecture Notes in Computer Science*, pp. 105–114, Springer Berlin / Heidelberg, 2007.
- [49] M. Guignard, “Lagrangian Relaxation: A Tutorial.” Department of OPIM, The Wharton School, University of Pennsylvania, February, 2008.
- [50] T.-C. Tsai, C.-H. Jiang, and C.-Y. Wang, “CAC and Packet Scheduling Using Token Bucket for IEEE 802.16 Networks,” *Journal of Communications*, vol. 1, pp. 30–37, May, 2006.
- [51] C.-H. Jiang and T.-C. Tsai, “Token bucket based CAC and packet scheduling for IEEE 802.16 broadband wireless access networks,” in *Consumer Communications and Networking Conference, 2006. CCNC 2006. 3rd IEEE*, vol. 1, pp. 183–187, January, 2006.
- [52] C. F. Leal, S. V. de Carvalho, A. N. F. da Rosa, C. R. L. Francês, e J. C. W. A. Costa, “Processo Markoviano de Decisão para Controle de Admissão de Conexão e Alocação Dinâmica de Recursos em Redes IEEE 802.16,” in *XXVII Simpósio Brasileiro de Telecomunicações, SBRT’09*, 2009.
- [53] M. Castrucci, F. D. Priscoli, C. Buccella, V. Puccio, and I. Marchetti, “Connection Admission Control in WiMAX Networks,” in *ICT’08 Conference on*, 2008.
- [54] J. Proakis, *Digital Communications*. McGraw-Hill Education, 4 ed., 2000.
- [55] M. Garg, “Linear Prediction Algorithms,” tech. rep., Indian Institute of Technology, Bombay, India, 2003.
- [56] F. Hou, P.-H. Ho, and X. Shen, “Performance Analysis of a Reservation Based Connection Admission Scheme in 802.16 Networks,” in *Global Telecommunications Conference, 2006. GLOBECOM ’06. IEEE*, pp. 1–5, December, 2006.
- [57] J. F. Borin and N. L. S. Fonseca, “Algoritmos Para Controle de Admissão e Escalonamento do Tráfego *Uplink* em Redes IEEE 802.16,” in *Workshop em Desempenho de Sistemas Computacionais e de Comunicação, SBC*, vol. 8, pp. 2273–2288, 2009.

- [58] *Network Simulator 2 - NS-2*, Disponível em: <http://www.isi.edu/nsnam/ns/>, Acessado em: 18/05/2011, 2010.
- [59] NIST, *The Network Simulator NS-2 NIST add on, IEEE 802.21 model (based on IEEE P802.21/D03.00)*, Acessado em: 7/12/2010, 2010.
- [60] C. Hoymann, “Analysis and Performance Evaluation of the OFDM-based Metropolitan Area Network IEEE 802.16,” *Computer Networks*, vol. 49, pp. 341–363, October, 2005.

Trabalhos Aceitos/Publicados Pelo Autor

- [61] E. C. Rosa and P. R. Guardieiro, “CAC and Uplink Scheduling Algorithms in WiMAX Networks,” (Artigo Convidado), *Revista Telecomunicações*, Santa Rita do Sapucaí, Brazil, 2011.
- [62] E. C. Rosa and P. R. Guardieiro, “An Uplink Scheduling and CAC Algorithms for IEEE 802.16 Networks,” in *4th International Workshop on Telecommunications, IWT’11*, Rio de Janeiro, Brazil, pp. 94–100, 2011.
- [63] E. C. Rosa and P. R. Guardieiro, “Algoritmos Para Controle de Admissão de Chamadas e Escalonamento em Redes IEEE 802.16,” in *9th International Information and Telecommunication Symposium, I2TS’10*, Rio de Janeiro, Brazil, p. 7p, 2010.
- [64] E. C. Rosa and P. R. Guardieiro, “CAC Using Delay Prediction and Uplink Scheduling Algorithm for IEEE 802.16 Networks,” Aceito para publicação em *2nd World Congress on Computer Science and Information Engineering, CSIE’11, IEEE Communication Society*, Changchun, China, p. 6p, 2011.
- [65] E. C. Rosa and P. R. Guardieiro, “A Delay Prediction Scheme for Admission Control and a Scheduling Algorithm for IEEE 802.16 Networks,” Aceito para publicação em *13th International Conference on Advanced Communication Technology, ICACT’11, IEEE Communication Society*, Phoenix Park, Korea, p. 6p, 2011.

- [66] E. C. Rosa and P. R. Guardieiro, “Esquemas para Controle de Admissão de Chamadas em Redes IEEE 802.16,” in *VIII Conferência de Estudos em Engenharia Elétrica, CEEL'10*, Uberlândia, Brazil, p. 7p, 2010.

Apêndice A

Neste apêndice são apresentados os algoritmos de escalonamento para as classes de serviço UGS, rtPS, nrtPS e BE. Em todos esses algoritmos, considera-se o elemento $Q_{i,j}$ como sendo o *data grant* ou requisição de largura de banda j referente à fila i , onde $i \in [1 = UGS, 2 = rtPS, 3 = nrtPS, 4 = BE]$. Define-se também o valor $|Q_i|$ como sendo a quantidade de elementos da fila i e uma constante P que denota o comprimento do preâmbulo transmitido em cada quadro de operação TDD. Os valores T e R são referentes a segmentação de largura de banda destinadas aos fluxos nrtPS e BE, respectivamente.

Algoritmo 6.1 Escalona_Fluxos_UGS

Entrada: N_b = Número de símbolos, DL-MAP = Mapa de *Uplink*

Saída: N_b atualizado, DL-MAP atualizado

```
1: para ( $j$  de 1 até  $|Q_1|$ ) faça
2:    $BW_{ugs} \leftarrow$  Símbolos OFDM alocados a cada frame para fluxos UGS.
3:   se ( $BW_{ugs} + P > (N_b - (T + R))$ ) então
4:     se ( $(N_b - (T + R)) < P$ ) então
5:       Realiza preempção na fila UGS (vai para passo 12). //Requer CAC para UGS
6:     fim se
7:      $BW_{ugs} \leftarrow (N_b - (T + R));$ 
8:   fim se
9:   UL-MAP  $\leftarrow$  Adiciona  $BW_{ugs}$  associado à conexão  $Q_{1,j}.conexao$ ;
10:   $N_b \leftarrow N_b - BW_{ugs};$ 
11: fim para
12: retorna  $N_b$ , UL-MAP;
```

No algoritmo A.1, para cada fluxo UGS calcula-se a quantidade de símbolos necessários para garantir o requisito de taxa constante dessa classe (linha 2). Esse cálculo é realizado em função do parâmetro de taxa máxima sustentável (tx_{ugs}) para os fluxos UGS,

a periodicidade (p) com que o escalonamento UGS é executado (1 se for a cada *frame*), a eficiência da modulação empregada em bits/símbolo (e) e também a duração do símbolo OFDM (d). A alocação é então dada por:

$$b_{ugs} = \frac{(tx_{ugs} * p * 8/e) * d}{d} \quad (6.1)$$

Algoritmo 6.2 Escalona_Fluxos_RTPS

Entrada: N_b = Número de símbolos, DL-MAP = Mapa de *Uplink*

Saída: N_b atualizado, DL-MAP atualizado

```

1: para ( $k_1$  de 1 até  $|Q_2| - 1$ ) faça
2:   para ( $k_2$  de  $k_1 + 1$  até  $|Q_2|$ ) faça
3:     se ( $Q_{2,k_1} \leq Q_{2,k_2}$ ) então
4:       temp =  $Q_{2,k_1}$ 
5:        $Q_{2,k_1} = Q_{2,k_2}$ 
6:        $Q_{2,k_2} = \text{temp}$ 
7:     fim se
8:   fim para
9: fim para
10: para ( $j$  de 1 até  $|Q_2|$ ) faça
11:    $BW_j \leftarrow$  Símbolos OFDM necessários para transmitir  $Q_{2,j}$ .
12:   se ( $BW_j + P > (N_b - (T + R))$ ) então
13:     se ( $(N_b - (T + R)) < P$ ) então
14:       Realiza preempção na fila rtPS.
15:     fim se
16:      $BW_j \leftarrow (N_b - (T + R));$ 
17:   fim se
18:   UL-MAP  $\leftarrow$  Adiciona  $BW_j$  associado à conexão  $Q_{2,j}.conexao$ ;
19:    $N_b \leftarrow N_b - BW_j$ ;
20: fim para
21: retorna UL-MAP;

```

Entre as linhas 1 e 9 do algoritmo A.2 ordena-se a fila das requisições rtPS a fim de atender primeiro as estações com uma quantidade maior de *bytes* em suas filas. A

alocação de largura de banda na linha 11 é calculada também com base na equação A.1 só que ao invés do valor tx_{ugs} utiliza-se o valor da requisição presente na fila.

Algoritmo 6.3 Escalona_Fluxos_NRTPS

Entrada: N_b = Número de símbolos, DL-MAP = Mapa de *Uplink*

Saída: N_b atualizado, DL-MAP atualizado

```
1: para ( $j$  de 1 até  $|Q_3|$ ) faça
2:    $BW_{reservada} \leftarrow (N_b - R)/|Q_3|$ 
3:   se ( $BW_{solicitada} \leq BW_{reservada}$ ) então
4:      $BW_{alocada} \leftarrow BW_{solicitada}$ 
5:   senão
6:      $BW_{alocada} \leftarrow BW_{reservada}$ 
7:   fim se
8:   se ( $BW_{alocada} + P > (N_b - R)$ ) então
9:     se ( $(N_b - R) < P$ ) então
10:      Realiza preempção na fila nrtPS (vai para passo 17).
11:    fim se
12:     $BW_{alocada} \leftarrow (N_b - R - P);$ 
13:  fim se
14:  UL-MAP  $\leftarrow$  Adiciona  $BW_{alocada}$  associado à conexão  $Q_{3,j}.conexao$ ;
15:   $N_b \leftarrow N_b - BW_{alocada}$ ;
16: fim para
17: retorna  $N_b$ , UL-MAP;
```

No algoritmo A.3, para cada estação nrtPS presente na rede, calcula-se a largura de banda alocada para o fluxo como sendo o valor resultante de $\min(banda_{reservada}, banda_{solicitada})$ nas linhas 2 até 7. A largura de banda residual será destinada aos fluxos BE que serão escalonados em seguida.

Algoritmo 6.4 Escalona_Fluxos_BE

Entrada: N_b = Número de símbolos, DL-MAP = Mapa de *Uplink*

Saída: N_b atualizado, DL-MAP atualizado

```
1:  $index\_ \leftarrow$  Índice da última estação BE atendida no frame anterior;
2: para ( $j$  de 1 até  $|Q_4|$ ) faça
3:    $i \leftarrow (j + index\_ ) \bmod |Q_4|$ ;
4:    $BW_{be} \leftarrow$  Símbolos OFDM para a transmissão da requisição referente à estação  $i$ ;
5:   se ( $BW_{be} + P > N_b$ ) então
6:     se ( $(N_b < P)$ ) então
7:       Realiza preempção na fila BE;
8:     fim se
9:      $BW_{be} \leftarrow (N_b - P)$ ;
10:  fim se
11:  UL-MAP  $\leftarrow$  Adiciona  $BW_{be}$  associado à conexão  $Q_{4,index\_}.conexao$ ;
12:   $N_b \leftarrow N_b - BW_{be}$ ;
13: fim para
14: retorna  $N_b$ , UL-MAP;
```

O algoritmo A.4 atende cada um dos fluxos BE presentes na rede começando pela última estação atendida no *frame* anterior. A partir disso, as demais estações vão sendo atendidas sequencialmente reservando-se a cada fluxo a quantidade de símbolos OFDM necessária para a transmissão de toda a requisição. A quantidade de símbolos OFDM é calculada com base na equação A.1.