

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

ELIANE GOBETTI DE CAMARGO

**Uma Arquitetura de Escalonamento
Hierárquica para Transmissões *Uplink* em
Redes IEEE 802.16e Baseadas em OFDMA**

Dissertação apresentada como requisito parcial
para a obtenção do grau de
Mestre em Ciência da Computação

Dra. Liane Margarida Rockenbach Tarouco
Orientador

Dr. Juergen Rochol
Co-orientador

Porto Alegre, maio de 2010

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Eliane Gobetti de Camargo,

Uma Arquitetura de Escalonamento Hierárquica para Transmissões *Uplink* em Redes IEEE 802.16e Baseadas em OFDMA /

Eliane Gobetti de Camargo. – Porto Alegre: PPGC da UFRGS, 2010.

82 f.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2010. Orientador: Liane Margarida Rockenbach Tarouco; Co-orientador: Juergen Rochol.

1. Escalonamento. 2. OFDMA. 3. IEEE 802.16e. 4. WiMAX. I. Tarouco, Liane Margarida Rockenbach. II. Rochol, Juergen. III. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos Alexandre Netto

Vice-Reitor: Prof. Rui Vicente Oppermann

Diretor do Instituto de Informática: Prof. Flávio Rech Wagner

Coordenadora do PPGC: Prof. Álvaro Freitas Moreira

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

Aos meus pais, pelo apoio.

AGRADECIMENTOS

A minha família, pelo apoio e paciência.

Aos professores do Departamento de Informática da UFRGS pelos ensinamentos. Em especial ao professor Juergen Rochol, pela orientação, pelas lições e pela disponibilidade.

Aos colegas do grupo de redes, em especial a Cristiano Both e Rafael Kunst, pelas discussões a respeito deste trabalho, pelas revisões e pela constante ajuda no desenvolvimento desta proposta.

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS	7
LISTA DE FIGURAS	10
LISTA DE TABELAS	11
RESUMO	12
ABSTRACT	13
1 INTRODUÇÃO	14
2 CONTEXTUALIZAÇÃO DA TECNOLOGIA IEEE 802.16E	16
2.1 Mecanismo AMC	17
2.2 Classes de Serviço	18
2.3 Multiplexação OFDMA	19
2.3.1 Subquadro DL	20
2.3.2 Subquadro UL	21
2.4 Mecanismos de Alocação de Banda	22
2.4.1 Oportunidades de Requisição de Banda	23
2.4.2 Oportunidades de Transmissão de Dados	25
3 TÉCNICAS DE ESCALONAMENTO	27
3.1 Algoritmos de Escalonamento Clássicos	27
3.2 Técnicas de Multiplexação de Canal e Escalonamento	28
3.3 Escalonamento Considerando OFDMA como Técnica de Multiplexação do Canal	30
3.3.1 Estratégias Híbridas Considerando OFDMA	30
3.3.2 Escalonamento UL Considerando OFDMA	32
4 UMA ARQUITETURA DE ESCALONAMENTO UPLINK	35
4.1 Utilização do Mecanismo de Alocação de Banda UL	36
4.2 Alocação de Rajadas	37
4.3 Gerenciamento de Requisições	40
4.3.1 Escalonamento Intra-classe	41
4.3.2 Escalonamento Inter-classes	41

5	FERRAMENTA DE SIMULAÇÃO	44
5.1	Ambientes de Programação	44
5.2	Modelos de Tráfego	45
5.2.1	Modelo de Tráfego VoIP sem Supressão de Silêncio	45
5.2.2	Modelo de Tráfego VoIP com Supressão de Silêncio	46
5.2.3	Modelo de Tráfego HTTP	47
5.2.4	Modelo de Tráfego FTP	48
5.2.5	Modelo de Tráfego Vídeo <i>Clip</i>	48
5.3	Fluxograma da Implementação	49
5.3.1	Perspectiva Geral	49
5.3.2	Perspectiva da BS	50
5.3.3	Perspectiva das MSs	51
6	RESULTADOS	52
6.1	Parâmetros de Simulação	52
6.2	Análise Considerando Cinco Classes de Serviço	53
6.3	Avaliação de Desempenho Comparativa	58
7	CONCLUSÕES E TRABALHOS FUTUROS	63
APÊNDICE A	ARTIGO PUBLICADO NO SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS - SBRC09	68

LISTA DE ABREVIATURAS E SIGLAS

AMC	<i>Adaptive Modulation and Coding</i>
AMR	<i>Adaptive Multi Rate</i>
BE	<i>Best Effort</i>
BPSK	<i>Binary Phase Shift Keying</i>
BS	<i>Base Station</i>
CAC	<i>Connection Admission Control</i>
CDMA	<i>Code Division Multiple Access</i>
CID	<i>Connection Identifier</i>
CP	<i>Cyclic Prefix</i>
DCD	<i>Downlink Channel Descriptor</i>
DL	<i>Downlink</i>
DMIA	<i>Dynamic MCS and Interference Aware Scheduling</i>
DRR	<i>Deficit Round Robin</i>
DSL	<i>Digital Subscriber Line</i>
EDF	<i>Earliest Deadline First</i>
ertPS	<i>extended real-time Polling Service</i>
FDD	<i>Frequency Division Duplex</i>
FEC	<i>Forward Error-Correction</i>
FHC	<i>Frame Control Header</i>
FIFO	<i>First-In-First-Out</i>
FTP	<i>File Transfer Protocol</i>
FUSC	<i>Full Usage of Subchannels</i>
GPS	<i>Generalized Processor Sharing</i>
HTTP	<i>Hypertext Transfer Protocol</i>
HUF	<i>Highest Urgency First</i>
HUMAN	<i>High-speed Unlicensed MAN</i>

IE	<i>Information Element</i>
IEEE	<i>Institute of Electrical and Eletronic Engineers</i>
MAC	<i>Medium Access Control</i>
MAN	<i>Metropolitan Area Network</i>
MCS	<i>Modulation and Coding Scheme</i>
MS	<i>Mobile Station</i>
MTU	<i>Maximum Transfer Unit</i>
M-LWDF	<i>Modified Largest Delay First</i>
nrtPS	<i>non-real-time Polling Service</i>
OFDM	<i>Orthogonal Frequency Division Multiplexing</i>
OFDMA	<i>Orthogonal Frequency Division Multiple Access</i>
PDF	<i>Probability Density Function</i>
PF	<i>Proportional Fair</i>
PFS	<i>Proportional Fairness Scheme</i>
PQ	<i>Priority Queueing</i>
PS	<i>Processor Sharing</i>
P2P	<i>Peer-to-Peer</i>
PUSC	<i>Partial Usage of Subchannels</i>
QAM	<i>Quadrature Amplitude Modulation</i>
QoS	<i>Quality of Service</i>
QPSK	<i>Quadrature Phase Shift Keying</i>
RR	<i>Round Robin</i>
RTG	<i>Receive/Transmit Transition Gap</i>
rtPS	<i>real-time Polling Service</i>
SC	<i>Single Carrier</i>
SCa	<i>Single Carrier adaptive</i>
TDD	<i>Time Division Duplex</i>
TTG	<i>Transmit/Receive Transition Gap</i>
TUSC	<i>Tile Usage of Subchannels</i>
UCD	<i>Uplink Channel Descriptor</i>
UGS	<i>Unsolicited Grant Service</i>
UL	<i>Uplink</i>
VI	<i>Virtual Instrument</i>
VoIP	<i>Voice over Internet Protocol</i>

WFQ *Weighted Fair Queueing*
WiMAX *Worldwide Interoperability for Microwave Access*
WRR *Weighted Round Robin*

LISTA DE FIGURAS

Figura 2.1:	Estrutura do quadro OFDMA	20
Figura 2.2:	Alocação de uma Rajada UL pela BS	22
Figura 2.3:	Acesso às filas de dados pelos escalonadores DL e UL	22
Figura 2.4:	Funcionamento do período de contenção	24
Figura 4.1:	Arquitetura de escalonamento <i>uplink</i>	36
Figura 4.2:	Estrutura de um símbolo OFDMA	37
Figura 5.1:	Fluxograma da implementação - perspectiva geral	49
Figura 5.2:	Fluxograma da implementação - perspectiva da BS	50
Figura 5.3:	Fluxograma da implementação - perspectiva das MSs	51
Figura 6.1:	Composição do subquadro UL	54
Figura 6.2:	Percentual de utilização do subquadro UL por classe de serviço	55
Figura 6.3:	Atraso médio no atendimento das requisições	56
Figura 6.4:	Percentual de requisições descartadas	57
Figura 6.5:	Comparativo entre os atrasos no atendimento das requisições BE	59
Figura 6.6:	Comparativo entre os atrasos no atendimento das requisições rtPS	60
Figura 6.7:	Percentual de requisições BE e rtPS descartadas	61

LISTA DE TABELAS

Tabela 2.1:	Perfis de rajada previstos pelo padrão IEEE 802.16e	17
Tabela 2.2:	Principais requisitos de QoS por classe de serviço	19
Tabela 2.3:	Utilização do mecanismo de alocação de banda em cada classe de serviço	25
Tabela 3.1:	Aspectos considerados pelos trabalhos relacionados	34
Tabela 4.1:	Utilização do mecanismo de alocação de banda UL na arquitetura proposta	36
Tabela 4.2:	Relação entre o fator de amostragem e a largura de banda	38
Tabela 4.3:	Duração do quadro para transmissões OFDMA	39
Tabela 5.1:	Aplicação associada a cada classe de serviço	45
Tabela 5.2:	Parâmetros codec G.711	45
Tabela 5.3:	Parâmetros codec AMR	46
Tabela 5.4:	Distribuição exponencial	46
Tabela 5.5:	Parâmetros para o tráfego HTTP	47
Tabela 5.6:	Distribuições lognormal e Pareto truncada	48
Tabela 5.7:	Parâmetros para o tráfego FTP	48
Tabela 5.8:	Parâmetros para o tráfego de vídeo <i>clip</i>	48
Tabela 6.1:	Parâmetros de Simulação	53

RESUMO

O padrão IEEE 802.16e é uma tecnologia emergente para redes de próxima geração com suporte à qualidade de serviço para usuários fixos e móveis. No contexto de aplicações móveis, a técnica OFDMA (*Orthogonal Frequency Division Multiple Access*) deve ser empregada para multiplexação de canais de rádio frequência, suportando a alocação simultânea do canal para múltiplos usuários. Porém, essa alocação, considerando a estrutura do quadro OFDMA, não é definida pelo padrão IEEE 802.16e. A maioria das pesquisas atuais sobre arquiteturas de escalonamento baseadas nessa técnica de multiplexação estão focadas na transmissão *downlink* ou consideram apenas parcialmente os aspectos decorrentes do emprego da técnica OFDMA. Sendo assim, neste trabalho é apresentada uma arquitetura de escalonamento hierárquica, com dois níveis, para transmissões *uplink* em redes IEEE 802.16e, baseadas em OFDMA. A arquitetura proposta aborda a divisão do subquadro *uplink* entre oportunidades de requisição de banda e oportunidades de transmissão de dados, no *Nível 1*, e a seleção das requisições de banda para atendimento, no *Nível 2*. A proposta considera, além da técnica de multiplexação do canal, as cinco classes de serviço previstas pelo padrão IEEE 802.16e e o mecanismo de alocação de banda definido para a tecnologia. Uma ferramenta de simulação foi desenvolvida para avaliar o desempenho da proposta, considerando os diferentes aspectos da arquitetura e comparando com trabalhos relacionados. A avaliação é realizada considerando métricas como atraso médio no atendimento de requisições, percentual de requisições descartadas e a utilização do quadro.

Palavras-chave: Escalonamento, OFDMA, IEEE 802.16e, WiMAX.

Proposal of an Hierarchical Uplink Scheduling Architecture for OFDMA based IEEE 802.16e Networks

ABSTRACT

The IEEE 802.16e standard is an emerging next generation network technology, designed to provide guaranteed QoS to both fixed and mobile users. In the context of mobile users, OFDMA technique must be used to multiplex the radio frequency channel. This technique allows to allocate the channel to multiple users at the same time. However, details about this allocation, considering the OFDMA frame structure, are not defined by the IEEE 802.16e standard. In this context, most of the current researches found on literature focus on proposing scheduling architectures for downlink transmissions only. Furthermore, researches that consider uplink transmissions, typically consider OFDMA aspects only partially. Therefore, this work presents an hierarchical scheduling architecture designed for uplink transmissions. The proposed architecture is composed of two levels. In the Level 1 the architecture addresses the division of the uplink subframe in bandwidth request opportunities, and data transmission opportunities. Level 2 is used to select among the bandwidth request, those that will be served on the next uplink subframe. The proposal considers OFDMA multiplexing technique, the five classes of service defined for QoS provisioning, and the bandwidth allocation mechanism defined in IEEE 802.16e standard. In order to evaluate the behavior of the proposed architecture, a simulation tool was developed. The approach used for evaluation was to compare the performance of the proposed architecture with related works found in the literature. The performance evaluation considers metrics such as average delay for serving requests, requests discarded, and the overall frame utilization.

Keywords: scheduling, OFDMA, IEEE 802.16e, WiMAX.

1 INTRODUÇÃO

O surgimento de dispositivos móveis, como *laptops* e a próxima geração de celulares está mudando o perfil de acesso dos usuários de redes de computadores, criando condições para o desenvolvimento de aplicações ubíquas, que precisam oferecer informações a qualquer tempo e em qualquer lugar. Nesse contexto, redes sem fio de banda larga, como IEEE (*Institute of Electrical and Eletronic Engineers*) 802.16e, conhecidas como WiMAX (*Worldwide Interoperability for Microwave Access*), devido ao fórum de fabricantes, são apontadas como uma alternativa de rede emergente, devido a algumas características da tecnologia.

O padrão IEEE 802.16e prevê o suporte a uma grande quantidade de usuários móveis e nomádicos (fixos) distribuídos em uma área metropolitana (*Metropolitan Area Network - MAN*). Além disso, essa tecnologia prevê garantias de QoS (*Quality of Service*) para aplicações de dados, voz e vídeo. A garantia de QoS é obtida principalmente pelo emprego de orientação à conexão no nível MAC (*Medium Access Control*), utilização de diferentes classes de serviço, mecanismo AMC (*Adaptive Modulation and Coding*) e da técnica de multiplexação de canal OFDMA (*Orthogonal Frequency Division Multiple Access*).

A implementação da técnica OFDMA é obrigatória em redes WiMAX móveis. Isso acontece porque essa técnica se aplica bem a cenários com mobilidade, pela utilização de subportadoras ortogonais, que tornam a tecnologia menos suscetível a ruídos e interferências no meio de transmissão (BERARDINELLI et al., 2008). Além disso, a tecnologia permite que múltiplos usuários transmitam em um mesmo símbolo OFDMA, provendo uma melhora na distribuição da largura de banda do canal de rádio frequência. Para que isso seja possível, a organização do quadro OFDMA precisa ser realizada nos domínios de tempo e frequência, como uma matriz bidimensional.

Um quadro OFDMA é dividido em dois subquadros: subquadro DL (*downlink*) e subquadro UL (*uplink*). O subquadro DL é utilizado para transmissão de informações em *broadcast*, da Estação Base (*Base Station - BS*) para as Estações Clientes (*Mobile Station - MS*). Já o subquadro UL, é utilizado para envio de informações das MSs para a BS, e precisa ser compartilhado pelas MSs. A organização das informações dentro dos subquadros e a seleção das informações para transmissão não é definida pelo padrão. Essas decisões fazem parte das tarefas de uma arquitetura de escalonamento, que também não são especificadas no padrão IEEE 802.16e (IEEE, 2005).

Pesquisas recentes, (PIZZI; MOLINARO; IERA, 2009), (WANG; IVERSEN, 2008) e (GAN et al., 2009), apresentam soluções para aspectos não definidos na especificação da tecnologia, como escalonamento, CAC (*Connection Admission Control*), policiamento e conformação de tráfego. Dentre esses aspectos, deve-se destacar a importância de considerar a técnica de multiplexação do canal e, conseqüentemente, a estrutura do quadro no projeto de uma arquitetura de escalonamento, como defendido por So-In et al. (SO-IN;

JAIN; TAMIMI, 2009) e Lei *et al.* (LEI *et al.*, 2008). Estratégias de escalonamento considerando a técnica OFDMA são sugeridas por Lin *et al.* (LIN *et al.*, 2008) e So (SO, 2008), além de Pang *et al.* (PANG *et al.*, 2008), que trabalham o escalonamento UL. Além disso, ainda existem questões em aberto, principalmente no escalonamento UL e em relação a aspectos do mecanismo de alocação de banda definido pelo padrão IEEE 802.16e.

Nesse contexto, é apresentada uma arquitetura de escalonamento *uplink* que considera aspectos resultantes do emprego da técnica de multiplexação de canal OFDMA na construção do quadro de transmissão e características do mecanismo de alocação de banda para transmissões UL, definido pelo padrão IEEE 802.16e. Além disso, considera-se a configuração AMC utilizada na transmissão das informações de um mesmo quadro. Esses aspectos diferenciam este trabalho da maior parte das propostas existentes na literatura. A maioria das propostas existentes considera apenas alguns aspectos do quadro OFDMA, como, por exemplo, a unidade mínima de alocação, chamada de *slot*, ou simplesmente não considera aspectos referentes a camada física, o que torna os algoritmos propostos pouco realistas para uma rede sem fio metropolitana.

A arquitetura proposta é definida em dois níveis. No primeiro nível, trata-se do problema da alocação de recursos para requisição de banda ou transmissão de dados, considerando a estrutura do quadro OFDMA. Já o segundo nível, é responsável pela seleção das informações que serão alocadas no intervalo de tempo reservado para transmissão de dados. A seleção das informações para transmissão é realizada com base no emprego de um algoritmo de escalonamento intra-classe e outro inter-classes. O algoritmo intra-classe seleciona a requisição para atendimento dentre as filas pertencentes a uma mesma classe de serviço enquanto o algoritmo inter-classes determina qual classe de serviço será atendida. As requisições são selecionadas para atendimento considerando um critério de urgência com o objetivo de garantir os requisitos de QoS das aplicações e evitar a postergação indefinida no atendimento as filas de menor prioridade.

Uma ferramenta de simulação foi desenvolvida com a finalidade de testar aspectos da arquitetura proposta, analisando o seu desempenho, além de compará-la a outras propostas. Tal ferramenta foi desenvolvida utilizando os ambientes de programação LabVIEW e MATLAB, pois esses ambientes possibilitam acesso a funções pré-definidas e outras facilidades de implementação que auxiliaram no desenvolvimento do ambiente de simulação.

O desempenho da proposta é analisado considerando diferentes modelos de tráfego, com variados requisitos de QoS, o que é caracterizado pelo emprego de cinco classes de serviço. Os resultados são avaliados considerando aspectos como atraso no atendimento, descarte de requisições e utilização do subquadro UL. A avaliação de desempenho é dividida em duas partes: primeiro o comportamento da arquitetura de escalonamento *uplink* proposta é analisado considerando as cinco classes de serviço, o mecanismo de alocação de banda e a estrutura do quadro OFDMA. Depois, o desempenho da proposta é comparado ao de trabalhos relacionados.

O restante deste trabalho está organizado da seguinte forma. No Capítulo 2, descreve-se aspectos relevantes da tecnologia relacionada com a proposta. No Capítulo 3, são abordados os principais trabalhos relacionados ao tema. No Capítulo 4, apresenta-se a arquitetura de escalonamento *uplink* sugerida. No Capítulo 5 é descrita a ferramenta de simulação desenvolvida para análise da proposta. No Capítulo 6, o desempenho da arquitetura proposta é avaliado, considerando o ganho obtido em relação a trabalhos relacionados e aspectos diferenciais da proposta apresentada. Finalmente, no Capítulo 7, são apresentadas conclusões e apontadas algumas direções para futuras investigações.

2 CONTEXTUALIZAÇÃO DA TECNOLOGIA IEEE 802.16E

O padrão IEEE 802.16e (IEEE, 2005) especifica as camadas física e de controle de acesso ao meio para a tecnologia WiMAX, que oferece uma alternativa aos sistemas de acesso tradicionais, como por exemplo, *cable modem*, DSL (*Digital Subscriber Line*) e sistemas 3G. Além disso, essa tecnologia prevê o suporte a aplicações de dados, voz e vídeo com garantias de QoS em abrangências metropolitanas.

A tecnologia suporta a utilização de dois tipos de topologia: ponto-multiponto e *mesh*. Na topologia ponto-multiponto, todo o tráfego da rede é controlado pela BS, não havendo comunicação direta entre as MSs. Já na topologia *mesh*, as MSs podem trocar informações entre si e realizar tarefas de roteamento, sem a necessidade do controle da BS. Além disso, a especificação da tecnologia prevê a utilização de duas técnicas de duplexação do canal: TDD (*Time Division Duplex*) ou FDD (*Frequency Division Duplex*). Na duplexação por tempo, as transmissões DL e UL são alternadas no tempo e um único canal é utilizado. Já na duplexação por frequência, as transmissões DL e UL ocorrem em canais de rádio frequência diferentes. Neste trabalho, considera-se a técnica de duplexação TDD, pois essa técnica possibilita um melhor aproveitamento da faixa de frequência. Além disso, emprega-se a topologia ponto-multiponto.

A garantia de QoS das aplicações suportadas pela rede é obtida pelo emprego de diferentes técnicas como: acesso ao meio controlado pela BS, emprego do mecanismo AMC para garantir controle de erros e confiabilidade na transmissão, tendo em vista que a transmissão em distâncias metropolitanas é mais suscetível aos problemas do meio físico e utilização de classes de serviço para agrupar fluxos com características semelhantes. Além disso, o padrão prevê a utilização da técnica de multiplexação OFDMA, que emprega subportadoras ortogonais, uma característica dos sistemas OFDM (*Orthogonal Frequency Division Multiplexing*), que diminui a interferência entre duas subportadoras adjacentes (BERARDINELLI et al., 2008). A técnica OFDMA permite uma alocação mais flexível da largura de banda disponível pela divisão do canal de rádio frequência em subcanais.

Todos esses aspectos precisam ser considerados no projeto de uma arquitetura de escalonamento para a tecnologia. Por isso, na Seção 2.1 é descrito o mecanismo AMC, utilizado para inserção de informações redundantes aos fluxos de informações. Na Seção 2.2, são abordadas as classes de serviço definidas pelo padrão, as quais devem ser empregadas para definir as estratégias de transmissão permitidas a cada fluxo de dados. Em seguida, na Seção 2.3, são estudadas as principais características da técnica de multiplexação OFDMA e na Seção 2.4 são apresentados os mecanismos de alocação de banda utilizados para controle de acesso ao meio. Maiores detalhes sobre a tecnologia podem ser obtidos em (SO-IN; JAIN; TAMIMI, 2009).

2.1 Mecanismo AMC

O mecanismo AMC deve ser empregado nas informações geradas pelas aplicações e que precisam ser transmitidos pela rede, para garantir maior confiabilidade à transmissão. Por isso, antes de estarem prontos para serem alocados no quadro OFDMA, os bits são agrupados em símbolos de modulação, organizados em constelações. Essa organização é chamada de esquema de modulação QAM (*Quadrature Amplitude Modulation*). Diversas constelações de modulações podem ser utilizadas, sendo que a constelação empregada indica a quantidade de bits (b_{simb}) que podem ser transmitidos em um símbolo de modulação, conforme equação 2.1, onde n indica o número de símbolos da constelação QAM. Depois de modulados, os bits são codificados utilizando técnicas FEC (*Forward Error-Correction*), que provêm detecção e correção de erros através da adição de bits redundantes à informação original (SALMON; OLIVIER, 2007) (KUNST et al., 2009). A combinação de um esquema de modulação e uma codificação FEC é conhecida como perfil de rajada.

$$b_{simb} = \log_2(n) \quad (2.1)$$

A configuração AMC para cada perfil de rajada tem o objetivo de garantir um equilíbrio entre confiabilidade e taxa de transmissão. Isto significa que se for utilizado um perfil de rajada mais robusto, a taxa de transmissão será reduzida. Entretanto, nesse caso, haverá um ganho na confiabilidade da transmissão, pois existirá uma maior capacidade de detecção e correção de erros. Por outro lado, a utilização de um perfil de rajada menos robusto reduzirá a confiabilidade da transmissão, enquanto aumenta a taxa de transmissão, uma vez que uma maior quantidade de dados será transmitida com uma menor quantidade de informações redundantes (GHOSH et al., 2005).

Os perfis de rajada suportados pelo padrão IEEE 802.16e são sete, como pode ser visto na Tabela 2.1. A implementação da modulação BPSK (*Binary Phase Shift Keying*) é obrigatória para transmissão de informações de controle, pois esse é o esquema de modulação mais robusto dentre os previstos para a tecnologia. Para a transmissão de dados, é obrigatória a implementação de QPSK (*Quadrature Phase Shift Keying*) e 16-QAM, para garantir a interoperabilidade entre dispositivos de diferentes fabricantes. Já a implementação da modulação 64-QAM é opcional.

Tabela 2.1: Perfis de rajada previstos pelo padrão IEEE 802.16e

Modulação	Bits por Símbolo QAM (b_{simb})	Codificação
BPSK	1	1/2
QPSK	2	1/2
QPSK	2	3/4
16-QAM	4	1/2
16-QAM	4	3/4
64-QAM	6	2/3
64-QAM	6	3/4

Mudanças no perfil de rajada empregado refletem variações das condições de propagação do canal de rádio frequência, ao longo do tempo (QIN; KUO, 2007). As condições

de propagação são afetadas pelas características do ambiente físico, que podem causar erros na transmissão. Porém, a capacidade de correção dos erros depende da quantidade de informações redundantes transmitidas em um quadro. Dessa forma, variações no perfil de rajada empregado afetam a largura de banda reservada para enviar dados das aplicações. Por isso, é necessário considerar o perfil de rajada empregado na alocação de oportunidades de transmissão de dados ou de requisição de banda (KIM; KIM; SUH, 2007). Essas alocações são atribuídas a uma aplicação, considerando, entre outros fatores, a classe de serviço em que essa aplicação é agrupada.

2.2 Classes de Serviço

O padrão IEEE 802.16e define cinco classes de serviço que representam as características de diferentes grupos de aplicação. Cada fluxo de serviço presente na rede deve ser mapeado para uma dessas classes e a alocação de banda deve ser realizada considerando o conjunto de regras definidos para cada classe. A seguir são descritas as cinco classes de serviço definidas pelo padrão:

- UGS (*Unsolicited Grant Service*): suporta fluxos com características de tempo real, com geração de pacotes de tamanho fixo em intervalos regulares, como, por exemplo, VoIP (*Voice over Internet Protocol*) sem supressão de silêncio (SHI; HU, 2008).
- ertPS (*extended real-time Polling Service*): essa classe de serviço suporta fluxos com características semelhantes aos suportados pela classe UGS. Possui necessidade de transmissão em intervalos constantes, porém com pacotes de tamanho variado. Podem ser classificadas nesta classe aplicações como VoIP com supressão de silêncio ou videoconferência (CICCONETTI; LENZINI; MINGOZZI, 2006).
- rtPS (*real-time Polling Service*): voltada para aplicações com características de tempo real, com fortes restrições quanto ao atraso e *jitter*. Os parâmetros de atraso máximo e *jitter* indicam, respectivamente, a latência máxima e a variação de atraso máximo suportados. Os pacotes são de tamanho variados em intervalos irregulares, como ocorre em aplicações do tipo *streaming* de vídeo ou vídeo *clip* (trata-se de um tipo de vídeo com curta duração) (SO-IN; JAIN; TAMIMI, 2009).
- nrtPS (*non-real-time Polling Service*): para fluxos sem características de tempo real, mas com necessidade de garantia de banda mínima para transmissão. Aplicações como HTTP (*Hypertext Transfer Protocol*), *Instant Messaging* e e-mail podem ser mapeadas para essa classe de serviço (TSAI et al., 2008).
- BE (*Best Effort*): direcionada a aplicações sem qualquer requisito de QoS. Podem ser agrupados nesta classe fluxos de aplicações FTP (*File Transfer Protocol*) e P2P (*Peer-to-Peer*) (TSAI et al., 2008).

Os principais requisitos previstos pelo padrão para cada classe de serviço são relacionados na Tabela 2.2. Nesse contexto, se a classe UGS possuir uma taxa mínima associada, a mesma deverá ser igual a taxa máxima suportada. A taxa máxima representa a largura de banda máxima destinada aos fluxos daquela classe de serviço e pode ser utilizada para policiamento de tráfego, por exemplo. Já o parâmetro priorização de tráfego, dentro da mesma classe de serviço, pode ser utilizado para conceder prioridades diferentes a fluxos

de dados pertencentes a uma mesma classe. Sendo que, essas prioridades podem variar em uma escala de 0 a 7 (IEEE, 2005). Finalmente, o parâmetro de taxa mínima indica que deve haver uma reserva de largura de banda mínima para transmissões de fluxos pertencentes a determinadas classes de serviço.

Tabela 2.2: Principais requisitos de QoS por classe de serviço

Parâmetros de QoS	UGS	ertPS	rtPS	nrtPS	BE
Taxa máxima	X	X	X	X	-
Latência máxima	X	X	X	-	-
<i>Jitter</i>	X	X	-	-	-
Priorização de tráfego na mesma classe	-	-	X	X	-
Taxa mínima reservada	X	X	X	X	-

Considerando a descrição das classes e os requisitos de QoS associados, pode-se verificar uma escala de prioridades entre as mesmas. A classe UGS agrupa as aplicações com maiores restrições e possui mais requisitos de QoS associados, por isso, essa deve ser a classe de maior prioridade para o escalonador, seguida pelas classes, ertPS, rtPS, nrtPS e BE. Essa priorização inicia pela forma como cada classe pode utilizar o mecanismo de alocação de banda. Porém, o mecanismo de alocação de banda sofre variações de acordo com a técnica de multiplexação do canal empregada pelo meio físico. Por isso, as características mais relevantes da técnica de multiplexação OFDMA para o escalonamento são descritas na seção a seguir.

2.3 Multiplexação OFDMA

O padrão IEEE 802.16e prevê a utilização de quatro configurações de camada física: SC (*Single Carrier*), SCa (*Single Carrier adaptive*), OFDM e OFDMA. As duas primeiras especificações de camada física não são multiplexadas, enquanto as duas últimas são baseadas na multiplexação do canal em diversas subportadoras. A organização em subportadoras tem o objetivo de possibilitar a otimização da largura de banda disponível, possibilitando a obtenção de taxas de transmissão mais elevadas. Além disso, a especificação da tecnologia define uma configuração específica para a transmissão em faixas de frequência não licenciadas, chamada de HUMAN (*High-speed Unlicensed MAN*) (KUNST, 2009).

A multiplexação OFDMA é a única de implementação obrigatória para garantia de interoperabilidade em redes WiMAX móveis, de acordo com o fórum de fabricantes. Por isso, essa técnica é considerada neste trabalho.

A técnica OFDMA aplica-se bem a cenários com mobilidade pela utilização de subportadoras ortogonais, que tornam a tecnologia menos suscetível a ruídos e interferências do meio de transmissão (BERARDINELLI et al., 2008). Além disso, OFDMA possui como principal característica o suporte a múltiplos usuários em um mesmo símbolo OFDMA, tal característica é obtida pela divisão do canal sem fio em subcanais. Segundo o padrão IEEE 802.16e, os subcanais são formados por conjuntos de subportadoras. O processo de distribuição das subportadoras em subcanais é conhecido como subcanalização ou zonas de permutação, sendo que cada zona de permutação emprega um método diferente de permutação das subportadoras em subcanais. O padrão especifica sete métodos

de subcanalização: DL PUSC (*Partial Usage of Subchannels*), DL FUSC (*Full Usage of Subchannels*), FUSC opcional, TUSC 1 (*Tile Usage of Subchannels*), TUSC 2, UL PUSC, PUSC opcional. Porém, dentre esses, apenas as técnicas PUSC são de implementação obrigatória, por isso, apenas esse método é considerado neste trabalho.

O processo de subcanalização torna o mecanismo de alocação de banda mais flexível, porque as alocações passam a ser realizadas nos domínios de tempo (símbolos OFDMA) e frequência (subcanais), utilizando uma unidade de alocação mínima chamada *slot* (BACIOCCOLA et al., 2009). Porém, o emprego da técnica PUSC é diferente nos subquadros DL e UL, por isso os chamados DL PUSC e UL PUSC. No DL PUSC, cada *slot* é formado por dois símbolos OFDMA e um subcanal, sendo que cada subcanal DL utiliza 24 subportadoras de dados. Já no UL PUSC, cada *slot* é formado por 3 símbolos OFDMA que utilizam 48 subportadoras de dados, totalizando 16 subportadoras de dados por subcanal UL. A subcanalização e o mecanismo de alocação de banda determinam a estrutura do quadro OFDMA, que pode ser representado como uma matriz bidimensional. Nessa matriz, os eixos x e y representam os domínios de tempo e frequência, respectivamente, como pode ser observado na Figura 2.1. Os componentes dos subquadros DL e UL representados na figura são descritos nas subseções seguintes.

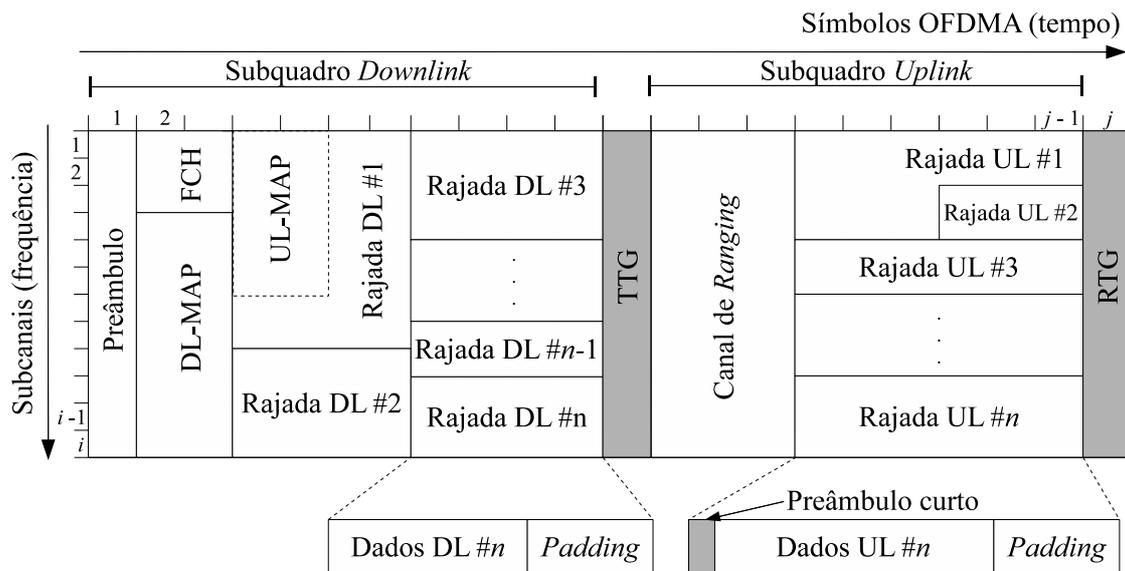


Figura 2.1: Estrutura do quadro OFDMA

2.3.1 Subquadro DL

O subquadro DL inicia com um preâmbulo utilizado para sincronização. Esse preâmbulo é transmitido utilizando todos os subcanais durante um símbolo OFDMA. No segundo símbolo inicia a transmissão do FCH (*Frame Control Header*), utilizando 4 subcanais, como pode ser observado na Figura 2.1. O FCH contém informações sobre a localização e duração da mensagem DL-MAP. Esse mapa faz parte de um conjunto de mensagens de gerenciamento, formado por: DL-MAP, UL-MAP, DCD (*Downlink Channel Descriptor*) e UCD (*Uplink Channel Descriptor*). Os descritores DCD e UCD podem não estar presentes em todos os quadros, pois sua transmissão é periódica e não está relacionada à frequência de geração dos quadros. A mensagem UL-MAP está presente

sempre que a transmissão de um subquadro UL seguir a transmissão de um subquadro DL.

Os mapas DL-MAP e UL-MAP são formados por conjuntos de informações conhecidos como IEs (*Information Elements*), referentes as regiões de dados dos subquadros DL e UL, respectivamente. Cada IE contém a localização, duração, tipo e configuração AMC de uma rajada. As mensagens descritoras de canal, DCD e UCD, contêm informações como código de duração do quadro, descrição dos perfis de rajada utilizados e demais informações sobre configurações físicas para os subquadros DL e UL.

Após as mensagens de gerenciamento, o subquadro DL é preenchido com rajadas de dados. A construção do subquadro DL deve seguir algumas regras definidas pelo padrão e descritas a seguir. Os dados destinados a uma ou mais MSs e que podem ser transmitidos utilizando o mesmo perfil de rajada, devem ser agrupados em uma mesma região de dados, de formato retangular (OHSEKI; MORITA; INOUE, 2007). Na Figura 2.1, cada região de dados do subquadro DL é representada como uma "Rajada DL". A organização dessas rajadas dentro do subquadro DL não é definida pelo padrão. Porém, o algoritmo implementado deve ter o objetivo de alocar as rajadas gerando o menor desperdício de banda possível, evitando espaços vazios entre as rajadas.

Quando a BS não possuir informações suficientes para utilizar todo o subquadro DL, o mesmo será preenchido com informações nulas ou *padding*. Após cada subquadro DL é inserido um TTG (*Transmit/Receive Transition Gap*), para permitir que a BS alterne do modo de transmissão para o modo de recepção em que atuará no subquadro UL.

2.3.2 Subquadro UL

Um subquadro UL é formado por dois tipos de informações: requisições de banda e transmissões de dados. A necessidade da MS de enviar uma requisição de banda pode ser sinalizada no canal de *ranging*. Por isso, um subquadro UL geralmente possui um canal de *ranging*, cujas demais funções são descritas na Subseção 2.4.1.

O canal de *ranging* deve ser formado por um ou mais conjuntos de seis subcanais adjacentes. Sendo que, a alocação do canal de *ranging* no início do subquadro UL, como no exemplo da Figura 2.1, minimiza a interferência na transmissão dos dados, conforme You et al. (YOU; KIM; KIM, 2005).

Após o canal de *ranging* são alocadas as "Rajadas UL", que contêm informações destinadas a uma MS ou a um grupo de MSs (no caso de uma transmissão *multicast*). A alocação das rajadas UL deve ser realizada pela BS, iniciando no primeiro subcanal e primeiro símbolo OFDMA livres, seguindo pelo eixo x até atingir o limite do quadro, ou o fim da zona de permutação. Quando o limite do quadro ou da zona for atingido, a alocação deve continuar no próximo subcanal livre, como exemplificado na Figura 2.2.

De forma semelhante ao que ocorre no subquadro DL, se não houverem informações suficientes para preencher todo o subquadro UL, o mesmo será completado com *padding*. Ao final do subquadro UL é inserido um RTG (*Receive/Transmit Transition Gap*), para permitir que a BS volte a atuar no modo de transmissão.

A técnica de multiplexação de canal empregada define o formato do quadro utilizado para as transmissões, mas influencia também o mecanismo de alocação de banda empregado. Por isso, o mecanismo de alocação de banda utilizado em redes IEEE 802.16e, quando a técnica OFDMA é empregada, é descrito na seção seguinte.

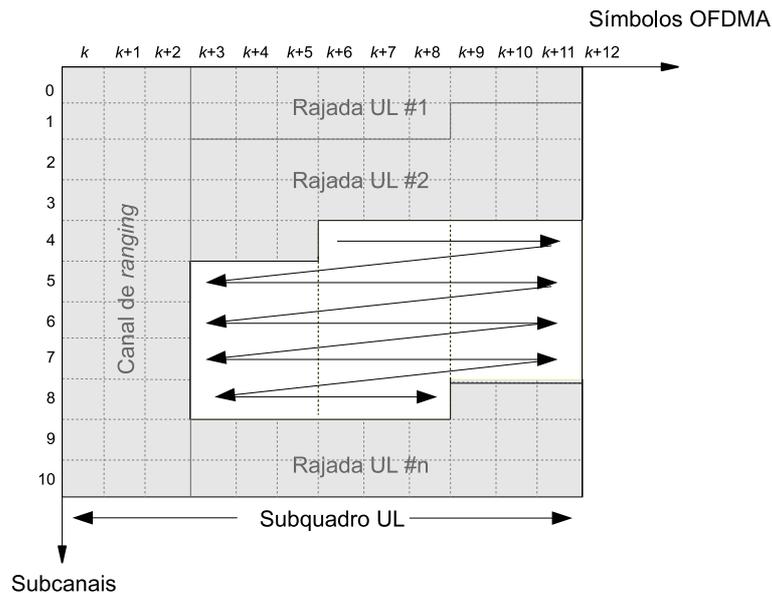


Figura 2.2: Alocação de uma Rajada UL pela BS

2.4 Mecanismos de Alocação de Banda

Os mecanismos de alocação de banda são diferentes para os subquadros DL e UL, devido à forma de acesso aos mesmos. No subquadro DL, apenas a BS transmite informações, já o acesso ao subquadro UL é compartilhado pelas MSs e precisa ser coordenado pela BS. A utilização dos subquadros DL e UL, através do mecanismo de alocação de banda, é definida pelo emprego de um escalonador. O escalonador responsável pelo fluxo DL é localizado na BS e possui acesso direto às filas de dados para as transmissões DL, como pode ser visto na Figura 2.3.

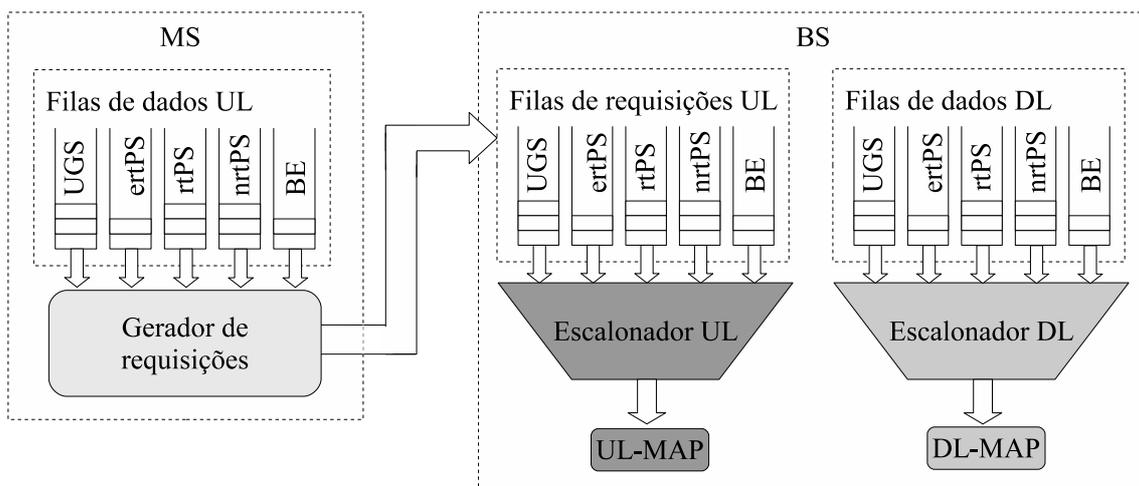


Figura 2.3: Acesso às filas de dados pelos escalonadores DL e UL

Por outro lado, o escalonador responsável pelo gerenciamento do subquadro UL é localizado na BS, e não possui acesso direto às filas de dados UL, localizadas nas MSs. Dessa forma, as MSs precisam informar à BS sobre o estado de suas filas através de

requisições de banda, que são agrupadas de acordo com as cinco classes de serviço e armazenadas na BS para atendimento. Cada requisição de banda possui o CID (*Connection Identifier*) da conexão e a informação sobre o tamanho, em bytes, da oportunidade de transmissão de dados que está sendo solicitada.

As filas de dados DL e UL são formadas por pacotes de dados gerados pelas aplicações e agrupados de acordo com a classe de serviço a que pertencem. Dessa forma, o mecanismo de alocação de banda UL pode ser considerado mais complexo do que o utilizado para o tráfego DL e, por isso, é detalhado nesta Seção.

A alocação de banda UL é um procedimento cíclico composto por duas fases: requisição e concessão de oportunidades de transmissão. Durante a fase de requisição, as MSs enviam à BS solicitações de oportunidades para transmissão de dados. Na fase de concessão, a BS destina porções do quadro para que as MSs enviem suas requisições de banda ou transmitam dados das aplicações. Ambas as fases podem ser divididas em diferentes tipos, descritos nas subseções seguintes.

2.4.1 Oportunidades de Requisição de Banda

As oportunidades concedidas pela BS para que as MSs façam seus pedidos de banda podem ser utilizadas de duas maneiras: com contenção ou livre de contenção. Uma porção do quadro destinada ao envio de requisições com contenção pode ser utilizada por qualquer MS interessada em requisitar banda. Desde que a MS possua fluxos de tráfego pertencentes às classes de serviços autorizadas a utilizar esse método de requisição. Com isso, várias MSs competem pelo meio e as requisições podem colidir. Se duas ou mais requisições colidirem, não serão recebidas pela BS. Ou seja, trata-se de um método de requisição de banda sem qualquer garantia de recebimento.

O período de contenção, em um quadro OFDMA, possui algumas particularidades. Trata-se do canal de *ranging*, que concentra quatro tipos de serviços: solicitação de banda, *ranging* inicial, *ranging* periódico e *handover*. Uma MS sinaliza a necessidade de utilizar algum desses serviços à BS pela informação de um código CDMA (*Code Division Multiple Access*) em uma das oportunidades do canal de *ranging*. A oportunidade de transmissão é escolhida aleatoriamente, assim como o código CDMA. O padrão IEEE 802.16e prevê a existência de 256 códigos CDMA diferentes, distribuídos entre os quatro serviços suportados pelo canal de *ranging*. Nesse contexto, colisões ocorrem quando mais de uma MS escolhe a mesma oportunidade de transmissão e o mesmo código CDMA (LEE; CHO, 2005). Nesse caso, a BS não é capaz de identificar corretamente o código informado e a MS não é atendida. Para diminuir a probabilidade de ocorrência das colisões, um algoritmo de *backoff* deve ser empregado.

O algoritmo de *backoff* previsto pelo padrão IEEE 802.16e é baseado em um algoritmo exponencial binário truncado, sendo que valores mínimo e máximo para a janela de *backoff* (J_b) são definidos pela BS em potências de 2, conforme equação 2.2, onde n varia entre 0 e 15. Sempre que uma MS deseja participar de um processo de contenção deverá escolher aleatoriamente um valor, dentre os permitidos pela janela de *backoff* definida pela BS. Esse valor indica o número de oportunidades de transmissão que a MS deve aguardar antes de iniciar sua participação no processo de contenção. Se ainda assim houver uma colisão quando o código CDMA for informado, as MSs participantes da colisão devem ter o valor máximo da sua janela de *backoff* acrescido em uma unidade, antes de elevá-lo à potência de 2 (SAYENKO; ALANEN; HAMALAINEN, 2007). Esse processo é repetido até que a MS consiga informar um código CDMA com sucesso, ou até que o número de tentativas possíveis seja atingido. O padrão define que um número mínimo de

16 tentativas deve ser suportado, mas não é definido um valor padrão ou máximo.

$$J_b = 2^n \quad (2.2)$$

Quando não ocorrem colisões, a BS deve adotar os procedimentos adequados ao serviço indicado pelo grupo de códigos a que pertence o código recebido. Por exemplo, se a BS receber um código CDMA pertencente ao grupo de códigos reservados para requisição de banda por contenção, deverá alocar uma oportunidade para que a MS envie seu pedido de banda em algum dos próximos subquadros UL. Essa oportunidade é chamada de *IE CDMA* ou *CDMA_allocation*, sendo destinada a apenas uma MS, não havendo mais probabilidade de colisão. Porém, a BS não indica, explicitamente, no *IE CDMA*, a que MS pertence tal alocação. Isso ocorre porque a BS não conhece ainda o CID da MS que informou o código. O *IE CDMA* indica apenas o código e a oportunidade do canal de *ranging* em que o mesmo foi informado. Nesse contexto, as MSs precisam identificar a qual delas pertence a oportunidade recebida, verificando se a oportunidade utilizada e o código CDMA informado pelas mesmas conferem com as informações do *IE CDMA*, como ilustrado pela Figura 2.4. Após a identificação da MS, a mesma gera um pedido de banda que é enviado na alocação indicada pelo *IE CDMA*. O funcionamento dos demais serviços do canal de *ranging* não é descrito nesta Seção por não fazerem parte do escopo deste trabalho.

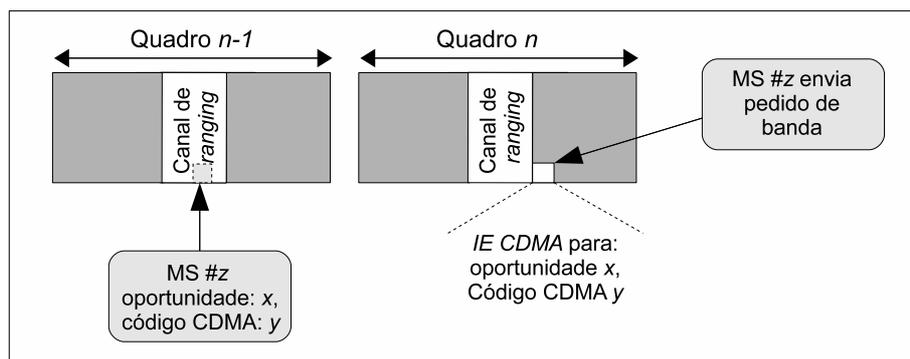


Figura 2.4: Funcionamento do período de contenção

Por outro lado, existem as oportunidades de requisição de banda sem contenção. Nesse caso, uma oportunidade de requisição é destinada a apenas uma MS, por isso, não existe probabilidade de colisão. Esse mecanismo é chamado de requisição por *polling unicast* ou simplesmente *polling*.

Uma oportunidade de *polling* também pode ser concedida a um grupo de MSs (*polling multicast*) ou a todas as MSs que fazem parte da rede (*polling broadcast*). Nesse caso, as MSs participantes do grupo devem empregar o mecanismo de *backoff* descrito anteriormente (ALANEN, 2007). Porém, essas duas variações de *polling* não são de implementação obrigatória de acordo com o padrão IEEE 802.16e.

Uma MS que necessita de oportunidade de transmissão e não recebe uma alocação para enviar sua requisição pode, ainda, enviá-la por *piggyback*. Ou seja, a MS pode utilizar parte de uma oportunidade de transmissão de dados recebida da BS para enviar, além dos dados, uma requisição de banda (KIM; CHO, 2007). Porém, essa estratégia não é de implementação obrigatória segundo a especificação da tecnologia.

A utilização da estratégia de requisição de banda é relacionada à classe de serviço a que pertence a aplicação que necessita de transmissão. Essa utilização é definida na

especificação da tecnologia e pode ser observada na Tabela 2.3, que indica a qual classe de serviço é permitido utilizar cada estratégia do mecanismo de alocação de banda.

Tabela 2.3: Utilização do mecanismo de alocação de banda em cada classe de serviço

Estratégias de alocação de banda	UGS	ertPS	rtPS	nrtPS	BE
<i>Polling unicast</i>	-	X	X	X	X
<i>Polling multicast</i> ou <i>broadcast</i>	-	X	-	X	X
Requisições por contenção	-	X	-	X	X
<i>Piggyback</i>	-	X	-	X	X

A utilização de oportunidades de *polling unicast*, *multicast*, *broadcast* contenção ou *piggyback* pela classe ertPS não indica a necessidade de requisitar oportunidades de transmissão, visto que o escalonador deve proporcionar aos fluxos desta classe, oportunidades de transmissão em intervalos regulares. Porém, o tamanho dessas oportunidades de transmissão é variável e as MSs, com fluxos desse tipo, podem utilizar qualquer estratégia de requisição para informar a BS sobre a necessidade de alterar o tamanho das oportunidades de transmissão de dados recebidas.

É possível verificar também, pela análise da Tabela 2.3, que é permitido às classes nrtPS e BE utilizar os mesmos recursos do mecanismo de alocação de banda. Nesse ponto, não ocorre diferenciação entre essas duas classes. A diferenciação ocorre apenas quanto à priorização dada pelo escalonador ao atendimento das requisições pertencentes a cada uma dessas classes.

2.4.2 Oportunidades de Transmissão de Dados

De forma semelhante ao que ocorre com as oportunidades de requisição, as oportunidades de transmissão de dados podem ser concedidas de duas formas: sem a necessidade de requisição pelas MSs ou concedidas mediante o recebimento de uma requisição. As oportunidades de transmissão de dados concedidas sem requisição são pré-estabelecidas. Nesse caso, quando um fluxo de dados que necessita realizar transmissões em intervalos constantes é aceito na rede, o escalonador deve ser informado sobre a periodicidade e o tamanho das necessidades de transmissão do mesmo. Não existe a necessidade da MS solicitar oportunidades de transmissão para fluxos desse tipo, pois suas transmissões são previamente alocadas pelo escalonador.

Alguns fluxos de dados possuem necessidades de transmissão em intervalos e tamanhos variados. Nesse caso, não é possível informar previamente ao escalonador sobre suas transmissões. Entretanto, o escalonador precisa ser informado sobre o estado das filas de dados das MSs e sobre suas necessidades de oportunidades de transmissão, periodicamente, pelo envio de requisições de banda. As requisições recebidas pela BS são selecionadas para atendimento com o objetivo de garantir os requisitos de QoS das aplicações. Nesse contexto, quando uma requisição é selecionada para atendimento, o escalonador aloca uma oportunidade de transmissão de dados, de tamanho correspondente ao solicitado, para o próximo subquadro UL.

Aspectos como o mecanismo de alocação de banda, classes de serviço, configuração AMC e multiplexação OFDMA, descritos nesse Capítulo têm sido considerados em trabalhos acadêmicos recentes (SO-IN; JAIN; TAMIMI, 2009), (WANG; IVERSEN, 2008), (LIN et al., 2008), abordando redes IEEE 802.16e. Essas características são relevantes

para o projeto de estruturas de provisão de QoS, como escalonadores, CAC, policiamento e conformação de tráfego. Dentre esses trabalhos, foram selecionadas os mais relevantes no escopo de estratégias de escalonamento para serem descritas no próximo capítulo.

3 TÉCNICAS DE ESCALONAMENTO

Estratégias de escalonamento são utilizadas para controlar o acesso a um recurso limitado e compartilhado (GIROUX; GANTI, 1999). No caso de redes de computadores, algoritmos desse tipo são empregadas com o objetivo de garantir os requisitos de QoS das aplicações que utilizam a rede. Nas redes IEEE 802.16e, o escalonamento é utilizado para controlar o acesso ao meio de transmissão, selecionando pacotes de dados para transmissão ou requisições de banda para atendimento.

O objetivo deste capítulo é apresentar as técnicas de escalonamento mais relevantes, considerando o contexto das redes IEEE 802.16e. Mas para isso, inicialmente, na Seção 3.1, são apresentadas conceitos básicos sobre escalonamento. Na Seção 3.2, são analisados alguns trabalhos que justificam a importância de considerar a técnica de multiplexação empregada na definição de estratégias de escalonamento. Finalmente, na Seção 3.3, são descritos os principais trabalhos desenvolvidos nesse contexto.

3.1 Algoritmos de Escalonamento Clássicos

Diversos algoritmos de escalonamento foram propostos e estudados ao longo do tempo. Esses algoritmos são adaptados e dão origem a várias estratégias de escalonamento propostas para tecnologias recentes. Por isso, os principais algoritmos clássicos de escalonamento são apresentados nessa seção, para facilitar a compreensão das arquiteturas de escalonamento descritas nas seções posteriores.

Um dos mais simples algoritmos de escalonamento e que não emprega qualquer tipo de priorização no atendimento, é o algoritmo FIFO (*First-In-First-Out*). Quando uma estratégia desse tipo é empregada, as informações recebidas são inseridas no final da fila, e no início da fila essas são selecionados para transmissão ou atendimento. Disciplinas FIFO podem ser empregadas para garantir que todos os usuários tenham as mesmas oportunidades de transmissão. Porém, não consideram características de geração dos pacotes ou aplicações com requisitos de QoS mais restritos.

A inserção de prioridades nas estratégias de escalonamento permite o favorecimento de alguns usuários ou aplicações em detrimento de outros. A maneira mais simples de considerar prioridades no escalonamento é empregar PQ (*Priority Queueing*), no qual as filas são servidas seguindo sua ordem estrita de prioridades. São selecionados para atendimento os pacotes da fila não vazia com maior prioridade. Uma fila com menor prioridade é atendida apenas quando não existem mais pacotes na fila de maior prioridade (GIROUX; GANTI, 1999). Essa é uma estratégia de implementação considerada simples, porém, pode levar a postergação indefinida no atendimento das filas de menor prioridade.

A postergação indefinida pode ser evitada com o emprego de outras disciplinas de escalonamento como RR (*Round Robin*), que atende todas as filas de forma cíclica. Cada

fila é atendida durante um intervalo de tempo calculado de acordo com o tamanho médio dos pacotes para transmissão. Essa disciplina proporciona o isolamento das filas, porém, se o tamanho dos pacotes para transmissão variam muito em relação ao tamanho médio estimado, uma fila pode receber mais ou menos oportunidades do que o necessário (ASH, 2007). Variações do algoritmo RR podem incluir a atribuição de pesos às filas, em um algoritmo conhecido como WRR (*Weighted Round Robin*). Para tratar pacotes com tamanhos variados, foi proposto o algoritmo DRR (*Deficit Round Robin*), para o qual não é necessário conhecer previamente o tamanho médio dos pacotes. Nesse algoritmo, o pacote do início da fila é escalonado se o seu tamanho não for superior ao tamanho máximo definido. Se o tamanho do pacote for superior, o tamanho máximo é adicionado ao *deficit* daquela fila. Quando a fila possuir um *deficit* que somado com o tamanho máximo seja superior ao tamanho do pacote, então o mesmo é selecionado para atendimento.

Uma alocação de recursos mais eficiente do que a obtida pelo emprego de RR, pode ser alcançada pelo emprego da técnica WFQ (*Weighted Fair Queueing*). O algoritmo WFQ pode ser considerado uma implementação prática do conceito de GPS (*Generalized Processor Sharing*), que divide a largura de banda disponível entre as filas de acordo com o peso atribuído a cada uma delas. Porém, o conceito de GPS não pode ser empregado na prática, pois assume que os pacotes são infinitamente divisíveis e que todas as filas não vazias podem ser servidas ao mesmo tempo. Diferentemente do que ocorre com o WFQ que atende cada uma das filas de forma cíclica, durante um tempo proporcional ao peso atribuído a cada fila. Entretanto, WFQ não prevê o processamento parcial de pacotes. Isso ocorre porque o mecanismo WFQ associa a cada pacote um *finish number* que corresponde ao tempo em que o processamento do pacote deveria ser concluído. O pacote da fila escolhida para atendimento, com o menor *finish number* é selecionado (COMER, 2004). A desvantagem no emprego do algoritmo WFQ está no fato de que não há uma estratégia definida para a atribuição dos pesos das filas, e a utilização de pesos pouco justos pode resultar em um escalonamento ineficiente.

Para o escalonamento de fluxos com características de tempo real, existem algoritmos baseados em tempo, como por exemplo, EDF (*Earliest Deadline First*). Nesse caso, um *deadline* é atribuído a cada pacote ou requisição da fila, que indica o seu tempo máximo para atendimento (CHENG, 2002). Após esse intervalo, a informação deve ser descartada, pois não é mais válida. Dessa forma, a cada ciclo de escalonamento é selecionado para atendimento o pacote ou requisição mais próximo de atingir o seu *deadline*.

Os algoritmos listados nessa seção dão origem a algoritmos adaptados, por exemplo, as características da técnica de multiplexação empregada pela camada física de uma tecnologia. Pois, a técnica de multiplexação utilizada influencia diretamente na forma como as informações serão distribuídas no quadro da camada MAC.

3.2 Técnicas de Multiplexação de Canal e Escalonamento

A importância de considerar as técnicas de multiplexação do canal no desenvolvimento de uma arquitetura de escalonamento para o padrão IEEE 802.16e é demonstrada por So-In, Jain e Tamimi (SO-IN; JAIN; TAMIMI, 2009) sob diferentes aspectos. Primeiro, os autores analisam as características chave da tecnologia e aspectos de projeto que devem ser considerados na proposta de uma estratégia de escalonamento DL e UL. Depois, realizam um levantamento dos principais trabalhos publicados, recentemente, sobre escalonamento para esta tecnologia.

Os autores acreditam que os algoritmos clássicos de escalonamento não podem ser

empregados diretamente em uma arquitetura de QoS para a tecnologia WiMAX, devido a especificidades da tecnologia como: utilização da técnica de multiplexação OFDMA, mecanismo de alocação de banda e unidade de alocação mínima (*slot*), definida nos domínios de tempo e frequência. Dessa forma, os algoritmos clássicos devem ser modificados para o emprego nesse contexto ou novas estratégias devem ser definidas.

Em um comparativo entre as funções de escalonamento, considerando as principais técnicas de multiplexação suportadas pela tecnologia, OFDM e OFDMA, os autores salientam que as mesmas estratégias não podem ser, diretamente, empregadas em ambas as técnicas de multiplexação. Pois, para OFDM, a tarefa do escalonador consiste em decidir quais porções do quadro serão dedicadas a cada MS, realizando essa alocação no domínio do tempo. Já para OFDMA, a alocação torna-se mais complexa, pois o escalonador destina porções do quadro para MSs, considerando não somente o domínio de tempo, mas também de frequência, com o objetivo de obter o menor desperdício de banda possível.

As diferenças entre o escalonamento DL e UL são evidenciadas pela descrição dos subquadros, pela forma de acesso ao meio em cada subquadro e pelo mecanismo de alocação de banda. Por isso, descreve-se a necessidade de três escalonadores distintos: dois escalonadores localizados na BS, sendo um para o tráfego DL e outro para o tráfego UL, e um terceiro escalonador em cada MS, responsável por distribuir, entre as conexões da MS, as oportunidades de transmissão recebidas da BS. Cada um desses escalonadores possui características diferentes, mas os autores criaram uma classificação em dois grupos, que engloba as principais propostas recentes de escalonadores, para a tecnologia WiMAX. Esses grupos são:

- Escalonadores que não consideram o meio físico: esse tipo de estratégia é considerada mais simples, pois o algoritmo não leva em conta as características da camada física em suas alocações. Exemplos dessa estratégia são o emprego dos algoritmos: RR, WFQ, algoritmos baseados no atraso como EDF e PQ. Apesar de bastante simples, os cenários considerados nessa classe de propostas não refletem de forma realista o ambiente de uma rede IEEE 802.16e.
- Escalonadores que consideram o meio físico: esses escalonadores consideram as características da camada física. Isso é feito, por exemplo, pela análise da razão sinal x ruído, ou pelo perfil de rajada empregado nas transmissões. Um exemplo dessa classes de escalonadores é o algoritmo M-LWDF (*Modified Largest Weighted Delay First*), que provê suporte aos requisitos de QoS pela garantia de taxas de transmissão. O algoritmo mantém o atraso das requisições em limites menores do que os definidos para cada classe de serviço, pela atribuição de probabilidades diferentes para cada usuário (PARAG; BHASHYAM; ARAVIND, 2005). Outro exemplo de estratégia desse tipo é o algoritmo PFS (*Proportional Fairness Scheme*), que tem o objetivo de garantir justiça a longo prazo no atendimento das requisições (WU; MO; WANG, 2007).

A importância das técnicas de multiplexação de canal empregadas no escalonamento é evidenciada no trabalho de Lei *et al.* (LEI *et al.*, 2008), que desenvolvem um modelo analítico para avaliação do desempenho de sistemas OFDM e OFDMA. O cenário analisado utiliza uma estratégia de escalonamento oportunístico, em nível de fluxos de serviço, para obter um partilhamento justo (*fair sharing*) dos recursos. A proposta estende o modelo teórico de escalonamento PS (*Processor Sharing*), sugerido para sistemas com uma única portadora, para um sistema multi-portadoras, onde a taxa total de serviço varia de acordo com o número total de fluxos na rede.

Utilizando o modelo desenvolvido, os autores calculam o "ganho" obtido no escalonamento utilizando partilhamento proporcionalmente justo (*proportional fair* - PF) em relação ao emprego de um algoritmo RR, nos sistemas OFDM e OFDMA. O ganho é calculado como a razão entre a vazão média obtida com o emprego do algoritmo PF e a vazão média obtida pelo emprego do algoritmo RR. Os resultados mostram que os ganhos obtidos no sistema OFDMA são sempre maiores, ou iguais, aos obtidos no sistema OFDM.

Através desses trabalhos, percebeu-se a necessidade de considerar as condições da camada física na arquitetura *uplink* proposta. Por isso, passam a ser estudadas as estratégias de escalonamento que consideram o emprego da técnica de multiplexação OFDMA na camada física.

3.3 Escalonamento Considerando OFDMA como Técnica de Multiplexação do Canal

Estratégias de escalonamento para redes IEEE 802.16e, considerando OFDMA como técnica de multiplexação de canal devem considerar as características de construção dos subquadros DL e UL em suas propostas, além do mecanismo de alocação de banda. O escalonamento DL é independente do escalonamento UL, e vice-versa, por isso ambos podem ser propostos separadamente. Entretanto, alguns autores apresentam estratégias híbridas, que poderiam ser adaptadas tanto para o escalonamento DL quando para o escalonamento UL. Algumas dessas abordagens, descritas nesta seção, apresentam aspectos relevantes para o escalonamento em redes IEEE 802.16e. Mas, tais propostas apresentam limitações para ambos os tráfegos, DL e UL, pois negligenciam aspectos característicos de cada subquadro. As propostas que apresentam algoritmos exclusivamente para o tráfego DL deixam de ser detalhadas aqui por não fazerem parte do escopo deste trabalho.

Para fornecer uma visão geral das principais pesquisas sobre escalonamento considerando OFDMA, são apresentadas nesta seção propostas híbridas, e portanto mais genéricas, para escalonamento DL e UL, na subseção 3.3.1 e estratégias específicas para escalonamento UL na subseção 3.3.2.

3.3.1 Estratégias Híbridas Considerando OFDMA

As propostas híbridas, para escalonamento DL e UL apresentadas nesta subseção, foram selecionadas porque abordam aspectos importantes do escalonamento UL, tratados também na arquitetura de escalonamento proposta neste trabalho.

Nesse contexto, Lin *et al.* (LIN *et al.*, 2008) desenvolvem um algoritmo de escalonamento para redes IEEE 802.16e utilizando a técnica de multiplexação de canal OFDMA. O algoritmo proposto é chamado de HUF (*Highest Urgency First*) e aloca oportunidades de transmissão nos subquadros DL e UL de acordo com um pré-determinado fator U, que considera urgência, prioridade e justiça no atendimento dos usuários.

A proporção do quadro destinada aos subquadros DL e UL é dinâmica e calculada de acordo com as necessidades de transmissão em cada sentido. Já a capacidade de transmissão e o número de *slots* em cada quadro, são calculados de acordo com a duração do quadro e com o perfil de rajada utilizado. Os autores não descrevem como esse cálculo é realizado, fornecem apenas a estratégia para o cálculo da capacidade de transmissão de cada *slot*, que é apresentada pela equação 3.1. Nesse caso, a variável *Mod_bits* indica o número de bits suportados em cada símbolo de modulação QAM, enquanto a variável T_c expressa a taxa de codificação empregada.

$$bytes_por_slot = \frac{48 * Mod_bits * T_c}{8} \quad (3.1)$$

As requisições recebidas são armazenadas em filas de acordo com a classe de serviço a que pertencem. Sendo que, o número de *slots* necessários para atendimento de cada requisição é dado pela equação 3.2. Onde, a variável Tam_req indica o tamanho da oportunidade de transmissão solicitada, em bytes.

$$num_slots = \frac{Tam_req}{bytes_por_slot} \quad (3.2)$$

Para as classes com requisitos de tempo real, os autores definem um *deadline* ou prazo máximo para atendimento das requisições, calculado de acordo com a equação 3.3. Nesse contexto, a variável LM indica a latência máxima definida para o fluxo, e a variável DQ indica a duração do quadro. Para fluxos de serviço que não possuem características de tempo real ou requisito de latência máxima, assume-se como *deadline* um valor equivalente ao maior valor entre os *deadlines* de todas as requisições inseridas na fila para atendimento.

$$deadline = \frac{LM}{DQ} \quad (3.3)$$

O cálculo do *deadline* é utilizado da seguinte forma: a cada quadro que a requisição permanece na fila aguardando atendimento, o valor do *deadline* é decrementado em uma unidade. Requisições com *deadline* igual a zero indicam violação do requisito de latência máxima daquele fluxo e a requisição é descartada. Para evitar esse tipo de violação, o algoritmo prevê a alocação prioritária das requisições com *deadline* igual a um, independente da classe de serviço a que a requisição pertence. A largura de banda restante, após essa alocação, é dividida entre as demais requisições considerando um fator U , que indica a urgência no atendimento. O fator U é calculado pela equação 3.4, na qual, a variável P indica a prioridade do fluxo a que pertence a requisição e pode variar entre 0 e 7, sendo que o maior valor é atribuído ao fluxo que deve receber a maior priorização.

$$fator_U = \frac{num_slots * P + 1}{deadline} \quad (3.4)$$

Após o cálculo do fator U , um valor médio é calculado para a fila de cada classe de serviço, conforme equação 3.5. Onde, a variável n indica o número de requisições na fila. É selecionada para atendimento a requisição mais antiga da fila que possuir o maior valor médio para os fatores U . O processo se repete, enquanto houverem *slots* disponíveis para alocação.

$$media = \frac{\sum_{i=1}^n fator_U}{n} \quad (3.5)$$

Por outro lado, Jae-Woo So (SO, 2008) realiza uma análise de desempenho das transmissões DL e UL considerando serviços VoIP. Para isso, o autor realiza uma análise mais detalhada da construção dos subquadros DL e UL do que a realizada por Lin *et al.*. Além de utilizar um esquema de escalonamento baseado no algoritmo FIFO.

A porção do quadro destinada aos subquadros DL e UL é dinâmica, como na proposta de HUF. As oportunidades de transmissão do subquadro UL são alocadas primeiro, e então, a parcela restante do quadro é utilizada para transmissões DL. O modelo do sistema

utilizado para a análise de desempenho considera a utilização de compressão nas mensagens DL-MAP e UL-MAP, o que é permitido pelo padrão IEEE 802.16e, mas não é obrigatório. Parâmetros como: número de símbolos por subquadro e número de subcanais não são derivados pelo autor, entretanto, são utilizados parâmetros empregados em um sistema real.

O autor considera a unidade de alocação mínima, o *slot*, na sua estratégia de preenchimento do quadro. Para isso, o número de *slots* no subquadro DL ($N_{slots,d}$) é obtido pelo emprego da equação 3.6.

$$N_{slots,d} = (N_{simb,d} - 1)/2 * N_{subc,d} \quad (3.6)$$

O número de *slots* do subquadro UL ($N_{slots,u}$), por outro lado, pode ser obtido a partir da equação 3.7.

$$N_{slots,u} = (N_{simb,u} - 3)/3 * N_{subc,u} \quad (3.7)$$

As requisições são escolhidas para atendimento empregando uma disciplina FIFO entre as filas de requisições dos diferentes usuários. Esse algoritmo é empregado, segundo o autor, porque provê atrasos e *jitter* constantes para os serviços VoIP.

De forma semelhante ao que ocorre na proposta de HUF, o trabalho de Jae-Woo So não considera as diferenças entre os subquadros DL e UL ao propor um modelo semelhante para ambos os subquadros. Outro aspecto relevante é a escolha do algoritmo FIFO, que não permite diferenciação entre os usuários e pode não garantir os requisitos de QoS dos serviços VoIP quando outras aplicações estão presentes na rede.

3.3.2 Escalonamento UL Considerando OFDMA

Aspectos específicos do escalonamento UL são considerados em alguns trabalhos. Considerar essas características é fundamental para propor um mecanismo de escalonamento UL adaptado a tecnologia IEEE 802.16e. Por isso, alguns dos principais trabalhos desenvolvidos, considerando as especificidades do escalonamento UL, são descritos nessa subseção.

Nesse contexto, o estudo de Pang *et al.* (PANG et al., 2008) resulta em um algoritmo de alocação de recursos guloso (*greedy*) de baixa complexidade para o subquadro UL. O mecanismo proposto considera apenas a classe rtPS, assim, apenas requisições por *polling* são tratadas. A estratégia sugerida consiste em alocar primeiro oportunidades de *polling* e utilizar o restante do subquadro UL para atender as requisições, escolhidas de acordo com a seguinte ordem de critérios:

1. Escolher o usuário com os requisitos de QoS mais urgentes.
2. Escolher o usuário que possui as melhores condições de canal para transmissão.
3. Escolher o usuário com a pior situação de pacotes na fila. Este critério é empregado quando muitos usuários possuem as mesmas condições para o canal de transmissão.

O desempenho da proposta é comparado com a utilização dos algoritmos RR e EDF. Para os cenários demonstrados pelos autores, o algoritmo proposto garante melhores resultados quanto a taxa de transmissão. Comparativamente, a proposta elaborada proporciona um menor atraso no atendimento das requisições.

Tarchi, Fantacci e Bonciani (TARCHI; FANTACCI; BONCIANI, 2008), em sua proposta de alocação de recursos, consideram oportunidades de requisição e de transmissão

de dados. Entretanto, apenas as oportunidades de requisição por contenção são consideradas.

O mecanismo proposto considera aplicações do tipo VoIP e BE. Para os fluxos VoIP emprega-se o algoritmo EDF. Já para a classe BE, quatro estratégias de escalonamento são consideradas: WFQ, onde o peso de cada fila é calculado dinamicamente de acordo com o número de requisições da fila, PQ, RR e PQ com pesos, para o qual são criados três pesos distintos, atribuídos às filas de acordo com a sua prioridade.

A avaliação do desempenho é realizada em relação aos quatro algoritmos previstos para a classe BE. Porém, os autores não concluem qual algoritmo seria a melhor opção, visto que, os algoritmos testados apresentam desempenhos superiores ou inferiores dependendo do critério analisado. A análise realizada também considera o impacto da variação do perfil de rajada no número de pacotes aguardando para serem transmitidos, mostrando a influência da utilização de um perfil mais ou menos robusto no desempenho do escalonador.

Por outro lado, o trabalho de Gan *et al.* (GAN et al., 2009) propõe um algoritmo de escalonamento com o principal objetivo de garantir uma solução com bom desempenho e de fácil implementação. Por isso, os autores propõem um algoritmo não-ótimo, chamado de DMIA (*Dynamic MCS and Interference Aware Scheduling*), onde MCS significa *Modulation and Coding Scheme*.

O algoritmo DMIA aloca banda utilizando a seguinte ordem de prioridades: UGS > ertPR > rtPS > nrtPS > BE. O algoritmo evita a postergação infinita no atendimento das filas de menor prioridade, além de reduzir a perda de pacotes pelas variações na configuração AMC e interferências do meio de transmissão.

O funcionamento do escalonador DMIA inicia com todas as filas não vazias sendo analisadas para consultar o tamanho das requisições aguardando atendimento. Com base nesse valor, na configuração AMC empregada e em um terceiro componente para cada classe de serviço, um *quantum* é calculado. Esse terceiro componente considera os seguintes fatores:

- Para as classes UGS, ertPS e rtPS, a taxa máxima de transmissão suportada para cada classe.
- Na classe nrtPS, a taxa mínima reservada para o tráfego de uma conexão dessa classe.
- Para a classe BE, o intervalo corrente de chegada de pacotes para conexões dessa classe.

No primeiro ciclo de escalonamento, a largura de banda disponível para cada classe de serviço não pode exceder ao *quantum* daquela classe. Para compensar a baixa prioridade no atendimento atribuída as classes nrtPS e BE, se houver largura de banda excedente após o primeiro ciclo de escalonamento, o segundo ciclo considera apenas essas duas classes.

O desempenho dessa proposta é analisado comparativamente aos algoritmos RR, WRR, e PQ+PF (*Proportional Fair*). Pelos testes realizados, os autores concluem que o algoritmo DMIA possibilita um desempenho superior aos demais algoritmos quanto a vazão. Por outro lado, o algoritmo evita a postergação indefinida no atendimento da fila BE e garante os requisitos de atraso máximo das filas de tempo real, quando existe largura de banda disponível. Além disso, o algoritmo possibilita uma menor taxa de pacotes perdidos

devido ao ajuste dinâmico do algoritmo a configuração AMC utilizada. Esse comportamento evita que as MSs recebam oportunidades para transmissão de informações quando as condições do canal são insuficientes para a transmissão.

O estudo das características da tecnologia IEEE 802.16e quando a técnica OFDMA é empregada, permite agrupar os aspectos relevantes para uma arquitetura de escalonamento em 4 categorias principais:

- Características do quadro de transmissão de acordo com a técnica de multiplexação
- Utilização do mecanismo de alocação de banda
- Emprego das cinco classes de serviço previstas pelo padrão IEEE 802.16e
- Configuração AMC empregada na transmissão das informações

Um resumo das características abordadas nos trabalhos relacionados listados na seção 3.3 é apresentado na Tabela 3.1. Nessa Tabela são assinalados os aspectos considerados e negligenciados em cada trabalho, além das características não analisadas completamente em cada proposta.

Tabela 3.1: Aspectos considerados pelos trabalhos relacionados

Proposta	Técnica de Multiplexação OFDMA	Mecanismo de alocação de banda	Cinco Classes de Serviço	Configuração AMC
Lin <i>et al.</i>	Parcialmente	Não	Não	Sim
Jae-Woo So	Sim	Parcialmente	Não	Não
Pang <i>et al.</i>	Parcialmente	Parcialmente	Não	Não
Tarchi, Fantacci e Bonciani	Parcialmente	Parcialmente	Não	Não
Gan <i>et al.</i>	Parcialmente	Não	Sim	Sim

Os trabalhos encontrados na literatura não abordam completamente os 4 grupos de características consideradas relevantes para o escalonamento em redes IEEE 802.16e baseadas em OFDMA. A maioria desses trabalhos considera apenas alguns aspectos, realizando simplificações de características importantes da tecnologia. Verifica-se que existem questões em aberto, principalmente no escalonamento do tráfego UL, por isso, uma arquitetura de escalonamento para o tráfego UL é apresentada no capítulo seguinte. O principal diferencial da arquitetura proposta consiste na utilização conjunta dos 4 grupos de características.

4 UMA ARQUITETURA DE ESCALONAMENTO *UPLINK*

Uma arquitetura de escalonamento hierárquica para transmissões *uplink* em redes WiMAX baseadas em OFDMA é proposta neste trabalho. A abordagem proposta considera o mecanismo de alocação de banda e a organização do subquadro UL, além da técnica de duplexação TDD e de uma topologia PMP. O objetivo é atender as funções do escalonamento UL citadas na Seção 3.3.2 considerando a configuração AMC utilizada para transmissão das informações em um mesmo quadro.

A estratégia utilizada para o desenvolvimento deste trabalho é hierárquica, empregando dois níveis que devem ser executados de forma sequencial. A divisão do escalonamento em dois níveis é inspirada no trabalho de Cohen e Katzir (COHEN; KATZIR, 2008), porém, essa proposta considera apenas o escalonamento DL. A divisão em dois níveis é realizada de acordo com as funcionalidades e com o objetivo de cada nível, possibilitando ainda que cada nível seja estudado e, se houver necessidade, modificado de forma independente.

O primeiro nível (*Nível 1 - Alocação de recursos*), trata o problema da alocação dos recursos disponíveis para transmissões UL. Esses recursos devem ser divididos considerando a capacidade de transmissão de cada *slot* UL e, então, alocados para oportunidades de requisição de banda ou de transmissão de dados pelas MSs. O segundo nível (*Nível 2 - Gerenciamento de filas*) é responsável pelo gerenciamento das filas de requisições de banda UL enviadas pelas MSs em quadros anteriores, que são armazenadas e aguardam atendimento pela BS.

A utilização da arquitetura de escalonamento proposta no contexto de uma rede IEEE 802.16e, considerando OFDMA como técnica de multiplexação de canal, pode ser visualizada na Figura 4.1. As requisições de banda recebidas em subquadros UL são armazenadas na BS em filas de acordo com as cinco classes de serviço definidas pelo padrão. As necessidades de transmissão de cada uma das classes precisam ser organizadas para atendimento nos próximos subquadros UL. Essa organização é realizada pela arquitetura de escalonamento proposta, que após a execução dos dois níveis de escalonamento fornece um mapa do próximo subquadro UL.

O UL-MAP, localizado no subquadro DL, é formado por IEs dos tipos: *polling*, *IE CDMA* e dados. Os IEs do tipo *polling* contém oportunidades de requisição de banda por *polling*. Os IEs do tipo *IE CDMA* contém oportunidades de requisição de banda concedidas em resposta a códigos CDMA informados no canal de *ranging* de quadros anteriores. Finalmente, os IEs do tipo dados contém oportunidades de transmissão de dados concedidas às MSs.

Cada MS é responsável por tratar os IEs destinados ao seu CID, fornecendo os dados para transmissão ou uma requisição de banda, de acordo com o tipo de oportunidade recebida. Dessa forma, o subquadro UL seguinte é formado por informações fornecidas

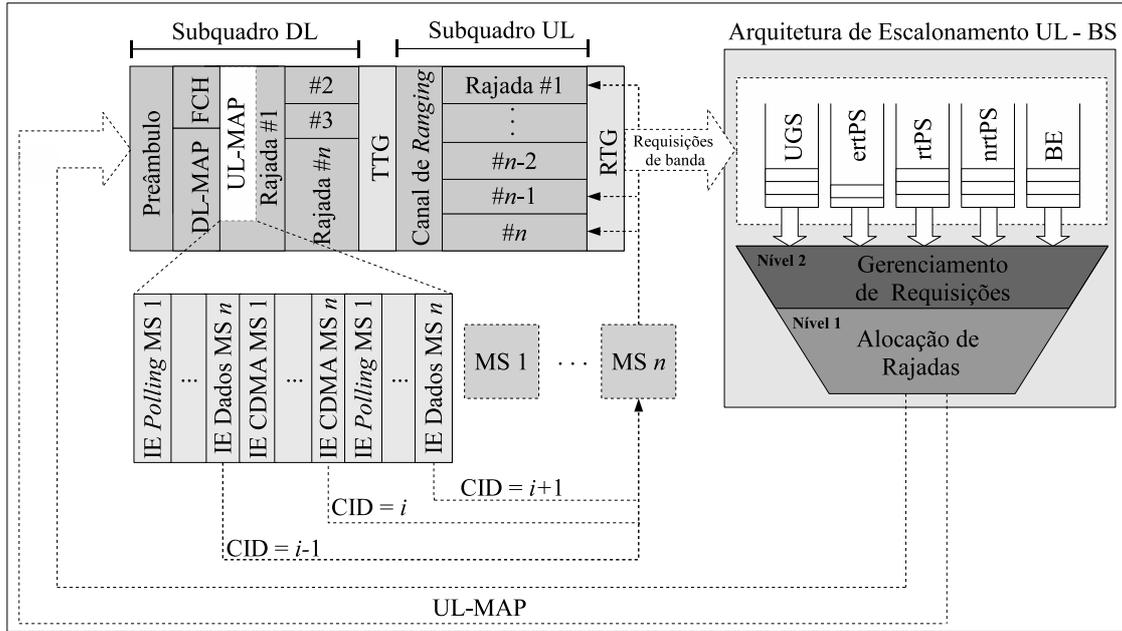


Figura 4.1: Arquitetura de escalonamento *uplink*

pelos MSs, transmitidas nas oportunidades indicadas pela UL-MAP. Observa-se que esse processo é cíclico e deve ser repetido para a geração de cada quadro.

A alocação de recursos (*Nível 1*) e o gerenciamento de filas (*Nível 2*) devem ser empregados para garantir os requisitos de QoS das aplicações agrupadas nas classes de serviço. A utilização do mecanismo de alocação de banda realizada em cada classe de serviço precisa ser apresentada. Dessa forma, a Subseção 4.1 descreve a forma de utilização do mecanismo de alocação de banda UL permitida a cada classe de serviço e as Subseções 4.2 e 4.3 descrevem, respectivamente, o *Nível 1* e o *Nível 2* da arquitetura proposta.

4.1 Utilização do Mecanismo de Alocação de Banda UL

O padrão IEEE 802.16e define as opções de utilização das estratégias previstas pelo mecanismo de alocação de banda UL, de acordo com cada classe de serviço. Na Tabela 4.1 são resumidos os mecanismos de alocação de banda implementados na arquitetura proposta, dentre as possibilidades previstas pelo padrão.

Tabela 4.1: Utilização do mecanismo de alocação de banda UL na arquitetura proposta

Estratégias de alocação de banda	UGS	ertPS	rtPS	nrtPS	BE
Oportunidades de transmissão não solicitadas	X	X	-	-	-
<i>Polling unicast</i>	-	-	X	-	-
Requisições por contenção	-	X	-	X	X

A Tabela 4.1 indica que as classes UGS e ertPS recebem oportunidades de transmissão, sem necessidade de requisição. Isso ocorre para atender as especificações do padrão IEEE 802.16e e para suprir as necessidades de transmissão em intervalos constantes, características de aplicações classificadas nesses dois grupos. É permitido à classe ertPS

informar código CDMA no canal de *ranging* para sinalizar a BS sobre a necessidade de alteração no tamanho das oportunidades de transmissão de dados que vêm sendo concedidas àquela conexão ertPS. Essa permissão é representada na Tabela 4.1 pela possibilidade das classes ertPS utilizarem requisições por contenção.

A classe rtPS é servida por oportunidades de requisição por *polling unicast* para garantia dos requisitos de atraso das aplicações classificadas nesse grupo. Já as classes nrtPS e BE devem utilizar apenas oportunidades de requisição por contenção. Tendo em vista que, requisições por contenção geram uma menor sobrecarga de informações de controle na rede. Essas oportunidades são utilizadas pelas MSs apenas quando as mesmas possuem conexões, pertencentes as classes nrtPS e BE, com informações aguardando transmissão. A utilização de requisições por contenção e, portanto, sem garantia de atendimento, pelas conexões nrtPS e BE é possível porque as aplicações agrupadas em ambas as classes não possuem fortes requisitos de QoS.

4.2 Alocação de Rajadas

O *Nível 1* da arquitetura proposta é responsável pela organização das rajadas a serem transmitidas no subquadro UL. Essas rajadas são alocadas considerando os domínios de tempo (símbolos OFDMA) e frequência (subcanais), além da configuração AMC utilizada na transmissão do quadro. Por isso, a composição dos símbolos OFDMA e dos subcanais, segundo o padrão IEEE 802.16e, é apresentada a seguir.

Cada símbolo OFDMA é formado por dois períodos distintos: o tempo de guarda (T_g) e o tempo de transmissão efetiva (T_b). Dessa forma, a duração total de um símbolo OFDMA (T_s) é dada pela equação 4.1.

$$T_s = T_b + T_g \quad (4.1)$$

O T_g é formado pela cópia cíclica da porção final do símbolo OFDMA, gerando o CP (*Cyclic Prefix*), conforme pode ser observado na Figura 4.2. O T_g é responsável pelo sincronismo entre os símbolos OFDMA no receptor e por minimizar os efeitos de fenômenos físicos como: desvanecimento, atraso por multi percurso e interferência entre símbolos adjacentes (KUNST, 2009).

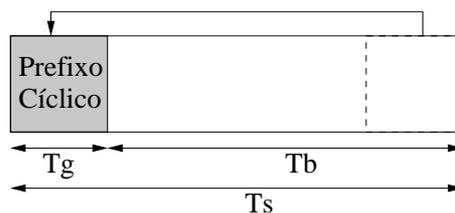


Figura 4.2: Estrutura de um símbolo OFDMA

O tamanho (G) do CP pode variar entre 1/4, 1/8, 1/16 e 1/32, conforme o padrão IEEE 802.16e, sendo que a proporção indica o tempo destinado ao CP em relação a porção do símbolo destinada a transmissão de dados úteis (T_b). Ou seja, quando a BS especifica a utilização de CP igual a 1/4, significa que 25% da duração do símbolo será destinada a transmissão de informações redundantes.

A duração efetiva do CP é calculada considerando a proporção de informações redundantes empregada (G) e o tempo utilizado para transmissão de informação úteis (T_b),

como pode ser visto na equação 4.2.

$$T_g = G.T_b \therefore G = \frac{T_g}{T_b} \quad (4.2)$$

Por outro lado, o intervalo do símbolo OFDMA destinado a transmissão de informações relevantes, T_b , é obtido em relação ao espaçamento existente entre as subportadoras ortogonais (Δ_f), conforme equação 4.6. Sendo que, o valor de Δ_f é definido em unidades de Hertz (Hz).

$$T_b = 1/\Delta_f \quad (4.3)$$

O espaçamento entre as subportadoras, Δ_f , é obtido pela equação 4.4. Nesse caso, N_{FFT} representa o número total de subportadoras, incluindo as subportadoras de dados, pilotos e as destinadas aos intervalos de guarda.

$$\Delta_f = \frac{F_s}{N_{FFT}} \therefore T_b = \frac{N_{FFT}}{F_s} \quad (4.4)$$

A frequência de amostragem (F_s) é definida em função da largura de banda (L_b) disponível no canal de rádio frequência e do fator de amostragem (n) associado, conforme equação 4.5.

$$F_s = \left\lfloor \frac{L_b \cdot n}{8000} \right\rfloor \cdot 8000 \quad (4.5)$$

Os fatores de amostragem, n , associados a cada largura de banda, de acordo com o padrão IEEE 802.16e podem ser vistos na Tabela 4.2. Um fator de amostragem é associado a uma largura de banda múltipla de cada valor definido na Tabela.

Tabela 4.2: Relação entre o fator de amostragem e a largura de banda

Largura de Banda (L_b)	Fator de Amostragem (n)
Múltiplo de 1,25 MHz	144/125
Múltiplo de 1,5 MHz	86/75
Múltiplo de 1,75 MHz	8/7
Múltiplo de 2 MHz	57/50
Múltiplo de 2,75 MHz	316/275
Demais casos	8/7

Utilizando a equação 4.5 para substituir as ocorrências de F_s na equação 4.4, obtêm-se a equação 4.6. A partir da qual é possível calcular, diretamente, o tempo destinado a transmissão de informações úteis em cada símbolo OFDMA (BOTH et al., 2008).

$$T_b = \frac{N_{FFT}}{\left\lfloor \frac{L_b \cdot n}{8000} \right\rfloor \cdot 8000} \quad (4.6)$$

Com base em T_s e no tempo de duração de cada quadro (T_q), é possível obter o número de símbolos OFDMA transmitidos em cada quadro, conforme equação 4.7.

$$N_{simb} = \left\lfloor \frac{T_q}{T_s} \right\rfloor \quad (4.7)$$

O padrão IEEE 802.16e permite a utilização de 7 tamanhos de quadros diferentes. A cada tamanho de quadro é associado a um código e uma duração, representada em ms, conforme pode ser verificado na Tabela 4.3.

Tabela 4.3: Duração do quadro para transmissões OFDMA

Código	Duração do Quadro (T_q) [ms]
0	2,5
1	4
2	5
3	8
4	10
5	12,5
6	20

Os subcanais utilizados para alocação das rajadas UL no domínio de frequência, são definidos em função da técnica de subcanalização empregada. Neste trabalho considera-se a técnica UL PUSC. O número de subcanais (N_{SC}) em cada quadro OFDMA é obtido pela razão entre o número de suportadoras destinadas a transmissão de dados (N_{dados}) e uma constante (α) que indica o número de subportadoras de dados associadas a cada subcanal, conforme equação 4.8. Nesse caso, α é igual a 16, conforme descrito na seção 2.3.

$$N_{sc} = \left\lfloor \frac{N_{dados}}{\alpha} \right\rfloor \quad (4.8)$$

A partir da definição do símbolo OFDMA e da construção dos subcanais, é possível determinar a capacidade de transmissão da unidade de alocação mínima utilizada no subquadro UL, isto é, o *slot* UL. Conforme detalhado na Seção 2.3, cada *slot* UL é formado por 3 símbolos OFDMA e um subcanal. A capacidade de transmissão de cada *slot* UL ($C_{slot-ul}$), em bits, é obtida em relação a N_{SC} e ao número de bits transmitidos em cada símbolo OFDMA (b_{simb}), esse cálculo pode ser realizado através da equação 4.9.

$$C_{slot-ul} = \left\lfloor \frac{3 \cdot b_{simb}}{N_{sc}} \right\rfloor \quad (4.9)$$

O valor de b_{simb} , por sua vez, é calculado em função de N_{simb} e da quantidade de bits transmitida em cada quadro OFDMA (b_q), conforme equação 4.10.

$$b_{simb} = \left\lfloor \frac{b_q}{N_{simb}} \right\rfloor \quad (4.10)$$

Nesse caso, b_q é obtido pela equação 4.11, a partir da vazão total da rede (V) e de T_q .

$$b_q = V \cdot T_q \cdot 10^{-3} \quad (4.11)$$

A vazão total da rede, V , é calculada pela equação 4.12 que considera a razão de codificação do canal (R_{cod}) e a quantidade de bits associada a cada símbolo de modulação (M), além dos valores de B_w , n , N_{dados} , N_{FFT} e G , definidos anteriormente.

$$V = \left\lfloor \frac{L_b \cdot n}{8000} \right\rfloor \cdot 8000 \cdot \frac{N_{dados}}{N_{FFT}} \cdot \frac{1}{1 + G} \cdot R_{Cod} \cdot M \quad (4.12)$$

A partir de $C_{slot-ul}$, o *Nível 1* da arquitetura proposta pode definir as áreas do subquadro UL utilizadas para requisição de banda e transmissão de rajadas de dados. As áreas para requisição de banda subdividem-se em: canal de *ranging*, requisições de banda utilizando *polling* e oportunidades de requisições de banda referentes a códigos CDMA informados no canal de *ranging* de quadros anteriores. Os recursos destinados a cada uma dessas regiões são definidos pelo algoritmo de escalonamento PQ, conforme a seguinte ordem de prioridade: (i) canal de *ranging*, (ii) IE *Polling*, (iii) IE CDMA e (iv) oportunidades para transmissão de dados. Porém, o *Nível 1* realiza a alocação apenas dos 3 primeiros grupos, já a porção do quadro destinada as oportunidades de transmissão de dados, deve ser gerenciada pelo *Nível 2* da arquitetura.

Dessa forma, o *Nível 1* deve fornecer ao *Nível 2* a quantidade de *slots* disponíveis para alocação de oportunidades de transmissão de dados ($N_{slot-dados}$), que serão distribuídas entre as MSs. Esse parâmetro é calculado pela equação 4.13.

$$N_{slot-dados} = \left\lfloor \frac{b_q}{C_{slot-ul}} \right\rfloor - L_{ranging} - 2 \cdot (L_{polling} + L_{CDMA}) \quad (4.13)$$

A quantidade de *slots* atribuídos ao canal de *ranging* ($L_{ranging}$) foi fixada em 12, distribuídos em 3 símbolos OFDMA e 12 subcanais. Cada *slot* do canal de *ranging* é capaz de receber a informação de até 30 códigos CDMA diferentes (YOU; KIM; KIM, 2005). Existe a probabilidade de que mais de uma MS informe o mesmo código CDMA utilizando a mesma oportunidade do canal de *ranging*, causando colisões.

Na equação 4.13, $L_{polling}$ representa o número de oportunidades de *polling* que deverão ser oferecidas naquele quadro e L_{CDMA} corresponde ao número de códigos CDMA que estão na fila da BS aguardando pela alocação de uma oportunidade de requisição de banda. Considera-se que cada conexão transmitindo fluxos nrtPS e BE, com pacotes enfileirados, envia apenas um código CDMA por canal de *ranging*, independente do número de pacotes em suas filas. Essa definição impede que o canal de *ranging* receba um número extremamente grande de códigos, o que poderia gerar colisões sucessivas. Os valores $L_{polling}$ e L_{CDMA} devem ser multiplicados por 2, tendo em vista que o atendimento de cada requisição é realizado em 2 *slots*: o primeiro contendo um preâmbulo e o segundo, o cabeçalho de requisição de banda.

4.3 Gerenciamento de Requisições

O *Nível 2* da arquitetura proposta foi projetado para definir quais requisições de banda devem ser atendidas no próximo subquadro UL. Essa seleção é baseada no emprego de um esquema de escalonamento intra-classe e outro inter-classe. O mecanismo intra-classe define qual requisição, dentre as requisições de uma mesma classe de serviço, será selecionada para atendimento. Já o algoritmo inter-classe seleciona qual classe de serviço deve ser atendida.

Os algoritmos intra-classe e inter-classes, propostos neste trabalho, foram inspirados na proposta de Lin *et al.* (LIN *et al.*, 2008) e adaptados para obter um melhor desempenho,

além de considerar aspectos negligenciados no trabalho original. O escalonamento intra-classe é descrito na Subseção 4.3.1, enquanto o escalonamento inter-classe é apresentado na Subseção 4.3.2.

4.3.1 Escalonamento Intra-classe

As requisições recebidas em cada quadro OFDMA são agrupadas em filas de acordo com a classe de serviço a que pertencem. Dentro dessas filas as requisições são ordenadas para atendimento de forma adequada as características dos fluxos de cada classe.

Para as classes UGS e ertPS não há requisições, uma vez que essas classes não necessitam solicitar oportunidades de transmissão, pois seus requisitos de QoS devem ser previamente conhecidas pelo escalonador. Portanto, conexões dessas classes devem receber oportunidades de transmissão previamente definidas pelo escalonador. Entretanto, podem ocorrer requisições de ertPS, solicitando alterações no tamanho das oportunidades de transmissão recebidas.

Nas classes rtPS, nrtPS e BE, as requisições são ordenadas de forma crescente considerando um prazo para atendimento (PA), atribuído a cada requisição no momento em que a mesma é recebida pela BS. O prazo para atendimento de cada requisição rtPS (PA_{rtPS}) é calculado pela equação 4.14. Onde A_{max} representa o atraso máximo em ms suportado pelas aplicações agrupadas nessa classe.

$$PA_{rtPS} = \frac{A_{max}}{T_q} \quad (4.14)$$

Por outro lado, o prazo para atendimento das requisições pertencentes a classe nrtPS (PA_{nrtPS}), é obtido pelo emprego da equação 4.15. Nesse caso, considera-se $\delta < 0$, essa característica é relevante para o estabelecimento de um PA_{nrtPS} superior ao PA_{rtPS} . Sendo que, neste trabalho, δ foi fixado em $\frac{3}{2}$.

$$PA_{nrtPS} = \delta \times PA_{rtPS} \quad (4.15)$$

Já o prazo para atendimento das requisições BE (PA_{BE}), é fornecido pela equação 4.16. O prazo para atendimento das requisições BE deve ser superior ao prazo para atendimento das requisições rtPS e nrtPS, por isso $\gamma < \delta$. Neste trabalho, considera-se $\gamma = 2$.

$$PA_{BE} = \gamma \times PA_{rtPS} \quad (4.16)$$

A cada ciclo de escalonamento em que as requisições permanecem nas filas aguardando atendimento, seus prazos para atendimento (PA_{rtPS} , PA_{nrtPS} e PA_{BE}), são decrementados em uma unidade. As requisições que possuírem PA_{rtPS} , PA_{nrtPS} ou $PA_{BE} \leq 0$, deverão ser descartadas. O descarte dessas requisições não implica, necessariamente, no descarte de pacotes. Entretanto, é importante ressaltar que o descarte das requisições é feito para que as requisições da fila rtPS não ultrapassem o critério de atraso máximo, e para retirar da fila as requisições nrtPS e BE antigas ou desatualizadas, para que a MS reenvie tais requisições informando a situação mais recente de suas filas.

4.3.2 Escalonamento Inter-classes

O Gerenciamento de Requisições recebe do *Nível 1* a quantidade de recursos disponíveis para alocação de oportunidades de transmissão de dados e, posteriormente, calcula a quantidade de *slots* necessária para atender a cada requisição ($N_{slot-req}$). Esse cálculo é realizado considerando a quantidade de bits total associada a cada requisição de banda

(b_{req}), incluindo os bits necessários para transmissão de informações redundantes, inseridas de acordo com o perfil de rajada empregado na transmissão, conforme equação 4.17.

$$N_{slot-req} = \left\lceil \frac{b_{req}}{C_{slot-ul}} \right\rceil \quad (4.17)$$

Após o cálculo da quantidade de slots necessária para o atendimento de cada requisição, deve ser selecionada a fila para atendimento. Essa escolha é realizada com base na aplicação do Algoritmo 1, que prioriza o atendimento das filas com requisitos de tempo real e com necessidades de transmissão constantes.

Algoritmo 1 Algoritmo Escalonamento inter-classes

- 1: **enquanto** $N_{slot-dados} > 0 \wedge fila_UGS \neq 0$ **faça**
 - 2: Aloca oportunidades de transmissão de dados para conexões UGS
 - 3: **fim enquanto**
 - 4: **enquanto** $N_{slot-dados} > 0 \wedge fila_ertPS \neq 0$ **faça**
 - 5: Aloca oportunidades de transmissão de dados para conexões ertPS
 - 6: **fim enquanto**
 - 7: **enquanto** $N_{slot-dados} > 0 \wedge fila_PA_{rtPS_1} \neq 0$ **faça**
 - 8: Aloca oportunidades de transmissão de dados para requisições com $PA_{rtPS} = 1$
 - 9: **fim enquanto**
 - 10: **enquanto** $N_{slot-dados} > 0 \wedge (fila_rtPS \neq 0 \vee fila_nrtPS \neq 0 \vee fila_BE \neq 0)$ **faça**
 - 11: Seleciona para atendimento a fila rtPS, nrtPS ou BE com maior valor para Md
 - 12: **fim enquanto**
-

A alocação de oportunidades de transmissão para as conexões UGS, representadas no algoritmo por $fila_UGS$, é constante, conforme definido pelo padrão IEEE 802.16e. Já as conexões ertPS alternam períodos de transmissão de pacotes com períodos sem transmissão. A alternância entre esses estados é sinalizada pela informação de códigos CDMA no canal de *ranging*. Dessa forma, oportunidades de transmissão são alocadas para todas as conexões ertPS que estiverem no período de geração de tráfego ($fila_ertPS$) naquele quadro.

Após a alocação das filas com necessidades de transmissão constantes, $fila_UGS$ e $fila_UGS$, são atendidas as solicitações das fila de requisições rtPS com $PA_{rtPS} = 1$ ($fila_PA_{rtPS_1}$). Essa definição foi estabelecida pois, essas requisições serão descartadas se não forem atendidas no quadro atual e neste caso, o requisito de atraso máximo desta classe não será atendido.

No passo seguinte, são selecionadas as filas rtPS, nrtPS ou BE, de acordo com o maior valor médio da urgência (Md) no atendimento de cada fila. Essa média é calculada de acordo com o fator de urgência (U) atribuído a cada requisição. O fator U de cada requisição na fila rtPS é calculado de acordo com a equação 4.18. Onde P_{rtPS} representa a prioridade atribuída a fila rtPS, no caso, $P_{rtPS} = 3$. Enquanto, a variável PA_{rtPS_i} representa o prazo para atendimento daquela requisição, obtido a partir do algoritmo de escalonamento intra-classe.

$$U_i = \frac{P_{rtPS}}{PA_{rtPS_i}} \quad (4.18)$$

De forma semelhante, o fator U para as requisições na fila nrtPS é dado pela equação 4.19. Onde a prioridade da fila nrtPS (P_{nrtPS}) é igual a 2.

$$U_i = \frac{P_{nrtPS}}{PA_{nrtPS_i}} \quad (4.19)$$

Para as requisições na fila BE, o fator U é fornecido pela equação 4.20 e a prioridade atribuída a fila (P_{BE}) é 1.

$$U_i = \frac{P_{BE}}{PA_{BE_i}} \quad (4.20)$$

O valor médio para a urgência no atendimento de cada fila é calculado pela equação 4.21. Onde n representa o número de requisições presentes na fila daquela classe de serviço. Será selecionada para atendimento, a cada ciclo, enquanto houverem *slots* disponíveis, a fila com o maior valor para Md .

$$Md = \frac{\sum_{i=1}^n U_i}{n} \quad (4.21)$$

Os algoritmos intra-classe e inter-classes, pertencentes ao *Nível 2* da arquitetura, são projetados com o objetivo de suprir as restrições de cada uma das cinco classes de serviço previstas pelo padrão. Além disso, o emprego dos *Níveis 1 e 2*, juntamente com a utilização das estratégias do mecanismo de alocação de banda, permitem o atendimento dos principais aspectos não definidos pelo padrão IEEE 802.16e, quanto ao escalonamento do tráfego UL em um canal sem fio multiplexado pela técnica OFDMA.

5 FERRAMENTA DE SIMULAÇÃO

A arquitetura de escalonamento *uplink* proposta no Capítulo 4 foi analisada através de uma ferramenta elaborada para esse fim. Essa ferramenta de simulação foi desenvolvida em diversos módulos para permitir sua integração com outras ferramentas, alterações ou mesmo o emprego de outras arquiteturas de escalonamento.

Além da ferramenta de simulação desenvolvida, são descritos neste Capítulo os modelos de tráfego empregados para caracterizar a utilização da rede pela transmissão de aplicações reais. Esses modelos são utilizados para representar a demanda pelos recursos que a arquitetura de escalonamento deve gerenciar. As tarefas mais relevantes, implementadas pela ferramenta de simulação, são apresentadas em fluxogramas considerando diferentes perspectivas.

O Capítulo está organizado da seguinte forma, a Seção 5.1 descreve os ambientes de programação utilizados, já a Seção 5.2 aborda os modelos de tráfego empregados para simulação de aplicações pertencentes as cinco classes de serviço. Finalmente, a Seção 5.3 contém os fluxogramas que representam as principais tarefas realizadas pela ferramenta e a ordem de execução das mesmas.

5.1 Ambientes de Programação

A ferramenta de simulação foi desenvolvida utilizando dois ambientes de programação diferentes. O LabVIEW (*Laboratory Virtual Instrument Engineering Workbench*), que é desenvolvido pela National Instruments e o MATLAB (*MATrix LABoratory*), pertencente a MathWorks.

O LabVIEW é um ambiente de programação visual que utiliza ícones gráficos, lembrando um fluxograma. A programação é realizada considerando o modelo de fluxo de dados em programas chamados de VIs (*Virtual Instruments*). Cada VI é composto por um painel frontal, que contém a interface, e por um diagrama de blocos, que contém o código do programa.

Os ícones gráficos utilizados para a construção do diagrama de blocos correspondem às funções pré-definidas, como por exemplo, a criação de vetores ou de laços de repetição. A interação entre esses ícones é realizada por conectores, que representam a entrada ou saída de dados de determinada função. A utilização desse tipo de ambiente permite uma melhor visualização do fluxo de informações na ferramenta de simulação e contribui para um desenvolvimento mais rápido, tendo em vista que diversas funções de uso comum são disponibilizadas como ícones gráficos. Um VI criado pelo usuário é associado a um novo ícone gráfico e pode ser utilizado como sub-VI na construção de um novo programa.

A escolha do ambiente LabVIEW justifica-se pela diversidade de funções de programação disponíveis no mesmo, pela linguagem de programação visual que possibilita um

desenvolvimento mais ágil e permite a integração da ferramenta desenvolvida com outros módulos criados em trabalhos relacionados ao tema e pertencentes ao mesmo grupo de pesquisa, do qual este estudo faz parte.

O MATLAB, utilizado para implementação dos modelos de tráfego, é um ambiente de programação que possibilita o uso de uma linguagem de alto nível para resolver problemas numéricos. Esse ambiente disponibiliza diversas funções matemáticas, previamente desenvolvidas. A utilização desse software para a implementação de funções probabilísticas, necessárias ao funcionamento da ferramenta de simulação construída, justifica-se pela diversidade de funções oferecidas e pela possibilidade de integração do mesmo com o ambiente LabVIEW.

5.2 Modelos de Tráfego

Nesta seção são descritos os modelos de tráfego utilizados nas simulações. Foram empregados 5 modelos de tráfego, correspondentes as 5 aplicações diferentes e adequados as características das classes de serviço, definidas pelo padrão IEEE 802.16e. A Tabela 5.1 apresenta a aplicação associada a cada classe de serviço. Essa associação entre classes de serviço e aplicações segue o sugerido pelo documento *System Evaluation Methodology*, publicado pelo WiMAX fórum (WiMAX Forum, 2008) para orientar avaliações de desempenho considerando o padrão IEEE 802.16e.

Tabela 5.1: Aplicação associada a cada classe de serviço

Classe	Aplicação
UGS	VoIP sem supressão de silêncio
erPS	VoIP com supressão de silêncio
rtPS	Vídeo Clip
nrtPS	HTTP
BE	FTP

5.2.1 Modelo de Tráfego VoIP sem Supressão de Silêncio

O modelo de tráfego para aplicações do tipo VoIP sem supressão de silêncio é baseado nos parâmetros do codec G.711, listados na Tabela 5.2, conforme (GARG; KAPPES, 2003).

Tabela 5.2: Parâmetros codec G.711

Parâmetro	Valor
Intervalo geração amostras de voz	10ms
Encapsulamento	RTP/UDP/IP
Tamanho do pacote	Cabeçalho MAC (6B) + RTP/UDP/IP comprimido (3B) + pacote de voz (160B) = 169B

Para o codec G.711 são gerados pacotes com tamanho de 169B em intervalos regulares

de 10ms. Não há variação no intervalo de geração de pacotes ou no tamanho dos mesmos porque esse codec não realiza operações de supressão de silêncio.

5.2.2 Modelo de Tráfego VoIP com Supressão de Silêncio

Para o modelo de tráfego de aplicações VoIP com supressão de silêncio, são utilizados os parâmetros do codec AMR (*Adaptive Multi Rate*), listados na Tabela 5.3.

Tabela 5.3: Parâmetros codec AMR

Parâmetro	Valor
Duração período de transmissão (On)	exponencial, média=1026ms
Duração período de silêncio (Off)	exponencial, média=1171ms
Intervalo geração amostras de voz (On)	20ms
Encapsulamento	RTP/UDP/IP
Tamanho do pacote (ON)	Cabeçalho MAC (6B) + RTP/UDP/IP comprimido (3B) + pacote de voz (33B) = 42B

Aplicações VoIP com supressão de silêncio, seguem o modelo de uma fonte ON/OFF. Durante o período ON, ou período de transmissão, há geração de pacotes com informações relevantes para o usuário. Já no período OFF, ou período de silêncio, são transmitidas apenas informações de controle ou ruído de conforto. Medições experimentais realizadas (ESTEPA; ESTEPA; VOZMEDIANO, 2004), indicam que, para o codec AMR, a duração do período ON segue uma distribuição de probabilidade do tipo exponencial, com valor médio de 1026ms, enquanto o período OFF pode ser modelado pelo mesmo tipo de distribuição exponencial, porém, com média igual a 1171ms.

A utilização de uma distribuição de probabilidade para geração de números aleatórios, implica na utilização de algum método de geração de números aleatórios, a partir da função PDF dessa distribuição. Por isso, a PDF (*Probability Density Function*) da distribuição exponencial é apresentada na Tabela 5.4, em que o parâmetro λ representa o valor médio da distribuição.

Tabela 5.4: Distribuição exponencial

Distribuição	PDF	Parâmetro
Exponencial	$f_x = \lambda e^{-\lambda x}, x \geq 0$	λ

Para a geração de números aleatórios usando a distribuição exponencial foi aplicado o método da transformação inversa, visto que, tal distribuição possui uma PDF que pode ser invertida analiticamente. Essa operação fornece a equação 5.1, onde x é uma variável aleatória exponencial e Un representa uma distribuição uniforme (DEVROYE, 1986).

$$x = -\frac{1}{\lambda} \ln(Un) \quad (5.1)$$

5.2.3 Modelo de Tráfego HTTP

O tráfego HTTP considera um MTU (*Maximum Transfer Unit*) de 576B e simula sessões típicas de navegação web (WiMAX Forum, 2008). Cada sessão é composta por chamadas de pacotes (que representam as requisições feitas pelo usuário) e tempos de leitura.

Tabela 5.5: Parâmetros para o tráfego HTTP

Parâmetro	Valor Médio	Valor Mínimo	Valor Máximo	Distribuição
Tamanho objeto principal	10710B	100B	2MB	lognormal truncada
Número de objetos embutidos	5,64		53	Pareto truncada
Tamanho dos objetos embutidos	7758B	50B	2MB	lognormal truncada
Tempo de leitura	30s			Exponencial
Tempo de <i>parsing</i>	0,13s			Exponencial

Cada chamada de pacotes pode ser subdividida em períodos ON e OFF. Nesse caso, os períodos ON e OFF existentes não são resultados da interação humana. Cada chamada de pacotes, em uma sessão, inclui o objeto principal da página web, os objetos embutidos e o tempo de *parsing*. O objeto principal é a página HTML inicial e os objetos embutidos são os objetos que estão inseridos na página inicial. Já o tempo de *parsing*, é o intervalo necessário para que o navegador apresente a página HTML com todos os objetos embutidos. A página inicial e os objetos embutidos representam o período ON e o tempo de *parsing* representa o período OFF, em cada chamada de pacotes. A cada chamada de pacotes segue-se um tempo de leitura, que representa o período em que o usuário analisa a página recebida, antes de fazer uma nova requisição. Esses e demais parâmetros utilizados para a simulação do tráfego HTTP são resumidos na Tabela 5.5.

Os valores para o tempo de leitura e tempo de *parsing*, obtidos pelo emprego da distribuição exponencial, utilizam a equação 5.1, da mesma forma que ocorre para a geração dos períodos ON e OFF no modelo de tráfego VoIP com supressão de silêncio. Já as PDFs das funções lognormal truncada e Pareto truncada, utilizadas para simular o tamanho e número dos objetos embutidos em uma página, são fornecidas na Tabela 5.6. Os parâmetros σ e μ empregados na distribuição lognormal representam o desvio padrão e a média, respectivamente. Já os parâmetros α , k , e m utilizados para a distribuição de Pareto representam, respectivamente, a forma, a escala e o limitante.

A geração de números aleatórios, segundo a distribuição de Pareto, é obtida pelo emprego do método da transformação inversa a partir da PDF da distribuição. Da aplicação do método resulta a equação 5.2, obtida em (DEVROYE, 1986), na qual x é uma variável aleatória que segue a distribuição de Pareto e U_n representa uma distribuição uniforme.

$$x = \frac{k}{U_n^{1/\alpha}} \quad (5.2)$$

Para a distribuição lognormal truncada, a geração de valores aleatórios foi obtida pelo emprego da função *longrnd*(μ , σ), disponível na ferramenta MATLAB.

Tabela 5.6: Distribuições lognormal e Pareto truncada

Distribuição	PDF	Parâmetros
lognormal truncada	$f_x = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], x \geq 0$	σ, μ
Pareto truncada	$f_x = \frac{\alpha k^\alpha}{x^{\alpha+1}}, k \leq x < m$ $f_x = \left(\frac{k}{m}\right)^\alpha, x = m$	α, k, m

5.2.4 Modelo de Tráfego FTP

Segundo Bassi (BASSI, 2006), uma sessão FTP consiste em uma sequência de transferências de arquivos, separados por tempos de leitura. Desta forma, existem apenas dois parâmetros para o modelo de tráfego de uma aplicação FTP: o tamanho do arquivo que será transferido e o tempo de leitura do usuário, que representa o tempo entre o término de uma transmissão até que uma nova requisição do usuário seja feita. Ambos os parâmetros são apresentados na Tabela 5.7.

Tabela 5.7: Parâmetros para o tráfego FTP

Parâmetro	Valor Médio	Valor Máximo	Distribuição
Tamanho do arquivo	2MB	5MB B	lognormal truncada
Tempo de leitura	180s		Exponencial

O modelo de tráfego FTP também considera o emprego de um MTU de 576B. Além disso, a utilização das distribuições de probabilidade lognormal truncada e exponencial no modelo de tráfego FTP é realizada da mesma forma que no modelo de tráfego HTTP.

5.2.5 Modelo de Tráfego Vídeo Clip

Uma aplicação do tipo vídeo *clip* caracteriza-se como um tipo de vídeo curto, em meio digital e, na maior parte das vezes, de natureza musical ou publicitária. Aspectos como o codec de vídeo utilizado e o tamanho do *display* influenciam no tamanho do arquivo de vídeo gerado. Esses e os demais parâmetros considerados para o modelo de tráfego de vídeo *clip* são listados na Tabela 5.8.

Tabela 5.8: Parâmetros para o tráfego de vídeo *clip*

Parâmetro	Valor
Duração do vídeo <i>clip</i>	exponencial, média=15s e máximo=60s
Codec de vídeo	MPEG4
Tamanho <i>display</i>	176 x 144
Tamanho do <i>frame</i> de vídeo comprimido	2,725KB

Os parâmetros considerados para o modelo de tráfego podem variar de um *trace* de vídeo para outro. Nesse caso, foram utilizados os valores de referência encontrados em

(WiMAX Forum, 2008). O codec MPEG4 é empregado pela sua ampla utilização e capacidade de compressão. Para esse codec, considera-se a geração de 25 quadros/s e a resolução do *display* empregada caracteriza o emprego de um dispositivo pequeno, com dimensões de 176 x 144.

Para o modelo de tráfego do tipo vídeo *clip*, considera-se vídeos com duração fornecida pelo emprego da distribuição exponencial, com média de 15s e duração máxima de 60s. Esses arquivos são transmitidos empregando o mesmo encapsulamento utilizado para transmissões FTP. Sendo que, o intervalo entre a transmissão de dois vídeo *clips* diferentes segue uma distribuição exponencial, com média de 180s.

5.3 Fluxograma da Implementação

O objetivo desta Seção é descrever o funcionamento da ferramenta de simulação desenvolvida para avaliar o desempenho da arquitetura de escalonamento *uplink* proposta neste trabalho e para realizar comparações com trabalhos relacionados. A descrição é feita por meio de fluxogramas, que representam a sequência de informações e a ordem de execução das tarefas.

A descrição da ferramenta é realizada considerando três perspectivas diferentes. A Subseção 5.3.1 contém o fluxograma que representa a visão geral da ferramenta de simulação. Já a Subseção 5.3.2, aborda o fluxo de informações e de execução de tarefas do ponto de vista da BS. Enquanto a Subseção 5.3.3 considera a geração de informações pelas MSs, ou seja, descreve o funcionamento da ferramenta de simulação do ponto de vista das MSs.

5.3.1 Perspectiva Geral

A ferramenta de simulação construída considera uma topologia PMP, em que a BS controla o acesso ao meio de transmissão, conforme Figura 5.1. O número de MSs presente em cada simulação é variável e informado como parâmetro. Cada MS possui uma conexão associada a uma classe de serviço. A cada classe de serviço é associada uma aplicação diferente, conforme os modelos de tráfego descritos.

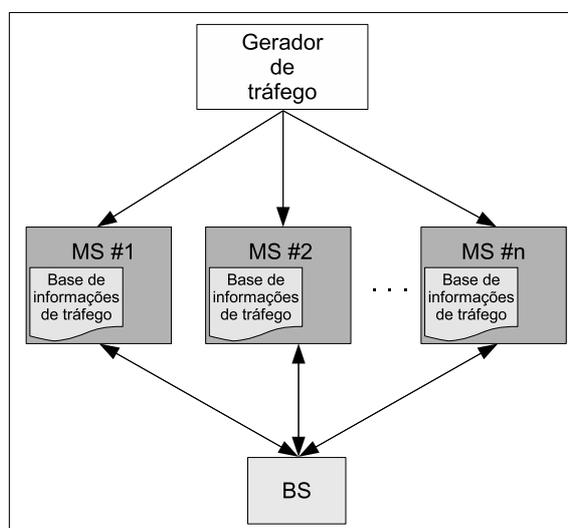


Figura 5.1: Fluxograma da implementação - perspectiva geral

Os módulos que representam as MSs e o módulo que representa a BS são descritos nas Subseções seguintes. Já o módulo Gerador de Tráfego é responsável por simular as necessidades de transmissão características de cada uma das cinco classes de serviço, conforme os modelos de tráfego descritos. Para isso, cada MS possui uma Base de Informações de Tráfego, que contém informações fornecidas pelo Gerador de Tráfego e que consistem em duas estruturas de dados contendo: o tamanho dos pacotes gerados e o tempo de chegada de cada pacote.

5.3.2 Perspectiva da BS

O módulo BS representa as atividades exercidas pela mesma para controlar o acesso das MSs ao subquadro UL e o seu funcionamento pode ser verificado no fluxograma da Figura 5.2. O processamento realizado pela BS é representado como uma sequência de tarefas repetidas a cada quadro OFDMA, até que a simulação seja finalizada.

As funções do módulo BS iniciam com o recebimento de um novo subquadro UL. Informações como requisições de oportunidades de transmissão de dados são armazenadas em filas de acordo com a classe de serviço a que pertencem. Os códigos CDMA informados no canal de *ranging* são armazenados em uma fila para atendimento. Após o processamento das novas informações recebidas, o escalonador UL é utilizado para definir a distribuição de informações no próximo subquadro UL e para gerenciar as filas de requisições. Ao final do ciclo de simulação do escalonador uma mensagem do tipo UL-MAP é gerada.

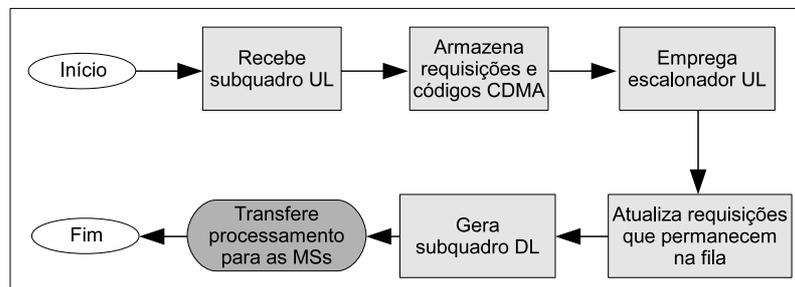


Figura 5.2: Fluxograma da implementação - perspectiva da BS

As requisições e os códigos CDMA que permanecerem na fila, após o ciclo de escalonamento têm o seu contador atualizado, pois permanecerão aguardando atendimento. Além disso, são verificadas as requisições cujo prazo para atendimento foi esgotado para que as mesmas sejam descartadas.

O subquadro DL é construído considerando a estrutura de quadro definida pela tecnologia. Ou seja, inicia com um preâmbulo de sincronização seguido pelas mensagens de gerenciamento DL-MAP e UL-MAP. A mensagem UL-MAP é gerada pelo escalonador UL, e a mensagem DL-MAP é gerada pelo módulo responsável pela construção do subquadro DL. Após as mensagens de gerenciamento, o restante do subquadro DL é preenchido com uma única rajada DL. Essa solução foi adotada porque o escalonamento DL não faz parte do escopo deste trabalho.

Após a geração do novo subquadro DL, a BS transfere o processamento para o módulo responsável pelas tarefas das MSs e só retorna quando o processamento for devolvido a BS.

5.3.3 Perspectiva das MSs

As MSs presentes na rede são representadas pelo módulo MSs, responsável pela simulação das funções das mesmas. Cada MS presente na simulação executa a sequência de atividades pertencentes ao fluxograma da Figura 5.3, sempre que o processamento está em controle das MSs.

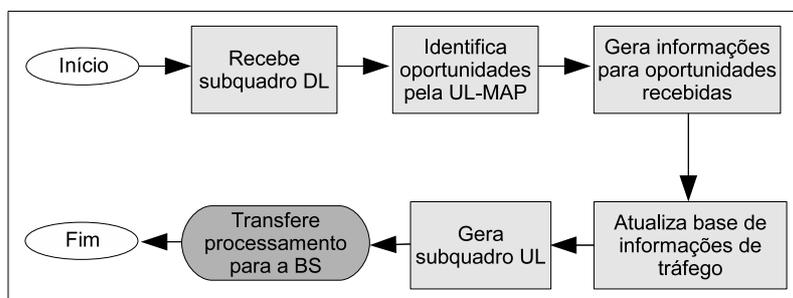


Figura 5.3: Fluxograma da implementação - perspectiva das MSs

As tarefas de uma MS iniciam com o recebimento de um novo subquadro DL. Nesse subquadro, a MS deve localizar a mensagem UL-MAP e identificar as oportunidades concedidas ao seu CID ou ao CID da conexão que a MS possui. Cada MS possui apenas uma conexão, essa característica foi implementada porque o escalonador UL da MS não faz parte do escopo deste trabalho.

Após a identificação das oportunidades recebidas, a MS deve consultar sua Base de Informações de Tráfego para verificar se há necessidade de informar um código CDMA no canal de *ranging*. A Base de Informações de Tráfego deve ser consultada para obter informações sobre o tamanho das oportunidades de transmissão de dados requeridas e para fornecer as informações referentes a eventuais oportunidades de transmissão de dados recebidas no subquadro atual. Ao final do processo, a Base de Informações de Tráfego deve ser atualizada para que deixem de constar as informações já transmitidas.

A construção do subquadro UL é realizada a partir dos dados de todas as MSs que receberam oportunidades na UL-MAP. Além disso, são reunidos os códigos CDMA informados no canal de *ranging*, compostos pelo código e pelas coordenadas da oportunidade do canal de *ranging* utilizada para informação do mesmo, para verificação de possíveis colisões. Após a construção do subquadro UL o processamento é devolvido a BS.

A descrição da ferramenta desenvolvida permite verificar que a mesma foi projetada para possibilitar a simulação do ambiente em que a arquitetura de escalonamento *uplink*, proposta neste trabalho, deve atuar. Esse cenário é criado considerando as características da tecnologia IEEE 802.16e e a demanda por recursos, gerada pelas aplicações que utilizam a rede. Por isso, esse ambiente é utilizado para avaliar o desempenho da proposta e os resultados dessa análise são descritos no capítulo seguinte.

6 RESULTADOS

O desempenho da arquitetura de escalonamento *uplink* proposta é avaliado neste capítulo, através da ferramenta de simulação desenvolvida para essa finalidade. Para isso, os parâmetros de simulação utilizados são descritos na Seção 6.1, enquanto na Seção 6.2 são avaliados diferentes aspectos da arquitetura de escalonamento *uplink* e a Seção 6.3 analisa a arquitetura proposta em relação a outras estratégias de escalonamento.

6.1 Parâmetros de Simulação

Os cenários utilizados para as avaliações de desempenho foram criados considerando o documento *System Evaluation Methodology* publicado pelo WiMAX fórum (WiMAX Forum, 2008). Esse documento contém sugestões de parametrização para avaliações de desempenho em simulações de redes baseadas no padrão IEEE 802.16e.

O documento WiMAX descreve que, para uma largura de banda (L_b) de 10MHz, deve-se considerar um número total de 1024 subportadoras (N_{FFT}), sendo que desse total, 720 subportadoras são utilizadas para transmissão de dados (N_{dados}). Para essa largura de banda, o fator de amostragem (n) é igual a 57/50. Além disso, o documento sugere o emprego de um prefixo cíclico (G) de 1/8.

Como técnica de subcanalização, considera-se o emprego da técnica UL PUSC, que possibilita a utilização de 16 subportadoras de dados por subcanal (α). Além disso, a estratégia de subcanalização utilizada implica na utilização de 3 símbolos OFDMA e um subcanal por *slot* UL. Essa técnica é empregada em um quadro com duração (T_q) de 10ms, dividido em iguais proporções entre as transmissões DL e UL. As informações alocadas no quadro são codificadas utilizando a configuração AMC 16-QAM 3/4, porque a mesma proporciona a maior vazão de dados úteis na rede, dentre as configurações de implementação obrigatória, previstas pelo padrão IEEE 802.16e.

Quanto a garantia de QoS, o atraso máximo (A_{max}) empregado no atendimento das requisições rtPS foi fixado em 100ms, conforme sugere Kang *et al.* (KANG *et al.*, 2007), para aplicações de vídeo. Considerando as características dos modelos de tráfego, empregados em cada classe de serviço, verifica-se que o modelo FTP possui o maior intervalo médio entre requisições, fornecido por uma distribuição exponencial com média de 180s. Por esse motivo, o tempo de simulação foi fixado em 360s, visando aumentar a probabilidade das conexões FTP alternarem entre o período de leitura e o período de transmissão de dados ao longo da simulação. Um resumo dos parâmetros utilizados nos cenários de simulação é apresentado na Tabela 6.1.

Os parâmetros de simulação descritos, possibilitam a criação de cenários que refletem as condições em que uma rede IEEE 802.16e é utilizada, considerando aspectos da camada física e da camada MAC, previstos pela tecnologia. Esses parâmetros foram utili-

Tabela 6.1: Parâmetros de Simulação

Parâmetro	Valor
Largura de banda (L_b)	10MHz
Taxa de amostragem (n)	57/50
Subportadoras total (N_{FFT})	1024
Subportadoras de dados (N_{dados})	720
Subportadoras de dados por subcanal UL (α)	16
Duração do prefixo cíclico (G)	1/8
Duração do símbolo OFDMA (T_s)	136 μ s
Duração do quadro (T_q)	10ms
Configuração AMC	16-QAM 3/4 ($M = 4, R_{Cod} = 3/4$)
Proporção do quadro (DL:UL)	1:1
Atraso máximo para atendimento rtPS (A_{max})	100ms
Tempo de simulação	360s
Intervalo de confiança	95%

zados para avaliar a arquitetura proposta e para comparação com trabalhos relacionados.

6.2 Análise Considerando Cinco Classes de Serviço

O desempenho da arquitetura proposta é analisado considerando a utilização de aplicações pertencentes as cinco classes de serviço previstas pelo padrão IEEE 802.16e, em iguais proporções. Utiliza-se, ainda, o mecanismo de alocação de banda previsto para a tecnologia, dividindo o quadro OFDMA entre oportunidades de requisição de banda e transmissão de dados. Para fornecer características mais realistas ao escalonamento *uplink*, as alocações são realizadas considerando a estrutura do quadro OFDMA e a configuração AMC utilizada nesse quadro.

A alocação de oportunidades de requisição de banda é de responsabilidade do *Nível 1*, na arquitetura proposta. Após a alocação dessas oportunidades, o *Nível 2* é informado sobre os *slots* disponíveis para transmissão de dados. Essa divisão do subquadro UL entre oportunidades de requisição de banda e transmissão de dados pode ser verificada na Figura 6.1. Nesse gráfico, o percentual do subquadro UL utilizado para requisições de banda, ou seja, para alocar o canal de *ranging*, as oportunidades de *polling* e os *IEs CDMA*, é representado pela linha chamada de sobrecarga. A linha que sinaliza os dados totais representa o percentual do subquadro utilizado para transmissão dos dados das cinco classes de serviço. Já a linha representada por *padding* indica o percentual do quadro preenchido com dados não úteis, para completar a duração do subquadro UL.

A análise do gráfico permite verificar que, quando há um pequeno número de conexões na rede, grande parte do subquadro é preenchido por *padding*, isso indica que não existem transmissões de dados e oportunidades de requisição de banda suficientes para preencher todo o subquadro UL. Por exemplo, quando há 20 conexões na rede, aproximadamente 80% do quadro é preenchido com *padding*, enquanto menos de 17% é utilizado para transmissão de dados. Além disso, a sobrecarga gerada pelas oportunidades de requisição de banda é de aproximadamente 3%.

O aumento do número de conexões na rede, faz com que o percentual do subquadro

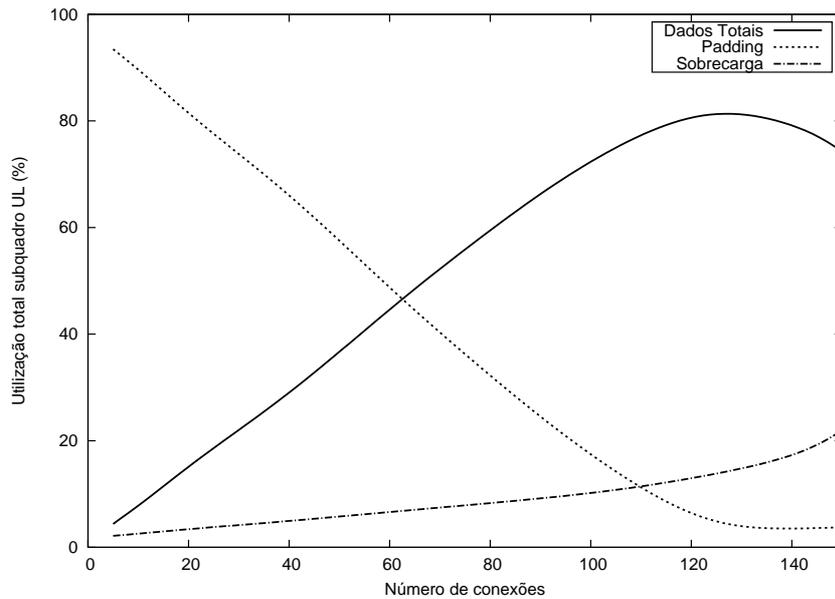


Figura 6.1: Composição do subquadro UL

UL preenchido por *padding* diminua gradativamente, enquanto o percentual utilizado para transmissão de dados aumenta. Nesse contexto, é possível verificar que, o percentual do subquadro utilizado pelas oportunidades de requisição de banda também aumenta. Esse comportamento justifica-se porque, com um número maior de conexões rtPS presentes na rede, é necessário alocar um número maior de oportunidades de *polling*. Uma vez que, conexões rtPS informam suas necessidades de transmissão de dados através de oportunidades desse tipo. Nesse cenário, também há mais códigos CDMA sendo informados no canal de *ranging* e aguardando atendimento, pois existem mais conexões nrtPS e BE na rede, e tais conexões notificam a BS sobre suas necessidades de transmissão pela informação de códigos CDMA no canal de *ranging*. Por exemplo, quando há 100 conexões na rede, a sobrecarga gerada pelas requisições de banda é de aproximadamente 10%, enquanto o *padding* inserido é próximo de 18% e o percentual do subquadro utilizado para transmissão de dados é aproxima-se de 72%.

É possível verificar que, a linha que representa o percentual do subquadro UL utilizado para transmissão de dados não apresenta um crescimento constante, com qualquer número de conexões na rede. A partir de 125 conexões na rede, passa a haver um decréscimo no percentual do subquadro utilizado para transmissão de dados. Esse comportamento ocorre porque foi atingida a capacidade máxima de transmissão em cada subquadro UL, porém o percentual do subquadro utilizado pelas oportunidades de requisição de banda continua crescendo, gerando um decréscimo no percentual utilizado para transmissão de dados. Nesse contexto, a observação da Figura 6.1 ressalta que a alocação das oportunidades de requisição de banda não pode ser ignorada na proposta de uma arquitetura de escalonamento e na avaliação de desempenho de tal arquitetura.

O percentual do subquadro UL utilizado para transmissão de dados é dividido entre as transmissões das cinco classes de serviço empregadas nesse cenário. Para analisar a transmissão dos fluxos de cada uma das classes elaborou-se o gráfico que pode ser verificado na Figura 6.2. Nesse gráfico, a soma dos percentuais de utilização do quadro por classe de serviço corresponde ao percentual total do subquadro UL alocado para transmissão de

dados.

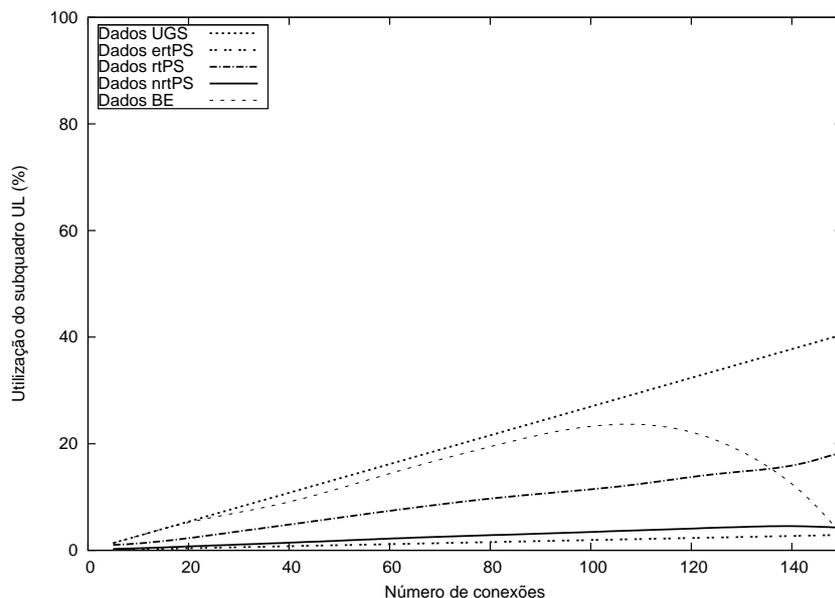


Figura 6.2: Percentual de utilização do subquadro UL por classe de serviço

A fração do subquadro UL utilizada para cada classe está diretamente relacionada às características da aplicação associada aquela classe e ao desempenho do escalonamento inter-classes. Quando existem até 100 conexões ativas, o tráfego da classe UGS utiliza o maior percentual do subquadro UL. Esse comportamento ocorre porque, a taxa de transmissão é constante, embora a aplicação que representa essa classe possibilite a geração de pacotes de dados de tamanho médio (169B). A segunda classe a ocupar o maior percentual do subquadro UL é a classe BE. Esse comportamento justifica-se pelo fato de que a aplicação FTP, atribuída a essa classe, utiliza arquivos de tamanhos grandes (em média 2MB), embora com grandes intervalos entre a requisição de um arquivo e outro (180s). A terceira classe de serviço na ocupação do subquadro UL é a classe rtPS, responsável pela inserção de arquivos menores do que os utilizados pela classe BE (em média 1,022MB), o que explica o percentual médio de ocupação do subquadro UL para transmissão de fluxos dessa classe. Por fim, as classes nrtPS e ertPS são responsáveis pela menor utilização do subquadro UL, o que ocorre pelas baixas necessidades de transmissão das aplicações atribuídas a essas classes.

Nesse cenário, o percentual do subquadro UL utilizado para transmissão dos fluxos pertencentes a cada uma das cinco classes de serviço cresce proporcionalmente ao aumento do número de conexões na rede, até que esse número atinja 100 conexões. A partir desse ponto, a rede passa a apresentar congestionamento. Nesse caso, o escalonador inter-classes prioriza o atendimento das requisições rtPS em detrimento das requisições BE. Verifica-se, por exemplo, que a classe rtPS mantém o crescimento no percentual de utilização do subquadro com o aumento no número de conexões, enquanto o percentual de utilização do subquadro pela classe BE decresce com o aumento no número de conexões na rede.

As classes de serviço atendidas por meio de requisições de oportunidades de transmissão estão sujeitas a atrasos nesse atendimento, o que caracteriza uma métrica importante na avaliação do desempenho de uma arquitetura de escalonamento. Por isso, o atraso

médio no atendimento das requisições de banda das classes rtPS, nrtPS e BE pode ser verificado na Figura 6.3. Considera-se que as cinco classes de serviço estão presentes na rede, mas o atraso médio é representado apenas para essas três classes porque, essas são as classes sujeitas a atrasos no atendimento de suas solicitações.

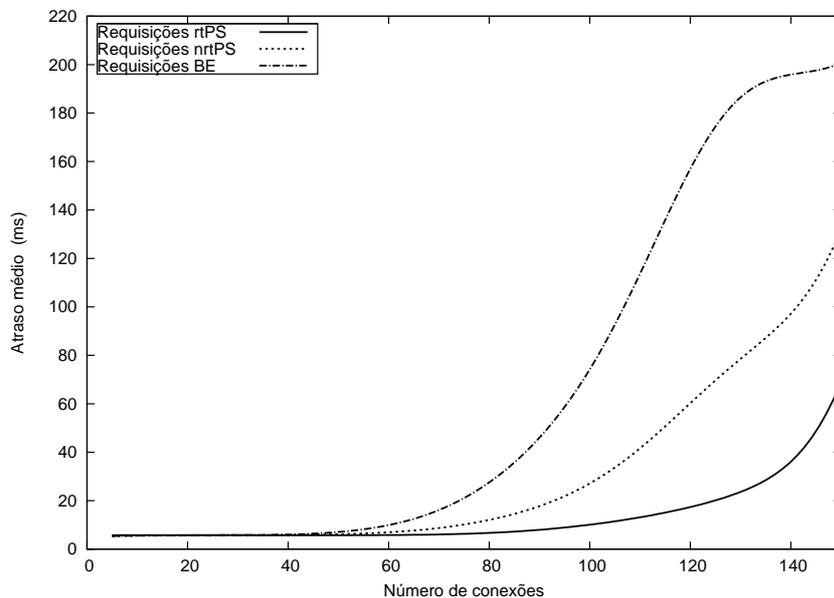


Figura 6.3: Atraso médio no atendimento das requisições

O gráfico mostra que, com até 40 conexões na rede, o atraso médio é mínimo, ou seja, todas as requisições recebidas em um quadro são atendidas no quadro seguinte. A medida que o número de conexões na rede aumenta, o atraso no atendimento das requisições também aumenta. Esse comportamento é devido a existência de um maior número de requisições aguardando atendimento. As primeiras requisições a experimentar atrasos maiores do que o atraso mínimo são as requisições da classe BE, seguidas pelas requisições da classe nrtPS, e finalmente pelas requisições rtPS. Sendo que, as requisições rtPS passam a apresentar atraso maior do que o atraso mínimo apenas quando há aproximadamente 80 conexões na rede.

A partir de 100 conexões na rede, o atraso no atendimento das requisições passa a crescer rapidamente, devido ao congestionamento da rede. Ainda assim, verifica-se que, com aproximadamente 150 conexões na rede, as requisições rtPS apresentam atraso médio de aproximadamente 70ms, enquanto as requisições nrtPS são atendidas com um atraso médio de 130ms e as requisições BE experimentam um atraso médio de 200ms. Em outras palavras, mesmo durante o período de congestionamento, a arquitetura de escalonamento proposta tenta garantir o atendimento dos requisitos de atraso máximo das conexões com características de tempo real.

O atraso médio no atendimento das requisições está relacionado ao algoritmo intra-classe, pertencente ao *Nível 2* da arquitetura proposta, que prevê o estabelecimento de um prazo para atendimento das requisições. Esse prazo é definido pela razão entre o atraso máximo suportado pela aplicação e a duração do quadro. O estabelecimento de um prazo para atendimento das requisições não implica em descarte de pacotes, apenas no descarte de requisições. Essa medida é adotada para evitar que as filas de requisições da BS fiquem sobrecarregadas com requisições antigas e para possibilitar que as MSs informem a BS,

a situação atual de suas filas. Nesse contexto, o percentual de requisições descartadas nas classes rtPS, nrtPS e BE a cada quadro, por ultrapassarem o prazo para atendimento, é apresentado na Figura 6.4.

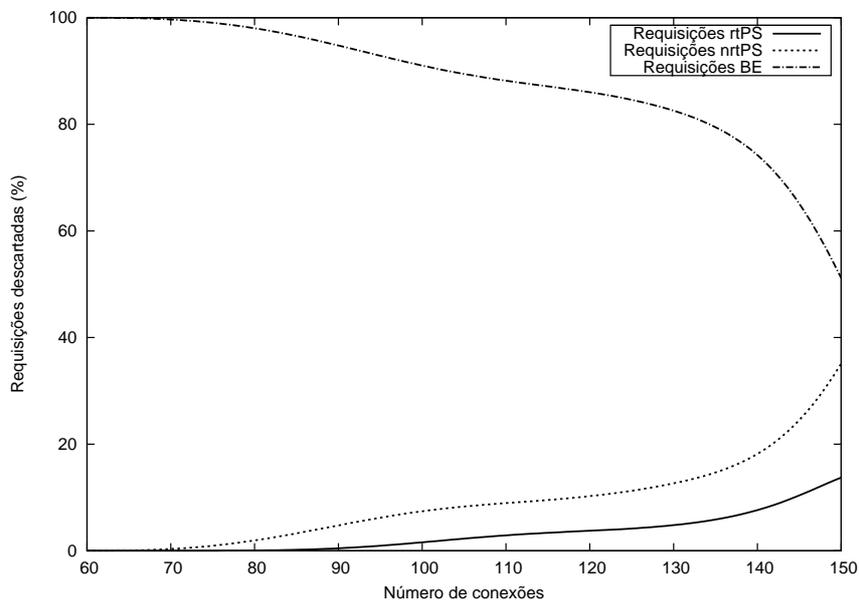


Figura 6.4: Percentual de requisições descartadas

O gráfico mostra o descarte das requisições apenas a partir de 60 conexões na rede porque, com menos de 60 conexões, nenhuma requisição é descartada. Cabe ressaltar que, são mostrados no gráfico os percentuais de descarte em cada classe de serviço, dentre o total de requisições descartadas a cada quadro. Como o descarte inicia pelas requisições BE, o gráfico inicia mostrando que 100% das requisições descartadas pertence a essa classe, pois apenas requisições BE estão sendo descartadas. Com o aumento no número de conexões na rede, passa a ocorrer descarte de requisições de outras classes de serviço, além da classe BE. Assim, o percentual de requisições BE descartadas diminui, em relação ao percentual de descarte das demais classes.

A análise do gráfico permite verificar que, quando há até 70 conexões na rede, apenas requisições BE são descartadas. A partir de 70 conexões, requisições nrtPS também passam a ser descartadas. Esse comportamento justifica-se pois, com o aumento no número de requisições aguardando atendimento, são descartadas requisições das classes nrtPS e BE, para evitar o descarte das requisições rtPS. Devido a esse mecanismo, passa a existir descarte de requisições rtPS apenas quando há mais de 85 conexões na rede. Além disso, dentre o total de requisições descartadas, o menor percentual pertence a classe rtPS, mesmo com um grande número de conexões ativas. Por exemplo, com 150 conexões na rede, 50% do total de requisições descartadas pertence a classe BE, enquanto aproximadamente 35% dos descartes são de requisições nrtPS e apenas 15% das requisições descartadas pertencem a classe rtPS.

As análises realizadas permitem verificar diversos aspectos da arquitetura de escalonamento *uplink* proposta, como composição do subquadro UL, atraso médio no atendimento das requisições e percentual de requisições descartadas, considerando a sua abrangência. Porém, para a comparação com os trabalhos relacionados é necessário limitar alguns aspectos, como o emprego das cinco classes de serviço previstas pelo padrão IEEE 802.16e.

Isso acontece porque os trabalhos relacionados não abordam algumas características consideradas nesta proposta.

6.3 Avaliação de Desempenho Comparativa

A arquitetura proposta é comparada ao algoritmo HUF, proposto por Lin *et al.* (LIN *et al.*, 2008) e ao trabalho de Jae-Woo So (SO, 2008), pois ambos apresentam estudos relevantes no contexto de escalonamento *uplink* ao qual essa proposta se aplica. Entretanto, esses projetos, bem como grande parte dos trabalhos estudados, apresentam limitações em suas arquiteturas de escalonamento. Por isso, o cenário utilizado para comparação considera apenas duas classes de serviço: a classe rtPS e a classe BE, em iguais proporções. Com base nessas duas classes, é possível avaliar o desempenho dos escalonadores quando aplicações com e sem requisitos de tempo real utilizam a rede. Além disso, as propostas utilizadas para comparação não consideram as cinco classes de serviço previstas pelo padrão IEEE 802.16e.

Outro aspecto relevante é que os trabalhos de Lin *et al.* e Jae-Woo So não consideram completamente o mecanismo de alocação de banda, usado quando a técnica OFDMA é empregada. Dessa forma, para tornar comparação de desempenho entre as propostas possível, foi necessário empregar o mecanismo de alocação de banda, definido na arquitetura apresentada neste trabalho, para as três estratégias de escalonamento consideradas. O mesmo ocorre quanto à estrutura do quadro OFDMA e a configuração AMC utilizada, tendo em vista que, os trabalhos de HUF e Jae-Woo So abordam apenas parcialmente os aspectos referentes a essa organização. Sendo assim, a comparação tornou-se possível apenas quanto ao *Nível 2* da arquitetura proposta, em relação as estratégias apresentadas por Lin *et al.* e Jae-Woo So

A análise comparativa inicia pelo atraso médio no atendimento das requisições BE, pois o tempo que as requisições passam nas filas da BS é uma métrica importante, para avaliar o comportamento geral dos escalonadores. Esse aspecto é demonstrado na Figura 6.5, onde são utilizadas três linhas para demonstrar o atraso experimentado em cada uma das arquiteturas de escalonamento consideradas. Cabe salientar que, o eixo y do gráfico está representado em escala logarítmica pois, a diferença entre os atrasos no atendimento proporcionados pela proposta de Jae-Woo So e o atrasos apresentados pelo algoritmo HUF e pela arquitetura proposta, é grande.

Pela observação do gráfico, é possível perceber que as linhas que representam o comportamento da arquitetura proposta e do algoritmo HUF se sobrepõem. Essa sobreposição ocorre porque a prioridade dada ao atendimento das requisições BE é semelhante em ambas as propostas. A diferença de priorização das requisições BE ocorre quando essas requisições apresentam um prazo para atendimento ou *deadline* igual a um. O prazo para atendimento é um critério definido pelo escalonamento intra-classe, do algoritmo HUF e da arquitetura proposta, em função do atraso máximo permitido, para o atendimento das requisições daquela classe e da duração do quadro. Nesse caso, se tais requisições não forem atendidas no quadro atual, serão descartadas. Para o algoritmo HUF, essas requisições são priorizadas, independente da classe a que pertencem. Por outro lado, como a priorização aplicada pela arquitetura proposta e pelo algoritmo HUF é semelhante quando o prazo para atendimento das requisições é diferente de um, o atraso médio experimentado pelas requisições BE em ambas as propostas é muito próximo.

Através da análise do gráfico pode-se observar que, quando há até 20 conexões na rede, o atraso no atendimento das requisições BE é mínimo, para as três propostas anali-

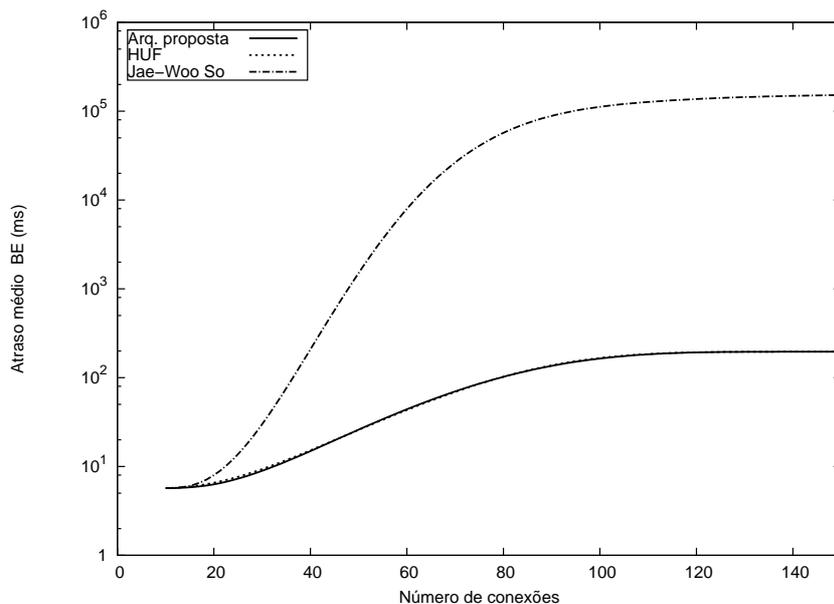


Figura 6.5: Comparativo entre os atrasos no atendimento das requisições BE

sadas. A partir de 20 conexões, o atraso sofrido por essas requisições, quando a proposta de Jae-Woo So é utilizada, passa a crescer rapidamente com o aumento no número de conexões ativas, tendendo a 100s, quando há mais de 100 conexões na rede. Cabe ressaltar que, o trabalho de Jae-Woo So não define qualquer critério de prazo máximo para atendimento de requisições. Por outro lado, quando a arquitetura proposta neste trabalho ou o algoritmo HUF são utilizados, o atraso no atendimento das requisições BE cresce lentamente, com o aumento do número de conexões na rede. O atraso aproxima-se do limite de 200ms, definido como prazo para atendimento dessas requisições, apenas quando há mais de 100 conexões ativas.

O atraso médio no atendimento das requisições BE deve ser analisado com conjunto com o atraso médio proporcionado às requisições rtPS, apresentado na Figura 6.6. Pela análise dos gráficos de atraso no atendimento das requisições BE e rtPS, é possível verificar que, quando a arquitetura proposta e o algoritmo HUF são empregados, o atraso no atendimento das requisições BE aumenta para que as requisições rtPS experimentem um atraso menor, com o mesmo número de conexões ativas. Entretanto, quando a proposta de Jae-Woo So é utilizada, as requisições rtPS e BE são atendidas em proporções semelhantes, devido a ausência de um mecanismo de diferenciação de classes. Nesse caso, o atraso no atendimento das requisições rtPS e BE é semelhante.

A análise da Figura 6.6 permite verificar que a arquitetura proposta mantém o atraso mínimo no atendimento das requisições rtPS, quando há até 50 conexões na rede. Por outro lado, com o emprego do algoritmo HUF, o atraso no atendimento das requisições rtPS é crescente a partir de 20 conexões ativas. O comportamento justifica-se pela maior prioridade concedida por HUF às requisições BE e rtPS com prazo para atendimento igual a um, em detrimento das demais requisições. Essa característica influencia no atendimento das requisições rtPS com prazo para atendimento maior do que um, gerando um maior atraso médio no atendimento das requisições rtPS. Por outro lado, é possível observar que a proposta de Jae-Woo So possui um atraso no atendimento das requisições crescente, a partir de, aproximadamente, 20 conexões ativas. Isso acontece, porque tal proposta não

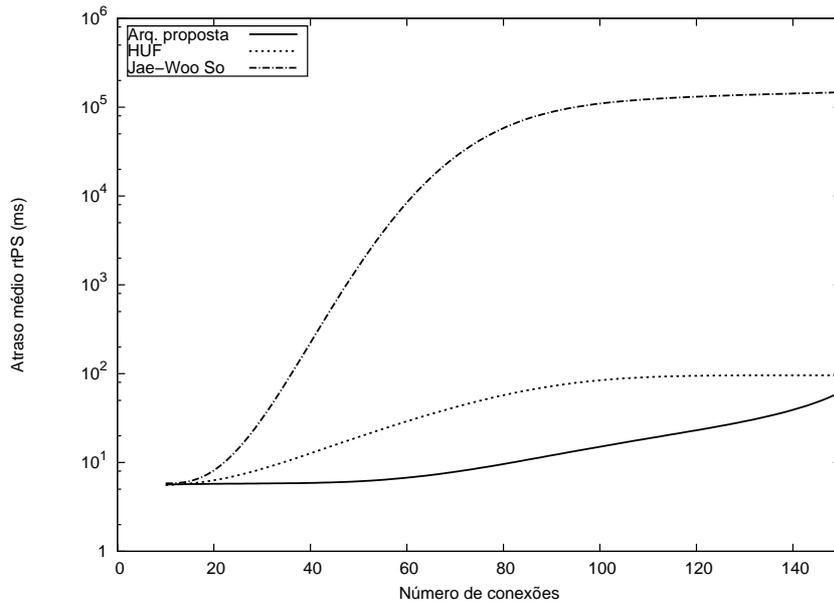


Figura 6.6: Comparativo entre os atrasos no atendimento das requisições rtPS

considera diferenciação de classes no seu algoritmo de escalonamento.

O comportamento das linhas permite verificar que, a arquitetura proposta proporciona o menor atraso médio no atendimento das requisições rtPS, em relação as demais propostas, mesmo quando há mais de 140 conexões na rede. Com o emprego da arquitetura proposta, o atraso no atendimento das requisições rtPS aproxima-se de forma mais lenta do limite de 100ms, definido como prazo máximo para atendimento das requisições dessa classe. Por exemplo, com 150 conexões ativas, a arquitetura proposta atende às solicitações da fila rtPS com um atraso médio de aproximadamente 65ms, enquanto o algoritmo HUF proporciona um atraso médio de 95ms e Jae-Woo So atende as requisições rtPS com um atraso médio aproximado de 150s.

A definição de prazos para atendimento das requisições, empregada pela arquitetura proposta e por HUF, mantém o atraso médio no atendimento das requisições dentro de limites definidos. Esse aspecto é utilizado para o descarte das requisições mais antigas. Nesse contexto, o percentual de requisições descartadas é outra métrica importante para a avaliação de desempenho. Por isso, o percentual de requisições descartadas nas classes BE e rtPS é apresentado na Figura 6.7. Nesse gráfico, considera-se o percentual de requisições descartadas pertencente a cada classe de serviço, em relação ao total de requisições descartadas em cada quadro. Entretanto, como a proposta de Jae-Woo So não define qualquer critério para o descarte das requisições antigas, a análise é realizada considerando apenas a arquitetura proposta e HUF. O gráfico inicia em 40 conexões porque, neste cenário, quando há menos de 40 conexões na rede, não há descarte de requisições.

O gráfico mostra que o descarte de requisições pela arquitetura proposta inicia pela fila de requisições BE, que recebem a menor prioridade no atendimento, por isso, quando inicia-se o descarte de requisições, o percentual de requisições BE descartadas é igual a 100%. A arquitetura proposta inicia o descarte de requisições BE com 40 conexões na rede, enquanto o algoritmo HUF inicia esse descarte com 50 conexões na rede. Por outro lado, é possível verificar que, o algoritmo HUF inicia o descarte de requisições rtPS com um menor número de conexões na rede do que a arquitetura proposta. Por exemplo,

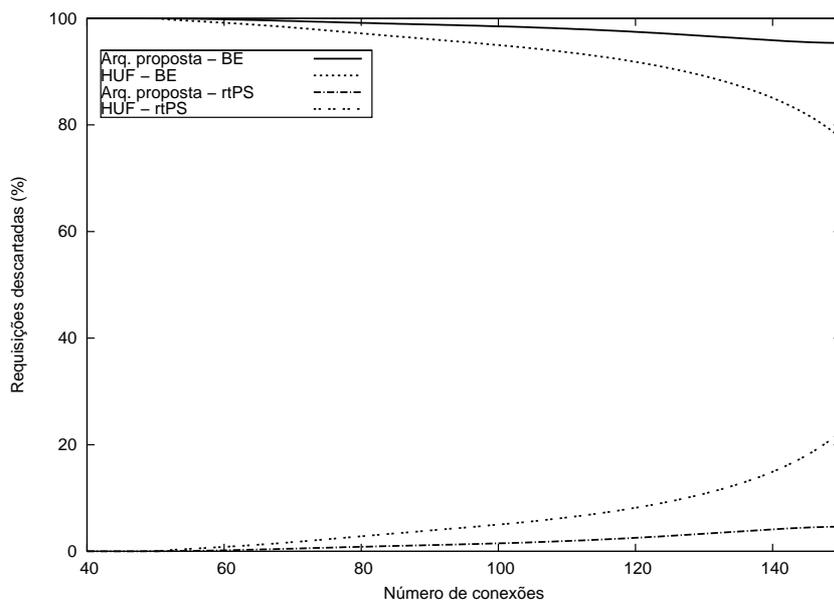


Figura 6.7: Percentual de requisições BE e rtPS descartadas

quando HUF é utilizado, o descarte de requisições rtPS inicia quando existem 50 conexões na rede. Enquanto isso, a arquitetura proposta inicia o descarte de requisições rtPS apenas quando existem 60 conexões ativas. Isso acontece, pelas diferenças na priorização das classes rtPS e BE empregadas pelas duas propostas.

O aumento no número de conexões gera um aumento no número total de requisições, e conseqüentemente, no número de requisições descartadas. Com isso, passam a ser descartadas requisições da classe rtPS, além de requisições BE, diminuindo o percentual de requisições BE descartadas em relação ao total de descartes. Entretanto, é possível observar que o aumento no número de conexões faz com que o algoritmo HUF passe a descartar um percentual maior de requisições rtPS do que a arquitetura proposta, enquanto o inverso ocorre com as requisições BE. Por exemplo, com 140 conexões na rede, na arquitetura proposta, um percentual inferior a 5% do total de requisições descartadas pertence a classe rtPS, enquanto para HUF, aproximadamente 17% do descarte ocorre na fila rtPS. Esse comportamento ocorre porque, o algoritmo HUF, apesar de priorizar o atendimento das requisições rtPS, também prioriza o atendimento das requisições que estão próximas de ultrapassar o prazo atendimento, independente da classe de serviço a que essa requisição pertence. Conseqüentemente, há um maior descarte das requisições da fila rtPS pelo algoritmo HUF do que pela arquitetura proposta.

As análises realizadas permitem verificar que a arquitetura proposta apresenta desempenho superior a proposta de Jae-Woo So. Além disso, proporciona atraso médio no atendimento de requisições BE muito semelhante ao obtido com o emprego do algoritmo HUF. Porém, com melhor aproveitamento em relação atraso médio no atendimento das requisições rtPS. Outro aspecto importante é que, a arquitetura *uplink* proposta neste trabalho mantém um menor percentual de requisições rtPS descartadas, em relação a HUF, mesmo com o aumento no número de conexões ativas. Por outro lado, o escopo considerado neste trabalho é mais abrangente do que o utilizado pelas propostas relacionadas. Dessa forma, a análise de desempenho apresentada foi obtida considerando aspectos do mecanismo de alocação de banda e organização do quadro OFDMA previstos pela arqui-

tetura proposta, uma vez que os trabalhos comparados não tratavam tais características.

7 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho foi proposta uma arquitetura de escalonamento hierárquica para transmissões *uplink*, em redes IEEE 802.16e, baseadas em OFDMA. O emprego da técnica de multiplexação OFDMA influencia na forma de utilização do quadro para alocar as informações que serão transmitidas. O emprego dessa técnica implica no uso de um mecanismo de alocação de banda específico, caracterizado pela informação de necessidades de transmissão através do canal de *ranging* e pela alocação dos *IEs CDMA*. Por isso, a técnica de multiplexação foi considerada na arquitetura de escalonamento *uplink*.

A arquitetura proposta foi organizada em dois níveis que atendem funcionalidades distintas. O *Nível 1* foi responsável pela alocação das rajadas UL, ou pela divisão do subquadro UL em oportunidades de requisição de banda e oportunidades de transmissão de dados. As oportunidades de requisição de banda são um aspecto importante do mecanismo de alocação de banda previsto pela tecnologia e são formadas por: canal de *ranging*, oportunidades de *polling* e *IE CDMA*. Após a alocação das oportunidades de requisição de banda, o *Nível 2* recebe a informação sobre a área do subquadro UL disponível para preenchimento com oportunidades de transmissão de dados. Com base nessa informação, o *Nível 2* foi projetado para ser responsável pelo gerenciamento das requisições de oportunidades de transmissão recebidas das MSs. O gerenciamento das requisições foi realizado com base em um algoritmo intra-classe, usado para ordenar as requisições pertencentes a uma mesma classe de serviço, e um algoritmo inter-classes, empregado para selecionar a classe de serviço para atendimento.

Os algoritmos empregados no *Nível 2*, foram baseados na proposta de Lin *et al.* (LIN *et al.*, 2008), pois essa apresenta uma estratégia de escalonamento eficiente, e adaptável as características da tecnologia. A arquitetura proposta neste trabalho realiza melhorias no algoritmo de escalonamento e aborda aspectos não considerados no projeto original, o que torna a proposta mais abrangente. Pois, são considerados na arquitetura de escalonamento *uplink*, a estrutura do quadro OFDMA, o mecanismo de alocação de banda definido pelo padrão IEEE 802.16e, além das cinco classes de serviço previstas pela tecnologia.

A avaliação de desempenho da proposta foi realizada com o emprego de uma ferramenta de simulação construída para esse fim. Essa avaliação considerou os diferentes aspectos da proposta e realizou uma análise comparativa com a proposta original de Lin *et al.* e o trabalho de Jae-Woo So (SO, 2008). Foram consideradas métricas como o atraso médio no atendimento das requisições, a utilização do subquadro UL, além do percentual de requisições descartadas.

A análise da utilização do subquadro UL foi realizada de duas formas, inicialmente, verificou-se o percentual do subquadro UL ocupado para transmissão de dados e pelo mecanismo de requisição de banda. Essa análise permitiu verificar que o percentual do subquadro UL utilizado pelo mecanismo de alocação de banda não foi desprezível, e por

isso, não pode ser ignorado pela estratégia de escalonamento. Além disso, o mecanismo de alocação de banda constitui um aspecto importante e diferencial da tecnologia IEEE 802.16e. Posteriormente, verificou-se o percentual do subquadro UL utilizado para transmissão de cada classe de serviço, em dois cenários diferentes, permitindo caracterizar a utilização do quadro, quando diferentes tipos de tráfego estão presentes na rede, ou quando diferentes estratégias de escalonamento são empregadas.

O percentual de requisições descartadas foi avaliado considerando a utilização da rede pelas cinco classes definidas pelo padrão e o emprego de diferentes arquiteturas de escalonamento. Essa análise permitiu verificar o impacto desses fatores no descarte das requisições quando a rede apresentava poucas ou muitas conexões ativas. Demonstrando a importância de empregar um mecanismo de priorização das requisições das filas com características de tempo real, quando a rede enfrenta congestionamentos.

O atraso médio no atendimento das requisições foi analisado para verificar o atendimento ou não dos requisitos de atraso máximo das requisições rtPS. Nesse aspecto, verificou-se um melhor desempenho da arquitetura proposta em relação aos trabalhos relacionados, em decorrência da maior prioridade concedida as requisições rtPS. Finalmente, a análise de desempenho realizada permitiu concluir que a arquitetura *uplink* apresentou um desempenho semelhante ou superior aos trabalhos relacionados, considerando as cinco classes de serviços, os mecanismos de requisição de banda e a construção do quadro OFDMA, especificados pelo padrão IEEE 802.16e.

Com relação a trabalhos futuros, a arquitetura *uplink* poderá ser expandida para considerar diferentes configurações AMC em um mesmo quadro. Essa característica tornará possível considerar a existência de MSs com diferentes condições de transmissão em um mesmo instante. A implementação dessa característica depende da associação de determinadas condições do canal a cada configuração AMC prevista pela tecnologia, o que não é definido pelo padrão.

O emprego de diferentes configurações AMC em um mesmo quadro, possibilitará incluir outro critério de escalonamento na arquitetura proposta. A seleção de requisições para atendimento poderia considerar as condições do canal de transmissão utilizado pela MS. Dessa forma, as decisões de escalonamento seriam adaptáveis as variações do canal de transmissão.

Finalmente, outros trabalhos futuros poderiam incluir módulos importantes em uma arquitetura de suporte a QoS como CAC, policiamento e conformação de tráfego. Pois, a inclusão desses módulos tornará o emprego da arquitetura de escalonamento mais eficiente, evitando, por exemplo, a sobrecarga de conexões na rede. Investigações nesse sentido estão em desenvolvimento pelo grupo de pesquisa do qual esse trabalho faz parte.

REFERÊNCIAS

ALANEN, O. Multicast Polling and Efficient VoIP Connections in IEEE 802.16 Networks. In: SYMPOSIUM ON MODELING, ANALYSIS, AND SIMULATION OF WIRELESS AND MOBILE SYSTEMS, 2007. **Proceedings...** [S.l.: s.n.], 2007. p.298–295.

ASH, G. R. **Traffic Engineering and QoS Optimization of Integrated Voice & Data Networks**. Amsterdam: Elsevier, 2007. 478p.

BACIOCCOLA, A.; CICONETTI, C.; ERTA, A.; LENZINI, L.; MINGOZZI, E.; MOILANEN, J. **A Downlink MAC Frame Allocation Framework in IEEE 802.16e OFDMA**: design and performance evaluation. New York: Springer, 2009. 300p.

BASSI, T. A. G. **Análise do desempenho do sistema 3G 1xEV-DV no enlace direto**. 2006. Dissertação de Mestrado — Universidade Estadual de Campinas - UNICAMP, Campinas.

BERARDINELLI, G.; TEMINO, L. A. M. R. D.; FRATTASI, S.; RAHMAN, M. I.; MOGENSEN, P. OFDMA vs. SC-FDMA: performance comparison in local area IMT-A scenarios. **IEEE Wireless Communications**, [S.l.]: IEEE, v.15, n.5, p.64–72, October 2008.

BOTH, C. B.; CAMARGO, E. G. de; KUNST, R.; GRANVILLE, L. Z.; ROCHOL, J. Analysis of WiMAX bandwidth allocation mechanism considering physical conditions. In: IEEE INTERNATIONAL SYMPOSIUM ON WIRELESS COMMUNICATION SYSTEMS, 2008, Reykjavik, Iceland. **Proceedings...** [S.l.]: IEEE, 2008. p.643–647.

CHENG, A. M. K. **Real-Time Systems**: scheduling, analysis and verification. Hoboken, NJ: Wiley-Interscience, 2002. 524p.

CICONETTI, C.; LENZINI, L.; MINGOZZI, E. Quality of Service Support in IEEE 802.16 Networks. **IEEE Network**, [S.l.]: IEEE, v.20, n.2, p.50–55, April 2006.

COHEN, R.; KATZIR, L. Computational Analysis and Efficient Algorithms for Micro and Macro OFDMA Scheduling. In: IEEE CONFERENCE ON COMPUTER COMMUNICATIONS, 2008, Phoenix, USA. **Proceedings...** [S.l.]: IEEE, 2008. p.511–519.

COMER, D. E. **Network Systems Design using Network Processors**. Upper Saddle River, NJ: Prentice Hall PTR, 2004. 514p.

DEVROYE, L. **Non-Uniform Random Variate Generation**. New York: Springer, 1986.

ESTEPA, A.; ESTEPA, R.; VOZMEDIANO, J. A new approach for VoIP traffic characterization. **IEEE Communications Letters**, [S.l.]: IEEE, v.8, n.10, p.644 – 646, October 2004.

GAN, W.; XIAN, J.; XIE, X.; RAN, J. A Cross-layer Designed Scheduling Algorithm for WiMAX Uplink. In: INTERNATIONAL CONFERENCE ON ELECTRONIC MEASUREMENT AND INSTRUMENTS, 2009, Beijing, China **Proceedings...** [S.l.]: IEEE, 2009. p.122–127.

GARG, S.; KAPPES, M. Can I add a VoIP call? In: IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS, 2003, Anchorage, Alaska. **Proceedings...** [S.l.]:IEEE, 2003. p.779–783.

GHOSH, A.; WOLTER, D.; ANDREWS, J.; CHEN, R. Broadband Wireless Access with WiMAX/802.16: current performance benchmarks and future potential. **IEEE Communications Magazine**, [S.l.]: IEEE, p.129–136, February 2005.

GIROUX, N.; GANTI, S. **Quality of Service in ATM Networks**. Upper Saddle River, NJ: Prentice Hall PTR, 1999. 252p.

IEEE. **IEEE Standard for Local and Metropolitan Area Networks - Part 16: air interface for fixed and mobile broadband wireless access systems - IEEE std. 802.16-2005**. New York, 2005.

KANG, K.; CHO, Y.; CHO, J.; SHIN, H. Scheduling Scalable Multimedia Streams for 3G Cellular Broadcast and Multicast Services. **IEEE Transactions on Vehicular Technology**, [S.l.]: IEEE, v.56, n.5, p.2655–2672, September 2007.

KIM, J.; CHO, D.-H. Piggybacking Scheme of MAP IE for Minimizing MAC Overhead in the IEEE 802.16e OFDMA Systems. In: VEHICULAR TECHNOLOGY CONFERENCE, 2007, Dublin, Ireland. **Proceedings...** [S.l.]: IEEE, 2007. p.284–288.

KIM, S.; KIM, W.; SUH, Y. An Efficient Bandwidth Request Mechanism for Non-Real-Time Services in IEEE 802.16 Systems. In: INTERNATIONAL CONFERENCE ON COMMUNICATION SYSTEMS SOFTWARE AND MIDDLEWARE, 2007, Bangalore, India. **Proceedings...** [S.l.]: IEEE, 2007. p.1–9.

KUNST, R. **Um injetor de erros aplicado à avaliação de desempenho do codificador de canal em redes IEEE 802.16**. 2009. Dissertação de Mestrado — Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre.

KUNST, R.; BOTH, C. B.; GRANVILLE, L. Z.; ROCHOL, J. Proposal of an error sequence generator applied to the performance analysis of IEEE 802.16. In: EUROPEAN WIRELESS CONFERENCE, 2009, Aalborg, Denmark. **Proceedings...** [S.l.]: IEEE, 2009. p.11–15.

LEE, H.; CHO, D. Reliable Multicast Services Using CDMA codes in IEEE 802.16 OFDMA system. In: VEHICULAR TECHNOLOGY CONFERENCE, 2005, Stockholm, Sweden. **Proceedings...** [S.l.]: IEEE, 2005. p.2349–2353.

LEI, L.; LIN, C.; CAI, J.; SHEN, X. S. Flow-level performance of opportunistic OFDM-TDMA and OFDMA networks. **IEEE Transactions on Wireless Communications**, [S.l.]: IEEE, v.7, n.12, p.5461–5472, December 2008.

LIN, Y.; WU, C.; LIN, Y.; LAI, Y. A Latency and Modulation Aware Bandwidth Allocation Algorithm for WiMAX Base Stations. In: WIRELESS COMMUNICATIONS AND NETWORKING CONFERENCE, 2008, Las Vegas, USA. **Proceedings...** [S.l.]: IEEE, 2008. p.1408–1413.

OHSEKI, T.; MORITA, M.; INOUE, T. Burst Construction and Packet Mapping Scheme for OFDMA Downlinks in IEEE 802.16 systems. In: IEEE GLOBAL TELECOMMUNICATIONS CONFERENCE, 2007, Washington, USA. **Proceedings...** [S.l.]: IEEE, 2007. p.4307–4311.

PANG, D.; HU, J.; ZHOU, J.; SHI, J.; DUTKIEWICZ, E. An uplink resource allocation scheme for SDMA-based IEEE 802.16 MIMO-OFDMA systems. In: IEEE GLOBAL COMMUNICATIONS CONFERENCE, 2008, New Orleans, USA. **Proceedings...** [S.l.]: IEEE, 2008. p.1–5.

PARAG, P.; BHASHYAM, S.; ARAVIND, R. A Subcarrier Allocation Algorithm for OFDMA using Buffer and Channel State Information. In: VEHICULAR TECHNOLOGY CONFERENCE, 2005, Stockholm, Sweden. **Proceedings...** [S.l.]: IEEE, 2005. p.622–625.

PIZZI, S.; MOLINARO, A.; IERA, A. On the performance of Compensation-Based and Greedy Scheduling Policies in IEEE 802.16 Networks. In: IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS, 2009, Dresden, Germany. **Proceedings...** [S.l.]: IEEE, 2009. p.1–6.

QIN, Z.; KUO, G.-S. Cross-Layer Design for QoS-Oriented Resource Allocation with Fairness Provision in IEEE 802.16 OFDMA Networks. In: INTERNATIONAL CONFERENCE ON NEXT GENERATION MOBILE APPLICATIONS, SERVICES AND TECHNOLOGIES, 2007, Cadiff, UK. **Proceedings...** [S.l.]: IEEE, 2007. p.287–291.

SALMON, B.; OLIVIER, J. Performance Analysis of Low Density Parity-Check Codes on a WiMAX Platform. In: WIRELESS COMMUNICATIONS AND NETWORKING CONFERENCE, 2007, Hong Kong, China. **Proceedings...** [S.l.]: IEEE, 2007. p.569–571.

SAYENKO, A.; ALANEN, O.; HAMALAINEN, T. Adaptive Contention Resolution for VoIP Services in the IEEE 802.16 Networks. In: IEEE INTERNATIONAL SYMPOSIUM ON A WORLD OF WIRELESS, MOBILE AND MULTIMEDIA NETWORKS, 2007, Helsinki, Finland. **Proceedings...** [S.l.]: IEEE, 2007. p.1–7.

SHI, J.; HU, A. Maximum Utility-based Resource Allocation Algorithm in the IEEE 802.16 OFDMA System. In: IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS, 2008, Beijing, China. **Proceedings...** [S.l.]: IEEE, 2008. p.311–316.

SO-IN, C.; JAIN, R.; TAMIMI, A.-K. Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey. **IEEE Journal on Selected Areas in Communications**, [S.l.]: IEEE, v.27, n.2, p.156–171, February 2009.

SO, J.-W. Performance Analysis of VoIP Services in the IEEE 802.16e OFDMA System With Inband Signaling. **IEEE Transactions on Vehicular Technology**, [S.l.]: IEEE, v.57, n.3, p.1876–1886, May 2008.

TARCHI, D.; FANTACCI, R.; BONCIANI, E. On the Ranging and Scheduling of data traffic in OFDMA Mobile Environments. In: WIRELESS COMMUNICATIONS AND NETWORKING CONFERENCE, 2008, Las Vegas, USA. **Proceedings...** [S.l.]: IEEE, 2008. p.1916–1921.

TSAI, C.; CHANG, C.; REN, F.; YEN, C. Adaptive Radio Resource Allocation for Downlink OFDMA/SDMA Systems with Multimedia Traffic. **IEEE Transactions on Wireless Communications**, [S.l.]: IEEE, v.7, n.5, p.1734–1743, May 2008.

WANG, H.; IVERSEN, V. B. Hierarchical Downlink Resource Management Framework for OFDMA based WiMAX Systems. In: WIRELESS COMMUNICATIONS AND NETWORKING CONFERENCE, 2008, Las Vegas, USA. **Proceedings...** [S.l.]: IEEE, 2008. p.1709–1715.

WiMAX Forum. **WiMAX System Evaluation Methodology V 2.1**. [S.l.: s.n.], 2008.

WU, J.; MO, J.; WANG, T. A method for non-real-time polling services in IEEE 802.16 Wireless Access Network. In: IEEE VEHICULAR TECHNOLOGY CONFERENCE, 2007, Dublin, Ireland. **Proceedings...** [S.l.]: IEEE, 2007. p.1518–1522.

YOU, J.; KIM, K.; KIM, K. Capacity Evaluation of the OFDMA-CDMA Ranging Subsystem in IEEE 802.16-2004. In: IEEE INTERNATIONAL CONFERENCE ON WIRELESS AND MOBILE COMPUTING, NETWORKING AND COMMUNICATIONS, 2005, Montreal, Canada. **Proceedings...** [S.l.]: IEEE, 2005. p.100–106.

APÊNDICE A ARTIGO PUBLICADO NO SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS - SBRC09

O artigo apresenta a idéia inicial para uma arquitetura de escalonamento *uplink*. Considera-se a divisão da arquitetura em dois níveis: alocação de rajadas e gerenciamento de requisições. Porém o gerenciamento de requisições é realizado pelo emprego de um algoritmo WFQ, para o escalonamento inter-classes, e dos algoritmos EDF e FIFO, para escalonamento intra-classe de requisições rtPS e BE, respectivamente. A arquitetura proposta nesse artigo considera o mecanismo de alocação de banda e a estrutura do quadro OFDMA e a avaliação de desempenho é realizada considerando duas classes de serviço, com e sem requisitos de QoS. A seguir são apresentados os dados do artigo:

- **Título:** Uma arquitetura de escalonamento hierárquica para transmissões *uplink* em redes IEEE 802.16e baseadas em OFDMA;
- **Autores:** Eliane Gobetti de Camargo, Cristiano Bonato Both, Rafael Kunst, Lisandro Zambenedetti Granville, Juergen Rochol
- **Local de submissão:** XXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos - SBRC09;

O artigo é reproduzido a seguir.

Uma Arquitetura de Escalonamento Hierárquica para Transmissões *Uplink* em Redes WiMAX Baseadas em OFDMA

Eliane Gobetti de Camargo, Cristiano Bonato Both, Rafael Kunst,
Lisandro Zambenedetti Granville, Juergen Rochol¹

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brazil

{egcamargo, cbboth, rkunst, granville, juergen}@inf.ufrgs.br

Abstract. *IEEE 802.16e is an emerging next generation network for providing access with QoS guarantees for both, fixed and mobile users. In the context of mobile applications, OFDMA should be applied as channel multiplexing technique, since it is able to carry data from many users within a transmission symbol. However, the resources allocation in OFDMA frame is not defined by IEEE 802.16e standard. Most of the current researches on OFDMA scheduling are focused in downlink transmissions or, otherwise, do not consider the frame structure of the technology. Therefore, this paper presents an uplink hierarchical scheduling architecture for OFDMA-based WiMAX networks. Moreover, we analyze the performance of the proposed architecture comparing the results with related works.*

Resumo. *A tecnologia IEEE 802.16e é um projeto emergente para redes de próxima geração com suporte à qualidade de serviço para usuários fixos e móveis. No contexto de aplicações móveis, a técnica OFDMA deve ser empregada para multiplexação de canais de rádio frequência, suportando múltiplos usuários em um mesmo símbolo. Entretanto, a alocação de recursos considerando a estrutura do quadro OFDMA não é definida pelo padrão IEEE 802.16e. A maioria das pesquisas atuais sobre arquiteturas de escalonamento baseadas em OFDMA estão focadas na transmissão downlink ou não consideram a estrutura do quadro definida para essa tecnologia. Sendo assim, neste artigo é apresentada uma arquitetura de escalonamento hierárquica para transmissões uplink em redes WiMAX baseadas em OFDMA. Finalmente, analisa-se o desempenho da arquitetura proposta, comparando com trabalhos relacionados.*

1. Introdução

IEEE 802.16e [IEEE 2005], também conhecido como WiMAX (*Worldwide Interoperability for Microwave Access*), é uma tecnologia emergente de próxima geração para redes sem fio, que suporta uma grande quantidade de usuários móveis e nomádicos (fixos) distribuídos através de uma extensa área geográfica. Além disso, esta tecnologia provê garantias estritas de QoS (*Quality of Service*) para aplicações de dados, voz e vídeo. Neste contexto, para garantir QoS, é necessário disponibilizar maiores quantidades de recursos, o que pode ser melhorado pela utilização da técnica de transmissão OFDMA (*Orthogonal Frequency Division Multiple Access*).

A implementação da técnica OFDMA é obrigatória em redes WiMAX móveis. Essa técnica se aplica bem a cenários com mobilidade, pois utiliza subportadoras ortogonais, tornando a tecnologia menos suscetível a ruídos e interferências do meio de transmissão [Berardinelli et al. 2008]. Além disso, a tecnologia permite que múltiplos usuários transmitam em um mesmo símbolo OFDMA, provendo uma melhora na distribuição da largura de banda do canal de rádio frequência. A organização do quadro OFDMA é realizada nos domínios tempo e frequência, dividindo o quadro em *downlink* (DL) e *uplink* (UL). As oportunidades de transmissão, no subquadro UL, utilizam o meio compartilhado e são divididas em duas partes: alocação de rajadas de dados e alocação de rajadas de requisições de banda. A organização de cada região do quadro é responsabilidade do escalonador e deve considerar os recursos físicos disponíveis, bem como os requisitos de QoS das aplicações. O padrão IEEE 802.16e define que uma arquitetura de escalonamento UL deve ser implementada para tratar as oportunidades de transmissão. Entretanto, a estrutura do escalonador e os algoritmos para atendimento das requisições não são definidos.

Este artigo propõe uma arquitetura de escalonamento hierárquica para transmissões UL em redes WiMAX baseadas em OFDMA. A proposta é organizada em dois níveis de escalonamento. No primeiro nível, trata-se o problema da alocação de recursos para requisição de banda e transmissão de dados, considerando a estrutura do subquadro UL. Para isso, é aplicado o algoritmo PQ (*Priority Queuing*). O segundo nível é responsável pelo gerenciamento das filas de requisições de banda recebidas em quadros anteriores. A implementação deste segundo nível combina os conceitos de WFQ (*Weighted Fair Queuing*) e EDF (*Earliest Deadline First*) para garantir os requisitos de QoS das aplicações.

O desempenho da arquitetura proposta é analisado considerando a transmissão de tráfego com requisitos de tempo real e BE (*Best Effort*). Os resultados mostram o atraso médio no atendimento das requisições de banda, o número de requisições atendidas e também o número de requisições descartadas. Finalmente, a partir das investigações realizadas pode-se observar o ganho de desempenho comparado com trabalhos relacionados.

O restante deste artigo está organizado da seguinte forma. Na Seção 2, descreve-se a fundamentação teórica sobre a estrutura do quadro OFDMA, além de apresentar alguns trabalhos relacionados ao escalonamento na tecnologia IEEE 802.16e. Na Seção 3, é apresentada a proposta de uma arquitetura de escalonamento hierárquica para transmissões UL. Na Seção 4, discute-se os resultados obtidos em relação aos trabalhos relacionados. Finalmente, na Seção 5, apresenta-se algumas conclusões e aponta-se direções para futuras investigações.

2. Fundamentação Teórica

O padrão IEEE 802.16e especifica as camadas física e de controle de acesso ao meio da tecnologia WiMAX, que oferece uma alternativa aos sistemas de acesso tradicionais, como por exemplo, *cable modem* e DSL (*Digital Subscriber Line*). Nesta seção são descritas as fundamentações teóricas sobre as características do quadro OFDMA, bem como trabalhos relacionados.

2.1. Estrutura do quadro OFDMA

A estrutura do quadro OFDMA precisa ser considerada em uma arquitetura de escalonamento, pois a composição dinâmica desse quadro é um recurso importante para o desempenho no atendimento de diferentes classes de serviços. Além disso, é importante ressaltar que outras técnicas de multiplexação, como por exemplo OFDM (*Orthogonal Frequency Division Multiplexing*), possuem uma estrutura de quadro diferente.

OFDMA é uma técnica de multiplexação de canal que possui como principal característica o suporte a múltiplos usuários em um mesmo símbolo OFDMA, tal característica é obtida pela divisão do canal sem fio em subcanais. Segundo o padrão IEEE 802.16e [IEEE 2005], os subcanais são formados por conjuntos de subportadoras. Esse processo de multiplexação é conhecido como subcanalização ou zonas de permutação. O padrão especifica 7 métodos de permutação das subportadoras, mas apenas o método PUSC (*Partially Used Subchannelization*) é de implementação obrigatória. Por isso, somente este método é considerado neste trabalho. O processo de subcanalização torna o mecanismo de alocação de banda mais flexível, porque as alocações passam a ser realizadas nos domínios de tempo (símbolos OFDMA) e frequência (subcanais), utilizando uma unidade de alocação mínima chamada *slot*. Na transmissão *downlink*, ou DL PUSC, cada *slot* é composto por 2 símbolos OFDMA e 1 subcanal, sendo que cada subcanal utiliza 24 subportadoras de dados. Já, na transmissão *uplink*, ou UL PUSC, cada *slot* é formado por 3 símbolos OFDMA que utilizam 48 subportadoras de dados, totalizando 16 subportadoras de dados por subcanal. A subcanalização e o mecanismo de alocação de banda determinam a estrutura do quadro OFDMA, que pode ser observada na Figura 1.

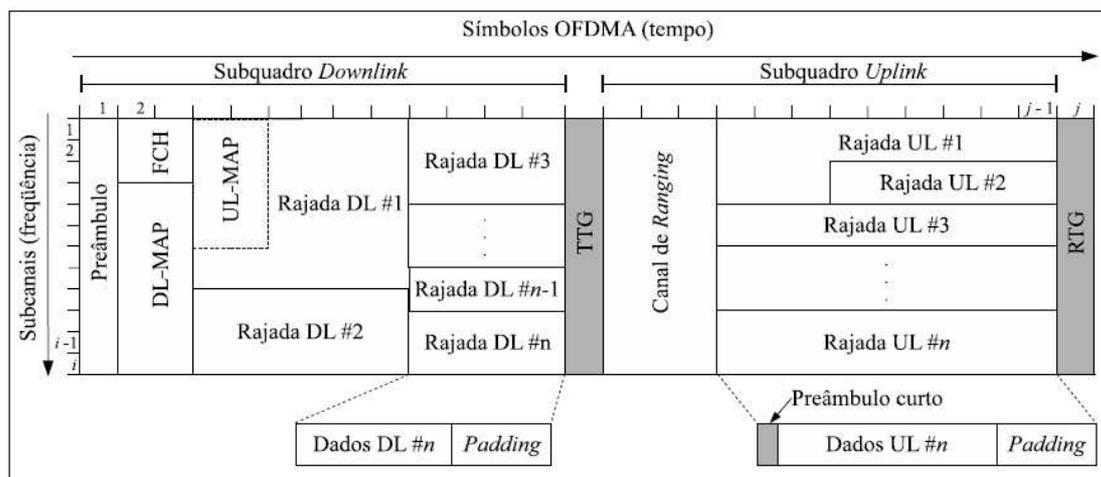


Figura 1. Estrutura do quadro OFDMA

O quadro OFDMA pode ser representado como uma matriz bidimensional, na qual os eixos x e y representam os domínios de tempo e frequência, respectivamente. A construção dos subquadros deve considerar, principalmente, as características de acesso ao meio. No subquadro DL apenas a BS (*Base Station*) transmite informações para todas as SSs (*Subscriber Station*), iniciando com um preâmbulo de sincronização formado por 1 símbolo de duração e transmitido em todos os subcanais. No segundo símbolo inicia a transmissão do FCH (*Frame Control Header*), utilizando 4 subcanais, como pode

ser observado na Figura 1. O FCH contém informações sobre a localização e duração da mensagem DL-MAP. Esse mapa contém informações sobre a localização, duração e configuração AMC (*Adaptive Modulation and Coding*) das rajadas DL. Uma configuração AMC indica o tipo de modulação e a razão de codificação utilizada na transmissão de uma determinada rajada de dados.

Já no subquadro UL, o meio de transmissão é compartilhado por todas as SSs que desejam transmitir informações para a BS. Por isso, o acesso das SSs ao subquadro UL precisa ser coordenado pela BS, através de um mecanismo de concessão de oportunidades de transmissão. Se um subquadro UL for alocado, a primeira rajada DL deve conter um mapa chamado UL-MAP. Esse mapa contém o mesmo tipo de informações que o DL-MAP, mas referente as rajadas UL. Cada rajada UL é formada por informações enviadas por apenas uma SS. Essas informações podem conter dados e nesse caso, a rajada UL é uma oportunidade de transmissão de dados concedida a uma determinada SS. Por outro lado, essa informação pode ser uma requisição de banda e dessa forma, a rajada UL é uma oportunidade de requisição de banda concedida a SS. Além das rajadas, um subquadro UL pode conter um canal de *ranging*, como no exemplo da Figura 1. Este canal, quando alocado no início do subquadro UL, conforme You *et al.* [You et al. 2005], minimiza a interferência na transmissão dos dados.

O canal de *ranging* é compartilhado por todas as SSs dentro da área de cobertura da BS, sendo utilizado para informação de códigos CDMA (*Code Division Multiple Access*). Uma SS informa um código CDMA no canal de *ranging* para alertar a BS que necessita realizar um dos quatro processos de *ranging*: (i) *ranging* inicial, (ii) *ranging* periódico, (iii) *handover* ou (iv) requisição de banda. A BS verifica a qual desses 4 grupos pertence o código CDMA informado e adota o procedimento adequado para atender a solicitação da SS. Por exemplo, se o código informado pertencer ao grupo de requisição de banda, a BS deverá alocar uma oportunidade, em um dos próximos subquadros UL, para que a SS envie seu pedido. Como o canal de *ranging* é compartilhado pelas SSs, podem ocorrer colisões. Por isso, este processo é conhecido como período de contenção. Além da contenção, os pedidos de banda podem ser enviados em oportunidades alocadas pela BS para uma determinada SS, sem a necessidade desta requisitar tal oportunidade e sem a probabilidade de ocorrerem colisões. Esse processo é conhecido como período de *polling* e destina-se aos fluxos de informações com altos requisitos de QoS.

Quando não houver informação suficiente para preencher completamente os subquadros DL ou UL, os mesmos serão preenchidos com *padding*. Além disso, entre cada subquadro existe um TTG (*Transmit/Receive Transition Gap*) alocado para permitir que a BS alterne do modo de transmissão para o de recepção. No final do subquadro UL, é alocado um RTG (*Receive/Transmit Transition Gap*) para que a BS retorne ao modo de transmissão.

Considerando o processo de subcanalização e o mecanismo de alocação de banda descritos anteriormente, uma arquitetura de escalonamento UL deve suportar, principalmente, três funções: (i) prover oportunidades de requisição de banda pelas SSs, (ii) gerenciar as filas de requisições de banda para transmissão de dados UL, e (iii) construir os subquadros UL. Na Subseção 2.2, são descritos trabalhos relacionados que propõem soluções para algumas características do escalonamento na tecnologia IEEE 802.16e.

2.2. Trabalhos Relacionados

Pesquisas recentes têm sido realizadas sobre escalonamento em OFDMA para redes WiMAX, abordando parcialmente as funções citadas na Subseção 2.1. Em Lin *et al.* [Lin *et al.* 2008] é proposto o algoritmo HUF (*Highest Urgency First*), projetado para alocar banda considerando um critério de urgência no atendimento de diferentes requisições. Além disso, o algoritmo considera as condições de propagação de um canal sem fio, através da análise da configuração AMC utilizada na transmissão. Jae-Woo So [So 2008], propõe uma arquitetura de escalonamento e, através de um modelo analítico, realiza uma análise de desempenho da sobrecarga de sinalização inserida para serviços VoIP em transmissões DL e UL. Einhaus e Klein [Einhaus e Klein 2006] conceberam um algoritmo de escalonamento básico para o desenvolvimento de um sistema de controle centralizado, considerando os efeitos do OFDMA. Pang *et al.* [Pang *et al.* 2008] apresentam um esquema para alocação dos recursos *uplink* de baixa complexidade. Entretanto, os trabalhos de Pang *et al.* e Einhaus e Klein não consideram a estrutura do quadro OFDMA.

As interfaces físicas previstas para redes WiMAX e a estrutura do quadro OFDMA são descritas por So-In *et al.* [So-In *et al.* 2009]. Os autores descrevem como essas características influenciam na proposta de uma estratégia de escalonamento adequada para a tecnologia. Além disso, são reunidas algumas propostas de mecanismos de escalonamento para o padrão IEEE 802.16e disponíveis na literatura. Da mesma forma, Lei *et al.* [Lei *et al.* 2008] apresentam um algoritmo de escalonamento oportunístico que pode ser adaptado para OFDM ou OFDMA. Porém, o algoritmo proposto precisa ser adaptado de acordo com a interface física empregada.

Outras abordagens propõem o gerenciamento dos recursos apenas para a transmissão DL. O trabalho proposto por Cohen e Katzir [Cohen e Katzir 2008] consiste em um algoritmo que busca resolver o problema do escalonamento DL dividindo-o em dois subproblemas: macro e micro escalonamento. Além disso, Wang e Iversen [Wang *et al.* 2008] sugerem um modelo para alocação dinâmica de recursos e para controle de admissão de conexões.

Analisando as pesquisas atuais, percebe-se que existem questões em aberto, principalmente porque há poucas investigações no sentido de propor uma solução para o problema do escalonamento de recursos em transmissões UL, considerando a estrutura do quadro OFDMA.

3. Proposta de uma Arquitetura de Escalonamento

Uma arquitetura de escalonamento hierárquica para transmissões UL em redes WiMAX baseadas em OFDMA, que atende as funções de escalonamento UL citadas na Subseção 2.1, é apresentada nesta seção. Essa arquitetura é voltada para a topologia PMP (*Point to Multipoint*), utilizando TDD (*Time Division Duplexing*) e deve ser implementada na BS, como pode ser observado na Figura 2. A proposta é composta de dois níveis hierárquicos para dividir o problema do escalonamento em dois subproblemas: (i) alocação dos recursos considerando a estrutura do quadro OFDMA, chamada de *Nível 1*, e (ii) gerenciamento das filas de requisição de banda, chamada de *Nível 2*.

O *Nível 1* aborda o problema da alocação de recursos para requisição de banda e transmissão de dados, considerando a estrutura do subquadro UL. Na Subseção 3.1

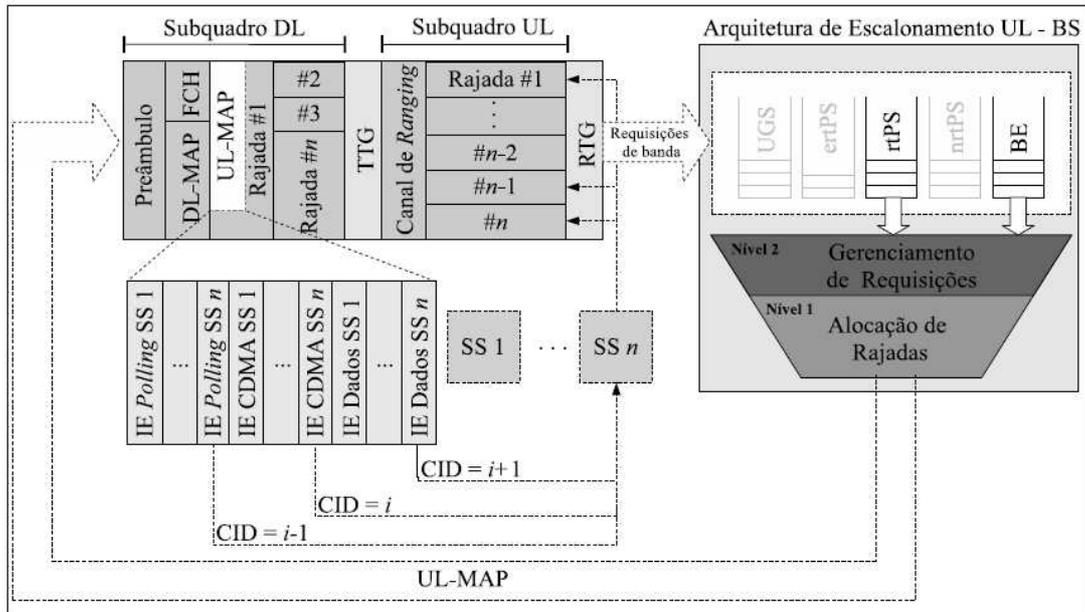


Figura 2. Estrutura do quadro OFDMA

são descritos o mecanismo e o algoritmo utilizados para garantir o funcionamento desse primeiro nível da arquitetura. Por outro lado, o *Nível 2* da arquitetura, apresentado na Subseção 3.2, é responsável pelo gerenciamento das filas de requisições de banda recebidas em quadros anteriores. No padrão IEEE 802.16e são definidas as filas: UGS (*Unsolicited Grant Service*), ertPS (*extended real-time Polling Service*), rtPS (*real-time Polling Service*), nrtPS (*non-real-time Polling Service*) e BE.

O objetivo da arquitetura proposta é a composição do UL-MAP referente ao próximo subquadro UL a ser transmitido. Esse subquadro possuirá as requisições de banda enviadas pelas SSs, que serão inseridas nas filas da BS. O UL-MAP é formado por estruturas de dados do tipo IE (*Information Element*). Na presente proposta, são considerados três tipos de IEs: IE *Polling*, IE CDMA e IE Dados que representam as oportunidades de transmissão concedidas a uma determinada conexão, de acordo com seu CID (*Connection Identifier*).

3.1. Alocação de Rajadas

O *Nível 1* da arquitetura proposta é responsável pela organização das rajadas a serem transmitidas no subquadro UL. Essa organização é realizada utilizando *slots* como unidade de alocação. A capacidade de transmissão de informações de cada *slot* é definida a partir dos símbolos OFDMA transmitidos em cada quadro. O número de símbolos OFDMA em cada quadro (N_{simb}) é calculado considerando a duração total de um símbolo (T_s), e a duração do quadro (T_q) [Both et al. 2008]. Sendo assim, N_{simb} é definido pela expressão 1.

$$N_{simb} = \left\lfloor \frac{T_q}{T_s} \right\rfloor \quad (1)$$

O número de bits transmitidos em cada símbolo OFDMA é obtido a partir da

vazão total da rede (V), calculada pela expressão 2 [Nuaymi e Noun 2006]. O valor de V é obtido em função da largura de banda disponível (B_W), do fator de amostragem empregado (n), do número total de subportadoras (N_{FFT}), da quantidade de subportadoras utilizadas para transmissão de dados (N_{dados}) e do percentual de informações redundantes inserido no símbolo OFDMA pelo prefixo cíclico (G). Além disso, a vazão é afetada pela configuração AMC utilizada, através da variação da razão de codificação do canal (R_{cod}) e da quantidade de bits associada a cada símbolo de modulação (M).

$$V = \left\lfloor \frac{B_W \cdot n}{8000} \right\rfloor \cdot 8000 \cdot \frac{N_{dados}}{N_{FFT}} \cdot \frac{1}{1 + G} \cdot R_{Cod} \cdot M \quad (2)$$

A partir do valor de V , dado em bits por segundo, pode-se calcular a quantidade de bits transmitida em um quadro OFDMA (b_q), pela expressão 3, onde T_q é medido em milissegundos.

$$b_q = V \cdot T_q \cdot 10^{-3} \quad (3)$$

Com os valores de b_q e N_{simb} , é possível obter a quantidade de bits transmitida em cada símbolo OFDMA (b_{simb}), conforme a expressão 4.

$$b_{simb} = \left\lfloor \frac{b_q}{N_{simb}} \right\rfloor \quad (4)$$

Um símbolo OFDMA compreende as informações transportadas pelos símbolos discretos de todas as subportadoras. O padrão IEEE 802.16e define que cada *slot* UL é composto por 3 símbolos OFDMA, transmitidos utilizando apenas as subportadoras de 1 subcanal. O número de subcanais utilizados em cada subquadro UL (N_{sc}), é dado pela expressão 5. Nesta expressão, o valor de α depende do método de subcanalização utilizado. No método UL PUSC, considerado neste trabalho, α é igual a 16 subportadoras de dados por subcanal.

$$N_{sc} = \left\lfloor \frac{N_{dados}}{\alpha} \right\rfloor \quad (5)$$

Considerando as expressões 4 e 5, pode-se calcular a capacidade de transmissão do *slot* UL ($C_{slot-ul}$), em bits, através da expressão 6.

$$C_{slot-ul} = \left\lfloor \frac{3 \cdot b_{simb}}{N_{sc}} \right\rfloor \quad (6)$$

A partir de $C_{slot-ul}$, o *Nível 1* da arquitetura proposta pode definir as áreas do subquadro UL utilizadas para requisição de banda e transmissão de rajadas de dados. Porém, as áreas para requisição de banda subdividem-se em: canal de *ranging*, requisições de banda utilizando *polling* e requisições de banda referentes aos códigos CDMA informados no canal de *ranging*, conforme pode ser visto na Figura 2. Os recursos destinados a cada uma dessas regiões são definidos pelo algoritmo de escalonamento PQ, conforme a

seguinte ordem de prioridade: (i) canal de *ranging*, (ii) IE *Polling*, (iii) IE CDMA e (iv) oportunidades para transmissão de dados.

Finalmente, o *Nível 1* deve fornecer ao *Nível 2* a quantidade de *slots* disponíveis para alocação de oportunidades de transmissão de dados ($N_{slot-dados}$), que serão distribuídas entre as SSs. Esse valor é calculado pela expressão 7.

$$N_{slot-dados} = \left\lfloor \frac{b_q}{C_{slot-ul}} \right\rfloor - L_{ranging} - 2 \cdot (L_{polling} + L_{CDMA}) \quad (7)$$

A quantidade de *slots* atribuídos ao canal de *ranging* ($L_{ranging}$) foi fixada em 12, distribuídos em 3 símbolos OFDMA e 12 subcanais. Cada *slot* do canal de *ranging* é capaz de receber a informação de até 30 códigos CDMA diferentes, porém, existe a probabilidade de que mais de uma SS informe o mesmo código utilizando a mesma oportunidade, causando colisões [You et al. 2005]. Na expressão 7, $L_{polling}$ representa o número de conexões transmitindo fluxos de dados baseados em *polling* e L_{CDMA} corresponde ao número de códigos CDMA que estão na fila da BS aguardando por uma oportunidade de requisição de banda. Os valores $L_{polling}$ e L_{CDMA} devem ser multiplicados por 2, tendo em vista que o atendimento de cada requisição é realizado em 2 *slots*: o primeiro contendo um preâmbulo e o segundo, o cabeçalho de requisição de banda.

3.2. Gerenciamento de Requisições

O *Nível 2* da arquitetura proposta foi projetado para definir quais requisições de banda devem ser atendidas no próximo subquadro UL. Essa definição é baseada na classificação dos tráfegos de acordo com as filas de requisição de banda, correspondendo às classes de serviços do padrão IEEE 802.16e. A arquitetura foi definida para suportar uma fila para cada classe de serviço. Neste artigo, considera-se as filas rtPS e BE, que correspondem aos tráfegos com maior e menor requisitos de QoS, dentre as classes de serviço que não possuem largura de banda garantida.

O Gerenciamento de Requisições recebe do *Nível 1* a quantidade de recursos disponíveis para alocação de oportunidades de transmissão de dados e, posteriormente, calcula a quantidade de *slots* necessária para cada transmissão de dados ($N_{slot-req}$). Esse cálculo é realizado considerando a quantidade de bits solicitada a cada requisição de banda (b_{req}), conforme a expressão 8.

$$N_{slot-req} = \left\lceil \frac{b_{req}}{C_{slot-ul}} \right\rceil \quad (8)$$

A seleção de qual fila deve ser atendida, dentre as que estão sendo utilizadas, é baseada na aplicação do algoritmo WFQ. A escolha desse algoritmo prioriza o tráfego rtPS, que demanda mais restrições de QoS em relação as aplicações com tráfego BE, além de evitar postergação infinita desse tráfego. A priorização é realizada pela atribuição de peso 4 para a fila de requisições rtPS, enquanto a fila de requisições BE possui peso 1. Dessa forma, 4 requisições rtPS serão atendidas para cada requisição BE. Esse algoritmo é aplicado enquanto existirem *slots* disponíveis em um subquadro UL para a transmissão de dados pelas SSs, respeitando a inequação 9, onde δ representa o número de requisições que estão aguardando atendimento em determinada fila da BS.

$$\left(\sum_{i=1}^{i=\delta_{rtPS}} \left\lceil \frac{b_{req_i}}{C_{slot-ul}} \right\rceil \right) + \left(\sum_{j=1}^{j=\delta_{BE}} \left\lceil \frac{b_{req_j}}{C_{slot-ul}} \right\rceil \right) \leq N_{slot-dados} \quad (9)$$

Além disso, a priorização das requisições de banda dentro da fila rtPS é obtida pelo emprego do algoritmo EDF, para garantir os requisitos de tempo real dessa classe de serviço. O algoritmo EDF prioriza as requisições pelo estabelecimento de prazos para atendimento (P_a), definidos pela expressão 10, que considera o atraso máximo em que uma requisição rtPS pode ser atendida (A_{max}) e a duração do quadro (T_q).

$$P_a = \frac{A_{max}}{T_q} \quad (10)$$

A cada requisição rtPS é associado um P_a , assim que esta é inserida na fila do escalonador. Após cada T_q em que a requisição permanece nessa fila, o prazo de atendimento é decrementado em uma unidade. Quando $P_a = 0$, a requisição é descartada, pois sua validade foi superada. Por outro lado, na fila BE, as requisições são classificadas utilizando o algoritmo FIFO (*First In First Out*) para garantir justiça no atendimento dos fluxos desse serviço.

Os dois níveis da arquitetura de escalonamento proposta são projetados para cobrir os aspectos que não são definidos pelo padrão IEEE 802.16e para uma transmissão UL sobre um canal sem fio baseado em OFDMA. Na Seção 4, é analisado o desempenho da arquitetura, discutindo aspectos como atraso e número de requisições atendidas.

4. Resultados

O desempenho da arquitetura proposta é analisado em comparação com o algoritmo HUF [Lin et al. 2008] e com o trabalho de Jae-Woo So [So 2008], uma vez que essas abordagens referem-se ao escalonamento UL e consideram aspectos dos dois níveis da arquitetura proposta. Os resultados são obtidos através de simulações implementadas no ambiente de programação Labview, levando em consideração o modelo apresentado na Seção 3. Como estudo de caso, considera-se as classes de serviço rtPS e BE, que representam tráfegos com e sem requisitos de QoS, respectivamente. O tráfego rtPS foi modelado por pacotes VoIP de 160 bytes, utilizando o CODEC G.711, gerados a cada 20ms. Além disso, foi adicionado tráfego de baixa prioridade composto por pacotes BE de 256 bytes, em uma proporção de 50% das conexões, utilizada em todos os cenários analisados, como sugerido por Both *et. al.* [Both et al. 2008]. Os demais parâmetros de simulação são apresentados na Tabela 1.

O primeiro aspecto considerado para análise de desempenho é o percentual de requisições rtPS atendidas por uma BS, em relação ao número de conexões ativas na rede, como pode ser verificado na Figura 3. O gráfico apresenta a diferença no percentual de atendimento das requisições entre a arquitetura proposta e as demais abordagens analisadas. Percebe-se que o percentual de requisições atendidas é maior quando uma pequena quantidade de conexões está ativa. A medida que aumenta o número de conexões, o percentual de requisições atendidas é reduzido. Esse comportamento ocorre porque havendo um maior número de conexões, mais recursos precisam ser destinados às oportunidades de requisição de banda. Conseqüentemente, mais requisições são recebidas pela

Tabela 1. Parâmetros de Simulação

Parâmetro	Valor
Largura de banda (B_w)	3,5MHz
Taxa de amostragem (n)	8/7
Subportadoras total (N_{FFT})	512
Subportadoras de dados (N_{dados})	360
Subportadoras de dados por subcanal UL (α)	16
Duração do prefixo cíclico (G)	1/16 ($8\mu s$)
Duração do símbolo OFDMA (T_s)	$136\mu s$
Duração do quadro (T_q)	20ms
Configuração AMC	16-QAM 3/4 ($M = 4$, $R_{Cod} = 3/4$)
Proporção do quadro (DL:UL)	1:1
Atraso máximo para atendimento rtPS (A_{max})	155ms
Intervalo de confiança	95%

BS, que necessita priorizar o atendimento, deixando de servir a totalidade das requisições no subquadro seguinte.

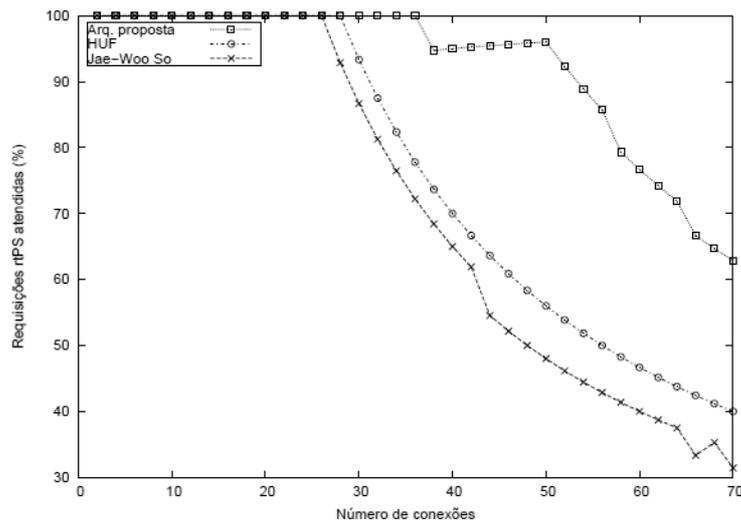


Figura 3. Requisições rtPS atendidas

A arquitetura proposta mantém o desempenho ótimo até aproximadamente 36 conexões, ou seja, todas as requisições rtPS recebidas em um subquadro são atendidas no próximo subquadro. Esse valor representa um ganho de aproximadamente 22% em relação ao algoritmo HUF e cerca de 28% sobre a abordagem de Jae-Woo So. Além disso, a maior vantagem é observada quando há 50 conexões, situação em que 96% das requisições são atendidas na arquitetura proposta, o que representa uma melhora em torno de 42% considerando HUF e 50% em comparação a Jae-Woo So. O pior caso apresentado ocorre quando há 70 conexões ativas. Nesse caso, a arquitetura proposta atende aproximadamente 63% das requisições, apresentando vantagens em torno de 36% e 50%, em relação às propostas HUF e Jae-Woo So, respectivamente.

O número de requisições atendidas é uma métrica importante para avaliação de desempenho. Entretanto, ela deve ser analisada em conjunto com a quantidade de requisições descartadas. O percentual de requisições rtPS descartadas, em relação ao número de conexões ativas, é mostrado na Figura 4. Nesse cenário, não é considerada a abordagem de Jae-Woo So, pois a mesma não define um critério para descarte de requisições. Para as demais abordagens, é utilizado o parâmetro de prazo para atendimento, calculado em função de A_{max} .

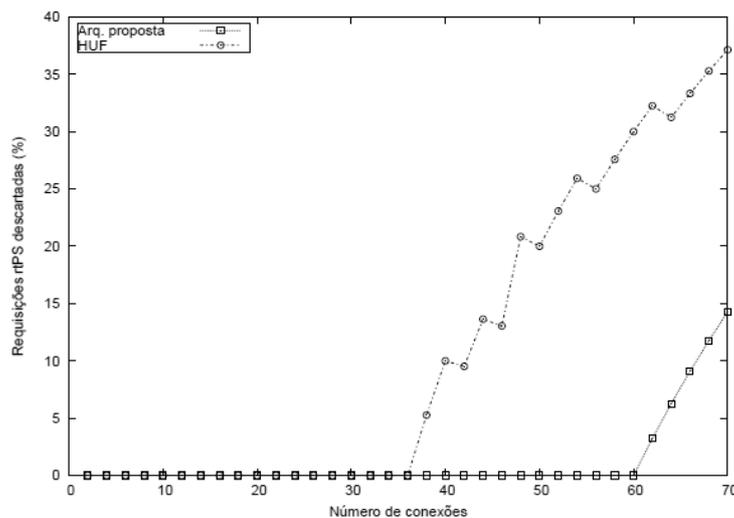


Figura 4. Requisições rtPS descartadas

A Figura 4 mostra que, até 36 conexões, todas as requisições são atendidas, nos dois casos analisados. Após esse ponto, o algoritmo HUF começa a descartar requisições, enquanto a arquitetura proposta só inicia esse processo a partir de 60 conexões, representando um ganho de 40%. O descarte ocorre pelo não atendimento das requisições rtPS dentro do atraso máximo, fixado em 155ms, conforme sugerido por Ali [Ali et al. 2008]. Além disso, percebe-se que o comportamento das curvas, após o primeiro descarte, possui um crescimento acentuado. Por exemplo, quando há 70 conexões, a arquitetura proposta mantém um desempenho cerca de 62% superior ao HUF.

Outra investigação realizada refere-se ao atraso médio no atendimento de requisições rtPS em relação ao número de conexões. Esse comportamento é apresentado na Figura 5. Pode-se observar que até 26 conexões ativas, o atraso é mínimo nos três casos, pois as requisições recebidas são atendidas no quadro seguinte. Após esse ponto, ocorre um crescimento no atraso médio das requisições.

Analisando a Figura 5, percebe-se que a arquitetura Jae-Woo So, após 26 conexões, apresenta um atraso médio sempre crescente. Esse comportamento ocorre devido a ausência de um critério de atraso máximo. Entretanto, nas outras propostas, o atraso médio, tende a convergir para o limite de 155ms, sendo que a proposta deste trabalho converge mais lentamente que o algoritmo HUF. A arquitetura apresentada neste artigo, mantém o atraso mínimo até 50 conexões, o que representa um ganho de aproximadamente 42% em comparação com o desempenho de HUF. Cabe ressaltar que o atraso apresentado no gráfico é inferior ao limite máximo definido, pois trata-se de um valor

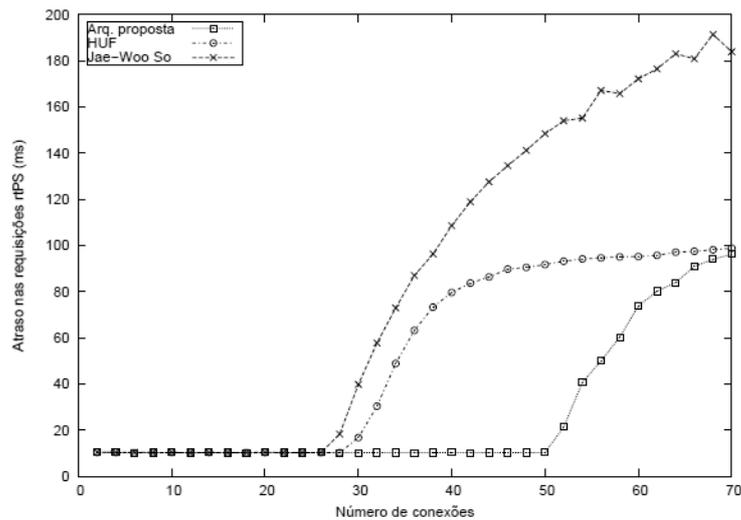


Figura 5. Atrazo médio no atendimento das requisições rtPS

médio.

A última investigação realizada diz respeito à comparação do desempenho entre alocações de oportunidades de transmissão de dados das classes rtPS e BE, considerando o número de conexões ativas, como pode ser observado na Figura 6. Nesse caso, o comportamento das alocações para transmissão de dados rtPS é apresentado na Figura 6 (a), enquanto as curvas relativas à classe de serviço BE são mostradas na Figura 6 (b).

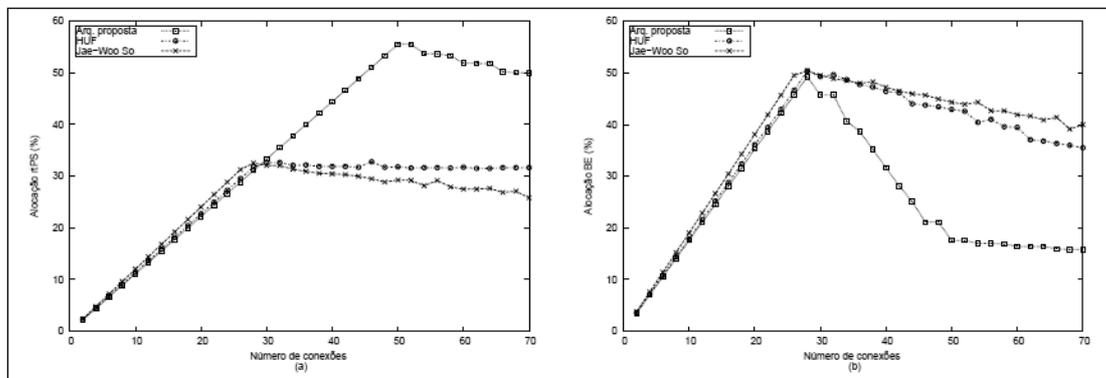


Figura 6. Percentual do subquadro UL alocado para transmissões rtPS e BE

O comportamento das três proposta é semelhante quando o número de conexões é inferior a 30. Nesse caso, observa-se que os recursos disponíveis na rede são suficientes para atender a todas as oportunidades de requisição rtPS e BE no quadro seguinte. Pode-se constatar que a arquitetura proposta prioriza a classe de tráfego de tempo real, diferentemente das demais abordagens. Jae-Wo So não diferencia as classes de serviço, enquanto o algoritmo HUF utiliza como um dos parâmetros para priorização de requisições o tamanho do pacote a ser transmitido. Analisando-se a Figura 6 (a), percebe-se que, a arquitetura proposta possui um ganho de até 43% em relação a HUF e 47% se comparado com Jae-Woo So.

Nessas investigações, considerou-se uma largura de banda de 3,5MHz em todos os cenários. Por isso, quando se prioriza uma classe de serviço, ocorre uma penalização da outra classe, de forma proporcional. A arquitetura proposta não prioriza o atendimento de requisições BE, pois essas não possuem requisitos de QoS, diferentemente das demais abordagens, como pode ser visto na Figura 6 (b).

O percentual do subquadro UL alocado para fluxos de dados rtPS e BE é menor do que 100%. Isso ocorre porque parte do subquadro é alocada para canal de *ranging* e oportunidades de requisição de banda. Cabe salientar, que o número de requisições cresce com o aumento do número de conexões. Sendo assim, se um mecanismo de controle de admissão não for implementado para limitar o número de conexões, a tendência é que todo o subquadro UL seja ocupado por oportunidades de requisição de banda, não ocorrendo transmissão de dados.

5. Conclusões e Trabalhos Futuros

Neste artigo, foi proposta uma arquitetura de escalonamento hierárquica para transmissões UL em redes WiMAX baseadas em OFDMA. O principal objetivo da arquitetura é o escalonamento UL, considerando requisições de banda para priorização de tráfego, a fim de prover estritos requisitos de QoS. Sendo assim, foi proposta uma hierarquia composta por dois níveis: o *Nível 1* é responsável pela alocação dos recursos considerando a estrutura do quadro OFDMA, enquanto o *Nível 2* é responsável pelo gerenciamento das filas de requisição de banda.

O desempenho da arquitetura proposta foi analisado, considerando as métricas de atraso médio no atendimento das requisições de banda, número de requisições atendidas e quantidade de requisições descartadas. Os resultados foram comparados com dois trabalhos relacionados, mostrando que o desempenho da arquitetura proposta é superior às demais abordagens considerando tráfego rtPS. Esse desempenho é alcançado porque os critérios utilizados na priorização de tráfego são otimizados para transmissões de tempo real. Além disso, este artigo demonstra a importância de considerar os detalhes da estrutura do quadro OFDMA no processo de escalonamento.

Direções para futuras investigações devem envolver a análise de desempenho da arquitetura proposta considerando as classes de serviço que não foram abordadas neste trabalho. Outra característica importante, que pode ser integrada à arquitetura proposta, é um algoritmo de escalonamento para transmissões DL que aloque as rajadas de dados do subquadro DL. Finalmente, sugere-se a definição de um mecanismo de controle de admissão que leve em conta a alocação de recursos em um quadro OFDMA, com o objetivo de evitar que todo o quadro seja utilizado para requisições de banda.

Agradecimentos

Os autores gostariam de agradecer à Universidade Federal do Rio Grande do Sul (UFRGS) e à Universidade de Santa Cruz do Sul (UNISC) pelo apoio financeiro.

Referências

- Ali, N., Hayajneh, M., and Hassanein, H. (2008). Cross layer scheduling algorithm for IEEE 802.16. In *Proceedings of IEEE International Conference on Communications, 2008. ICC '08*.

- Berardinelli, G., Temino, L. A. M. R. D., Frattasi, S., Rahman, M. I., and Mogensen, P. (2008). OFDMA vs. SC-FDMA: performance comparison in local area IMT-A scenarios. *IEEE Wireless Communications*, 15(5):64–72.
- Both, C. B., de Camargo, E. G., Kunst, R., Granville, L. Z., and Rochol, J. (2008). Analysis of WiMAX bandwidth allocation mechanism considering physical conditions. In *Proceedings of the IEEE International Symposium on Wireless Communication Systems 2008 - ISWCS*.
- Cohen, R. and Katzir, L. (2008). Computational analysis and efficient algorithms for micro and macro OFDMA scheduling. In *Proceedings of the IEEE INFOCOM - Conference on Computer Communications*.
- Einhaus, M. and Klein, O. (2006). Performance evaluation of a basic OFDMA scheduling algorithm for packet data transmissions. In *Proceedings of the 11th IEEE Symposium on Computers and Communications (ISCC)*.
- IEEE (2005). IEEE standard for local and metropolitan area networks - part 16: Air interface for fixed and mobile broadband wireless access systems - IEEE std. 802.16-2005.
- Lei, L., Lin, C., Cai, J., and Shen, X. S. (2008). Flow-level performance of opportunistic OFDM-TDMA and OFDMA networks. *IEEE Transactions on Wireless Communications*, 7(12):5461–5472.
- Lin, Y., Wu, C., Lin, Y., and Lai, Y. (2008). A latency and modulation aware bandwidth allocation algorithm for WiMAX base stations. In *Proceedings of the WCNC - Wireless Communications and Networking Conference*.
- Nuaymi, L. and Noun, Z. (2006). Enabling large-scale wireless broadband: The case for TAPs. In *Proceedings of 17th Annual IEEE Symposium of Personal, Indoor and Mobile Radio Communications*, Finland.
- Pang, D., Hu, J., Zhou, J., Shi, J., and Dutkiewicz, E. (2008). an uplink resource allocation scheme for SDMA-based IEEE 802.16 MIMO-OFDMA systems. In *Proceedings of the IEEE GLOBECOM - Global Communications Conference*.
- So, J.-W. (2008). Performance analysis of VoIP services in the IEEE 802.16e OFDMA system with inband signaling. *IEEE Transactions on Vehicular Technology*.
- So-In, C., Jain, R., and Tamimi, A.-K. (2009). Scheduling in IEEE 802.16e mobile WiMAX networks: key issues and a survey. *IEEE Journal on Selected Areas in Communications*, 27(2):156–171.
- Wang, H., , and Iversen, V. B. (2008). Hierarchical downlink resource management framework for OFDMA based WiMAX systems. In *Proceedings of the WCNC - Wireless Communications and Networking Conference*.
- You, J., Kim, K., and Kim, K. (2005). Capacity evaluation of the OFDMA-CDMA ranging subsystem in IEEE 802.16-2004. In *Proceedings of the IEEE International Conference on Wireless And Mobile Computing, Networking And Communications - WiMob*.