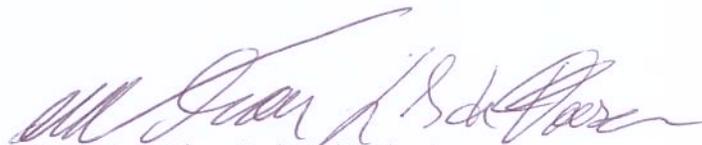


Mecanismos para provisão de Qualidade de Serviço em redes IEEE 802.16

Este exemplar corresponde à redação final da Tese devidamente corrigida e defendida por Juliana Freitag Borin e aprovada pela Banca Examinadora.

Campinas, 25 de janeiro de 2010.



Prof. Dr. Nelson Luis Saldanha da Fonseca
IC, UNICAMP (Orientador)

Tese apresentada ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Doutora em Ciência da Computação.

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**
Bibliotecária: Miriam Cristina Alves – CRB8 / 5094

Borin, Juliana Freitag

B644m Mecanismos para a provisão de qualidade de serviço em redes
IEEE 802.16/Juliana Freitag Borin-- Campinas, [S.P. : s.n.], 2010.

Orientador : Nelson Luis Saldanha da Fonseca

Tese (Doutorado)- Universidade Estadual de Campinas, Instituto de
Computação.

1.Redes de Computação. 2. Sistemas de comunicação sem fio. 3.
Redes metropolitanas de computação. . I. Fonseca, Nelson Luis Saldanha
da. II. Universidade Estadual de Campinas. Instituto de Computação.
III. Título.

Título em inglês: Mechanisms for quality of service provision in IEEE 802.16 networks

Palavras-chave em inglês (Keywords): 1. Computer networks. 2. Wireless communication
systems. 3. Metropolitan area networks.

Área de concentração: Redes de computadores

Titulação: Doutora em Ciência da Computação

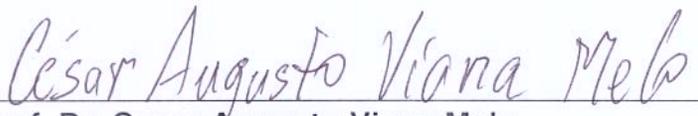
Banca examinadora: Prof. Dr. Nelson Luis Saldanha da Fonseca (IC-UNICAMP)
Prof. Dr. Cesar Augusto Viana Melo (UFAM)
Prof. Dr. Edmundo Roberto Mauro Madeira (IC-UNICAMP)
Prof. Dr. Omar Branquinho (PUCC)
Prof. Dr. Ricardo de Oliveira Anido (IC-UNICAMP)

Data da defesa: 25/01/2010

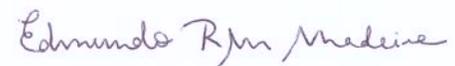
Programa de Pós-Graduação: Doutorado em Ciência da Computação

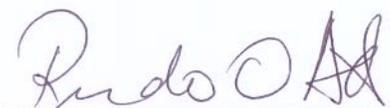
TERMO DE APROVAÇÃO

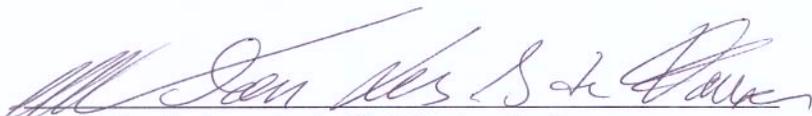
Tese Defendida e Aprovada em 25 de janeiro de 2010, pela Banca examinadora composta pelos Professores Doutores:


Prof. Dr. Cesar Augusto Viana Melo
DCC / UFAM


Prof. Dr. Omar Carvalho Branquinho
PUCC


Prof. Dr. Edmundo Roberto Mauro Madeira
IC / UNICAMP


Prof. Dr. Ricardo de Oliveira Anido
IC / UNICAMP


Prof. Dr. Nelson Luis Saldanha da Fonseca
IC / UNICAMP

Mecanismos para provisão de Qualidade de Serviço em redes IEEE 802.16

Juliana Freitag Borin¹

Janeiro de 2010

Banca Examinadora:

- Prof. Dr. Nelson Luis Saldanha da Fonseca
IC, UNICAMP (Orientador)
- Prof. Dr. César Augusto Viana Melo
DCC, UFAM
- Prof. Dr. Edmundo Roberto Mauro Madeira
IC, UNICAMP
- Prof. Dr. Omar Branquinho
PUCC
- Prof. Dr. Ricardo de Oliveira Anido
IC, UNICAMP

¹Financiado por CAPES e UOL Bolsa Pesquisa

Resumo

O padrão IEEE 802.16 para acesso sem fio em banda larga, também conhecido como WiMAX (*Worldwide Interoperability for Microwave Access Forum*), é uma solução de baixo custo para o acesso à Internet em áreas metropolitanas e rurais. Outro atrativo desta tecnologia está no fato de incluir uma arquitetura para provisão de Qualidade de Serviço (QoS) para aplicações de tempo real e aplicações que demandam altas taxas de transmissão. Esta arquitetura de QoS inclui um mecanismo de sinalização entre a estação base e as estações cliente, cinco níveis de serviço, para priorização e diferenciação do tráfego, um mecanismo de escalonamento e um mecanismo de controle de admissão de novas conexões. Entretanto, as políticas que devem ser utilizadas na implementação dos mecanismos de escalonamento e controle de admissão não foram definidas pelo padrão.

O mecanismo de escalonamento tem como objetivo garantir a banda passante requisitada pelas estações cliente bem como promover o uso eficiente do enlace sem fio, enquanto o mecanismo de controle de admissão restringe o número de conexões simultaneamente presentes na rede, de forma a evitar a saturação do enlace sem fio e, conseqüentemente, a violação dos requisitos de QoS.

Esta tese propõe políticas para os mecanismos de escalonamento e de controle de admissão para a provisão de QoS em redes baseadas no padrão IEEE 802.16.

A política de escalonamento proposta nesta tese é compatível com as especificações do padrão IEEE 802.16 e é capaz de atender os cinco tipos de serviço bem como o conjunto de parâmetros de QoS a eles associado. Requisitos de latência máxima e taxa mínima são garantidos para cada conexão aceita na rede, sem que os limites impostos pelos requisitos de taxa máxima e tamanho máximo da rajada sejam violados.

Cinco soluções são propostas para o mecanismo de controle de admissão. Estas soluções visam atingir três objetivos principais: restringir o número de conexões simultaneamente presentes na rede para que o escalonador tenha recursos suficientes para atender os requisitos de QoS de cada conexão, atender as expectativas do provedor de serviço através da maximização do lucro e maximizar a satisfação dos usuários através da disponibilização de recursos extras.

As soluções propostas foram avaliadas através de experimentos de simulação, utilizando um módulo para simulação de redes IEEE 802.16 desenvolvido para conduzir esta pesquisa.

Abstract

The IEEE 802.16 standard for broadband wireless access, also known as WiMAX (Worldwide Interoperability for Microwave Access Forum), is a low cost solution for Internet access in metropolitan and rural areas. Another advantage of this technology is the fact that it includes an architecture for the provision of Quality of Service (QoS) for real time and bandwidth demanding applications. This QoS architecture includes a signaling mechanism between base station and subscriber stations, five types of service, for traffic prioritization and differentiation, a scheduling mechanism and a connection admission control mechanism. However, policies for the implementation of these mechanisms are not defined by the standard.

The scheduling mechanism aims at guaranteeing the bandwidth required by the subscriber stations as well as enabling the efficient wireless link usage, while the connection admission control restricts the number of users simultaneously present in the network so as to avoid the wireless link saturation and, consequently, violation of QoS contracts.

This theses proposes policies for scheduling and connection admission control mechanisms for the provision of QoS in IEEE 802.16 networks.

The proposed scheduling policy is standard compliant and supports the five types of service as well as the QoS parameters associated with them. Maximum latency and minimum rate requirements are guaranteed without violating maximum rate and maximum traffic burst requirements.

Five solutions are proposed for the connection admission control mechanism. These solutions aim to support three main goals: to restrict the number of simultaneous connections in the system so that the resources available to the scheduler are sufficient to guarantee the QoS requirements of each connection, to support the service provider expectations by maximizing the revenue, and to maximize the users satisfaction by providing extra resources.

The proposed solutions were evaluated through simulation experiments using a module for the simulation of IEEE 802.16 networks developed to conduct this research.

À minha mãe Leonida e ao meu pai Lotário.

Agradecimentos

Se cheguei até aqui, no final dessa importante etapa da minha vida, é porque eu não estive sozinha nesse período em que realizei meu Doutorado. Neste espaço quero agradecer a todos que de alguma forma contribuíram para a conclusão desse trabalho.

A Deus, principal responsável por essa conquista. Aquele que me ofereceu um caminho repleto de amigos, oportunidades e dificuldades que me ajudaram a evoluir tanto no lado pessoal como no lado profissional.

Aos meus pais, Lotário e Leonida, que nunca mediram esforços para me dar uma boa educação. Aos meus pais e meus irmãos, Fabiano e Rafael, por me oferecerem um porto seguro, para onde eu sabia que sempre poderia voltar e que sempre serviu de fonte para a recarga das minhas energias.

Ao Borin, meu esposo, pelo amor, pela paciência, pela amizade, por vibrar com minhas conquistas, secar minhas lágrimas nos momentos de tristeza, me encorajar nos momentos de desânimo. Também por, ao final dessa jornada, me dar um lindo presente, nosso primeiro filho, que ainda não nasceu, mas que tem se comportado muito bem permitindo que eu mantivesse meu ritmo de trabalho. Finalmente, pelo apoio técnico, em especial nos problemas relacionados a compilação de código C++.

Ao meu orientador, Prof. Nelson Fonseca, pela confiança, pelos direcionamentos, por compartilhar seu conhecimento e sua experiência como pesquisador. Com certeza, esse foi um período de grande aprendizado.

Aos amigos do LRC: André, Carlos Froidi, Cesar Melo, Cláudio, Daniel Batista, Daniel Manzato, Flávio, Gustavo Bittencourt, Juliana, Jorge, Luiz, Luciano e Neumar, pelas discussões técnicas, pelas conversas divertidas e, em especial, pelo suporte no período em que trabalhei a distância.

Aos amigos Sil e Pará, Carmen e Zeh, Daniela e Bazinho, Thaisa e Sandro, Daniele e Glauber, pelo companheirismo. Vocês se tornaram minha família em Campinas.

Aos docentes e funcionários do Instituto de Computação, pelo conhecimento transmitido e pela infraestrutura disponibilizada.

A CAPES e a UOL pelo apoio financeiro.

Sumário

Resumo	v
Abstract	vi
1 Introdução	1
1.1 Contribuições da Tese	4
1.2 Organização da Tese	5
1.3 Publicações Realizadas pela Autora	5
2 Padrão IEEE 802.16	7
2.1 Arquitetura	7
2.2 Escopo	8
2.3 Camada Física	9
2.3.1 Subframe Downlink	12
2.3.2 Subframe Uplink	13
2.4 Camada de Acesso ao Meio	14
2.4.1 Mecanismos para a provisão de Qualidade de Serviço	15
2.4.2 Mecanismos para Requisição e Alocação de Banda	17
2.5 Considerações Finais do Capítulo 2	18
3 Módulo para simulação de redes WiMAX no simulador ns-2	20
3.1 Modelagem e Implementação	21
3.1.1 Requisição e alocação de largura de banda	23
3.1.2 Serviço UGS	24
3.1.3 Serviço rtPS	26
3.1.4 Serviço ertPS	27
3.1.5 Serviço nrtPS e BE	28
3.2 Teste de Conformidade	30
3.2.1 Frames	30
3.2.2 Serviço UGS	32

3.2.3	Serviço rtPS	33
3.2.4	Serviço ertPS	34
3.2.5	Serviço nrtPS	36
3.2.6	Serviço BE	37
3.3	Considerações Finais do Capítulo 3	38
4	Escalonamento em redes IEEE 802.16	39
4.1	Estado da arte da pesquisa em escalonamento em redes IEEE 802.16	41
4.1.1	Propostas que não consideram a camada física	41
4.1.2	Propostas que consideram a camada física	44
4.2	Escalonador <i>Uplink</i> para Provisão de QoS em redes IEEE 802.16	47
4.3	Considerações Finais do Capítulo 4	51
5	Controle de Admissão em redes IEEE 802.16	53
5.1	Estado da arte da pesquisa em controle de admissão em redes IEEE 802.16	54
5.2	Controle de Admissão baseado na Taxa Mínima	58
5.3	Controle de Admissão Ótimo para Maximização do Lucro	60
5.4	Controle de Admissão Guloso para Maximização do Lucro	61
5.5	Controle de Admissão Ótimo para Maximização do Lucro e da Utilidade	62
5.6	Controle de Admissão Heurístico para Maximização do Lucro e da Utilidade	64
5.7	Considerações Finais do Capítulo 5	65
6	Avaliação das Políticas de Escalonamento e Controle de Admissão	69
6.1	Descrição do Cenário de Simulação	69
6.2	Avaliação dos Resultados	71
6.3	Considerações Finais do Capítulo 6	81
7	Conclusões e Trabalhos Futuros	82
7.1	Conclusões	82
7.2	Trabalhos Futuros	84
7.2.1	Extensão dos mecanismos propostos para considerar a camada física	84
7.2.2	Extensão dos mecanismos propostos para considerar mobilidade	85
7.2.3	Testes com diferentes valores de lucro e de utilidade	85
7.2.4	Mecanismos de QoS para as estações cliente	86
	Bibliografia	87

Lista de Tabelas

2.1	Parâmetros de QoS fornecidos por cada tipo de serviço	17
4.1	Vantagens e desvantagens das propostas discutidas na Seção 4.1.1	43
4.2	Vantagens e desvantagens das propostas discutidas na Seção 4.1.2	46
5.1	Vantagens e desvantagens das propostas discutidas na Seção 5.1	67
6.1	Características dos traces de vídeo usados nas simulações	70

Lista de Figuras

2.1	Tecnologias de Comunicação Sem Fio	8
2.2	Arquitetura de uma rede IEEE 802.16	9
2.3	Escopo do padrão IEEE 802.16	10
2.4	Estrutura do frame FDD [1]	11
2.5	Estrutura do frame TDD [1]	12
2.6	Formato do subframe downlink [1]	12
2.7	Formato do subframe uplink [1]	13
3.1	Estrutura do módulo para simulação do padrão IEEE 802.16	22
3.2	Máquina de estados para o serviço UGS [46].	25
3.3	Máquina de estados para o serviço rtPS. [46].	26
3.4	Máquina de estados para o serviço ertPS	27
3.5	Máquina de estados para o serviço BE [46]	28
3.6	Divisão do tempo em frames	31
3.7	Divisão dos frames em subframe <i>downlink</i> e <i>uplink</i>	31
3.8	Verificação do serviço UGS no sentido <i>uplink</i>	32
3.9	Verificação do serviço UGS no sentido <i>downlink</i>	33
3.10	Verificação do serviço rtPS no sentido <i>uplink</i>	34
3.11	Verificação do serviço rtPS no sentido <i>downlink</i>	34
3.12	Verificação do serviço ertPS no sentido <i>uplink</i>	35
3.13	Verificação do serviço ertPS no sentido <i>downlink</i>	35
3.14	Verificação do serviço nrtPS no sentido <i>uplink</i>	36
3.15	Verificação do serviço nrtPS no sentido <i>downlink</i>	37
3.16	Verificação do serviço BE	37
4.1	Comparação entre as camadas física SC, OFDM e OFDMA (as letras A, B, C, D e E representam diferentes usuários)	40
4.2	Arquitetura do escalonador proposto	48
4.3	Regulador <i>dual leaky bucket</i>	49
6.1	Probabilidade de Bloqueio para a política CA Simples	71

6.2	Probabilidade de Bloqueio para a política CA Ótimo Max. Lucro	72
6.3	Probabilidade de Bloqueio para a política CA Guloso Max. Lucro	72
6.4	Probabilidade de Bloqueio para a política CA Ótimo Max. Lucro e Utilidade .	73
6.5	Probabilidade de Bloqueio para a política CA Guloso Max. Lucro e Utilidade .	73
6.6	Lucro fornecido pelas políticas de controle de admissão	74
6.7	Utilidade fornecida pelas políticas CA Ótimo Max. Lucro e Utilidade e CA Guloso Max. Lucro e Utilidade	74
6.8	Latência das conexões UGS, ertPS e rtPS com a política CA Simples	76
6.9	Latência das conexões UGS, ertPS e rtPS com a política CA Ótimo Max. Lucro	76
6.10	Latência das conexões UGS, ertPS e rtPS com a política CA Guloso Max. Lucro	77
6.11	Latência das conexões UGS, ertPS e rtPS com a política CA Ótimo Max. Lucro e Utilidade	77
6.12	Latência das conexões UGS, ertPS e rtPS com a política CA Guloso Max. Lucro e Utilidade	78
6.13	Vazão das conexões rtPS que transmitiram o vídeo Ice Age	78
6.14	Vazão das conexões rtPS que transmitiram o vídeo Aladdin	79
6.15	Vazão das conexões nrtPS	80
6.16	Vazão das conexões BE	80

Capítulo 1

Introdução

O rápido crescimento na demanda por acesso residencial e comercial de alta velocidade à Internet tem impulsionado o desenvolvimento de tecnologias de acesso em banda larga. Prover acesso em banda larga utilizando infra-estrutura com fibra ótica ou cabo pode ser muito caro ou até mesmo inviável em lugares de difícil acesso. Uma solução de menor custo consiste no uso da tecnologia de acesso sem fio de banda larga, do inglês *Broadband Wireless Access* (BWA), dado que esta é capaz de prover altas taxas de transferência de dados para amplas áreas geográficas, alto nível de escalabilidade e baixo custo de instalação e manutenção.

O padrão IEEE 802.16 [1] vem sendo desenvolvido pelo IEEE (*Institute of Electrical and Electronics Engineers*) com a finalidade de padronizar a tecnologia BWA. O padrão define a interface aérea e o protocolo de acesso ao meio para redes metropolitanas sem fio (*Wireless Metropolitan Area Network* - WMAN), fornecendo altas taxas de transmissão para o acesso comercial e residencial à Internet. Redes baseadas nesse padrão são, frequentemente, chamadas de redes WiMAX (*Worldwide Interoperability for Microwave Access Forum*), em uma referência ao *WiMAX Forum*, um grupo composto por centenas de membros que inclui a maioria dos operadores e fabricantes de componentes e equipamentos para sistemas de comunicação [57]. O objetivo desse fórum é certificar e promover a compatibilidade e interoperabilidade dos produtos para redes BWA. Produtos implementados de acordo com o padrão IEEE 802.16 que passam pelos testes de compatibilidade e interoperabilidade conduzidos pelo fórum recebem o selo de certificação WiMAX.

Para dar suporte à grande diversidade de aplicações disponíveis na Internet, tais como serviços de voz, vídeo, multimídia e transferência de arquivos, o padrão IEEE 802.16 e sua extensão, IEEE 802.16e [2], definem mecanismos de sinalização entre a estação base e as estações cliente e cinco níveis de serviço. Em ambos os sentidos do tráfego, *uplink* (das estações cliente para a estação base) e *downlink* (da estação base para as estações cliente), os pacotes são associados a um fluxo de serviço específico pela camada de controle de acesso ao meio. Um conjunto de parâmetros de Qualidade de Serviço (QoS), dentre eles latência máxima e taxa mínima, é

associado a cada fluxo. Dessa forma, aplicações com necessidades distintas podem receber tratamento diferenciado para que alcancem o desempenho esperado.

Embora esses serviços forneçam a base para a provisão de QoS, uma solução completa exige mecanismos de controle de tráfego, tais como escalonamento e controle de admissão. Os mecanismos de controle de tráfego promovem a utilização eficiente dos recursos da rede bem como a provisão de Qualidade de Serviço. Mecanismos conservadores podem aumentar o nível de QoS oferecido aos usuários, mas, em contrapartida, podem resultar em baixa utilização da rede. Um controle de tráfego agressivo, por outro lado, pode aumentar a utilização da rede ao custo da deterioração da Qualidade de Serviço. Esse compromisso entre utilização e Qualidade de Serviço é de fundamental importância nas redes WiMAX, que agregam diferentes tipos de tráfego em um sistema com recursos limitados.

A arquitetura de QoS definida pelo padrão IEEE 802.16 inclui mecanismos de escalonamento e controle de admissão, entretanto, as políticas que devem ser usadas na implementação desses mecanismos não são especificadas. O fato das políticas de escalonamento e controle de admissão serem deixadas em aberto permite que os fabricantes diferenciem seus produtos.

O mecanismo de escalonamento tem como objetivo garantir a banda passante requisitada pelas estações cliente bem como promover o uso eficiente do enlace sem fio. Em uma rede WiMAX com topologia ponto-multiponto, o escalonamento do tráfego *downlink* requer um único escalonador na estação base, enquanto o escalonamento do tráfego *uplink* precisa de dois componentes, um na estação base e um na estação cliente. O escalonador *downlink* gerencia a transmissão dos pacotes destinados às estações cliente. O escalonador *uplink* na estação base aloca largura de banda para as estações cliente, enquanto o escalonador *uplink* nas estações cliente determina quais pacotes serão enviados nas oportunidades de transmissão recebidas.

O mecanismo de controle de admissão restringe o número de usuários simultaneamente presentes na rede, de forma a evitar a saturação do enlace sem fio e, conseqüentemente, a violação dos requisitos de QoS.

Esta tese introduz uma política para o escalonador *uplink* da estação base e um conjunto de políticas para o controle de admissão das conexões *uplink* em redes baseadas no padrão IEEE 802.16.

Dentre os três escalonadores incluídos na arquitetura de QoS do padrão IEEE 802.16, o escalonador *uplink* da estação base é o mais complexo. O escalonador *downlink* e o escalonador *uplink* das estações cliente têm acesso direto às filas que armazenam os pacotes a serem enviados, logo, têm à disposição informações atualizadas sobre as necessidades de cada fluxo de tráfego. O escalonador *uplink* da estação base, por outro lado, depende de mensagens de requisição de banda enviadas pelas estações cliente para conhecer o estado das filas dos fluxos *uplink* e, baseado nessa informação, fazer a alocação dos recursos. As mensagens de requisição de banda, no entanto, podem não refletir o estado atual das filas no instante em que forem atendidas, em especial quando os fluxos que as geraram servem aplicações com tráfego de taxa

variável.

Diversas propostas, disponíveis na literatura, introduzem mecanismos de escalonamento baseados em políticas originalmente propostas para redes cabeadas [17, 27, 20, 34, 37]. A grande maioria destes mecanismos utiliza uma hierarquia de políticas, tais como *Earliest Deadline First* (EDF), *Deficit Round Robin* (DRR), *Weighted Fair Queueing* (WFQ) e *Worst-case Weighted Fair Queueing* (W^2FQ), o que resulta em mecanismos de implementação complexa. Por outro lado, existem propostas de soluções mais simples, mas que não atendem a todas as especificações do padrão IEEE 802.16 [10, 45]. Existem, ainda, trabalhos [22, 39] que propõem um mecanismo de escalonamento apenas para o tráfego de tempo real.

A política de escalonamento proposta nesta tese é compatível com as especificações do padrão IEEE 802.16 e é capaz de atender os cinco tipos de serviço bem como o conjunto de parâmetros de QoS a eles associado. O escalonador utiliza três filas de prioridade para armazenar as mensagens de requisição de banda enviadas pelas estações cliente e um esquema de migração das mensagens entre essas filas para garantir a provisão dos requisitos de QoS de cada fluxo de serviço.

As políticas de controle de admissão propostas visam atingir três objetivos principais: 1) restringir o número de conexões simultaneamente presentes na rede para que o escalonador tenha recursos suficientes para atender os requisitos de QoS de cada fluxo de serviço, 2) atender as expectativas do provedor de serviço através da maximização do lucro e 3) maximizar a satisfação dos usuários através da disponibilização de recursos extras, além do combinado no contrato de admissão.

A primeira política de controle de admissão proposta é a mais simples, pois atende apenas o primeiro dentre os três objetivos mencionados. A estratégia de decisão sobre a admissão de novas conexões consiste em garantir que o somatório dos requisitos de taxa mínima de todas as conexões aceitas na rede seja menor do que a capacidade do enlace sem fio. Essa estratégia garante que, mesmo em situações de alta carga na rede, o escalonador terá a sua disposição recursos suficientes para satisfazer os requisitos mínimos de QoS de cada conexão aceita.

A segunda e a terceira políticas de controle de admissão atendem ao primeiro e ao segundo objetivos. A segunda política resolve o problema de admissão de novas conexões e maximização do lucro de forma ótima, utilizando programação linear inteira. A terceira política fornece uma solução de implementação mais simples, pois utiliza uma técnica gulosa para a resolver o mesmo problema.

A quarta e a quinta políticas de controle de admissão atendem os três objetivos. A quarta política decide sobre a admissão de novas conexões ao mesmo tempo em que tenta maximizar o lucro e a satisfação dos usuários de, maneira ótima, utilizando programação inteira mista, enquanto a quinta política fornece uma solução menos complexa através do uso de um algoritmo heurístico.

A grande maioria das soluções de controle de admissão disponíveis na literatura utilizam

o requisito de taxa mínima das conexões na tomada de decisões, entretanto, de acordo com as especificações do padrão IEEE 802.16, os recursos que devem ser reservados para garantir a QoS solicitada pela nova conexão ultrapassam o valor de taxa mínima requisitado. Isso se deve a sobrecargas características do tipo de serviço ao qual a conexão está associada, como por exemplo, a carga gerada pelo envio de mensagens de requisição de banda. As políticas de controle de admissão propostas, nesta tese, consideram a taxa mínima requisitada pela nova conexão bem como a sobrecarga gerada pelo tipo de serviço solicitado. Essa estratégia garante uma estimativa mais precisa da quantidade de recursos que devem ser reservados a cada nova admissão. Além disso, as políticas propostas fornecem estratégias de maximização do lucro e da satisfação dos usuários, que são elementos igualmente importantes para o sucesso das redes WiMAX.

A política de escalonamento e as políticas de controle de admissão propostas foram avaliadas através de experimentos de simulação utilizando modelos de tráfego que representam o tráfego esperado para cada um dos cinco tipos de serviço definidos no padrão IEEE 802.16. Para que este estudo fosse possível, foi necessário projetar e implementar um módulo para simulação de redes baseadas no padrão IEEE 802.16.

1.1 Contribuições da Tese

Esta tese contribui com os seguintes aspectos inovadores:

- Apresentação do estado da arte das pesquisas relacionadas ao desenvolvimento de políticas de controle de admissão e de escalonamento do tráfego *uplink* em redes baseadas no padrão IEEE 802.16;
- Proposta de uma política de escalonamento do tráfego *uplink* para a estação base de redes baseadas no padrão IEEE 802.16. Além de compatível com o padrão, a política proposta é capaz de prover garantias de Qualidade de Serviço para aplicações com necessidades distintas;
- Proposta de um conjunto de políticas de controle de admissão para redes baseadas no padrão IEEE 802.16. As políticas propostas permitem a provisão de garantias de Qualidade de Serviço além de oferecerem meios para a maximização do lucro obtido pelo provedor de serviços e da satisfação dos usuários;
- Projeto e implementação de um módulo para simulação de redes baseadas no padrão IEEE 802.16. O módulo desenvolvido está disponível para a comunidade científica em [58].

1.2 Organização da Tese

Esta tese está organizada da seguinte forma: o Capítulo 2 descreve as principais características e o funcionamento de uma rede baseada no padrão IEEE 802.16. Maior destaque é dado para a camada de acesso ao meio onde está localizada a arquitetura de Qualidade de Serviço especificada pelo padrão.

O Capítulo 3 apresenta a modelagem e as características de implementação do módulo desenvolvido para simulação de redes WiMAX bem como os testes conduzidos para verificar sua conformidade com as especificações do padrão IEEE 802.16. Esse módulo foi utilizado para derivar os resultados apresentados nesta tese e está disponível para os pesquisadores que desejarem utilizá-lo.

O Capítulo 4 discute o funcionamento e os desafios do escalonamento em redes baseadas no padrão IEEE 802.16, com ênfase no problema do escalonamento do tráfego *uplink* realizado pela estação base. São apresentadas algumas das soluções mais significativas propostas para esse problema bem como as vantagens e desvantagens dessas propostas. Em seguida, o capítulo introduz uma nova política para o escalonador *uplink* da estação base. Essa política é compatível com as especificações do padrão IEEE 802.16 e capaz de atender os requisitos de QoS de cada fluxo de serviço presente na rede.

O Capítulo 5 apresenta o problema do controle de admissão em redes baseadas no padrão IEEE 802.16 e discute as principais soluções propostas na literatura. Um conjunto de novas políticas de controle de admissão para provisão de Qualidade de Serviço, maximização do lucro e maximização da satisfação dos usuários é introduzido neste mesmo capítulo.

A primeira seção do Capítulo 6 descreve os experimentos de simulação realizados para avaliar a política de escalonamento e as políticas de controle de admissão propostas. Essa descrição apresenta os modelos de tráfego utilizados, os parâmetros de configuração da rede e os valores atribuídos para os elementos do conjunto de parâmetros de QoS de cada um dos cinco tipos de serviço definidos no padrão. Nas seções seguintes, os resultados obtidos são apresentados e discutidos.

No Capítulo 7, são apresentadas as considerações finais e os trabalhos futuros.

1.3 Publicações Realizadas pela Autora

Capítulo de Livro

- J. F. Borin e N. L. S. da Fonseca, Admission Control and Scheduling for QoS provisioning in WiMAX Networks. *QoS Architecture for Wireless Networks: Performance Metrics and Management*, editado por Sasan Adibi, Raj Jain, Mostafa Tofigh e Shyam Parekh, Editora IGI Global, 2009.

Periódicos

- J. F. Borin e N. L. S. da Fonseca, Scheduler for IEEE 802.16 networks. *IEEE Communications Letters*, vol. 12, no. 4, pg. 274-276, 2008.
- J. F. Borin e N. L. S. da Fonseca, Simulator for WiMAX networks. *Simulation Modeling Practice and Theory*, vol. 16, no. 7, pg. 817-833, 2008.

Eventos Internacionais

- J. F. Borin e N. L. S. da Fonseca, Uplink Scheduler and Admission Control for the IEEE 802.16 standard. *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, pg. 1-5, 2009.
- J. F. Borin e N. L. S. da Fonseca, A Standard-Compliant Scheduler for WiMAX Networks. *Proceedings of the 4th IEEE Broadband Wireless Access Workshop*, pg. 1-5, 2008.
- J. Freitag e N. L. S. da Fonseca, Uplink Scheduling with Quality of Service in IEEE 802.16 Networks. *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, pg. 2503-2508, 2007.
- J. Freitag e N. L. S. da Fonseca, WiMAX Module for the ns-2 Simulator. *Proceedings of the 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pg. 1-6, 2007.

Eventos Nacionais

- J. F. Borin e N. L. S. da Fonseca, Controle de admissão em redes IEEE 802.16. *Aceito para publicação nos Anais do 28º Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*, 2010.
- J. F. Borin e N. L. S. da Fonseca, Algoritmos para controle de admissão e escalonamento do tráfego uplink em redes IEEE 802.16. *Anais do VIII Workshop em Desempenho de Sistemas Computacionais e de Comunicação (WPERFORMANCE)*, pg. 1-14, 2009.
- J. F. Borin e N. L. S. da Fonseca, Um Módulo para Simulação de Redes WiMAX no Simulador NS-2. *Anais do VII Workshop em Desempenho de Sistemas Computacionais e de Comunicação (WPERFORMANCE)*, pg. 1-14, 2007.
- J. Freitag e N. L. S. da Fonseca, Escalonamento com Qualidade de Serviço em redes IEEE 802.16. *Anais do 25º Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*, pg. 1-14, 2007.

Capítulo 2

Padrão IEEE 802.16

O IEEE (*Institute of Electrical and Electronics Engineers*) tem desenvolvido uma hierarquia de padrões complementares para o acesso sem fio. Esta hierarquia, ilustrada na Figura 2.1, inclui os padrões IEEE 802.15 para redes pessoais (*Wireless Personal Area Network* - WPAN), IEEE 802.11 para redes locais (*Wireless Local Area Network* - WLAN), IEEE 802.16 para redes metropolitanas (*Wireless Metropolitan Area Network* - WMAN) e IEEE 802.20 para redes geograficamente distribuídas (*Wireless Wide Area Network* - WWAN).

O padrão IEEE 802.16, foco da pesquisa desta tese, vem sendo projetado para prover acesso sem fio em banda larga com desempenho comparável às formas de acesso tradicionais via cabo, DSL (*Digital Subscriber Line*) ou T1. Dentre as vantagens dos sistemas baseados no padrão 802.16 estão a rápida disponibilização dos serviços, mesmo em áreas de difícil acesso, e o baixo custo. Fornecer acesso em banda larga utilizando cabo ou DSL pode ser demorado e caro devido a infraestrutura necessária; por esse motivo, diversas áreas no mundo inteiro não têm acesso à esse tipo de conexão [25]. O padrão IEEE 802.16 permitirá acelerar a introdução da tecnologia de banda larga sem fio no mercado, bem como aumentar o desempenho e a confiabilidade dos serviços oferecidos pelos provedores de acesso. Outra vantagem deste padrão é a especificação de mecanismos para provisão de Qualidade de Serviço. Conforme será apresentado na Seção 2.4.1, o padrão fornece meios para diferenciação e priorização de tráfego bem como para a garantia de requisitos de Qualidade de Serviço, tais como latência máxima e taxa mínima.

O restante deste capítulo apresenta as principais características e o funcionamento do padrão IEEE 802.16.

2.1 Arquitetura

A arquitetura de uma rede que utiliza o padrão IEEE 802.16 [1], apresentada na Figura 2.2, é composta por dois elementos: estação base (*Base Station* - BS) e estação cliente (*Subscriber*

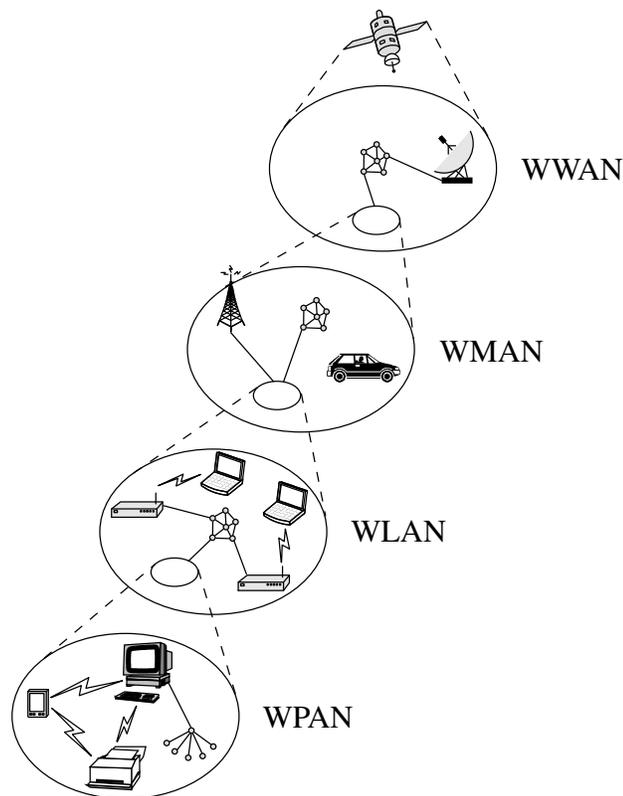


Figura 2.1: Tecnologias de Comunicação Sem Fio

Station - SS). A estação base realiza a comunicação entre a rede sem fio e a rede núcleo e suporta interfaces IP, ATM, Ethernet e E1/T1. A estação cliente fornece ao usuário acesso à rede núcleo através do estabelecimento de conexões com a estação base em uma topologia Ponto-Multiponto (PMP). O padrão ainda permite a implementação de uma topologia *Mesh* (opcional). A principal diferença entre as topologias PMP e *Mesh* está no fato de que em uma rede PMP o tráfego flui apenas entre a estação base e as estações cliente, enquanto no modo *Mesh* o tráfego pode ser roteado através das SSs e pode ocorrer diretamente entre duas estações cliente. O estudo apresentado nesta tese tem como foco redes com topologia PMP.

2.2 Escopo

A Figura 2.3 ilustra o escopo do padrão IEEE 802.16. A definição da camada física inclui múltiplas especificações, cada qual apropriada a uma faixa de frequência e aplicação. Estas especificações serão apresentadas na Seção 2.3.

A camada de acesso ao meio (*Medium Access Control - MAC*) compreende três subcamadas:

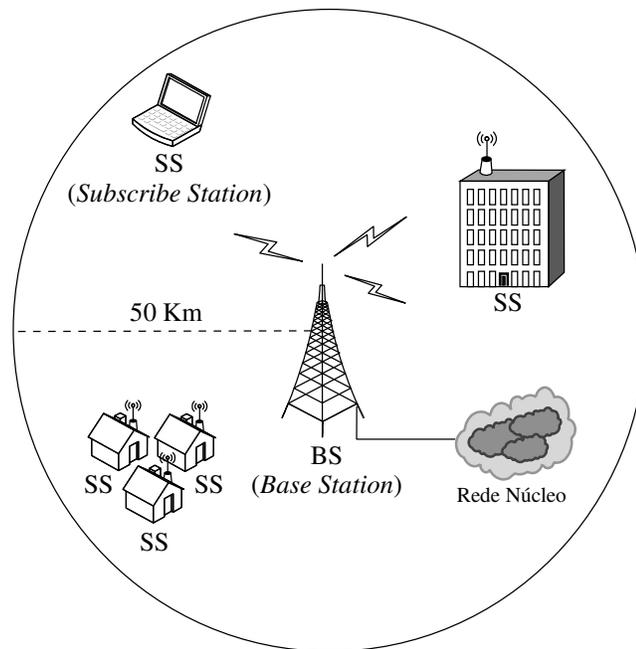


Figura 2.2: Arquitetura de uma rede IEEE 802.16

- *Service-Specific Convergence Sublayer (CS)*: faz a interface entre a camada de transporte e a camada MAC. O padrão especifica dois tipos de subcamada CS. A ATM CS é definida para serviços ATM e a *Packet CS* é definida para mapear serviços, tais como IPv4, IPv6, Ethernet e redes locais virtuais (VLAN). A principal função da subcamada CS é classificar unidades de dados de serviço (SDUs) e associá-las ao identificador do fluxo de serviço (*Service Flow Identifier - SFID*) e ao identificador da conexão (*Connection Identifier - CID*) apropriados.
- *MAC Common Part Sublayer (MAC CPS)*: provê as funcionalidades para o sistema de acesso ao meio, alocação de largura de banda, provisão de QoS, estabelecimento e manutenção das conexões.
- *Security Sublayer*: responsável pelas funções de autenticação para o acesso à rede e troca de chaves de segurança e criptografia para privacidade dos dados.

2.3 Camada Física

A primeira versão do padrão IEEE 802.16 [3], estabelecida em 2001, incluía apenas a faixa de frequências de 10 GHz a 66 GHz e operava em linha de visada (*Line Of Sight - LOS*) [14], o que significa que o receptor precisa estar visível para o emissor. Entretanto, a necessidade de

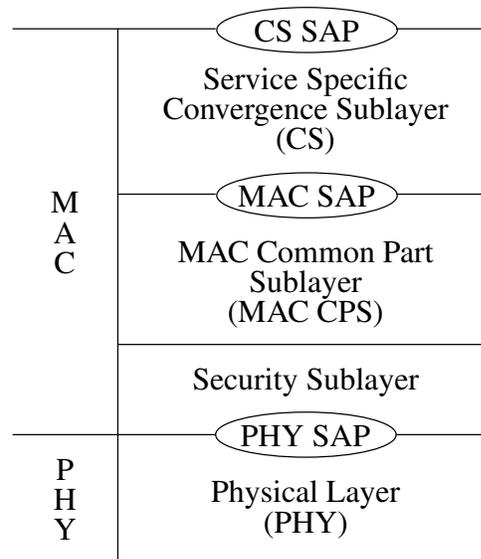


Figura 2.3: Escopo do padrão IEEE 802.16

operar em LOS limita a adoção da tecnologia, principalmente em áreas urbanas. Com o objetivo de superar esta limitação, o IEEE publicou em 2003 uma nova especificação para operar sem linha de visada. Esta versão do padrão, chamada 802.16a [5], utiliza faixas de frequências entre 2 e 11 GHz, incluindo bandas licenciadas e não licenciadas. Em 2004, foi publicada a versão 802.16-2004, que incorpora a versão 802.16a e especifica as regras para interoperabilidade nas frequências até 66 GHz [7]. Esta versão especifica cinco tipos de camada física que podem ser utilizadas em conjunto com a MAC [1]:

- **WirelessMAN-SC:** versão de portadora única projetada para operação com linha de visada na faixa de frequências de 10 a 66 GHz. Esta versão garante um alto grau de flexibilidade no planejamento das células, capacidade, serviços e custo.
- **WirelessMAN-SCa:** versão de portadora única para frequências até 11 GHz. Inclui um conjunto de funcionalidades para suportar operação sem linha de visada, tais como: estrutura de quadros robusta a multipercurso, estimativa e equalização de canal, modulação adaptativa, múltiplos esquemas de codificação, antenas adaptativas, diversidade de transmissão, controle de potência e *Automatic Repeat Request (ARQ)*.
- **WirelessMAN-OFDM:** versão que utiliza modulação multiportadoras (*Orthogonal Frequency Division Multiplexing - OFDM*) para frequências abaixo de 11 GHz e operação sem linha de visada. Além das funcionalidades definidas na versão WirelessMAN-SCa, suporta topologia *Mesh* e subcanalização no enlace *uplink* (das SSs para a BS).

- **WirelessMAN-OFDMA:** suporta operação sem linha de visada em frequências abaixo de 11 GHz. Utiliza o esquema de múltiplo acesso denominado *Orthogonal Frequency Division Multiple Access* (OFDMA), que é uma extensão da técnica de modulação OFDM para permitir o compartilhamento do canal por múltiplos usuários. Além das funcionalidades fornecidas pela versão WirelessMAN-SCa, permite subcanalização dos enlaces *uplink* e *downlink* (da BS para as SSs).
- **WirelessHUMAN:** inclui funcionalidades para operação em bandas não licenciadas nas faixas de frequência de 5 e 6 GHz. Utiliza um esquema flexível de canalização que inclui canais de 10 e 20 MHz de largura, com espaçamento de 5 MHz. Esta especificação se aplica às versões SCa, OFDM e OFDMA.

A camada física opera em um formato de *frames*. A duração de um *frame* pode variar de 0,5 ms a 20 ms, dependendo da tecnologia em uso na camada física. Os *frames* são subdivididos em intervalos de tempo chamados *slots* físicos (*Physical Slots - PSs*). O número de *slots* físicos (n) em um *frame* é uma função da taxa de símbolos (número de símbolos transmitidos por segundo) e da duração do *frame*, isto é, $n = (TaxaDeSimbolos \times DuracaoDoFrame)/4$.

Em cada *frame* há um *subframe downlink* e um *subframe uplink*. O *subframe downlink* é utilizado pela BS para a transmissão de dados e de informações de controle para as SSs. O *subframe uplink* é compartilhado entre todas as SSs. A BS é responsável por atribuir largura de banda para cada SS através da técnica DAMA-TDMA (*Demand Assigned Multiple Access - Time Division Multiple Access*).

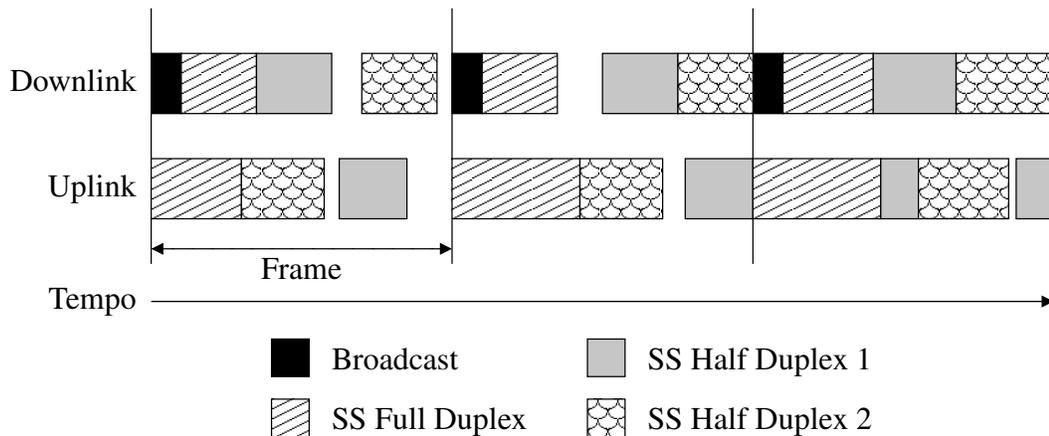


Figura 2.4: Estrutura do frame FDD [1]

O padrão IEEE 802.16 permite dois modos de acesso ao meio físico: duplexação por divisão de frequência (*Frequency Division Duplexing - FDD*) e duplexação por divisão de tempo (*Time Division Duplexing - TDD*). No modo FDD, os canais *downlink* e *uplink* operam, simultaneamente, em frequências diferentes. Portanto, a duração de um *subframe* (*uplink* ou *downlink*) é

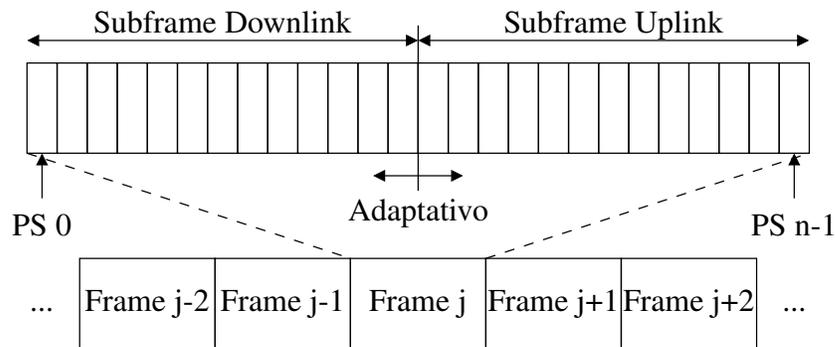


Figura 2.5: Estrutura do frame TDD [1]

igual a duração do *frame*. A estrutura de um *frame* FDD é apresentada na Figura 2.4. No modo TDD, os *subframes uplink* e *downlink* podem compartilhar a mesma frequência, porém, não podem realizar transmissões simultâneas. Cada *frame* TDD tem um *subframe downlink* seguido por um *subframe uplink*, como mostrado na Figura 2.5. O número de *slots* físicos alocados para cada *subframe* pode variar, dinamicamente, para acomodar a quantidade de tráfego em cada direção.

2.3.1 Subframe Downlink

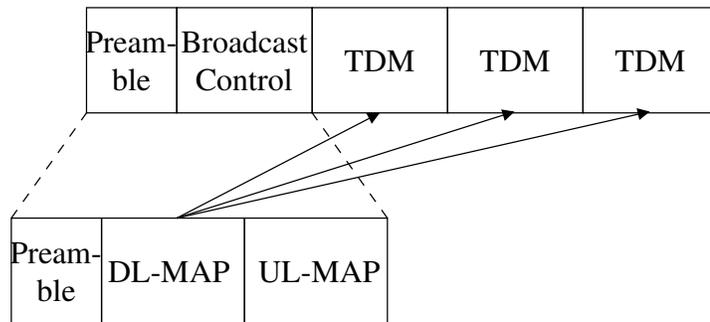


Figura 2.6: Formato do subframe downlink [1]

O *subframe downlink*, mostrado na Figura 2.6, inicia com um preâmbulo utilizado pela camada física para sincronização. Esse campo é seguido pela seção de controle do *frame* que contém um mapa *downlink* (*Downlink Map* - DL-MAP), para o *subframe downlink* em questão, e um mapa *uplink* (*Uplink Map* - UL-MAP) para o *subframe uplink* em um determinado instante do futuro. O DL-MAP informa a duração do *frame*, o número do *frame*, o identificador do canal *downlink* e o instante no qual as transições da camada física (modificações na modulação e *Forward Error Correction* - FEC) ocorrem dentro do *subframe downlink*. O UL-MAP especifica

o identificador do canal *uplink*, o instante inicial do *subframe uplink* e as oportunidades de transmissão (*grants*) alocadas para as SSs. A largura de banda é alocada para as SSs em *mini-slots*, sendo que cada *mini-slot* é igual a 2^m *slots* físicos ($0 \leq m \leq 7$). A seção de controle também pode conter um descritor do canal *downlink* (*Downlink Channel Descriptor* - DCD) e um descritor do canal *uplink* (*Uplink Channel Descriptor* - UCD). Os descritores DCD e UCD definem as características dos canais físicos *downlink* e *uplink*, respectivamente.

A seção de controle é seguida por uma porção TDM (*Time Division Multiplexing*) que carrega dados organizados em rajadas com diferentes perfis. Os dados são transmitidos para cada SS utilizando o perfil previamente negociado. Cada estação recebe e decodifica as informações de controle e busca cabeçalhos MAC que indicam dados endereçados a ela no restante do *subframe downlink*. Em sistemas FDD, a porção TDM pode ser seguida por um segmento TDMA (*Time Division Multiple Access*) para dar suporte a SSs que operam no modo *half-duplex*. O segmento TDMA contém um preâmbulo extra no início de cada rajada que permite às estações recuperar a sincronização. O *subframe downlink* TDD possui a mesma estrutura do *subframe* FDD, porém sem o segmento TDMA.

2.3.2 Subframe Uplink

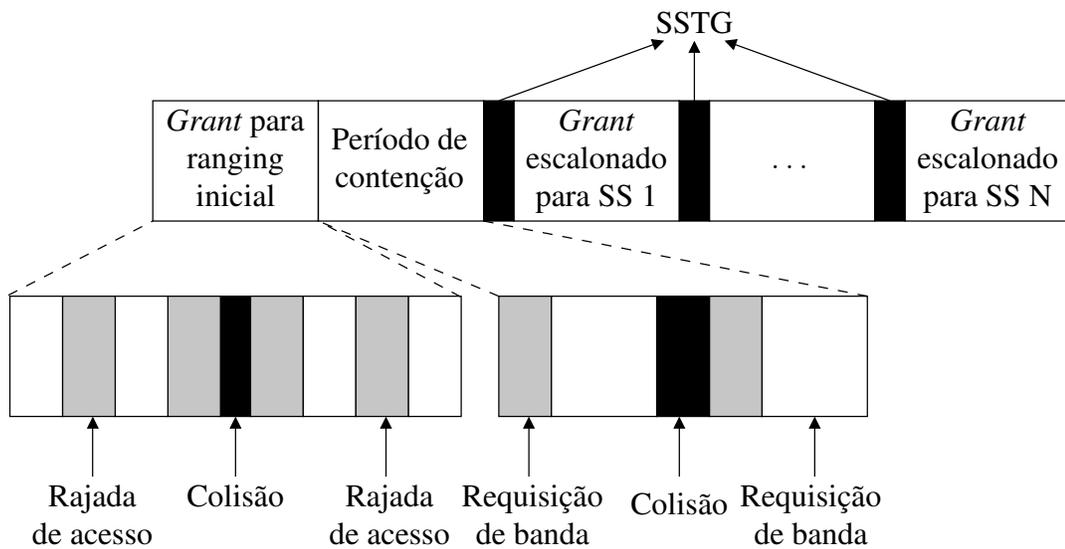


Figura 2.7: Formato do subframe uplink [1]

A Figura 2.7 mostra a estrutura do *subframe uplink* utilizado pelas SSs para realizar transmissões para a BS. São especificados três tipos de rajadas que podem ser transmitidas pelas SSs:

- rajadas que são transmitidas em *slots* de contenção reservados para o processo de *ranging* inicial;
- rajadas que são transmitidas em *slots* de contenção ou *unicast* reservados para mensagens de requisição de largura de banda;
- rajadas que são transmitidas em *slots unicast* alocados para SSs individuais para transmissão de dados no sentido *uplink*.

Qualquer um desses tipos de rajada pode estar presente em um *subframe uplink*. As rajadas podem ocorrer em qualquer ordem e em qualquer quantidade, limitada pelo número de *slots* alocados para transmissão *uplink* pela BS. As SSs transmitem nos *slots* alocados utilizando o perfil de rajada indicado na mensagem UL-MAP. Entre as rajadas, existe um intervalo chamado *SS Transition Gap* (SSTG) seguido por um preâmbulo para que a BS possa sincronizar-se com a nova SS.

2.4 Camada de Acesso ao Meio

Como mostrado na Seção 2.2, a camada de acesso ao meio (MAC) subdivide-se em 3 subcamadas. Dado que questões de segurança e interface com a camada de transporte estão fora do escopo deste trabalho, esta seção tem seu foco voltado para a subcamada MAC CPS.

A subcamada MAC CPS especifica o mecanismo para o acesso ao meio compartilhado. No sentido *downlink*, apenas a BS transmite para as SSs; assim sendo, não há necessidade de coordenação com outras estações. A BS possui uma antena capaz de transmitir para múltiplos setores simultaneamente. Todas as SSs que operam em uma determinada frequência e setor recebem a mesma transmissão. A BS pode transmitir mensagens através de *unicast* para uma única SS, *multicast* para um grupo de SSs ou *broadcast* para todas as SSs. As SSs compartilham o canal no sentido *uplink* utilizando o protocolo DAMA-TDMA (*Demand Assigned Multiple Access - Time Division Multiple Access*).

A camada MAC é orientada a conexão. Cada conexão tem um identificador (*Connection Identifier - CID*) de 16 bits e cada SS possui um endereço MAC único que a identifica e é utilizado para registrá-la e autenticá-la na rede. Durante o processo de entrada na rede, dois pares de conexões de gerenciamento devem ser estabelecidos entre a SS e a BS e um terceiro par de conexões de gerenciamento pode ser opcionalmente gerado. Cada par compreende uma conexão *uplink* e uma conexão *downlink*. Os três pares de conexões refletem o fato de que há três níveis de QoS para o tráfego de gerenciamento entre uma SS e a BS. A conexão básica é utilizada pela camada MAC da BS e das SSs para a troca de mensagens de gerenciamento curtas e urgentes. A conexão primária é utilizada pela camada MAC da BS e das SSs para a troca de mensagens mais longas e mais tolerantes ao retardo, enquanto que a conexão secundária é utilizada pela

BS e pelas SSs para a transferência de mensagens tolerantes ao retardo e padronizadas, tais como mensagens DHCP (*Dynamic Host Configuration Protocol*), TFTP (*Trivial File Transfer Protocol*) e SNMP (*Simple Network Management Protocol*) entre outras.

As SSs, também, estabelecem conexões de transporte unidirecionais para os fluxos de serviço contratado. Um fluxo de serviço define os parâmetros de QoS da conexão, os quais são utilizados por um mecanismo de controle de admissão para decidir se a conexão poderá ser estabelecida. Conexões são, também, reservadas para o acesso inicial baseado em contenção, transmissões *broadcast* no sentido *downlink* e *polling broadcast* e *multicast* das SSs.

A unidade de dados trocada entre as camadas MAC da BS e das SSs é a PDU (*Protocol Data Unit*). A subcamada CS recebe SDUs (*Service Data Units*) através da interface CS SAP (*CS Service Access Point*) e as associa ao fluxo de serviço e ao identificador da conexão apropriados. A subcamada MAC CPS recebe esses dados da subcamada CS e os encapsula em uma PDU que será enviada para o destinatário. A PDU consiste em um cabeçalho de tamanho fixo, um campo de tamanho variável para os dados (*payload*) e um campo CRC (*Cycle Redundancy Check*) opcional. O cabeçalho pode ser de dois tipos: genérico e de requisição de largura de banda. A PDU com cabeçalho genérico contém uma mensagem de gerenciamento da MAC ou dados, enquanto a PDU com cabeçalho de requisição de largura de banda é utilizada pela SS para requisitar banda para uma conexão e não carrega dados.

2.4.1 Mecanismos para a provisão de Qualidade de Serviço

A camada MAC provê também mecanismos para fornecer QoS aos tráfegos *uplink* e *downlink*. O principal mecanismo para a provisão de QoS consiste em associar os pacotes que passam pela camada MAC a um fluxo de serviço. O fluxo de serviço é um serviço da camada MAC que fornece transporte unidirecional aos pacotes. Durante a fase de estabelecimento da conexão, estes fluxos de serviço são criados e ativados pela BS e pela SS. Cada fluxo de serviço ativo possui um identificador da conexão único. Várias seções das camadas superiores podem operar sobre o mesmo identificador da conexão na camada MAC caso seus requisitos de QoS sejam os mesmos. Cada fluxo de serviço deve definir seu conjunto de parâmetros de QoS:

- *minimum reserved traffic rate*: especifica a taxa mínima reservada para o fluxo de serviço. Se a quantidade de banda requisitada for menor do que a taxa mínima reservada, a BS deve alocar a banda excedente para outra finalidade.
- *maximum sustained traffic rate*: define a taxa de pico do serviço. A taxa é informada em bits por segundo e não inclui os cabeçalhos da MAC e CRCs. A SS deve controlar o serviço no sentido *uplink* para que a taxa seja, na média, igual ao valor definido por esse parâmetro. Na BS, no sentido *downlink*, assume-se que o tráfego é controlado na entrada da rede não sendo necessário fazer um controle adicional. Esse parâmetro define apenas

um limite e não uma garantia de que a taxa estará disponível. O algoritmo para o controle desse parâmetro não é definido pelo padrão.

- *maximum traffic burst*: define o tamanho máximo da rajada que deve ser disponibilizado para o serviço.
- *maximum latency*: especifica a latência máxima entre a recepção do pacote na interface de rede da BS ou da SS e a transmissão do pacote para a interface de rádio frequência.
- *tolerated jitter*: define a variação máxima permitida para a latência da conexão.
- *service flow scheduling type*: o valor deste parâmetro determina o tipo de serviço ao qual o fluxo deve ser associado. Se o parâmetro é omitido, usa-se o serviço de melhor esforço.
- *request/transmission policy*: este parâmetro permite configurar alguns atributos para o fluxo, tais como opções sobre a formatação da PDU e, para fluxos *uplink*, restrições nos métodos de requisição de banda que podem ser utilizados.
- *unsolicited grant interval*: intervalo entre *grants* sucessivos alocados para um fluxo de serviço UGS ou ertPS.
- *unsolicited polling interval*: intervalo máximo entre *grants* sucessivos alocados para um fluxo rtPS enviar requisição de banda.
- *traffic priority*: o valor desse parâmetro especifica a prioridade atribuída ao tráfego. Dados dois fluxos de serviço idênticos em todos os parâmetros de QoS exceto na prioridade, o fluxo de serviço com maior prioridade deve ter latência menor e prioridade maior no processo de armazenamento no *buffer*. Por outro lado, para fluxos de serviço não idênticos, a prioridade não deve preceder outros parâmetros de QoS conflitantes.

O padrão e sua extensão IEEE 802.16e (2005) [2] especificam cinco tipos de fluxo de serviço. Cada fluxo é mapeado para um desses serviços e um escalonador na BS aloca largura de banda para as SSs, seguindo o conjunto de regras definido por eles.

O serviço UGS (*Unsolicited Grant Service*) suporta fluxos de tempo real que geram pacotes de dados com tamanho fixo periodicamente, tal como voz sobre IP. Conexões UGS recebem *grants* periódicos de tamanho fixo sem a necessidade de requisitar largura de banda.

O segundo tipo de serviço é o rtPS (*Real-Time Polling Service*), projetado para aplicações com requisito de tempo real que geram pacotes com tamanho variável periodicamente, como por exemplo aplicações de vídeo MPEG. Fluxos rtPS requisitam banda periodicamente através de *polling unicast* e a QoS é garantida satisfazendo-se os requisitos de latência máxima e de taxa mínima.

O serviço ertPS (*Extended Real-Time Polling Service*) é projetado para tráfego de tempo real com taxa variável, como por exemplo aplicações de voz sobre IP com supressão de silêncio. Este serviço usa um mecanismo de *grant* similar àquele utilizado pelas conexões UGS. No entanto, os *grants* alocados periodicamente podem ser usados para enviar requisições de banda para informar a BS sobre a necessidade de um novo tamanho de *grant*. A BS não muda o tamanho dos *grants* até que receba uma requisição de largura de banda da SS.

O serviço nrtPS (*Non-real-time Polling Service*) suporta tráfego insensível ao retardo que requer *grants* de tamanho variável regularmente, tal como tráfego FTP. A QoS é garantida satisfazendo-se o requisito de taxa mínima. O serviço oferece *polling unicast*, porém, com menor frequência do que o serviço rtPS, e permite que a SS utilize os *slots* de contenção reservados para requisição de banda.

O serviço BE (*Best Effort*) suporta tráfego de melhor esforço sem quaisquer garantias de QoS. A SS pode utilizar tanto *slots unicast* quanto *slots* de contenção para requisitar largura de banda.

A Tabela 2.1 mostra quais parâmetros de QoS cada tipo de serviço deve informar durante o estabelecimento da conexão.

Tabela 2.1: Parâmetros de QoS fornecidos por cada tipo de serviço

	UGS	ertPS	rtPS	nrtPS	BE
<i>minimum reserved traffic rate</i>		✓	✓	✓	
<i>maximum sustained traffic rate</i>	✓	✓	✓	✓	
<i>maximum traffic burst</i>	✓	✓	✓	✓	✓
<i>maximum latency</i>	✓	✓	✓		
<i>tolerated jitter</i>	✓	✓			
<i>service flow scheduling type</i>	✓	✓	✓	✓	✓
<i>request/transmission policy</i>	✓	✓	✓	✓	✓
<i>unsolicited grant interval</i>	✓	✓			
<i>unsolicited polling interval</i>			✓		
<i>traffic priority</i>	✓	✓	✓	✓	

2.4.2 Mecanismos para Requisição e Alocação de Banda

A alocação de recursos para as SSs é realizada sob demanda. Quando uma SS precisa requisitar banda para uma conexão, ela envia uma mensagem para a BS contendo o pedido. Uma requisição de banda pode ser enviada como um pacote, cujo cabeçalho indica a presença de um pedido de banda, ou pode ser enviada juntamente com um pacote de dados (*Piggyback*). Todas as requisições devem indicar o número de bytes necessários para transmitir os dados (*payload*) e o cabeçalho MAC.

A requisição de largura de banda pode ser incremental ou agregada. Quando a BS recebe uma requisição incremental, ela deve adicionar a quantidade de banda requisitada à sua percepção atual sobre a banda que deve ser alocada para a conexão. Quando a requisição é agregada, a BS deve substituir a percepção de banda que deve ser alocada para a conexão pela quantidade de banda requisitada. O campo *type* no cabeçalho de uma mensagem de requisição indica se a requisição é incremental ou agregada. Dado que requisições *Piggyback* não possuem este campo, o pedido deve ser sempre incremental.

Para a SS, as requisições de banda sempre são referentes à uma determinada conexão, enquanto os *grants* são alocados para a SS e não para uma conexão em particular. Dessa forma, a SS pode utilizar o *grant* recebido para uma conexão diferente daquela para a qual a requisição foi feita.

A BS destina parte da banda para as SSs enviarem suas requisições. A alocação de *grants* para o envio de requisições pode ser para uma SS particular ou para um grupo de SSs. Esta alocação é informada para as SSs através da mensagem UL-MAP e recebe o nome de *polling*. O padrão define dois mecanismos de *polling*:

- **Unicast:** a SS recebe um *grant* cujo tamanho é suficiente para o envio de uma requisição de banda.
- **Baseado em contenção:** utilizado quando não há banda disponível para fazer o *polling unicast* de todas as SSs. Neste caso, a BS aloca um *grant* para um grupo de SSs, as quais devem competir pela oportunidade de enviar a mensagem de requisição. Para reduzir a probabilidade de colisão, apenas as SSs que necessitam de banda participam da contenção. A resolução da contenção se dá com um algoritmo de *backoff* exponencial. O tamanho da janela mínima e da janela máxima de contenção é controlado pela BS.

A alocação dos recursos nas redes IEEE 802.16 é realizada por três escalonadores. A estação base possui um escalonador para o tráfego *downlink* (da estação base para as estações cliente) e um escalonador para o tráfego *uplink* (das estações cliente para a estação base). O escalonador *downlink* decide quais conexões terão direito de transmitir em cada *subframe downlink*. O escalonador *uplink* atribui *grants* para as estações cliente enviarem requisições de banda e dados nos *subframes uplink*. Cada estação cliente possui um escalonador local que decide quais pacotes serão enviados nas oportunidades de transmissão recebidas nos *subframes uplink*.

2.5 Considerações Finais do Capítulo 2

Este capítulo apresentou as principais características e o funcionamento do padrão IEEE 802.16. Mostrou-se que redes baseadas neste padrão podem possuir topologia Ponto-Multiponto ou

topologia *Mesh* e que o escopo do padrão envolve a camada física e a camada de acesso ao meio (MAC).

Os cinco tipos de camada física especificados pelo padrão foram descritos bem como os dois modos de acesso possíveis: FDD e TDD. Mostrou-se ainda que a camada física opera em *frames* e que cada *frame* é subdividido em um *subframe downlink* e um *subframe uplink*.

Maior destaque, entretanto, foi dado à descrição da camada de acesso ao meio, onde concentra-se a pesquisa desta tese. As principais funcionalidades da camada MAC envolvem o estabelecimento e a manutenção das conexões, a alocação de banda e a provisão de Qualidade de Serviço para as conexões aceitas na rede. Os mecanismos responsáveis por essas funcionalidades foram descritos sendo que dois deles serão detalhados nos capítulos seguintes: o mecanismo de escalonamento e o mecanismo de controle de admissão. Embora esses mecanismos sejam de fundamental importância para o correto funcionamento das redes IEEE 802.16 e exerçam importante papel no desempenho da rede, as políticas que devem ser usadas em sua implementação não são definidas pelo padrão.

Capítulo 3

Módulo para simulação de redes WiMAX no simulador ns-2

A simulação é uma ferramenta essencial na pesquisa de redes de computadores, pois permite o desenvolvimento e a análise de novos protocolos e mecanismos em topologias complexas, sem a necessidade de implementá-las fisicamente. Dentre as ferramentas disponíveis para simulação de redes de computadores, o *Network Simulator* (ns-2) é a ferramenta de maior popularidade entre os pesquisadores [51]. Esta popularidade deve-se ao fato do ns-2 ser de domínio público e permitir a simulação de grande parte dos protocolos utilizados na Internet em redes cabeadas e sem fio.

Dois módulos foram propostos na literatura para simulação de redes IEEE 802.16 usando o ns-2. Um deles, implementado pelo *National Institute of Standards and Technology* (NIST) [48], provê, entre outras funcionalidades, camada física OFDM com modulação configurável, duplexação por divisão de tempo (*Time Division Duplexing* - TDD), topologia ponto-multiponto (*Point-to-Multipoint* - PMP) e fragmentação de pacotes. No entanto, este módulo não implementa os mecanismos para suporte de QoS na camada MAC, ou seja, os diferentes tipos de serviço e mecanismos de escalonamento com QoS. O segundo módulo, proposto em [26], também implementa o modo de duplexação TDD e a topologia PMP, mas utiliza a implementação do canal sem fio já fornecida pelo ns-2. Embora este módulo implemente os cinco tipos de serviço definidos pelo padrão IEEE 802.16, o mecanismo para requisição de largura de banda não contempla a especificação do padrão. Adicionalmente, o módulo não permite que o usuário configure os requisitos de QoS, tais como latência máxima e banda mínima, dos serviços de alta prioridade. A ferramenta de simulação OPNET [38], de domínio privado, inclui um módulo para simulação de redes WiMAX. No entanto, o acesso a esse módulo ainda é restrito aos membros do grupo responsável por seu desenvolvimento.

Este capítulo apresenta a especificação e a implementação de um módulo para simulação de redes IEEE 802.16 no simulador ns-2. O desenvolvimento do módulo tem seu foco na camada

de acesso ao meio e utiliza a camada física sem fio já disponível no ns-2. A implementação contempla os mecanismos de reserva e alocação de banda e de provisão de QoS especificados no padrão permitindo diversos estudos sobre a camada de acesso ao meio, como por exemplo, a investigação de mecanismos para provisão de QoS. O módulo implementado foi uma ferramenta essencial para os estudos apresentados nesta tese e tem sido de extrema valia para pesquisadores da área visto que quase 2.000 pessoas de diversas partes do mundo efetuaram a cópia do código disponibilizado em [58]. Embora o código seja extenso, totalizando 17 classes e mais de 17.000 linhas de código, a modularização obtida com a implementação orientada a objetos facilita a inclusão de novas funcionalidades.

O restante do capítulo está organizado da seguinte forma. A Seção 3.1 apresenta a modelagem e a implementação do módulo para simulação de redes WiMAX. A Seção 3.2 apresenta os experimentos de simulação realizados para verificar a conformidade do módulo implementado com as especificações do padrão IEEE 802.16. Finalmente, a Seção 3.3 traz as considerações finais.

3.1 Modelagem e Implementação

O módulo para simulação de redes WiMAX foi desenvolvido na versão 2.28 do simulador ns-2 e segue as especificações do padrão IEEE 802.16 para a topologia PMP e o modo de acesso TDD [1]. A implementação foi realizada na linguagem C++ utilizando o paradigma da programação orientada a objetos.

O módulo WiMAX baseou-se em um módulo [46] desenvolvido para o ns-2 para simulação do padrão DOCSIS [13], utilizado para transmissão de vídeo em redes cabeadas. A reutilização desse código foi possível pois a especificação da camada MAC do padrão IEEE 802.16 é baseada no padrão DOCSIS [14]. Assim como no padrão IEEE 802.16, no padrão DOCSIS existe uma entidade central que controla o acesso ao meio fazendo a alocação de banda para as estações cliente. A alocação da banda baseia-se nas requisições enviadas pelas estações cliente e nos requisitos de QoS das conexões.

Embora a camada MAC de ambos os padrões seja semelhante, foram necessárias várias modificações no código do módulo DOCSIS para adequá-lo ao padrão IEEE 802.16. Dentre essas modificações, destacam-se: a implementação do serviço nrtPS para o tráfego *uplink*, a implementação dos cinco tipos de serviço para o tráfego *downlink*, a alteração da interface entre a camada MAC e a camada física para que o módulo WiMAX utilizasse o meio sem fio disponível no ns-2, a implementação da divisão do tempo em *frames* e dos *frames* em *subframe downlink* e *subframe uplink* e a inclusão dos parâmetros de QoS latência máxima, para o serviço rtPS, taxa mínima e taxa máxima, para os serviços rtPS e nrtPS.

A Figura 3.1 apresenta a estrutura do módulo implementado. Tanto na BS quanto nas SSS o tráfego proveniente da camada superior é classificado para o fluxo de serviço apropriado (as-

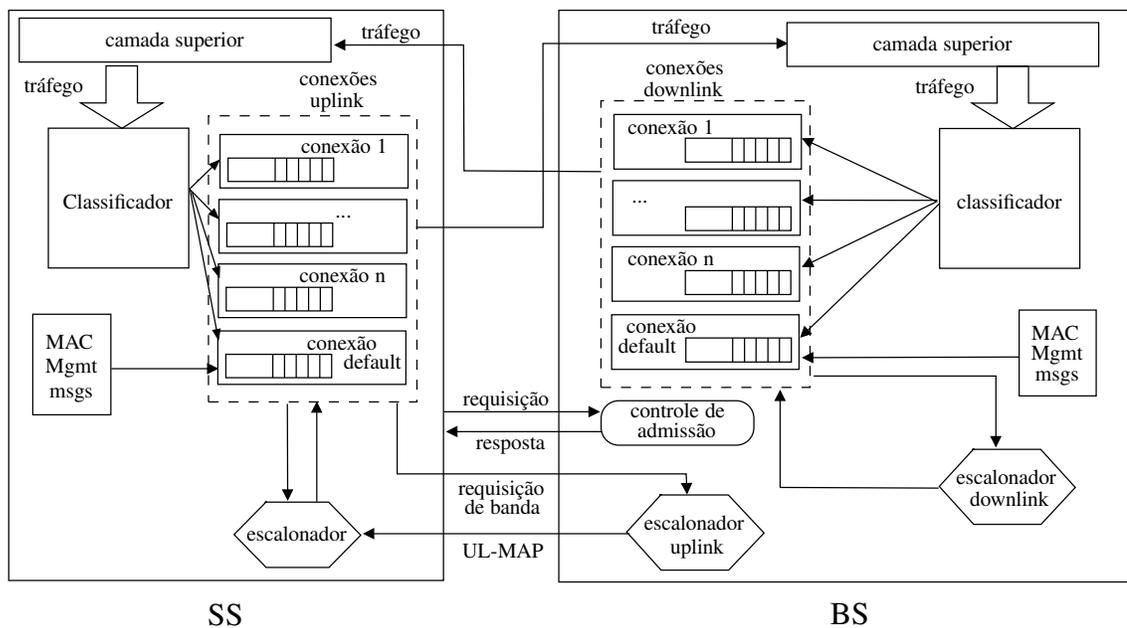


Figura 3.1: Estrutura do módulo para simulação do padrão IEEE 802.16

sociado a uma conexão), a partir do tipo do tráfego e do par origem e destino. Os fluxos de serviço são configurados pelo usuário, que entre outras informações, determina os requisitos de QoS. Cada nó da rede deve ter uma conexão *default* para transmitir mensagens de gerenciamento e o tráfego que não puder ser mapeado para nenhuma das outras conexões. No início da simulação, todas as SSs registram-se com a BS para o estabelecimento das conexões com seus respectivos parâmetros de QoS. A BS atribui um identificador único para cada conexão e armazena as informações em uma tabela, dentre elas: o tipo do serviço e seus parâmetros de QoS, se a conexão utiliza as funcionalidades de fragmentação, concatenação e *piggybacking*, e o identificador único.

A BS possui um controlador de admissão, um escalonador *downlink* e um escalonador *uplink*. O controlador de admissão decide quais conexões *uplink* poderão ser estabelecidas. O escalonador *downlink* determina quais pacotes enviados pela camada superior serão transmitidos em cada *subframe downlink*. Para tomar essa decisão, o escalonador baseia-se nos requisitos de QoS e na situação da fila de cada conexão. O escalonador *uplink* decide quais SSs terão direito de transmitir em cada *subframe uplink* e quantos *slots* cada uma dessas SSs poderá utilizar. A decisão desse escalonador leva em consideração os requisitos de QoS das conexões *uplink*, bem como as requisições de banda recebidas.

Nas SSs existe um escalonador que decide quais pacotes provenientes da camada superior serão enviados nos *slots* recebidos. Para tomar essa decisão, o escalonador baseia-se nas informações trazidas pelo UL-MAP, nos requisitos de QoS e na situação das filas das conexões.

No padrão IEEE 802.16, no início de cada *subframe downlink*, a BS envia um mapa *downlink* (DL-MAP) e um mapa *uplink* (UL-MAP), para informar quais SSs receberão pacotes no *subframe downlink*, quais poderão transmitir no *subframe uplink* e o número de *slots* reservados para cada transmissão. No módulo implementado, o DL-MAP não é enviado, dado que as informações que ele conteria não são necessárias em um sistema de simulação¹. No início do *subframe downlink*, porém, são reservados alguns *slots* para simular a sobrecarga do envio dessa mensagem. O UL-MAP é enviado e contém a descrição dos *slots* alocados no *subframe uplink* subsequente.

No padrão IEEE 802.16, os *frames* têm uma duração pré-definida, enquanto os *subframes uplink* e *downlink* podem ter tamanho variáveis dependendo da necessidade de cada um. O módulo de simulação permite que o usuário defina a duração dos *frames*. Dado esse valor, o módulo atribui durações iguais para cada um dos *subframes*. O ajuste dinâmico da duração dos *subframes* será implementado em um trabalho futuro.

Em um sistema real, a BS e as SSs trocam em torno de 50 tipos de mensagens de gerenciamento, como por exemplo mensagens de *-ranging*, mensagens com informações sobre o canal, mensagens para garantir a segurança na rede, entre outras. A maioria dessas mensagens, em especial aquelas que envolvem informações sobre o canal, autenticação e segurança na rede, não é implementada por estarem fora do foco deste trabalho. Entretanto, mensagens de gerenciamento são geradas periodicamente para representar a sobrecarga necessária.

3.1.1 Requisição e alocação de largura de banda

Quando um pacote chega na camada MAC, a partir das camadas superiores, ele é classificado para um fluxo de serviço (associado a uma conexão) baseado no endereço IP de origem e destino e no tipo do pacote (HTTP, FTP, CBR, etc.). Na sequência, o cabeçalho MAC de 6 bytes é incluído no pacote². Cada fluxo de serviço possui uma fila onde os pacotes são inseridos caso não haja um *grant* disponível. Os pacotes que chegam pelo canal são classificados como uma mensagem de gerenciamento ou de dados através das informações disponíveis no cabeçalho e manipulados adequadamente.

Cada fluxo de serviço possui quatro componentes principais:

- **Classificador:** o classificador usa os endereços de origem e destino e o tipo do pacote para classificar um pacote para um fluxo de serviço.

¹Uma das funções da mensagem DL-MAP é informar para a SS em que período ela deve “escutar” o canal para detectar os pacotes enviados pela BS. No ns-2 as estações sem fio escutam o canal o tempo inteiro, portanto, o não envio desta mensagem não prejudica o recebimento dos pacotes do tráfego *downlink*.

²O tamanho do cabeçalho está de acordo com o tamanho definido pelo padrão IEEE 802.16 para o cabeçalho MAC genérico.

- Fila: todos os pacotes classificados para um fluxo de serviço são inseridos na fila caso não possam ser imediatamente enviados.
- Tabela de Alocação: mantém as informações sobre os *grants* alocados para o fluxo de serviço. Esta tabela é atualizada sempre que um UL-MAP é recebido.
- Máquina de estados finita (FSM - *Finite State Machine*): determina as ações de um fluxo de serviço de acordo com os eventos que ocorrem na rede. Sempre que um evento ocorre, dependendo do estado em que a máquina se encontra, a função apropriada é chamada para tratar o evento de entrada e a variável de estado é atualizada.

A definição das máquinas de estado para os serviços UGS, rtPS e BE no sentido *uplink* foram baseadas nas máquinas de estado propostas em [46]. A máquina de estados para o serviço nrtPS utiliza o mesmo modelo da máquina de estados do serviço BE, dado que ambos os serviços utilizam *polling unicast* e de contenção. A diferença está no fato de que a BS aloca *grants unicast*, freqüentemente, para o serviço nrtPS, enquanto o serviço BE recebe *grants unicast* apenas quando há sobra de recursos. A máquina de estados para o serviço ertPS foi definida de forma a permitir o envio tanto de dados como de requisições de banda nos *grants* alocados periodicamente pela BS.

Assim como o serviço UGS no sentido *uplink*, os fluxos de serviço no sentido *downlink* não precisam enviar requisições, sua tarefa consiste em enviar os dados nos *grants* alocados pela BS. Desta forma, todos os serviços no sentido *downlink* usam a mesma máquina de estados modelada para o serviço UGS no sentido *uplink*.

Note que a QoS fornecida para os diferentes tipos de serviço não depende do modelo da máquina de estados, mas dos mecanismos de controle de admissão e de escalonamento implementados na BS e nas SSs. As máquinas de estado determinam o envio de um pacote de dados ou de requisição de banda de acordo com as informações armazenadas na Tabela de Alocação.

Os mecanismos de escalonamento implementados no escalonador das SSs e no escalonador *downlink* da BS seguem a política *Strict Priority*, ou seja, a fila de um serviço de maior prioridade é sempre servida antes da fila de um serviço de menor prioridade. Políticas de escalonamento mais eficientes para esses dois escalonadores serão parte de um estudo futuro.

O escalonador *uplink* da BS utiliza a política apresentada no Capítulo 4.

As subseções seguintes apresentam as máquinas de estados para os serviços UGS, rtPS, ertPS, nrtPS e BE no sentido *uplink*.

3.1.2 Serviço UGS

Para a máquina de estados do serviço UGS foram definidos os seguintes estados:

- *Idle*: não há nenhum pacote para transmitir;

- *Decision*: estado temporário no qual se verifica a existência de *grants* na Tabela de Alocação;
- *Wait-For-Map*: um mapa com novas alocações é aguardado;
- *To-Send*: há uma transmissão de pacote pendente.

Foram definidos, adicionalmente, os seguintes eventos:

- *Packet*: um pacote chegou na fila para ser enviado;
- *Map*: um mapa foi recebido;
- *SendTimer*: o *timer* (referenciado como *send_timer* na descrição da máquina de estados) para envio de um pacote expirou;
- *SendPacket*: um pacote deve ser enviado.

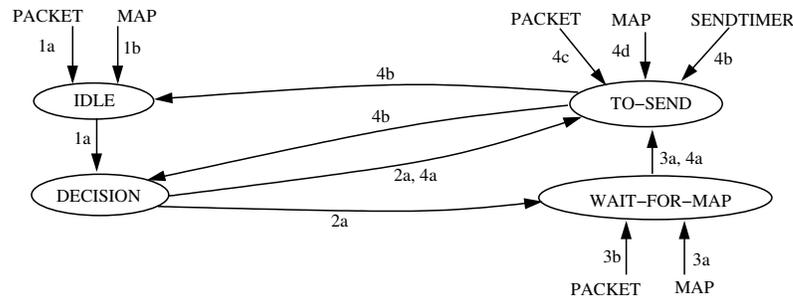


Figura 3.2: Máquina de estados para o serviço UGS [46].

A Figura 3.2 apresenta a máquina de estados finita para o serviço UGS. Quando o estado é *Idle* e um pacote chega (evento 1a), o cabeçalho MAC é adicionado ao pacote, o pacote é armazenado em uma variável chamada *current_pkt* e a FSM entra no estado *Decision*. Quando um mapa chega (evento 1b), a Tabela de Alocação do fluxo é atualizada e o estado continua o mesmo.

No estado *Decision*, se não há *grant* para o fluxo de serviço na Tabela de Alocação, a FSM entra no estado *Wait-For-Map*; caso contrário, um evento *SendPacket* ocorre e há uma transição para o estado *To-Send*.

Quando um mapa chega com um *grant* para o fluxo no estado *Wait-For-Map*, a FSM entra no estado *To-Send* e um evento *SendPacket* ocorre. Caso o mapa não contenha um *grant* para o fluxo, não há mudança de estado. Quando um pacote chega, o cabeçalho MAC é adicionado e o pacote é inserido na fila.

Quando o evento *SendPacket* ocorre no estado *To-Send*, o *send_timer* é configurado para expirar no instante de início do *grant*. Quando o *timer* expira, o evento *SendTimer* é disparado e o pacote em *current_packet* é transmitido. Se a fila está vazia, a FSM entra no estado *Idle*, caso contrário, um pacote é retirado da fila, armazenado na variável *current_packet* e a FSM entra no estado *Decision* com o evento *Packet*.

Quando um pacote chega e a FSM está no estado *To-Send*, o cabeçalho MAC é adicionado e o pacote é inserido na fila. Quando um mapa chega, a Tabela de Alocação do fluxo é atualizada.

3.1.3 Serviço rtPS

Todos os estados e eventos definidos para a máquina de estados do serviço UGS têm o mesmo significado para a máquina de estados do serviço rtPS. Além disso, foram definidos o estado *To-Send-Request*, que indica que há um *grant unicast* para envio de requisição de banda na Tabela de Alocação, e os seguintes eventos:

- *ReqTimer*: o *timer* (referenciado como *req_timer* na descrição da máquina de estados) para o envio da requisição expirou.
- *SendReq*: indica que uma requisição deve ser enviada.

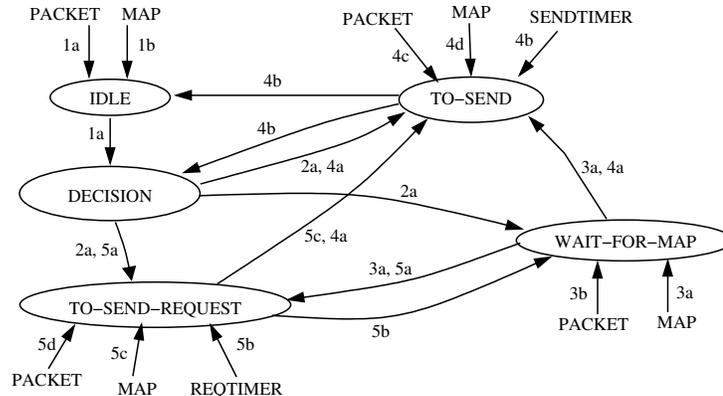


Figura 3.3: Máquina de estados para o serviço rtPS. [46]

A Figura 3.3 apresenta a máquina de estados finita para o serviço rtPS. Quando a FSM está no estado *Decision* e há um *grant unicast* para o fluxo na Tabela de Alocação, a FSM muda para o estado *To-Send-Request* e um evento *SendReq* ocorre.

Quando a FSM entra no estado *To-Send-Request* e um evento *SendReq* ocorre, o *req_timer* é configurado para expirar no instante de início do *grant unicast*. Quando o *timer* expira, um evento *ReqTimer* é disparado, uma requisição de banda é enviada e a FSM entra no estado *Wait-For-Map*. Se, entretanto, um mapa chega trazendo um *grant* para envio de dados para o fluxo

antes do *timer* expirar, o *timer* pára, a FSM faz uma transição para o estado *To-Send* e um evento *SendPacket* ocorre. Se um pacote chega antes do *timer* expirar, o cabeçalho MAC é adicionado e o pacote é inserido na fila.

Caso a FSM esteja no estado *Wait-For-Map* e a SS recebe um mapa com um *grant* para envio de dados para o fluxo, então a FSM entra no estado *To-Send* e um evento *SendPacket* ocorre. No entanto, caso o *grant* seja para envio de requisição de banda, a FSM entra no estado *Decision* e um evento *Packet* ocorre. Quando um pacote chega e a FSM está no estado *Wait-For-Map*, o cabeçalho MAC é inserido e o pacote é incluído na fila.

A máquina de estados proposta por Shrivastav [46] não permite fragmentação e concatenação de pacotes no serviço rtPS. No módulo WiMAX essa deficiência foi corrigida. Desse modo, no estado *To-Send* vários pacotes podem ser concatenados e transmitidos em um único *grant*, ou um pacote pode ser fragmentado de forma que possa ser transmitido em um *grant* de tamanho menor.

3.1.4 Serviço ertPS

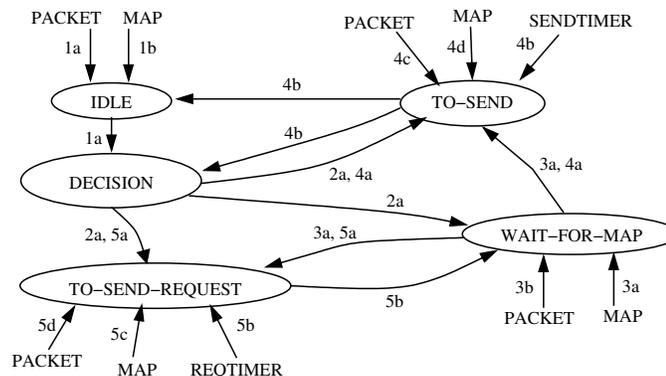


Figura 3.4: Máquina de estados para o serviço ertPS

A Figura 3.4 apresenta a máquina de estados finita definida para o serviço ertPS. Esta FSM tem os mesmos estados e eventos usados pela FSM do serviço rtPS, porém, as transições e as ações tomadas em alguns estados são diferentes.

Caso a FSM esteja no estado *Decision* e a Tabela de Alocação do fluxo tem um *grant* com tamanho menor do que o tamanho do pacote em *current_pkt*, a FSM entra no estado *To-Send-Request* e um evento *SendReq* ocorre. No entanto, caso o *grant* seja maior ou igual ao tamanho do pacote, há uma transição para o estado *To-Send* e o evento *SendPacket* é disparado. A mesma ação é tomada quando um mapa chega e o estado é *Wait-For-Map*.

No estado *To-Send-Request*, o *grant* é utilizado para requisitar um novo tamanho de *grant* através do envio de uma requisição de banda para a BS. Depois de enviar a requisição de banda, a FSM entra no estado *Wait-For-Map*.

No estado *To-Send*, antes de enviar um pacote, a FSM compara o tamanho do pacote (incluindo os cabeçalhos) com o tamanho do *grant*. Caso os tamanhos sejam iguais, o pacote é enviado. No entanto, caso o tamanho do *grant* seja maior do que o tamanho do pacote, uma requisição de banda *piggyback* é enviada com o pacote de dados para informar a BS sobre o novo tamanho de pacote.

3.1.5 Serviço nrtPS e BE

Além dos estados e eventos usados para os serviços UGS, rtPS e ertPS, foi definido o estado *Contention*, que implementa o algoritmo de *backoff*, o estado *Req-Sent*, que indica que uma requisição de banda foi enviada, e os seguintes eventos:

- *Unicast-Req*: um *grant unicast* para o envio de banda foi alocado para o fluxo de serviço;
- *Contention-Req*: um *grant* de contenção foi disponibilizado para o fluxo de serviço;
- *Contention-On*: a fase de contenção deve iniciar;
- *Contention-Slots*: a fase de contenção já foi iniciada;
- *Contention-Bkoff*: a fase de *backoff* é necessária pois a requisição enviada no *grant* de contenção não foi enviada com sucesso;

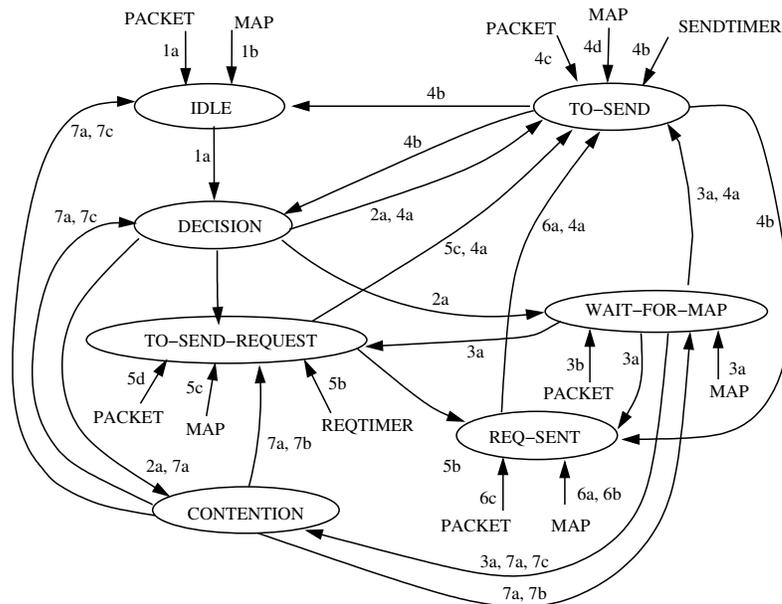


Figura 3.5: Máquina de estados para o serviço BE [46]

Quando a FSM entra no estado *Decision* com um evento *Packet*, a SS verifica se a Tabela de Alocação do fluxo tem algum *grant* disponível. Caso haja um *grant* para envio de dados, a FSM entra no estado *To-Send* e um evento *SendPacket* ocorre. Caso haja um *grant unicast* para envio de requisição de banda, há uma transição para o estado *To-Send-Request* e um evento *Unicast-Req* é disparado. Em ambas as situações, o processo de contenção é finalizado, caso esteja em andamento. Se há um *grant* de contenção para envio de requisição de banda e o processo de contenção não está em andamento, a FSM entra no estado *Contention* e um evento *Contention-On* ocorre. Se o processo de contenção já tiver iniciado, a FSM entra no estado *Contention* com o evento *Contention-Bkoff*.

Quando um mapa chega e a FSM está no estado *Wait-For-Map*, as ações tomadas são as mesmas descritas para o estado *Decision*, exceto quando o mapa traz um *grant* de contenção para o fluxo e o processo de contenção já tiver iniciado. Neste caso, a FSM entra no estado *Contention* e o evento *Contention-Slots* ocorre.

No estado *To-Send*, vários pacotes podem ser concatenados e transmitidos em um único *grant* ou um pacote pode ser fragmentado tal que possa ser enviado em um *grant* de tamanho menor. Caso um *grant* de dados não seja suficiente para enviar todos os pacotes armazenados na fila e não haja mais *grants* de dados pendentes na Tabela de Alocação, o fluxo de serviço envia uma requisição de banda *piggyback* e a FSM entra no estado *Req-Sent*. Caso haja pacotes na fila e a requisição *piggyback* não foi enviada, ocorre uma transição para o estado *Decision* e um evento *Packet* ocorre. Se a fila está vazia, há uma transição para o estado *Idle*.

No estado *To-Send-Request*, o *req_timer* está ativado. Quando o *timer* expira, uma requisição de banda *unicast* ou de contenção será enviada e a FSM entra no estado *Req-Sent*. Quando um mapa chega com um *grant* de dados para o fluxo de serviço no estado *Req-Sent*, o processo de contenção é finalizado, a FSM entra no estado *To-Send* e um evento *SendPacket* ocorre. Se o mapa não tem um *grant* de dados para o fluxo de serviço e o *timer* T16³ expirou, a janela de contenção dobra de tamanho e a FSM entra no estado *Decision* com o evento *Packet*.

A transição para o estado *Contention* e a ocorrência do evento *Contention-On* acontece quando o serviço inicia o processo de contenção. Neste caso, um número aleatório r é selecionado dentro da janela de contenção. Se o número de *slots* alocados para contenção é maior do que r , o *req_timer* é configurado para expirar no *slot* de contenção $r + 1$ e a FSM entra no estado *To-Send-Request*. Caso contrário, uma variável chamada *skipped* recebe $r - \text{número de slots de contenção}$ e a FSM entra no estado *Wait-For-Map*.

A entrada no estado *Contention* e a ocorrência do evento *Contention-Slots* indica que o processo de contenção já foi iniciado e que um *grant* de contenção foi alocado para o fluxo de serviço. Caso o número de *slots* alocados para contenção seja maior do que o valor da variável

³T16 é um *timer* definido no padrão IEEE 802.16 [1]. Quando uma SS envia uma requisição através de um *grant* de contenção e nenhum *grant* de dados é recebido dentro do intervalo T16, a SS considera que a transmissão foi perdida.

skipped, o *req_timer* é configurado para expirar no *slot* de contenção *skipped* + 1 e a FSM entra no estado *To-Send-Request*. Caso contrário, *skipped* é atualizada para *skipped* - número de *slots* de contenção e a FSM faz uma transição para o estado *Wait-For-Map*.

Caso o evento *Contention-Bkoff* ocorra no estado *Contention* e o número máximo de reenvios da requisição de banda foi atingido, o pacote armazenado em *current_packet* é descartado e o processo de contenção pára. Se a fila estiver vazia, a FSM entra no estado *Idle*; caso contrário, um pacote é retirado da fila e armazenado em *current_packet*, a FSM faz uma transição para o estado *Decision* e um evento *Packet* ocorre. Caso o número máximo de reenvios para requisição de banda não for atingido, a FSM executa as mesmas ações definidas para o evento *Contention-On*.

3.2 Teste de Conformidade

Os experimentos apresentados nesta seção têm como objetivo verificar se o modelo implementado está de acordo com as especificações do padrão IEEE 802.16 para topologias PMP que utilizam o modo de acesso TDD. Em particular, verifica-se se o módulo WiMAX implementa a divisão do tempo em *frames* e dos *frames* em *subframe downlink* e *subframe uplink* e se a alocação de *grants* para a transmissão de requisição de banda e de dados para os cinco tipos de serviço segue as regras do padrão. Os cinco tipos de serviço foram verificados tanto no sentido *uplink* como no sentido *downlink*.

O cenário de simulação consiste em uma BS com as SSs uniformemente distribuídas ao redor dela. A duração do frame é de 1 ms e a capacidade do canal é de 40 Mbps, assumindo uma divisão TDD 1:1 para *downlink* e *uplink*. O módulo usa o canal sem fio fornecido pelo simulador ns-2 que implementa a técnica de espalhamento espectral por seqüência direta (DSSS - *Direct-Sequence Spread-Spectrum*). Os cenários não foram escolhidos com intenção de simular redes operacionais. O objetivo é analisar o funcionamento dos mecanismos de acesso ao meio e de requisição e alocação de *slots* em diferentes situações de carga na rede, ou seja, tanto em situações de pouca carga oferecida como em situações de sobrecarga na rede. Para tal, foram utilizadas fontes CBR para simular o tráfego dos quatro tipos de serviço. Modelos de tráfego determinísticos são importantes na etapa de verificação da implementação pois facilitam a análise dos resultados [47]. Modelos de tráfego mais realistas devem ser considerados na análise dos mecanismos de provisão de QoS, tais como os mecanismos de escalonamento e de controle de admissão propostos nesta tese, e são utilizados no Capítulo 6.

3.2.1 Frames

Para verificar a implementação da divisão do tempo em frames, simulou-se um cenário com três estações cliente conectadas a uma estação base. A Figura 3.6 apresenta o envio das mensagens

UL-MAP para as estações cliente. Nota-se que a cada 1 ms um novo UL-MAP é enviado. Cada UL-MAP é enviado no início de um novo *frame* o que indica que os frames são gerados corretamente.

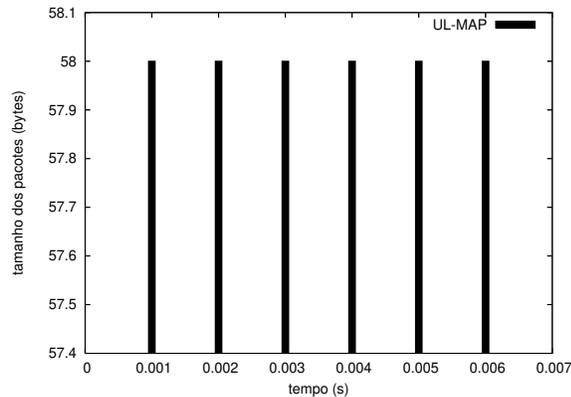


Figura 3.6: Divisão do tempo em frames

Para verificar a divisão dos frames em *subframe downlink* e *subframe uplink* simulou-se um cenário com uma estação base e duas estações cliente. Dois fluxos CBR com taxa de 5 Mbps geram o tráfego na rede, um no sentido *downlink* e outro no sentido *uplink*. Ambos os fluxos pertencem ao serviço BE. O tamanho do *frame* foi configurado em 1 ms, portanto, cada subframe tem duração de 0,5 ms.

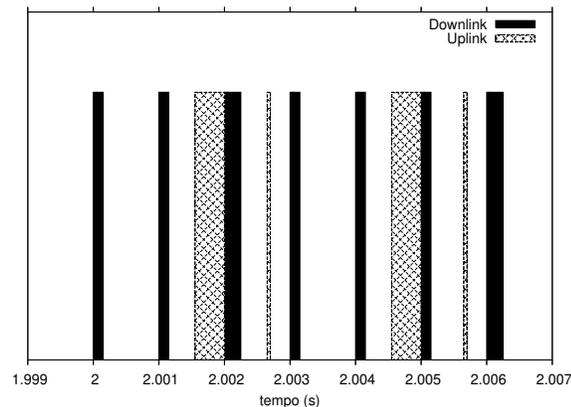


Figura 3.7: Divisão dos frames em subframe *downlink* e *uplink*

A Figura 3.7 mostra o tráfego enviado em um intervalo da simulação. O tráfego inclui os pacotes de dados, as mensagens UL-MAP e as mensagens de requisição de banda. Nota-se que a divisão dos *frames* se dá de maneira correta, dado que a primeira transmissão é sempre no sentido *downlink*, enquanto que a transmissão *uplink* inicia na segunda metade do *frame*.

Estes resultados evidenciam que o controle de acesso ao meio implementado no módulo WiMAX está de acordo com as especificações do padrão IEEE 802.16 para o modo de acesso TDD. Nota-se que a BS é capaz de coordenar as transmissões nos sentidos *downlink* e *uplink*, para que elas aconteçam de maneira sucessiva respeitando os limites de duração dos *subframes* e do *frame*.

3.2.2 Serviço UGS

Assim como no padrão IEEE 802.16, ao configurar uma conexão do tipo UGS no módulo WiMAX deve-se informar o intervalo entre *grants* desejado. Para verificar se as conexões UGS no sentido *uplink* recebem *grants* periódicos para o envio de dados, fez-se um experimento de simulação com um cenário consistindo em uma estação base e duas estações cliente. Uma das estações cliente possui um fluxo UGS que inicia a transmissão dos pacotes no instante 0,5 s. O tráfego é gerado por uma fonte CBR com taxa de 500 Kbps e o intervalo entre os *grants* fornecidos pela estação base é de 15 ms. A segunda estação cliente possui um fluxo BE que inicia a transmissão no instante 1 s. O tráfego é gerado por uma fonte CBR com taxa de 2 Mbps.

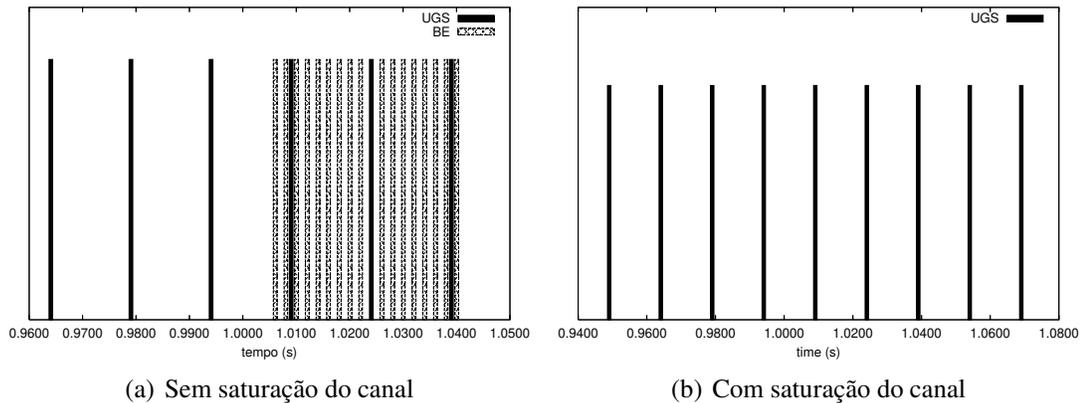


Figura 3.8: Verificação do serviço UGS no sentido *uplink*

A Figura 3.8(a) mostra o envio de pacotes de dados em um intervalo do tempo de simulação. Nota-se que mesmo com a entrada de um tráfego de melhor esforço na rede, a estação com serviço UGS obteve os *grants* no intervalo esperado.

A alocação de *grants* para o serviço UGS também foi avaliada em um cenário onde o canal é saturado por 10 fluxos BE, cada um gerando um tráfego CBR com taxa de 2 Mbps. Para facilitar a visualização, a Figura 3.8(b) apresenta apenas os pacotes enviados pela conexão UGS em um intervalo de tempo de simulação. Pode-se notar, novamente, que a entrada de fluxos de melhor esforço não prejudica a alocação de recursos para o serviço UGS.

Para verificar o serviço UGS no sentido *downlink* simulou-se uma rede com uma estação base e duas estações sem fio. A estação base possui dois fluxos *downlink*, um para cada uma das estações sem fio. Um dos fluxos é do tipo UGS e inicia a transmissão dos pacotes no instante 0,5 s. O tráfego é gerado por uma fonte CBR com taxa de 500Kbps e o intervalo entre os *grants* fornecidos pela estação base é de 15 ms. O segundo fluxo é do tipo BE e inicia a transmissão dos pacotes no instante 1 s. Os pacotes são gerados por uma fonte CBR com taxa de 2 Mbps.

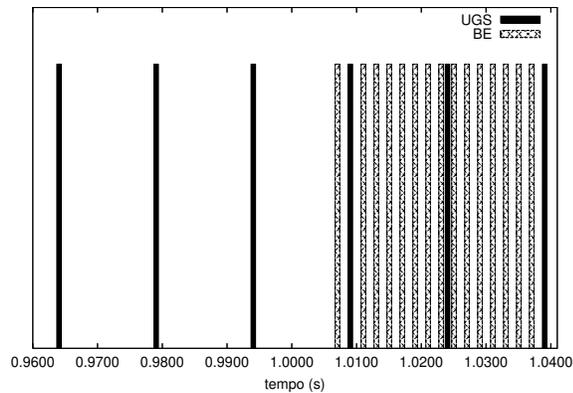


Figura 3.9: Verificação do serviço UGS no sentido *downlink*

A Figura 3.9 mostra o envio de pacotes de dados em um intervalo do tempo de simulação. Nota-se que mesmo com a entrada de um fluxo de melhor esforço na rede, o fluxo UGS obteve os *grants* no intervalo esperado.

3.2.3 Serviço rtPS

O cenário de simulação utilizado para verificar a alocação de recursos para o serviço rtPS no sentido *uplink* consiste em uma estação base e uma estação cliente. A estação cliente possui um fluxo *uplink* do tipo rtPS com pacotes gerados por uma fonte CBR com taxa de 1 Mbps. A estação base fornece *grants* periódicos para o envio de requisição de banda a cada 15 ms. O valor deste intervalo deve ser definido pelo usuário no *script* de configuração da rede.

A Figura 3.10 ilustra o envio de pacotes de requisição de largura de banda e de pacotes de dados em um intervalo da simulação. Nota-se que os *grants* para envio de requisição são recebidos no intervalo esperado e, após cada requisição, a estação recebe *grants* para o envio dos dados.

Dado que para conexões *downlink* que pertencem ao serviço rtPS não é necessário fornecer *grants* periódicos para o envio de requisição de banda, o teste para este tipo de serviço no sentido *downlink* verifica se um fluxo rtPS recebe *grants* para o envio de dados e se consegue ter prioridade sobre um fluxo BE.

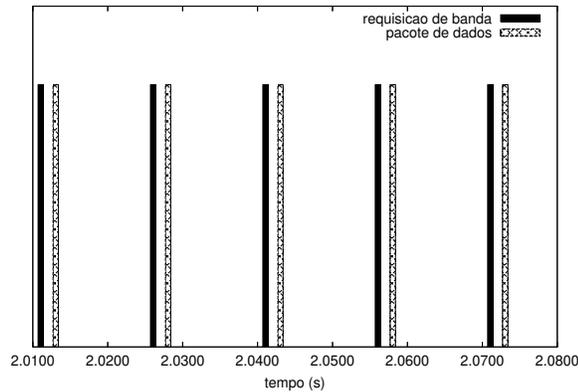


Figura 3.10: Verificação do serviço rtPS no sentido *uplink*.

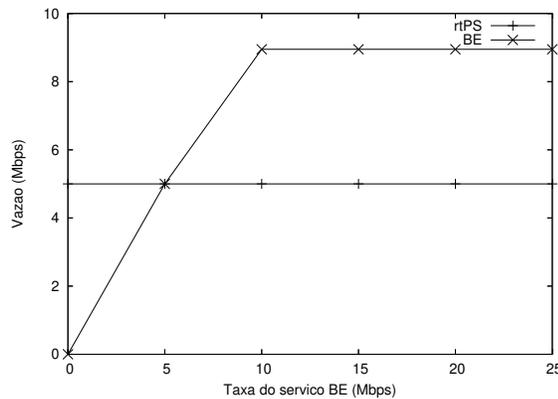


Figura 3.11: Verificação do serviço rtPS no sentido *downlink*.

O cenário de simulação consiste em uma estação base e duas estações cliente. A estação base possui dois fluxos *downlink*, um para cada estação cliente. Um dos fluxos é do tipo rtPS com pacotes gerados por uma fonte CBR com taxa de 5 Mbps. O segundo fluxo é da classe BE com pacotes gerados por uma fonte CBR. Para verificar a influência do tráfego de melhor esforço sobre o tráfego de tempo real, a taxa da fonte BE varia de 0 a 25 Mbps.

A Figura 3.11 ilustra a vazão de ambos os fluxos. Nota-se que o aumento na carga oferecida pelo fluxo de melhor esforço não interfere no serviço oferecido ao fluxo rtPS, que obteve vazão igual a taxa de dados gerada pela fonte em todas as situações.

3.2.4 Serviço ertPS

O cenário de simulação para testar o serviço ertPS consiste em uma estação base e uma estação cliente. A estação cliente tem uma conexão ertPS no sentido *uplink* que gera os pacotes em intervalos constantes de 20 ms. Para avaliar o mecanismo de alocação de banda variou-se o

tamanho dos pacotes. No início os pacotes têm tamanho de 200 bytes. No instante 5 s, o tamanho dos pacotes muda para 500 bytes e no instante 5,2 s ele é reduzido para o valor inicial. O intervalo entre *grants* foi configurado em 20 ms.

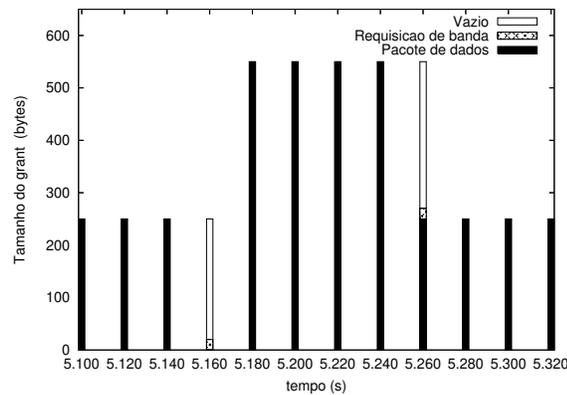


Figura 3.12: Verificação do serviço ertPS no sentido *uplink*.

A Figura 3.12 apresenta o tamanho dos *grants* alocados para a conexão ertPS durante um período da simulação. No início da simulação, quando o tamanho do pacote é 200 bytes, a BS aloca um *grant* de 270 bytes para a transmissão dos dados e dos cabeçalhos. Quando o tamanho do pacote aumenta para 500 bytes, a conexão ertPS envia uma requisição de banda no instante 5.16 e a BS aloca um *grant* de 570 bytes. No intervalo [5,18; 5,24], a conexão envia pacotes com tamanho de 500 bytes. No instante 5,26, a conexão envia um pacote de 200 bytes com uma requisição de banda *piggyback* para solicitar a redução do tamanho do *grant*. Embora os *grants* tenham sido alocados periodicamente em intervalos de 20 ms, há um retardo na transmissão do primeiro pacote de 500 bytes. Isso acontece porque a aplicação CBR na SS e o escalonador na BS não estão sincronizados.

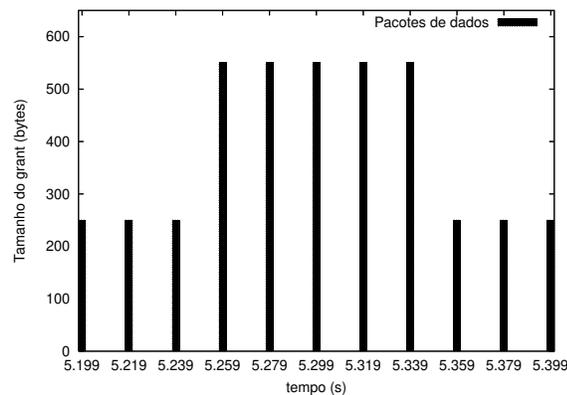


Figura 3.13: Verificação do serviço ertPS no sentido *downlink*.

O mesmo cenário foi utilizado para verificar o serviço ertPS no sentido *downlink*. A Figura 3.13 apresenta o mesmo comportamento verificado na Figura 3.12, exceto pelo fato de que as requisições de banda não são necessárias neste caso. Esses resultados mostram que o serviço ertPS foi implementado de acordo com as especificações do padrão IEEE 802.16.

3.2.5 Serviço nrtPS

O experimento realizado para verificar a alocação de recursos para o serviço nrtPS no sentido *uplink* utiliza um cenário que consiste em uma estação base e 11 estações cliente. Uma das estações cliente possui um fluxo de serviço do tipo nrtPS com pacotes gerados por uma aplicação CBR com taxa de 1 Mbps. O fluxo recebe *grants unicast* para envio de requisição a cada 15 ms. Este intervalo deve ser definido pelo usuário no *script* de configuração da rede. As outras 10 estações cliente têm um fluxo de serviço do tipo BE com pacotes gerados por uma aplicação CBR com taxa de 2 Mbps.

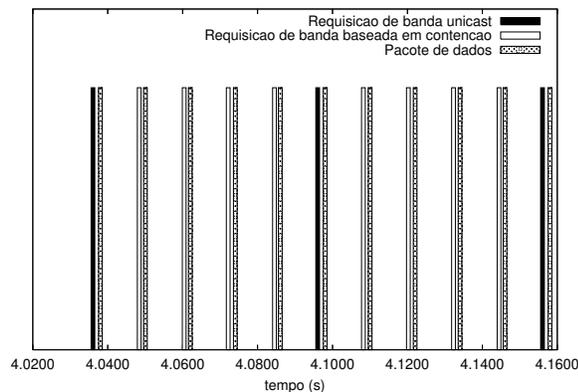


Figura 3.14: Verificação do serviço nrtPS no sentido *uplink*.

A Figura 3.14 mostra os pacotes de requisição e de dados enviados pelo fluxo nrtPS em um intervalo da simulação. Nota-se que mesmo com uma alta carga na rede, o fluxo nrtPS consegue enviar requisições através de *grants unicast* e de contenção, bem como pacotes de dados. Alguns *grants unicast* não foram utilizados dado que no instante em que foram alocados a fila do fluxo de serviço estava vazia.

Para verificar o serviço nrtPS no sentido *downlink* utilizou-se um cenário com 1 estação base e 2 estações cliente. Uma das estações cliente possui uma conexão nrtPS no sentido *downlink* com tráfego CBR sendo gerado a uma taxa de 5 Mbps. A outra estação cliente possui uma conexão BE no sentido *downlink* com tráfego CBR cuja taxa varia de 0 a 25 Mbps.

A Figura 3.15 ilustra a vazão de ambos os tipos de serviço. Pode-se observar que a alocação de recursos para o serviço nrtPS no sentido *downlink* satisfaz as necessidades do fluxo mesmo quando a rede é sobrecarregada por um fluxo de melhor esforço.

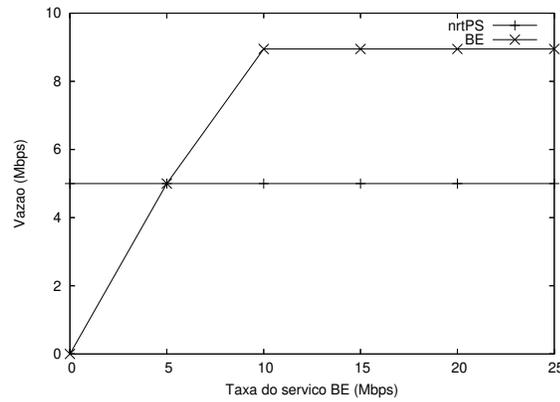


Figura 3.15: Verificação do serviço nrtPS no sentido *downlink*.

3.2.6 Serviço BE

O experimento realizado para verificar o serviço BE no sentido *uplink* verifica se um fluxo BE consegue enviar pacotes de requisição de banda e pacotes de dados na presença de fluxos de maior prioridade. O cenário consiste em cinco estações cliente conectadas a uma estação base. Duas estações cliente possuem serviço UGS, duas possuem serviço rtPS e uma possui serviço BE. O tráfego é gerado por aplicações CBR com taxa de 2 Mbps para os serviços UGS e rtPS e taxa de 1 Mbps para o serviço BE.

A Figura 3.16 apresenta os pacotes de requisição de largura de banda e os pacotes de dados enviados pela estação com serviço BE em um intervalo da simulação. Nota-se que mesmo na presença de tráfego de maior prioridade o serviço BE recebe *grants* para o envio de requisição de banda e para o envio de dados.

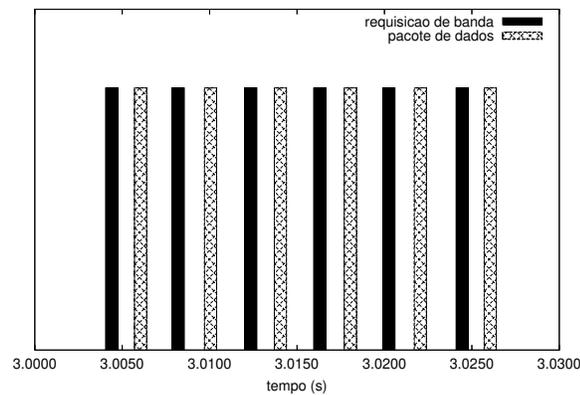


Figura 3.16: Verificação do serviço BE

A partir dos testes realizados com os outros tipos de serviço no sentido *downlink* é possível notar que o serviço de melhor esforço no sentido *downlink* obtém oportunidades de transmissão

mesmo na presença de tráfego de maior prioridade.

3.3 Considerações Finais do Capítulo 3

Este capítulo apresenta a implementação e o teste de conformidade de um módulo para simulação de redes WiMAX. O módulo é implementado na ferramenta ns-2 e atende as especificações do padrão IEEE 802.16 para redes com topologia PMP e modo de acesso ao meio TDD. O módulo WiMAX inclui as funcionalidades de requisição e alocação de banda e provisão de QoS. Adicionalmente, permite que os usuários configurem requisitos de QoS para aplicações com diferentes demandas.

Experimentos foram projetados para avaliar a implementação das principais funcionalidades da camada de acesso ao meio do padrão IEEE 802.16. Os resultados indicam que a BS é capaz de gerenciar o acesso ao meio nos sentidos *downlink* e *uplink*, bem como fazer a alocação de *grants* para envio de requisição de banda e de dados conforme as regras estabelecidas para cada um dos cinco tipos de serviço. Além disso, é possível concluir que as estações cliente são capazes de enviar requisição de banda e pacotes de dados de acordo com a alocação de *grants* informada pela estação base através da mensagem UL-MAP.

O módulo implementado pode ser empregado em diversos estudos sobre redes IEEE 802.16, em especial estudos sobre alocação de banda e provisão de QoS, inclusive facilitando a inclusão de novas funcionalidades e mecanismos, dada a estrutura modularizada do código.

Trabalhos futuros irão focar a implementação da camada física OFDM especificada no padrão IEEE 802.16, de escalonadores mais eficientes para as estações cliente e para o tráfego *downlink* na estação base.

Capítulo 4

Escalonamento em redes IEEE 802.16

O escalonamento em redes IEEE 802.16 compreende o escalonamento do tráfego *downlink*, realizado pela BS, e escalonamento do tráfego *uplink*, realizado por dois escalonadores, um na BS e outro nas SSs. Para executar a alocação dos recursos, os escalonadores usam informações sobre os requisitos de Qualidade de Serviço e a situação das filas das conexões.

O escalonador *downlink* e o escalonador *uplink* das SSs têm acesso direto às filas das conexões. O escalonador *uplink* localizado na BS, por outro lado, depende das mensagens de requisição de banda enviadas pelas SSs para manter-se informado sobre o estado das filas de cada conexão. Essas requisições, além de aumentarem a carga na rede, podem sofrer atrasos, gerados pelo mecanismo de contenção, por exemplo, resultando na entrega de informações desatualizadas. Adicionalmente, diferente dos outros dois escalonadores, o escalonador *uplink* da BS deve alocar recursos não só para transmissão de dados, mas também para o envio das requisições de banda. Em ambos os casos, a alocação dos recursos deve ser feita de tal forma que os requisitos de QoS de todas as conexões sejam garantidos.

Assim sendo, pode-se afirmar que um dos maiores desafios no projeto da camada MAC dos equipamentos para redes WiMAX está na implementação do escalonador *uplink* da BS. Políticas de escalonamento não são especificadas pelo padrão IEEE 802.16, o que fornece uma oportunidade para os fabricantes pesquisarem algoritmos inovadores que diferenciem seus produtos.

Independente da política de escalonamento adotada para o tráfego *uplink*, os seguintes aspectos definidos no padrão devem ser levados em consideração na implementação do escalonador:

- a distribuição dos recursos deve ser feita com base nas requisições de banda enviadas pelas SSs e nos parâmetros de QoS de cada conexão, sendo que conexões distintas que utilizam um mesmo tipo de serviço podem ter diferentes valores para um mesmo parâmetro de QoS;

- a alocação de banda deve permitir não só a transmissão de dados, mas também a transmissão de requisições de banda de acordo com o mecanismo de requisição definido para cada tipo de serviço;
- todos os parâmetros de QoS definidos pelo padrão devem ser garantidos.

Além dos pontos citados, espera-se que o escalonador use de forma eficiente a largura de banda disponível de forma que um maior número de usuários possa ser admitido, consequentemente, resultando em altos níveis de utilização da rede.

Embora o escalonador seja implementado na camada MAC, a tecnologia utilizada na camada física pode influenciar o seu projeto. Quando se utiliza a tecnologia de portadora única (*Single Carrier - SC*), há apenas uma portadora e toda a frequência é dada para uma SS. Na tecnologia OFDM múltiplas subportadoras formam um *slot* físico, mas pelo fato de serem transparentes para a camada MAC, as subportadoras podem ser vistas como um canal lógico, sob o ponto de vista do escalonador. Entretanto, cada subcanal pode usar diferentes modulações de forma que SSs distintas podem apresentar diferentes taxas de transmissão. Ao contrário das tecnologias SC e OFDM, nas quais apenas o domínio do tempo é considerado, a tecnologia OFDMA requer alocação dos recursos em duas dimensões: frequência e tempo. Em outras palavras, o escalonador deve decidir não só sobre a alocação de *slots*, mas também de subportadoras para cada usuário. Dado que mais de um usuário pode usar o canal em um mesmo instante de tempo, é fundamental que o escalonamento em sistemas OFDMA considere as características da camada física. A Figura 4.1 ilustra as diferenças entre os três tipos de camada física.

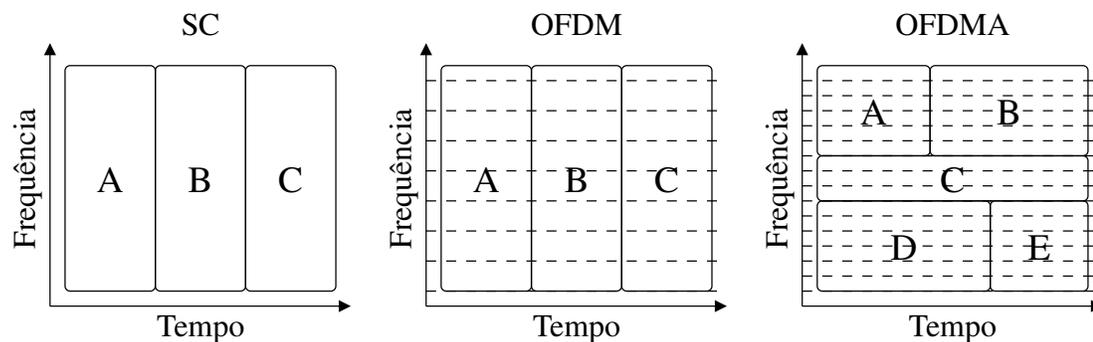


Figura 4.1: Comparação entre as camadas física SC, OFDM e OFDMA (as letras A, B, C, D e E representam diferentes usuários)

A Seção 4.1 apresenta algumas das principais soluções propostas na literatura para o escalonamento *uplink* na BS. Essas propostas são divididas em dois grupos. O primeiro grupo inclui os escalonadores que não utilizam informações da camada física, tendo como principal objetivo

garantir a Qualidade de Serviço requisitada por cada fluxo de serviço. O segundo grupo compreende os escalonadores que consideram a tecnologia usada na camada física para auxiliar na tomada de decisões.

A Seção 4.2 introduz um novo escalonador *uplink* para a BS. Este escalonador não utiliza informações da camada física na alocação dos recursos, entretanto, diferente da maioria das propostas disponíveis na literatura, ele é totalmente compatível com o padrão IEEE 802.16, dado que atende os cinco tipos de fluxo de serviço bem como seus respectivos parâmetros de QoS.

4.1 Estado da arte da pesquisa em escalonamento em redes IEEE 802.16

4.1.1 Propostas que não consideram a camada física

A tendência nas primeiras soluções propostas na literatura [20, 34, 37] consistia em adaptar para redes IEEE 802.16 políticas de escalonamento clássicas propostas para redes cabeadas. Esses trabalhos, bem como [17] e [27], que seguem a mesma tendência, utilizam uma combinação de políticas tais como *Strict Priority*, *Weighted Fair Queueing* (WFQ) [8] e *Earliest Deadline First* (EDF), resultando em esquemas de escalonamento complexos. Propostas mais recentes disponíveis na literatura [10, 45], por outro lado, têm investido em idéias de implementação mais simples. Dado que o escalonador *uplink* da BS é executado a cada *frame* e, em determinadas configurações, pode-se ter até 400 *frames* por segundo, soluções simples tornam-se mais atrativas.

Hawa e Petr [37] propõem um mecanismo para escalonamento do tráfego *uplink* em redes DOCSIS [13] e em redes IEEE 802.16. O escalonador implementa uma combinação de escalonamento baseado em prioridades e da disciplina WFQ para atribuir *grants* às SSs. O serviço UGS tem a prioridade mais alta para alocação de banda, o restante dos fluxos são servidos utilizando WFQ com prioridade, ou seja, se dois *grants* têm tempo final igual, o de maior prioridade é servido primeiro. Entretanto, este mecanismo aloca *grants* para as conexões e não para SSs, o que era permitido em versões anteriores do padrão. Não são apresentados em [37] resultados que comprovem a eficiência e o desempenho do mecanismo proposto.

Wongthavarawat e Ganz [34] propõem um mecanismo de escalonamento que utiliza uma combinação das disciplinas *Strict Priority*, EDF e WFQ. A alocação de largura de banda acontece em dois passos. No primeiro passo, o escalonador distribui a banda entre os diferentes tipos de serviço utilizando a disciplina *Strict Priority*. No segundo passo, a banda recebida por cada serviço é distribuída entre as conexões. Nesta fase, o serviço UGS usa uma política de banda fixa, o serviço rtPS utiliza a disciplina EDF, o serviço nrtPS utiliza a disciplina WFQ e

o serviço BE distribui a banda igualmente entre todas as conexões. O mecanismo proposto é avaliado através de experimentos de simulação, porém apenas tráfego rtPS e BE são utilizados.

Chen, Jiao e Wang [27] utilizam a mesma estrutura proposta por Wongthavarawat e Ganz, porém, no primeiro passo, a alocação de banda é feita com a política *Deficit Fair Priority Queue* (DFPQ). Os resultados de simulação mostram que o algoritmo DFPQ é mais justo do que o algoritmo *Strict Priority* e evita a inanição do tráfego de melhor esforço. Entretanto, os autores não mostram resultados sobre o retardo. Dado que para serviços UGS e rtPS o retardo é um parâmetro de QoS de suma importância, sua análise torna-se indispensável para verificação do nível de QoS oferecido pelo mecanismo.

Tarchi, Fantacci e Bardazzi [17] também propõem uma combinação de disciplinas de escalonamento. No primeiro passo, a largura de banda é distribuída entre os serviços UGS, rtPS, nrtPS e BE utilizando a técnica *strict semi-preemptive priority*. No passo seguinte, a banda é distribuída entre as conexões utilizando o algoritmo *Packet Based Round Robin* (PBRR) para o serviço UGS, a disciplina EDF para o serviço rtPS e a disciplina WFQ para os serviços nrtPS e BE.

Sayenko, Alanen e Hämäläinen [10] apresentam uma solução baseada na disciplina *Round Robin* (RR) que compreende três estágios. No primeiro estágio, a BS calcula o número mínimo de slots que deve ser fornecido para cada conexão de modo que os requisitos de QoS sejam garantidos. Esse cálculo utiliza equações diferenciadas para cada tipo de serviço que levam em consideração as requisições de banda enviadas pelas SSs bem como os parâmetros de QoS fornecidos pelas conexões. No segundo estágio, caso haja *slots* não alocados, equações que emulam a disciplina *Round Robin* distribuem os *slots* entre as conexões rtPS, nrtPS e BE. Entretanto, diferente do algoritmo RR padrão, a solução proposta utiliza apenas um passo, ao invés de múltiplos passos, e evita que o requisito de taxa máxima seja excedido. Finalmente, no terceiro estágio, a BS faz a ordenação dos *slots* dentro da mensagem UL-MAP para que os requisitos de tempo sejam garantidos.

Masri, Abdellatif e Juanole [45] propõem uma solução para o gerenciamento dos recursos no sentido *uplink* que compreende um mecanismo de escalonamento para a BS, um mecanismo de escalonamento para as SSs, um mecanismo de controle de admissão e um mecanismo para requisição de largura de banda diferente daquele proposto no padrão IEEE 802.16. Nesta proposta, as mensagens de requisição de banda enviadas por uma SS indicam a quantidade de banda requisitada por todas as conexões *uplink* estabelecidas por ela e não por uma conexão individual, como especifica o padrão. As mensagens de requisição possuem dois campos: o primeiro, chamado CB (*Contracted Bytes*) contém o número de *bytes* requisitado pela SS que estão dentro do contrato estabelecido pelo controle de admissão; o segundo campo, chamado AB (*Additional Bytes*), contém os *bytes* requisitados pela SS que excedem o valor estabelecido no contrato. O escalonador *uplink* da BS aloca primeiro o número de *slots* necessários para transmitir os *bytes* indicados no campo CB de cada requisição de banda recebida. No passo

seguinte, se os *slots* restantes forem suficientes para satisfazer todas as requisições adicionais, os recursos são distribuídos de acordo com os pedidos incluídos no campo AB das mensagens de requisição de banda. Caso contrário, os *slots* restantes são distribuídos proporcionalmente ao número de *bytes* solicitados no campo AB de cada requisição de banda. Garantias de latência máxima e taxa mínima são implementadas pelo escalonador das SSs.

Além dos trabalhos discutidos nesta seção, existem propostas que lidam ou apenas com tráfego de tempo real [9, 16, 22, 39] ou com tráfego de melhor esforço [19, 50]. A Tabela 4.1 resume as vantagens e desvantagens das propostas discutidas nesta seção bem como do escalonador proposto nesta tese.

Tabela 4.1: Vantagens e desvantagens das propostas discutidas na Seção 4.1.1

Proposta	Vantagens	Desvantagens
Hawa e Petr, 2002 [37]	Garante latência para o serviço UGS e taxa mínima para os demais serviços.	Algoritmo complexo baseado em uma hierarquia de escalonadores; não inclui o serviço ertPS.
Wongthavarawat e Ganz, 2003 [34]	Garantia de latência para os serviços UGS e rtPS.	Algoritmo complexo baseado em uma hierarquia de escalonadores; não inclui o serviço ertPS; não garante taxa mínima.
Chen, Jiao e Wang, 2005 [27]	Garantias de latência, taxa mínima e taxa máxima.	Algoritmo complexo baseado em uma hierarquia de escalonadores; não inclui o serviço ertPS.
Tarchi, Fantacci e Bardazzi, 2006 [17]	Garantias de latência para os serviços rtPS e UGS.	Algoritmo complexo baseado em uma hierarquia de escalonadores; não inclui o serviço ertPS; não fornece garantia de taxa mínima.
Sayenko, Alanen e Hämäläinen, 2008 [10]	Algoritmo simples baseado na política RR; inclui os 5 tipos de serviço; garante taxa mínima e taxa máxima;	Não fornece garantias de latência máxima.
Masri, Abdelatif e Juanole, 2009 [45]	Algoritmo simples que trabalha em conjunto com o escalonador das SSs e com o controle de admissão.	Exige modificações nas especificações do padrão IEEE 802.16; não considera o serviço ertPS.
Borin e Fonseca	Algoritmo simples; inclui os 5 tipos de serviço; garante latência máxima e taxa mínima; faz controle de taxa máxima e tamanho máximo da rajada.	Conexões BE são servidas por uma fila FIFO, o que pode gerar distribuição desigual dos recursos para as conexões que usam esse serviço.

4.1.2 Propostas que consideram a camada física

Singh e Sharma [52] introduzem um conjunto de algoritmos para a BS alocar subcanais e *slots* para diferentes SSs em uma rede que utiliza a tecnologia OFDMA e o modo de acesso ao meio TDD. Uma heurística é proposta para alocar diferentes subcanais para as SSs a cada *slot* com o objetivo de maximizar a vazão na rede. Entretanto, os autores mostram que esta heurística não é justa dado que, para maximizar o desempenho do sistema, ela favorece as SSs com melhores condições de canal. A heurística é, então, modificada, para se obter uma solução mais justa e quatro novas heurísticas são apresentadas. Diferentes combinações destas heurísticas são propostas para o escalonador *uplink* da BS baseadas na carga e nas características do canal da rede. O escalonador proposto tenta satisfazer, primeiro, as necessidades das conexões UGS. No próximo passo, o escalonador serve as conexões rtPS seguidas pelas conexões nrtPS. Conexões nrtPS sem requisitos de taxa mínima e conexões BE compartilham os recursos restantes. Os autores não apresentam resultados de desempenho para as soluções propostas.

O escalonamento *uplink* em sistemas OFDM é analisado por Huang *et alli* [32]. O escalonador proposto aloca largura de banda e potência com o objetivo de maximizar uma função utilidade que modela a Qualidade de Serviço que deve ser garantida aos usuários. Uma solução ótima bem como uma família de algoritmos sub-ótimos de baixa complexidade são introduzidos. O algoritmo ótimo determina a alocação ótima da portadora ordenando os usuários em cada tom. Dada uma alocação ótima da portadora, a alocação de potência ótima é dada utilizando uma técnica conhecida como *waterfilling*. Em cada algoritmo sub-ótimo, as mesmas duas fases são utilizadas, mas com algumas modificações para reduzir a complexidade da computação na alocação ótima da portadora. Mais especificamente, os algoritmos começam com uma fase de alocação da portadora que atribui cada subportadora a no máximo um usuário. Ao invés de usar métricas dadas por valores ótimos, métricas baseadas em uma alocação constante de potência para todas as portadoras atribuídas a um usuário são consideradas. Na fase seguinte, a alocação de potências utiliza o esquema *waterfilling* assim como no algoritmo ótimo. Segundo os autores, a solução ótima tem uma complexidade computacional proibitiva e foi desenvolvida apenas para ser usada como base para derivar as soluções sub-ótimas. Resultados de simulação em termos de taxa e justiça entre os diversos usuários são apresentados para os algoritmos sub-ótimos. Entretanto, os autores não avaliam a capacidade das soluções propostas para prover Qualidade de Serviço.

Bai, Shami e Ye [53] propõem um conjunto de mecanismos para provisão de QoS em redes IEEE 802.16 que utilizam a tecnologia de portadora única na camada física. A solução inclui um mecanismo de escalonamento para a BS que utiliza a abordagem *cross-layer*. Além das requisições de banda, o escalonador considera o tipo de modulação utilizado pelas SSs na camada física e, desse modo, faz a alocação de símbolos modulados ao invés de *slots*. Em modulações diferentes, o número de símbolos necessários para o envio de uma mesma quantidade de bits varia. A escolha das requisições a serem atendidas se dá seguindo um valor de prioridade de-

finido pelas SSs. Embora essa técnica resulte em um escalonador menos complexo na BS, ela limita a interoperabilidade entre equipamentos de diferentes fabricantes, dado que todas as SSs devem ser capazes de calcular os valores de prioridade. Ao final do processo de alocação, um módulo para criação da mensagem UL-MAP converte os símbolos alocados para número de *slots*. Resultados obtidos a partir de experimentos de simulação com os serviços rtPS, nrtPS e BE mostram que a solução proposta é capaz de garantir os requisitos de QoS das conexões.

Chen *et alli* [15] utilizam controladores *fuzzy* para prover justiça e QoS em redes WiMAX com camada física OFDMA. O escalonador proposto, chamado FQFC (*fairness and QoS guaranteed scheduling approach with fuzzy controls*), atribui para cada conexão uma prioridade e uma oportunidade de transmissão (TXOP) a partir de valores inferidos por um estimador *fuzzy*. Primeiro, o escalonador inicializa as duas variáveis com base nas características das conexões e os ajusta para que se adaptem à dinâmica do sistema. O valor de prioridade é usado para determinar a ordem com que as conexões realizarão as transmissões e este valor é ajustado de acordo com a qualidade do canal, requisitos de QoS e o tipo do serviço. O valor da variável TXOP é configurado, considerando a taxa de transmissão e a diferença no tamanho da fila entre duas transmissões consecutivas na camada MAC, o que fornece o número máximo de pacotes que uma conexão consegue transmitir em um *frame*. Para conexões de tempo real, o valor de prioridade é adicionalmente ajustado de acordo com a taxa de perda de pacotes, de forma que todas as conexões associadas ao mesmo tipo de serviço tenham mesma taxa de perda de pacotes. A justiça entre conexões nrtPS se dá garantindo que todas as conexões associadas à esse serviço tenham a mesma média para a razão entre vazão e taxa mínima. O escalonador, também, prevê justiça entre os diferentes tipos de serviço garantindo que a vazão média das conexões não exceda o requisito de taxa máxima. Resultados obtidos, via simulação, com diferentes cenários, incluindo situações diversas de canal, mostram que a solução é capaz de fornecer QoS bem como justiça entre as conexões.

Gidlund e Wang [44] introduzem um algoritmo *cross-layer* para o escalonamento *uplink* na BS que suporta os serviços UGS, rtPS, nrtPS e BE. O algoritmo calcula a quantidade ótima de *slots* que deve ser alocada em cada *frame*, de forma a minimizar a probabilidade de bloqueio dos diferentes tipos de serviço. Ele integra, adicionalmente, um esquema de equidade proporcional (*proportional fairness*) às informações de qualidade do canal das SSs, para fazer a alocação dos *slots* entre as conexões associadas ao mesmo tipo de serviço. Para cada conexão UGS, o escalonador calcula um valor de prioridade a partir do requisito de latência máxima e um fator que indica o grau de justiça (*proportional fairness*) entre os usuários. O grau de justiça de uma conexão é igual a razão entre a largura de banda da conexão no *frame* atual, que indica a qualidade do canal entre a BS e a SS, e a vazão média alcançada pela conexão dentro de uma janela de tempo. Para o cálculo do valor de prioridade para as conexões rtPS, além dos parâmetros utilizados no cálculo da prioridade das conexões UGS, o escalonador considera o tamanho da fila da conexão. Para conexões nrtPS, o valor de prioridade é calculado com base

no grau de justiça e no tamanho da fila da conexão. Para conexões BE apenas a qualidade do canal é considerada no cálculo da prioridade. Os autores mencionam que o requisito de taxa mínima é garantido para conexões UGS, rtPS e nrtPS, porém não explicam como essa garantia é alcançada. Além disso, os autores não informam como se dá a alocação de recursos para o envio de mensagens de requisição de banda.

Embora tanto o WiMAX Forum quanto os fabricantes de equipamentos WiMAX tenham dado maior atenção para as tecnologias OFDM e OFDMA, poucos são os trabalhos disponíveis, na literatura, que fornecem soluções para o escalonamento *uplink* com Qualidade de Serviço nessas redes. A grande maioria das soluções propostas foca no escalonamento *downlink* [21, 23, 28, 33, 36, 55, 56], que, conforme discutido anteriormente, apresenta menor complexidade.

A Tabela 4.2 apresenta um resumo das vantagens e desvantagens das propostas discutidas nesta seção.

Tabela 4.2: Vantagens e desvantagens das propostas discutidas na Seção 4.1.2

Proposta	Vantagens	Desvantagens
Singh e Sharma, 2006 [52]	Tenta satisfazer os requisitos de QoS ao mesmo tempo em que explora a diversidade nos canais dos usuários.	Melhores canais são alocados para conexões UGS; autores não apresentam avaliação de desempenho da proposta.
Huang <i>et alli</i> , 2007 [32]	Considera limitações de potência por usuário.	Autores não incluem resultados que evidenciem a capacidade do escalonador em prover QoS; provisão de QoS não inclui o conjunto de parâmetros definido no padrão.
Bai, Shami e Ye, 2008 [53]	Utiliza uma abordagem <i>cross-layer</i> que considera as diferentes modulações utilizadas pelas conexões; fornece uma solução de escalonamento <i>uplink</i> integrada para a BS e as SSs.	Garantias de latência, taxa mínima e taxa máxima fornecidas apenas pelas SSs.
Chen <i>et alli</i> , 2009 [15]	Considera requisitos de latência, taxa mínima e taxa máxima e provê justiça entre as conexões.	O requisito de taxa mínima do serviço rtPS e o serviço ertPS não são incluídos na proposta.
Gidlund e Wang, 2009 [44]	Utiliza uma abordagem <i>cross-layer</i> que considera a qualidade do canal das SSs; fornece justiça entre SSs com qualidade de canal diferente; minimiza a probabilidade de bloqueio dos diferentes tipos de serviço.	Não considera o serviço ertPS; não informa como se dá a alocação de recursos para o envio de requisição de banda.

4.2 Escalonador *Uplink* para Provisão de QoS em redes IEEE 802.16

Seguindo a tendência das soluções mais recentes disponíveis, na literatura, para o escalonamento *uplink* em redes IEEE 802.16, o escalonador proposto, nesta tese, utiliza um algoritmo de implementação simples, ao invés de utilizar uma hierarquia de políticas desenvolvidas para redes cabeadas. Entretanto, diferente das propostas de Sayenko *et alli* [10] e de Masri *et alli* [45], discutidas na Seção 4.1.1, o escalonador proposto é compatível com as especificações do padrão IEEE 802.16. A solução apresentada por Sayenko *et alli* não provê garantias de latência máxima, enquanto a solução introduzida por Masri *et alli* não inclui o serviço ertPS e exige modificações no mecanismo de requisição de banda definido no padrão.

Dentre os mecanismos que consideram informações da camada física durante o escalonamento, observa-se também que as soluções propostas negligenciam partes da especificação da camada MAC. Embora o escalonador proposto, nesta tese, não utilize dados da camada física na alocação dos recursos, ele pode ser facilmente estendido para considerar usuários com diferentes taxas de transmissão devido ao uso de modulações distintas, como acontece em um canal OFDM, por exemplo. A adaptação do escalonador para o uso em redes com canal OFDMA seria mais complexa dado que além de *slots* deve-se alocar subportadoras. Entretanto, a política proposta pode servir como base para o desenvolvimento de um escalonador OFDMA que forneça QoS e seja totalmente aderente ao padrão.

A solução proposta utiliza três filas com prioridades diferentes, como mostra a Figura 4.2. A fila de baixa prioridade armazena as requisições de banda enviadas por conexões BE. A fila intermediária armazena as requisições de banda enviadas por conexões rtPS e nrtPS. Essas requisições podem migrar para a fila de alta prioridade para que seus requisitos de QoS sejam garantidos. Além das requisições provenientes da fila intermediária, a fila de alta prioridade armazena os *grants* periódicos para envio de dados e os *grants* periódicos para envio de requisição de banda que devem ser escalonados no *frame* seguinte. A BS executa o escalonador a cada *frame* e a agenda de escalonamento é enviada para as SSs na mensagem UL-MAP.

Grants periódicos são gerados e inseridos na fila de alta prioridade em intervalos pré-definidos. Os intervalos entre *grants* UGS e *grants* ertPS são especificados pelas conexões através do parâmetro *unsolicited grant interval*. O intervalo entre *grants* periódicos para envio de requisição de banda é definido pela BS para conexões nrtPS e pelo parâmetro *unsolicited polling interval* para conexões rtPS. Dessa forma, *grants* periódicos para conexões UGS, ertPS, rtPS e nrtPS são garantidos conforme especificado no padrão. Para garantir o requisito de latência máxima, a BS atribui *deadlines* para cada requisição rtPS que entra na fila intermediária. Cada vez que o escalonador é executado, as requisições cujos *deadlines* expiram dois *frames* adiante e cujas conexões ainda não receberam a taxa mínima em uma janela de duração T migram da fila intermediária para a fila de alta prioridade. Dessa maneira, o escalonador garante

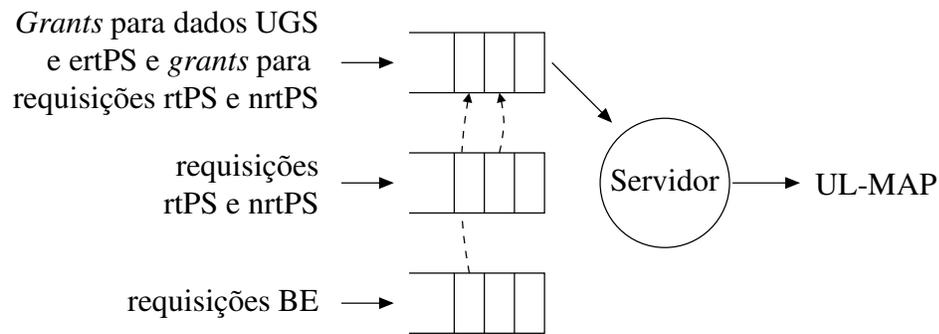


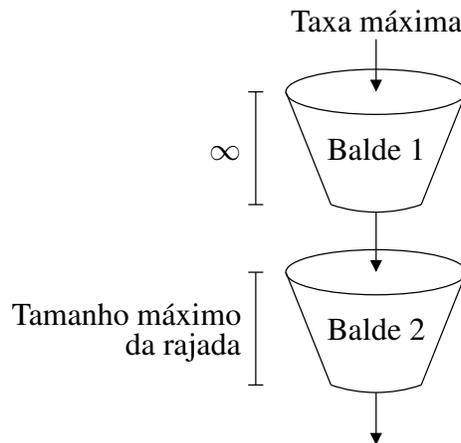
Figura 4.2: Arquitetura do escalonador proposto

a latência máxima para o tráfego que não excede a taxa mínima, conforme especificado pelo padrão. Para determinar o *deadline* de uma requisição, é necessário saber o instante de chegada dos pacotes na fila da SS. Dado que a BS não tem acesso a essa informação, considera-se que o pacote que gerou a requisição chegou na fila da SS imediatamente após o envio da última requisição de banda. Assim sendo, o *deadline* de uma requisição é igual ao instante de chegada da última requisição enviada pela conexão correspondente mais o requisito de latência máxima da conexão.

O escalonador garante o requisito de taxa mínima das conexões rtPS e nrtPS em janelas de duração T . Sempre que o escalonador é executado, calcula-se um valor de prioridade para cada requisição de banda na fila intermediária, considerando-se as seguintes informações da conexão: i) requisito de taxa mínima, ii) quantidade de bytes requisitados pela conexão e ainda não atendidos e iii) taxa recebida na janela T atual. Requisições enviadas por conexões que já atingiram a taxa mínima na janela atual recebem valores baixos de prioridade. Para o restante das requisições, quanto menor a taxa atingida pela conexão, maior é a prioridade de suas requisições.

Todos os *slots* alocados para uma SS em um *frame* devem ser combinados em um único *grant* na mensagem UL-MAP. Assim sendo, o escalonador deve garantir que o somatório da banda alocada para uma única conexão seja menor ou igual ao tamanho máximo da rajada (*maximum traffic burst*). O escalonador, adicionalmente, não deve alocar banda para uma conexão caso resulte em violação do requisito de taxa máxima (*maximum sustained traffic rate*). O policiamento de ambos os parâmetros no escalonador proposto é feito por um regulador do tipo duplo balde furado (*dual leaky bucket*), apresentado na Figura 4.3. O primeiro balde (*bucket1*) tem capacidade infinita e taxa de geração de fichas igual ao parâmetro *maximum sustained traffic rate*. O segundo balde (*bucket2*) tem capacidade igual ao parâmetro *maximum traffic burst* e, a cada *frame*, é preenchido com fichas provenientes do primeiro balde. Dessa forma, antes de alocar um *grant*, o escalonador verifica se há fichas suficientes no *bucket2*.

O Algoritmo *Scheduling* apresenta o mecanismo de escalonamento proposto. Depois de

Figura 4.3: Regulador *dual leaky bucket*

ALGORITMO Scheduling

1. insere na fila de alta prioridade requisições para os *grants* periódicos que devem ser alocados no próximo *frame*
 2. *CheckDeadline*
 3. *CheckMinimumRate*
 4. *DistributeFreeResources*
 5. escalona as requisições da fila de alta prioridade começando pela cabeça da fila
-

inserir os *grants* periódicos na fila de alta prioridade, o algoritmo verifica quais requisições *rtPS* e *nrtPS* devem migrar da fila intermediária para a fila de alta prioridade (linhas 2 e 3). Na linha 4, o escalonador distribui os *slots* não utilizados pelas conexões de mais alta prioridade entre as conexões *BE* através da migração de requisições da fila de baixa prioridade para a fila de alta prioridade. No último passo, o escalonador serve todas as requisições armazenadas na fila de alta prioridade.

No procedimento *CheckDeadline*, se há banda disponível, o escalonador usa o procedimento *MigrateBWRequest* para migrar da fila intermediária para a fila de alta prioridade as requisições *rtPS* que satisfazem as seguintes condições: o *deadline* da requisição expira no *frame* que segue o próximo *frame* e a taxa alcançada pela conexão na janela atual ($TR[CID]$, onde CID é o identificador da conexão) é menor do que a taxa mínima requisitada ($minTR[CID]$).

O procedimento *MigrateBWRequest* verifica se a quantidade de *bytes* solicitada pela requisição i sendo migrada ($BR[i]$) está disponível no *subframe uplink*. Caso a quantidade de banda disponível seja menor do que o requisitado, apenas parte da requisição será atendida nesta rodada de escalonamento. O atendimento de parte da requisição também será adiado caso a quantidade de *bytes* requisitados seja maior do que o tamanho do *bucket2*. Uma vez que a quantidade de banda a ser alocada (*grantSize*) é definida, se *grantSize* é menor do que $BR[i]$,

PROCEDIMENTO *CheckDeadline*

1. **para** cada requisição i na fila intermediária **faça**
 2. **se** $availableBw == 0$ **então**
 3. **break**
 4. **se** $service[CID] == rtPS$ **então**
 5. $frame \leftarrow \lfloor (deadline[i] - currentTime) \div frameDuration \rfloor$
 6. **se** $frame == 3$ e $TR[CID] < minTR[CID]$ **então**
 7. **MigrateBWRequest**(i)
-

PROCEDIMENTO *MigrateBWRequest*(i)

1. **se** $BR[i] > availableBw$ **então**
 2. $grantSize \leftarrow availableBw$
 3. **senão**
 4. $grantSize \leftarrow BR[i]$
 5. **se** $grantSize > bucket2[CID]$ **então**
 6. $grantSize \leftarrow bucket2[CID]$
 7. **se** $0 < grantSize < BR[i]$ **então**
 8. crie uma nova requisição j para a conexão CID
 9. $BR[j] \leftarrow BR[i] - grantSize$
 10. insira a requisição j no final da fila intermediária
 11. $BR[i] \leftarrow grantSize$
 12. mova a requisição i para a fila de alta prioridade
 13. $TR[CID] \leftarrow TR[CID] + grantSize$
 14. $bucket2[CID] \leftarrow bucket2[CID] - grantSize$
 15. $backlogged[CID] \leftarrow backlogged[CID] - grantSize$
 16. $availableBw \leftarrow availableBw - grantSize$
-

uma nova requisição j é criada com $BR[j]$ igual a $BR[i] - grantSize$ e inserida no final da fila intermediária. A quantidade de banda solicitada pela requisição i é atualizada para $grantSize$ e i migra para a fila de alta prioridade. Finalmente, o procedimento atualiza: i) a taxa de tráfego recebida pela conexão na janela T ($TR[CID]$), ii) a quantidade de bytes requisitados pela conexão e ainda não atendidos ($backlogged[CID]$), iii) o tamanho do $bucket2$ da conexão ($bucket2[CID]$) e iv) o número de bytes disponíveis no *subframe uplink* ($availableBW$).

O procedimento *CheckMinimumRate* calcula um valor de prioridade para cada requisição na fila intermediária (linhas 6-9). Em seguida, ele ordena a fila intermediária de acordo com os valores de prioridade atribuídos para as requisições (linha 13). No último passo, o escalonador tenta migrar as requisições para a fila de alta prioridade utilizando o procedimento *MigrateBWRequest* seguindo a ordem de prioridade estabelecida.

O procedimento *DistributeFreeResources* distribui os *slots* não utilizados pelos serviços de mais alta prioridade para as conexões BE migrando as requisições da fila de baixa prioridade

pelos Ss e nos requisitos de QoS de cada conexão. Cada conexão está associada a um dos cinco tipos de serviço disponíveis no padrão, sendo que cada serviço é caracterizado por um conjunto de parâmetros de QoS. Muitos dos mecanismos de escalonamento propostos na literatura não consideram todos os possíveis parâmetros de QoS. Na realidade, existem poucas propostas que se preocupam com a complexidade do mecanismo, ainda que esse seja um aspecto essencial, dada a frequência com que o escalonador deve ser executado.

Este capítulo introduziu um novo escalonador *uplink* para a BS. A proposta apresentada segue as especificações do padrão IEEE 802.16 e fornece uma solução de implementação simples baseada em três filas de prioridade. Garantias de latência máxima e taxa mínima são fornecidas respeitando os parâmetros de QoS de cada conexão sem que os limites impostos pelos requisitos de taxa máxima e de tamanho máximo da rajada sejam violados. Uma análise da eficácia e do desempenho deste escalonador será apresentada no Capítulo 6.

Capítulo 5

Controle de Admissão em redes IEEE 802.16

Enquanto o escalonamento garante que a quantidade de recursos requisitada seja alocada para as conexões, de modo que os requisitos de QoS sejam atendidos, o mecanismo de controle de admissão limita o número de conexões para que a rede não seja sobrecarregada por um número excessivo de usuários. Sempre que um usuário deseja estabelecer uma nova conexão, uma requisição é enviada para a BS para que o mecanismo de controle de admissão decida se a nova conexão pode ou não ser aceita. Para tomar essa decisão, o controle de admissão deve verificar se existem recursos suficientes para satisfazer os requisitos de QoS da nova conexão sem comprometer o atendimento dos requisitos mínimos de QoS das conexões em andamento.

A escolha da política de controle de admissão a ser adotada em uma rede IEEE 802.16 está fortemente associada ao mecanismo de escalonamento utilizado. Por exemplo, ao se adotar um mecanismo de controle de admissão que estima os recursos disponíveis a partir da diferença entre a capacidade total do enlace e a soma dos requisitos de taxa mínima das conexões já admitidas, é preciso garantir-se de que o escalonador não alocará mais do que a taxa mínima para uma conexão, quando outras conexões ainda não tiveram seus requisitos mínimos atendidos. Além disso, a integração dos mecanismos de escalonamento e controle de admissão pode resultar em soluções mais simples, pois quando um dos mecanismos é capaz de garantir o atendimento à um requisito de QoS a mesma garantia não precisa ser implementada pelo outro mecanismo.

O padrão IEEE 802.16 não define políticas de controle de admissão, o que tem incentivado pesquisadores do meio acadêmico e da indústria a investigarem soluções para esse problema. A Seção 5.1 discorre sobre algumas das principais propostas existentes na literatura, enquanto as Seções 5.2, 5.3, 5.4, 5.5 e 5.6 introduzem novas soluções para o problema.

5.1 Estado da arte da pesquisa em controle de admissão em redes IEEE 802.16

Chen, Jiao e Wang [27] propõem um algoritmo de controle de admissão que utiliza o requisito de taxa mínima para verificar se uma nova conexão pode ser aceita sem comprometer o atendimento aos requisitos mínimos de QoS do tráfego presente na rede. O algoritmo faz o somatório da taxa mínima requisitada por todas as conexões já admitidas e subtrai este valor da capacidade total da rede, obtendo, dessa forma, a capacidade disponível (C_a). Quando uma nova conexão chega ou uma conexão já ativa solicita alterações nos requisitos de QoS, o algoritmo verifica se a condição $C_a \geq 0$ é satisfeita. Em caso afirmativo, a solicitação é aceita. Entretanto, é importante lembrar que conexões rtPS e nrtPS necessitam de banda não só para a transmissão de dados, mas também para o envio de requisições de banda, logo, considerar apenas o requisito de taxa mínima para calcular a capacidade disponível pode resultar em uma estimativa incorreta.

Em [35], os autores propõem um algoritmo de controle de admissão que utiliza os requisitos de taxa mínima e taxa máxima no processo de decisão. Conexões UGS são aceitas sempre que o requisito de taxa máxima for menor ou igual a capacidade disponível. Conexões rtPS, ertPS e nrtPS são aceitas quando o requisito de taxa mínima pode ser atendido e conexões BE sempre são aceitas. Quando uma conexão é admitida, a BS reserva uma quantidade de banda para a conexão. Para uma conexão UGS a banda reservada é igual ao requisito de taxa máxima e para conexões rtPS, ertPS ou nrtPS a banda reservada é igual ao mínimo entre a banda disponível e o requisito de taxa máxima sustentada. O mecanismo de controle de admissão também faz uma reserva de banda para conexões provenientes de outras células (*handoff*) com o objetivo de reduzir a taxa de bloqueio de conexões em andamento. Além disso, o mecanismo utiliza um esquema de empréstimo de banda, no qual conexões de baixa prioridade emprestam banda para a admissão de conexões de alta prioridade. Os autores assumem que a capacidade do canal é C bits por segundo, a quantidade de banda disponível é bl bits por segundo, e uma porcentagem n da capacidade do canal é reservada como canal de guarda para conexões em *handoff*. Quando uma nova conexão chega com requisito de banda br , se $(bl - br) > C.n\%$, a conexão é admitida. Para conexões em *handoff*, se $(bl - br) > 0$, admite-se a conexão. Os autores não determinam como as variáveis C , bl e n devem ser estimadas.

Wang, He e Agrawal [24] propõem um mecanismo de controle de admissão que dá maior prioridade para conexões UGS e utiliza um esquema de empréstimo e degradação de taxa com o objetivo de maximizar a utilização do canal. Neste mecanismo, conexões UGS sempre são aceitas quando há capacidade disponível, ou seja, quando a capacidade utilizada somada à taxa requisitada pela conexão é menor ou igual à capacidade disponível (B). Uma conexão rtPS é aceita quando a capacidade utilizada somada à taxa requisitada pela conexão é menor ou igual a $B - U$, onde U é a capacidade reservada exclusivamente para conexões UGS. Se a condição não for satisfeita, o algoritmo degrada a taxa das conexões nrtPS até o valor mínimo requisitado

e verifica a condição de admissão novamente. Quando uma conexão nrtPS requisita admissão, o algoritmo verifica se a capacidade utilizada somada à taxa requisitada pela conexão menos um coeficiente de degradação é menor ou igual a $B - U$. Se a condição não for satisfeita, o algoritmo tenta degradar a taxa das outras conexões nrtPS até o valor mínimo requisitado. Conexões BE são sempre aceitas. A reserva de uma parte da capacidade para serviços UGS é uma desvantagem neste algoritmo. Os autores justificam a adoção dessa estratégia baseados na idéia de que, do ponto de vista do usuário, bloquear um novo fluxo UGS, que normalmente serve tráfego de voz, traz mais problemas do que bloquear um fluxo não-UGS. Entretanto, reservar parte da capacidade para um único tipo de serviço pode causar subutilização da rede, quando a taxa requisitada pelos fluxos UGS é menor do que a capacidade reservada para este serviço.

Niyato e Hossain [18] propõem um mecanismo de alocação de banda e controle de admissão baseado em análise de filas e teoria dos jogos. A modelagem das filas é usada para calcular a vazão dos serviços nrtPS e BE e a latência dos serviços de tempo real, para serem usados pelo modelo de teoria dos jogos na admissão de novas conexões. A formulação do jogo baseia-se em dois fatos: i) a BS pode admitir uma nova conexão somente se o desempenho das conexões já admitidas não for degradado para níveis abaixo do desejado e ii) a nova conexão aceita os serviços oferecidos pela BS apenas se os requisitos de latência e taxa puderem ser garantidos. Quando uma nova conexão solicita admissão, ela informa qual o tipo de serviço que deseja, os parâmetros do tráfego e os requisitos de QoS. A BS estabelece um conjunto de estratégias e calcula o *payoff* esperado para cada estratégia. Em seguida, o jogo é resolvido para obter o equilíbrio de Nash, que é usado para decidir se a nova conexão será ou não aceita. Caso a conexão seja aceita, a quantidade de banda a ser reservada para a nova conexão é determinada.

No mecanismo de controle de admissão proposto por Chandra e Sahoo [49], as requisições para o estabelecimento de novas conexões são classificadas em filas, de acordo com o tipo de serviço requisitado. O mecanismo atende primeiro a fila UGS, seguida, respectivamente, pela fila rtPS e pela fila nrtPS. Conexões BE não passam pelo controle de admissão dado que não necessitam de nenhum tipo de garantia de QoS. A cada intervalo de escalonamento, o mecanismo de controle de admissão percorre as filas de requisições para novas conexões e decide quais conexões serão aceitas. Se a requisição é do tipo UGS, a seguinte condição deve ser satisfeita: o número de *slots* requisitados, de acordo com os requisitos de taxa máxima e intervalo entre grants, deve ser menor ou igual ao número total de *slots* que podem ser acomodados de acordo com o *jitter* máximo tolerado pela conexão. Se esta condição é satisfeita, o mecanismo ainda verifica se o número de *slots* requisitados pode ser atendido dentro de um intervalo chamado *HyperInterval*. Para conexões UGS, o valor de *HyperInterval* é igual ao mínimo múltiplo comum do parâmetro *unsolicited grant interval* de todas as conexões, enquanto que para conexões rtPS e nrtPS, o valor é igual ao mínimo múltiplo comum do parâmetro *unsolicited polling interval* de todas as conexões. Nessa etapa do algoritmo, o mecanismo de controle de admissão utiliza a rotina *search(no_of_slots, initial_slot, final_slot)*, que busca o número de *slots* requi-

sitados (*no_of_slots*) dentro do intervalo [*initial_slot*, *final_slot*]. Se a requisição para uma nova conexão é do tipo rtPS, o número de *slots* requisitados dentro de um intervalo de *polling*, de acordo com o requisito de taxa mínima, deve ser menor ou igual ao número de *slots* que podem ser acomodados pela capacidade disponível dentro deste mesmo intervalo. Em seguida, o mecanismo verifica a disponibilidade de *slots*, tanto para envio de requisição de banda como para envio de dados, em cada intervalo de *polling* dentro do *HyperInterval*. A admissão de conexões nrtPS é semelhante a admissão de conexões rtPS. Embora essa solução seja interessante, dado que considera a maioria dos parâmetros de QoS definidos no padrão bem como a taxa necessária para o envio de requisições banda, os autores não detalham o funcionamento da rotina *search*, que tem papel essencial no mecanismo.

O mecanismo de controle de admissão proposto por Chang, Chen e Chou [11] utiliza dois níveis de prioridade: um para as SSs e outro para os serviços UGS, rtPS, nrtPS e BE. Os níveis de prioridade são usados para estabelecer o valor do parâmetro $RW_{r,k}$ que define a recompensa da rede ao admitir uma conexão do serviço k para um nó com prioridade r . O objetivo do mecanismo de controle de admissão é aumentar a recompensa da rede, porém, os autores não informam como configurar o valor do parâmetro $RW_{r,k}$. O mecanismo de controle de admissão utiliza, também, uma função de custo chamada COL (*Competitive On-Line*). Quando a BS recebe uma requisição para o estabelecimento de uma conexão, o mecanismo calcula um valor de custo baseado no tipo do serviço, na taxa mínima requisitada, na capacidade do enlace e em uma constante pré-definida. A conexão será aceita se a recompensa obtida for maior do que o custo.

Guo *et alli* [54] observam dois problemas principais no projeto de mecanismos de controle de admissão e de reserva de recursos. Um deles é reservar recursos para o tráfego de tempo real de forma que se faça um uso eficiente da largura de banda. O segundo problema consiste em reservar recursos para conexões provenientes de outras células (*handoff*), de forma a reduzir a probabilidade de bloqueio destas conexões. Para tratar esses dois problemas, os autores propõem o mecanismo *Dynamic Bandwidth Reservation Admission Control* (DBRAC). Para atender o serviço rtPS, o mecanismo reserva uma quantidade de banda igual a soma de $MinTR$ com um valor dentro da faixa $[0, MaxTR - MinTR]$, onde $MaxTR$ e $MinTR$ são, respectivamente, os requisitos de taxa máxima e taxa mínima. A soma dos recursos extra (além de $MinTR$) reservados para as conexões rtPS é denominada Rrt_vbr . Para conexões em *handoff* o mecanismo reserva uma quantidade de banda Rhf calculada por uma cadeia de Markov unidimensional, na qual a variável de estado representa o número de conexões em *handoff*. O mecanismo utiliza ainda uma terceira variável, R , que representa o valor máximo entre Rrt_vbr e Rhf . O mecanismo DBRAC faz a decisão de admissão da seguinte forma: se a banda disponível (capacidade da célula menos a soma dos requisitos de taxa mínima das conexões já admitidas) é maior do que a taxa mínima requisitada pela conexão somada a R , então, a conexão é aceita, caso contrário, ela é rejeitada. Conexões em *handoff* são rejeitadas apenas quando

a taxa mínima requisitada não pode ser atendida.

Rong *et alli* [12] propõem uma solução ótima bem como uma solução aproximada para maximizar o lucro, a utilidade e a justiça na rede. A maximização da utilidade consiste em maximizar a quantidade de banda disponibilizada para cada usuário, enquanto a maximização da justiça consiste em manter taxas de bloqueio uniformes para os diferentes tipos de serviço. Os autores dividem o controle de admissão em N módulos de forma que o k -ésimo módulo faz a admissão das conexões do k -ésimo usuário. A solução de controle de admissão apresentada aplica-se a cada módulo individualmente. O algoritmo proposto é executado em duas fases. A primeira fase consiste em dividir a banda disponível para o k -ésimo usuário em M partes utilizando-se a técnica de partição completa (CP - *Complete Partition*), onde M é o número de classes de serviço que compartilham a banda. Essa divisão é feita de forma que a restrição de justiça seja respeitada. Na segunda fase, a banda restante é distribuída de forma a maximizar a utilidade e o lucro. Para atender a restrição de utilidade, apenas as classes de serviço que oferecem utilidade maior do que U^{th} (*utility threshold*) são escolhidas como candidatas para receber recursos. Dentre as classes candidatas, o algoritmo seleciona aquelas que oferecem maior lucro.

Em [45], os autores usam critérios diferentes para a admissão de conexões sensíveis à latência (UGS e rtPS) e conexões nrtPS. Para conexões UGS e rtPS, o algoritmo baseia-se na taxa requisitada pela conexão para calcular quantos *slots* devem ser alocados em cada *frame*. Se a quantidade de *slots* calculada estiver disponível, a conexão é aceita. Para cada conexão aceita, serão reservados *slots* em cada *subframe uplink* durante todo o período em que a conexão estiver ativa para que se garanta a latência máxima. Para conexões nrtPS, o algoritmo verifica se a taxa requisitada pela nova conexão agregada à taxa já reservada para conexões nrtPS pode ser servida em um espaço de tempo maior do que um *frame*, dado que essas conexões não têm requisito de latência máxima. Conexões BE não passam pelo controle de admissão, entretanto, o mecanismo de controle de admissão reserva parte da banda para este tipo de serviço.

Dois dos esquemas discutidos nesta seção apresentam características únicas. O esquema proposto por Chandra e Sahoo [49] leva em consideração não só a taxa mínima requisitada pela nova conexão, mas também a taxa utilizada pelo mecanismo de requisição de banda associado ao tipo de serviço solicitado. Para conexões UGS, o esquema proposto garante que o número de *slots* requisitados está disponível dentro do *jitter* tolerado para cada intervalo entre *grants*. Estas características mostram a preocupação dos autores em assegurar que o mecanismo proposto atenda às especificações do padrão IEEE 802.16. No esquema proposto por Rong *et alli* [12], um dos objetivos é maximizar o lucro na rede. Dado que o WiMAX é uma rede multi-serviço, espera-se que usuários de serviços distintos paguem preços diferentes. Do ponto de vista do provedor de serviços, é importante maximizar o lucro ao mesmo tempo em que se garante Qualidade de Serviço para o usuário. Um mecanismo que integre controle de admissão e contabilidade de lucro, portanto, é bastante promissor. A combinação das principais caracte-

rísticas dos dois esquemas mencionados em uma política de controle de admissão resulta em um mecanismo que atende três importantes expectativas tanto por parte dos usuários quanto por parte dos provedores de serviço: Qualidade de Serviço, lucro e compatibilidade com o padrão IEEE 802.16.

Embora as conexões de tempo real tenham requisito de latência máxima, a maioria das propostas discutidas usam apenas o requisito de taxa mínima no processo de decisão. Essa abordagem não é inválida se considerarmos que, de acordo com o padrão IEEE 802.16, a latência máxima precisa ser garantida apenas para o tráfego que não ultrapassa o requisito de taxa mínima da conexão. Com um controle de admissão que garante capacidade suficiente para satisfazer a taxa mínima de cada conexão, o escalonador pode se encarregar de garantir a latência máxima conforme especificado no padrão. Algoritmos de controle de admissão que incluem garantias de latência para todo o tráfego de tempo real tendem a ser mais complexos, como a proposta apresentada em [18]. A Tabela 5.1 apresenta um resumo das vantagens e desvantagens das soluções apresentadas nesta seção bem como das soluções propostas nesta tese.

5.2 Controle de Admissão baseado na Taxa Mínima

A política de admissão proposta nesta seção baseia-se na garantia do requisito de taxa mínima para decidir quais conexões devem ser aceitas. Uma vez que esse requisito possa ser garantido para a nova conexão bem como para as conexões previamente admitidas, a conexão é aceita. Uma nova conexão é aceita sempre que a seguinte condição é satisfeita:

$$(C_{reserved} + TR_{ij} \leq C) \quad (5.1)$$

TR_{ij} representa a taxa que deve ser garantida para a nova conexão j associada ao tipo de serviço i ; $C_{reserved}$ representa a capacidade reservada para as conexões já admitidas na rede e é igual a $\sum_{i=1}^4 \sum_{j=1}^n TR_{ij}$ ¹; e C é a capacidade disponível para o escalonador, ou seja, a quantidade de banda que o escalonador pode alocar para transmissão de dados e requisição de banda através de *grants unicast*.

Embora esta seja uma condição de admissão bastante simples, ela atende às especificações do padrão IEEE 802.16 com a vantagem de ser facilmente implementada na estação base. A condição proposta garante recursos suficientes para o escalonador atender o requisito de taxa mínima das conexões UGS², rtPS, ertPS e nrtPS bem como a latência máxima das conexões de tempo real cujo tráfego não ultrapassa essa taxa.

¹ i varia de 1 à 4, pois apenas os serviços UGS (1), rtPS (2), ertPS (3) e nrtPS (4) passam pelo controle de admissão.

²No caso do serviço UGS, o padrão estabelece que se deve garantir o requisito de taxa máxima e que o requisito de taxa mínima, quando definido, deve ser igual ao valor da taxa máxima. Dessa forma, sempre que nos referirmos à taxa mínima do serviço UGS, estamos supondo que ela é igual à taxa máxima.

A principal diferença entre o esquema proposto e outros esquemas de controle de admissão propostos para o padrão IEEE 802.16 está no cálculo do parâmetro TR_{ij} utilizado na equação 5.1. Ao enviar um pedido de admissão, a conexão fornece, entre outros parâmetros, o requisito de taxa mínima ($minTR_{ij}$), entretanto, se considerarmos a sobrecarga gerada pelo tipo de serviço solicitado (por exemplo, com o envio de requisições de banda), essa taxa não será suficiente para atender as necessidades da conexão. Assim sendo, ao invés de considerar apenas a taxa mínima, o parâmetro TR_{ij} inclui também uma estimativa da sobrecarga gerada pelo tipo de serviço associado à conexão.

Conexões UGS e ertPS recebem *grants* em intervalos definidos pelo parâmetro *unsolicited grant interval* (ugi). O tamanho do *grant* para conexões UGS é fixo e definido de acordo com o requisito de taxa máxima da conexão (que é igual ao requisito de taxa mínima para esse serviço). Para conexões ertPS, embora o tamanho dos *grants* varie de acordo com a quantidade de banda requisitada, o tamanho mínimo a ser garantido deve basear-se no requisito de taxa mínima. Desse modo, para ambos os tipos de serviço, a BS deve garantir a alocação de *grants* periódicos com tamanho igual ao valor fornecido pela seguinte equação:

$$grantSize_{ij} = minTR_{ij} * ugi_{ij} \quad (5.2)$$

A equação 5.2 fornece o tamanho dos *grants* em bytes, entretanto, na mensagem UL-MAP os *grants* são descritos em número de *slots*, obtido a partir do cálculo do menor inteiro maior ou igual a $grantSize_{ij}/slotSize$. Esta normalização pode resultar em desperdício de banda quando $grantSize_{ij}$ não é um múltiplo do número de bytes que pode ser transmitido em um *slot* ($slotSize$). A taxa real utilizada por uma conexão UGS (ertPS) pode, conseqüentemente, ser maior do que a taxa máxima (taxa mínima) requisitada. Para contabilizar esta sobrecarga, o valor do parâmetro TR_{ij} para conexões UGS e ertPS é calculado da seguinte forma:

$$TR_{ij} = \frac{\lceil \frac{grantSize_{ij}}{slotSize} \rceil * slotSize}{ugi_{ij} + tj_{ij}} \quad (5.3)$$

Na equação 5.3, $\lceil \frac{grantSize_{ij}}{slotSize} \rceil$ fornece o tamanho de um *grant* em número de *slots*. Os *grants* devem ser fornecidos em intervalos de duração ugi_{ij} com *jitter* máximo tj_{ij} (definido pelo parâmetro *tolerated jitter*). Multiplicando-se o número de *slots* em um *grant* pelo tamanho de um *slot* em número de bytes e dividindo este valor pela soma $ugi_{ij} + tj_{ij}$, obtém-se a taxa mínima que deve ser garantida para conexões UGS e ertPS.

Além da taxa mínima, conexões rtPS e nrtPS precisam de *grants* periódicos para o envio de requisição de banda. Assim sendo, a taxa que deve ser garantida para estas conexões é igual a soma do requisito de taxa mínima ($minTR_{ij}$) com a taxa usada pelo mecanismo de requisição de banda, conforme apresentado na equação 5.4.

$$TR_{ij} = \min TR_{ij} + \frac{upSlots * slotSize}{upi_{ij}} \quad (5.4)$$

onde $upSlots$ representa o número de *slots* usados para o envio de uma mensagem de requisição de banda e upi_{ij} é o valor fornecido pelo parâmetro *unsolicited polling interval*. Conexões nrtPS não definem o parâmetro upi_{ij} , entretanto, o padrão IEEE 802.16 estabelece que este intervalo deve ser da ordem de 1 segundo e pode ser definido pela BS.

Conexões BE são sempre aceitas dado que este serviço não tem requisitos de QoS.

5.3 Controle de Admissão Ótimo para Maximização do Lucro

A segunda política de controle de admissão proposta nesta tese estende a política apresentada na Seção 5.2 com o objetivo de maximizar o lucro obtido pelo provedor.

Da mesma forma que a política anterior, antes de admitir uma nova conexão, o controle de admissão verifica se é possível satisfazer o requisito de banda mínima desta nova conexão sem prejudicar as conexões já admitidas na rede. A sobrecarga gerada por cada tipo de serviço é igualmente considerada.

Para maximizar o lucro, o controlador de admissão deve coletar requisições durante intervalos de tempo, para, então, decidir quais conexões devem ser aceitas. Uma política de admissão ótima deve encontrar uma combinação de conexões que, a longo prazo, forneça o maior lucro possível respeitando a capacidade da rede. Esse problema pode ser mapeado para o problema clássico da mochila, no qual a capacidade da rede é a mochila e a quantidade de banda requisitada por cada conexão representa o tamanho de cada objeto. A modelagem deste problema de otimização é apresentada a seguir.

$$\max \sum_{i=1}^4 \sum_{j=m_i+1}^{n_i} x_{ij} r_i \min TR_{ij}$$

Sujeito a:

$$x_{ij} \in \{0, 1\} \quad \forall i \in \{1, 2, 3, 4\} \quad \forall j \in \{m_i + 1, \dots, n_i\} \quad (1)$$

$$\sum_{i=1}^4 \sum_{j=m_i+1}^{n_i} TR_{ij} \leq C_{available} \quad (2)$$

onde:

- n_i : número total de conexões associadas ao serviço i (conexões admitidas + conexões solicitantes);

- m_i : número de conexões admitidas e ainda ativas na rede associadas ao serviço i ;
- $\min TR_{ij}$: requisito de taxa mínima da conexão j associada ao serviço i ;
- TR_{ij} : taxa que deve ser reservada para a conexão j associada ao serviço i e que é calculada através da Equação 5.3 para os serviços UGS e ertPS e através da Equação 5.4 para os serviços rtPS e nrtPS;
- r_i : lucro associado ao serviço i ;
- x_{ij} : variável que receberá valor 1 caso a conexão j associada ao serviço i deva ser admitida e valor 0 caso contrário;
- $C_{available}$: $C - \sum_{i=1}^4 \sum_{j=1}^{m_i} TR_{ij}$, onde C é a capacidade de que o escalonador dispõe para alocação de *grants* para transmissão de dados e requisição de banda.

Na formulação do problema, i varia entre 1 e 4, pois apenas conexões UGS, rtPS, ertPS e nrtPS estão sujeitos ao controle de admissão. A função de lucro $r_i * \min TR_{ij}$ indica que quanto maior o valor da taxa mínima requisitada pela conexão, maior é o preço que o usuário deverá pagar e, conseqüentemente, maior o lucro obtido com sua admissão. Esta função garante maior justiça no preço pago por usuários de um mesmo tipo de serviço, dado que o valor cobrado é proporcional a taxa mínima requisitada. A restrição 2 garante que a soma das taxas que devem ser alocadas (TR_{ij}) para todas as conexões aceitas é menor do que a capacidade disponível.

5.4 Controle de Admissão Guloso para Maximização do Lucro

A terceira política de controle de admissão utiliza uma técnica gulosa para a maximização do lucro. Assim como no controle de admissão ótimo apresentado na Seção 5.3, neste mecanismo o controlador de admissão realiza a admissão de novas conexões em intervalos de tempo pré-definidos. As conexões são admitidas seguindo a ordem decrescente dos valores de lucro associados as conexões e respeitando a capacidade da rede em atender o requisito de taxa mínima de cada conexão.

O Algoritmo *GreedyMaxRevenueCAC* apresenta os passos realizados a cada intervalo de admissão para decidir quais conexões serão aceitas. Durante um intervalo de admissão, as requisições para o estabelecimento de novas conexões são armazenadas em um *buffer*. No primeiro passo, o algoritmo calcula um valor de lucro para cada nova conexão utilizando a função de lucro $r_i * \min TR_{ij}$. Em seguida, as requisições para o estabelecimento de novas conexões são ordenadas seguindo a ordem decrescente dos valores de lucro calculados. Nos

passos seguintes, o algoritmo utiliza a equação 5.1 para decidir quais conexões serão aceitas, começando pelo primeiro elemento do *buffer* ordenado.

ALGORITMO GREEDYMAXREVENUECAC

1. **para** cada conexão j armazenada no *buffer* **faça**
 2. $revenue_{ij} \leftarrow r_i * \min TR_{ij}$
 3. ordene o *buffer* seguindo a ordem decrescente dos valores de $revenue_{ij}$
 4. **para** cada conexão j armazenada no *buffer* **faça**
 5. **se** tipo de serviço de j for UGS ou ertPS **então**
 6. calcule TR_{ij} usando a equação 5.3
 7. **senão**
 8. calcule TR_{ij} usando a equação 5.4
 9. **se** $(C_{reserved} + TR_{ij} \leq C)$ **então**
 10. aceite a conexão j
 11. $C_{reserved} \leftarrow C_{reserved} + TR_{ij}$
 12. **senão**
 13. rejeite a conexão j
-

5.5 Controle de Admissão Ótimo para Maximização do Lucro e da Utilidade

O quarto mecanismo de controle de admissão proposto nesta tese tem como objetivos maximizar o lucro e a utilidade da rede. Sob o ponto de vista do usuário, a melhor política de admissão é aquela que lhe permite atingir as maiores taxas de transmissão. Dessa forma, nesta política, além do valor de lucro r_i , cada serviço i está associado a um valor u_i que reflete um ganho de utilidade ou, em outras palavras, a satisfação do cliente. O valor u_i é multiplicado por $extraTR_{ij}$ que indica o quanto de taxa extra ao requisito de taxa mínima, será alocado para a conexão j . $extraTR_{ij}$ deve ser menor ou igual a diferença entre os requisitos de taxa máxima e taxa mínima da conexão para que o requisito de taxa máxima não seja violado. Assim sendo, quanto maior o somatório dos valores de $u_i * extraTR_{ij}$ de todas as conexões admitidas, maior será o ganho de utilidade e, conseqüentemente, maior será a satisfação dos usuários. A formulação deste problema de maximização é apresentada a seguir.

$$max \sum_{i=1}^4 \sum_{j=m_i+1}^{n_i} x_{ij} r_i \min TR_{ij} + \sum_{i=1}^4 \sum_{j=1}^{n_i} u_i extraTR_{ij}$$

Sujeito a:

$$x_{ij} \in \{0, 1\} \quad \forall i \in \{1, 2, 3, 4\} \quad \forall j \in \{m_i + 1, \dots, n_i\} \quad (1)$$

$$0 \leq extraTR_{ij} \leq maxTR_{ij} - minTR_{ij} \quad \forall i \in \{1, 2, 3, 4\} \text{ e } \forall j \in \{1, \dots, m_i\} \quad (2)$$

$$0 \leq extraTR_{ij} \leq (maxTR_{ij} - minTR_{ij})x_{ij} \quad \forall i \in \{1, 2, 3, 4\} \text{ e } \forall j \in \{m_i + 1, \dots, n_i\} \quad (3)$$

$$\sum_{i=1}^4 \sum_{j=1}^{n_i} (TR_{ij} + extraTR_{ij}) \leq C \quad (4)$$

onde

- n_i : número total de conexões associadas ao serviço i (conexões admitidas + conexões solicitantes);
- m_i : número de conexões admitidas e ainda ativas associadas ao serviço i ;
- $minTR_{ij}$: requisito de taxa mínima da conexão j associada ao serviço i ;
- $maxTR_{ij}$: requisito de taxa máxima da conexão j associada ao serviço i ;
- TR_{ij} : taxa mínima que deve ser reservada para a conexão j associada ao serviço i e que é calculada através da Equação 5.3 para os serviços UGS e ertPS e através da Equação 5.4 para os serviços rtPS e nrtPS;
- $extraTR_{ij}$: variável que indicará a taxa extra, além da mínima, que será alocada para a conexão j associada ao serviço i ;
- r_i : lucro associado ao serviço i ;
- u_i : ganho em utilidade associado ao serviço i ;
- x_{ij} : variável que receberá valor 1 caso a conexão j associada ao serviço i deva ser admitida e valor 0 caso contrário;
- C : capacidade de que o escalonador dispõe para alocação de *grants* para transmissão de dados e requisição de banda.

Na formulação apresentada, a restrição 2 garante que o valor de taxa extra ($extraTR_{ij}$) que será alocada para as conexões previamente aceitas na rede não viola o requisito de taxa máxima. Esse valor será atualizado a cada rodada de admissão, de forma a se adaptar às mudanças causadas pela entrada e saída de conexões da rede. A restrição 3 garante não só que o valor de taxa extra alocado para as novas conexões aceitas não viola o requisito de taxa máxima, mas também que o valor de $extraTR_{ij}$ seja igual a zero para conexões rejeitadas. É importante observar que o algoritmo nunca aloca taxa extra para conexões UGS dado que esse tipo de serviço tem mesmo valor para os requisitos de taxa mínima e taxa máxima. A restrição 4 garante que o somatório das taxas que devem ser reservadas para cada conexão (TR_{ij}) e das taxas extras

alocadas ($extraTR_{ij}$) é menor ou igual a capacidade de que o escalonador dispõe para alocação de *grants*.

5.6 Controle de Admissão Heurístico para Maximização do Lucro e da Utilidade

O mecanismo de controle de admissão proposto nesta seção utiliza uma heurística para maximizar o lucro e a utilidade da rede. Assim como na solução proposta na Seção 5.5, este mecanismo decide quais conexões poderão entrar na rede bem como a quantidade de taxa extra que será alocada para cada conexão.

Intuitivamente, para que se admita, de maneira gulosa, o conjunto de novas conexões que maximize o lucro e a utilidade, deve-se escolher aquelas que fornecem maior valor para a soma $r_i * minTR_{ij} + u_i * extraTR_{ij}$, entretanto, essa abordagem pode não retornar o melhor conjunto, quando a quantidade de recursos requisitada pelas novas conexões for maior do que a capacidade disponível. Note que as conexões que fornecem maior valor para a soma $r_i * minTR_{ij} + u_i * extraTR_{ij}$ são, normalmente, aquelas que têm maiores requisitos de taxa mínima ($minTR_{ij}$) e que podem receber maior taxa extra ($extraTR_{ij} \leq maxTR_{ij} - minTR_{ij}$). Ao se admitir uma conexão j com essas características, por um lado, tem-se na rede uma conexão que fornece alto valor de lucro e utilidade, mas, por outro lado, conexões que exigem menos recursos terão que ser rejeitadas. Embora individualmente essas conexões rejeitadas forneçam menores valores de lucro e utilidade, se agregadas, elas podem trazer maior ganho para a rede.

Desse modo, esta política utiliza os conceitos de custo e benefício para tomar as decisões de admissão. O benefício trazido por uma nova conexão é igual a soma do lucro com a utilidade ($r_i * minTR_{ij} + u_i * extraTR_{ij}$) alcançada na sua admissão. O custo imposto pela admissão de uma conexão é igual a taxa que deverá ser alocada, ou seja, $TR_{ij} + extraTR_{ij}$. Serão admitidas as conexões que oferecerem maior ganho para a rede, sendo que o ganho é igual a razão entre o benefício e o custo.

Para conexões já admitidas, utiliza-se o mesmo raciocínio para decidir o quanto de taxa extra deve-se alocar para maximizar a utilidade. Neste caso, o benefício de se alocar taxa extra para uma conexão é igual a $u_i * extraTR_{ij}$, enquanto o custo é igual a $extraTR_{ij}$. Se o ganho obtido ao se alocar taxa extra para uma conexão já ativa na rede for maior do que o ganho alcançado com a admissão de uma nova conexão, a nova conexão será rejeitada e a conexão ativa receberá taxa extra.

O Algoritmo *GreedyMaxRevenueUtilityCAC* apresenta os passos realizados a cada intervalo de admissão. O algoritmo calcula, inicialmente, o ganho fornecido por cada nova conexão com valor de TR_{ij} menor do que a capacidade disponível e por cada conexão ativa não associada ao serviço UGS. Conexões novas que requerem taxa maior do que a capacidade disponível são

rejeitadas, enquanto conexões ativas associadas ao serviço UGS são desconsideradas dado que possuem requisito de taxa máxima igual ao requisito de taxa mínima e, portanto, não recebem taxa extra. O cálculo do ganho baseia-se no benefício e no custo oferecido por cada conexão, conforme explicado anteriormente. Para a taxa extra ($extraTR_{ij}$), utiliza-se o valor máximo que poderia ser alocado para a conexão. Para uma nova conexão esse valor é igual ao mínimo entre a capacidade disponível após sua admissão ($C_{available} - TR_{ij}$) e a diferença entre os requisitos de taxa máxima e taxa mínima ($maxTR_{ij} - minTR_{ij}$). Para uma conexão já ativa na rede, o valor máximo para $extraTR_{ij}$ é igual ao mínimo entre a capacidade disponível e a diferença entre os requisitos de taxa máxima e taxa mínima.

Calculados os valores de ganho ($reward_{ij}$) para cada conexão, o algoritmo escolhe as conexões que oferecem maior ganho para admitir, no caso de novas conexões, ou para alocar banda extra, caso seja uma conexão ativa na rede. Sempre que uma conexão é escolhida, a capacidade disponível é atualizada subtraindo-se $TR_{ij} + extraTR_{ij}$ ou $extraTR_{ij}$ para conexões novas e ativas, respectivamente. Quando a conexão que oferece maior ganho não puder ser admitida, calcula-se novo valor de ganho para as conexões com base no valor de $C_{available}$ atualizado. O processo termina quando a capacidade disponível for igual a zero ou quando o controle de admissão tiver finalizado suas decisões.

5.7 Considerações Finais do Capítulo 5

Este capítulo discutiu o mecanismo de controle de admissão em redes baseadas no padrão IEEE 802.16. Assim como acontece com o mecanismo de escalonamento, o padrão não define uma política para a implementação do controle de admissão.

Algumas das principais soluções disponíveis na literatura para este problema foram apresentadas e cinco novas políticas foram propostas. A primeira delas, tem como único objetivo garantir que a capacidade disponível seja suficiente para satisfazer o requisito de taxa mínima de cada conexão admitida. Para isso, além do requisito de taxa mínima, o mecanismo considera a sobrecarga gerada pelo tipo de serviço associado a conexão. Para os serviços UGS e ertPS, essa sobrecarga envolve o desperdício de banda que pode ocorrer ao se mapear o tamanho dos *grants* em bytes para número de *slots*. No caso dos serviços rtPS e nrtPS, essa sobrecarga reflete a taxa que é utilizada na alocação de *grants* para o envio de mensagens de requisição de banda.

A segunda política de controle de admissão estende a primeira proposta com o objetivo de maximizar o lucro na rede. Com uma formulação baseada no problema da mochila, essa política escolhe o conjunto de conexões capaz de oferecer maior lucro garantindo que a capacidade da rede seja suficiente para atender o requisito de taxa mínima de cada conexão. A terceira proposta tem o mesmo objetivo da segunda política de controle de admissão, entretanto, utiliza a técnica gulosa para a solução do problema.

A quarta e a quinta propostas têm como objetivo satisfazer tanto as expectativas do provedor

de serviços como as expectativas dos usuários da rede. Para tanto, essas políticas buscam maximizar o lucro e a utilidade ao mesmo tempo em que garantem a provisão do requisito de taxa mínima das conexões. A diferença entre as duas propostas está na maneira como o problema é solucionado. A quarta política é formulada como um problema de otimização inteira mista, enquanto a quinta política utiliza uma heurística baseada na técnica gulosa.

Uma análise do desempenho de cada uma das políticas de controle de admissão propostas, neste capítulo, será apresentada no Capítulo 6.

Tabela 5.1: Vantagens e desvantagens das propostas discutidas na Seção 5.1

Proposta	Vantagens	Desvantagens
Chen, Jiao e Wang, 2005 [27]	Algoritmo simples.	Não inclui o serviço ertPS; não considera a taxa usada pelo mecanismo de requisição de banda.
Wang <i>et alli</i> , 2007 [35]	Inclui os 5 tipos de serviço; considera conexões em <i>handoff</i> .	Não informa como estimar variáveis importantes no algoritmo.
Wang, He e Agrawal, 2007 [24]	Tenta maximizar a utilização do canal com um esquema de empréstimo e degradação de taxa.	Reserva parte da banda para o serviço UGS, o que pode gerar desperdício de recursos.
Niyato e Hossain, 2007 [18]	Considera os requisitos de latência e de taxa mínima.	Algoritmo complexo; não considera o serviço ertPS.
Chandra e Sahoo, 2007 [49]	Considera os requisitos de latência e de taxa mínima bem como a taxa necessária para o mecanismo de requisição de banda.	Não inclui o serviço ertPS; não detalha o algoritmo utilizado pela rotina <i>search</i> , que é parte importante do mecanismo proposto.
Chang, Chen e Chou, 2007 [11]	Usa a idéia de custo e recompensa, o que facilita a inclusão de variáveis de preço na admissão das conexões para se maximizar o lucro do provedor.	Assume que as estações têm diferentes prioridades, o que não está de acordo com o padrão; não define como configurar o valor da recompensa obtida pela rede ao se admitir uma determinada conexão.
Guo <i>et alli</i> , 2007 [54]	Inclui os 5 tipos de serviço; considera conexões em <i>handoff</i> ; faz reserva dinâmica de recursos para as conexões em <i>handoff</i> diminuindo o desperdício de banda.	Depois de atender o requisito de taxa mínima das conexões e reservar parte da capacidade para conexões em <i>handoff</i> , a capacidade disponível é repartida apenas entre as conexões rtPS, sendo que conexões ertPS e nrtPS também poderiam ser beneficiadas com taxa extra.
Rong <i>et alli</i> , 2008 [12]	Inclui os 5 tipos de serviço; busca maximizar lucro, utilidade e justiça.	Considera requisito de taxa igual para todas as conexões de um mesmo tipo de serviço, diferente do que especifica o padrão.
Masri, Abdelatif e Juanole, 2009 [45]	Algoritmo simples.	Aloca <i>slots</i> para conexões UGS e rtPS em cada <i>frame</i> para garantir latência, podendo gerar retardos desnecessários para conexões nrtPS quando o requisito de latência das conexões de tempo real é maior do que a duração de um <i>frame</i> .
Borin e Fonseca	Além da taxa mínima, considera a sobrecarga associada ao tipo de serviço solicitado pela conexão; permite maximizar lucro e utilidade; inclui os 5 tipos de serviço.	Serviços que fornecem maiores ganhos de lucro e utilidade são beneficiados com menores taxas de bloqueio.

ALGORITMO GREEDYMAXREVENUEUTILITYCAC

1. insira no vetor *reward_array* as conexões novas e as conexões ativas
 2. **para** cada conexão *j* armazenada no vetor *reward_array* **faça**
 3. **se** *j* é uma nova conexão **então**
 4. **se** $TR_{ij} > C_{available}$ **então**
 5. rejeite a conexão *j* e a exclua do vetor *reward_array*
 6. **senão**
 7. $extraTR_{ij} \leftarrow \min(C_{available} - TR_{ij}, maxTR_{ij} - minTR_{ij})$
 8. $benefit \leftarrow r_i * minTR_{ij} + u_i * extraTR_{ij}$
 9. $cost \leftarrow TR_{ij} + extraTR_{ij}$
 10. $reward_{ij} \leftarrow benefit \div cost$
 11. **senão**
 12. **se** *j* é uma conexão UGS **então**
 13. $extraTR_{ij} \leftarrow 0$
 14. exclua *j* do vetor *reward_array*
 15. **senão**
 16. $extraTR_{ij} \leftarrow \min(C_{available}, maxTR_{ij} - minTR_{ij})$
 17. $benefit \leftarrow u_i * extraTR_{ij}$
 18. $cost \leftarrow extraTR_{ij}$
 19. $reward_{ij} \leftarrow benefit \div cost$
 20. **enquanto** $C_{available} > 0$ e *reward_array* $\neq \emptyset$ **faça**
 21. busque no vetor *reward_array* a conexão *j* com maior valor para $reward_{ij}$
 22. **se** *j* é uma nova conexão **então**
 23. **se** $TR_{ij} + extraTR_{ij} \leq C_{available}$ **então**
 24. aceite a conexão *j* com banda extra igual a $extraTR_{ij}$
 25. $C_{available} \leftarrow C_{available} - (TR_{ij} + extraTR_{ij})$
 26. exclua *j* do vetor *reward_array*
 27. **senão**
 28. volte para o passo 2
 29. **senão**
 30. **se** $extraTR_{ij} \leq C_{available}$ **então**
 31. configure a taxa extra de *j* para $extraTR_{ij}$
 32. $C_{available} \leftarrow C_{available} - extraTR_{ij}$
 33. exclua *j* do vetor *reward_array*
 34. **senão**
 35. volte para o passo 2
-

Capítulo 6

Avaliação das Políticas de Escalonamento e Controle de Admissão

Este capítulo apresenta a avaliação da política de escalonamento proposto no Capítulo 4 e das políticas de controle de admissão propostas no Capítulo 5. A avaliação foi realizada através de experimentos de simulação, utilizando o módulo para simulação de redes WiMAX apresentado no Capítulo 3.

A Seção 6.1 descreve o cenário utilizado nas simulações, a Seção 6.2 apresenta uma análise dos resultados obtidos e a Seção 6.3 traz as considerações finais deste capítulo.

6.1 Descrição do Cenário de Simulação

O cenário de simulação consiste em um conjunto de SSs uniformemente distribuídas ao redor de uma BS. A duração do *frame* é de 5 ms e a capacidade do canal é de 40 Mbps, sendo utilizado o modo de duplexação TDD com razão 1:1 para *downlink* e *uplink*. Cada *slot* é capaz de transportar 100 bytes. Os experimentos utilizam tráfego de voz, voz com supressão de silêncio, vídeo, FTP e WEB que são servidos, respectivamente, pelos serviços UGS, ertPS, rtPS, nrtPS e BE [1] [2].

O tráfego de voz é gerado por uma fonte *on/off* [40]. A duração dos períodos *on* e *off* segue uma distribuição exponencial com médias 1,2 e 1,8 segundos, respectivamente. Durante o período *on* são gerados pacotes de 66 bytes a cada 20 ms. O tráfego de voz com supressão de silêncio utiliza a codificação EVRC (*Enhanced Variable Rate Codec*) [6]. Os pacotes são gerados a cada 20 ms utilizando taxa 1 (171 bits/pacote), taxa 1/2 (80 bits/pacote), taxa 1/4 (40 bits/pacote) ou taxa 1/8 (16 bits/pacote). O tráfego de vídeo é obtido a partir de traces de vídeo reais, cujas características são apresentadas na Tabela 6.1 [42]. O tráfego WEB é gerado a partir de uma distribuição híbrida Lognormal/Pareto. O corpo da distribuição, correspondendo à uma área de 0,93, é modelado por uma distribuição Lognormal com média de 27,6 KBytes, enquanto

a cauda é modelada por uma distribuição Pareto com média de 146,3 KBytes [41]. O tráfego FTP é gerado utilizando uma distribuição exponencial com média de 512 KBytes.

Tabela 6.1: Características dos traces de vídeo usados nas simulações

vídeo	taxa de pico (Mbps)	CoV do tamanho dos frames	Hurst
Aladdin	2.242800	1.217128	0.876541
Friends vol4 DVD	1.175520	1.820629	0.818063
Gupta Lecture	2.218080	1.398482	0.777175
Baseball	1.859760	1.284597	0.684708
MTV	2.976240	1.448513	0.842930
Star Wars V	1.720800	1.492767	0.767742
Lady Tramp DVD	0.910080	1.763723	0.813108
Ice Age DVD	1.058400	1.532578	0.834963
Dreamcatcher DVD	0.932400	1.400108	0.700975

O intervalo entre *grants* para os serviços UGS e ertPS é de 20 ms. O intervalo entre *grants unicast* para o envio de mensagens de requisição de banda é de 20 ms para o serviço rtPS e de 1 s para o serviço nrtPS. O requisito de latência máxima do serviço rtPS é de 100 ms e o requisito de taxa mínima e de taxa máxima varia para cada conexão, de acordo com a taxa média do vídeo transmitido. O serviço nrtPS tem requisito de taxa mínima de 200 Kbps e de taxa máxima de 300 Kbps. O serviço BE não tem nenhum requisito de QoS.

O tempo de vida das conexões UGS, ertPS, rtPS e nrtPS é determinado por uma distribuição exponencial com média de 600 s para o tráfego de vídeo e 300 s para os demais tipos de tráfego. Da mesma forma, a taxa de chegada das conexões segue uma distribuição exponencial com médias que variam entre 2 s e 60 s para cada tipo de serviço. Desse modo, na situação de menor carga, chega em média 1 conexão de cada tipo de serviço por minuto, enquanto na situação de maior carga chegam em média 30 conexões de cada tipo de serviço por minuto.

O número de conexões BE é fixo em todas as situações simuladas e igual a 20, dado que esse serviço não passa pelo processo de controle de admissão.

Para as políticas de controle de admissão que maximizam o lucro, a taxa de lucro associada a cada tipo de serviço é de 4 para o serviço UGS, 3 para o serviço rtPS, 2 para o serviço ertPS e 1 para o serviço nrtPS. Para as políticas de controle de admissão que maximizam a utilidade, os mesmos valores são usados para a taxa de utilidade fornecida por cada tipo de serviço, com exceção do serviço UGS que tem valor de utilidade 0, pois não recebe taxa extra.

Para cada valor de taxa de chegada de conexões testado, foram executadas 10 replicações com sementes diferentes, para gerar o intervalo de confiança de 95%. Cada simulação teve duração de 3.600s. Os resultados apresentados na Seção 6.2 mostram a média obtida e o intervalo de confiança de 95%.

6.2 Avaliação dos Resultados

A política de escalonamento e as políticas de controle de admissão propostas, nesta tese, atuam de maneira complementar para a provisão de QoS em redes baseadas no padrão IEEE 802.16, logo, serão avaliadas, conjuntamente, nesta seção.

A política de controle de admissão que se baseia no requisito de taxa mínima das conexões, introduzida na Seção 5.2, será referenciada nesta seção como CA Simples. A política de controle de admissão, proposta na Seção 5.3, que tem o objetivo de maximizar o lucro da rede de maneira ótima será referenciada pelo termo CA Ótimo Max. Lucro, enquanto a política que tem o mesmo objetivo, mas usa a técnica gulosa, introduzida na Seção 5.4, será chamada CA Guloso Max. Lucro. A política ótima (Seção 5.5) e a política gulosa (Seção 5.6) para maximização do lucro e da utilidade serão referenciadas, respectivamente, pelos termos CA Ótimo Max. Lucro e Utilidade e CA Guloso Max. Lucro e Utilidade.

O primeiro conjunto de resultados tem o objetivo de avaliar a taxa de bloqueio e o lucro obtido com cada uma das políticas de controle de admissão introduzidas no Capítulo 5 bem como a utilidade fornecida pelas políticas CA Ótimo Max. Lucro e Utilidade e CA Guloso Max. Lucro e Utilidade.

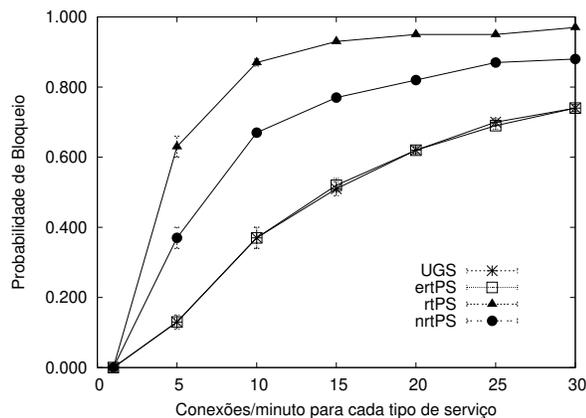


Figura 6.1: Probabilidade de Bloqueio para a política CA Simples

As Figuras 6.1 a 6.5 apresentam a taxa de bloqueio dos serviços UGS, ertPS, rtPS e nrtPS para cada uma das políticas de controle de admissão propostas. O bloqueio implica no descarte do pedido de estabelecimento da conexão.

Com a política de controle de admissão CA Simples, os serviços UGS e ertPS têm valores de probabilidade de bloqueio idênticos em todos os cenários simulados, como mostra a Figura 6.1. Os valores de probabilidade de bloqueio para esses dois tipos de serviço são mais baixos do que para os serviços rtPS e nrtPS, pois essa política tende a privilegiar conexões com menores valores para o requisito de taxa mínima, dado que esse é o único parâmetro considerado na

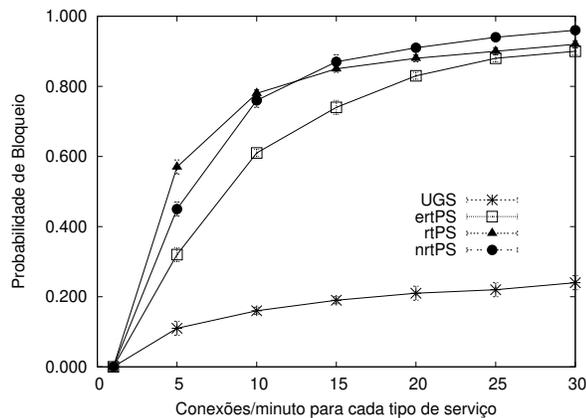


Figura 6.2: Probabilidade de Bloqueio para a política CA Ótimo Max. Lucro

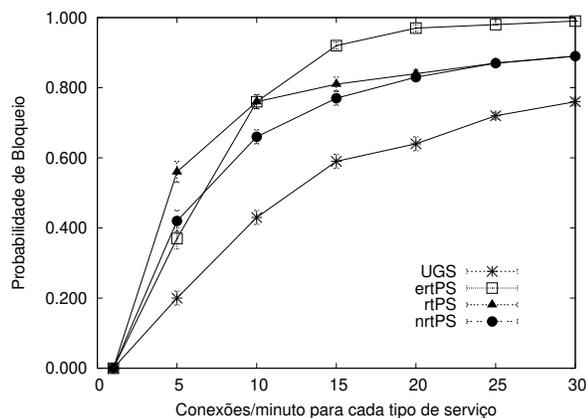


Figura 6.3: Probabilidade de Bloqueio para a política CA Guloso Max. Lucro

admissão. O serviço rtPS apresenta os maiores valores de probabilidade de bloqueio por ser o serviço com requisito mais alto de taxa mínima.

As políticas CA Ótimo Max. Lucro e CA Guloso Max. Lucro fornecem menor probabilidade de bloqueio para o serviço UGS, cujas conexões retornam altos valores de lucro exigindo baixas taxas de transmissão. No entanto, como mostram as Figuras 6.2 e 6.3, a política CA Ótimo Max. Lucro aceita mais conexões UGS do que a política CA Guloso Max. Lucro. No cenário de mais alta carga na rede, a política CA Ótimo Max. Lucro chega a aceitar até três vezes mais conexões UGS. Da mesma forma, na política ótima são aceitas até 10% mais conexões ertPS do que na política gulosa. No entanto, o fato da política CA Guloso Max. Lucro admitir um menor número de conexões UGS e ertPS, faz com que ele seja capaz de admitir um maior número de conexões rtPS e nrtPS. Essas diferenças influenciam os resultados de lucro obtidos com ambas as políticas, como será explicado mais adiante.

As Figuras 6.4 e 6.5 mostram que as políticas CA Ótimo Max. Lucro e Utilidade e CA

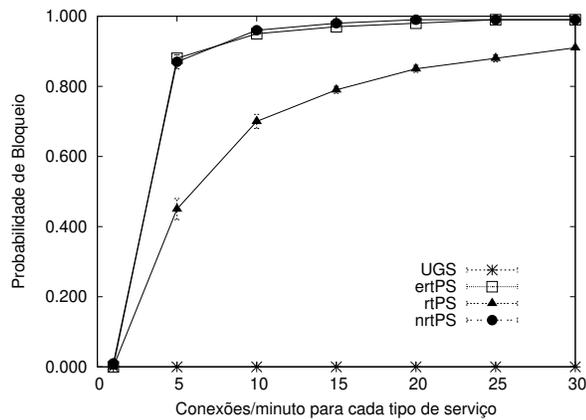


Figura 6.4: Probabilidade de Bloqueio para a política CA Ótimo Max. Lucro e Utilidade

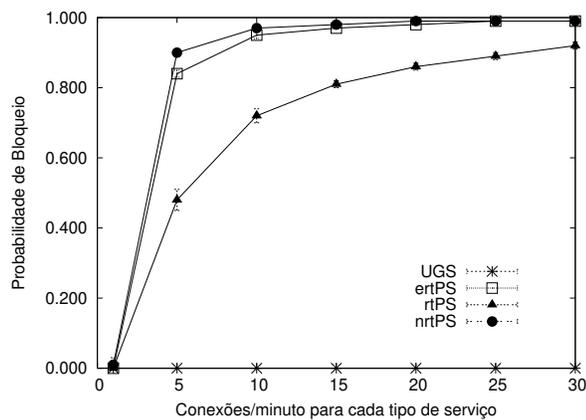


Figura 6.5: Probabilidade de Bloqueio para a política CA Guloso Max. Lucro e Utilidade

Guloso Max. Lucro e Utilidade têm resultados de probabilidade de bloqueio muito semelhantes em todos os cenários simulados. Em ambos os casos, todas as conexões UGS são admitidas para qualquer que seja a carga na rede e os serviços ertPS e nrtPS têm valores de probabilidade bloqueio muito próximos e maiores do que os valores de probabilidade de bloqueio do serviço rtPS. Em comparação com as outras três políticas de controle de admissão, as políticas CA Ótimo Max. Lucro e Utilidade e CA Guloso Max. Lucro e Utilidade bloqueiam um número maior de conexões ertPS e nrtPS e um número semelhante de conexões rtPS e, dessa maneira, conseguem admitir mais conexões UGS, resultando em maiores valores de lucro do que as políticas que maximizam apenas o lucro, como será mostrado a seguir.

A Figura 6.6 apresenta o valor do lucro acumulado em cada uma das situações de carga testadas para as cinco políticas de controle de admissão. Embora a política CA Simples não tente maximizar o lucro, mediu-se os valores de lucro fornecidos por ela utilizando-se a mesma função de lucro ($r_i * \min T R_{ij}$) considerada pelas outras quatro políticas para cada conexão

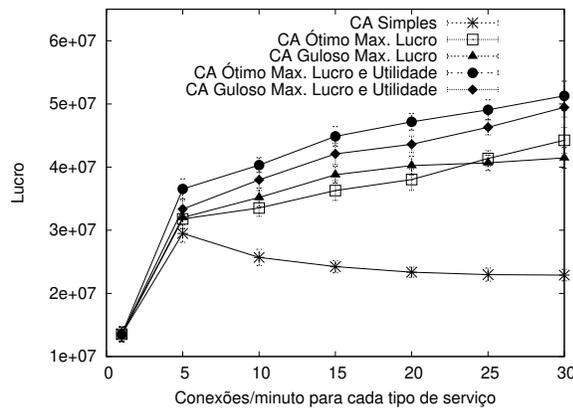


Figura 6.6: Lucro fornecido pelas políticas de controle de admissão

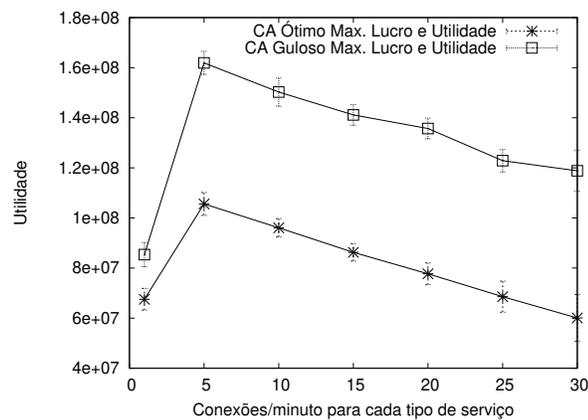


Figura 6.7: Utilidade fornecida pelas políticas CA Ótimo Max. Lucro e Utilidade e CA Guloso Max. Lucro e Utilidade

admitida.

A política CA Simples fornece os menores valores de lucro e, pelo fato de considerar apenas a taxa mínima das conexões na admissão, esses valores tendem a se estabilizar conforme a rede fica sobrecarregada. A queda nos valores de lucro, observada no cenário onde a taxa de chegada de conexões é de 10 conexões/minuto para cada tipo de serviço, deve-se ao alto número de conexões rtPS bloqueadas neste cenário. As conexões rtPS são capazes de fornecer altas taxas de lucro por requisitarem altas taxas mínimas de transmissão e a política CA Simples é a que apresenta os maiores valores de probabilidade de bloqueio para este tipo de serviço.

As políticas CA Ótimo Max. Lucro e CA Guloso Max. Lucro fornecem valores próximos de lucro. Nos cenários onde a taxa de chegada para cada tipo de serviço é de 10, 15 e 20 conexões por minuto, a política CA Guloso Max. Lucro fornece lucro médio ligeiramente maior do que a política ótima. Isso acontece porque as políticas dão soluções locais, cada uma delas baseada na

entrada da rodada atual de admissão, dessa forma, decisões diferentes tomadas pelas políticas gulosa e ótima em um instante t , podem fazer com que no instante $t+x$ a política gulosa consiga um melhor resultado do que a política ótima. Para os cenários nos quais a taxa de chegada para cada tipo de serviço é maior do que 20 conexões por minuto, os valores de lucro estabilizam para a política gulosa e continuam aumentando para a política ótima.

Da mesma forma, a política CA Guloso Max. Lucro e Utilidade provê valores de lucro próximos àqueles fornecidos pela política CA Ótimo Max. Lucro e Utilidade. No entanto, em todos os cenários, a política ótima fornece lucro um pouco maior do que a política que utiliza a técnica gulosa.

As políticas que maximizam lucro e utilidade apresentam melhor desempenho em termos de lucro do que as políticas que maximizam apenas lucro. Os resultados de probabilidade de bloqueio mostram que as quatro políticas tomam decisões de admissão diferentes em um mesmo cenário. Embora, intuitivamente, espera-se que as políticas que maximizam apenas lucro forneçam os maiores valores de lucro, a Figura 6.6 mostra que, a longo prazo, as decisões tomadas pelas políticas que maximizam lucro e utilidade provêm os melhores resultados.

A Figura 6.7 mostra o ganho de utilidade obtido com as políticas CA Ótimo Max. Lucro e Utilidade e CA Guloso Max. Lucro e Utilidade. Os valores apresentados na figura representam a média do ganho de utilidade acumulado em cada execução de simulação, sendo que o ganho fornecido por cada conexão para a qual o controle de admissão reserva taxa extra é dado pela função $u_i * extraTR_{ij}$. Em todos os cenários simulados, a política gulosa forneceu maior ganho de utilidade do que a política ótima. Em contrapartida, conforme mostrado na Figura 6.6, a política ótima permite que o provedor obtenha maiores valores de lucro. Para ambas as políticas, o ganho de utilidade reduz com o aumento da taxa de chegada de conexões. Isso acontece porque, quando a rede está sobrecarregada, o ganho obtido com a admissão de novas conexões UGS é maior do que o ganho obtido com a alocação de banda extra para uma conexão admitida. Uma nova conexão UGS fornece um ganho de 4 unidades para cada 1 Kbps, enquanto a alocação de taxa extra para uma conexão rtPS, que é o serviço que oferece o maior ganho de utilidade, fornece um ganho de 3 unidades para cada 1 Kbps. Esse fato se reflete nos resultados de probabilidade de bloqueio do serviço UGS quando se utilizam as políticas CA Ótimo Max. Lucro e Utilidade e CA Guloso Max. Lucro e Utilidade. Caso o provedor de serviço tenha interesse em obter maior ganho de utilidade, ainda que sacrificando o lucro, ele pode aumentar as taxas de ganho de utilidade fornecidas por cada tipo de serviço.

O segundo conjunto de resultados compara os parâmetros de QoS medidos nas simulações com os requisitos correspondentes definidos na Seção 6.1 para cada tipo de serviço. O objetivo é verificar a capacidade das políticas de controle de admissão e do escalonador de prover QoS para as conexões. Espera-se, mais especificamente, que: 1) conexões UGS e ertPS recebam *grants* periódicos conforme o intervalo definido pelo parâmetro *unsolicited grant interval*, 2) o tráfego que não excede a taxa mínima de uma conexão rtPS tenha latência menor do que a

latência máxima requisitada pela conexão, e 3) a vazão média de uma conexão rtPS ou nrtPS não seja menor do que seu requisito de taxa mínima, nem maior do que seu requisito de taxa máxima.

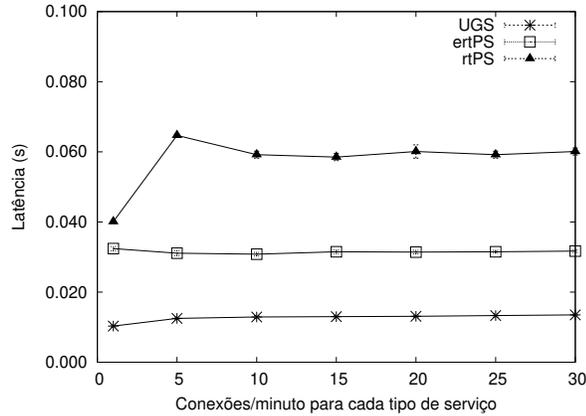


Figura 6.8: Latência das conexões UGS, ertPS e rtPS com a política CA Simples

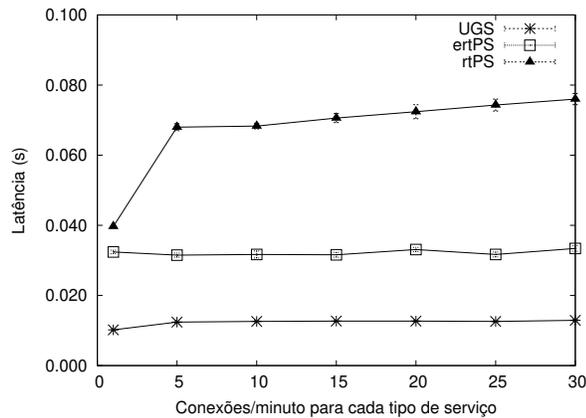


Figura 6.9: Latência das conexões UGS, ertPS e rtPS com a política CA Ótimo Max. Lucro

As Figuras 6.8 a 6.12 mostram os resultados de latência para os serviços UGS, ertPS e rtPS obtidos com cada uma das cinco políticas de controle de admissão utilizadas em conjunto com o escalonador proposto. Os valores de latência dos serviços UGS e ertPS permanecem, praticamente, constantes em todos os cenários simulados e apresentam intervalo de confiança bastante reduzido, o que mostra que a variação na latência dos pacotes transmitidos por conexões associadas a esses dois tipos de serviço foi pequena e, portanto, o escalonador foi capaz de alocar oportunidades de transmissão periodicamente como especificado no padrão. Os valores de latência das conexões rtPS são menores do que 100 ms em todos os cenários, o que significa que

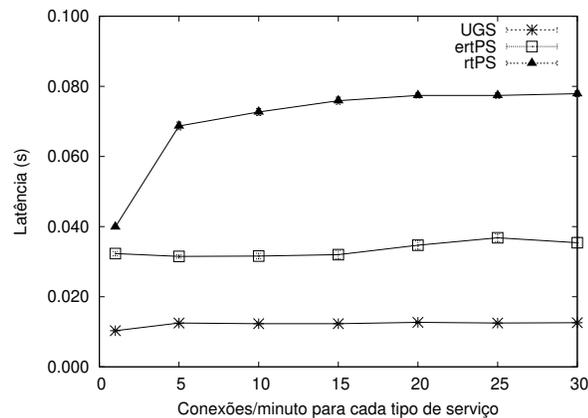


Figura 6.10: Latência das conexões UGS, ertPS e rtPS com a política CA Guloso Max. Lucro

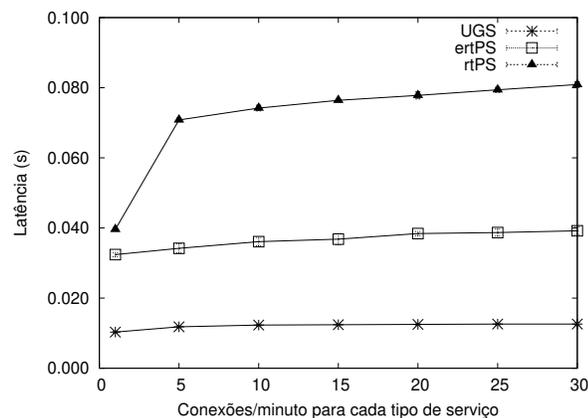


Figura 6.11: Latência das conexões UGS, ertPS e rtPS com a política CA Ótimo Max. Lucro e Utilidade

o escalonador, independente da política de controle de admissão adotada, foi capaz de garantir o requisito de latência máxima desse serviço.

Observou-se duas tendências nos resultados de vazão das conexões rtPS. Desse modo, a vazão média do serviço rtPS será analisada para conexões que transmitem o vídeo Ice Age e o vídeo Alladin, que são representativos para os dois conjuntos de resultados.

A Figura 6.13 mostra os valores médios de carga oferecida e de vazão das conexões que transmitem o vídeo Ice Age, as quais têm requisito de taxa mínima igual a 200 Kbps e requisito de taxa máxima igual a 300 Kbps. Em todas as situações de carga e independente da política de admissão, as conexões rtPS que transmitem o vídeo Ice Age obtiveram vazão média dentro do intervalo estabelecido pelos requisitos de taxa mínima e taxa máxima.

A Figura 6.14 mostra os valores médios de carga oferecida e de vazão das conexões que transmitem o vídeo Aladdin, as quais têm requisito de taxa mínima igual a 900 Kbps e requisito

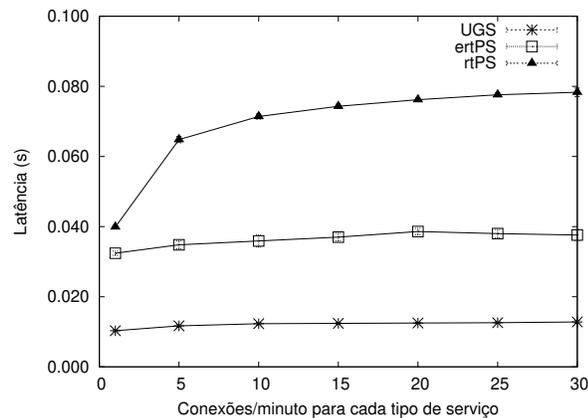


Figura 6.12: Latência das conexões UGS, ertPS e rtPS com a política CA Guloso Max. Lucro e Utilidade

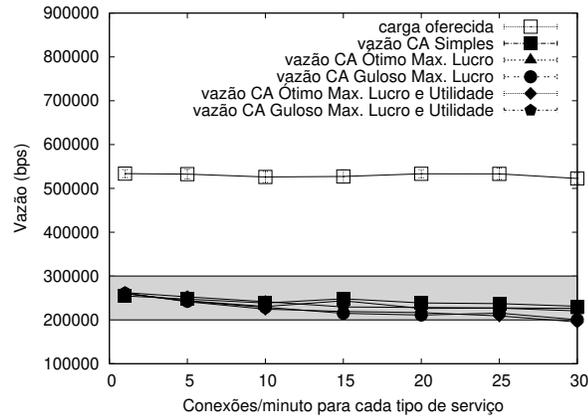


Figura 6.13: Vazão das conexões rtPS que transmitiram o vídeo Ice Age

de taxa máxima igual a 1.1 Mbps. Para os cenários com taxa de chegada de até 10 conexões de cada tipo de serviço, para qualquer que seja a política de controle de admissão utilizada, a vazão das conexões não é menor do que o requisito de taxa mínima e nem maior do que o requisito de taxa máxima. Para os cenários nos quais a carga na rede é maior, a vazão das conexões foi um pouco inferior ao requisito de taxa mínima quando se utilizou as políticas CA Ótimo Max. Lucro, CA Guloso Max. Lucro e CA Ótimo Max. Lucro e Utilidade. Embora a capacidade da rede seja suficiente para o escalonador garantir a taxa mínima de todas as conexões aceitas, o vídeo Aladdin, como outros que mostram resultados de vazão semelhantes, geram rajadas mais frequentes e mais longas do que vídeos com características semelhantes ao vídeo Ice Age. Com isso, se a latência máxima de uma rajada que está armazenada na fila da estação cliente é violada em uma janela de tempo na qual a conexão já recebeu a taxa mínima, há um grande descarte de pacotes na fila da estação cliente, reduzindo a vazão dessas conexões. Esse fator, associado à

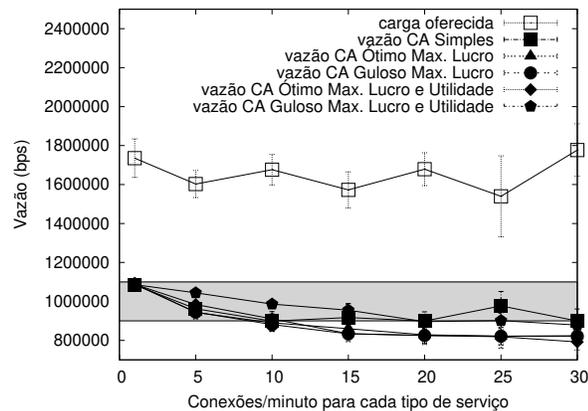


Figura 6.14: Vazão das conexões rtPS que transmitiram o vídeo Aladdin

um grande número de conexões rtPS aceitas, dificulta a provisão de taxa mínima para conexões com tráfego semelhante ao produzido pelo vídeo Aladdin. A política CA Simples é a política que admite o menor número de conexões rtPS, logo, o escalonador tem maior facilidade em atender as conexões associadas a esse serviço. A política CA Guloso Max. Lucro e Utilidade, conforme mostrado na Figura 6.7, provê maior utilidade do que a política CA Ótimo Max. Lucro e Utilidade, pois aloca mais banda extra para as conexões rtPS, resultando em um menor número de pacotes descartados nas estações cliente.

Mecanismos de moldagem de tráfego para suavização de rajadas, com conseqüente redução da taxa de perda de pacotes e melhor utilização da banda, têm sido amplamente estudados. Em [43], os autores investigam mecanismos de moldagem de tráfego para transmissão de vídeos MPEG. Conforme o padrão IEEE 802.16, as estações cliente também devem contribuir na provisão de Qualidade de Serviço, inclusive através do policiamento do tráfego para que, na média, não ultrapasse a taxa máxima acordada durante o processo de controle de admissão. Assim sendo, mecanismos de moldagem de tráfego poderiam ser utilizados nas estações cliente para suavizar o tráfego com alta variabilidade em rajadas e, dessa forma, reduzir a taxa de perda de pacotes.

A vazão média das conexões nrtPS mantém-se dentro do intervalo definido pelos requisitos de taxa mínima e taxa máxima, delimitado pela região cinza na Figura 6.15, em todas as situações de carga simuladas para as cinco políticas de controle de admissão. Embora as fontes FTP sejam configuradas para gerar uma carga com média de 300 Kbps, igual ao requisito de taxa máxima do serviço, o mecanismo de controle de congestionamento do protocolo TCP reduz a taxa do tráfego gerado em resposta ao policiamento da taxa máxima feito pelo escalonador. Desse modo, o escalonador consegue servir todo o tráfego gerado pelas fontes em todas as situações de carga simuladas.

A Figura 6.16 mostra a carga oferecida e a vazão das conexões BE para cada um dos me-

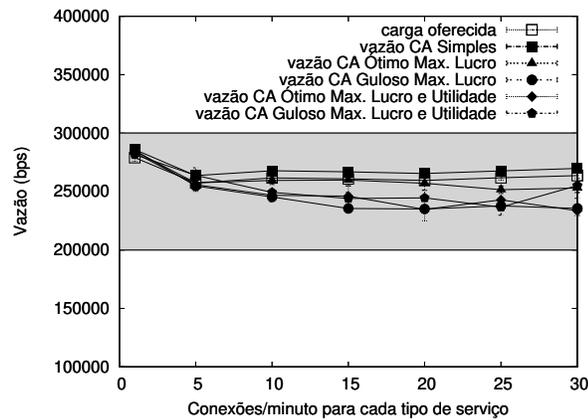


Figura 6.15: Vazão das conexões nrtPS

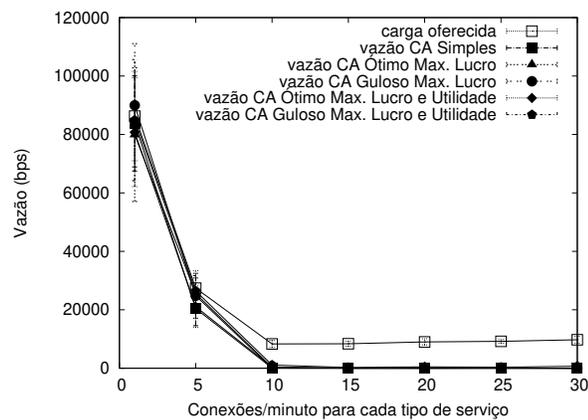


Figura 6.16: Vazão das conexões BE

canismos de controle de admissão propostos. Quando a rede não está sobrecarregada, o escalonador aloca para as conexões BE os *slots* não utilizados pelas conexões de mais alta prioridade. A partir do momento em que a rede está sobrecarregada, os *slots* que eram alocados para as conexões BE são utilizados para satisfazer as necessidades das conexões associadas aos outros quatro tipos de serviço. Embora os mecanismos propostos possam resultar em baixa ou nenhuma vazão para o serviço BE em situações nas quais a rede está sobrecarregada com conexões de mais alta prioridade, as especificações do padrão IEEE 802.16 não são violadas, dado que ele não especifica quaisquer garantias para o tráfego BE. Caso o provedor tenha interesse em garantir uma taxa mínima de transmissão para as conexões BE em qualquer que seja a condição de carga na rede, os mecanismos de controle de admissão propostos podem ser alterados para que parte da capacidade da rede seja reservada para esse serviço.

6.3 Considerações Finais do Capítulo 6

Este capítulo apresentou uma avaliação do mecanismo de escalonamento proposto no Capítulo 4 e das políticas de controle de admissão propostas no Capítulo 5.

Dentre as políticas de controle de admissão propostas, as políticas CA Ótimo Max. Lucro e Utilidade e CA Guloso Max. Lucro e Utilidade fornecem os melhores resultados de lucro com a vantagem de também maximizarem a utilidade da rede. Para as outras políticas de controle de admissão, quando há sobra de recursos, o escalonador faz o compartilhamento desses recursos sem levar em conta ganhos de utilidade.

Os resultados de vazão e latência apresentados mostram que o escalonador em conjunto com qualquer uma das cinco políticas de controle de admissão fornece uma solução compatível com o padrão IEEE 802.16 capaz de prover Qualidade de Serviço para os diferentes tipos de serviço. Entretanto, quando existem fluxos de tráfego com alta variabilidade em rajada, como o tráfego gerado por vídeos MPEG, comumente servido pelo serviço rtPS, torna-se imprescindível a cooperação do policiamento de tráfego nas estações cliente.

Os resultados mostram, ainda, que o escalonador faz uso eficiente dos recursos disponíveis, alocando para o tráfego de melhor esforço os *slots* não utilizados pelos serviços de mais alta prioridade.

Capítulo 7

Conclusões e Trabalhos Futuros

7.1 Conclusões

Por ser um padrão recente e deixar, propositalmente, vários pontos em aberto para que os fabricantes possam diferenciar seus produtos, o padrão IEEE 802.16 tem despertado grande interesse na comunidade científica. Em particular, diversas pesquisas têm sido desenvolvidas envolvendo os mecanismos para provisão de Qualidade de Serviço (QoS) da camada de acesso ao meio.

O Capítulo 2 apresentou uma descrição do padrão IEEE 802.16, que define a camada física e a camada de acesso ao meio (MAC) das redes metropolitanas para acesso sem fio em banda larga. A camada MAC define uma arquitetura de QoS que inclui cinco tipos de serviço, para a diferenciação e priorização do tráfego, um mecanismo de escalonamento e um mecanismo de controle de admissão de novas conexões. No entanto, o padrão não define políticas para os mecanismos de escalonamento e controle de admissão, os quais são de fundamental importância para que os requisitos de QoS de cada um dos tipos de serviço sejam garantidos bem como para promover a utilização eficiente dos recursos da rede.

Desse modo, esta tese concentrou-se na investigação e proposta de mecanismos de escalonamento e controle de admissão para provisão de Qualidade de Serviço em redes baseadas no padrão IEEE 802.16.

O Capítulo 4 discutiu os desafios relacionados ao escalonamento nas redes IEEE 802.16. O escalonamento nessas redes é realizado por três escalonadores: i) o escalonador *downlink* da estação base, responsável pela transmissão do tráfego proveniente da rede núcleo e destinado às estações cliente, ii) o escalonador *uplink* da estação base, que faz o compartilhamento dos recursos do canal *uplink* entre todas as estações cliente e iii) o escalonador *uplink* das estações cliente, responsável pela transmissão do tráfego *uplink* nas oportunidades de transmissão alocadas pela estação base. Conforme discutido no Capítulo 4, o escalonador *uplink* da estação base é o que apresenta os maiores desafios, pois não tem acesso direto às filas que armazenam os dados a serem transmitidos e, ainda assim, deve ser capaz de satisfazer os requisitos de QoS

das conexões estabelecidas pelas estações cliente ao mesmo tempo em que utiliza de maneira eficiente os recursos do canal *uplink*.

Os desafios encontrados no projeto do escalonador *uplink* da estação base motivaram a proposta de uma solução para este problema, a qual foi introduzida no Capítulo 4. O escalonador proposto utiliza três filas de prioridade para armazenar as requisições de banda enviadas pelas estações cliente. Estas requisições podem migrar das filas de mais baixa prioridade para a fila de alta prioridade, quando precisarem ser atendidas para que os requisitos de QoS das conexões que as geraram sejam garantidos. O escalonador fornece garantias de latência máxima e taxa mínima, sem violar os limites impostos pelos requisitos de taxa máxima e tamanho máximo da rajada.

O Capítulo 5 concentrou-se no controle de admissão de conexões *uplink* em redes IEEE 802.16. Depois de apresentar um levantamento sobre as principais soluções propostas, na literatura, o capítulo introduziu cinco novas políticas de controle de admissão. As políticas propostas visam atingir três objetivos: 1) restringir o número de conexões simultaneamente presentes na rede, para que o escalonador proposto tenha recursos suficientes para atender os requisitos de QoS de cada conexão *uplink*, 2) atender as expectativas do provedor de serviço através da maximização do lucro e 3) maximizar a satisfação dos usuários através da disponibilização de recursos extras, além do combinado no contrato de admissão.

A primeira política proposta serviu como base para o desenvolvimento das outras quatro propostas. Esta política é a mais simples dentre as cinco e atende apenas o primeiro objetivo. Uma nova conexão é aceita sempre que há recursos suficientes para atender seu requisito de taxa mínima somado à sobrecarga gerada pelo tipo de serviço solicitado. A segunda e a terceira políticas propostas estendem a primeira atendendo, também, o segundo objetivo. A diferença entre as duas soluções está na forma como o problema é resolvido. A segunda política utiliza programação linear inteira para fornecer uma solução ótima para a maximização do lucro na admissão das novas conexões, enquanto a terceira política fornece uma solução menos complexa, pois utiliza uma técnica gulosa para escolher as conexões que fornecem maior lucro. A quarta e a quinta políticas atendem os três objetivos. A quarta política resolve o problema da maximização do lucro e da utilidade de forma ótima, utilizando programação linear inteira mista. Uma alternativa menos complexa é apresentada pela quinta política, que utiliza uma técnica gulosa para resolver o mesmo problema de maximização.

Dado que a provisão de Qualidade de Serviço em redes baseadas no padrão IEEE 802.16 depende da ação conjunta dos mecanismos de escalonamento e controle de admissão, a avaliação do desempenho das soluções propostas, apresentada no Capítulo 6, foi realizada testando-se cada uma das cinco políticas de controle de admissão com o escalonador proposto em cenários idênticos de simulação. Todas as combinações testadas se mostraram compatíveis com as especificações do padrão IEEE 802.16 e capazes de atender os requisitos de QoS das conexões *uplink*. A compatibilidade com o padrão é essencial para que uma solução possa ser adotada e

implementada em uma estação base, a qual deverá passar pelo processo de certificação do Fórum WiMAX, antes de entrar no mercado. A provisão de QoS é uma das principais expectativas dos usuários, logo, fundamental para o sucesso dessa tecnologia de acesso a Internet.

A partir dos testes realizados com traces de vídeo real, servidos pelo serviço rtPS (*real time Polling Service*), constatou-se que, embora a estação base seja o elemento principal na provisão de QoS, mecanismos que auxiliem nessa tarefa, também, devem ser incluídos nas estações cliente. As aplicações de vídeo, cada vez mais comuns na Internet, impõem grandes desafios para os mecanismos de provisão de QoS, pois geram tráfego com alta variabilidade em rajada, sensível ao atraso e que exige altas taxas de transmissão. O conjunto de parâmetros de QoS fornecido por uma conexão rtPS, que inclui os requisitos de taxa mínima, taxa máxima e latência máxima, não fornecem dados suficientes para que a estação base faça uma estimativa precisa de como será o comportamento do tráfego. Assim sendo, mecanismos de policiamento e moldagem de tráfego devem ser incluídos nas estações cliente para que conexões que transmitam tráfego com as características mencionadas acima não sejam penalizadas com altas taxas de perda de pacotes.

Para realizar o estudo proposto nesta tese, foi necessário projetar e desenvolver um módulo para simulação de redes IEEE 802.16. O projeto, características da implementação bem como o teste de conformidade do módulo foram apresentados no Capítulo 3. Essa parte do trabalho serviu como ferramenta para a pesquisa realizada nesta tese, mas, ainda assim, fornece uma contribuição importante para a comunidade científica visto a quantidade de pesquisadores de diversas partes do mundo que efetuaram a cópia do código disponibilizado em [58].

7.2 Trabalhos Futuros

Esta seção apresenta questões que não foram tratadas nesta tese, mas que são relevantes e podem trazer resultados importantes, bem como alguns direcionamentos para a realização de trabalhos futuros.

7.2.1 Extensão dos mecanismos propostos para considerar a camada física

Conforme discutido no Capítulo 4, a tecnologia utilizada na camada física pode influenciar o projeto do escalonador *uplink* da estação base. O escalonador proposto nesta tese fornece uma solução compatível com o padrão IEEE 802.16 capaz de fornecer QoS e que pode ser estendida para se adaptar às características da camada física utilizada. Dentre as tecnologias de camada física especificadas no padrão, espera-se que as tecnologias OFDM e OFDMA sejam as mais utilizadas. Para redes com camada física OFDM, o escalonador poderia ser estendido para considerar as diferentes modulações utilizadas pelas estações cliente. Diferentes modulações

resultam em diferentes taxas de transmissão. Assim sendo, além dos parâmetros de QoS das conexões, o escalonador poderia utilizar os valores de taxa de transmissão de cada conexão para decidir a maneira mais eficiente de compartilhar os recursos em cada *frame*. Para redes com camada física OFDMA, a extensão do escalonador proposto também é possível, porém mais complexa, pois, além de *slots*, o escalonador deve alocar subportadoras para cada usuário.

Da mesma forma, as políticas de controle de admissão podem ser estendidas para considerar a camada física. Isso pode ser feito, por exemplo, dando maior prioridade de admissão para conexões provenientes de estações cliente que apresentam maior taxa de transmissão.

7.2.2 Extensão dos mecanismos propostos para considerar mobilidade

Uma das grandes expectativas em relação às redes baseadas no padrão IEEE 802.16 é que elas permitam o acesso em banda larga a qualquer hora e em qualquer lugar. Essa questão envolve a capacidade de se prover mobilidade para os usuários.

Em redes baseadas na extensão IEEE 802.16e [2], que provê mobilidade, o mecanismo de controle de admissão deve atender as solicitações de novas conexões, bem como de conexões em andamento provenientes de outras células. Conexões provenientes de outras células, normalmente, têm maior prioridade na admissão do que novas conexões, pois o bloqueio de uma conexão em andamento costuma trazer mais insatisfação para um usuário do que o bloqueio de uma nova conexão. Assim sendo, as políticas de controle de admissão propostas podem ser estendidas para se adaptar a cenários com mobilidade. Conexões provenientes de outras células e conexões novas podem, por exemplo, fornecer diferentes ganhos em relação a satisfação dos usuários e de acordo com o tipo de serviço solicitado.

Em 2007, o IEEE publicou o padrão IEEE 802.16j [4], que inclui um novo elemento na rede IEEE 802.16: a estação *relay*. Essa estação estende o alcance de uma estação base, provendo maior mobilidade para os usuários. Usuários fora do alcance da estação base, podem se conectar a ela através da estação *relay*. Do ponto de vista da estação base, a estação *relay* é considerada uma estação cliente. No entanto, a estação *relay* deve possuir um escalonador mais inteligente do que aquele implementado na estações cliente, dado que deve compartilhar os recursos requisitados à estação base entre vários usuários. Desse modo, um estudo interessante seria investigar a viabilidade de se usar o escalonador proposto em estações *relay*.

7.2.3 Testes com diferentes valores de lucro e de utilidade

Nos testes feitos com as políticas de controle de admissão propostas, considerou-se os mesmos valores para o ganhos de lucro e de utilidade para um mesmo tipo de serviço; com exceção do serviço UGS, que fornecia ganho de lucro igual a 4 e ganho de utilidade igual a 0, dado que esse tipo de serviço serve tráfego de taxa constante e, portanto, não recebe taxa extra. Os valores

utilizados são interessantes, pois refletem as diferentes prioridades dos serviços e dão mesmo peso para lucro e utilidade. No entanto, testes adicionais poderiam ser realizados com valores de lucro e de utilidade variados para analisar o impacto desses valores nas probabilidades de bloqueio e nos ganhos de lucro e utilidade. Baseado nestes testes, seria possível, ainda, delinear regras de como configurar os valores dos ganhos de lucro e de utilidade para atender diferentes objetivos e necessidades dos provedores.

7.2.4 Mecanismos de QoS para as estações cliente

Esta tese concentrou-se nos mecanismos para provisão de QoS que devem ser implementados na estação base, no entanto, de acordo com as especificações do padrão IEEE 802.16 e com os resultados apresentados no Capítulo 6, as estações cliente também devem cooperar nesta tarefa. Segundo o padrão, as estações cliente devem possuir um escalonador para o tráfego *uplink* para decidir quais pacotes devem ser transmitidos nas oportunidades de transmissão alocadas pelo escalonador *uplink* da estação base. Além disso, elas devem realizar o policiamento do tráfego para que, na média, ele não ultrapasse o requisito de taxa máxima acordado no estabelecimento das conexões. Os resultados apresentados nesta tese mostram a necessidade de haver esse policiamento nas estações cliente bem como mecanismos de moldagem de tráfego para tráfegos com alta variabilidade em rajada.

Assim sendo, seria interessante investigar mecanismos para provisão de QoS para as estações cliente. Embora importante, essa questão tem recebido pouco destaque na literatura. Para validar tais mecanismos, experimentos de simulação poderiam ser realizados em conjunto com os mecanismos propostos nesta tese para a estação base.

Referências Bibliográficas

- [1] IEEE Standard 802.16-2004, Part 16. *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, 2004.
- [2] IEEE Standard 802.16-2005, Part 16. *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, 2005.
- [3] IEEE Standard 802.16-2001, *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, 2002.
- [4] IEEE Standard 802.16j. *Air Interface for Fixed and Mobile Broadband Wireless Access Systems: Multihop Relay Specification*, 2007.
- [5] IEEE P802.16a/D3-2001, *Draft Amendment to IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems - Medium Access Control Modifications and Additional Physical Layer Specifications for 2-11 GHz*, 2002.
- [6] 3GPP2 C.S0014-0, *Enhanced Variable Rate Codec (EVRC)*.
- [7] A. Ghosh, D.R. Wolter, J.F Andrews e R. Chen, Broadband Wireless Access with WiMax/802.16: Current Performance Benchmarks and Future Potential. *IEEE Communications Magazine*, vol. 43, no. 2, pg. 129-136, 2005.
- [8] A. K. Parekh, R. G. Gallager, A generalized processor sharing approach to flow control in integrated services networks: the single-node case. *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pg. 344-357, 1993.
- [9] A. Mohammadi, S. G. Akl e F. Behnamfar, Optimal linear-time QoS-based scheduling for WiMAX. *Em Proceeding of the 21st Canadian Conference on Electrical and Computer Engineering*, pg. 1811-1814, 2008.
- [10] A. Sayenko, O. Alanen e T. Hamalainen, Scheduling solution for the IEEE 802.16 base station. *Computer Networks*, vol. 52, no. 1, pg. 96-115, 2008.

- [11] B.-J. Chang, Y.-L. Chen e C.-M. Chou, Adaptive hierarchical polling and cost-based call admission control in IEEE 802.16 WiMAX networks. Em *Proceedings of the Wireless Communications and Networking Conference*, pp. 1954-1958, 2007.
- [12] B. Rong *et alli*, Call Admission Control Optimization in WiMAX Networks *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pg. 2509-2522, 2008.
- [13] Cable Television Labs Inc. *Data Over Cable Service Interface Specifications - Radio Frequency Interface Specification*. SP-RFIV2.0.
- [14] C. Eklund, R.B. Marks, K.L. Stanwood e S. Wang, IEEE Standard 802.16: A Technical Overview of the WirelessMAN Air Interface for Broadband Wireless Access. *IEEE Communications Magazine*, vol. 40, no. 6, pg. 98-107, 2002.
- [15] C.-L. Chen, J.-W. Lee, C.-Y. Wu e Y.-H. Kuo, Fairness and QoS Guarantees of WiMAX OFDMA Scheduling with Fuzzy Controls. *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, Article ID 512507, 14 pages, 2009.
- [16] D.-J. Deng, L.-W. Chang, C.-H. Ke, Y.-M. Huang e J. M. Chang, Delay constrained uplink scheduling policy for rtPS/ertPS service in IEEE 802.16e BWA systems *International Journal of Communication Systems*, vol. 22, no. 2, pg. 119-133, 2008.
- [17] D. Tarchi, R. Fantacci, and M. Bardazzi, Quality of Service management in IEEE 802.16 wireless metropolitan area networks. Em *Proceedings of the IEEE International Conference on Communications*, pg. 1789-1794, 2006
- [18] D. Niyato e E. Hossain, QoS-aware bandwidth allocation and admission control in IEEE 802.16 broadband wireless access networks: A non-cooperative game theoretic approach. *Computer Networks*, vol. 55, no. 11, pg. 3305-3321, 2007.
- [19] F. Hou, J. She, P.-H. Ho e X. Shen, Performance Analysis of Weighted Proportional Fairness Scheduling in IEEE 802.16 Networks. Em *Proceedings of the IEEE International Conference on Communications*, pg. 3452-3456, 2008.
- [20] G. Chu, D. Wang e S. Mei, A QoS Architecture for the MAC Protocol of IEEE 802.16 BWA System. Em *Proceedings of the IEEE International Conference on Communications, Circuits and Systems and West Sino Expositions*, pg. 435-439, 2002.
- [21] H. Gowda *et alli*, A slot allocation mechanism for diverse QoS types in OFDMA based IEEE 802.16e systems, Em *Proceedings of the 9th International Conference on Advanced Communication Technology*, pg. 13-17, 2007.

- [22] H. Lee, T. Kwon e D. Cho, An Enhanced Uplink Scheduling Algorithm Based on Voice Activity for VoIP Services in IEEE 802.16d/e System. *IEEE Communications Letters*, vol. 9, no. 8, pg. 691-693, 2005.
- [23] H. J. Zhu, R. H. M. Hafez, Scheduling schemes for multimedia service in wireless OFDM systems *IEEE Wireless Communications*, vol. 14, no. 5, 2007.
- [24] H. Wang, B. He, and D. P. Agrawal, Above packet layer level admission control and bandwidth allocation for IEEE 802.16 wireless MAN. *Simulation Modeling Practice and Theory*, vol. 15, no. 14, pg. 266-382, 2007.
- [25] Intel Corporation, IEEE 802.16 and WiMAX. White Paper disponível em http://www.intel.com/netcomms/events/downloads/ieee80216_wp.pdf, 2003.
- [26] J. Chen *et al.* The design and implementation of WiMAX Module for ns-2 simulator. Em *Proceedings of the Workshop on ns-2: the IP network*, paper nº 5, 2006.
- [27] J. Chen, W. Jiao e H. Wang, A Service Flow Management Strategy for IEEE 802.16 Broadband Wireless Access Systems in TDD Mode. Em *Proceedings of the IEEE ICC*, pp. 3422-3426, 2005.
- [28] J. Chen e J.-Y. Wu, A Downlink Delay-Minimized Scheduling Scheme for OFDMA WiMAX Systems. Em *Proceedings of the Tenth International Conference on Mobile Data Management: Systems, Services and Middleware*, pg. 459-464, 2009.
- [29] J. F. Borin, and N. L. S. Fonseca, *Scheduler for IEEE 802.16 networks*. *IEEE Communications Letters*, vol. 12, no. 4, pp. 274-276, 2008.
- [30] J. F. Borin, and N. L. S. Fonseca, *A Standard-Compliant Scheduler for WiMAX Networks*. In *Proceedings of the 4th IEEE Broadband Wireless Access Workshop*, pp. 1-5, 2008.
- [31] J.F. Borin, and N.L.S. Fonseca, *Simulator for WiMAX networks*. *Simulation Modeling Practice and Theory*, vol. 16, no. 7, pp. 817-833, 2008.
- [32] J. Huang, V. G. Subramanian, R. Berry e R. Agrawal, Joint Scheduling and Resource Allocation in Uplink OFDM Systems. Em *Proceedings of the 41st Asilomar Conference on Signals, Systems and Computers*, pg. 265-269, 2007.
- [33] J. Lee, J. Lee, T. Kim e H. Kim, A radio resource scheduling algorithm considering wireless channel characteristics in IEEE 802.16 OFDMA systems, Em *Proceedings of the Wireless Telecommunications Symposium*, pg. 1-4, 2009.

- [34] K. Wongthavarawat e A. Ganz, IEEE 802.16 based last mile broadband wireless military networks with Quality of Service support. Em *Proceedings of the IEEE MILCOM'03*, pp. 779-784, 2003.
- [35] L. Wang, F. Liu, Y. Ju, e N. Ruangchaijatupon, Admission control for non-preprovisioned service flow in wireless metropolitan area networks. Em *Proceedings of the Fourth European Conference on Universal Multiservice Networks*, pg. 243-249, 2007.
- [36] L. Wan, W. Ma, Z. Guo, A Cross-layer Packet Scheduling and Subchannel Allocation Scheme in 802.16e OFDMA System. Em *Proceedings of the Wireless Communications and Networking Conference*, pg. 1865-1870, 2007.
- [37] M. Hawa e D. W. Petr, Quality of Service Scheduling in Cable and Broadband Wireless Access Systems. Em *Proceedings of the 10th IEEE International Workshop on Quality of Service*, pp. 247-255, 2002.
- [38] *OPNET*. <http://www.opnet.com/WiMAX/index.html>.
- [39] O. Yang e J. Lu, A new scheduling and CAC scheme for real-time video application in fixed wireless networks. Em *Proceedings of the 3rd IEEE CCNC*, pp. 303-307, 2006.
- [40] P. Brady, A model for generating On-Off speech patterns in Two-Way conversations. *Bell System Tech. Journal*, vol. 48, pp. 2445-2472, 1969.
- [41] P. Barford et al, *Changes in web client access patterns: Characteristics and caching implications*. Tech. Report 1998-023, Boston Uni., 1998
- [42] P. Seeling, M. Reisslein e B. Kulapala, Network Performance Evaluation Using Frame Size and Quality Traces of Single-Layer and Two-Layer Video: A Tutorial. *IEEE Communications Surveys and Tutorials*, vol. 6, no. 2, pp. 58-78, 2004.
- [43] M. F. Alam, M. Atiquzzaman e M. A. Karim, Traffic shaping for MPEG video transmission over the next generation internet. *Computer Communications*, vol. 23, no. 14, pp. 1336-1348, 2000.
- [44] M. Gidlund e G. Wang, Uplink Scheduling Algorithms for QoS Support in Broadband Wireless Access Networks. *Journal of Communications*, vol. 4, no. 2, pg. 133-142, 2009.
- [45] M. Masri, S. Abdellatif e G. Juanole, An Uplink Bandwidth Management Framework for IEEE 802.16 with QoS Guarantees. *Lecture Notes in Computer Science*, vol. 5550, pg. 651-663, 2009.

- [46] N. Shrivastav, *A Network Simulator Model of the DOCSIS Protocol and a Solution to the Bandwidth-hog Problem in the Cable Networks*. Dissertação de Mestrado, North Carolina State University, Raleigh, NC, EUA, 2003.
- [47] R. Jain. *The Art of Computer Systems Performance Analysis*. John Wiley & Sons, Inc., 1991.
- [48] R. Rouil, *The Network Simulator ns-2 NIST add-on - IEEE 802.16 model (MAC+PHY)*. <http://www.antd.nist.gov/seamlessandsecure>, 2007.
- [49] S. Chandra e A. Sahoo, An efficient call admission control for IEEE 802.16 networks. Em *Proceedings of the 15th IEEE Workshop on Local & Metropolitan Area Networks*, pg. 188-193, 2007.
- [50] S. Kim e I. Yeom, TCP-Aware Uplink Scheduling for IEEE 802.16. *IEEE Communications Letters*, vol. 11, no. 2, pg. 146-148, 2007.
- [51] *The Network Simulator - ns-2*. <http://www.isi.edu/nsnam/ns/>.
- [52] V. Singh e V. Sharma, Efficient and fair scheduling of uplink and downlink in IEEE 802.16 OFDMA networks. Em *Proceedings of the Wireless Communications and Networking Conference*, pg. 984-990, 2006.
- [53] X. Bai, A. Shami e Y. Ye, Robust QoS control for single carrier PMP mode IEEE 802.16 systems. *IEEE Transactions on Mobile Computing*, vol. 7, no. 4, pg. 416-429, 2008.
- [54] X. Guo, W. Ma, Z. Guo e Z. Hou, Dynamic bandwidth reservation admission control scheme for the IEEE 802.16e broadband wireless access systems. Em *Proceedings of the Wireless Communications and Networking Conference*, pg. 3420-3425, 2007.
- [55] X. Zhu *et alli*, QoS-Guaranteed Scheduling and Resource Allocation Algorithm for IEEE 802.16 OFDMA System, Em *Proceedings of the International Conference on Communications*, pg. 3463-3468, 2008.
- [56] W.-H. Park, S. Cho e S. Bahk, Scheduler Design for Multiple Traffic Classes in OFDMA Networks Em *Proceedings of the International Conference on Communications*, pg. 790-795, 2006.
- [57] *WiMAX Forum*. <http://www.wimaxforum.org>.
- [58] *WiMAX module for the ns-2 simulator*. http://www.lrc.ic.unicamp.br/wimax_ns2.