

**José Coelho de Melo Filho**

**MECANISMOS DE CONTROLE DE QUALIDADE  
DE SERVIÇO EM REDES IEEE 802.11**

Instituto de Matemática - Mestrado em Informática

Prof<sup>ª</sup>. Luci Pirmez,  
D.Sc. - COPPE/UFRJ - Brasil - 1996

Prof. José Ferreira de Rezende,  
Docteur, Université Pierre et Marie-Curie - França - 1997

RIO DE JANEIRO

2003

**MECANISMOS DE CONTROLE DE QUALIDADE DE SERVIÇO EM  
REDES IEEE 802.11**

**José Coelho de Melo Filho**

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO  
INSTITUTO DE MATEMÁTICA DA UNIVERSIDADE FEDERAL  
DO RIO DE JANEIRO - UFRJ, COMO PARTE DOS REQUISITOS  
NECESSÁRIOS À OBTENÇÃO DO GRAU DE MESTRE.

Aprovada por:

---

Prof<sup>ta</sup> Luci Pirmez,  
D.Sc. COPPE/UFRJ

---

Prof. José Ferreira de Rezende,  
Docteur, Université Pierre et Marie Curie

---

Prof. Markus Endler,  
Dr. rer. nat. Technical University Berlin

---

Luís Henrique Maciel Kosmalski Costa,  
Docteur, Université Pierre et Marie Curie

---

Prof. Oswaldo Vernet de Souza Pires,  
D.Sc. COPPE/UFRJ

RIO DE JANEIRO, RJ - BRASIL

JANEIRO DE 2003

MELO FILHO, JOSÉ COELHO

Mecanismos de Controle de Qualidade de Serviço em Redes  
IEEE 802.11 [Rio de Janeiro] 2003

xi, 90 p. 29,7 cm (IM/NCE/UFRJ, MSc., Informática,  
2003)

Dissertação (Mestrado) - Universidade Federal do Rio de  
Janeiro, IM/NCE

1. Redes de Computadores 2. Comunicação Sem Fio 3. QoS

I. IM/NCE/UFRJ II. Título (série)

*À minha esposa e filhas, pela compreensão e carinho prestados enquanto me dediquei a este trabalho.*

# Agradecimentos

A Deus, por ter me dado oportunidade para iniciar e saúde para concluir este trabalho.

Ao professor Luiz Fernando Rust, pelo acolhimento cordial em seu grupo de pesquisa no início do curso de mestrado.

À professora Luci Pirmez, por sua zelosa orientação acadêmica e pelo continuado interesse no progresso do meu trabalho de pesquisa.

Ao professor José Ferreira de Rezende, por sua dedicação, disponibilidade e exemplo, e pela valiosa condução em uma área de pesquisa ainda pouco conhecida para o nosso grupo. Ao pessoal do Grupo de Teleinformática e Automação (GTA) da COPPE/Elétrica, do qual faz parte o professor Rezende, pelo compartilhamento de equipamentos e pela troca de conhecimento.

Aos companheiros do Laboratório de Redes do NCE (LabNet), Werner, Reinaldo, Renata, Ana Paula, Cecílio, Flávia, Noel, Leonardo e Alexandre, pelos momentos tanto de alegria quanto de dificuldades compartilhados e pela inabalável disposição em ajudar.

Ao Instituto de Matemática e ao Núcleo de Computação Eletrônica, pelas instalações e equipamentos utilizados e pela excelência do ensino.

Ao Instituto Militar de Engenharia e ao Exército Brasileiro, pela liberação em tempo integral para o desenvolvimento deste trabalho e pelo estímulo ao constante aperfeiçoamento de seus profissionais.

Resumo da Tese apresentada ao IM/NCE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## **MECANISMOS DE CONTROLE DE QUALIDADE DE SERVIÇO EM REDES IEEE 802.11**

**José Coelho de Melo Filho**

Janeiro/2003

Orientadores: Luci Pirmez

José Ferreira de Rezende

Departamento: Informática

O padrão IEEE 802.11, criado pelo *Institute of Electrical and Electronics Engineers - IEEE* para as redes locais sem fio (*Wireless Local Area Networks - WLANs*), constitui uma das tecnologias mais promissoras entre as opções de acesso móvel aos serviços de dados das redes fixas, como a Internet. Esse padrão é estendido com a especificação 802.11e, desenvolvida com o intuito de introduzir nas WLANs um suporte à Qualidade de Serviço (QoS), importante para as aplicações multimídia e para as de tempo real. A extensão 802.11e cria uma nova funcionalidade de QoS, mas o gerenciamento da qualidade de serviço está fora do escopo da especificação. Este trabalho propõe dois mecanismos de controle para complementar a funcionalidade de QoS: um mecanismo de controle de admissão, adaptado de estudos realizados em redes fixas, e um mecanismo de controle de carga, que seleciona dinamicamente os parâmetros de diferenciação de serviços usados no método de acesso com contenção. Os mecanismos propostos contribuem para uma melhor utilização dos recursos da rede, ao mesmo tempo em que mantêm a diversidade de serviços. São realizadas simulações a fim de avaliar o desempenho dos dois mecanismos.

Abstract of Thesis presented to IM/NCE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

**QUALITY OF SERVICE CONTROL MECHANISMS FOR IEEE  
802.11 NETWORKS**

**José Coelho de Melo Filho**

January/2003

Advisors: Luci Pirmez

José Ferreira de Rezende

Department: Informatics

The IEEE 802.11 standard, created by the Institute of Electrical and Eletronics Engineers for the Wireless Local Area Networks (WLANs), is one of the most promising technologies among the alternatives for mobile access to data services provided by fixed networks, like the Internet. This standard is extended with the 802.11e specification, aiming to provide a QoS support in WLANs, which is mandatory for multimedia and real-time applications. The 802.11e extension creates a new QoS functionality, but QoS management issues are out of the scope of its specification. This work proposes two control mechanisms for QoS functionality enhancement: an admission control mechanism, adapted from fixed networks, and a load control mechanism, which dinamically selects the service differentiation parameters used in the contention-based access method. Both mechanisms contribute for a better network resource utilization while maintaining the service diversity. Simulations have been executed to evaluate the performance of both mechanisms.

# Lista de Figuras

2.1	Componentes de uma Rede 802.11 Infra-Estruturada . . . . .	15
2.2	Modos de Acesso do MAC 802.11 . . . . .	18
2.3	Esquema Básico de Acesso no DCF. . . . .	20
2.4	Acesso DCF com RTS / CTS. . . . .	21
2.5	Superquadro e Divisão de Períodos (CP e CFP). . . . .	23
2.6	Relações dos Intervalos Entre-Quadros do MAC 802.11. . . . .	23
2.7	Formas de Acesso Móvel à Internet. . . . .	26
3.1	Superquadro no MAC 802.11e . . . . .	44
3.2	MAC Proposto no EDCF . . . . .	45
3.3	Relações dos Intervalos Entre-Quadros no 802.11e . . . . .	45
3.4	Conjunto de Parâmetros de QoS do EDCF . . . . .	46
3.5	Mecanismo de Medição por Janela de Tempo . . . . .	52
4.1	Posicionamento do Controle de Carga . . . . .	58
4.2	Vazão de 1 Fonte CBR em Função do Tamanho de AIFS . . . . .	61
4.3	Vazão Agregada de 8 Fontes CBR em Função do Tamanho de AIFS . . . . .	61
4.4	Vazão Agregada de 16 Fontes CBR em Função do Tamanho de AIFS . . . . .	62
4.5	Vazão de 1 Fonte CBR em Função do Tamanho de CWMin . . . . .	62

4.6	Vazão Agregada de 8 Fontes CBR em Função do Tamanho de CWMin	63
4.7	Vazão Agregada de 16 Fontes CBR em Função do Tamanho de CWMin	63
4.8	Número Médio de Tentativas de Transmissão em Função do Tamanho de CWMin . . . . .	64
4.9	Diferenciação por Variação Conjunta de AIFS e CWMin . . . . .	65
4.10	Vazão de Duas Fontes de Categorias Distintas em Função da Sobreposição de Janelas . . . . .	66
4.11	Módulo Principal do Controle de Carga . . . . .	69
4.12	Ajuste de Contenção na Categoria de Tempo Real . . . . .	70
4.13	Ajuste da Carga da Categoria de Melhor Esforço . . . . .	71
4.14	Módulo Principal do Controle de Carga para $n$ Categorias . . . . .	74
5.1	Cenário usado nas simulações. . . . .	78
5.2	Carga do tráfego agregado de voz no sentido de descida, medida pelo estimador no AP em cada intervalo de amostra. . . . .	82
5.3	Retardo máximo dos pacotes de voz do tráfego de descida, medido no AP a cada intervalo de amostra. . . . .	83
5.4	Retardo máximo dos pacotes de voz do tráfego de subida, medido na 1ª STA a cada intervalo de amostra. . . . .	83
5.5	Distribuição do Retardo do Tráfego de Tempo Real . . . . .	88
5.6	Vazão Agregada do Tráfego FTP sem Controle de Carga . . . . .	88
5.7	Vazão Agregada do Tráfego FTP com Controle de Carga . . . . .	88
5.8	Retardo de Acesso do Tráfego de Voz sem Controle de Carga . . . . .	89
5.9	Retardo de Acesso do Tráfego de Voz com Controle de Carga . . . . .	89
5.10	Vazão Agregada do Tráfego FTP x Intervalo de Monitoração . . . . .	90

5.11 Retardo Médio do Tráfego de Voz x Intervalo de Monitoração . . . . .	90
5.12 Vazão Agregada do Tráfego FTP x Delta . . . . .	91
5.13 Retardo de Acesso do Tráfego de Voz x Delta . . . . .	91

# Lista de Tabelas

4.1	Valores de SIFS, SlotTime e DIFS usados no 802.11 . . . . .	56
5.1	Teste do Controle de Admissão . . . . .	81
5.2	Retardo de Acesso dos Quadros de Voz . . . . .	86
5.3	Vazão Média do Tráfego FTP . . . . .	87

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação . . . . .	1
1.2	Objetivos . . . . .	3
1.3	Trabalhos Relacionados . . . . .	5
1.4	Organização da Dissertação . . . . .	6
<b>2</b>	<b>Conceitos Básicos</b>	<b>7</b>
2.1	Redes Móveis . . . . .	8
2.2	Redes Sem Fio . . . . .	10
2.3	O Padrão IEEE 802.11 . . . . .	13
2.3.1	Arquitetura da Rede . . . . .	14
2.3.2	Camada Física . . . . .	15
2.3.3	Subcamada de Controle de Acesso ao Meio (MAC) . . . . .	17
	Função de Coordenação Distribuída - DCF . . . . .	18
	Função de Coordenação Centralizada - PCF . . . . .	22
	Suporte à Fragmentação . . . . .	24
2.4	802.11 em Cenários de Internet Móvel . . . . .	25
2.5	Considerações Finais . . . . .	27

<b>3</b>	<b>802.11 e Qualidade de Serviço</b>	<b>28</b>
3.1	Qualidade de Serviço . . . . .	29
3.1.1	Classificação das Aplicações . . . . .	30
3.1.2	Qualidade de Serviço na Internet . . . . .	32
3.1.3	Qualidade de Serviço em Redes Móveis . . . . .	33
3.2	Diferenciação de Serviços no MAC 802.11 . . . . .	34
3.2.1	Limitações do PCF e do DCF . . . . .	35
3.2.2	Técnicas de Diferenciação no Modo DCF . . . . .	38
	Variação do Tamanho do DIFS . . . . .	38
	Alteração no Procedimento de <i>Backoff</i> . . . . .	39
	Variação na Característica de Persistência do CSMA . . . . .	40
	Variação do Tamanho dos Quadros de Dados . . . . .	41
	Escopo da Diferenciação . . . . .	42
3.2.3	A Extensão 802.11e . . . . .	42
3.3	Necessidades Adicionais . . . . .	47
3.3.1	Determinação dos Parâmetros de QoS e Controle de Carga . . . . .	47
3.3.2	Controle de Admissão . . . . .	49
	Controle de Admissão Baseado em Medidas . . . . .	50
	Componentes do Controle de Admissão . . . . .	51
3.4	Considerações Finais . . . . .	53
<b>4</b>	<b>Mecanismos de Controle Propostos</b>	<b>54</b>
4.1	Adaptação do MBAC para o Caso Sem Fio . . . . .	55
4.2	Controle de Carga . . . . .	57

4.2.1	Posicionamento do Controle de Carga . . . . .	58
4.2.2	Influência dos Parâmetros AIFS e CWMin . . . . .	59
4.2.3	Sobreposição das Janelas de Contenção . . . . .	65
4.2.4	Restrições . . . . .	68
4.2.5	Algoritmos . . . . .	69
4.2.6	Configuração do Controle de Carga . . . . .	72
4.3	Considerações Finais . . . . .	73
<b>5</b>	<b>Simulações e Resultados</b>	<b>75</b>
5.1	Ambiente de Simulação . . . . .	75
5.1.1	Preparação do Ambiente de Simulação . . . . .	76
5.1.2	Cenário de Simulação . . . . .	77
5.2	Avaliação do Controle de Admissão . . . . .	78
5.2.1	Modelo de Tráfego . . . . .	79
5.2.2	Configuração do MBAC . . . . .	80
5.2.3	Resultados . . . . .	80
5.3	Avaliação do Controle de Carga . . . . .	83
5.3.1	Modelos de Tráfego . . . . .	83
5.3.2	Configuração do Controle de Carga . . . . .	84
5.3.3	Resultados . . . . .	85
5.3.4	Variações do Controle de Carga . . . . .	89
	Intervalo de Monitoração . . . . .	89
	Sobreposição de Janelas ( $\delta$ ) . . . . .	91
5.4	Considerações Finais . . . . .	92

<i>SUMÁRIO</i>	xv
<b>6 Conclusões e Trabalhos Futuros</b>	<b>94</b>
6.1 Conclusões . . . . .	95
6.2 Sugestões para Trabalhos Futuros . . . . .	97
<b>Referências Bibliográficas</b>	<b>99</b>

# Capítulo 1

## Introdução

### 1.1 Motivação

Os anos 90 foram marcados pela explosão do número de usuários tanto da Internet quanto dos sistemas celulares ao redor do mundo, saltando da ordem de centenas de milhares no início da década para centenas de milhões nos dias atuais. Isso provocou uma explosão da procura por serviços tanto de acesso a dados quanto de comunicações pessoais. As necessidades desses usuários apontam para serviços multimídia (voz, dados e imagem) funcionais e de alta qualidade, associados a um sistema de acesso com elevada disponibilidade. O que se vislumbra é uma Internet Móvel que, convergindo os serviços de comunicação de voz e dados conhecidos hoje, e empregando técnicas de transmissão digitais e comutadas por pacotes, seja capaz de realizar o ideal do acesso a qualquer hora e em qualquer lugar, conhecido como acesso ubíquo.

Possibilitar Internet Móvel com acesso ubíquo é uma tarefa que traz enormes desafios para as redes de computadores, pois envolve:

- desenvolvimento de aplicações e protocolos novos, preservando alguma compatibilidade com os já existentes;
- desenvolvimento de terminais de acesso com grande portabilidade e versatilidade;
- suporte aos grandes volumes de tráfego gerados pelas aplicações multimídia;

- suporte à Qualidade de Serviço (*Quality of Service - QoS*), requerida pelas aplicações multimídia e de tempo real;
- suporte amplo à mobilidade, dada a necessidade dos usuários de permanecerem conectados à rede enquanto se movem;
- utilização de comunicações sem fio confiáveis, capazes de possibilitar o acesso móvel e onipresente, com integração de sistemas heterogêneos e convergência tecnológica.

Diversas tecnologias de comunicação sem fio podem ser empregadas (e combinadas) para atender a essas necessidades, como as redes locais sem fio, os sistemas celulares com serviços de transmissão de dados e os sistemas que usam transmissão digital via satélite. O padrão do IEEE (*Institute of Electrical and Electronics Engineers*) para as redes locais sem fio (*Wireless Local Area Networks - WLANs*), conhecido como IEEE 802.11 [1], é uma dessas tecnologias. Esse padrão se destaca das demais opções de acesso móvel por sua elevada popularidade, pelas altas taxas de transmissão de dados que oferece e pelo custo relativamente baixo. O padrão permite o estabelecimento tanto de redes sem infra-estrutura, que utilizam apenas estações sem fio e são denominadas redes *Ad Hoc*, quanto de redes infra-estruturadas, nas quais um nó especial, chamado ponto de acesso, fornece às estações sem fio o acesso à rede fixa.

O 802.11 está cada vez mais presente nas redes de computadores das universidades de países desenvolvidos, oferecendo a seus usuários uma contínua possibilidade de acesso à rede a partir de qualquer ponto do *campus*. O padrão também é empregado por provedores privados que distribuem pontos de acesso em locais de grande concentração de usuários, como as salas de espera de aeroportos, centros de convenções e hotéis. Esses pontos de concentração são conhecidos como *hotspots*. Uma vez dentro da área de cobertura de um desses provedores, o usuário tem a possibilidade de alugar uma interface - ou utilizar a que esteja disponível em seu dispositivo móvel - para fazer a conexão com a rede, pagando uma taxa de acesso proporcional ao tempo de conexão ou ao volume de dados trafegados. O preço das interfaces de redes 802.11, atualmente abaixo de cem dólares americanos, permite assumir que essas interfaces estarão disponíveis em uma parte considerável dos dispositivos móveis utilizados, sejam eles computadores portáteis, assistentes pessoais digitais (*Personal*

*Digital Assistants - PDAs*) ou outros quaisquer que venham a ser desenvolvidos.

Porém, a ausência de um suporte à QoS para as aplicações multimídia e de tempo real no 802.11 diminui o seu valor como opção de acesso sem fio. Essa deficiência crítica é corrigida pelo IEEE com o desenvolvimento da especificação 802.11e [2], denominada *MAC Enhancements for Quality of Service*, como uma extensão à especificação original. Tal extensão adiciona uma nova função de coordenação para regular o acesso das estações ao meio sem fio, chamada *Hybrid Coordination Function - HCF*. A função HCF acrescenta dois novos métodos de acesso aos já disponíveis na subcamada de controle de acesso ao meio (MAC) das redes locais sem fio, sendo um com contenção (*Enhanced Distribution Coordination Function - EDCF*) e outro sem contenção (*HCF polled channel access*), ambos com funcionalidades de QoS. Esses novos métodos permitem aplicar tratamento diferenciado a categorias de tráfego com requisitos distintos de QoS.

No 802.11e, o controle da funcionalidade de QoS tanto no acesso sem contenção quanto no acesso com contenção é responsabilidade de um nó especial chamado coordenador híbrido (*Hybrid Coordinator - HC*). Esse coordenador é geralmente o próprio ponto de acesso, que passa a exercer essa função em adição àquelas que já executa em sua área de cobertura e torna-se capaz de regular o acesso ao meio sem fio seguindo regras de diferenciação de serviços e prioridades pré-estabelecidas. No novo método de acesso com contenção (EDCF), em particular, o coordenador híbrido pode emitir quadros de controle contendo um conjunto de parâmetros de QoS (*QoS Parameter Set* ou *EDCF Parameter Set*), usado pelas estações móveis para configurar as regras de acesso ao meio na subcamada MAC de suas interfaces de rede. Dessa forma, o coordenador é capaz de alterar a carga imposta à rede pelas estações, adaptando-se a mudanças nas condições de tráfego e estabelecendo ou modificando as políticas de acesso.

## 1.2 Objetivos

A extensão 802.11e cria o suporte à Qualidade de Serviço nas redes locais sem fio, mas não define como esse suporte deve ser utilizado. No acesso com contenção,

por exemplo, fica em aberto a questão de como o conjunto de parâmetros de QoS deve ser composto para cada situação particular de carga e quando esse conjunto deve ser alterado.

Outras questões relacionadas ao gerenciamento de Qualidade de Serviço de forma geral também estão fora do escopo dessa especificação. Isso motiva o desenvolvimento de novos mecanismos, voltados para a monitoração e o controle dos níveis de serviço, complementando a funcionalidade de QoS introduzida com a extensão 802.11e. Esses mecanismos são importantes para garantir a eficiência no uso dos recursos de comunicação disponíveis na rede e, simultaneamente, permitir o estabelecimento e a manutenção dos níveis de Qualidade de Serviço adequados aos requisitos das diversas aplicações presentes.

O trabalho descrito nesta dissertação aborda o desenvolvimento de alguns desses mecanismos, com ênfase particular em duas questões: o controle de admissão do tráfego de tempo real (*soft real-time*), e o controle da carga imposta pelo tráfego de melhor esforço no acesso com contenção (EDCF). O foco são as redes locais sem fio infra-estruturadas, usadas para fornecer acesso móvel às redes fixas, como a Internet.

O controle de admissão tem o propósito de evitar que a carga imposta pelo tráfego de tempo real seja superior à capacidade máxima da rede, para evitar que as garantias de QoS sejam perdidas nas situações de sobrecarga.

O controle de carga tem como objetivo forçar o tráfego de melhor esforço a ceder o acesso ao meio para o tráfego de tempo real nas situações de sobrecarga da rede. Ao mesmo tempo, busca permitir que o tráfego de melhor esforço explore ao máximo a largura de banda não utilizada pelo tráfego de tempo real. Para atingir esses objetivos, o controle de carga busca capacitar o ponto de acesso 802.11e a selecionar dinamicamente os parâmetros de QoS a serem empregados pelas estações móveis, durante o acesso com contenção, de acordo com a situação de carga presente na rede a cada instante.

## 1.3 Trabalhos Relacionados

O controle de admissão é uma questão amplamente estudada nos cenários de redes fixas IP e ATM [3, 4, 5, 6]. Poucos trabalhos, no entanto, trazem para o ambiente sem fio a experiência adquirida com esses estudos. Alguns trabalhos desenvolvem esquemas de controle de admissão específicos para o caso sem fio como, por exemplo, o modelo de rede SWAN (*Stateless Wireless Ad Hoc Networks*), apresentado por Ahn *et al* [7]. Esse modelo propõe um controle de admissão distribuído, baseado nas fontes geradoras de tráfego e idealizado para redes *Ad Hoc*. Embora possa ser adaptado para uso em redes infra-estruturadas, como é o caso das redes de acesso em um cenário de Internet Móvel, o controle de admissão do modelo SWAN adiciona às estações móveis o esforço computacional relativo à função de admissão, o que nem sempre é desejável.

O controle de admissão descrito nesta dissertação busca adaptar ao caso sem fio os esquemas desenvolvidos para as redes fixas (baseadas em IP ou ATM), mantendo a complexidade relacionada às operações de controle restrita ao ponto de acesso 802.11.

O modelo de rede SWAN também propõe um mecanismo de controle de carga que, a exemplo do controle de admissão, é executado nas estações, de forma distribuída. Esse mecanismo desconsidera modificações importantes na subcamada MAC, como a nova funcionalidade de QoS criada pela extensão 802.11e. Em um outro trabalho, Romdhani *et al* [8] apresentam um esquema chamado AEDCF (*Adaptive Enhanced Distributed Coordination Function*), que considera o novo método de acesso com contenção do 802.11e e o torna mais adaptativo, obtendo bons resultados com relação à eficiência na utilização do canal. Esse esquema, no entanto, também foi idealizado para redes *Ad Hoc* e para execução distribuída.

No controle de carga descrito nesta dissertação, o ponto de acesso toma as decisões de controle de forma centralizada, baseado na informação que detém sobre os fluxos que entram e saem da rede sem fio e nas restrições de cada tipo de tráfego. O mecanismo utilizado para transmitir às estações os novos parâmetros de acesso é o próprio (*QoS Parameter Set*) criado pela extensão 802.11e. Assim, o controle busca

promover a diferenciação de serviços sem comprometer a eficiência na utilização do canal e sem adicionar complexidade às estações móveis, exigindo dessas estações apenas a compatibilidade com a extensão 802.11e.

## 1.4 Organização da Dissertação

A presente dissertação está estruturada em seis capítulos. O capítulo 2 relaciona conceitos básicos sobre redes móveis e sem fio e sobre o padrão 802.11. No capítulo 3 são apresentados conceitos sobre Qualidade de Serviço em geral e são descritas as principais técnicas elaboradas para a provisão de QoS em sistemas 802.11, as quais culminaram com a elaboração da extensão 802.11e. O capítulo 4 descreve os mecanismos de controle de admissão e de controle de carga propostos. O capítulo 5 apresenta simulações realizadas para verificar o funcionamento dos mecanismos e avaliar a viabilidade da proposta. Finalmente, o capítulo 6 traz conclusões e sugestões para trabalhos futuros.

# Capítulo 2

## Conceitos Básicos

O UNIVERSO das redes móveis tem características específicas que as distinguem das redes de computadores tradicionais, compostas por nós fixos. A mobilidade impõe o uso de tecnologias de comunicações de dados sem fio e de equipamentos de acesso desenvolvidos com um maior grau de miniaturização. A localização do usuário adquire uma dimensão temporal que deve ser tratada pela própria rede. Problemas relacionados à segurança, interferência e consumo de energia crescem de importância nos cenários móveis.

Este capítulo aborda os conceitos básicos relacionados às redes móveis e à comunicação sem fio, com particular enfoque sobre o padrão 802.11, objeto de estudo do trabalho descrito nesta dissertação. A Seção 2.1 traz considerações sobre redes móveis em geral. A Seção 2.2 apresenta características das redes sem fio e lista alguns dos padrões e tecnologias disponíveis. A Seção 2.3 fornece uma visão geral da especificação do 802.11 e de suas extensões 802.11a e 802.11b. A Seção 2.4 exemplifica como o 802.11 pode ser utilizado em sistemas de acesso sem fio com ampla cobertura, idealizados como suporte à Internet Móvel. Por último, a Seção 2.5 traz as considerações finais do capítulo.

## 2.1 Redes Móveis

As redes móveis são aquelas nas quais um subconjunto dos nós que a compõem mudam de localização durante o seu funcionamento. Normalmente, existe uma parte da rede que é fixa, compondo uma infra-estrutura de acesso e interligação com outras redes, enquanto outra parte é móvel, composta pelos terminais de acesso empregados pelos usuários. Este é o caso das redes de telefonia móvel celular e das redes locais sem fio interligadas a redes fixas padrão *Ethernet*. O equipamento do usuário é chamado, de forma geral, de estação móvel. O equipamento que integra a parte fixa com a parte móvel, e que portanto fornece o acesso às estações móveis, é chamado de ponto de acesso (nas redes locais sem fio) ou estação-base (nos sistemas celulares).

Existem ainda redes em que todos os nós que a compõem são móveis, não havendo uma infra-estrutura identificável e uma topologia estaticamente estabelecida. A transmissão de dados entre um nó de origem e um de destino é feita de forma colaborativa, utilizando os nós intermediários, como nas redes fixas. A diferença é que esses nós podem também se mover durante a transmissão, causando falhas de rotas freqüentes, o que provoca a necessidade de emprego de protocolos de roteamento específicos. Essas redes são chamadas de Redes Móveis *Ad Hoc*, ou MANETs (*Mobile Ad Hoc Networks*) [9], e constituem uma área de pesquisa muito explorada atualmente. Existe um grupo de trabalho sobre MANETs na IETF (*Internet Engineering Task Force*), que propõe padrões para as redes *Ad Hoc*, principalmente na área de roteamento. As redes *Ad Hoc* estão fora do escopo do trabalho descrito nesta dissertação.

Nas redes fixas, é possível associar o endereço de rede à localização do nó na rede, o que facilita o roteamento. Essa premissa de localização fixa foi usada na concepção do protocolo de rede utilizado na Internet, o IP - Internet Protocol [10], que emprega um endereçamento hierárquico dependente de localização, com endereços de 32 bits. Nas redes móveis, porém, isso não é viável, já que a localização dos nós é variável. Isso faz com que, em uma rede IP, um usuário móvel não obtenha acesso aos serviços da rede enquanto não reconfigurar sua interface para um endereço compatível com sua nova posição. Ele também não é localizável pelos demais enquanto não divulgar,

por algum meio, o seu novo endereço. No caso específico das redes IP, uma solução para contornar esse problema foi proposta em 1996 e é chamada de IP Móvel [11].

Cabe aqui fazer uma distinção entre estação móvel e estação portátil, segundo encontrado na literatura e na especificação do 802.11 [1]. A estação portátil é aquela que, por suas dimensões e peso reduzidos, pode ser desconectada, conduzida pelo usuário a partir de sua localização original e reconectada à rede ao chegar em uma nova localização. É o caso, por exemplo, de um computador pessoal portátil, utilizado pelo usuário para fazer o acesso à rede de sua empresa, podendo estar em casa, na própria empresa ou em uma rede externa. A estação portátil não se comunica durante o movimento, ou seja, ela usa a rede somente quando está parada em suas possíveis localizações. Por isso, ela pode até dispensar o uso de comunicação sem fio: o acesso à rede em cada localização pode ser feito tanto utilizando cabos para ligação direta com pontos da rede fixa quanto empregando interfaces de rede sem fio para se comunicar com os pontos de acesso disponíveis. Deve ser notado que a presença de estações portáteis em redes IP, embora não envolva comunicação durante a mudança de posição, já é o suficiente para gerar falhas de roteamento, devido à questão dos endereços dependentes de localização discutida acima, o que já motiva o emprego de técnicas como o IP Móvel.

Por outro lado, uma estação móvel, além de ser portátil, deve possuir a capacidade de não interromper qualquer conexão existente enquanto se move; isso envolve o uso obrigatório de comunicação sem fio e de mecanismos para transferência de conexões ativas quando a estação passa da área de cobertura de um ponto de acesso para a de outro. Esse procedimento de transferência é conhecido como *handoff*. Também se deseja que ela seja continuamente localizável durante o movimento, tornando necessário o uso de esquemas para gerência de localização. Um aparelho de telefonia móvel celular, por exemplo, deve satisfazer plenamente o conceito de estação móvel. Algumas considerações sobre gerência de localização e *handoff* em redes móveis de diferentes tecnologias podem ser encontradas em [12].

Além da mobilidade dos usuários e dos terminais de acesso, outros problemas como o consumo de bateria e a interface com o usuário são típicos de redes móveis. O primeiro deve-se ao fato de que a estação, justamente por ser móvel, fica

impossibilitada de receber um suprimento contínuo de energia a partir de fontes fixas, devendo carregar consigo seu próprio suprimento e garantir um uso eficiente do mesmo. O segundo está ligado à miniaturização dos equipamentos empregados, que faz com que o usuário tenha que lidar com elementos de interface (como o teclado e os sistemas de vídeo e de áudio) muitas vezes mais restritos do que os oferecidos pelos equipamentos de acesso fixo.

A segurança é outra questão-chave a ser considerada, uma vez que os sinais no meio sem fio podem ser captados por nós que não pertencem à estrutura lógico-administrativa da rede e que têm, portanto, uma possibilidade de acesso não autorizado.

Uma visão mais abrangente desses e de outros problemas relacionados às redes móveis pode ser encontrada no trabalho de G. Forman e J. Zahorjan [13].

## 2.2 Redes Sem Fio

As tecnologias de comunicações de dados sem fio permitem que os nós permaneçam conectados enquanto se movimentam, dispensando o uso de cabos. Estas tecnologias são também úteis nas situações em que as áreas a serem cobertas pela rede são de difícil acesso, ou ainda quando a instalação do cabeamento é economicamente inviável.

Uma vez que existem várias tecnologias disponíveis para comunicação sem fio, todas usando o ar como meio de transmissão, há necessidade de se regular as faixas do espectro de frequências que cada uma dessas tecnologias emprega. Existem faixas que são licenciadas e outras que são não-licenciadas. As licenciadas dependem de autorização de um órgão regulador dos serviços de telecomunicações de cada país para serem usadas, mediante pagamento de taxas. Este é o caso das faixas usadas pelos sistemas celulares. As não-licenciadas podem ser utilizadas sem pagamento de licença. As faixas específicas e as normas de uso podem variar de um país para outro.

Os padrões para comunicação de dados sem fio disponíveis atualmente têm carac-

terísticas técnicas bastante diversificadas e foram idealizados para atender a diferentes demandas. Esses padrões são criados por órgãos de padronização internacional da área de telecomunicações. O IEEE (*Institute of Electrical and Electronics Engineers*) é um desses órgãos, e possui três grupos de trabalhos (*Working Groups ou WGs*) dedicados a redes sem fio. Esse grupos são integrantes do Comitê de Padrões 802, que elabora as normas para redes locais (*Local Area Networks ou LANs*) e metropolitanas (*Metropolitan Area Networks ou MANs*). Os 3 grupos de trabalho são descritos a seguir:

a) Grupo de Trabalho 11 [14]: é o responsável pelo padrão 802.11, para as redes locais sem fio. As redes locais sem fio (*Wireless Local Area Networks - WLANs*), foram desenvolvidas para oferecer aos usuários móveis uma experiência de uso similar às das redes locais cabeadas. O alcance típico de uma WLAN varia de 100 a 500 metros, dependendo das condições de transmissão do ambiente. O 802.11 só usa faixas de frequência não-licenciadas. A especificação original [1] fornece taxas de transmissão de dados de 1 ou 2 Mbps. A extensão 802.11b [15] suporta taxas adicionais de 5,5 e 11 Mbps na faixa situada em torno de 2,4 GHz, chamada de "banda ISM" (*Industrial, Scientific and Medical*). A extensão 802.11a [15] oferece taxas que variam de 6 a 54 Mbps na faixa situada em torno de 5 GHz, conhecida como "banda U-NII" (*Unlicensed National Information Infrastructure*). Porém, o alcance de uma estação 802.11a é normalmente menor que o de uma 802.11b;

b) Grupo de Trabalho 15 [16]: atua na área das redes de área pessoal (*Wireless Personal Area Networks ou WPANs*). O conceito de WPAN foi criado para designar as redes de até dez metros de raio, que envolvem a interligação de dispositivos portáteis e de computação móvel, como computadores pessoais (PCs), periféricos, telefones celulares, pagers e dispositivos de eletrônica de consumo. Uma das principais tecnologias adotadas nesta área é o Bluetooth, criado pelo Bluetooth SIG (*Special Interest Group*) [17], um consórcio de empresas de telecomunicações e informática, e incorporado ao padrão 802.15 pelo Grupo de Tarefas (*Task Group*) 1 do Grupo de Trabalho 15. O Bluetooth utiliza a mesma banda ISM empregada pelo 802.11 e oferece taxas de até 1 Mbps. Outros grupos de tarefas dentro do WG 15 estudam tecnologias para WPANs de altas (TG3) e baixas taxas de transmissão

(TG4);

c) Grupo de trabalho 16 [18]: elabora as especificações para as redes metropolitanas sem fio (*Broadband Wireless Metropolitan Area Networks ou WirelessMANs*), objetivando a implantação de sistemas de acesso fixo em banda larga (*fixed BWA - Broadband Wireless Access*). O 802.16 utiliza tecnologias como o WLL (*Wireless Local Loop*) / LMDS (*Local Multipoint Distribution System*) para estabelecer sistemas de distribuição de serviços de voz, dados, Internet e vídeo em banda larga, usando uma arquitetura de rede similar à das redes celulares (porém para acesso fixo). Normalmente usa faixas licenciadas e situadas entre 10 e 66 GHz, mas também prevê o uso de faixas não-licenciadas<sup>1</sup>.

A lista de padrões acima não é exaustiva, pois engloba apenas o trabalho desenvolvido no IEEE. Outros padrões como, por exemplo, o HIPERLAN2, HIPERLINK e HIPERACCESS, desenvolvidos pelo *European Telecommunications Standards Institute - ETSI* [19], o WirelessATM, desenvolvido pelo *ATM Forum* [20] e os utilizados nas redes de comunicação de dados via satélites digitais não foram apresentados por questões de brevidade.

Igualmente importantes são os padrões para as Redes Celulares de 3<sup>a</sup> Geração [21], ou Redes 3G, evolução das redes celulares digitais atuais, consideradas de 2<sup>a</sup> Geração. As Redes 3G oferecem serviços de acesso a redes de dados (baseadas em IP ou X.25) utilizando técnicas de comutação de pacotes e valendo-se da infraestrutura dos próprios sistemas de telefonia móvel celular. Um exemplo de Rede 3G é o UMTS (*Universal Mobile Telecommunications System*) [22]. O sistema é subdividido em Rede de Acesso Rádio (*Radio Access Network - RAN*) e Rede de Núcleo (*Core Network*). O núcleo é baseado no protocolo IP e pode ser interligado a outras redes IP públicas ou privadas, como as redes IP corporativas (*Intranets*) e a própria Internet. As redes de acesso suportam taxas de transmissão de dados que vão de 144 Kbps a 2 Mbps, para usuários estacionários, em movimento lento ou embarcados em veículos. A adoção dos padrões para as Redes 3G envolve a substituição dos terminais de acesso e dos equipamentos de rede usados nas redes

---

<sup>1</sup>Apesar de a mobilidade não ter sido o foco inicial desse padrão, alguns esforços recentes vêm sendo desenvolvidos para possibilitar acesso móvel empregando o 802.16 (Mobile BWA).

de 2ª Geração, bem como a aquisição de novas licenças sobre faixas do espectro de frequências. Por esse motivo, alguns desenvolvimentos intermediários foram realizados com a intenção de acrescentar os serviços baseados em comutação de pacotes aproveitando a infra-estrutura já disponível. As redes que incorporam esses desenvolvimentos são comumente chamadas de Redes de "Geração 2,5" ou "Redes 2,5G". Exemplos dessas redes são o *GPRS - General Packet Radio Service* [23] e o *EDGE - Enhanced Data Rate for GSM Evolution* [24].

As Redes 2,5 e 3G podem ser consideradas como um primeiro esforço na direção da Internet Móvel e do acesso ubíquo, porque buscam tirar partido da ampla cobertura e do grande número de usuários alcançados pelos sistemas celulares ao redor do mundo.

Embora todos esses padrões tenham sido concebidos para satisfazer necessidades específicas de comunicação, o grau de desenvolvimento que alguns deles atingiram fez com que ocorresse uma sobreposição nas aplicações a que eles inicialmente se destinavam, o que pode sugerir que alguns prevalecerão em detrimento de outros. Porém, mais do que soluções simplesmente competidoras, esses padrões podem ser vistos como complementares, cada uma sendo capaz de atender a uma parcela das demandas futuras dos sistemas de comunicação.

## 2.3 O Padrão IEEE 802.11

Como já foi mencionado na seção anterior, o 802.11 foi criado pelo IEEE (*Institute of Electrical and Electronics Engineers*), no Grupo de Trabalho 11 do Comitê de Padrões 802. Os padrões da família 802 envolvem as camadas física e de enlace do Modelo Referencial Básico para Interconexão de Sistemas Abertos (ISO/IEC 7498-1: 1994) (*Open Systems Interconnection Basic Reference Model*), conhecido como Modelo OSI e criado pela ISO (*International Organization for Standardization*) [25].

No trabalho do comitê 802, a camada de enlace de dados foi subdividida em uma subcamada de controle do enlace lógico (*Logical Link Control - LLC*) e outra de controle de acesso ao meio (*Medium Access Control - MAC*). O LLC é descrito

na especificação 802.2 e é comum a todos os demais padrões da família 802. Esses outros padrões, por sua vez, especificam o nível físico e a subcamada MAC do nível de enlace. Assim foram geradas, por exemplo, as especificações para as redes *Ethernet* (802.3), *Token Bus* (802.4), *Token Ring* (802.5) e *DQDB* (802.6).

O 802.11 foi concebido para ser compatível com os demais padrões da família 802, ou seja, deve ser visto pelo LLC como qualquer outra rede 802. A versão atual da especificação foi editada em 1999 e adotada também pela ISO como padrão internacional no documento ISO/IEC 8802-11:1999. Conforme será visto a seguir, algumas extensões à especificação básica foram também elaboradas, como o 802.11a e 802.11b, que adicionam maiores capacidades ao nível físico, permitindo taxas de transmissão mais altas.

### 2.3.1 Arquitetura da Rede

No 802.11 [1], o Conjunto Básico de Serviço (*Basic Service Set* - BSS) é o bloco fundamental de construção da arquitetura. Um BSS é definido como um grupo de estações (STAs) que estão sob o controle direto de uma mesma função de coordenação (Seção 2.3.3). Tal função de coordenação determina quando cada estação pode transmitir e receber dados usando o meio sem fio. A área ocupada pelos membros de um BSS é chamada de BSA (*Basic Service Area*).

O padrão permite que estações sem fio possam se organizar compondo um BSS e comunicar-se umas com as outras diretamente [26], sem contar com nenhuma infra-estrutura, formando uma rede *Ad Hoc*. Nesse caso, o BSS é chamado de BSS independente (iBSS). Um iBSS pode ser criado com apenas duas estações 802.11.

Quando as estações dispõem de uma infra-estrutura para se comunicar com outros BSS fora de seu alcance ou com redes fixas, a rede 802.11 é chamada de infra-estruturada. Nessas redes, vários BSSs podem ser interligados através de um sistema de distribuição (*Distribution System* - DS), de modo a aumentar a cobertura da rede. São utilizados pontos de acesso (*Access Points* - AP) que interligam as estações sem fio de um BSS ao sistema de distribuição. De acordo com a especificação, o sistema de distribuição pode utilizar qualquer tecnologia de rede. O conjunto dos pontos

de acesso e dos sistemas de distribuição é definido com um Conjunto Estendido de Serviço (*Extended Service Set - ESS*). Dentro de um ESS, as estações podem se comunicar umas com as outras e se movimentar de um BSS para outro. Os serviços e funcionalidades fornecidos pelo ESS em sua área permitem essa mobilidade, e fazem o ESS parecer ao LLC como uma única rede. Esta pode ainda ser interligada a redes locais que seguem outros padrões (802.x) através de componentes chamados portais. A Figura 2.1 ilustra esses componentes. O foco do presente trabalho está nas redes 802.11 infra-estruturadas.

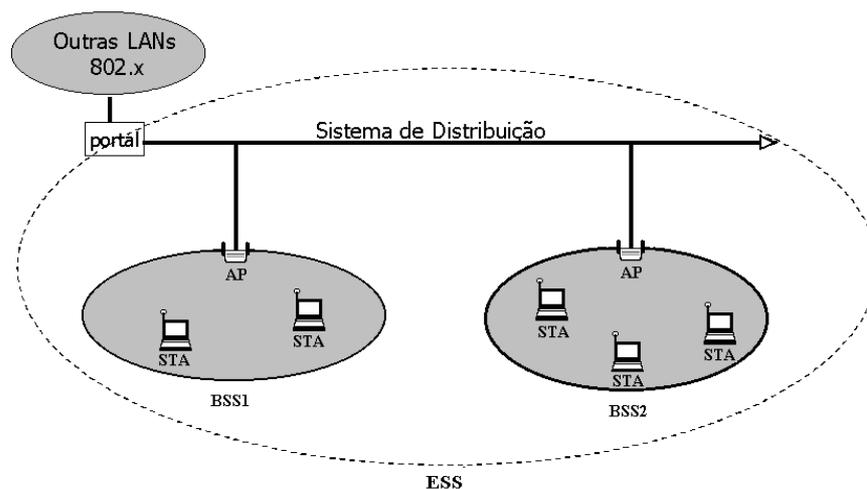


Figura 2.1: Componentes de uma Rede 802.11 Infra-Estruturada

### 2.3.2 Camada Física

O 802.11 define três tipos de camada física [1]: espalhamento de espectro por salto em frequências (*Frequency Hopping Spread Spectrum - FHSS*), espalhamento de espectro por sequência direta (*Direct Sequence Spread Spectrum - DSSS*) e infravermelho. Todas as camadas físicas do 802.11 incluem a provisão de um sinal de avaliação de canal livre (*Clear Channel Assessment signal - CCA*) que indica o estado atual do uso do meio sem fio. Esse sinal é utilizado pelo MAC (Seção 2.3.3) para verificar se o meio está livre. Cada camada física oferece uma taxa de 1 ou 2 Mbps.

O FHSS é uma técnica de espalhamento de espectro na qual a largura de banda

total disponível é dividida em vários canais de frequência; o transmissor e o receptor transmitem utilizando um desses canais por um certo tempo e depois saltam para outro canal, seguindo padrões pseudo-aleatórios de uso dos canais chamados de seqüências de saltos. Esse espalhamento proporciona maior segurança e menor susceptibilidade à interferência eletro-magnética. O FHSS usa a banda ISM (*Industrial, Scientific and Medical*) que vai de 2,4000 a 2,4835 GHz. Nos EUA e em quase toda a Europa, são definidos 79 canais. O primeiro canal tem uma frequência central de 2,402 GHz e os canais subseqüentes estão separados por 1 MHz. Cada canal possui uma banda de 1 Mbps. São definidos três diferentes conjuntos com 26 seqüências de saltos. Os três conjuntos existem para evitar períodos de colisões entre diferentes seqüências [26]. Com isso, permite-se a coexistência de até três redes (BSSs) em uma mesma área geográfica.

A taxa básica de 1 Mbps usa uma modulação gaussiana por chaveamento de frequência (*Gaussian Frequency Shift Keying - GFSK*) de dois níveis, na qual o dado passa por um filtro gaussiano em banda base e é modulado em frequência (um 1 lógico é codificado usando uma frequência  $F_c + f$  e um 0 lógico usa uma frequência  $F_c - f$ ). A taxa de acesso opcional de 2 Mbps usa um GFSK de quatro níveis, no qual dois bits são codificados por vez usando quatro frequências.

O DSSS é um outro método de espalhamento de espectro, no qual diferentes transmissões simultâneas são separadas por códigos e não por frequência como no FHSS. Nesse método, a duração de um símbolo de dados é subdividida em intervalos menores chamados *chips*. O DSSS também usa a banda ISM de 2,4 GHz. A taxa básica de 1 Mbps é gerada através de uma modulação diferencial binária por chaveamento de fase (*Differential Binary Phase Shift Keying - DBPSK*) e a taxa de 2 Mbps usa uma modulação diferencial quaternária por chaveamento de fase (*Differential Quadrature Phase Shift Keying - DQPSK*). O espalhamento é feito através da divisão da banda disponível em 11 subcanais, cada um com 11 MHz; cada símbolo de dados é espalhado usando uma seqüência de Barker de 11 *chips* (+1,-1,+1,+1,-1,+1,+1,+1,-1,-1,-1). Redes localizadas em uma mesma área devem usar canais cujas frequências centrais sejam separadas de pelo menos 30 Mhz, para evitar interferência mútua. Assim, de forma similar ao que ocorre no FHSS, no

máximo três redes podem coexistir em uma mesma área.

A especificação de infravermelho utiliza comprimentos de onda de 850 a 950 nm. O infravermelho foi projetado para ser usado em áreas fechadas. Opera com transmissões não direcionadas com alcance máximo de aproximadamente 10 metros, ou 20 metros caso sejam utilizados receptores mais sensíveis. A transmissão é sujeita à interferência causada por fontes de calor ou pela luz do Sol. As estações podem receber dados em suas linhas de visada e por transmissões refletidas. A codificação da taxa básica de 1 Mbps é realizada através de uma modulação por posição de pulso (*16-Pulse Position Modulation* - PPM), na qual quatro bits de dados são mapeados em 16 bits codificados para transmissão. A taxa opcional de 2 Mbps usa uma 4-PPM, na qual dois bits de dados são mapeados em 4 bits codificados para transmissão.

Os padrões 802.11a e 802.11b alteram a camada física do 802.11 para proverem taxas de transmissão mais altas. O padrão 802.11b [15] especifica, além das taxas de 1 e 2 Mbps, taxas de transmissão mais altas na banda de 2,4 GHz, de 5,5 e 11 Mbps, através do uso de um chaveamento de código complementar (*8-chip Complementary Code Keying* - CCK) no DSSS. Um protocolo de convergência de nível físico (*Physical Layer Convergence Protocol* - PLCP) permite que estações 802.11 e 802.11b se comuniquem utilizando as taxas de 1 e 2 Mbps. O padrão 802.11a [27] utiliza a banda de 5 GHz para prover taxas de 6 até 54 Mbps, utilizando uma técnica mais avançada de multiplexação, chamada multiplexação por divisão ortogonal em frequência (*Orthogonal Frequency Division Multiplexing* - OFDM). O sistema usa 52 subportadoras que são moduladas usando BPSK ou QPSK, modulação 16-QAM (*Quadrature Amplitude Modulation*) ou 64-QAM. As estações 802.11a não são compatíveis com as que atendem somente às especificações 802.11 ou 802.11b.

### 2.3.3 Subcamada de Controle de Acesso ao Meio (MAC)

O controle de acesso ao meio do 802.11 é baseado em funções de coordenação. Uma função de coordenação determina qual estação tem permissão para transmitir e receber dados utilizando o meio sem fio. A especificação 802.11 define duas funções

de coordenação: uma distribuída, conhecida como DCF (*Distributed Coordination Function*), de implementação obrigatória, que proporciona um acesso com contenção, e outra centralizada, conhecida como PCF (*Point Coordination Function*), de implementação opcional, que provê um acesso sem contenção. Essa última utiliza as regras de acesso da função distribuída para estabelecer os serviços de acesso sem contenção, como mostrado na Figura 2.2.

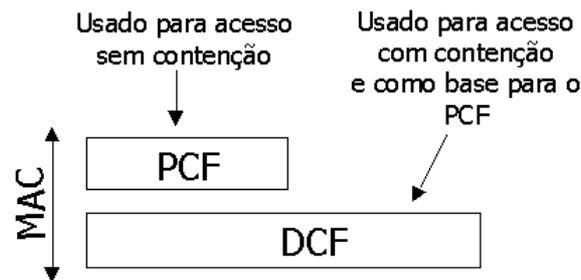


Figura 2.2: Modos de Acesso do MAC 802.11

### Função de Coordenação Distribuída - DCF

A Função de Coordenação Distribuída utiliza um protocolo de acesso múltiplo com detecção de portadora, conhecido como CSMA/CA (*Carrier Sense Multiple Access with Collision Avoidance*), semelhante ao CSMA/CD usado nas redes IEEE 802.3 (*Ethernet*). Nesse tipo de protocolo de acesso múltiplo, um nó que queira transmitir dados "escuta" o meio por um período fixo de tempo, antes de iniciar a transmissão, para detectar se ele está ou não ocupado. O nó transmite caso o meio não esteja ocupado durante esse período ou aguarda um tempo adicional aleatório no caso contrário. A diferença do CSMA/CA para o CSMA/CD é que a detecção de colisão não é realizada no meio sem fio. Além de supor que todas as estações sejam capazes de escutar a transmissão das demais, o que nem sempre é verdade em redes móveis, a detecção de colisão exigiria o projeto de equipamentos de rádio mais complexos (e mais caros), com capacidade de transmitir e receber ao mesmo tempo [28]. Ao invés disso, é utilizado um mecanismo de prevenção de colisão reforçado com o reconhecimento positivo (ACK) dos quadros de dados.

A detecção de portadora pode ser feita através de um mecanismo físico, usando o

sinal CCA descrito em 2.3.2, ou virtual. O mecanismo de detecção virtual, explicado mais adiante nesta Seção, utiliza um vetor de alocação da rede, localizado em cada nó, em conjunto com uma informação de duração da transmissão contida nos quadros de dados e em alguns quadros de controle.

O acesso utilizando a função DCF pode ainda ser realizado de duas formas distintas [1]: um esquema básico usando CSMA, de implementação obrigatória, e um esquema de acesso opcional que adiciona ao esquema básico o emprego de quadros de pedidos (*Request To Send - RTS*) e permissões (*Clear To Send - CTS*) para transmitir dados.

No esquema básico, uma estação que deseja transmitir algum quadro ouve o meio, detectando ou não a portadora. Caso o meio esteja livre durante um período de tempo chamado DIFS - *Distributed (Coordination Function) Interframe Space*, definido no padrão, a estação transmite o quadro de dados. O receptor, ao receber o quadro de dados, usa o método de verificação cíclica (CRC) para detectar erros no quadro e, caso ele esteja correto, envia um quadro de reconhecimento positivo (ACK). Esse quadro é enviado após um tempo chamado SIFS (*Short Interframe Space*), contado a partir do recebimento do quadro de dados. SIFS é sempre menor que DIFS, ou seja, a estação receptora ouve o meio por um tempo menor para enviar o ACK (Figura 2.3), ganhando a disputa sobre outras estações que queiram transmitir.

Caso o meio seja ocupado durante a espera inicial por DIFS, a estação transmissora adia a transmissão do quadro de dados e executa um procedimento de recuo exponencial binário (*Binary Exponential Backoff*). Nesse procedimento, a estação escolhe um valor aleatório, a partir de uma distribuição uniforme no intervalo entre zero e o tamanho de sua janela de contenção (*Contention Window - CW*). Na situação inicial,  $CW$  tem um tamanho mínimo, configurado pelo parâmetro ( $CW_{min}$ ). O valor escolhido é multiplicado pelo tempo de duração de um *slot*, denominado  $slottime^2$ , para criar um temporizador de *backoff*. Esse temporizador é decrementado a cada *slot*, caso não haja transmissão no meio. Se alguma transmis-

---

<sup>2</sup>Um *slottime* corresponde ao atraso máximo de propagação de ida e volta dentro de um BSS, incluindo ainda um tempo de processamento no receptor.

são for detectada, o temporizador é paralisado e só é reiniciado quando o meio ficar livre novamente por um tempo DIFS. Quando finalmente o temporizador expira, a estação envia o seu quadro. Dessa forma, o procedimento de *backoff* evita sucessivas colisões entre estações que estão tentando transmitir simultaneamente, pois distribui no tempo as tentativas de transmissão.

Caso o ACK não seja recebido no tempo esperado, a estação transmissora deduz que o quadro colidiu ou foi recebido com erro e escalona uma retransmissão entrando no procedimento de *backoff* descrito anteriormente. O valor de  $CW$  é dobrado e um novo temporizador é criado. A cada transmissão não sucedida, o valor de  $CW$  é aumentado para a próxima potência de 2 (menos 1 unidade), até que seja atingido um valor máximo pré-definido chamado  $CW_{max}$  (no 802.11b DSSS,  $CW_{min}$  é igual a 31 e  $CW_{max}$  é igual a 1023). Caso um número máximo de tentativas de transmissões seja alcançado (7, por *default*), o quadro é descartado e o nível superior é informado sobre a falha de transmissão.

O procedimento de *backoff* também é executado após cada transmissão bem sucedida de uma estação, para evitar que ela capture o meio, impedindo as demais de transmitir.

O esquema básico do DCF é apresentado na Figura 2.3. No 802.11b DSSS, a duração do SIFS (*sifstime*) é  $10\mu s$  e a do *slottime* é  $20\mu s$ . O DIFS é definido por  $SIFS + 2 * slottime$ , o que resulta em um tempo de espera de  $50\mu s$ .

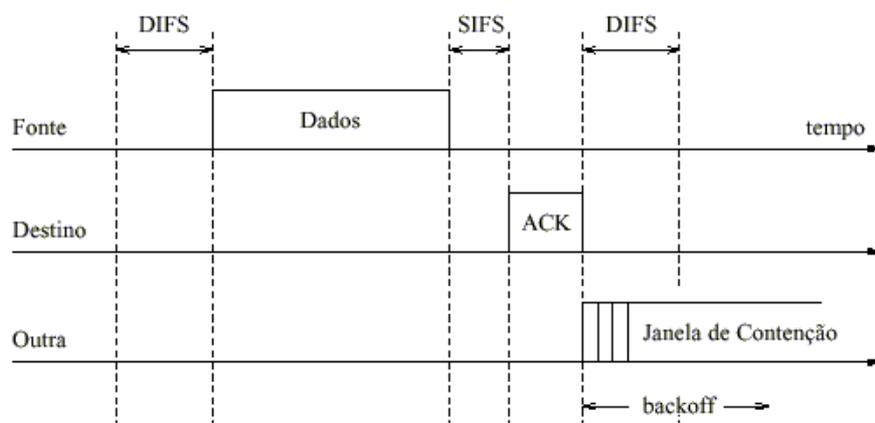


Figura 2.3: Esquema Básico de Acesso no DCF.

O segundo esquema de acesso usado no modo DCF inclui pacotes de solicitação (RTS) e permissão de transmissão (CTS) para evitar problemas gerados por terminais “escondidos”<sup>3</sup>, típico da comunicação sem fio. Nesse esquema, os pacotes

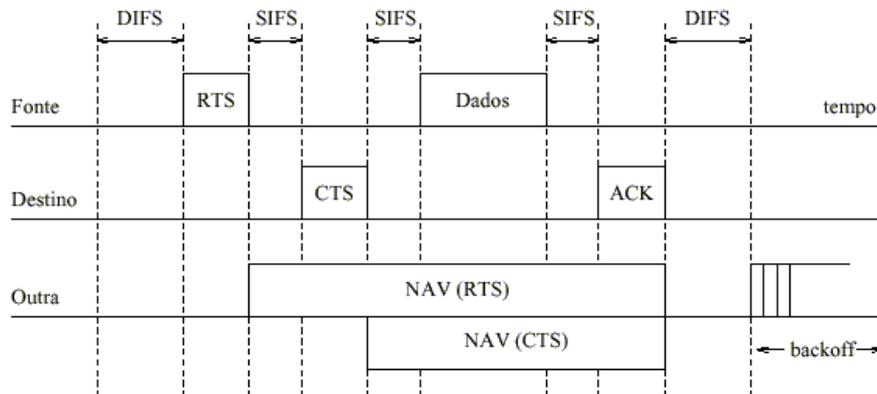


Figura 2.4: Acesso DCF com RTS / CTS.

tes RTS e CTS contêm informações a respeito do nó de destino e uma informação de duração da transmissão que inclui o tempo necessário ao recebimento do ACK referente ao pacote de dados que se quer transmitir. A estação transmissora, após perceber o meio livre por um tempo equivalente a DIFS, envia um RTS ao receptor antes da transmissão do quadro de dados, a fim de reservar o meio (Figura 2.4). O receptor responde com um CTS, após o meio estar livre por um tempo SIFS, caso esteja pronto para receber. Ao receber o CTS e esperar o meio estar livre por SIFS segundos (Figura 2.4), o transmissor inicia o envio do quadro de dados e passa a aguardar o ACK correspondente, como no DCF básico. Caso não receba o CTS, o transmissor executa o procedimento de *backoff* e retransmite o RTS.

Todas as estações que ouvirem o RTS, o CTS, ou o próprio quadro de dados utilizam a informação da duração da transmissão contida nesses quadros para atualizar o seu vetor de alocação de rede (*Network Allocation Vector* - NAV), que é utilizado para a detecção virtual da portadora (Figura 2.4). Nesse tipo de detecção, a informação de duração funciona como um pedido de reserva do meio. Com essa

<sup>3</sup>O problema do terminal escondido surge quando uma estação *B* é capaz de receber quadros de dois diferentes transmissores, *A* e *C*, porém estes transmissores estão fora do alcance um do outro. Diz-se que *A* está escondido para *C* e vice-versa. Nesse caso, o transmissor *A* pode achar que o meio está livre mesmo que *C* esteja transmitindo, o que resultará em colisão no receptor *B*.

informação, as demais estações não envolvidas na transmissão evitam fazer acesso ao meio no período de tempo reservado, não importando se o CCA indique que o meio está livre. Desse modo, qualquer terminal escondido poderá adiar a sua transmissão para evitar colisões. As estações só voltam a tentar novas transmissões quando seus vetores de alocação da rede indicam que não há mais transmissões pendentes.

O uso de RTS e CTS é controlado em cada estação através de um parâmetro configurável chamado de limiar de RTS ( $RTS_{threshold}$ ). A estação irá utilizar o RTS somente na transmissão de quadros com tamanhos maiores que o limiar, caso contrário, utilizará o esquema de acesso básico. Uma estação que tenha a opção de RTS e CTS implementada mas não esteja usando RTS em suas transmissões deve ainda ser capaz de responder com o CTS para outras que estejam usando esse recurso.

A colisão de um quadro RTS de 20 *bytes* é menos severa e menos provável que uma colisão dos quadros de dados, que podem ter até 2346 *bytes*. No entanto, quando a estação está usando tamanhos de quadros pequenos (por exemplo, de 64 ou 128 *bytes*), a sobrecarga adicionada pela transmissão dos quadros RTS e CTS pode reduzir o benefício trazido pelo uso desses quadros. O parâmetro  $RTS_{threshold}$  permite decidir quando usar ou não o esquema de acesso com RTS/CTS, controlando o efeito dessa sobrecarga.

### **Função de Coordenação Centralizada - PCF**

A Função de Coordenação Centralizada, ou PCF, diferentemente da DCF, é de implementação opcional. No modo PCF, um ponto de coordenação controla o acesso ao meio, através de consulta periódica (*polling*) às estações relacionadas em uma lista de consulta, proporcionando a essas estações uma oportunidade de transmitirem sem contenção. Não é necessário que todas as estações associadas a um BSS estejam na lista de consulta, apenas aquelas que necessitam dos serviços de acesso sem contenção.

O ponto de coordenação, geralmente localizado no próprio ponto de acesso, divide o tempo de acesso em períodos chamados de superquadros. Cada superquadro

compreende um período livre de contenção ou CFP (*Contention Free Period*), que usa o modo PCF, e um período com contenção ou CP (*Contention Period*), que usa o modo DCF. O superquadro é mostrado na Figura 2.5.

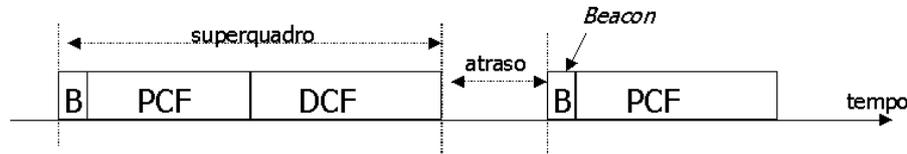


Figura 2.5: Superquadro e Divisão de Períodos (CP e CFP).

O ponto de coordenação inicia e controla o CFP. Ele escuta o meio por um tempo PIFS - *Point (Coordination Function) Interframe Space* e então começa o período através da difusão de um sinal de *beacon* (Figura 2.5). Como, por definição, PIFS é igual a  $SIFS + slottime$ , sendo portanto menor que DIFS, nenhuma estação pode começar a enviar dados no modo DCF antes do ponto de coordenação. Ao receberem o *beacon*, todas as estações (inscritas ou não na lista de consulta) adicionam a duração máxima do período livre de contenção ( $CFP_{maxduration}$ ) aos seus respectivos vetores de alocação da rede (NAVs). A relação entre os diversos intervalos entre-quadros (PIFS, SIFS, DIFS) utilizados no MAC 802.11 é mostrada na Figura 2.6. Essa figura também mostra a seqüência das ações tomadas pela estação ao realizar uma transmissão.

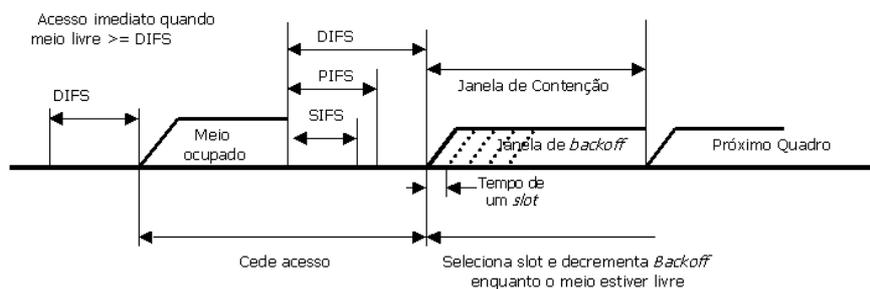


Figura 2.6: Relações dos Intervalos Entre-Quadros do MAC 802.11.

Durante o período livre de contenção (CFP), o coordenador consulta as estações inscritas no modo PCF através de quadros de consulta (*CF-Poll*) para saber se elas têm algo a transmitir. O ponto de coordenação percorre a lista e envia um

quadro de consulta para as estações, uma de cada vez. Se há dados para transmitir para a estação sendo consultada, eles são enviados junto com o quadro de consulta (*piggyback*). O receptor envia de volta um ACK, também incluindo dados se for o caso, após um tempo SIFS. Ao encerrar a transmissão a todas as estações contidas na lista, o ponto de coordenação reinicia o processo de consulta após um tempo PIFS e até que o período livre de contenção seja terminado. As estações que ficam sem transmitir por alguns ciclos são retiradas da lista e só voltam a ser consultadas no início do próximo CFP.

O período livre de contenção pode terminar a qualquer momento através do envio de um quadro  $CF_{end}$  pelo ponto de coordenação. O término antecipado do CFP ocorre freqüentemente quando a rede está com pouca carga. Terminado o CFP, inicia-se um período com contenção (CP) onde as estações utilizarão apenas as regras do modo DCF para disputar o acesso. O intervalo esperado entre períodos livre de contenção consecutivos é definido por um parâmetro  $CFP_{interval}$ . Esse intervalo nem sempre é respeitado, uma vez que o início de um CFP pode ser adiado se alguma estação no modo DCF do superquadro anterior ainda não tiver concluído a sua transmissão (atraso mostrado na Figura 2.5).

O modo PCF está fora do escopo deste trabalho. Apesar de ter sido idealizado para oferecer serviços de acesso sem contenção, teoricamente mais adequados ao tráfego que possui requisitos de QoS, esse modo possui limitações no cumprimento desses requisitos (essas limitações são discutidas em seção posterior). Além disso, sua implementação é opcional, e muitos fabricantes de interfaces de rede 802.11 optaram por não implementar o PCF em seus equipamentos.

### **Suporte à Fragmentação**

O MAC 802.11 provê também suporte para a fragmentação e remontagem de quadros em transmissões ponto-a-ponto, permitindo a transmissão de um quadro grande em uma seqüência de quadros menores. A fragmentação é útil nas situações onde ocorrem altas taxas de erros por enfraquecimento do sinal ou por ruído. A implementação é obrigatória nos receptores, mas opcional nos transmissores. Um

limiar de fragmentação configurável ( $Fragmentation_{threshold}$ ) é estabelecido, de modo que um quadro seja fragmentado somente quando for maior do que esse limiar. O tamanho máximo de um fragmento também é dado por  $Fragmentation_{threshold}$ .

No período com contenção, os fragmentos de um mesmo quadro são enviados em rajada, utilizando uma única invocação do procedimento de acesso ao meio do DCF. Um fragmento é enviado em um tempo SIFS após o recebimento do ACK relativo ao fragmento anterior. No período livre de contenção, os fragmentos são enviados como quadros individuais.

## 2.4 802.11 em Cenários de Internet Móvel

A Internet Móvel consiste essencialmente na expansão da Internet atual para permitir a mobilidade dos usuários. No cenário ideal, o usuário deve ter a possibilidade de acesso a qualquer hora e em qualquer lugar, usando qualquer dispositivo de comunicação de dados disponível. Uma estação móvel deve ser capaz de estabelecer comunicação com todos os demais equipamentos fixos ou móveis conectados à Internet, mesmo que estejam sendo usadas tecnologias de acesso totalmente distintas. O uso do IP como protocolo comum da camada de rede facilita essa integração, mas não é o suficiente. Os problemas inerentes às redes móveis, como gerência de localização e *handoff*, citados na Seção 2.1, são mais graves quando tecnologias heterogêneas são empregadas em conjunto.

Nesse contexto, o padrão 802.11 pode ser empregado seguindo-se pelo menos duas abordagens: 1) conexão direta a redes IP públicas ou privadas; 2) integração com a Rede de Núcleo baseada em IP dos sistemas celulares 2,5 e 3G, como GPRS e UMTS, citados na Seção 2.2.

Na primeira abordagem, a rede sem fio é uma extensão natural da rede IP fixa. A principal vantagem é a menor complexidade envolvida nos níveis inferiores da rede, uma vez que o padrão para WLAN já foi desenvolvido para ser compatível com os demais padrões para LANs e MANs (802.x). A desvantagem é que deve ser garantida uma grande densidade de pontos de acesso para que o usuário possa

utilizá-los "em qualquer lugar". Uma tendência para superar essa limitação é a formação de um grande sistema cooperativo, no qual os detentores de redes 802.11 compartilham seus pontos de acesso com os usuários de outras redes em visita à sua área de cobertura, desde que seus usuários tenham o mesmo tratamento quando estiverem migrando para as áreas de cobertura dessas outras redes. A idéia de construção de um sistema de acesso de ampla cobertura, sem controle centralizado ou proprietário único, construído de baixo para cima e de uma forma colaborativa recorda a formação da própria Internet.

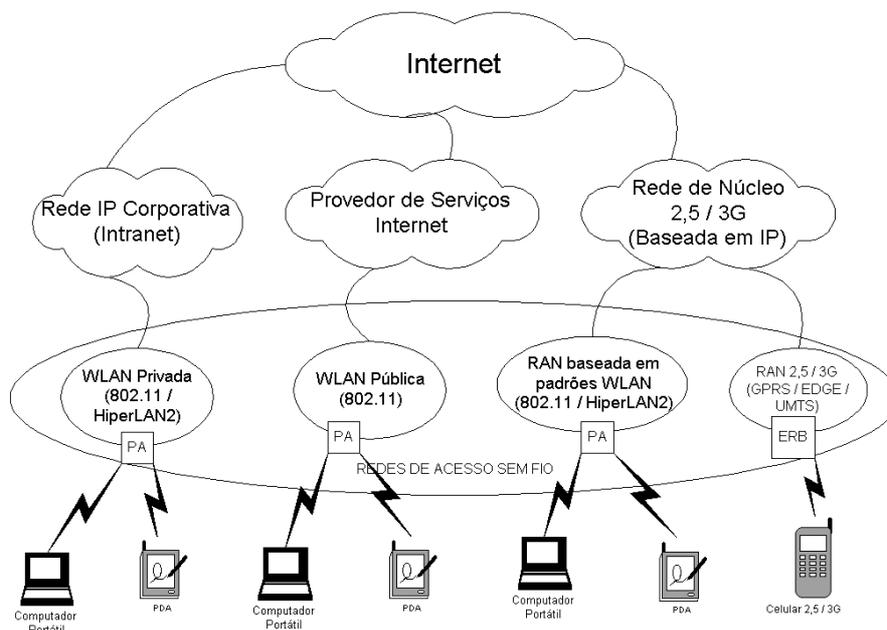


Figura 2.7: Formas de Acesso Móvel à Internet.

Em uma outra proposta, Leung *et al* [29], sugerem o uso de células 802.11 com raio de até 6 quilômetros, em áreas abertas. Esse alcance, cerca de 10 vezes maior do que o normal, seria possível substituindo-se o SIFS por DIFS na espera de um quadro CTS ou ACK e por aprimoramentos no nível físico [30].

Na segunda abordagem, o 802.11 é empregado como uma rede de acesso rádio (*RAN - Radio Access Network*) alternativa, trazendo como principal vantagem a oferta de taxas de transmissão de dados mais altas, já que as WLANs possibilitam taxas de até 54 Mbps, enquanto as RANs estão limitadas a 2Mbps. A necessidade de RANs com maior capacidade foi identificada pelas próprias entidades padroni-

zadoras de redes 3G. A especificação do HIPERLAN2, padrão WLAN desenvolvido pelo grupo de redes de acesso em banda larga (*BRANs - Broadband Radio Access Networks*) do *ETSI - European Telecommunications Standards Institute*, busca atender a essa demanda. O 802.11 também tem sido considerado, principalmente suas extensões 802.11b e 802.11a. Em [31], por exemplo, é proposta uma arquitetura chamada *OWLAN - Operator WLAN* a qual provê mecanismos para autenticação, segurança, contabilidade dos serviços e migração das estações em um cenário de integração entre redes de acesso 802.11b e sistemas celulares GPRS.

Outras vantagens da integração WLAN-celular são o baixo custo de instalação, com o uso de pontos de acesso baratos em relação a uma estação rádio-base (ERB) celular, e o aproveitamento dos mecanismos de gerência de localização já existentes no sistema celular. A principal desvantagem é o alcance mais curto dos pontos de acesso 802.11 em relação às ERBs celulares, o que restringe o uso desse padrão apenas aos pontos de grande concentração de usuários (*hotspots*) anteriormente citados. A Figura 2.7 ilustra como as várias tecnologias de comunicação sem fio podem ser utilizadas para formar tipos diferentes de redes de acesso.

## 2.5 Considerações Finais

O presente capítulo relacionou os principais aspectos acerca das redes móveis e sem fio, e apresentou a arquitetura, a camada física e a subcamada MAC do padrão 802.11, objeto de estudo deste trabalho. Foram ainda apresentadas possíveis formas de emprego de redes 802.11 infra-estruturadas em um contexto de acesso ubíquo, tirando partido das altas taxas de transmissão de dados e do baixo custo de instalação e operação dessas redes.

Independentemente da forma particular como o 802.11 será empregado, e a despeito de sua relevância como opção de acesso, há ainda a necessidade de que algumas questões sejam analisadas, como a segurança, a vulnerabilidade à interferência eletro-magnética, o suporte à mobilidade em larga escala e o suporte à Qualidade de Serviço. No capítulo a seguir, as principais questões relacionadas ao suporte à QoS no 802.11 serão discutidas.

## Capítulo 3

### 802.11 e Qualidade de Serviço

O PROBLEMA da provisão de Qualidade de Serviço, nas redes fixas, é normalmente resolvido na camada de rede do modelo OSI ou nas camadas superiores, chegando a ser tratado até mesmo na camada de aplicação. Nas redes móveis, no entanto, esse problema deve ser considerado já na camada de enlace, dadas as limitações dos recursos de transmissão disponíveis na rede sem fio. O meio sem fio apresenta taxas de transmissão mais baixas e taxas de erros mais altas que as observadas nas redes fixas, e sofre ainda variações na capacidade de transmissão de dados provocadas por fatores externos, como interferência eletro-magnética e condições atmosféricas.

Neste capítulo, são apresentadas as questões relativas ao suporte à Qualidade de Serviço no padrão 802.11 para as redes locais sem fio. A Seção 3.1 cita aspectos sobre QoS em geral. A Seção 3.2 relaciona as principais técnicas de diferenciação de serviços desenvolvidas para o 802.11 e apresenta a extensão 802.11e. A Seção 3.3 aborda mecanismos adicionais que buscam aprimorar a Qualidade de Serviço oferecida pela rede sem fio. Por último, a Seção 3.4 traz as considerações finais do capítulo.

## 3.1 Qualidade de Serviço

O conceito de Qualidade de Serviço (*Quality of Service - QoS*) foi introduzido pela ISO para mensurar a qualidade dos serviços oferecidos por uma rede de comunicações, ou seja, refletir o quanto ela é capaz de atender às expectativas de seus usuários através dos serviços que oferece [32]. Esse conceito, inicialmente focado na rede, evoluiu para uma noção mais ampla, contemplando as múltiplas camadas da interação usuário-sistema. Nahrstedt [33] apresenta um modelo que envolve aspectos relacionados à própria percepção do usuário (*Perceptual QoS*), aos requisitos das aplicações (*Application QoS*) e aos recursos disponíveis no sistema (*System QoS*), seja no equipamento do usuário (*Device QoS*) ou na rede em si (*Network QoS*).

A Qualidade de Serviço envolve ainda questões de ordem organizacional e administrativa, como segurança, privacidade, contabilidade/política de preços dos serviços, estabelecimento e monitoração de contratos de serviços e grau de disponibilidade da rede (em termos de tempo médio entre falhas e tempo médio de reparo).

A forma mais simples de oferecer Qualidade de Serviço é o superdimensionamento da rede [34], no qual se procura garantir que a quantidade de recursos disponíveis seja sempre superior à demanda. Essa primeira abordagem, embora simples, tem como desvantagem o custo elevado, pois exige investimentos altos na infra-estrutura. Tais investimentos devem considerar um excesso de demanda, além das estimativas de demanda atual e futura, para permitir que a rede continue atendendo às expectativas mesmo em situações excepcionais, difíceis de prever, como um crescimento repentino no número de usuários ou o desenvolvimento de uma nova aplicação que consome um volume maior de recursos da rede.

Uma segunda abordagem, ao mesmo tempo mais barata e mais complexa que a primeira, envolve a inclusão de mecanismos que buscam racionalizar o uso dos recursos disponíveis na rede. Esses mecanismos estabelecem níveis de serviço e permitem a convivência na mesma rede de tráfegos com requisitos distintos de qualidade. Tráfegos pertencentes a níveis de serviço diferentes são tratados de forma que o nível mais prioritário possa sempre dispor dos recursos de que necessita, ainda que em detrimento dos níveis menos prioritários. Ao mesmo tempo, tráfegos pertencen-

tes a um mesmo nível de serviço são tratados de forma que suas demandas sejam atendidas de forma justa.

Os mecanismos de diferenciação, os níveis de serviço oferecidos e os parâmetros de desempenho da rede utilizados para satisfazer os requisitos das aplicações são geralmente definidos em uma Arquitetura ou uma Estrutura de QoS (*QoS Framework*). Existem várias propostas, criadas para cada tipo de rede: para Redes ATM (*Asynchronous Transfer Mode*), para redes IP (Seção 3.1.2) e para redes celulares de 3<sup>a</sup> Geração, por exemplo. Como há várias questões em aberto em muitos dos aspectos relacionados à QoS, as divergências de concepção dessas arquiteturas geram problemas de mapeamento entre níveis de serviço quando é necessária a comunicação entre redes com arquiteturas diferentes.

Os tópicos restantes desta seção discorrem sobre outros conceitos relacionados à Qualidade de Serviço. A Seção 3.1.1 apresenta uma classificação das aplicações quanto às suas necessidades de QoS; a Seção 3.1.2 apresenta duas arquiteturas de QoS empregadas na Internet; finalmente, a Seção 3.1.3 apresenta comentários sobre QoS em redes móveis.

### 3.1.1 Classificação das Aplicações

As aplicações multimídia e de tempo real possuem requisitos de QoS que devem ser traduzidos para parâmetros de desempenho da rede. Tais parâmetros são normalmente relacionados à capacidade de transmissão de dados, ao tempo consumido nas transmissões e à confiabilidade. Alguns dos parâmetros comumente empregados [35] são descritos a seguir.

- Largura de banda (*bandwidth*): quantidade de dados que podem ser transmitidos por unidade de tempo, normalmente expressa em bits por segundo (bits/s, *bps*) ou em múltiplos dessa unidade (Kilobits/s, Megabits/s).
- Vazão (*throughput*): quantidade de dados transmitida com sucesso por unidade de tempo, também expressa em bits/s.
- Retardo (*delay*): tempo consumido na transmissão de uma mensagem. Na

camada de aplicação, o retardo é a diferença de tempo (fim-a-fim) transcorrida entre a geração do dado no transmissor e a sua apresentação no receptor. Essa diferença inclui parcelas referentes ao processamento nos nós finais (*endpoints*) e intermediários (roteadores, comutadores), à disputa pelo acesso ao meio nos enlaces compartilhados e ao tempo de propagação no meio físico. Caso a conexão entre transmissor e receptor envolva múltiplos saltos, como é comum em redes comutadas por pacotes, a soma de todos os retardos salto-a-salto mais o retardo de processamento deve ser igual ou inferior ao retardo fim-a-fim desejado. Geralmente deseja-se limitar algum parâmetro relativo à curva de distribuição do retardo, como um valor médio, máximo, ou um percentil.

- Variação do retardo (*jitter*): variação observada no conjunto de valores de retardo de unidades de dados consecutivas. Como no caso do retardo, deseja-se limitar valores estatísticos da distribuição.
- Taxa de erros de bits (*bit error rate - BER*): razão entre a quantidade de bits recebidos com erro e a quantidade de bits enviados.
- Taxas de erros ou de perdas de pacotes (*packet error/loss rate - PLR/PER*): razão entre a quantidade de pacotes perdidos e a quantidade de pacotes enviados. Os pacotes podem ser perdidos na rede por descarte nas filas dos nós intermediários, ou podem ser corrompidos por colisão com outros pacotes em enlaces compartilhados e ainda por variações no meio físico (tendo, nesse caso, relação com a taxa de erros de bits).

As necessidades relativas à QoS variam de acordo com a natureza da aplicação. Aplicações de áudio e vídeo, por exemplo, exigem a garantia de uma vazão mínima, baixo retardo e baixa variação de retardo, mas são ligeiramente tolerantes a erros e perdas. A perda de alguns quadros em um fluxo de vídeo, por exemplo, não é suficiente para comprometer a percepção do usuário com relação ao conteúdo da informação audiovisual. Por outro lado, aplicações de transferência de arquivo como FTP são intolerantes a erros, mas não necessitam de garantias de retardo ou vazão, embora uma transferência mais rápida traga mais satisfação para o usuário.

No tocante às necessidades de QoS, a *Internet Engineering Task Force - IETF* classifica as aplicações em dois grandes grupos [36]: as de tempo real, nas quais os dados são esperados em um certo tempo para que sejam considerados úteis, e as elásticas, nas quais a utilidade dos dados não está relacionada ao seu tempo de chegada. As de tempo real são ainda subdivididas em intolerantes - que não suportam violações do limite de retardo - e tolerantes - que suportam violações ocasionais. Essa classificação é baseada principalmente na sensibilidade ao retardo.

Existem ainda aplicações adaptativas, que procuram adequar-se às condições da rede a cada instante. Um exemplo seria uma aplicação de vídeo que varia a sua taxa de geração de dados trocando o codificador de vídeo ou reduzindo a resolução, o número de cores e o número de quadros de vídeo apresentados por segundo.

### 3.1.2 Qualidade de Serviço na Internet

A Internet foi inicialmente concebida para oferecer um único nível de serviço, o de melhor esforço (*Best-Effort*), sem preocupações com QoS. Nesse modelo não existe reserva de recursos ou prioridades entre tráfegos. Em caso de congestionamento da rede, todas as aplicações sofrem degradação na qualidade, independentemente de seus requisitos particulares. Porém, a explosão do número de usuários e a evolução das aplicações, saindo de um padrão de comunicação textual para padrões de comunicação multimídia, motivaram o surgimento de propostas no intuito de incluir o suporte à QoS nas redes IP. Duas propostas padronizadas pelo IETF e bastante discutidas são a arquitetura de Serviços Integrados (*Integrated Services - IntServ*) [36] e a arquitetura de Serviços Diferenciados (*Differentiated Services - DiffServ*) [37].

O IntServ é baseado em reserva de recursos e é aplicado sobre os fluxos. Na RFC 1633 [36], um fluxo é definido como "um conjunto de datagramas correlacionados, gerados por uma mesma atividade de um usuário, que exige um mesmo nível de QoS". Um protocolo de reserva (*ReSerVation Protocol - RSVP*) é usado para configurar a chamada e reservar os recursos necessários como, por exemplo, o espaço de armazenamento nos *buffers* dos roteadores. O esquema de reserva provê garantias de QoS mais absolutas para as aplicações, porém requer o armazenamento do estado

de cada fluxo individual nos nós intermediários.

O DiffServ, por outro lado, é baseado em um esquema de prioridades, aplicado sobre os agregados de fluxos. Um agregado reúne fluxos com requisitos similares para serem tratados (servidos) em conjunto. Para distinguir o tratamento a ser aplicado ao agregado, os nós da rede utilizam a informação contida no campo DS do cabeçalho dos pacotes IP [38]. Diferentemente do IntServ, o DiffServ não fornece garantias absolutas de QoS. O enfoque está na diferenciação relativa entre classes (ou categorias) de tráfego. Em compensação, não necessita armazenar nos roteadores as informações de estado dos agregados. Por esse motivo, o DiffServ é considerado mais escalável, sendo mais adequado às redes de núcleo e às espinhas dorsais (*backbones*) da Internet.

As arquiteturas IntServ e DiffServ podem ser combinadas, o que permite tirar partido das vantagens de cada uma delas. A RFC 2998 fornece linhas gerais para essa integração [39].

### 3.1.3 Qualidade de Serviço em Redes Móveis

A Qualidade de Serviço em redes móveis está sujeita às peculiaridades desse tipo de rede. Chalmers e Sloman [35] ressaltam os efeitos da mobilidade e o tipo do meio empregado como diferenças importantes entre as redes móveis e as redes fixas no que diz respeito à QoS.

A mobilidade provoca a necessidade de uma reserva adicional de recursos para suportar a transferência de conexões ativas entre células (*handoff*). Quanto maior a mobilidade, maior a probabilidade de ocorrência de *handoff* entre células vizinhas, e maior a quantidade de recursos que devem ser reservados. A mobilidade pode ainda fazer com que as estações fiquem sujeitas temporariamente a um nível mais baixo de recepção de sinal. Isso ocorre, por exemplo, quando as estações se movem para longe do centro da célula. Como a potência do sinal diminui à medida que cresce a distância entre transmissor e receptor, os quadros podem chegar abaixo do limiar de recepção, não podendo ser corretamente interpretados. O aumento da distância entre origem e destino degrada a QoS oferecida pela rede e pode, em alguns casos,

causar até mesmo a perda da conexão.

Quanto ao tipo do meio empregado, a necessidade imperativa de se utilizar comunicação sem fio nas redes móveis também traz dificuldades para a provisão de QoS. O meio sem fio apresenta limitações quando comparado com os meios de transmissão usados nas redes fixas, como os cabos de cobre e os de fibra ótica. Entre essas limitações estão as taxas de transmissão mais baixas, em geral uma ordem de grandeza menores que as suportadas nas transmissões via cabo. As redes locais sem fio 802.11, por exemplo, suportam taxas de 1 até 54 Mbps, enquanto as redes locais fixas 802.3 suportam taxas de 10 Mbps até 1 Gbps. Outra limitação diz respeito à taxa de erros de bits mais alta. Enquanto essa taxa é praticamente desprezível na fibra ótica, no meio sem fio ela pode ser de 1 bit a cada  $10^5$  transmitidos [28] ou pior. Além disso, a qualidade dos enlaces sem fio pode apresentar uma variação ao longo do tempo, causada por condições atmosféricas adversas, como chuva forte e raios, e por fontes de interferência eletro-magnética [26]. O usuário pode ainda passar por áreas que possuem barreiras físicas à transmissão sem fio, como túneis, áreas de vegetação densa ou regiões localizadas atrás de grandes elevações do terreno.

## 3.2 Diferenciação de Serviços no MAC 802.11

Uma vez entendidos os principais conceitos relacionados à provisão de Qualidade de Serviços de um modo geral, é possível agora discutir os problemas específicos do MAC 802.11 e os esquemas de diferenciação empregados nesse tipo de rede.

Muitos trabalhos foram realizados após a publicação das especificações 802.11, 802.11b e 802.11a, em 1999, identificando problemas no controle de acesso ao meio e sugerindo soluções. A maioria está relacionada à utilização do TCP em redes 802.11, como o problema da injustiça no compartilhamento da banda passante disponível [40, 41, 42, 43] e o da instabilidade provocada no TCP em redes sem fio de múltiplos saltos [43]. Esses problemas, aliados à necessidade de provisão de QoS nas redes locais sem fio, sinalizaram a necessidade de evolução do MAC 802.11.

Com relação à diferenciação de serviços, diversas propostas foram elaboradas [41,

42, 44, 45]. Essas propostas geralmente envolvem a configuração de um ou mais parâmetros básicos do método de acesso para promover diferenciação, alterando a forma com que cada uma das estações tem acesso ao meio e estabelecendo prioridades de acesso entre elas. Algumas técnicas foram elaboradas para o acesso com contenção (DCF) e outras para o acesso sem contenção (PCF).

As propostas de QoS para o PCF envolvem normalmente esquemas avançados de consulta (*polling*), que levam em consideração a prioridade das estações. As técnicas particulares idealizadas para o PCF estão fora do escopo do trabalho descrito nesta dissertação.

O enfoque deste trabalho está nas técnicas de diferenciação empregadas no acesso com contenção (DCF). As principais técnicas para esse método são descritas na Seção 3.2.2. Antes de apresentar essas técnicas, é oportuno tecer algumas considerações sobre as limitações do PCF e sobre a capacidade do MAC 802.11 em modo DCF.

### 3.2.1 Limitações do PCF e do DCF

A capacidade de garantir QoS no PCF é limitada pelo número de estações inscritas na lista de consulta, apesar de esse método ter sido projetado para suportar um serviço de acesso sem contenção, supostamente adequado a aplicações sensíveis ao retardo. Quanto maior for o número de estações inscritas na lista de consulta, maior será a dificuldade enfrentada pelo ponto de acesso no atendimento aos limites particulares de retardo ou vazão dessas estações. Além disso, existem dois outros problemas [46]: a imprevisibilidade do intervalo entre os quadros de *beacon* e o desconhecimento da duração das transmissões das estações consultadas.

O primeiro problema ocorre sempre que uma estação que utiliza o método DCF no período com contenção (CP) obtém acesso ao meio já no final desse período. Como essa transmissão não pode ser interrompida, o ponto de acesso fica incapacitado de enviar o *beacon* até que essa estação conclua a sua transmissão. Isso provoca um atraso (mostrado na Figura 2.5) no início do próximo período livre de contenção (CFP). Esse atraso é proporcional ao tamanho do quadro empregado pela estação em modo DCF, e pode provocar uma violação dos requisitos de retardo e *jitter* das

estações em modo PCF (embora o atraso seja limitado ao tempo de transmissão de um quadro de tamanho máximo, seu valor é indeterminado).

O segundo problema decorre do fato de que uma estação em modo PCF, uma vez consultada, pode enviar um quadro de até 2304 *bytes*, inteiro ou fragmentado, empregando qualquer uma das taxas de transmissão de dados disponíveis no BSS (1, 2, 5,5 ou 11 Mbps no 802.11b e de 6 a 54 Mbps no 802.11a). Isso faz com que o tempo de transmissão dessa estação esteja fora do controle do ponto de acesso, reduzindo sua capacidade de cumprir os limites de retardo das demais estações inscritas no CFP.

Quanto às limitações do DCF, algumas análises do desempenho disponíveis na literatura [47, 48, 49] mostram que a vazão máxima obtida usando esse método de acesso é dependente do tipo de tráfego, do tamanho dos quadros e do número de estações presentes na rede. O tempo total gasto para transmitir cada quadro de dados usando o acesso básico (sem RTS/CTS) é dado pela equação abaixo (adaptada de [49]):

$$t_m = DIFS + t_{dados} + SIFS + t_{ACK} + t_{backoff}(m) \quad (3.1)$$

onde  $m$  é o número de estações transmitindo,  $DIFS$  e  $SIFS$  são os valores de intervalos entre-quadros definidos no padrão, e  $t_{dados}$  e  $t_{ACK}$  são os tempos de transmissão dos quadros de dados e de reconhecimento, respectivamente.  $t_{backoff}(m)$  é o tempo médio gasto com o procedimento de recuo exponencial binário ou *backoff* (a equação desconsidera o retardo de propagação em um BSS, que é de 2  $\mu$ s).

Seja, por exemplo, uma única estação ( $m=1$ ), transmitindo continuamente em um BSS 802.11b com nível físico DSSS. Para esse caso,  $SIFS$  é igual a 10  $\mu$ s e  $DIFS$  é igual a 50  $\mu$ s. O tempo de um *slot* é 20  $\mu$ s e  $CW_{Min}$  é 31; multiplicando-se a média do intervalo uniforme [0,31] pelo tempo de um *slot*, tem-se que  $t_{backoff}(m)$  é igual a 320  $\mu$ s (mesmo havendo uma única estação transmitindo, o *backoff* é sempre executado após a transmissão bem sucedida de um quadro, para evitar que a estação capture o meio).

Para todos os quadros, de dados ou de controle, a especificação requer que o preâmbulo do nível físico e o cabeçalho do protocolo de convergência de nível físico

(*Physical Layer Convergence Protocol* - PLCP) sejam sempre transmitidos usando uma taxa básica (por exemplo, 1 Mbps), para que todas as estações pertencentes a um BSS sejam capazes de recebê-los corretamente, independentemente das taxas de dados que cada uma delas esteja usando (1, 2, 5,5 ou 11 Mbps). Esse campos somam juntos 192 bits, que são transmitidos em 192  $\mu$ s.

Os quadros de controle, como RTS, CTS e ACK, devem ser também transmitidos com uma taxa básica. O tempo de transmissão de um ACK (14 bytes) é, portanto,  $t_{ACK} = (14 * 8 / 1000000) + 192 = 304 \mu$ s.

Se a estação estiver usando a taxa de dados máxima de 11 Mbps e quadros de tamanho máximo (2304 bytes), o tempo para transmitir um quadro de dados é  $t_{dados} = ((2304 + 34) * 8 / 11000000) + 192 = 1892 \mu$ s. Os 34 bytes a mais são do cabeçalho MAC e do campo *Frame Check Sequence* - FCS, adicionados pela subcamada MAC antes da transmissão do quadro.

O tempo total para a transmissão ( $t_1$ ), dado pela equação 3.1 quando o número de estações ( $m$ ) é igual a 1, é de 2576  $\mu$ s, no exemplo em questão. A razão  $t_{dados}/t_1$  dá uma noção da sobrecarga imposta pelo método de acesso. Nesse caso, dos 2576  $\mu$ s totais gastos na transmissão, apenas 1892  $\mu$ s se referem à transmissão do quadro de dados em si. Assim, a banda útil máxima, medida na subcamada MAC para uma única estação em um canal DSSS, é cerca de 73,4% ( $1892 / 2576 * 100$ ) dos 11 Mbps, ou seja, 8,079 Mbps. Essa banda útil diminui à medida que são empregados quadros de dados menores.

À medida que cresce o número de estações transmitindo simultaneamente ( $m$ ), cresce a probabilidade de uma estação encontrar o meio ocupado, acarretando um aumento do tamanho médio da janela de contenção das estações e de  $t_{backoff}(m)$ . O DCF busca, através da execução do procedimento de *backoff*, distribuir no tempo as tentativas de transmissão e minimizar o efeito da contenção. Porém, se o número de estações crescer de forma ilimitada, haverá situações em que  $t_{backoff}(m)$  será excessivamente alto, o que reduz drasticamente a eficiência na utilização do canal.

A partir das análises de desempenho do DCF disponíveis [47, 48, 49], percebe-se que tanto a vazão agregada da rede quanto a obtida pelas fontes individualmente

sofrem grande influência da carga total imposta à rede pelas estações, do tamanho dos quadros, dos intervalos entre-quadros usados no protocolo e do procedimento de *backoff*. As técnicas de diferenciação no DCF, apresentadas a seguir, exploram esses efeitos.

### 3.2.2 Técnicas de Diferenciação no Modo DCF

As principais técnicas de diferenciação no modo DCF envolvem a variação do tamanho do DIFS, as alterações no procedimento de *backoff*, as variações na característica de persistência do protocolo de acesso múltiplo e a variação no tamanho dos quadros de dados.

#### Variação do Tamanho do DIFS

Conforme explicado na Seção 2.3.3, o DIFS (*DCF Interframe Space*) controla o tempo que uma estação espera antes de iniciar uma transmissão de dados. Variando-se o tamanho do DIFS, pode-se estabelecer a prioridade relativa entre as estações. Uma estação de baixa prioridade, que usa um valor de DIFS maior, terá menor probabilidade de encontrar o meio livre em uma disputa com outra, de alta prioridade, que esteja usando um DIFS menor. Como o DIFS pode ser definido por  $SIFS + N * SlotTime$ , a diferenciação consiste em variar o valor de  $N$ , ou seja, aumentar ou reduzir o número de *slots* que a estação aguarda além do tempo SIFS (no padrão 802.11, o valor de  $N$  é, por definição, igual a 2). Alguns cuidados devem ser observados na escolha do valor de  $N$ . Esse valor não pode ser 0 ou 1, pois isso faria com que o valor de DIFS equivalesse ao de SIFS ou ao de PIFS, respectivamente, causando falha no funcionamento do MAC. Também devem ser evitados valores de  $N$  excessivamente altos para um tráfego de baixa prioridade, porque isso pode fazer com que esse tráfego nunca transmita na presença de outro de alta prioridade. Esse efeito negativo é chamado de inanição (*starvation*).

### Alteração no Procedimento de *Backoff*

O procedimento de *backoff* também controla o tempo que uma estação espera antes de tentar uma transmissão. Esse procedimento é executado pelas estações em diversas situações: na transmissão de cada quadro caso a estação encontre o meio ocupado; entre as transmissões de quadros adjacentes, evitando, assim, a captura do meio; nas retransmissões de quadros que não foram reconhecidos pelo receptor.

A técnica de diferenciação consiste na escolha de intervalos distintos para a seleção do tamanho da janela de contenção (*CW*) de cada estação, controlados pelos parâmetros *CWMin* e *CWMax* do MAC. Sejam, por exemplo, uma estação de alta prioridade com *CWMin*=31 e outra de baixa com *CWMin*=63. A primeira escolherá valores para *CW* no intervalo uniformemente distribuído  $[0, 31]$ , enquanto a segunda escolherá valores no intervalo  $[0, 63]$ . Esperando mais tempo em média, a segunda estação terá menos chances de obter o meio que a primeira, ficando portanto com prioridade mais baixa. *CWMin* deve ser o menor possível para tráfego com restrições de retardo. Pode-se empregar, por exemplo, um valor igual a 7 para a classe mais prioritária, reduzindo assim o tempo gasto em *backoff* e mantendo-se um número mínimo de slots selecionáveis.

A alteração no valor do *CWMax* também provoca diferenciação já que esse parâmetro regula o limite máximo para o tamanho da janela de contenção. Essa técnica, no entanto, é menos efetiva que a alteração do *CWMin*, porque só começa a fazer diferença quando há muita retransmissão. Alterar o valor do *CWMax* pode também ser útil para reduzir a probabilidade de que as janelas de contenção de estações com prioridades diferentes se sobreponham. Por exemplo, uma estação de alta prioridade com *CWMin*=7 e *CWMax*=31 usará, no pior caso, 31 *slots* no procedimento de *backoff*, enquanto uma de baixa, com *CWMin*=31 e *CWMax*=1023 usará até 1023. Deve-se observar que, mesmo nessa situação, ainda há uma pequena possibilidade de que a estação de baixa prioridade escolha valores menores que a de alta prioridade. Como *CW* é escolhido no intervalo  $[0, CWMin]$ , pode acontecer, por exemplo, que a estação de alta prioridade escolha *CW*=6 no intervalo  $[0, 7]$  enquanto a de baixa escolhe 5 no intervalo  $[0, 31]$ .

As variações do procedimento de *backoff* foram empregadas em estudos relacionados à justiça de compartilhamento, ao aprimoramento do desempenho do MAC e à diferenciação. Os estudos de diferenciação seguem abordagens bem distintas. Na proposta de Tang e Gerla [42], o intervalo de *backoff* calculado é proporcional ao tamanho do quadro a ser enviado e inversamente proporcional ao peso atribuído ao fluxo ao qual pertence o quadro. Esse cálculo faz com que estações de maior peso (maior prioridade) escolham tempos de *backoff* menores. A inclusão do tamanho do quadro faz com que o compartilhamento da banda passante ponderado pelo peso seja justo. Esse esquema implementa, de maneira distribuída, o mecanismo de escalonamento conhecido como *fair-queuing*. Na proposta de Aad e Castellucia [44], os intervalos da janela de contenção são calculados de acordo com a prioridade estabelecida para cada estação. Kanodia *et al*, na proposta chamada *Distributed Priority Scheduling* [45], sugerem um esquema onde a prioridade do próximo quadro a ser enviado seja incluída nos quadros RTS, CTS, Dados e ACK. Assim, cada estação mantém uma tabela com os quadros presentes no topo da fila de todas as estações em disputa pelo meio. O intervalo da janela de contenção é então calculado em cada estação de acordo com a posição (*rank*) do seu quadro, em termos de prioridade, nessa tabela. Bensaou *et al* [41] apresentam um esquema de *backoff* diferente no qual cada estação ajusta a sua janela de contenção de acordo com a estimativa do seu compartilhamento do meio e do compartilhamento das outras estações. A idéia é permitir que todas as estações transmitam normalmente se a carga total das estações for menor do que a capacidade do meio, mas, caso isso não ocorra, cada estação deve obter um acesso ao meio proporcional a um índice de compartilhamento previamente estabelecido.

### Variação na Característica de Persistência do CSMA

Em protocolos com detecção de portadora (CSMA) do tipo  $p$ -persistente, a estação escuta o canal. Se ele estiver livre, transmite com probabilidade  $p$  ou adia a transmissão para o próximo *slot* com uma probabilidade  $(1-p)$ , repetindo o processo *slot a slot* até transmitir ou até que outra estação o faça. Nesse último caso, a estação procede como em uma colisão, aguardando um tempo aleatório para tentar depois.

Se o canal estiver ocupado, a estação permanece escutando o canal até este tornar-se livre e aplica o algoritmo. O CSMA/CD, utilizado nas redes 802.3 (*Ethernet*), por exemplo, é 1-persistente, ou seja, a estação transmite sempre que encontra o meio livre. Caso encontre o canal ocupado, a estação permanece escutando e transmite com probabilidade 1 tão logo ele fique livre. Por outro lado, em protocolos do tipo não-persistente, como o CSMA/CA, a estação, ao encontrar o canal ocupado, pára de escutar e aguarda um intervalo aleatório para tentar novamente.

A técnica baseada na variação das características de persistência do 802.11<sup>1</sup> tem como princípio tornar as estações de alta prioridade "mais persistentes" do que as de baixa prioridade, ou seja, fazê-las esperar um menor número de *slots* após encontrar o meio ocupado. Esse princípio pode ser implementado de formas variadas. A proposta do *p*-DCF [50] sugere que a estação calcule a probabilidade de colisão de acordo com a carga da rede e assim regule o número de *slots* que aguarda. A proposta inicial do 802.11e [46] define um Fator de Persistência (PF) diferente para cada nível de prioridade, a ser usado no cálculo do incremento do CW. A cada incremento, o novo valor de CW é dado por  $CW_{novo} \geq ((CW_{antigo} + 1) * PF) - 1$ . Os níveis de prioridade mais altos usam PF menor, executando procedimentos de *backoff* mais curtos. Note-se que PF é igual a 2 no DCF padrão.

### Variação do Tamanho dos Quadros de Dados

O tamanho do quadro de dados também promove diferenciação [44], uma vez que a disputa pelo acesso é realizada quadro a quadro. Considerando-se que, no padrão, toda a sobrecarga do protocolo CSMA/CA é fixa (tamanho dos quadros RTS/CTS/ACK e dos intervalos entre-quadros), o tamanho do quadro de dados passa a ter o efeito de regular o tempo que uma estação mantém o controle sobre o meio após vencer uma disputa. Estações com maior prioridade utilizam quadros de maior tamanho nas suas transmissões, garantindo um maior tempo de utilização do meio. Aad e Castelluccia [44], no entanto, ressaltam que os quadros maiores estão mais sujeitos a erros, o que reduz a eficácia dessa técnica quando empregada em

---

<sup>1</sup>Tal variação também tem sido proposta com finalidade de melhorar o desempenho do protocolo, em termos de vazão máxima, e não apenas com intuito de diferenciação de serviços.

meios com taxas de erros altas (esse é exatamente um dos motivos da existência do suporte à fragmentação no 802.11). Tal técnica também não é adequada para algumas aplicações de tempo real, como as de voz, que utilizam quadros de tamanho pequeno para reduzir o retardo de cada unidade de dados.

### Escopo da Diferenciação

As técnicas apresentadas podem ser utilizadas para estabelecer diferenciação tanto por fluxo individual quanto por categoria de tráfego e não apenas por estação. O DIFS e o *backoff* podem ser usados, por exemplo, para diferenciação de categorias de tráfego distintas de uma mesma estação. Até mesmo uma diferenciação por quadro é possível, de forma que quadros de uma mesma estação ou até de um mesmo fluxo possam ser transmitidos com prioridades diferentes, como ocorre na proposta de Kanodia *et al* [45]. As diversas técnicas podem ainda ser empregadas em conjunto, compondo uma solução de diferenciação mais elaborada.

### 3.2.3 A Extensão 802.11e

Atento às necessidades de padronização e aos desenvolvimentos recentes na área de QoS em redes 802.11, o IEEE criou o Grupo de Tarefas E [2] dentro do Grupo de trabalho 11 para especificar uma extensão ao padrão que introduzisse mecanismos de provisão de QoS no MAC 802.11. Essa extensão, denominada *MAC Enhancements for Quality of Service*, ainda está em desenvolvimento, de forma que as definições que serão aqui apresentadas estão sujeitas a mudanças.

No 802.11e, as estações (STAs) são chamadas de QSTAs (*QoS Stations*), o ponto de acesso de QAP (*QoS Access Point*) e o BSS de QBSS (*QoS Basic Service Set*). A extensão incorpora uma nova função de coordenação, a Função de Coordenação Híbrida (*Hybrid Coordination Function - HCF*). Nessa função, um Coordenador Híbrido (*Hybrid Coordinator - HC*), normalmente localizado no ponto de acesso, é responsável pelo estabelecimento e coordenação da funcionalidade de QoS dentro do QBSS. A função híbrida combina as características do PCF e do DCF e introduz outras novas, específicas para QoS, bem como novos tipos de quadros para permitir

transferências de dados com restrições de QoS tanto no período com contenção (CP) quanto no período livre de contenção (CFP). A especificação permite coexistência com estações legadas (802.11/b/a) de forma transparente para essas estações.

Um conceito importante no 802.11e é o de oportunidade de transmissão, ou TxOp (*Transmission Opportunity*). Uma TxOp é definida como um intervalo de tempo no qual uma estação tem o direito de iniciar transmissões. Ela é marcada por um tempo de início e uma duração. Durante uma TxOp, uma estação pode enviar vários quadros em rajada, separados por SIFS, sem ter que disputar o meio quadro a quadro. A duração de uma TxOp é limitada, para evitar retardo excessivo nas outras estações que estão aguardando o acesso.

A função HCF usa dois métodos de acesso: um baseado em consulta, chamado *HCF polled channel access*, que é executado de forma centralizada no HC, e outro, baseado em contenção, chamado *HCF contention-based channel access*, que é executado nas estações.

No acesso baseado em consulta, as TxOps são atribuídas explicitamente pelo QAP (*polled TxOp*), de forma centralizada, através de quadros especiais. Isso pode ser realizado tanto durante o CFP quanto durante o CP, para satisfazer prioridades de entrega, taxas de serviço ou requisitos de retardo e *jitter* pré-definidos. Tirando partido de seu acesso privilegiado, o QAP cria períodos de acesso controlado com duração limitada (*CAPs - Controlled Access Periods*). O QAP usa o conhecimento que tem dos fluxos presentes no QBSS para criar esses períodos e exercer o controle da funcionalidade de QoS em qualquer instante do superquadro. A QSTA, ao receber uma TxOp, pode transferir múltiplos quadros em rajada, dependendo do tamanho da TxOp que lhe foi atribuída. Esse tamanho é informado no quadro de atribuição.

A Figura 3.1 (extraída de [46]) mostra o superquadro no MAC 802.11e. Como pode ser visto na figura, o HC usa os quadros de *Beacon* para iniciar e *CF-End* para finalizar o CFP. Quadros especiais *QoS CF-Poll* são usados para atribuição direta de TxOps às STAs, mas há também TxOps obtidas por contenção no CP. Em cada TxOp, o transmissor realiza a troca de quadros RTS/CTS/DATA/ACK com o receptor, podendo enviar um ou mais quadros de dados, dependendo da duração

máxima permitida para a oportunidade de transmissão.

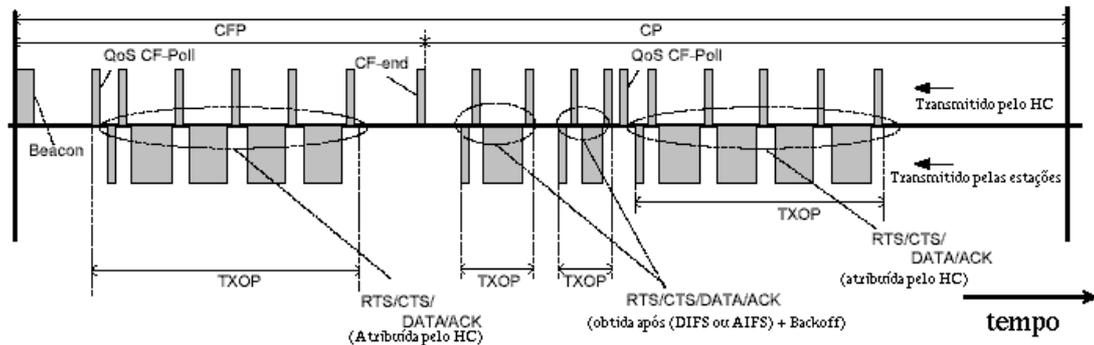


Figura 3.1: Superquadro no MAC 802.11e

No acesso baseado em contenção, as estações competem pelas TxOps de uma forma distribuída, semelhante ao que ocorre no DCF. Para isso, elas usam uma variante avançada do DCF, chamada *Enhanced DCF* ou EDCF. O EDCF incorpora o conceito de múltiplas filas de transmissão em uma mesma estação, cada uma com um nível de prioridade. O mecanismo é baseado em uma proposta anterior chamada Virtual-DCF ou VDCF [51, 52]. Nessa proposta, dentro de uma mesma estação, várias instâncias do DCF são executadas em paralelo, como se fossem "MACs virtuais", com filas diferentes e parâmetros de QoS (DIFS e CWMin) distintos para cada instância (Figura 3.2). Essas instâncias, definidas como categorias de acesso (*Access Categories - ACs*), executam o procedimento de *backoff* independentemente umas das outras, e competem entre si pelas oportunidades de transmissão. Cada estação 802.11e pode suportar de 2 a 8 categorias de acesso. O QAP, no entanto, deve suportar no mínimo 4.

A inclusão de múltiplas filas em uma mesma estação introduz um novo conceito, o de colisão virtual. A colisão virtual, interna a uma estação, ocorre sempre que duas categorias de uma mesma estação têm seus temporizadores expirados simultaneamente. Os conflitos entre essas categorias são resolvidos internamente pela estação, de forma que a categoria de maior prioridade receba a TxOp e a de menor prioridade execute o procedimento de *backoff*, de maneira semelhante ao que ocorre nas colisões reais no meio físico (externas à estação). Obviamente, o quadro transmitido fica sujeito ainda às colisões externas.

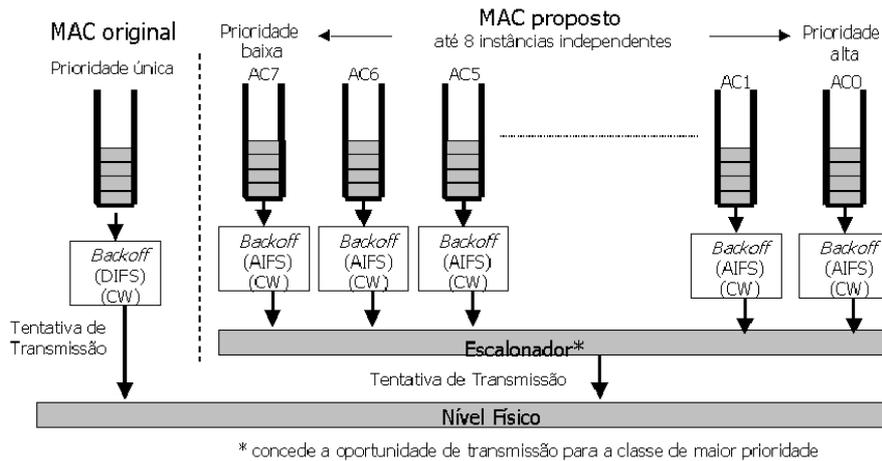


Figura 3.2: MAC Proposto no EDCF

A diferenciação entre ACs é obtida com o uso conjunto da variação da duração mínima de espera por meio livre e do tamanho mínimo da janela de contenção (DIFS e CWMin do DCF). O DIFS de cada categoria é chamado de *AIFS - Arbitration Interframe Space*. As categorias com prioridades mais altas usam valores menores de AIFS. A Figura 3.3 ilustra as relações dos valores de AIFS com os outros intervalos entre-quadros já definidos.

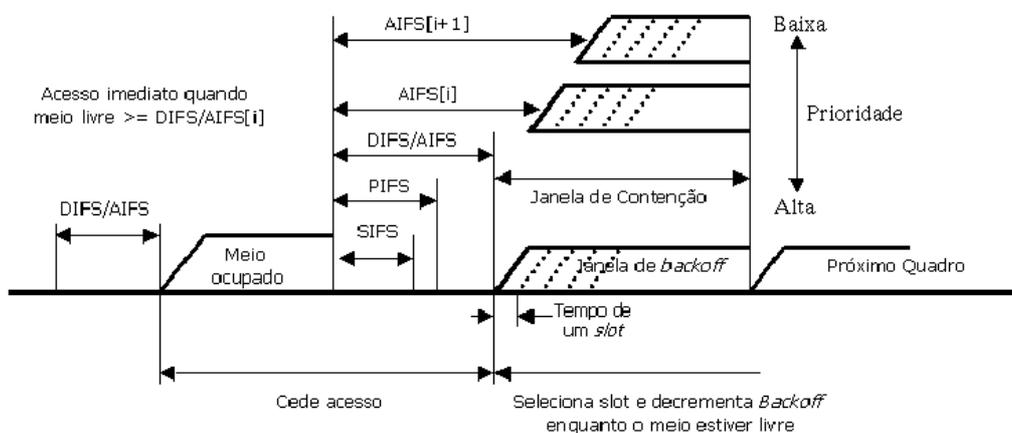


Figura 3.3: Relações dos Intervalos Entre-Quadros no 802.11e

Os parâmetros de QoS a serem usados são definidos pelo HC sempre com base em 8 prioridades de usuários (*User Priorities - UPs*). Um conjunto de parâmetros de QoS (*QoS Parameter Set ou EDCF Parameter Set*), mostrado na Figura 3.4, é usado para informar às QSTAs os parâmetros necessários para o estabelecimento da

funcionalidade de QoS. Esse conjunto, que é transmitido pelo HC encapsulado em quadros de controle como os quadros de *Beacon* [1], fornece o valor de  $AIFS(i)$  e do  $CWMin(i)$  a ser usado para as prioridades de 0 a 7.

Os dois primeiros campos, *Element ID* e *Length*, identificam o *QoS Parameter Set* e informam o seu comprimento. Os campos de AIFS e CWMin têm 8 octetos; cada octeto carrega um inteiro sem sinal que indica o valor de AIFS ou CWMin de cada prioridade  $i$ . A duração máxima das oportunidades de transmissão obtidas por contenção é definida em um terceiro parâmetro, o *TxOp Limit*. Esse parâmetro é informado em um campo de 2 octetos que indica um número de períodos de  $16 \mu s$  correspondente à duração máxima das TxOps. Ao contrário do  $AIFS(i)$  e do  $CWMin(i)$ , esse valor é configurado para todo o QBSS, e não por prioridade de usuário.

Element ID (12 bits)	Length (18 bits)	CP TxOP Limit (2 octetos)	CWMin[UP] CWMin[0]...CWMin[7] (8 octetos)	AIFS[UP] AIFS[0]...AIFS[7] (8 octetos)
-------------------------	---------------------	------------------------------	---	--

Figura 3.4: Conjunto de Parâmetros de QoS do EDCF

Outros parâmetros podem ainda ser incluídos no *QoS Parameter Set*, como um tamanho máximo de janela de contenção ( $CWMax(i)$ ), um fator de persistência ( $PF(i)$ ) e um limite da oportunidade de transmissão por categoria. Na fase atual da padronização, ainda não há consenso no Grupo de Tarefas "E" sobre a inclusão de tais parâmetros.

Como uma estação pode suportar de 2 a 8 categorias de acesso, cabe às estações realizar um mapeamento das 8 prioridades de usuário definidas pelo HC para o número de categorias de acesso que ela suporta. Quando múltiplas UPs estão associadas a uma mesma AC, os parâmetros a serem usados são os correspondentes aos da UP de mais baixa prioridade associada à categoria.

## 3.3 Necessidades Adicionais

A inclusão do suporte à QoS no 802.11 traz para as redes locais sem fio toda a problemática do gerenciamento da Qualidade de Serviço. Muitas das questões envolvidas nessa problemática já foram bastante estudadas em redes fixas. Ao abordar o gerenciamento de tráfego em redes fixas, Keshav [34] afirma que o objetivo central desse gerenciamento é estabelecer um compromisso entre a diversidade de serviços e a eficiência da rede. Uma preocupação similar passa a valer nas redes 802.11e, provocando a necessidade de desenvolver mecanismos para resolver questões como controle de admissão, escalonamento entre as categorias de tráfego, gerenciamento de filas e controle de carga dos fluxos presentes na rede.

As variações na qualidade dos enlaces sem fio e o impacto da mobilidade sobre a QoS, citados na Seção 3.1.3, trazem dificuldades ainda maiores para o gerenciamento de QoS em ambientes móveis. Chalmers e Sloman [35] apontam a necessidade de um gerenciamento adaptativo, capaz de determinar um conjunto de níveis aceitáveis de QoS, ao invés de tentar garantir valores específicos para os parâmetros de qualidade.

Uma das questões está relacionada à definição dos parâmetros de QoS para o acesso com contenção, informados às estações através do conjunto de parâmetros de QoS (Seção 3.2.3).

### 3.3.1 Determinação dos Parâmetros de QoS e Controle de Carga

O conjunto de parâmetros de QoS pode ser usado pelo HC para estabelecer políticas, modificar políticas ao admitir novos fluxos ou ainda adaptar-se às mudanças na carga oferecida à rede. Porém, a definição de como ele deve ser composto para cada situação particular de carga está fora do escopo da especificação 802.11e. É esperado que o HC o faça de duas formas: 1) seguindo regras estáticas e pré-definidas, extraídas de uma política administrativa de utilização da rede; 2) utilizando alguma inteligência ou algum mecanismo de controle que avalie cada situação e encontre dinamicamente os parâmetros mais adequados. Cada uma dessas abordagens tem

vantagens e desvantagens.

A primeira envolve uma menor complexidade no ponto de acesso - que normalmente desempenha o papel de coordenador híbrido - porém requer um estudo amplo das situações de carga mais prováveis. Seja, por exemplo, uma WLAN que oferece dois níveis de serviço, implementados em duas categorias de acesso, uma com prioridade alta e outra com prioridade baixa. Os valores de AIFS(i) e CWMin(i) de cada categoria podem ser definidos estaticamente, com base em uma estimativa da carga que cada categoria poderá impor à rede. Dessa forma, o HC pode exercer o controle da funcionalidade de QoS facilmente, apenas informando esses valores a cada nova estação que entra no QBSS, sem necessidade de modificá-los no decorrer do tempo.

No entanto, essa definição estática pode causar ineficiência no uso dos recursos da rede nas situações que se desviam das estimativas iniciais. No exemplo da WLAN com duas categorias, caso não exista tráfego de alta prioridade presente na rede por um certo período, não é possível para o tráfego de baixa prioridade explorar nesse período a vazão máxima do canal, uma vez que os valores mais altos de AIFS e CWMin foram estaticamente estabelecidos para a sua categoria. Tal efeito é ainda mais grave em redes que ofereçam mais do que dois níveis de serviço, já que haverá uma maior diversidade nas situações de carga possíveis.

Ao ser empregada a segunda abordagem, é possível tornar a definição desses parâmetros menos dependente das premissas sobre o uso da rede, buscando-se uma maximização do nível de utilização, ao mesmo tempo em que são mantidos os níveis de serviço desejados. A desvantagem, nesse caso, é a maior complexidade adicionada ao ponto de acesso, uma vez que agora ele deve selecionar dinamicamente os parâmetros de diferenciação do conjunto de parâmetros de QoS. Para que essa seleção dinâmica seja possível, é necessário um mecanismo para monitoração e controle de carga dos tráfegos que utilizam o método de acesso com contenção (EDCF).

O projeto do mecanismo de controle de carga pode ser guiado por dois enfoques diferentes, dependendo da previsão de existência, no acesso com contenção, de tráfego caracterizado como sendo de "tempo real". Embora se espere que esse tipo de tráfego utilize o novo método de acesso sem contenção (*HCF polled channel access*),

não há nada que impeça o uso do EDCF pelo tráfego de tempo real. Nesse caso, o enfoque do controle de carga é a preservação dos requisitos de vazão, retardo e perdas desse tipo de tráfego. A definição dinâmica dos elementos do *QoS Parameter Set* permite ajustar o tráfego de melhor esforço para, quando necessário, conceder acesso ao tráfego de tempo real e, quando possível, explorar ao máximo a largura de banda não utilizada por esse tipo de tráfego.

Caso não seja previsto tráfego de tempo real no acesso com contenção, o enfoque do controle de carga passa a ser apenas a maximização da utilização da rede através da definição dinâmica dos parâmetros de QoS, com o cuidado de não alterar as prioridades relativas entre as categorias de acesso ao redefinir esses parâmetros.

No trabalho descrito nesta dissertação, em particular, é proposto um controle de carga que pressupõe a coexistência de tráfegos de tempo real e de melhor esforço no acesso com contenção, tendo enfoque portanto na preservação de requisitos de qualidade como retardo e perdas.

Apenas controlar a carga do tráfego de melhor esforço não é suficiente para garantir QoS. É necessário ainda limitar o número de fluxos de tempo real presentes na rede, para evitar que a capacidade máxima do canal seja exaurida, o que acarreta perda das garantias de qualidade.

A função desempenhada pelos nós da rede, de forma centralizada ou distribuída, no intuito de determinar se uma nova chamada ou um novo fluxo pode ser aceito em uma rede com recursos limitados é chamada de controle de admissão (*Admission Control*).

### 3.3.2 Controle de Admissão

O objetivo principal de qualquer mecanismo de controle de admissão é a preservação da QoS dos fluxos já existentes na rede, quando da chegada de uma nova solicitação. Um novo fluxo só é admitido se, por algum critério de decisão previamente estabelecido, não comprometer as garantias de qualidade já oferecidas aos demais fluxos. O controle de admissão tem ainda como objetivo secundário a maxi-

mização dos níveis de utilização da rede. Também é desejável que tenha baixo custo computacional, já que deve oferecer uma resposta em tempo real às solicitações de entrada dos novos fluxos.

Uma vez que muitos esquemas de controle de admissão já foram propostos em estudos de redes fixas, a adaptação de alguns desses mecanismos para uso em cenários de redes móveis passa a ser uma alternativa atraente. Um tipo de controle de admissão muito estudado em redes IP e ATM é o controle de admissão baseado em medidas (*MBAC - Measurement-Based Admission Control*) [3, 4, 5, 6].

### Controle de Admissão Baseado em Medidas

O controle de admissão baseado em medidas não requer uma especificação de tráfego muito precisa para exercer sua função. Isso é uma vantagem em relação aos métodos tradicionais de controle de admissão, baseados em parâmetros (*PBAC-Parameter-Based Admission Control*), os quais utilizam descrições de tráfego que seguem modelos determinísticos ou estocásticos. Grossglauser e Tse [5] apontam dois problemas no controle baseado em parâmetros: a dificuldade de se caracterizar o tráfego de fontes com comportamento em rajadas, o que pode levar a erros que superestimam ou subestimam a necessidade real de recursos, e a dificuldade de policiar este tráfego estatisticamente modelado para evitar que utilize mais recursos do que o que foi alocado por ocasião da admissão. O MBAC é mais imune a esses problemas, embora tenha maior custo computacional e ofereça garantias menos absolutas que os métodos baseados em parâmetros.

O MBAC também não necessita guardar informações sobre os fluxos já admitidos para avaliar a quantidade de recursos disponível na rede. Essa quantidade é estimada usando-se medidas realizadas diretamente sobre o tráfego presente na rede a cada instante. As decisões são feitas comparando-se os parâmetros do novo fluxo com as estimativas. Essa abordagem ameniza os problemas decorrentes das imprecisões na caracterização de fluxos, mas diminui a capacidade do controle de admissão de oferecer garantias absolutas a aplicações pouco tolerantes às variações do retardo e da taxa de perdas.

### Componentes do Controle de Admissão

Um controle de admissão baseado em medidas tem dois componentes básicos: um mecanismo de medição, usado para estimar a carga atual da rede, e um algoritmo de decisão, que usa a estimativa levantada para decidir se o fluxo será ou não aceito. Ambos podem usar técnicas variadas.

Jamin *et al* [3] citam como exemplos de técnicas usadas nos mecanismos de medição as técnicas de janelas de tempo (*Time-Window*), amostras de pontos (*Point Samples*) e média exponencial (*Exponential Averaging*). No mesmo trabalho, são citadas para algoritmos de decisão as técnicas de soma medida (*Measured Sum*), região de aceitação (*Acceptance Region*) e banda equivalente (*Equivalent Bandwidth*). Breslau *et al* [6] mostram que as diversas técnicas tanto de algoritmos de decisão quanto de mecanismos de medição apresentam desempenho muito similar na tarefa de garantir QoS e ao mesmo tempo promover uma alta utilização da rede. Os autores mostram ainda que essas técnicas são intercambiáveis, ou seja, pode-se utilizar praticamente qualquer mecanismo de medição com qualquer algoritmo de decisão.

Considerando a equivalência de desempenho apresentada por Breslau *et al*, a escolha de uma ou outra técnica para os componentes do MBAC não é uma questão crítica. Um esquema de controle de admissão muito comum utiliza decisão por soma medida e medição por janelas de tempo [4].

No algoritmo de soma medida, a carga do tráfego existente é estimada através de medições feitas diretamente na rede e é somada à vazão solicitada pelo novo fluxo para decidir sobre a admissão. Um novo fluxo será aceito se

$$V + V_f < C \cdot \mu \quad (3.2)$$

ou seja, a soma da estimativa de carga ( $V$ ) com a vazão solicitada pelo novo fluxo ( $V_f$ ) for menor que a capacidade do canal ( $C$ ) multiplicada por um fator de utilização ( $\mu$ ) pré-definido. O fator de utilização é um valor entre 0 e 1. Jamin *et al* [4] afirmam que "a variância do retardo em filas M/M/1 diverge à medida que o sistema se aproxima da utilização máxima. Uma abordagem MBAC é fadada a falhas quando

as variações de retardo são excessivamente grandes, o que irá ocorrer quando o nível de utilização for muito alto. Assim sendo, é necessário identificar um alvo de utilização e fazer com que o algoritmo de controle de admissão mantenha a utilização abaixo desse nível". O fator de utilização é usado para estabelecer esse nível no algoritmo de soma medida.

Ao ser aceito o novo fluxo, a estimativa de carga é atualizada somando-se os valores de  $V$  e  $V_f$ .

Na técnica de medição por janelas de tempo (Figura 3.5), o mecanismo de medição, também chamado de estimador, divide o tempo em janelas de tamanho  $T$  segundos, e subdivide cada janela em intervalos de amostra de tamanho  $S$ . A carga média observada na rede é computada a cada intervalo de amostra. Ao final de uma janela de tempo, a maior das amostras de carga obtidas nessa janela é tomada como a estimativa de carga inicial a ser usada pelo algoritmo de decisão na janela seguinte. No decorrer de cada janela, se for obtida uma amostra de carga maior que a estimativa atual, o valor dessa amostra passa a ser considerado a estimativa válida para a janela atual. Quando um novo fluxo é admitido, a estimativa de carga é atualizada somando-se a ela o valor de  $V_f$ . Nesse caso, uma nova janela é iniciada.

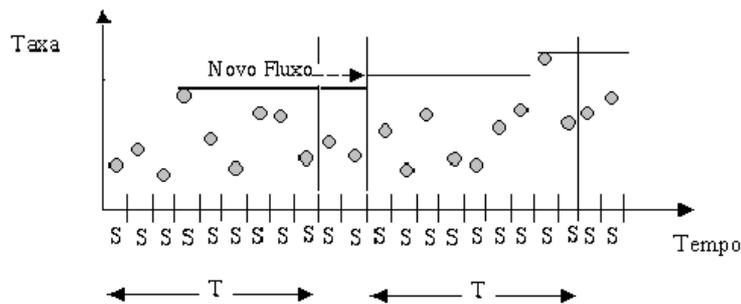


Figura 3.5: Mecanismo de Medição por Janela de Tempo

A Figura 3.5 ilustra o mecanismo de medição por janelas de tempo. Os pontos de cor cinza representam as amostras de carga de cada intervalo  $S$ . A amostra de valor mais alto em cada janela de tamanho  $T$  é a estimativa de carga válida para a janela atual e para a seguinte. Essa estimativa é representada na figura por uma

linha horizontal. Seu valor aumenta quando é encontrada uma amostra de valor mais alto ou quando um novo fluxo é admitido. Esse valor diminui quando, transcorrida uma janela inteira, nenhuma amostra superior à estimativa foi obtida. Nesse caso, a maior das amostras da janela atual (que é menor que a estimativa) será usada para iniciar a próxima janela.

## 3.4 Considerações Finais

Neste capítulo foram abordados os principais conceitos relacionados à qualidade de serviço, bem como as técnicas de diferenciação elaboradas para o MAC das redes locais sem fio. O estudo de tais técnicas as revela como formas bastante efetivas de estabelecimento de diferenciação nesse tipo de rede. Esse estudo culmina com a elaboração da extensão 802.11e, um grande incremento ao valor do padrão 802.11 como opção tecnológica para o acesso móvel.

Foi mostrado ainda que a inclusão de um suporte à QoS no 802.11 motiva o desenvolvimento de mecanismos adicionais que garantam um uso eficiente da nova funcionalidade e dos recursos da rede. Tais mecanismos buscam resolver questões relacionadas ao gerenciamento de QoS de uma forma geral. Dessas questões, o controle de carga das categorias de acesso é de interesse particular para o trabalho descrito nesta dissertação. Um mecanismo de controle é proposto para resolver essa questão, complementado por um controle de admissão para o tráfego de tempo real. Nos capítulos seguintes, esses mecanismos serão apresentados e avaliados.

## Capítulo 4

# Mecanismos de Controle Propostos

**O**S mecanismos de controle de admissão e de controle de carga propostos neste trabalho procuram manter a complexidade do controle de QoS restrita ao ponto de acesso 802.11e. Esses mecanismos são executados em conjunto pelo ponto de acesso, que se utiliza apenas da informação localmente disponível sobre os fluxos que entram e saem na rede sem fio para tomar suas decisões de controle.

Assume-se que as estações que geram tráfegos com requisitos de QoS sejam compatíveis com a especificação 802.11e. Essas estações são capazes de interpretar as informações do conjunto de parâmetros de QoS usado no método de acesso com contenção (EDCF). Também supõe-se que esteja disponível um protocolo de sinalização e estabelecimento de conexões ou chamadas, usado pelas estações para requisitar a admissão de novos fluxos e pelo ponto de acesso para informar a decisão do controle de admissão. Entre o estabelecimento da conexão e a desconexão, nenhuma comunicação sobre o estado dos fluxos é requerida entre as estações e o ponto de acesso.

A adaptação de um controle de admissão baseado em medidas é descrita na Seção 4.1 deste capítulo. O controle de carga por parametrização dinâmica de QoS é apresentado na Seção 4.2. A Seção 4.3 traz as considerações finais do capítulo.

## 4.1 Adaptação do MBAC para o Caso Sem Fio

O mecanismo de controle de admissão proposto é uma adaptação de um controle de admissão baseado em medidas muito comum em redes fixas. Esse controle de admissão usa como componentes básicos um algoritmo de decisão por soma medida e um mecanismo de medição por janelas de tempo.

O algoritmo de soma medida não pode ser utilizado diretamente em uma rede 802.11 sem antes sofrer uma adaptação, uma vez que a capacidade do canal ( $C$ ) nesse tipo de rede não é fixa. Conforme discutido na Seção 3.2.1, fatores como a duração dos intervalos entre-quadros (IFS), as taxas de dados básica e máxima em uso no BSS, o tamanho dos quadros de dados e o tempo médio consumido em *backoff* afetam essa capacidade. Na proposta deste trabalho, a inequação

$$V + V_f < C \cdot \mu \quad (4.1)$$

usada no algoritmo original, passa a ser

$$V + V_f < B \cdot \alpha \cdot \mu \quad (4.2)$$

onde a capacidade do canal ( $C$ ) é substituída pelo produto de  $B$  - a taxa máxima de dados em uso no BSS - por um fator  $\alpha$  - uma estimativa de eficiência do canal. Essa estimativa é calculada através da relação  $\alpha = t_{MSDU}/t_m$ , onde  $t_{MSDU}$  é um tempo teórico de transmissão do quadros de dados sem a sobrecarga do método de acesso e  $t_m$  é o tempo de transmissão incluindo essa sobrecarga.  $t_{MSDU}$  é igual ao tamanho (em bits) da unidade de dados de serviço do MAC (MSDU) dividido pela taxa de transmissão de dados máxima ( $B$ ). O tempo de transmissão com sobrecarga ( $t_m$ ) é estimado pela equação<sup>1</sup>

$$t_m = DIFS + t_{dados} + SIFS + t_{ACK} + t_{backoff}(m) \quad (4.3)$$

<sup>1</sup>Esse cálculo desconsidera o retardo de propagação no BSS e o uso de RTS/CTS. Se o acesso com RTS/CTS for usado, deve-se somar o tempo de transmissão desses dois quadros e mais duas vezes SIFS.

introduzida na Seção 3.2.1.  $t_{dados}$  e  $t_{ACK}$ , como demonstrado naquela seção, são influenciados pela taxa básica de dados. Essa taxa é usada na transmissão dos pacotes de controle, do preâmbulo do nível físico e do cabeçalho do protocolo de convergência de nível físico (PLCP), a fim de proporcionar compatibilidade entre estações que usam taxas de dados diferentes em um mesmo BSS.

*DIFS* e *SIFS* são definidos no padrão, de acordo com o nível físico em uso. A Tabela 4.1 mostra os valores usados no 802.11b e no 802.11a.

Tabela 4.1: Valores de SIFS, SlotTime e DIFS usados no 802.11

Valores	802.11b DSSS	802.11a
SIFS	10 $\mu s$	16 $\mu s$
<i>SlotTime</i>	20 $\mu s$	9 $\mu s$
DIFS ( $SIFS + 2 \cdot SlotTime$ )	50 $\mu s$	34 $\mu s$

O tempo médio gasto em *backoff* ( $t_{backoff}(m)$ ) é dependente do número de instâncias do EDCF competindo pelo meio ( $m$ ) e das taxas de geração de dados das fontes. Esses dois fatores afetam a carga imposta à rede. Quanto maior a carga, maior a probabilidade de colisão e maior o número de retransmissões necessárias para uma transmissão bem sucedida.

O tamanho dos quadros de dados gerados pelas fontes pode ser fornecido através da especificação de tráfego, pois geralmente é um parâmetro bem conhecido dos codificadores (de voz, vídeo ou áudio, por exemplo). Caso essa informação não esteja disponível, pode ser feita uma média observando-se o tamanho dos quadros na fila de transmissão do ponto de acesso.

A vazão solicitada pelo novo fluxo ( $V_f$ ) na camada de aplicação deve ser corrigida para refletir a verdadeira carga imposta pelo fluxo ao canal sem fio na subcamada MAC, uma vez que todas as medidas do controle de admissão proposto são feitas na fila de transmissão da interface sem fio do ponto de acesso. A correção consiste em incluir no tamanho dos pacotes a sobrecarga dos protocolos envolvidos na transmissão, como os protocolos de transporte UDP (*User Datagram Protocol*) e TCP (*Transmission Control Protocol*), o protocolo de rede IP (*Internet Protocol*) e o protocolo de transporte em tempo real RTP (*Real-time Transport Protocol*).

O padrão de tráfego da aplicação também deve ser considerado no controle de admissão em redes 802.11. Aplicações com padrão bidirecional ou conversacional, como telefonia IP e videoconferência, geram dados nos dois sentidos se forem aceitas. Aceitar uma chamada de voz em uma rede 802.11 implica assumir que a capacidade disponível é suficiente para acomodar tanto o tráfego de subida (*uplink* ou *upstream*, isto é, das estações para o ponto de acesso) quanto o de descida (*downlink* ou *downstream*, do ponto de acesso para as estações). Enquanto em uma rede celular, por exemplo, esses tráfegos ocupariam canais separados, no 802.11 eles competem pelo mesmo canal. Como a estimativa de carga é coletada pelo estimador na fila de transmissão da interface sem fio do ponto de acesso, refletindo apenas o tráfego de descida, o mecanismo de controle de admissão proposto só pode alocar, para o caso do tráfego bidirecional, metade da capacidade máxima da rede, já que está implicitamente admitindo o tráfego de subida ao admitir o de descida.

Serviços de vídeo ou áudio (*video/audio streaming*), por outro lado, geram um tráfego assimétrico, predominantemente no sentido de descida, já que os servidores de áudio e vídeo ficam geralmente localizados na rede fixa. O tráfego de subida compreende apenas os quadros de reconhecimento e de controle dos protocolos usados na transmissão e as interações dos usuários com o servidor, como a solicitação de um arquivo de vídeo em particular. A admissão, nesse caso, pode ser feita levando em conta apenas o sentido predominante do fluxo de dados.

## 4.2 Controle de Carga

O mecanismo de controle de carga proposto é baseado na seleção e configuração dinâmica dos parâmetros de QoS utilizados no acesso com contenção (EDCF), de acordo com a situação de carga. Os parâmetros são informados às estações através da emissão periódica do conjunto de parâmetros de QoS (*QoS Parameter Set*), criado com a extensão 802.11e. A exemplo do controle de admissão, o mecanismo tem o duplo objetivo de garantir os requisitos de QoS e manter elevada a utilização da rede.

O ponto central da proposta de controle de carga contida no presente trabalho

é o desenvolvimento dos algoritmos para a seleção dinâmica dos parâmetros que afetam o acesso das categorias, em particular o  $AIFS(i)$  e o  $CWMin(i)$ . Antes que esses algoritmos sejam apresentados, é feito um estudo da influência dos dois parâmetros sobre a vazão e o retardo das categorias de acesso e de seus efeitos sobre a diferenciação. São ainda levantadas algumas restrições a serem observadas na seleção dos dois parâmetros.

O controle assume a coexistência de tráfego de tempo real com tráfego de melhor esforço no EDCF, tendo um enfoque direcionado para a preservação de limites sobre os valores como retardo e perdas das categorias mais prioritárias. O posicionamento do controle de carga em relação a outros mecanismos e outras entidades do sistema é mostrado a seguir.

### 4.2.1 Posicionamento do Controle de Carga

O mecanismo de controle de carga se comunica com outros mecanismos de controle disponíveis e outras entidades de gerenciamento a fim de cumprir sua função. A Figura 4.1 mostra essa comunicação.

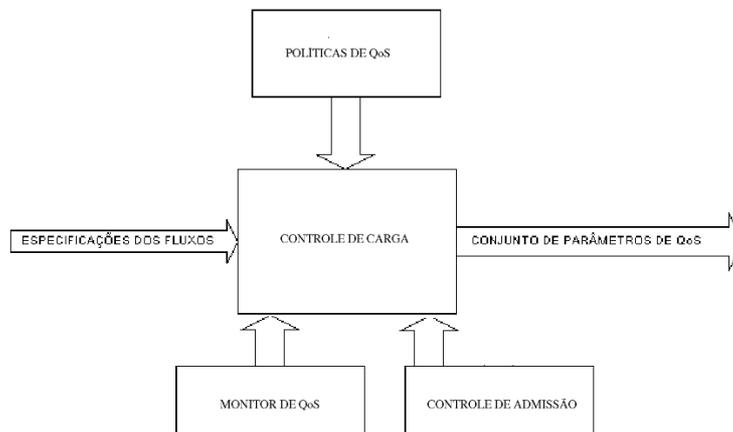


Figura 4.1: Posicionamento do Controle de Carga

A saída do mecanismo de controle de carga é o conjunto de parâmetros de QoS. Para que esse conjunto possa ser composto, o controle de carga recebe como entrada as especificações de tráfego, as políticas de QoS, dados do controle de admissão e a

situação atual da rede, fornecida por um módulo monitor de QoS.

As especificações de tráfego contêm, por exemplo, os valores de retardo, vazão e perdas esperados para cada categoria de acesso  $i$ , ou seja, os limites ou faixas de valores a serem garantidos.

As políticas de QoS configuradas pelo administrador da rede definem, por exemplo, o número de categorias suportadas pelo ponto de acesso, os parâmetros iniciais de AIFS e CWMin dessas categorias e, se for o caso, uma percentagem máxima de utilização do enlace para cada uma delas.

O controle de admissão fornece dados ao controle de carga como o número de fontes admitidas na(s) categoria(s) mais prioritária(s) e a carga medida pelo estimador (mecanismo de medição) do controle de admissão.

O módulo monitor de QoS fornece a situação atual da rede. Esse módulo é um conjunto de rotinas que inspeciona periodicamente as filas de transmissão das categorias. A informação retornada pelo monitor de QoS inclui dados como o tamanho das filas, a taxa de perdas de pacotes por descarte, a latência (retardo) de acesso e a carga de cada categoria. Com base na situação de carga, nas especificações de tráfego e nas políticas, o controle emite um novo conjunto de parâmetros de QoS que provoca aumento ou redução da carga das categorias.

Os dois parâmetros de QoS de particular interesse para a proposta de controle desta dissertação são o  $AIFS(i)$  e o  $CWMin(i)$ . Antes que se descreva como são empregados pelo controle de carga, é necessário conhecer a influência desses parâmetros sobre o retardo de acesso e a vazão de fontes individuais, bem como sobre a vazão agregada de uma categoria de acesso.

### 4.2.2 Influência dos Parâmetros AIFS e CWMin

As técnicas de diferenciação usando o tamanho do intervalo entre-quadros do EDCF (AIFS) e o tamanho mínimo da janela de contenção (CWMin) produzem efeitos aparentemente equivalentes, pois ambas regulam o tempo de espera. Há, no entanto, algumas diferenças a serem consideradas.

O AIFS controla uma parte determinística da espera, enquanto o CWMin controla uma parte probabilística. Essa diferença traz conseqüências para o controle de carga: as mudanças de valor de AIFS provocam uma resposta mais imediata, enquanto os reflexos das mudanças de CWMin só são perceptíveis tomando-se uma média sobre um intervalo de tempo.

O CWMin também pode afetar o retardo médio de acesso de uma fonte muito mais gravemente que o AIFS. Uma vez que esse parâmetro só assume valores que são potências de 2 subtraídas de 1 unidade (31,63,127...), a cada mudança de CWMin o número de *slots* usados na espera dobra. Com AIFS, é possível acrescentar um *slot* de cada vez ao tempo de espera, o que permite um controle menos agressivo.

Outras diferenças podem ser constatadas estudando-se algumas situações de sobrecarga da rede, nas quais se espera que o controle de carga tenha maior utilidade. As Figuras 4.2, 4.3 e 4.4 mostram o efeito da alteração do número de *slots* usados no AIFS sobre a vazão agregada obtida por conjuntos de 1, 8 e 16 fontes, respectivamente, em experimentos realizados neste trabalho a fim de explorar os cenários de sobrecarga.

As fontes pertencem a uma mesma categoria de acesso e geram dados a uma taxa de bits constante (*Constant Bit Rate - CBR*), sendo cada fonte iniciada em uma estação distinta. A taxa de geração de cada fonte é tal que a soma das taxas de todas as fontes pertencentes ao conjunto seja igual à capacidade máxima do canal. Todas as estações pertencem a um mesmo BSS 802.11b e utilizam a taxa de transmissão de dados máxima de 11 Mbps. Nas três figuras, cada curva mostra a soma das vazões obtidas pelas fontes em um experimento isolado. Em cada experimento é utilizado um determinado tamanho de pacote ( $p$ ), que varia de 64 a 2304 bytes. Como mostrado na Seção 3.2.1, a vazão máxima obtida é dependente do tamanho de pacote empregado.

Observando-se as Figuras 4.2, 4.3 e 4.4, percebe-se que o aumento do número de *slots* utilizados no AIFS provoca uma redução linear na vazão agregada de cada conjunto, independentemente do número de fontes presentes no BSS. Isso se deve ao fato de que, a cada *slot* adicionado a  $AIFS(i)$ , há um acréscimo do tempo exato

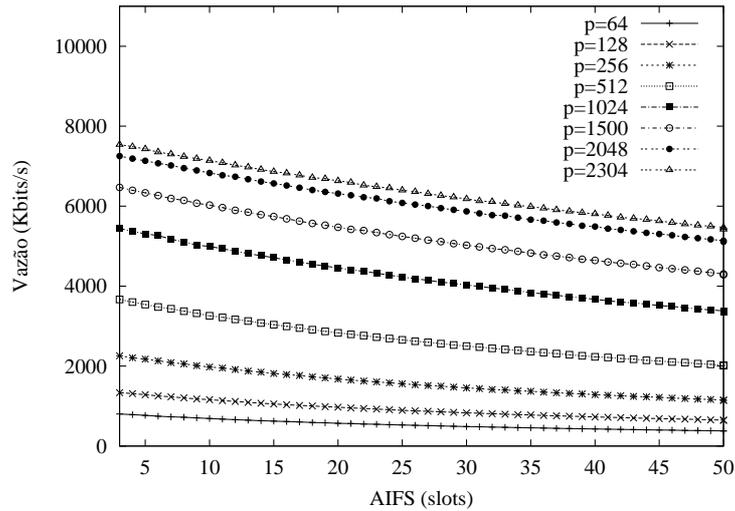


Figura 4.2: Vazão de 1 Fonte CBR em Função do Tamanho de AIFS

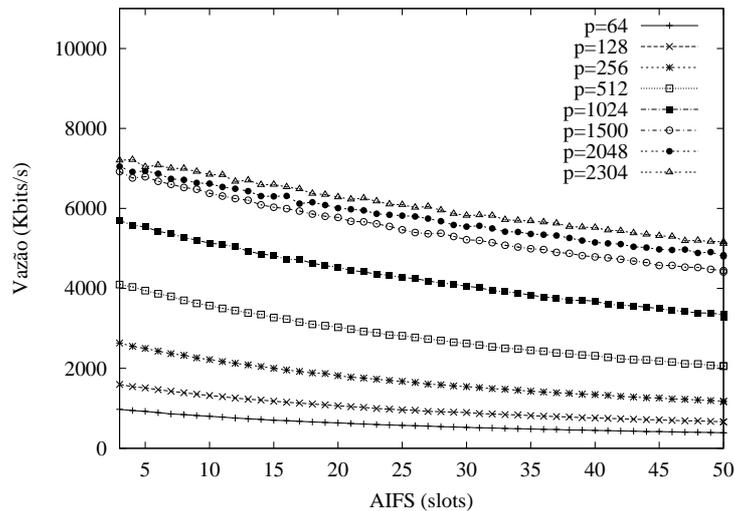


Figura 4.3: Vazão Agregada de 8 Fontes CBR em Função do Tamanho de AIFS

de um *slot* (no caso,  $20\mu s$ ) ao tempo de espera de todas as instâncias do EDCF que competem na categoria  $i$ .

As Figuras 4.5, 4.6 e 4.7 mostram os mesmos experimentos anteriores, realizados agora com a variação do tamanho mínimo da janela de contenção. Os valores para  $CW_{min}$  evoluem de 31 a 4095, fazendo-se  $CW_{novo} = (CW_{antigo} * 2) + 1$  periodicamente. O AIFS é mantido fixo e igual a 2. Através desses experimentos, verifica-se que o efeito do aumento do  $CW_{min}$  sobre a vazão máxima não é tão previsível quanto o do aumento do AIFS.

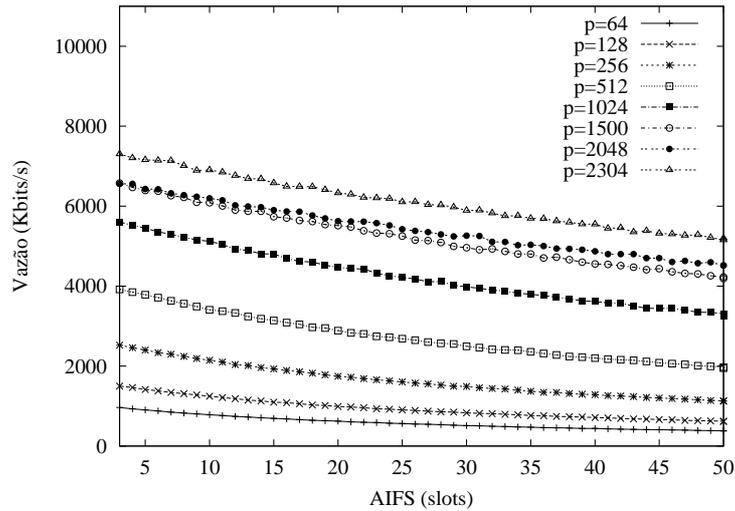


Figura 4.4: Vazão Agregada de 16 Fontes CBR em Função do Tamanho de AIFS

Para uma única fonte, o aumento exponencial do tempo de espera provoca uma redução inversamente proporcional na vazão máxima.

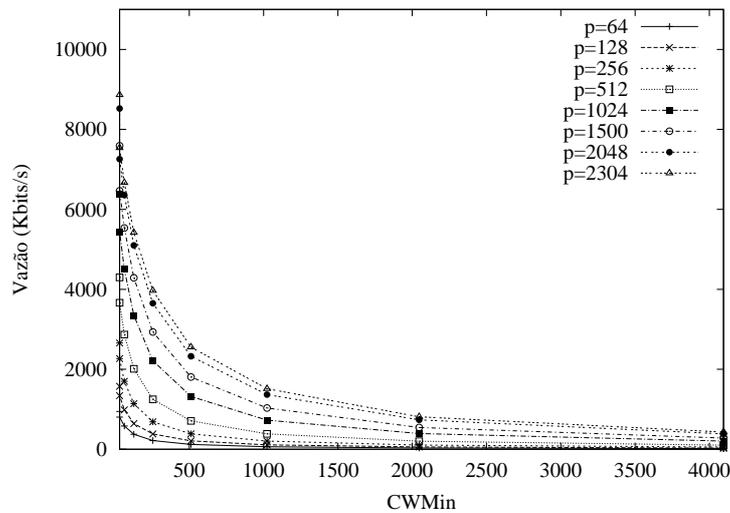


Figura 4.5: Vazão de 1 Fonte CBR em Função do Tamanho de CWMin

Para 8 e 16 fontes, no entanto, observa-se uma curva de perfil diferente, mais imune às variações de CWMin. Isso pode ser explicado considerando-se a importância que o tamanho da janela de contenção tem na prevenção de colisão do CSMA/CA. Em situações de sobrecarga, como as exploradas nesses experimentos, a probabilidade de colisão aumenta com o número de instâncias do MAC que competem pelo meio. Cada instância que perde uma oportunidade de transmissão é obrigada a

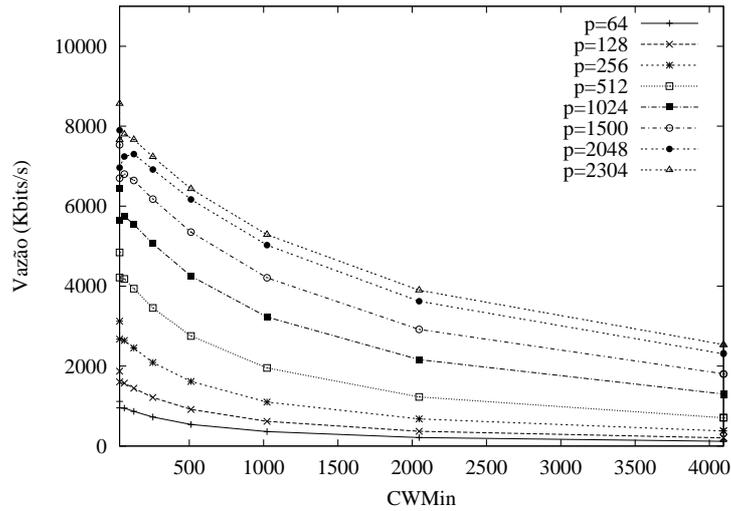


Figura 4.6: Vazão Agregada de 8 Fontes CBR em Função do Tamanho de CWMin

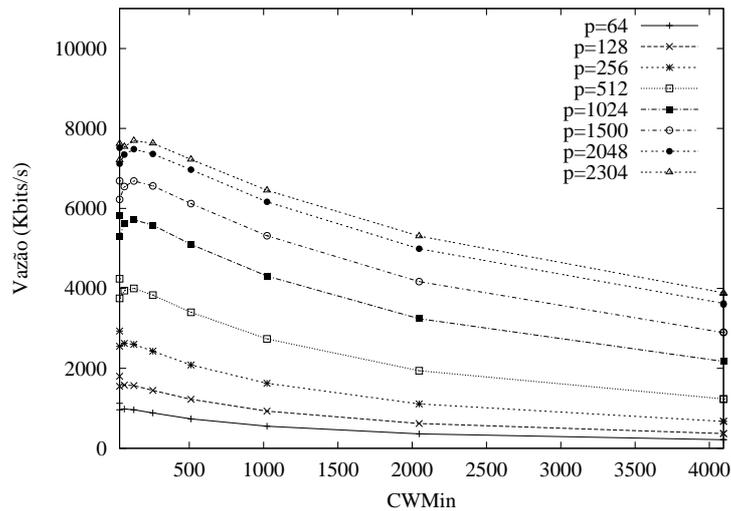


Figura 4.7: Vazão Agregada de 16 Fontes CBR em Função do Tamanho de CWMin

entrar em *backoff* usando um valor maior de CW. O tempo gasto em *backoff* até que se consiga transmitir um quadro é em média

$$\frac{1}{2} \cdot \sum_{i=1}^t [(CWMin + 1) \cdot 2^{i-1}] \cdot SlotTime$$

onde  $t$  é o número de tentativas necessárias para que a transmissão seja bem sucedida. O número médio de tentativas cai quando se usa um valor de CWMin maior, como pode ser visto na Figura 4.8.

Elevando-se o valor de CWMin, reduz-se a probabilidade de que duas instâncias

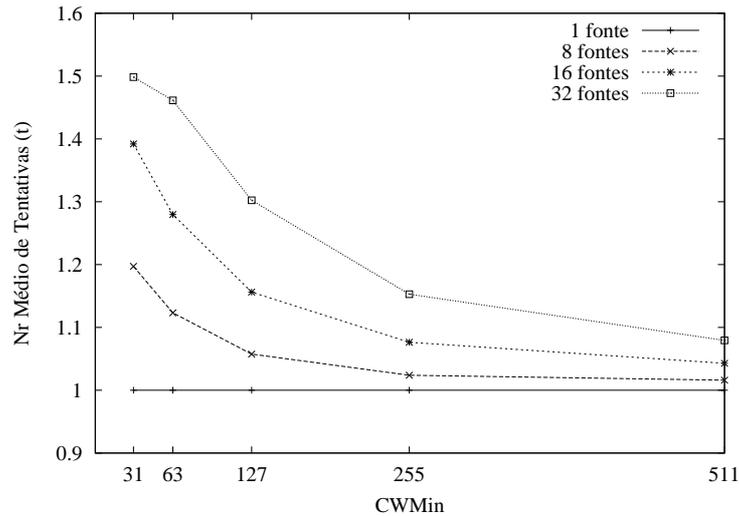


Figura 4.8: Número Médio de Tentativas de Transmissão em Função do Tamanho de CWMin

escolham o mesmo *slot*, aumentando a chance de transmissão na primeira tentativa e nas subsequentes. Por esse motivo, ocorre uma elevação da vazão agregada nas situações com 8 e 16 fontes ao ser alterado o valor de CWMin de 31 para 63, por exemplo. Nessas situações, a vazão agregada só começa a cair quando CWMin assume valores excessivos em relação ao valor ideal para a prevenção de colisão.

Os mesmos efeitos das variações de AIFS e CWMin sobre a vazão agregada são observados quando as fontes CBR são trocadas por fontes FTP (*File Transfer Protocol*) transmitindo continuamente. A única diferença observada é que a vazão agregada é sempre menor em valores absolutos que a vazão obtida com fontes CBR, por causa da atuação do controle de congestionamento do protocolo de transporte TCP (*Transmission Control Protocol*) e da sobrecarga adicionada com o tráfego dos reconhecimentos (ACKs) de segmentos desse mesmo protocolo.

Os experimentos acima mostram que a alteração do CWMin deve ser usada de forma cuidadosa. Elevar o valor de CWMin em uma categoria de acesso em que há muitas instâncias do EDCF competindo pelo meio pode servir para aumentar a vazão agregada dessa categoria, ao invés de diminuí-la.

Com base nas observações sobre a influência dos dois parâmetros, pode-se afirmar que as variações de AIFS e CWMin devem ser empregadas em conjunto, de uma

forma complementar. O controle de carga pode usar preferencialmente o AIFS para definir o grau de separação entre os níveis de prioridade, enquanto o CWMin é alterado de acordo com o número de instâncias do EDCF competindo em um mesmo nível. A Figura 4.9 mostra os dois parâmetros sendo utilizados em conjunto.

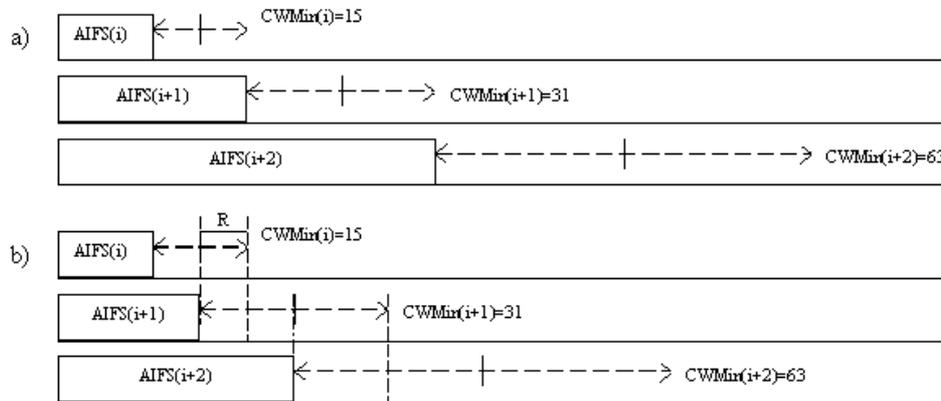


Figura 4.9: Diferenciação por Variação Conjunta de AIFS e CWMin

### 4.2.3 Sobreposição das Janelas de Contenção

A relação entre os diversos valores de AIFS( $i$ ) e CWMin( $i$ ) das categorias de acesso tem reflexos sobre o funcionamento do controle de carga.

Se o AIFS de uma categoria de acesso ( $i+1$ ) for igual ao AIFS + CWMin usados na categoria de acesso( $i$ ), mais prioritária do que ( $i+1$ ), como mostrado na situação "a" da Figura 4.9, uma separação total entre as categorias é garantida. Não há sobreposição entre as janelas dos diversos níveis de prioridade, reduzindo a probabilidade de que um quadro menos prioritário (no caso, de índice  $i$  maior) seja transmitido antes de outro mais prioritário. Uma inversão de prioridade ainda pode ocorrer, embora pouco provável, caso a categoria de acesso de baixa prioridade esteja concluindo a execução de um *backoff* anterior quando a de alta prioridade recebe um novo quadro para transmitir.

A separação sem sobreposição de janelas traz algumas desvantagens. Uma dessas desvantagens é a possibilidade de inanição (*starvation*) da categoria menos prioritária. Se a categoria mais prioritária apresentar carga alta, de forma que haja sempre

algum quadro a transmitir nas filas desse nível, a categoria menos prioritária nunca irá obter uma oportunidade de transmissão e irá sofrer inanição. Esse efeito é mostrado na Figura 4.10, onde duas fontes de categorias diferentes tentam obter a vazão máxima do canal. Nessa figura, à medida que a diferença dos intervalos entre quadros das duas categorias aumenta (conseqüentemente, a sobreposição diminui), a fonte 1, mais prioritária, obtém mais vazão, até dominar totalmente o acesso ao meio.

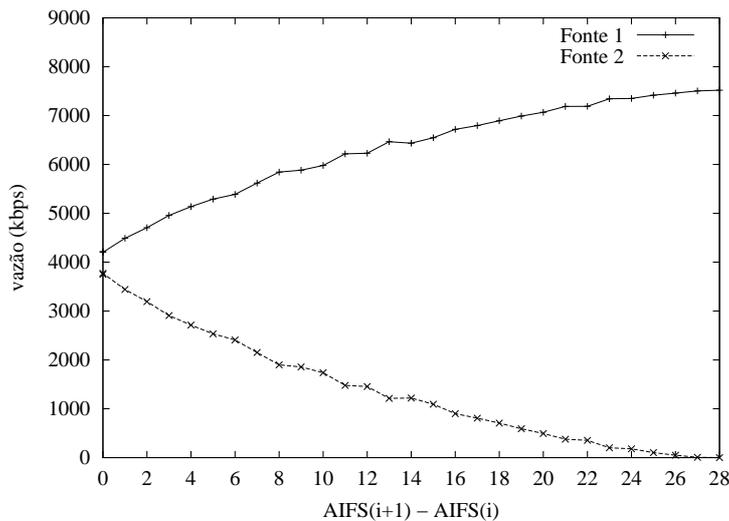


Figura 4.10: Vazão de Duas Fontes de Categorias Distintas em Função da Sobreposição de Janelas

Outra desvantagem da separação sem sobreposição de janelas é que o número de níveis suportáveis fica reduzido, porque o nível mais baixo irá usar valores de AIFS e CWMin muito elevados, tornando impraticável o suporte a até oito prioridades, como previsto na especificação 802.11e.

Se para a categoria  $(i+1)$  o  $AIFS(i+1)$  é menor que  $AIFS(i) + CWMin(i)$ , como mostrado na situação "b" da Figura 4.9, a inanição das categorias de menor prioridade pode ser evitada e um maior número de níveis pode ser suportado. Em compensação, ocorre uma probabilidade de inversão de prioridades ( $P_{inv}$ ) proporcional ao tamanho da região de sobreposição ( $R$ ), que é dado por  $AIFS(i) + CWMin(i) - AIFS(i+1)$ . Sejam, por exemplo, duas categorias de acesso,  $AC(i)$  e  $AC(i+1)$ , que possuem um quadro para transmitir em um mesmo instante. A categoria de menor prioridade ( $AC(i+1)$ ) poderá enviar seu quadro antes da de

maior prioridade se seu temporizador de *backoff* expirar antes do temporizador de  $AC(i)$ . Isso ocorrerá se ambas escolherem valores dentro da região de sobreposição para definir seus temporizadores e se  $AC(i+1)$  escolher um valor cujo índice é menor que o escolhido por  $AC(i)$  nesse conjunto de valores comuns.

Na situação com duas fontes, uma em cada categoria, a probabilidade de que a inversão de prioridade ocorra em uma disputa é

$$P_{inv} = \sum_{j=0}^{R-1} (P_{i+1} = j) \cdot (P_i > (j + AIFS(i+1))), \quad (4.4)$$

onde  $j$  é o índice dos valores na região de sobreposição,  $R$  é o tamanho dessa região,  $(P_{i+1} = j)$  é a probabilidade de que a categoria  $i+1$  escolha um valor igual a  $j$  no intervalo  $[0, CWMin(i+1)]$  e  $(P_i > j + AIFS(i+1))$  é a probabilidade de que a categoria  $i$  escolha um valor maior que  $j + AIFS(i+1)$  no intervalo  $[0, CWMin(i)]$ .

Se há mais de uma estação transmitindo na categoria  $i+1$ , a inversão pode ser causada por qualquer uma dessas estações, aumentando ainda mais a probabilidade de que um quadro da categoria  $i$  seja preterido. A consequência disso é que o controle de carga deve regular o tamanho da região de sobreposição, levando em conta o número de níveis de prioridades a ser suportado, a quantidade de fontes em cada categoria e o grau de diferenciação desejado.

Considerando o que foi discutido nesta seção e na anterior, já é possível visualizar a idéia básica da determinação dos valores dos dois parâmetros pelo controle de carga. Sempre que possível, o controle irá buscar os menores valores de  $AIFS(i)$  e  $CWMin(i)$  que não provoquem violação dos requisitos de qualidade. Se não há carga na categoria  $i$ , o controle pode minimizar a parte determinística da espera das categorias desde  $i+1$  até a de mais baixa prioridade, reduzindo os valores de  $AIFS$  dessas categorias. À medida que a carga na categoria  $i$  aumenta, o controle aumenta o  $CWMin$  da categoria, o que provoca a necessidade de aumentar a parte determinística da espera das categorias menos prioritárias que  $i$ , para evitar uma sobreposição excessiva da parte probabilística. O controle repete esse procedimento periodicamente e para cada categoria, iniciando sempre pela de maior prioridade.

Os algoritmos do controle de carga utilizados para determinar os valores de  $AIFS$

e CWMin são mostrados na Seção 4.2.5. Algumas restrições utilizadas pelo controle são apresentadas a seguir.

#### 4.2.4 Restrições

O controle de carga deve observar algumas restrições na determinação dos valores de AIFS e CWMin.

- Restrição 1: Em qualquer situação, o valor de AIFS( $i$ ) de qualquer categoria  $i$  deve ser sempre maior ou igual a 2, para evitar falha no funcionamento do MAC. Como o intervalo entre-quadros da categoria é definido por  $SIFS + AIFS(i) \cdot SlotTime$ , o valor de AIFS( $i$ ) não pode ser 0, para evitar que esse intervalo fique igual a SIFS, nem pode ser 1, para não se igualar ao intervalo entre-quadros do PCF (PIFS), que é igual a  $SIFS \cdot SlotTime$ .
- Restrição 2: Em qualquer situação, o valor de AIFS( $i+1$ ) nunca poderá ser menor que o de AIFS( $i$ ), para preservar a diferenciação, devendo ser maior ou igual. Se AIFS( $i+1$ ) for igual a AIFS( $i$ ), então CW( $i+1$ ) deve ser obrigatoriamente maior que CW( $i$ ). Isso permite que as categorias menos prioritárias explorem ao máximo a capacidade do canal ao mesmo tempo em que mantém a possibilidade de diferenciação caso a atividade nas categorias mais prioritárias reinicie.
- Restrição 3: AIFS( $i+1$ ) deve ser menor que AIFS( $i$ ) + CW( $i$ ) se a carga da categoria  $i$  for alta, para evitar inanição na categoria  $i + 1$ .
- Restrição 4: O CWMin de qualquer categoria nunca pode ser maior que o valor de CWMax.
- Restrição 5: Nas situações de pouca carga na categoria mais prioritária, o CWMin dessa categoria não deve ser menor do que 7, considerando-se que esse valor é um mínimo razoável para permitir um compromisso entre baixo retardo e prevenção de colisão. A categoria de melhor esforço (menos prioritária) não deverá usar CWMin menor que 31, o mesmo empregado pelas estações legadas

(802.11/11b/11a), para evitar que tal categoria tenha prioridade sobre essas estações, deixando assim de ser de melhor esforço. Em geral, uma categoria  $i$  não pode ter CWMin menor que  $2^{3+i} - 1$ , para que a restrição 2 também seja atendida.

### 4.2.5 Algoritmos

O controle de carga é executado periodicamente, de acordo com um intervalo de monitoração. A cada execução, o controle verifica se o tamanho do CWMin usado nas categorias está adequado, utilizando a função `ajustarContencao(i)`, a partir da prioridade mais alta. Em seguida, realiza o ajuste da carga imposta pelas categorias de menor prioridade, aumentando ou diminuindo os valores de AIFS e CWMin dessas categorias, sempre tendo em vista as restrições e os objetivos de diferenciação, como retardo limitado ou vazão mínima em uma determinada categoria. Após ter executado essas duas operações em todas as categorias, o controle emite um novo conjunto de parâmetros de QoS.

As Figuras 4.11, 4.12 e 4.13 mostram os algoritmos utilizados para o caso de duas categorias, explorando a coexistência de tráfego de tempo real com tráfego de melhor esforço.

```

controlarCarga() /* executado a cada intervalo de monitoração */
{
    ajustarContencao(i)
    ajustarCarga(i+1)
    emitirConjuntoDeParametrosDeQoS()
}

```

Figura 4.11: Módulo Principal do Controle de Carga

Na Figura 4.11 está o módulo principal do controle. No caso de duas categorias, o módulo chama o procedimento `ajustarContencao` para a categoria mais prioritária, mostrado na Figura 4.12.

No procedimento `ajustarContencao()` é verificado se o tamanho mínimo da janela de contenção da categoria deve ser maior para reduzir o número de tentativas de

```

ajustarContencao(i) /* ajusta o CWMin da categoria dependendo do
                    número de instâncias do EDCF nessa categoria */
{
  CWAntigo = CWMin(i)
  se FontesAceitas > CWMin(i) /* FontesAceitas vem do Controle de Admissão */
    CWMin(i) = CWMin(i) * 2 + 1
  se FontesAceitas < CWMin(i) / 2
    CWMin(i) = (CWMin(i) - 1) / 2
  se CWMin(i) < 2**(3+i) - 1
    CWMin(i) = 2**(3+i) - 1 /* restrição 5: CWMin=7 para i=0 */
  se CWMin(i) > CWAntigo /* houve aumento no CWMin da categoria */
    AIFS(i+1) = AIFS(i+1) + [(CWMin(i) - CWAntigo) * δ]
  se CWMin(i) < CWAntigo /* houve redução no CWMin da categoria */
    AIFS(i+1) = AIFS(i+1) - [(CWAntigo - CWMin(i)) * δ]
  se AIFS(i+1) < 2
    AIFS(i+1) = 2 /* restrição 1 */
}

```

Figura 4.12: Ajuste de Contenção na Categoria de Tempo Real

transmissão. Isso é feito comparando-se o CWMin com o número de fontes aceitas na categoria, que é informado pelo controle de admissão do tráfego de tempo real. É utilizada uma regra simples: se o número de fontes for maior que o número de *slots* usados no CWMin, então CWMin deve ser maior. Essa regra poderia ser mais elaborada, levando em conta um cálculo de probabilidades de colisão, mas isso seria dependente das características do tráfego gerado. O procedimento verifica também se CWMin pode ser reduzido, comparando a metade desse valor com o número de fontes aceitas.

Caso o tamanho de CWMin da categoria de tempo real seja alterado, haverá uma ampliação ou redução da parte determinística da espera da categoria de melhor esforço. Isso é feito somando-se ao AIFS da categoria  $i+1$  a diferença entre o CWMin antigo e o novo da categoria  $i$ . Se for somada toda a diferença ao  $AIFS(i+1)$ , o tamanho da região de sobreposição das partes probabilísticas da espera de  $i$  e  $i+1$  se mantém fixo em número de *slots*. No entanto, a probabilidade de inversão de prioridade diminui, porque a categoria mais prioritária passa a ter menor risco de selecionar um *slot* dentro da região de sobreposição quando usa um valor maior de CWMin. Conseqüentemente, o aumento de  $AIFS(i+1)$  usando toda a diferença provoca uma ampliação do grau de diferenciação entre as categorias, sacrificando excessivamente a vazão da categoria menos prioritária.

É possível utilizar apenas uma fração dessa diferença, multiplicando-a por um

fator  $\delta$ , que varia de 0 a 1. Isso é feito no algoritmo usando-se a expressão

$$\lceil (CWMin(i) - CWAntigo) * \delta \rceil.$$

Um valor de  $\delta$  igual a 0,1 indica que apenas 10% da diferença entre os valores antigo e novo do CWMin da categoria mais prioritária será utilizado para aumentar ou diminuir o valor de AIFS da categoria menos prioritária. Um  $\delta$  igual a 1,0 significa que toda a diferença será utilizada. A avaliação do impacto do controle de sobreposição de janelas proporcionado pelo parâmetro  $\delta$  é realizada em seção posterior, variando-se esse parâmetro de 0,1 até 1,0, com incrementos de 0,1.

Em seguida, o módulo principal chama o procedimento ajustarCarga(i+1), mostrado na Figura 4.13. Esse procedimento verifica se a carga da categoria de melhor esforço pode ser elevada ou reduzida. Se o retardo experimentado pela categoria de tempo real estiver acima de um máximo, a carga do melhor esforço deve ser reduzida. Se estiver abaixo de um "mínimo", que representa uma condição de sobra de recursos para o tipo de tráfego de tempo real, a carga do melhor esforço pode ser elevada.

```

ajustarCarga(i+1) /* verifica os limites de qualidade */
{
    se (RetardoAtual(i) > RetardoMáximo(i))
        reduzirCarga(i+1)
    senão se (RetardoAtual(i) < RetardoMínimo(i)) e (CargaAtual(i) <= CargaAnterior(i)
- CargaMediaDaFonte(i))
        aumentarCarga(i+1)
}
reduzirCarga(i+1) /* baixa a carga da categoria i */
{
    se (AIFS(i+1) < (AIFS(i)+CWMin(i))) /* restrição 3 */
        AIFS(i+1) = AIFS(i+1) + SlotsReducao
    senão se (CWMin(i+1) < CWMax) /* restrição 4 */
        CWMin(i+1) = (CWMin(i+1) * 2) + 1
}
aumentarCarga(i+1) /* eleva a carga da categoria i */
{
    se (CWMin(i+1) > 31) /* restrição 5 */
        CWMin(i+1) = (CWMin(i+1) - 1) / 2
    senão se (AIFS(i+1) > 2) /* restrição 1 */
        AIFS(i+1) = AIFS(i+1) - SlotsAumento
}

```

Figura 4.13: Ajuste da Carga da Categoria de Melhor Esforço

Para evitar instabilidade do controle, uma condição adicional ( $CargaAtual(i) <=$

$CargaAnterior(i) - CargaMediaDaFonte(i)$  é colocada no aumento de carga. Essa condição sugere que houve a saída de uma fonte de tempo real do período anterior para o atual. A carga média da fonte pode ser obtida a partir das especificações de tráfego.

A redução da carga é feita inicialmente somando-se um número de *slots* a AIFS (*SlotsReducao*), até o limite da restrição 3. A partir desse ponto, CWMin passa a ser dobrado, até o limite da restrição 4. A elevação é feita de modo inverso, respeitando as restrições 1 e 5 e utilizando um número de *slots* (*SlotsAumento*) diferente. Utilizar *SlotsAumento* menor que *SlotsReducao* faz com que o controle seja menos agressivo na elevação do que na redução de carga, o que também evita instabilidade do mecanismo.

Depois que esses procedimentos são executados nas categorias, um novo conjunto de parâmetros de QoS foi composto e deve ser transmitido para as estações.

#### 4.2.6 Configuração do Controle de Carga

Os valores de alguns parâmetros utilizados pelo próprio controle podem ter influência no desempenho do mecanismo, tais como:

- intervalo de monitoração: a princípio, um intervalo menor reduz a possibilidade de que as garantias de qualidade do tráfego mais prioritário sejam violadas por períodos mais longos. Em compensação, ocorre um maior número de execuções do módulo de controle e uma maior carga de processamento no ponto de acesso;
- tamanho da sobreposição de janelas: sobreposições maiores elevam a probabilidade de colisão e reduzem as garantias de QoS, mas permitem que um maior número de níveis seja utilizado. O parâmetro  $\delta$ , usado em `ajustarContencao()`, permite estabelecer um compromisso entre o tamanho da janela de sobreposição e a probabilidade de inversão de prioridade. Se for usado  $\delta$  igual a 1, a sobreposição se mantém fixa em número de *slots*, mas a probabilidade de inversão passa a ser menor, porque a categoria  $i$  tem menor probabilidade de

escolher um *slot* dentro da região de sobreposição quando  $CWMin(i)$  é dobrado. Utilizando-se um  $\delta$  menor que 1, é possível tentar manter a probabilidade de inversão igual após dobrar  $CWMin(i)$ , sem necessidade de aumentar demasiadamente o valor de  $AIFS(i+1)$ ;

- número de *slots* usados na redução e no aumento de carga (*SlotsReducao* e *SlotsAumento*): dependem do quanto da carga do melhor esforço se deseja reduzir ou elevar a cada execução.

O último item da lista acima levanta a questão de como estimar a quantidade de *slots* que devem ser adicionados ou subtraídos ao valor de  $AIFS$  para que se provoque um impacto específico de  $c$  kbps na vazão de uma categoria. Neste trabalho, essa questão foi resolvida através de um artifício algébrico que usa a fórmula da estimativa de eficiência ( $\alpha$ ) do controle de admissão. Basta comparar o número de *slots* usados no  $AIFS$  para se obter uma vazão  $X$  kbps com o número de *slots* usados em um  $AIFS$  maior que dá uma vazão  $X - c$ , considerando todos os outros fatores da expressão iguais nos dois casos.

As condições que disparam a redução e o aumento de  $CWMin$  e a redução e o aumento de carga também podem variar, dependendo do objetivo particular de QoS que se quer garantir com o controle.

### 4.3 Considerações Finais

Os mecanismos propostos reúnem idéias úteis para as duas funções em questão, o controle de carga e o controle de admissão. Algumas simplificações foram assumidas, considerando cenários de estudo, de modo que algumas observações ainda podem ser feitas, tendo em vista cenários mais realísticos.

No controle de admissão,  $\alpha$  representa uma estimativa de eficiência. O cálculo dessa eficiência pode ser pouco preciso se as fontes presentes na categoria cuja admissão está sendo controlada forem heterogêneas. Por exemplo, a categoria pode estar sendo utilizada por um serviço de voz, mas que permite o uso de codificadores

diferentes, com tamanhos de pacotes e taxas de geração diferentes. Neste trabalho, assume-se que as fontes em cada categoria sejam do mesmo tipo.

A especificação de tráfego deve informar ao controle de admissão a natureza da aplicação. É recomendável que as aplicações com padrão conversacional utilizem a categoria de acesso mais alta, pois o retardo e sua variação têm limites mais restritos em tais aplicações, para evitar que se perca a interatividade entre os interlocutores. Na classificação da IETF, essas aplicações podem ser consideradas como "de tempo real e intolerantes". As do tipo *streaming* podem utilizar a segunda categoria, abaixo das intolerantes, porém acima das elásticas.

Os algoritmos mostrados na Seção 4.2.5 podem ser generalizados para o caso de mais de duas categorias, fazendo-se o `ajustarContencao()` para cada categoria  $i$  e `ajustarCarga()` para cada categoria menos prioritária que  $i$ , conforme mostrado na Figura 4.14.

```
controlarCarga(n) /* n é o número de categorias */
{
  para cada categoria i, i de 0 até n-2 /* 0 é a mais prioritária, n-1 a de melhor esforço */
    ajustarContencao(i)
    para cada categoria j, j de i+1 até n-1
      ajustarCarga(j)
  emitirConjuntoDeParametrosDeQoS()
}
```

Figura 4.14: Módulo Principal do Controle de Carga para  $n$  Categorias

Deve-se, no entanto, tomar cuidado com as restrições. Por exemplo: toda vez que o `CWMin` de uma categoria  $i$  for aumentado, em qualquer parte do código, os AIFS de todas as categorias menos prioritárias do que  $i$  devem ser também incrementados, para que se mantenha limitada a sobreposição de janelas e se preserve o grau de diferenciação estabelecido entre as categorias.

No capítulo a seguir, os mecanismos propostos são avaliados.

# Capítulo 5

## Simulações e Resultados

OS mecanismos propostos precisam ser avaliados para que se ateste o quanto esses mecanismos se adequam às redes locais sem fio e o quanto cumprem os objetivos pretendidos de controle de Qualidade de Serviço. Como os cenários de sobrecarga da rede podem envolver um grande número de estações móveis, a melhor maneira de realizar essas avaliações é através de simulações por software.

Este capítulo descreve as avaliações desses mecanismos. A Seção 5.1 descreve o ambiente de simulação, relacionando as ferramentas utilizadas, a preparação do ambiente e o cenário montado para avaliar as propostas. A Seção 5.2 apresenta a avaliação da adaptação às redes sem fio do controle de admissão baseado em medidas (MBAC). A Seção 5.3 avalia o controle de carga proposto, submetendo-o a diversas situações de carga e explorando a coexistência de tráfego de tempo real com tráfego de melhor esforço. Por último, a Seção 5.4 descreve as considerações finais do capítulo.

### 5.1 Ambiente de Simulação

A ferramenta de simulação empregada foi o Network Simulator (NS) [53], versão 2.1b8a. O Network Simulator é um simulador orientado a eventos, desenvolvido no Projeto VINT (*Virtual InterNetwork Testbed*) com o suporte da DARPA (*Defense Advanced Research Projects Agency*), e com a colaboração de pesquisadores de cen-

tros de pesquisa e universidades como UCB (*University of California in Berkeley*), LBL (*Lawrence Berkeley National Laboratory*), USC/ISI (*University of Southern California / Information Sciences Institute*), e Xerox PARC (*Palo Alto Research Center*).

O simulador é escrito em C++. Uma outra linguagem orientada a objetos chamada OTcl (*MIT Object Tcl*) [54] é utilizada para estabelecer uma interface de comando e configuração. Essa linguagem auxiliar é interpretada, permitindo que os *scripts* de simulação sejam escritos e modificados sem necessidade de recompilar todo o código do simulador.

O código do MAC 802.11 é originário do Projeto Monarch, desenvolvido na CMU (*Carnegie Mellon University*). Esse projeto incorpora uma extensão às versões mais recentes do NS, introduzindo o suporte às redes sem fio e à mobilidade. O código do MAC contido nessa extensão suporta apenas o método de acesso com contenção (DCF).

### 5.1.1 Preparação do Ambiente de Simulação

Para que as simulações deste trabalho pudessem ser realizadas, foram necessárias modificações na configuração e em algumas partes do código do simulador.

Como a extensão Monarch só continha protocolos de roteamento para redes *Ad Hoc*, foi preciso agregar ao código do NS um protocolo específico, já que nas redes 802.11 infra-estruturadas não existe comunicação direta entre estações móveis de um mesmo BSS. Isso foi conseguido adaptando-se para a versão 2.1b8a um agente de roteamento chamado NOAH (*NO n Ad Hoc routing agent*) [55], desenvolvido por Joerg Widmer para a versão 2.1b7 do NS. Diferentemente dos protocolos de roteamento *Ad Hoc*, o agente NOAH faz com todos os quadros de e para as estações móveis passem pelo sistema de distribuição, através do ponto de acesso.

O código do MAC 802.11 contido na extensão Monarch foi também modificado, a fim de prover as funcionalidades do EDCF. Inicialmente foi desenvolvido um código que permitia a alteração dos parâmetros (como DIFS, CWMin e RTSThreshold)

do MAC para realizar a diferenciação utilizando uma única fila em cada nó. Essa implementação tornou possível o estudo da diferenciação baseada em estação móvel (*station-based*). Esse estudo serviu de base para o desenvolvimento do controle de admissão e do controle de carga propostos nesta dissertação, e pode ser encontrados em [56].

Para o estudo da diferenciação por categorias de acesso (*queue-based*), surgiu a necessidade de implementação do conceito de múltiplas filas do EDCF. A implementação para múltiplas filas foi realizada tendo como base o código do Virtual DCF [52], desenvolvido por Aman Singla e derivado do código original do Monarch. O código do Virtual DCF, composto de aproximadamente duas mil linhas, permite simular o modo PCF, a diferenciação por categorias de acesso (múltiplas filas na mesma estação) e as alterações na característica de persistência do MAC 802.11. Como esse código foi desenvolvido para a versão 2.1b7a do NS, foi preciso compatibilizá-lo com a versão 2.1b8a. Foi necessário ainda adaptá-lo ao agente de roteamento NOAH e a outras necessidades particulares do trabalho em questão.

A implementação da funcionalidade de um monitor de QoS exigiu a monitoração das filas de transmissão sem fio do ponto de acesso e das estações móveis. Na hierarquia de classes do NS já existe uma classe QueueMonitor, que permite a monitoração de uma fila única em enlaces fixos. Baseado nessa classe, foi criada uma nova, chamada de IfQMonitor, para permitir a monitoração das múltiplas filas das estações 802.11e e coletar separadamente as informações de retardo, perdas e bytes transmitidos em cada categoria de acesso.

O controle de admissão e o controle de carga foram inteiramente implementados em OTcl. Os scripts usados para implementar a lógica dos controles, montar os cenários e controlar o andamento das simulações perfazem aproximadamente 850 linhas de código OTcl.

### 5.1.2 Cenário de Simulação

O cenário de simulação consiste em um ponto de acesso conectado a um nó fixo por meio de um enlace de 100 Mbps, com retardo de 2 ms. O ponto de acesso (AP)

está localizado no centro de uma área de 350 x 350 metros, de forma que todos os nós móveis contidos na área estejam a uma distância menor que 250 metros, que é o alcance padrão no NS para nós móveis. O enlace sem fio tem largura de banda de 11 Mbps. A taxa de dados utilizada é de 11 Mbps e a taxa básica é de 1 Mbps.

O cenário simula a situação de um *hotspot* 802.11 fornecendo acesso à rede fixa, por exemplo, uma rede IP ou uma rede celular 2,5 ou 3G com acesso a redes IP. Neste último caso, o 802.11 estaria sendo utilizado na RAN (*Radio Access Network*) da arquitetura 2,5/3G, como uma alternativa de acesso de curto alcance e banda larga. Os nós móveis são posicionados de forma aleatória no interior da topologia de 350 x 350 metros. A Figura 5.1 ilustra essa configuração.

As simulações são feitas sem o uso de RTS/CTS, ou seja, utilizando o esquema de acesso básico do 802.11. O acesso com RTS/CTS não foi usado, uma vez que o benefício decorrente do uso desse esquema não é suficiente para compensar o custo da transmissão dos quadros RTS e CTS no caso de fontes que usam quadros de dados pequenos, como é o caso das fontes de voz empregadas nas simulações descritas a seguir.

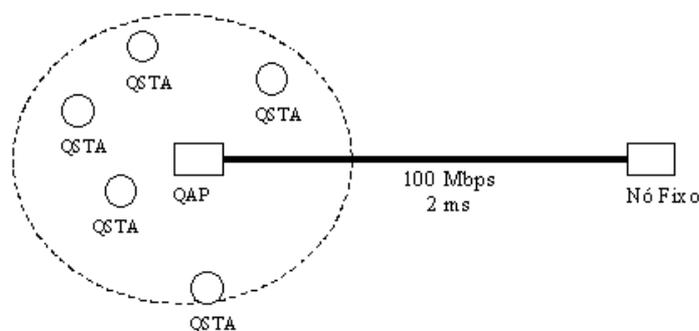


Figura 5.1: Cenário usado nas simulações.

## 5.2 Avaliação do Controle de Admissão

Inicialmente foram feitas simulações para verificar se a adaptação do controle de admissão baseado em medidas (MBAC) está correta e apresenta, para o caso sem fio, um desempenho comparável ao verificado no caso tradicional de redes fixas.

Essas simulações foram realizadas somente com fontes de voz.

### 5.2.1 Modelo de Tráfego

As fontes de voz geram dados a uma taxa de 64 kbps, com pacotes de 210 bytes, e têm períodos de atividade e de silêncio seguindo distribuições exponenciais com médias 1,2 e 1,8 segundos, respectivamente. Essas distribuições seguem o modelo de tráfego de voz criado por Brady [57], utilizado em vários estudos de Voz sobre IP, como por exemplo [58, 59] e de voz em 802.11 [60]. Com essa configuração, o tráfego gerado pelas fontes de voz equivale ao tráfego gerado por um codificador PCM (*Pulse Code Modulation*) com supressão de silêncio.

As fontes de voz têm um intervalo entre chegadas que segue uma distribuição exponencial com média 4,0 segundos. Em outros estudos de controle de admissão baseado em medidas [3, 4, 6], o intervalo entre chegadas segue uma distribuição exponencial com média 400 ms. No entanto, isso leva a uma taxa de chegada de fluxos muita alta e pouco razoável para o caso de uma rede de acesso, sendo mais adequada aos cenários de redes de núcleo explorados naqueles trabalhos. A duração das conexões de voz segue uma distribuição exponencial com média de 300 segundos, semelhante ao empregado em [3, 4, 6].

Para simular o padrão conversacional de cada conexão de voz, foi empregada uma fonte exponencial (*onoff*) na direção do nó móvel para o nó fixo (sentido de subida ou *uplink*) e outra na direção do nó fixo para o móvel (sentido de descida ou *downlink*), que são iniciadas com uma diferença de até 1 segundo. Essa bidirecionalidade, embora pouco explorada em outros estudos, é necessária para tornar a simulação mais próxima da situação real já que, nas redes 802.11, o tráfego do sentido de descida compete pelo meio sem fio com o tráfego do sentido de subida.

O protocolo de transporte utilizado é o UDP, com tamanho de pacote igual ao usado pela fonte, para evitar fragmentação e remontagem na camada de transporte. Por simplicidade, não foi utilizado um protocolo de sinalização de chamadas específico, uma vez que é possível, no *script* de controle de simulação, iniciar a fonte apenas se a conexão for admitida. No entanto, deve ficar claro que tal protocolo é

indispensável em uma implementação real.

### 5.2.2 Configuração do MBAC

No mecanismo de medição foi empregado um intervalo de amostra ( $S$ ) de 0,4 segundos. O tamanho do intervalo de amostra tem relação com o tempo de transmissão de um pacote. Em [4] é utilizado um tamanho de  $S$  de pelo menos 100 vezes esse tempo, mas em [3] são utilizados valores de 500, 1000 e 5000 vezes. O tamanho de janela ( $T$ ) é de 4,0 segundos, para manter a relação  $T/S$  igual a 10, conforme [3]. Nessa referência, afirma-se que valores menores de  $S$  tendem a proporcionar estimativas máximas mais altas dentro de uma janela de medição, fazendo com que o algoritmo de controle de admissão apresente um comportamento mais conservador, isto é, passe a admitir menos fontes. Valores maiores de  $T$  mantêm um histórico de medidas mais longo no tempo, o que também torna o controle mais conservador.

A carga média do tráfego de voz, medida pelo estimador a cada intervalo de amostra, é dada pela soma dos bytes dos pacotes de voz que passaram pela fila de transmissão sem fio (IfQ) do AP no intervalo, dividida pelo tamanho do intervalo (0,4 s).

O fator de utilização ( $\mu$ ) foi fixado em 0,9 [3], o que faz com que o controle de admissão permita que no máximo 90% da capacidade do enlace sem fio seja ocupada pelo tráfego de voz. A estimativa de eficiência do canal ( $\alpha$ ) é calculada em 0,2044, considerando os fatores que afetam a capacidade do canal (tamanho dos quadros, intervalos entre-quadros, taxas básica e de dados, tempo de *backoff*). Multiplicando-se  $\alpha$  pela taxa nominal de 11 Mbps e pelo fator de utilização de 0,9, o MBAC assume uma utilização máxima de 1011,6 Kbps para cada sentido do tráfego de voz.

### 5.2.3 Resultados

A avaliação do controle de admissão foi feita após 10 rodadas do cenário de simulação, utilizando apenas fontes de voz. Cada rodada teve 400 segundos de duração. Os resultados são mostrados na Tabela 5.1.

Nessa Tabela, a coluna "carga média" reflete a carga média do tráfego de voz no sentido de descida, para o período da simulação entre 150 e 400 s, conforme as medidas tomadas pelo estimador na fila de transmissão sem fio do ponto de acesso. A coluna "# fontes" indica o número médio de fontes simultâneas presentes na rede no mesmo período.

As colunas "retardo AP" e "retardo STA" mostram a média das amostras de retardo dos quadros de voz na fila de transmissão sem fio do ponto de acesso e da primeira estação móvel, respectivamente. Cada amostra de retardo corresponde ao intervalo de tempo decorrido desde a entrada do quadro na fila da categoria de acesso até a sua saída. O valor dessas amostras é influenciado por dois fatores: o tamanho da fila e o retardo de acesso que a categoria observa na disputa pelo meio físico. O primeiro é função da carga imposta pela categoria de acesso na estação que a implementa e o segundo é função da carga total presente na rede.

Cada linha numerada de 1 a 10 na Tabela mostra o resultado particular de uma rodada. As duas últimas linhas mostram, respectivamente, a média e um intervalo de confiança de 95% para as 10 simulações.

Tabela 5.1: Teste do Controle de Admissão

rodada	carga média (Kbps)	# fontes	retardo AP (ms)	retardo STA (ms)
1	789,56	22,44	4,76	1,34
2	781,50	22,32	4,06	1,12
3	839,61	24,42	5,13	1,48
4	866,68	24,09	5,49	1,46
5	756,69	21,87	3,91	1,08
6	856,52	24,17	5,31	1,31
7	784,48	23,06	4,27	1,38
8	803,22	23,12	5,04	0,98
9	762,22	22,77	4,10	1,66
10	794,37	22,98	4,28	1,11
média	803,48	23,12	4,64	1,29
IC-95	27,32	0,61	0,41	0,15

A média da carga observada no ponto de acesso (803,48 Kbps), mostrada na Tabela 5.1, é de 79,4% da utilização máxima pretendida (1011,6 Kbps). Este resultado está coerente com o obtido em [3], onde o algoritmo de soma medida possibilitou 79% de utilização com fontes equivalentes às usadas neste trabalho.

O resultado relativo à carga mostra que o controle de admissão teve compor-

tamento conservador: um número de fontes menor que o esperado foi admitido e a carga de voz foi mantida abaixo da utilização máxima. Apesar disso, os valores de retardo observados indicam que o controle contribuiu para manter o retardo de acesso no sentido de descida ("retardo AP") na mesma ordem de grandeza do retardo no sentido de subida ("retardo STA"), o que é importante para as aplicações de voz, que são bidirecionais e tendem a impor à rede uma carga simétrica nos dois sentidos da conversação.

As Figuras 5.2, 5.3 e 5.4 ilustram as medidas realizadas em uma rodada particular de simulação. A Figura 5.2 mostra as estimativas de carga de voz feitas pelo estimador a cada intervalo de amostra ( $S$ ). Observa-se uma grande variação nas amostras, decorrente da natureza em rajada do tráfego que está sendo medido e da dinâmica de chegada e partida das fontes de voz.

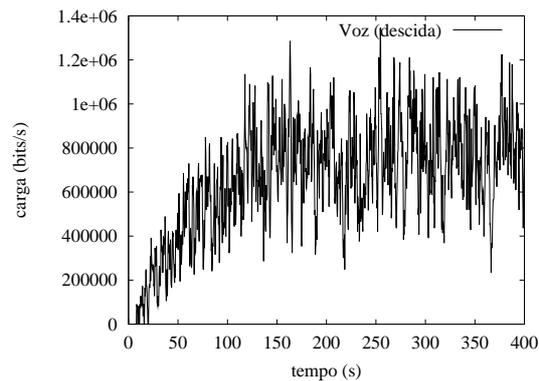


Figura 5.2: Carga do tráfego agregado de voz no sentido de descida, medida pelo estimador no AP em cada intervalo de amostra.

As Figuras 5.3 e 5.4 mostram o valor máximo das amostras de retardo dos pacotes de voz no ponto de acesso e na primeira estação móvel, calculado a cada intervalo de amostra ( $S$ ), para uma rodada particular da simulação. Nessas figuras observa-se que, para a rodada em questão, os valores máximos de retardo estiveram sempre abaixo de 20 ms no ponto de acesso e abaixo de 12 ms na estação móvel. Os valores nulos da Figura 5.4 correspondem aos períodos em que a fonte de voz da primeira estação está em silêncio.

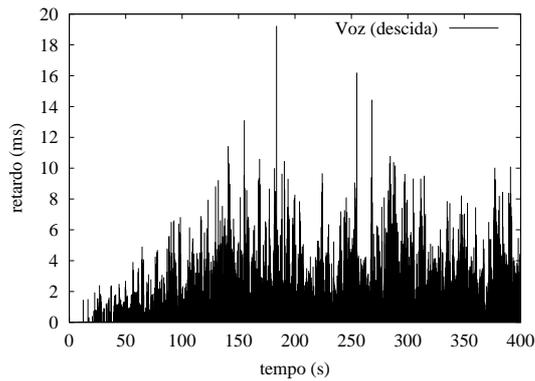


Figura 5.3: Retardo máximo dos pacotes de voz do tráfego de descida, medido no AP a cada intervalo de amostra.

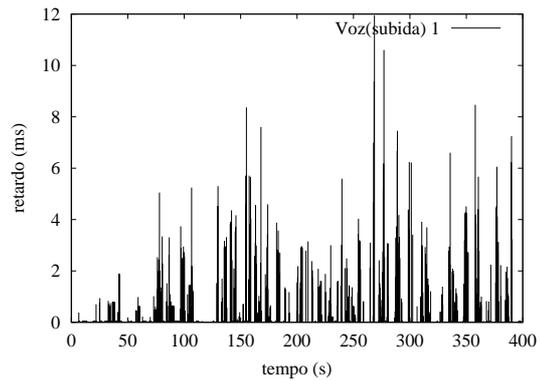


Figura 5.4: Retardo máximo dos pacotes de voz do tráfego de subida, medido na 1ª STA a cada intervalo de amostra.

## 5.3 Avaliação do Controle de Carga

A avaliação do controle de carga explorou a coexistência de tráfego de tempo real com tráfego de melhor esforço no acesso com contenção (EDCF). O objetivo é permitir que o maior número possível de fontes de tempo real cheguem e partam da rede, mantendo o retardo limitado, e ainda permitir ao tráfego de melhor esforço explorar ao máximo a banda passante não utilizada pelo tráfego de tempo real.

### 5.3.1 Modelos de Tráfego

O tráfego de tempo real foi modelado por fontes de voz iguais às utilizadas na avaliação do controle de admissão. A entrada das fontes de voz é controlada pelo mecanismo de controle de admissão, que é executado em conjunto com o controle de carga.

O tráfego de melhor esforço simula conexões TCP de longa duração, empregando fontes FTP que realizam uma transferência de dados contínua, gerando pacotes de 576 bytes<sup>1</sup>. A origem das conexões FTP está no nó fixo e o destino nos nós móveis, representando a situação em que os usuários móveis estão transferindo arquivos da rede fixa para suas estações móveis (*download*).

<sup>1</sup>O tamanho de pacote de 576 bytes é considerado típico de aplicações de navegação na Internet (Web) [29].

O tráfego de voz usa a categoria de mais alta prioridade (AC0) e o tráfego FTP usa a categoria menos prioritária (AC1). AC0 emprega AIFS=2 e CWMin=7 e AC1 inicia com AIFS=9 e CWMin=31.

Para estabelecer uma base de comparação para o desempenho do controle de carga, foram também realizadas simulações sem o uso desse controle, isto é, com definição estática dos parâmetros de QoS tanto para a voz quanto para o FTP.

Em cada simulação são medidos: 1) a vazão obtida pela fonte FTP, com base na contagem dos pacotes de reconhecimento do TCP (ACK da camada de transporte) recebidos pela fonte, medida essa conhecida como *goodput*; 2) o retardo dos pacotes de voz na fila de transmissão da interface sem fio do ponto de acesso e de um dos nós móveis, obtido de forma similar ao que foi feito na avaliação do controle de admissão; 3) a carga agregada do tráfego de voz no sentido de descida, medida no AP pelo estimador do controle de admissão, através de monitoração da fila da categoria de tempo real.

### 5.3.2 Configuração do Controle de Carga

Nas simulações com controle de carga, o intervalo de monitoração é de 3 segundos. Como não havia informação sobre um valor ideal para esse parâmetro, algumas simulações foram feitas empregando valores diferentes para esse intervalo (Seção 5.3.4).

Os retardos de acesso máximo e mínimo para os quadros de voz, usados em `ajustarCarga()` como condição de aumento e redução da carga do FTP, são de 20 ms e 4 ms, respectivamente. Esses valores podem variar de acordo com os requisitos do tráfego de tempo real e com as políticas de QoS estabelecidas para a utilização da rede. A escolha desses limites estabelece um compromisso entre a qualidade do serviço oferecido ao tráfego de tempo real e a vazão que pode ser explorada pelo tráfego de melhor esforço.

No caso das aplicações de voz, deve ser garantido um retardo fim-a-fim de 150 ms (para um serviço de voz com qualidade excelente) a 400 ms (para qualidade

de voz aceitável) [61]. O retardo medido na fila do ponto de acesso será ainda somado ao retardo de processamento na camada de rede dos nós intermediários, ao retardo de propagação nas outras partes (fixas ou sem fio) da rede e ao retardo de processamento dos codificadores e decodificadores de voz, dos sistemas operacionais e dos dispositivos de áudio das estações móveis. Sendo assim, supondo um serviço de qualidade excelente, o retardo na rede de acesso deve contribuir o mínimo possível nessa soma para que se garanta um retardo fim-a-fim de 150 ms.

Neste trabalho, optou-se por fixar um limite de retardo máximo de 20 ms, menor que o intervalo de geração dos pacotes da fonte de voz (26 ms). O valor mínimo, usado para disparar a elevação de carga do FTP, é de 4 ms, coerente com a média de retardo observada na avaliação do controle de admissão, onde somente as fontes de voz estão presentes na rede.

Ainda no procedimento `ajustarCarga()` são utilizados 1 *slot* para *SlotsAumento* e 4 para *SlotsReducao*. Convém que o número de *slots* usados na operação de aumento de carga seja menor que o usado na redução, para evitar uma alternância desnecessária entre essas operações, o que provoca instabilidade no mecanismo.

O parâmetro  $\delta$  usado em `ajustarContenção()` é igual a 0.8. Simulações com outros valores de  $\delta$  foram também realizadas, no intuito de estudar a influência desse parâmetro sobre o desempenho do controle de carga.

### 5.3.3 Resultados

Os resultados obtidos para as situações com e sem controle de carga refletem medidas realizadas em 10 execuções do cenário de simulação. Cada execução dura 600 segundos. A fim de facilitar a observação do efeito da chegada e partida das fontes de voz sobre a vazão obtida pelo tráfego FTP, as fontes de voz só são admitidas até 150 s de simulação e têm duração fixa de 250 s. Assim podem ser observadas 4 fases: uma fase de admissão (Fase 1) entre 0 e 150 s, uma fase estável (Fase 2) entre 150 e 300 s, onde o número máximo de fontes de voz está presente, uma fase de saída das fontes (Fase 3), a partir de 300 s e, por último, uma fase onde somente o tráfego FTP e a primeira fonte de voz estão presentes (Fase 4). Esta primeira fonte

é usada, durante toda a simulação, para monitorar o retardo do tráfego de subida.

A Tabela 5.2 mostra o efeito do controle de carga sobre o retardo dos quadros de voz no sentido de descida. As linhas numeradas de 1 a 10 mostram o resultado de cada rodada. As duas últimas linhas apresentam a média das 10 rodadas e um intervalo de confiança de 95%.

Na coluna "retardo" está a média de todas as amostras de retardo da rodada. As colunas "P-90", "P-95" e "P-99" mostram, respectivamente, os percentis 90, 95 e 99 dessas amostras.

Tabela 5.2: Retardo de Acesso dos Quadros de Voz

	Sem Controle de Carga				Com Controle de Carga			
	retardo(ms)	P-90	P-95	P-99	retardo(ms)	P-90	P-95	P-99
1	110,49	3,71	8,35	46,38	4,37	1,52	2,61	4,99
2	133,86	5,08	20,34	312,65	2,92	1,74	2,93	6,14
3	133,48	3,61	7,21	67,31	4,06	1,83	3,29	8,42
4	37,94	2,70	6,07	38,46	2,63	2,16	3,53	14,42
5	234,65	7,76	36,84	217,16	3,28	1,79	3,09	6,51
6	58,89	3,64	9,45	205,05	6,32	2,12	3,60	9,31
7	43,58	5,49	31,70	168,86	3,47	1,94	3,23	6,72
8	57,64	3,28	9,84	74,74	9,39	1,65	2,75	5,78
9	165,31	3,72	9,30	80,03	2,56	1,62	2,70	5,40
10	80,13	3,62	7,21	25,29	2,98	1,86	2,94	6,08
média	105,60	4,26	14,63	123,59	4,20	1,82	3,07	7,38
IC-95	44,84	1,06	7,97	68,90	1,53	0,15	0,24	2,01

Os dados obtidos revelam que o controle de carga contribuiu para manter o retardo de acesso limitado, abaixo dos 20 ms pretendidos. Tanto o retardo médio quanto o percentil-99 de retardo ficaram abaixo do limite máximo em todas as 10 execuções, o que não ocorreu nas situações sem controle de carga. Nessas situações, a definição estática dos valores de AIFS e CWMin mostrou-se insuficiente para garantir qualquer previsibilidade sobre os limites de retardo do tráfego de tempo real na presença do tráfego de melhor esforço. Caso fossem utilizados valores maiores para AIFS e CWMin para a categoria do FTP (AC1), seria possível obter valores de retardo menores para o tráfego de voz mesmo sem controle de carga, porém a vazão do tráfego FTP seria excessivamente reduzida.

O controle de carga, por outro lado, torna possível manter o retardo limitado

sem sacrificar excessivamente o tráfego de melhor esforço, pois reduz a carga do FTP apenas nos instantes mais críticos para o tráfego de voz e a eleva sempre que possível. A Tabela 5.3 mostra a vazão média agregada do tráfego FTP. A coluna "média" expressa a média das 10 execuções (com e sem controle de carga) e a coluna "IC-95" revela o intervalo de confiança de 95%. A primeira linha mostra as medidas considerando a duração total da simulação e as quatro restantes mostram os resultados para cada uma das fases anteriormente citadas.

Na primeira linha dessa tabela observa-se que o controle provocou uma redução de apenas 8,5% na vazão do tráfego de melhor esforço, considerando todo o período de simulação. Nas medidas das fases 1, 2 e 3 verifica-se que a vazão foi sempre menor no caso com controle de carga, especialmente na fase 2 quando há o maior número de fontes de voz presentes na rede. Em contrapartida, na fase 4 a vazão do tráfego FTP chega a ser maior no caso com controle, porque o procedimento `ajustarCarga()` reconhece que há pouco tráfego de voz e é capaz de selecionar valores menores de AIFS e CWMin até chegar a AIFS=2 e CWMin=31 (valores *default* usados em uma rede 802.11b).

Tabela 5.3: Vazão Média do Tráfego FTP

período	Sem Controle de Carga		Com Controle de Carga	
	média(Kbps)	IC-95	média(Kbps)	IC-95
0 a 600 s	1319,23	24,69	1206,54	47,32
Fase 1	1025,74	36,96	816,75	40,29
Fase 2	625,47	51,40	341,05	46,51
Fase 3	1609,62	36,33	1533,26	125,87
Fase 4	2010,43	11,72	2186,46	33,83

Os dados das Tabelas 5.2 e 5.3 mostram que uma redução de cerca de 95% do retardo do tráfego de tempo real (caindo de uma média de 105,60 ms para 4,20 ms) foi obtida com pouca redução da vazão do tráfego de melhor esforço (8,5%). Esse resultado, bastante significativo, é similar ao atingido por Ahn *et al* com o modelo de rede SWAN [7], onde um retardo 60 a 75% menor para o tráfego de tempo real foi obtido com redução de apenas 2% da vazão do tráfego de melhor esforço. A Figura 5.5 permite comparar a distribuição de retardo entre uma situação sem e outra com controle de carga.

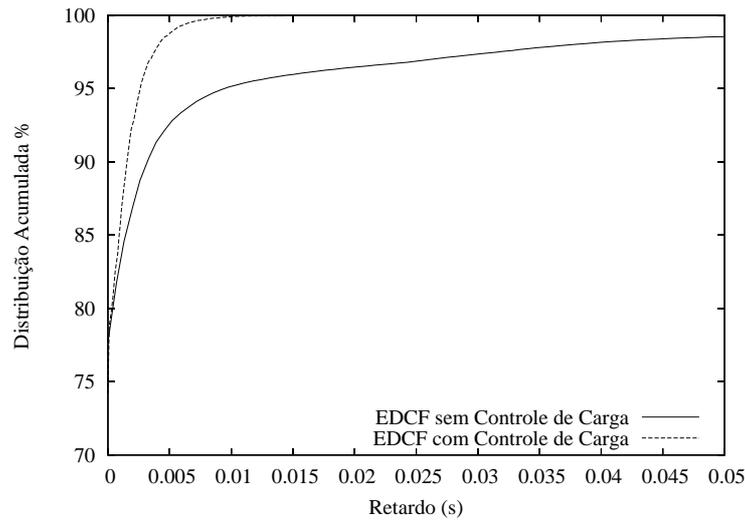


Figura 5.5: Distribuição do Retardo do Tráfego de Tempo Real

Exemplos das medidas de vazão agregada do tráfego FTP para um rodada sem e outra com controle de carga são mostrados nas Figuras 5.6 e 5.7, respectivamente.

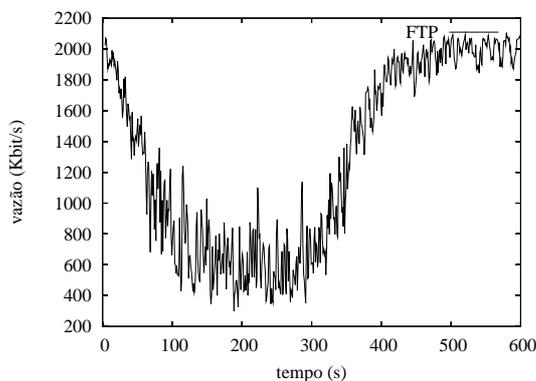


Figura 5.6: Vazão Agregada do Tráfego FTP sem Controle de Carga

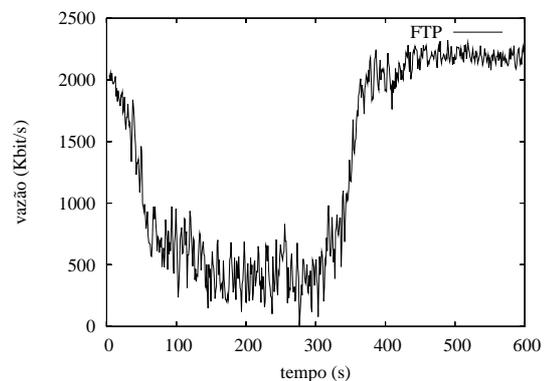


Figura 5.7: Vazão Agregada do Tráfego FTP com Controle de Carga

As Figuras 5.8 e 5.9 mostram o valor máximo do retardo de acesso dos pacotes de voz em duas execuções particulares da simulação, uma sem e outra com controle de carga. Os valores são coletados periodicamente no AP, por monitoração da fila de transmissão da categoria AC0.

Observa-se que, sem controle de carga, o limite de retardo pretendido (20 ms) é violado várias vezes entre 100 e 300 s de simulação, chegando a picos de mais de 600 ms. Essa situação é inaceitável para o tráfego de voz, considerando-se que para

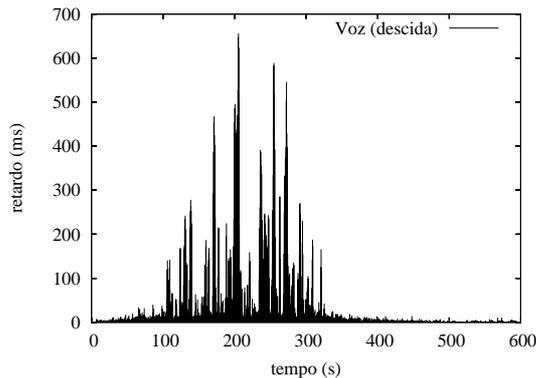


Figura 5.8: Retardo de Acesso do Tráfego de Voz sem Controle de Carga

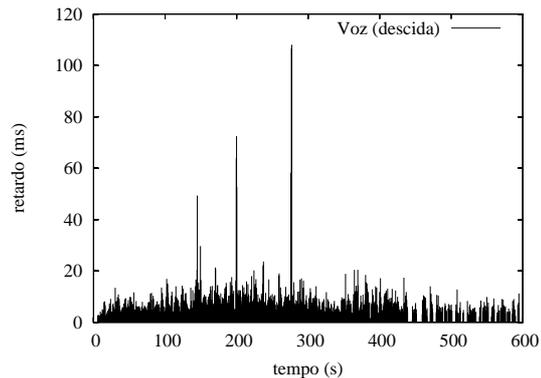


Figura 5.9: Retardo de Acesso do Tráfego de Voz com Controle de Carga

esse tipo de aplicação deve ser garantido um retardo fim-a-fim de 150 ms a 400 ms [61], e que o retardo medido na fila do ponto de acesso é apenas uma das parcelas do retardo fim-a-fim. Se os interlocutores estiverem cada um em uma rede de acesso sem fio com carga elevada, a imprevisibilidade do limite de retardo nas duas redes de acesso trará efeitos ainda mais danosos.

Fica claro que a rede sem fio deve possuir mecanismos para garantir que o retardo dos quadros pertencentes ao tráfego de tempo real seja o menor possível. Com o controle de carga, é possível confinar mais facilmente esse retardo abaixo de um limite máximo.

### 5.3.4 Variações do Controle de Carga

Foram realizadas ainda simulações variando alguns dos parâmetros de configuração do próprio controle de carga, como o intervalo de monitoração e o parâmetro de controle da sobreposição de janelas ( $\delta$ ). Os resultados são comentados a seguir.

#### Intervalo de Monitoração

A avaliação do impacto do intervalo de monitoração sobre a vazão do tráfego FTP e sobre o retardo médio do tráfego de voz foi realizada variando-se esse intervalo de 0,25 s até 3 s, com incrementos de 0,25. Para cada valor de intervalo são feitas 10 execuções distintas, retirando-se a média e um intervalo de confiança de 95% para

as dez execuções. No total, 120 rodadas do cenário foram necessárias. Os resultados são mostrados nas Figuras 5.10 e 5.11.

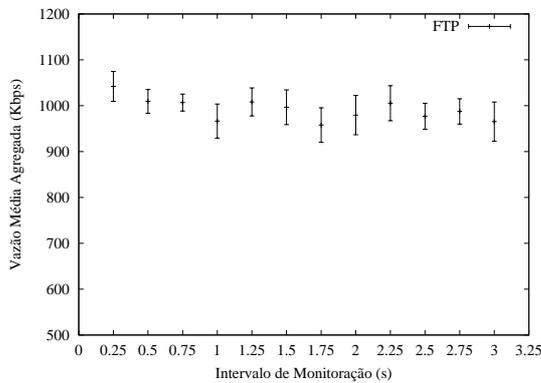


Figura 5.10: Vazão Agregada do Tráfego FTP x Intervalo de Monitoração

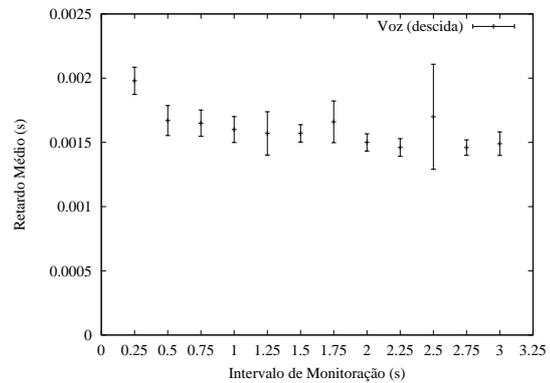


Figura 5.11: Retardo Médio do Tráfego de Voz x Intervalo de Monitoração

Para a situação de coexistência de voz com FTP, ambos com carga elevada, como a explorada nas simulações deste trabalho, os resultados obtidos na avaliação do intervalo de monitoração não permitem afirmar que um intervalo menor que 3,0 s traga algum ganho de vazão do tráfego de melhor esforço sem comprometer o retardo do tráfego de tempo real. Diminuir o intervalo de monitoração faz com que o controle de carga (e conseqüentemente, os procedimentos `reduzirCarga()` e `aumentarCarga()`) seja executado mais vezes ao longo do tempo, aumentando o esforço computacional exercido no ponto de acesso, sem no entanto trazer um benefício que possa ser claramente quantificado. Independentemente de ser executado  $K$  vezes ou  $K \cdot 100$  vezes, o controle de carga só irá agir sobre o tráfego de melhor esforço quando as condições que disparam o aumento ou a redução de carga forem satisfeitas.

Embora valores maiores que 3 s não tenham sido avaliados, é possível estimar um limite superior para o intervalo de monitoração. Tal intervalo deve ser menor que o intervalo médio entre chegadas das fontes de tempo real, para evitar que a situação em que a frequência de chegada de tais fontes é maior que a frequência das operações de controle de carga. Nessa situação, o controle de admissão admite mais fontes de tempo real em um dado intervalo de tempo, sem que o controle de carga tenha tido oportunidade de, no mesmo intervalo, ajustar a contenção da categoria mais prioritária ou de forçar o tráfego de melhor esforço a ceder acesso em proporção suficiente para garantir o limite de retardo desejado para o tráfego de tempo real.

### Sobreposição de Janelas ( $\delta$ )

A avaliação do impacto do controle de sobreposição de janelas proporcionado pelo parâmetro  $\delta$  do procedimento `ajustarContencao()` foi realizada variando-se esse parâmetro de 0,1 até 1,0, com incrementos de 0,1. Um valor de 0,1 indica que apenas 10% da diferença entre os valores antigo e novo do  $CW_{min}$  da categoria mais prioritária será utilizado para aumentar ou diminuir o valor de  $AIFS$  da categoria menos prioritária. Um  $\delta$  igual a 1,0 significa que toda a diferença será utilizada.

Para cada valor de  $\delta$  são feitas 10 execuções distintas, retirando-se a média e um intervalo de confiança de 95% para as dez execuções. No total, 100 rodadas do cenário foram necessárias. Os resultados são mostrados nas Figuras 5.12 e 5.13.

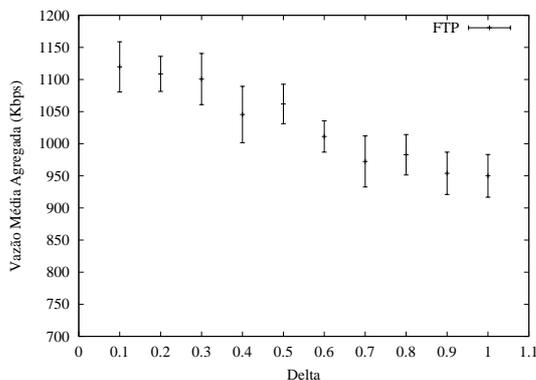


Figura 5.12: Vazão Agregada do Tráfego FTP x Delta

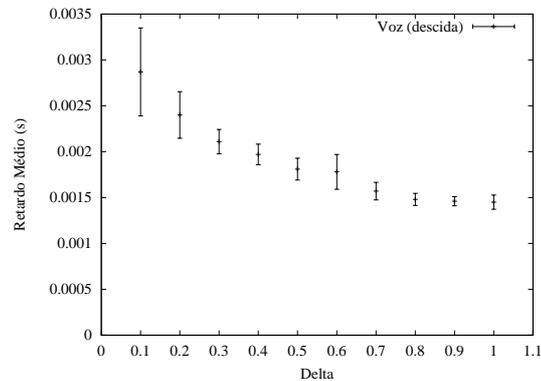


Figura 5.13: Retardo de Acesso do Tráfego de Voz x Delta

Os resultados mostram que um valor menor de  $\delta$  faz com que tanto a vazão do tráfego FTP quanto o retardo do tráfego de voz sejam maiores, o que já era esperado. Quanto menor o  $\delta$ , maior a coincidência das janelas de contenção das duas categorias de acesso e maior a probabilidade de inversão de prioridades entre elas. Uma boa solução para a definição do valor desse parâmetro se evidencia a partir de  $\delta=0,5$ , onde o retardo de voz já está dentro do limite de 20 ms pretendido. Valores maiores que 0,8 também não apresentam ganho significativo na redução de retardo, mas podem ainda reduzir ligeiramente a vazão obtida pelo tráfego FTP.

## 5.4 Considerações Finais

Este capítulo abordou a avaliação dos mecanismos de controle propostos no presente trabalho, mostrando que a adaptação de mecanismos de controle de admissão projetados para redes fixas é viável, desde que sejam consideradas as particularidades das redes locais sem fio, como a variação da capacidade do canal. Os resultados do controle de carga são bastante promissores, revelando que o duplo objetivo de manter o retardo do tráfego de tempo real limitado e permitir que o tráfego de melhor esforço explore o máximo de banda pode ser atingido com uma boa configuração do controle.

Cabem ainda algumas considerações sobre mobilidade. Nas simulações realizadas, considerou-se uma baixa mobilidade no interior do BSS. As estações móveis estão paradas ou em velocidade compatível com a velocidade de um usuário deslocando-se a pé dentro da área coberta pelo ponto de acesso 802.11 (BSA). Essa premissa é coerente com a idéia de um *hotspot*, destinado a prover o acesso em lugares como salas de espera de aeroportos e centros de convenções. Como o escopo do trabalho foi o estudo dos mecanismos de controle em uma única rede sem fio, não está sendo considerada a mobilidade do usuário entre BSSs distintos. Assume-se que as áreas cobertas (BSAs) não têm intersecção, estando distanciadas umas das outras, o que também é coerente com a idéia dos *hotspots*.

Em cenários com múltiplos BSSs onde as BSAs se sobreponham, deve ser considerada a possibilidade de transferência de conexões ativas (*handoff*) entre pontos de acesso. Nesses cenários, espera-se que a mobilidade entre BSSs tenha pouco impacto sobre o mecanismo de controle de carga, pois a saída ou a entrada de uma estação com conexões ativas em um BSS equivalem, para esse mecanismo, aos eventos de partida ou chegada de novas fontes.

Para o controle de admissão, no entanto, a mobilidade do usuário entre os BSSs provocará um impacto que deve ser levado em conta. A conexão ou o fluxo já admitido em um BSS não deve perder continuidade na passagem para outro BSS, de forma que devem ser garantidamente admitidos (ou ter baixa probabilidade de rejeição), o que quer dizer que o gerenciamento de localização das estações deve

distinguir e dar preferência às conexões que estão migrando de um BSS para outro. Isso só é possível mantendo-se uma reserva de recursos em cada BSS, destinada para *handoff*, o que faz com que o controle de admissão não possa utilizar toda a estimativa de capacidade do canal para tomar suas decisões. A quantidade de recursos a ser reservada deve ser proporcional à probabilidade de ocorrência de *handoff* entre cada par de BSS vizinhos, o que implica estudos extensivos dos padrões de mobilidade dos usuários. O fator de utilização ( $\mu$ ) pode ser empregado para limitar ainda mais o percentual do enlace que pode ser utilizado pelo controle de admissão, garantindo essa reserva.

# Capítulo 6

## Conclusões e Trabalhos Futuros

**O**BJETIVO do trabalho descrito nesta dissertação foi o desenvolvimento de mecanismos de controle para complementar a funcionalidade de qualidade de serviço (QoS) introduzida no padrão 802.11 com a extensão 802.11e.

As principais contribuições deste trabalho foram:

- a adaptação para o caso sem fio de um mecanismo de controle de admissão usado nas redes fixas;
- a proposta de um controle de carga para as categorias de acesso do 802.11e, explorando cenários de coexistência de tráfego de tempo real com tráfego de melhor esforço.

Este capítulo descreve as conclusões sobre o desempenho dos mecanismos e sobre a experiência adquirida no decorrer deste trabalho e apresenta sugestões para trabalhos futuros. A Seção 6.1 descreve as observações finais do trabalho. A Seção 6.2 relaciona trabalhos futuros que podem ser iniciados a partir das questões motivadas e das observações extraídas da pesquisa desenvolvida.

## 6.1 Conclusões

A proposta do mecanismo de controle de admissão buscou adaptar para o caso das redes sem fio um esquema de controle de admissão já estudado em redes fixas. A adaptação consistiu em alterar o algoritmo de soma medida, usado para decidir sobre a admissão de novos fluxos, substituindo-se a capacidade do canal ( $C$ ), considerada naquele algoritmo, por uma estimativa de eficiência do canal ( $\alpha$ ), multiplicada pela taxa de transmissão de dados em uso no BSS ( $B$ ). Essa adaptação foi necessária, uma vez que a capacidade da rede 802.11 é variável, sofrendo influência de fatores como o tamanho dos quadros, os intervalos entre-quadros e a sobrecarga imposta pelo método de acesso usado nessas redes.

A avaliação do mecanismo de controle de admissão demonstrou que, respeitadas as peculiaridades das redes locais sem fio, como a variação da capacidade do canal, é possível adaptar para essas redes os esquemas de controle de admissão baseados em medidas empregados nos cenários de redes fixas e ainda obter desses esquemas um desempenho similar ao que apresentam nesses cenários. Apesar de ter sido conservador na utilização de largura de banda, o controle de admissão contribuiu para que fosse preservada uma simetria no tráfego de tempo real, mantendo o retardo do tráfego do sentido de subida na mesma ordem de grandeza do tráfego do sentido de descida, o que é importante para as aplicações com padrão de tráfego conversacional, como as de voz.

O mecanismo de controle de carga proposto teve o duplo objetivo de possibilitar que o tráfego de melhor esforço explorasse o máximo da largura de banda disponível na rede e de evitar que um limite de retardo máximo para o tráfego de tempo real fosse ultrapassado. A intenção desse controle de carga foi regular a carga imposta por todas as fontes de dados associadas a uma classe ou categoria de tráfego, e não por fontes individuais.

A avaliação do mecanismo de controle de carga permitiu constatar que, através da seleção dinâmica dos parâmetros que controlam a diferenciação, como o AIFS e o CWMin, é possível proporcionar uma melhor previsibilidade de retardo e uma melhor utilização do canal, mesmo em cenários que envolvam sobrecarga da rede e

coexistência de tráfego de tempo real com tráfego de melhor esforço. Os algoritmos e princípios elaborados para o controle de carga podem ser estendidos para o suporte a mais de duas categorias de acesso (até oito), permitindo à rede local sem fio abrigar uma maior diversidade de serviços.

Os dois mecanismos de controle apresentam como benefícios a utilização das novas funcionalidades introduzidas pela extensão 802.11e no padrão 802.11 e a restrição ao ponto de acesso de toda a complexidade relacionada às funções de controle de QoS.

A definição dos elementos do conjunto de parâmetros de QoS revelou-se uma tarefa difícil. Encontrar valores ideais para esses parâmetros depende do número de níveis de prioridade que se pretende suportar, das restrições de qualidade que se quer atender, das características das fontes geradoras de tráfego (tamanho e intervalo de geração dos quadros) e das premissas sobre a carga esperada em cada categoria.

Como última observação, e com base na experiência adquirida neste trabalho, é válido afirmar que o *tempo* é o conceito central na alocação dos recursos de transmissão em redes 802.11. Todas as decisões sobre a diferenciação de serviços e o gerenciamento da qualidade de serviço nessas redes, incluindo os controles de admissão e de carga, são influenciadas pela noção temporal (tempo de espera para tentar uma transmissão, tempo consumido na transmissão, tempo consumido em procedimentos de *backoff*, etc). O tempo é a verdadeira quantidade limitada a ser alocada nessas redes (e não a largura de banda, que é variável). Abstraindo-se de todos os detalhes técnicos, pode-se dizer que o papel da função de coordenação, seja ela centralizada, distribuída ou híbrida, é o mesmo papel desempenhado pela função de alocação de canais em sistemas que usam acesso múltiplo por divisão no tempo (*Time-Division Multiple Access - TDMA*). A diferença em relação ao TDMA é que, ao invés de alocar fatias uniformes de tempo para as estações, a fim de satisfazer requisitos bem definidos de vazão e retardo (como no caso dos sistemas de voz que usam TDMA), nas redes locais sem fio são alocadas fatias de tempo de tamanho variável, para satisfazer requisitos diversificados e muitas vezes conflitantes.

## 6.2 Sugestões para Trabalhos Futuros

Como sugestão para trabalhos futuros, recomenda-se o estudo do método de acesso sem contenção da Função de Coordenação Híbrida (*HCF polled-based channel access*), que esteve fora do escopo deste trabalho. O problema de prover regras para que o ponto de acesso possa atribuir satisfatoriamente as oportunidades de transmissão (TxOps) às estações é de natureza similar ao que foi estudado neste trabalho. Outras questões particulares do acesso sem contenção, não relacionadas às questões do acesso com contenção, podem ainda ser investigadas.

De forma idêntica ao que se verificou no EDCF, haverá também uma necessidade de mecanismos de controle de admissão para o acesso livre de contenção. As alternativas básicas são a adaptação de outros mecanismos de controle de admissão usados em redes fixas ou o desenvolvimento de novos mecanismos específicos para o caso sem fio.

O efeito da duração das oportunidades de transmissão no EDCF não foi explorado, uma vez que ainda há indefinição na padronização do 802.11e sobre a fixação de durações das TxOps independentes para cada categoria de acesso ou de uma única duração, válida para todas as categorias. Como a duração das TxOps regula o tempo que uma estação mantém o controle sobre o meio, é de se esperar que esse parâmetro tenha um efeito similar ao provocado pela técnica de variação do tamanho dos quadros. Há, no entanto, uma diferença básica que estabelece uma vantagem do conceito de TxOp sobre aquela técnica: a aplicação não fica restrita à necessidade de enviar quadros grandes, que são mais sujeitos a erros e são pouco adequados a aplicações de voz, por exemplo. A possibilidade de enviar uma rajada de quadros menores durante uma TxOp, a exemplo do que é feito na fragmentação, libera a aplicação dessa restrição.

O controle de carga pode ser beneficiado pela utilização de técnicas e princípios de Inteligência Computacional como a Lógica Difusa ou Nebulosa (*Fuzzy Logic*). A Lógica Difusa é bastante adequada como técnica de implementação de mecanismos de controle em sistemas complexos. Com o uso dessa técnica, noções intuitivas como "carga alta", "carga baixa", "retardo alto", "retardo baixo" podem ser incorporadas

às regras de controle de uma forma menos dependente de faixas de valores ou limiares rígidos, muitas vezes de difícil definição, permitindo assim maior flexibilidade e adaptabilidade desse controle às condições de carga do sistema. Essa implementação, no entanto, deve ser acompanhada de um cuidadoso estudo de complexidade de processamento, para evitar que o ponto de acesso apresente um atraso excessivo na definição dos elementos do conjunto de parâmetros de QoS. A base de regras a ser utilizada pelo controle difuso deve ser exaustivamente refinada, uma vez que podem ser geradas regras inadequadas que, embora pareçam corretas, conflitam com outras regras presentes na própria base, não contribuam para a eficiência do sistema ou até mesmo a comprometam.

A interoperabilidade com redes que utilizam outras arquiteturas de QoS também é importante, tendo em vista que a rede sem fio, no contexto de acesso móvel à Internet, estará sempre conectada a redes fixas que possuem suas próprias arquiteturas de QoS, como as redes baseadas em IP ou ATM ou ainda as redes de sistemas celulares 2,5 / 3G. O controle de admissão, por exemplo, deve levar em consideração as particularidades da arquitetura da rede fixa à qual está conectada a rede móvel. No caso da integração das WLANs com os sistemas celulares, existem preocupações adicionais. Nesses sistemas, dado o histórico dos serviços de voz, os usuários estão mais habituados a parâmetros de desempenho bem definidos (como taxa de perda e retardo máximos) do que a prioridades relativas. Os mecanismos de controle de QoS devem, nesse caso, focalizar as garantias absolutas para esses parâmetros, ainda que isso acarrete alguma redução na eficiência de utilização do canal.

Outras questões situadas fora da camada de enlace de dados também são relevantes, como as regras para o estabelecimento das políticas de uso da rede e para a definição de acordos de nível de serviço compatíveis com os níveis de prioridade suportados, bem como a tradução dessas políticas para os parâmetros empregados nos mecanismos de controle de QoS.

# Referências Bibliográficas

- [1] INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Standard 802.11, 1999.
- [2] INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. Medium access control (MAC) enhancements for quality of service (QoS). IEEE Draft for Standard 802.11e/D3.2, março de 2002.
- [3] JAMIN, S., SHENKER, S., E DANZIG, P. Comparison of measurement-based admission control algorithms for controlled-load service. *IEEE Joint Conference of Computer and Communications Societies (INFOCOM'97)* (Kobe, Japan, abril de 1997).
- [4] JAMIN, S., DANZIG, P., SHENKER, S., E ZHANG, L. A measurement-based admission control algorithm for integrated services packet networks. *IEEE/ACM Transactions on Networking* (fevereiro de 1997).
- [5] GROSSGLAUSER, M., E TSE, D. A framework for robust measurement-based admission control. *IEEE/ACM Transactions on Networking* 7, 3 (junho de 1999).
- [6] BRESLAU, L., JAMIN, S., E SHENKER, S. Performance of measurement-based admission control algorithms. *IEEE Joint Conference of Computer and Communications Societies (INFOCOM'2000)* (Tel Aviv, Israel, março de 2000).
- [7] AHN, G.-S., CAMPBELL, A. T., VERES, A., E SUN, L.-H. Service differentiation in stateless wireless ad hoc networks. *IEEE Joint Conference of Computer*

- and Communications Societies (INFOCOM'2002)* (New York, USA, junho de 2002).
- [8] ROMDHANI, L., NI, Q., E TURLETTI, T. Enhanced service differentiation for IEEE 802.11 wireless ad-hoc networks. *Wireless Communications and Network Conference (WCNC'2003)* (New Orleans, USA, março de 2003).
- [9] INTERNET ENGINEERING TASK FORCE. IETF mobile ad-hoc networks (MANet) working group. Página da World Wide Web. url: <http://www.ietf.org/html.charters/manet-charter.html>, novembro de 2002.
- [10] DEFENSE ADVANCED RESEARCH PROJECTS AGENCY. DARPA internet program protocol specification. Request For Comments - RFC 791, setembro de 1981.
- [11] PERKINS, C. IP mobility support. Request for Comments - RFC 2002, outubro de 1996.
- [12] PAHLAVAN, K., KRISHNAMURTHY, P., E HATAMI, A. Handoff in hybrid mobile data networks. *IEEE Personal Communications* (abril de 2000), 34–46.
- [13] FORMAN, G. H., E ZAHORJAN, J. The challenges of mobile computing. *IEEE Computer* (março de 1994).
- [14] INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. IEEE 802.11 wireless local area networks. Página da World Wide Web. url: <http://www.ieee802.org/11/>, novembro de 2002.
- [15] INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band. IEEE Standard 802.11b, 1999.
- [16] INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. IEEE 802.15 working group for wireless personal area networks. Página da World Wide Web. url: <http://www.ieee802.org/15/>, novembro de 2002.

- [17] Bluetooth Special Interest Group. Página da World Wide Web. url: <http://www.bluetooth.com>, novembro de 2002.
- [18] INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. The IEEE 802.16 working group on broadband wireless access standards. Página da World Wide Web. url: <http://www.ieee802.org/16/>, novembro de 2002.
- [19] European Telecommunications Standard Institute. Página da World Wide Web. url: <http://www.etsi.org/>, novembro de 2002.
- [20] The ATM Forum. Página da World Wide Web. url: <http://www.atmforum.org/>, novembro de 2002.
- [21] INTERNATIONAL TELECOMMUNICATION UNION. Itu-t special study group imt-2000 and beyond. Página da World Wide Web. url: <http://www.itu.int/ITU-T/studygroups/ssg/index.asp>, novembro de 2002.
- [22] EUROPEAN TELECOMMUNICATIONS STANDARDS INSTITUTE. Universal mobile telecommunications system (UMTS); general UMTS architecture. ETSI Technical Specification 123.101 v. 4.0.0 release 4, abril de 2001.
- [23] BETTSTETTER, C., VOGEL, H., E EBERSPACHER, J. GSM phase 2+ general packet radio service GPRS: Architecture, protocols and air interface. *IEEE Communications Surveys* 2, 3 (1999), 2–14.
- [24] SCHRAMM, P., JAVERBRING, S., EDVARDSSON, N., HOOK, M., EDHOLM, C., ANDREASSON, H., MULLER, F., E SKOLD, J. Radio interference performance of EDGE, a proposal for enhanced data rates in existing digital cellular systems. *IEEE Vehicular Technology Conference (VTC'1998)* (Ottawa, Canada, maio de 1998).
- [25] International Organization for Standardization. Página da World Wide Web. url: <http://www.iso.org/>, novembro de 2002.
- [26] CROW, B. P., WIDJAJA, I., KIM, J. G., E SAKAI, P. T. IEEE 802.11 wireless local area networks. *IEEE Communications Magazine* 35, 9 (setembro de 1997), 116–126.

- [27] INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS. Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-speed Physical Layer in the 5 GHz Band. IEEE Standard 802.11a, 1999.
- [28] SCHILLER, J. *Mobile Communications*, 1<sup>st</sup> ed. Addison-Wesley, 2000. ISBN 0201398362.
- [29] LEUNG, K., MCNAIR, B., CIMINI, L., E WINTERS, J. Outdoor IEEE 802.11 cellular networks: MAC protocol design and performance. *IEEE International Conference on Communications (ICC'2002)* (New York, USA, abril de 2002).
- [30] CLARK, M. V., LEUNG, K. K., MCNAIR, B., E KOSTIC, Z. Outdoor IEEE 802.11 cellular networks: Radio link performance. *IEEE International Conference on Communications (ICC'2002)* (New York, USA, abril de 2002).
- [31] ALA-LAURILA, J., MIKKONEN, J., E RINNEMAA, J. Wireless lan access network architecture for mobile operators. *IEEE Communications Magazine* (novembro de 2001), 82–88.
- [32] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. Quality of service basic framework - outline. ISO/ IEC JTC1/ SC21/ WG1 N1145, 1994.
- [33] NAHRSTEDT, K., E STEINMETZ, R. Resource management in networked multimedia systems. *IEEE Computer* (maio de 1995), 52–63.
- [34] KESHAV, S. *An Engineering Approach to Computer Networking: ATM Networks, the Internet, and the Telephone Network*. Addison-Wesley, 1997. ISBN 0201634422.
- [35] CHALMERS, D., E SLOMAN, M. A survey of quality of service in mobile computing environments. *IEEE Communications Surveys* 2, 2 (1999), 2–10.
- [36] BRADEN, R., CLARK, D., E SHENKER, S. Integrated services in the internet architecture: an overview. Request for Comments - RFC 1633, junho de 1994.
- [37] BLAKE, S., BLACK, D., CARLSON, M., DAVIES, E., WANG, Z., E WEISS, W. An architecture for differentiated services. Request for Comments - RFC 2475, dezembro de 1998.

- [38] NICHOLS, K., BLAKE, S., BAKER, F., E BLACK, D. Definition of the differentiated services field (ds field) in the ipv4 and ipv6 headers. Request for Comments - RFC 2474, dezembro de 1998.
- [39] BERNET, Y., FORD, P., YAVATKAR, R., BAKER, F., ZHANG, L., SPEER, M., BRADEN, R., DAVIE, B., WROCLAWSKI, J., E FELSTAIN, E. A framework for integrated services operation over diffserv networks. Request for Comments - RFC 2998, novembro de 2000.
- [40] TANG, K., E GERLA, M. Fair sharing of MAC under TCP in wireless ad hoc networks. *IEEE Multiaccess, Mobility and Teletraffic for Personal Communications (MMT'99)* (Veneza, Itália, outubro de 1999).
- [41] BENSOU, B., WANG, Y., E KO, C. C. Fair medium access in 802.11 based wireless ad-hoc networks. *IEEE/ACM The First Annual Workshop on Mobile Ad Hoc Networking & Computing (MobiHoc'00)* (Boston, EUA, agosto de 2000).
- [42] VAIDYA, N. H., BAHL, P., E GUPTA, S. Distributed fair scheduling in a wireless LAN. *ACM The Sixth Annual Conference on Mobile Computing and Networking (MobiCom'00)* (Boston, EUA, agosto de 2000).
- [43] XU, S., E SAADAWI, T. Does the IEEE 802.11 MAC protocol work well in multihop wireless ad hoc networks? *IEEE Communications Magazine* 39, 6 (junho de 2001), 130–137.
- [44] AAD, I., E CASTELLUCCIA, C. Differentiation mechanisms for IEEE 802.11. *IEEE Joint Conference of Computer and Communications Societies (INFOCOM'2001)* (Anchorage, Alasca, abril de 2001).
- [45] KANODIA, V., LI, C., SABHARWAL, A., SADEGHI, B., E KNIGHTLY, E. Distributed multi-hop scheduling and medium access with delay and throughput constraints. *ACM The Seventh Annual Conference on Mobile Computing and Networking (MobiCom'01)* (Roma, Itália, julho de 2001).

- [46] MANGOLD, S., CHOI, S., MAY, P., KLEIN, O., HIERTZ, G., E STIBOR, L. IEEE 802.11e wireless LAN for quality of service. *European Wireless (EW'2002)* (Florence, Italy, fevereiro de 2001).
- [47] BIANCHI, G. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications - Wireless Series* 18, 3 (2000).
- [48] CALI, F., CONTI, M., E GREGORI, E. IEEE 802.11 wireless lan: Capacity, analysis and protocol enhancement. *IEEE Joint Conference of Computer and Communications Societies (INFOCOM'1998)* (San Francisco, USA, março de 1998).
- [49] CALI, F., CONTI, M., E GREGORI, E. Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit. *IEEE/ACM Transactions on Networking* 8, 6 (dezembro de 2000), 785–799.
- [50] HO, J., SCHRUM, S., TURKI, K., SHAVER, D. P., E SHOEMAKE, M. B. P-DCF for prioritized MAC service - IEEE document 802.11-00/367. Relatório técnico, IEEE 802.11 TGe, novembro de 2000.
- [51] CHESSON, G., DIEPSTRATEN, W., KITCHIN, D., TEUNISSEN, H., E WENTINK, M. Baseline D-QoS proposal - IEEE document 802.11-00/399. Relatório técnico, IEEE 802.11 TGe, 2000.
- [52] CHESSON, G., DIEPSTRATEN, W., HOEBEN, M., SINGLA, A., TEUNISSEN, H., E WENTINK, M. VDCF presentation - IEEE document 802.11-01/132. Relatório técnico, IEEE 802.11 TGe, março de 2001.
- [53] FALL, K., E VARADHAN, K. The NS Manual. Relatório técnico, The VINT Project, janeiro de 2002.
- [54] MASSACHUSETTS INSTITUTE OF TECHNOLOGY. Otcl - object tcl: The faq manual. Página da World Wide Web. url: <ftp://ftp.tns.lcs.mit.edu/pub/otcl/README.html>, dezembro de 2002.

- [55] WIDMER, J. Extensions to the ns network simulator. Página da World Wide Web. url: <http://www.icsi.berkeley.edu/widmer/mnav/ns-extension/>, janeiro de 2002.
- [56] MELO FILHO, J. C., REZENDE, J. F., PIRMEZ, L., E CARMO, L. F. R. C. Controle de admissão e controle de carga para redes IEEE 802.11 infra-estruturadas. *Workshop de Comunicação Sem Fio e Computação Móvel - WCSF'2002* (São Paulo, Brasil, outubro de 2002).
- [57] BRADY, P. A technique for investigating on-off patterns of speech. *Bell Systems Technical Journal* 44 (1969), 1–22.
- [58] MARKOPOULOU, A. P., TOBAGI, F. A., E KARAM, M. J. Assesment of VoIP quality of internet backbones. *International Conference on Computer Communications and Networks (ICCCN'2000)* (Las Vegas, USA, outubro de 2000).
- [59] JIANG, W., E SHULZRINNE, H. Analysis of on-off patterns in VoIP and their effect on voice traffic aggregation. *International Conference on Computer Communications and Networks (ICCCN'2000)* (Las Vegas, USA, outubro de 2000).
- [60] LIU, H.-H., E WU, J.-L. C. A scheme for supporting voice over IEEE 802.11 wireless local area network. *National Science Council* 25 (2001), 259–268.
- [61] INTERNATIONAL TELECOMMUNICATION UNION - TELECOMMUNICATION STANDARDIZATION SECTOR. One way transmission time. ITU-T Recommendation G.114., maio de 2000.