

UNIVERSIDADE FEDERAL DO CEARÁ
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

Controle de Congestionamento para Voz sobre IP em HSDPA

Autor

André Ribeiro Braga

Orientador

Prof. Dr. Francisco Rodrigo Porto Cavalcanti

*Dissertação apresentada à Coordenação
do Programa de Pós-graduação em
Engenharia de Teleinformática da
Universidade Federal do Ceará como
parte dos requisitos para obtenção do
grau de **Mestre em Engenharia de
Teleinformática.***

FORTALEZA – CEARÁ

MAIO 2006

Resumo

O crescimento do número dos usuários do serviço de *Voice over IP* (VoIP) faz dele o serviço com o maior interesse de ser provido por operadoras de telefonia celular. Por outro lado, este demanda um controle de *Quality of Service* (QoS) bastante rígido, o que torna-se ainda mais complicado em redes sem-fio, porque, além de congestionamentos na rede, os pacotes podem ser perdidos devido à erros nas transmissões no enlace de rádio. Dentro deste paradigma, estratégias de controle de congestionamento aparecem como uma boa solução para lidar com as garantias de QoS em situações de sobrecarga do sistema, onde os recursos se encontram exauridos e os requerimentos de qualidade se encontram ameaçados.

Este trabalho consiste na avaliação de algoritmos de controle de congestionamento objetivando um aumento da capacidade e das garantias de QoS para serviços de voz. Os algoritmos avaliados neste trabalho são o escalonamento de pacotes e o controle de admissão.

A análise em cenários de serviços mistos composto por usuários VoIP e *Web* também está contida neste trabalho. O maior foco está no controle do atraso de pacote, já que este é um requerimento crucial para serviços de tempo-real, como o VoIP.

Os resultados mostram que um arcabouço de controle de congestionamento projetado para este serviço é capaz de melhorar o desempenho do sistema e mitigar os efeitos de congestionamento da rede. No cenário de serviços mistos, os algoritmos são capazes de efetuar reserva de recursos dependendo da prioridade definida para cada serviço, levando a um aumento na qualidade percebida pelo serviço mais sensível através de uma leve degradação no serviço mais robusto.

Abstract

The growth in the number of *Voice over IP* (VoIP) users on the internet makes it the service with the highest interest to be provided by cellular operators. On the other hand, it demands very strict *Quality of Service* (QoS) control, which becomes even more complicated in wireless networks, because packets can be lost due to radio link transmission errors, as well as network congestion. Within this paradigm, congestion control strategies appear as a good solution to cope with QoS guarantees under high loads, where the resources are exhausted and the service quality is threatened.

This work comprises the evaluation of congestion control algorithms aiming to improve system capacity and QoS guarantees for speech users. The evaluated algorithms within this work are packet scheduling and admission control.

The analysis in mixed services scenarios composed of VoIP and Web users is also provided in this work. The main focus of the framework is to control the packet delay, since it is a crucial requirement for real-time services.

The results show that a suitable congestion control framework is able to provide performance improvements and mitigation of the effects from overloaded conditions. In the mixed services scenario, the algorithms are capable to perform resource reservation depending on the priority defined to each service, leading to an increase in the quality of the more sensitive service by degrading the more robust service.

Agradecimentos

Ao prof. Dr. Francisco Rodrigo Porto Cavancanti por me conceder a oportunidade e a orientação no decorrer deste trabalho.

A todos os colegas do Grupo de Pesquisas em Telecomunicações Sem-Fio (GTEL) que contribuíram com comentários, discussões técnicas e sugestões, além do companheirismo e um agradável convívio.

Aos amigos da *Ericsson Research* em Luleå, em especial a Stefan Wänstedt, Mårten Ericson e Mats Nordberg, quem me acolheram de forma bastante calorosa durante minha estada, da mesma maneira que me proporcionaram excelentes condições de trabalho e valiosas contribuições técnicas.

A meus pais e irmãos, pelo incentivo e apoio incondicional.

À Camila por todo o amor, incentivo e compreensão.

À *Ericsson do Brasil* pelo suporte financeiro.

A todos aqueles que, direta ou indiretamente, colaboraram para a concretização deste trabalho e, no entanto, não foram citados.

Sumário

Acrônimos	xii
Lista de Figuras	xv
Lista de Tabelas	xvii
1 Introdução	1
1.1 Motivação e Objetivos	1
1.2 Metodologia	3
1.3 Contexto de Desenvolvimento	3
1.4 Contribuições e Produção Científica	3
1.5 Estrutura do Trabalho	5
2 Sistema UMTS/HSDPA e Serviço VoIP	7
2.1 <i>Universal Mobile Telecommunications System</i> (UMTS)	7
2.1.1 <i>High Speed Downlink Packet Access</i> (HSDPA): Evolução do Padrão	11
2.2 Voz sobre IP	13
2.2.1 Qualidade para VoIP	14
2.3 Controle de Congestionamento	15
2.4 Medidas de Desempenho	16
3 Escalonamento de Pacotes para VoIP em HSDPA	19
3.1 Algoritmos de Escalonamento	20
3.1.1 Algoritmos de QoS-não-diferenciado	21
3.1.2 Algoritmos de QoS-diferenciado	21
3.2 Modelos de Simulação	24
3.2.1 <i>Grid</i> Celular	24
3.2.2 Propagação	24
3.2.3 Mobilidade	25
3.2.4 Tráfego	26
3.2.5 Camadas RLC/MAC	27
3.2.6 Tempo de simulação	28

3.3	Resultados	28
3.3.1	Cenário de serviço simples (VoIP)	28
3.3.2	Cenário misto de serviços (VoIP e <i>Web</i>)	38
3.3.3	Regiões de capacidade conjunta	46
3.4	Sumário	48
4	Controle de Admissão para VoIP em HSDPA	53
4.1	Algoritmos de Controle de Admissão	54
4.1.1	Controle de Admissão de Enlace	54
4.1.2	Controle de Admissão de Sessão	55
4.2	Modelos de Simulação	59
4.2.1	<i>Grid</i> Celular	59
4.2.2	Propagação	60
4.2.3	Mobilidade	60
4.2.4	Tráfego	61
4.2.5	Camadas RLC/MAC	62
4.2.6	Tempo de simulação	63
4.3	Resultados	63
4.3.1	Cenário de serviço simples (VoIP)	64
4.3.2	Cenário misto de serviços (VoIP e <i>Web</i>)	77
4.4	Sumário	85
5	Conclusões e Perspectivas	89
5.1	Conclusões	89
5.2	Perspectivas	90
	Referências Bibliográficas	98

Acrônimos

- 16QAM** *16-Quadrature Amplitude Modulation*
- 3GPP** *3rd. Generation Partnership Project*
- 3G** *Terceira Geração*
- A-DPCH** *Associated Dedicated Physical Channel*
- AD** *Attack-Decay*
- ADS** *Asymptotic Delay Scheduler*
- AM** *Acknowledged Mode*
- AMR** *Adaptive Multirate*
- ARMA** *Autoregressive Moving Average*
- ARQ** *Automatic Repeat Request*
- AS** *Active Set*
- FB** *Função Barreira*
- BLER** *Block Error Rate*
- CAC** *Controle de Admissão de Chamadas*
- CC** *Controle de Congestionamento*
- CDF** *Cummulative Distribution Function*
- CIR** *Carrier-to-Interference Ratio*
- CN** *Core Network*
- CODEC** *Codificador-Decodificador*

CPICH *Common Pilot Channel*

CQI *Channel Quality Indicator*

DCH *Dedicated Channel*

DS-CDMA *Direct Sequence Code Division Multiple Access*

DTX *Discontinuous Transmission*

FER *Frame Erasure Rate*

GSM *Global System for Mobile communication*

GTEL *Grupo de Pesquisas em Telecomunicações Sem-Fio*

HARQ *Hybrid Automatic Repeat Request*

HSDPA *High Speed Downlink Packet Access*

HS-DSCH *High Speed Downlink Shared Channel*

HS-SCCH *High Speed Shared Control Channel*

ILPC *Inner-loop Power Control*

IP *Internet Protocol*

IPv4 *Internet Protocol Version 4*

L2 *Camada de Enlace*

LDS *Linear Delay Scheduler*

MAC *Medium Access Control*

MRC *Maximal Ratio Combining*

NBAP *Node B Application Part*

OLPC *Outer-loop Power Control*

PDU *Protocol Data Unit*

PF *Proportional Fair*

PSTN *Public Switched Telephone Network*

QBMC *Queue Based Max CIR*

QoS *Quality of Service*

QPSK *Quadrature Phase Shift Key*

R99 *Release '99*

R4 *Release 4*

RLC *Radio Link Control*

RNC *Radio Network Controller*

ROHC *Robust Header Compression*

RR *Round Robin*

RRM *Radio Resource Management*

RTP *Real-Time Transport Protocol*

SAW *Stop-And-Wait*

SDU *Service Data Unit*

SHO *Soft Handover*

TCP *Transport Control Protocol*

TM *Transparent Mode*

TTI *Transmission Time Interval*

UDP *User Datagram Protocol*

UE *User Equipment*

UM *Unacknowledged Mode*

UMTS *Universal Mobile Telecommunications System*

UTRAN *UMTS Terrestrial Radio Access Network*

VoIP *Voice over IP*

WCDMA *Wideband Code Division Multiple Access*

WPF *Weighted Proportional Fair*

Lista de Figuras

2.1	Arquitetura UTRAN.	9
3.1	Funcionamento do esquema de escalonamento.	20
3.2	Comportamento da função barreira.	23
3.3	Usuários VoIP satisfeitos com um requerimento de atraso de 150 ms.	29
3.4	Usuários VoIP satisfeitos com um requerimento de atraso de 130 ms.	30
3.5	Usuários VoIP satisfeitos com um requerimento de atraso de 110 ms.	31
3.6	Taxa de perda de pacotes devido a erros no canal para um requerimento de atraso de 150 ms.	32
3.7	Taxa de perda de pacotes devido a erros no canal para um requerimento de atraso de 130 ms.	33
3.8	Taxa de perda de pacotes devido a erros no canal para um requerimento de atraso de 110 ms.	34
3.9	Taxa de descarte de pacotes na MAC-hs para um requerimento de atraso de 150 ms.	34
3.10	Taxa de descarte de pacotes na MAC-hs para um requerimento de atraso de 130 ms.	35
3.11	Taxa de descarte de pacotes na MAC-hs para um requerimento de atraso de 110 ms.	36
3.12	Taxa de descarte de pacotes no <i>buffer</i> de <i>play-out</i> para um requerimento de atraso de 150 ms.	36
3.13	Taxa de descarte de pacotes no <i>buffer</i> de <i>play-out</i> para um requerimento de atraso de 130 ms.	37
3.14	Taxa de descarte de pacotes no <i>buffer</i> de <i>play-out</i> para um requerimento de atraso de 110 ms.	38
3.15	Percentual de usuários satisfeitos considerando um requerimento de atraso de 150 ms.	39
3.16	Percentual de usuários satisfeitos considerando um requerimento de atraso de 130 ms.	40
3.17	Percentual de usuários satisfeitos considerando um requerimento de atraso de 110 ms.	41
3.18	Percentual de usuários satisfeitos considerando um requerimento de atraso de 150 ms.	42

3.19	Percentual de usuários satisfeitos considerando um requerimento de atraso de 130 ms.	43
3.20	Percentual de usuários satisfeitos considerando um requerimento de atraso de 110 ms.	43
3.21	Percentual de usuários satisfeitos considerando um requerimento de atraso de 150 ms.	44
3.22	Percentual de usuários satisfeitos considerando um requerimento de atraso de 130 ms.	45
3.23	Percentual de usuários satisfeitos considerando um requerimento de atraso de 110 ms.	45
3.24	Regiões de capacidade considerando um requerimento de atraso para VoIP de 150 ms.	46
3.25	Regiões de capacidade considerando um requerimento de atraso para VoIP de 130 ms.	47
3.26	Regiões de capacidade considerando um requerimento de atraso para VoIP de 110 ms.	48
3.27	Capacidade absoluta considerando um requerimento de atraso para VoIP de 150 ms.	49
3.28	Capacidade absoluta considerando um requerimento de atraso para VoIP de 130 ms.	50
3.29	Capacidade absoluta considerando um requerimento de atraso para VoIP de 110 ms.	50
4.1	Ilustração do funcionamento do controle de admissão.	54
4.2	Esquema de CAC proposto.	56
4.3	Média e 98 ^o percentil do atraso medido e estimado.	58
4.4	Desempenho dos escalonadores para tráfego VoIP.	65
4.5	98 ^o percentil da taxa de pacotes perdidos.	66
4.6	98 ^o percentil da taxa de descarte de pacotes.	67
4.7	98 ^o percentil e média da potência non-HS.	68
4.8	Satisfação de usuários utilizando RR.	69
4.9	Satisfação de usuários conectados utilizando RR.	70
4.10	Taxa de bloqueio utilizando RR.	70
4.11	98 ^o percentil da taxa de pacotes perdidos utilizando RR.	71
4.12	98 ^o percentil da taxa de descarte de pacotes utilizando RR.	72
4.13	Satisfação de usuários utilizando PF.	72
4.14	Satisfação de usuários conectados utilizando PF.	73
4.15	Taxa de bloqueio utilizando PF.	73
4.16	98 ^o percentil da taxa de perda de pacotes utilizando PF.	74
4.17	98 ^o percentil da taxa de descarte de pacotes utilizando PF.	75
4.18	Satisfação de usuários utilizando ADS.	75
4.19	Satisfação de usuários conectados utilizando ADS.	76
4.20	Taxa de bloqueio utilizando ADS.	76
4.21	98 ^o percentil da taxa de descarte de pacotes utilizando ADS.	77
4.22	Carga servida em diferentes <i>mix</i> de serviços.	78
4.23	Satisfação de serviços VoIP e <i>Web</i> x Carga oferecida.	79

4.24	Satisfação de serviços VoIP e <i>Web</i> x Carga servida.	79
4.25	Satisfação de serviços VoIP e <i>Web</i> User utilizando PF.	80
4.26	Satisfação de usuários VoIP e <i>Web</i> conectados utilizando PF.	81
4.27	Taxas de bloqueio para serviços VoIP e <i>Web</i> utilizando PF.	82
4.28	Efeito do controle de admissão na carga servida utilizando PF.	82
4.29	Satisfação de serviços VoIP e <i>Web</i> utilizando RR.	83
4.30	Satisfação de usuários VoIP e <i>Web</i> conectados utilizando RR.	83
4.31	Taxas de bloqueio para serviços VoIP e <i>Web</i> utilizando RR.	84
4.32	Efeito do controle de admissão na carga servida utilizando RR.	85
4.33	Capacidade relativa de carga oferecida empregando.	86
4.34	Capacidade relativa de carga servida empregando.	86
4.35	Capacidade relativa de carga oferecida empregando.	87
4.36	Capacidade relativa de carga servida empregando.	88

Lista de Tabelas

3.1	Parâmetros do <i>grid</i>	24
3.2	Principais modelos de propagação.	25
3.3	Parâmetros do tráfego <i>Web</i>	26
3.4	Parâmetros do tráfego de voz.	27
3.5	Principais parâmetros de simulação.	29
4.1	Parâmetros do <i>grid</i>	60
4.2	Principais modelos de propagação.	61
4.3	Parâmetros do tráfego de voz.	62
4.4	Parâmetros do tráfego <i>Web</i>	62
4.5	Tempos de simulação.	64
4.6	Principais parâmetros de simulação.	65

Introdução

1.1 Motivação e Objetivos

A Terceira Geração (3G) dos sistemas de comunicação móvel [1] está em processo de implantação na Europa, Japão e Estados Unidos. Isto possibilita o oferecimento de uma gama de serviços que usualmente só poderiam ser providos em redes de comunicação fixa.

Juntamente com a implantação dos sistemas avançados de telefonia celular, se populariza rapidamente a utilização de serviços que proporcionam a transmissão de dados de voz sobre a *internet* utilizando o *Internet Protocol* (IP). Este serviço é chamado de voz sobre IP (VoIP, do inglês *Voice over IP*).

Dentro deste contexto, os sistemas de comunicação móvel celular foram inicialmente projetados para prover dois modos de conexão:

- ▶ Por chaveamento de circuitos, no qual uma conexão fixa é estabelecida durante todo o tempo de utilização do enlace. É onde o serviço de voz é comumente alocado.
- ▶ Por chaveamento de pacotes, em que as conexões são estabelecidas e liberadas de acordo com a necessidade de transmissão de cada usuário. É mais indicado para serviços de dados onde o padrão de tráfego se comporta por meio de rajadas de dados, o que permite o compartilhamento de recursos de forma mais adequada.

Inicialmente o modo de comunicação que vem sendo implantado se baseia neste paradigma. No entanto, com o advento da telefonia IP, o serviço de voz tende a ser transferido para o modo de operação por chaveamento de pacotes. Seguindo a

tendência da convergência de redes para o cenário *all-IP* [2,3]. Como este serviço apresenta requerimentos especiais com relação à interatividade na comunicação, ele necessita de um tratamento diferenciado no esquema de compartilhamento de recursos.

Como o modo de comunicação por chaveamento de pacotes foi basicamente concebido para serviços de melhor esforço, a motivação deste trabalho tem como base a possibilidade de se prover garantias de *Quality of Service* (QoS) para o serviço VoIP com a adoção de técnicas de controle de congestionamento. Os métodos abordados neste trabalho são o escalonamento de pacotes e o Controle de Admissão de Chamadas (CAC).

Espera-se que os métodos mencionados, sozinhos ou trabalhando conjuntamente, possam ser capazes de controlar o atraso atribuído aos usuários VoIP sob condições de altas cargas onde o limite de QoS seja afetado.

O objetivo deste trabalho tem como base a elaboração de algoritmos de escalonamento de pacotes e controle de admissão que visam a garantia de QoS para o serviço de VoIP e baseando-se no controle de atraso e da perda de quadros de voz.

Primeiramente, uma vasta análise bibliográfica foi efetuada tendo em vista o estado da arte dos algoritmos de controle de congestionamento. Com relação ao escalonamento, vários algoritmos foram propostos com o intuito de prover QoS para serviços de tempo real [4], no entanto nenhum foca explicitamente em cenários mistos, em que serviços com diferentes requerimentos compartilham os recursos. Dessa maneira, o problema de escalonamento em cenários mistos envolvendo o serviço VoIP mostra-se como um nicho bastante promissor e inexplorado.

Com relação ao CAC, até o presente momento, nenhum estudo explícito de controle de admissão foi realizado com o intuito de proporcionar QoS para o serviço VoIP. Nos trabalhos realizados em [5,6], esquemas de CAC baseados em métricas de QoS são avaliados, mas sem um foco no serviço VoIP, muito menos em cenários mistos contendo este serviço.

Dentro de todo este contexto detalhado, técnicas de escalonamento e CAC surgem como protagonistas de esquemas inteligentes de gerenciamento de recursos com o intuito do aumento da capacidade.

1.2 Metodologia

A metodologia adotada consiste basicamente em um estudo aprofundado de várias técnicas de controle de congestionamento aplicáveis ao problema de se garantir QoS a serviços VoIP. Logo após, algumas técnicas e algoritmos são formulados com o intuito de serem avaliadas. A verificação da validade dos esquemas propostos é efetuada por meio de simulações computacionais. Estas procuram emular todos os efeitos dinâmicos de um sistema real.

Foram utilizadas duas ferramentas de simulação. A primeira foi desenvolvida dentro do Grupo de Pesquisas em Telecomunicações Sem-Fio (GTEL) e foi destinada aos estudos de escalonamento de pacotes. A outra ferramenta utilizada foi a existente nos laboratórios da *Ericsson Research* e foi responsável pelos resultados referentes ao CAC.

1.3 Contexto de Desenvolvimento

O presente trabalho está em sincronismo com um projeto de parceria técnico-científica firmado entre o GTEL e a *Ericsson Research* que vislumbra o estudo de técnicas de gerenciamento de recursos para redes avançadas de telefonia celular.

Vale ressaltar que uma parte dos resultados obtidos neste trabalho foi obtida por meio de um estágio realizado em um laboratório de pesquisa da *Ericsson Research* na cidade de Luleå, na Suécia.

1.4 Contribuições e Produção Científica

No decorrer do desenvolvimento deste trabalho algumas contribuições técnico-científicas podem ser ressaltadas:

- ▶ construção de uma ferramenta de simulação sistêmica baseada no padrão do *Universal Mobile Telecommunications System* (UMTS) [1]. Esta pode ser utilizada em uma diversidade de estudos referentes a este sistema e com uma considerável capacidade de se aplicar os resultados obtidos nos sistemas reais, devido a natureza bastante fiel dos modelos utilizados na ferramenta. Dessa maneira, esta ferramenta possibilitou a avaliação dos mecanismos de controle de congestionamento propostos e a comparação com esquemas convencionais;
- ▶ confecção de três relatórios técnicos, sendo dois referentes a atividades realizadas no decorrer do projeto e o outro referente ao estágio realizado na

Suécia.

A seguir, os artigos relacionados ao tema deste trabalho de dissertação e que já foram publicados são apresentados:

Fast Cell Selection considering network constraints in HSDPA, André R. Braga, Leonardo S. Cardoso, Francisco R. P. Cavalcanti, XXII Simpósio Brasileiro de Telecomunicações-SBT'05, 4-8 de Setembro de 2005, Campinas, SP.

Evaluation of Fixed Thresholds for Allocation and Management of Dedicated Channels Transmission Power in WCDMA Networks, Carlos H. M. de Lima, Emanuel B. Rodrigues, Vicente A. de Sousa Jr., Francisco R. P. Cavalcanti, André R. Braga, Lecture Notes in Computer Science, v. 3124, p. 1122-1127, 2004.

QoS and Load Management via Admission Control in UMTS Forward Link, Emanuel B. Rodrigues, Carlos H. M. de Lima, Vicente A. de Sousa Jr., Francisco R. P. Cavalcanti e André R. Braga, IEEE Vehicular Technology Conference Spring, Maio, 2004.

Já outros artigos foram submetidos a conferências internacionais e estão em processo de avaliação:

Admission Control for VoIP over HSDPA in a Mixed Traffic Scenario, André R. Braga, Stefan Wänstedt e Märten Ericson, IEEE International Telecommunications Symposium, 3-6 de Setembro de 2006, Fortaleza, Brasil.

Novel Scheduling Algorithms Aiming for QoS Guarantees for VoIP over HSDPA, André R. Braga, Emanuel B. Rodrigues e Francisco R. P. Cavalcanti, IEEE International Telecommunications Symposium, 3-6 de Setembro de 2006, Fortaleza, Brasil.

Packet Scheduling for Voice over IP over HSDPA in Mixed Traffic Scenarios with Different End-to-End Delay Budgets, André R. Braga, Emanuel B. Rodrigues e Francisco R. P. Cavalcanti, IEEE International Telecommunications Symposium, 3-6 de Setembro de 2006, Fortaleza, Brasil.

Admission Control for VoIP over HSDPA, André R. Braga, Stefan Wänstedt e Märten Ericson, IEEE International Symposium on Personal Indoor and Mobile Radio Communications, 11-14 de Setembro de 2006, Helsinque, Finlândia.

Packet Scheduling for VoIP over HSDPA in a Mixed Traffic Scenario, André R. Braga, Emanuel B. Rodrigues e Francisco R. P. Cavalcanti, IEEE

International Symposium on Personal Indoor and Mobile Radio Communications, 11-14 de Setembro de 2006, Helsinque, Finlândia.

1.5 Estrutura do Trabalho

Esta dissertação está estruturada da seguinte forma:

- ▶ Capítulo 2 - o sistema onde este trabalho se focaliza além da contextualização técnica, como a descrição dos serviços e métricas de desempenho são abordados.
- ▶ Capítulo 3 - uma das técnicas abordada neste estudo é o escalonamento de pacotes. Neste capítulo, alguns algoritmos de escalonamento projetados para lidar com o problema de QoS para VoIP são formulados e comparados a algoritmos clássicos.
- ▶ Capítulo 4 - a outra técnica avaliada é o CAC. Neste capítulo, o esquema é proposto e aplicado com o intuito de se garantir o QoS para o serviço VoIP.
- ▶ Capítulo 5 - neste capítulo as considerações finais, conclusões e perspectivas são apresentadas.

Sistema UMTS/HSDPA e Serviço VoIP

Este capítulo tem como finalidade a descrição do sistema em que este trabalho é baseado, além das propriedades do serviço de VoIP. Os requerimentos e métricas de avaliação do sistema também são abordados.

2.1 *Universal Mobile Telecommunications System (UMTS)*

O *Wideband Code Division Multiple Access* (WCDMA) é a interface aérea adotada pelo *3rd. Generation Partnership Project* (3GPP) para a 3G, também chamado de UMTS. Ele é caracterizado pelo múltiplo acesso por divisão de código em seqüência direta (DS-CDMA, do inglês *Direct Sequence Code Division Multiple Access*), onde os bits de informação são espalhados em uma banda de 5 MHz através da multiplicação da seqüência de bits de informação por uma seqüência de bits chamada código de espalhamento [7].

O WCDMA apresenta uma flexibilidade nas taxas de transmissão, cuja variação torna-a capaz de prover os múltiplos tipos de serviços e aplicações listados para serem oferecidos pelo UMTS. Tais serviços são classificados como: conversacional, *streaming*, interativo e *background* [1].

Arquitetura de Rede

O UMTS é composto por um certo número de elementos lógicos de rede, cada um com uma funcionalidade bem definida [7]. A Figura 2.1 apresenta uma ilustração da estrutura da rede de uma maneira resumida¹. O maior foco deste trabalho reside em três elementos da rede:

- ▶ **UMTS Terrestrial Radio Access Network (UTRAN)** - que lida com todas as funcionalidades de rádio referentes à rede de acesso;
- ▶ **Core Network (CN)** - responsável pela interface com redes externas, como a *Public Switched Telephone Network* (PSTN) e a *internet*;
- ▶ **User Equipment (UE)** - que intermedia a rede de rádio e o usuário assinante.

O UTRAN consiste basicamente de dois elementos principais: a *Radio Network Controller* (RNC) e o *Node B*. O primeiro é responsável por aplicar funcionalidades de gerenciamento de recursos de rádio (RRM, do inglês *Radio Resource Management*) nos *Node Bs* controlados e também age como um ponto de acesso à CN. Já o *Node B* tem como principal função a conversão do fluxo de dados provenientes da RNC em sinais de rádio com destino aos UEs. Com a padronização do *High Speed Downlink Packet Access* (HSDPA) [9], o *Node B* fica também responsável, além do controle de potência em laço interno, em prover algumas funcionalidades de RRM como *Hybrid Automatic Repeat Request* (HARQ) e escalonamento.

São também padronizadas na arquitetura as interfaces de comunicação entre os elementos da rede. A interface *Iur* tem a função de intermediar a comunicação entre diferentes RNCs. Já a interface *Iub* é responsável de comunicar a RNC e os *Node Bs*. A interface aérea propriamente dita é a *Uu*, que comunica o *Node B* ao UE.

Os sistemas WCDMA que vêm sendo implantados oferecem serviços alocando canais dedicados, *Dedicated Channels* (DCHs), que apresentam uma alocação fixa de códigos de canalização e o controle de potência. Os quadros de voz são entregues pelo Codificador-Decodificador (CODEC) de voz e transmitidos através do enlace de rádio com praticamente nenhum *overhead*, ou seja, nenhuma adição extra de cabeçalhos para endereçamento pelo fato de se utilizar o modo de chaveamento de circuito.

Com relação ao fluxo de dados no sistema, as principais entidades da Camada de Enlace (L2) são a *Radio Link Control* (RLC) e *Medium Access Control* (MAC).

A primeira situa-se na RNC e é responsável por diversas funcionalidades no fluxo de pacotes como: segmentação, contacenação, retransmissão, controle de fluxo, dentre outras. Sua funcionalidade básica é de receber pacotes das camadas

¹Figura 2.1 foi extraída de [8].

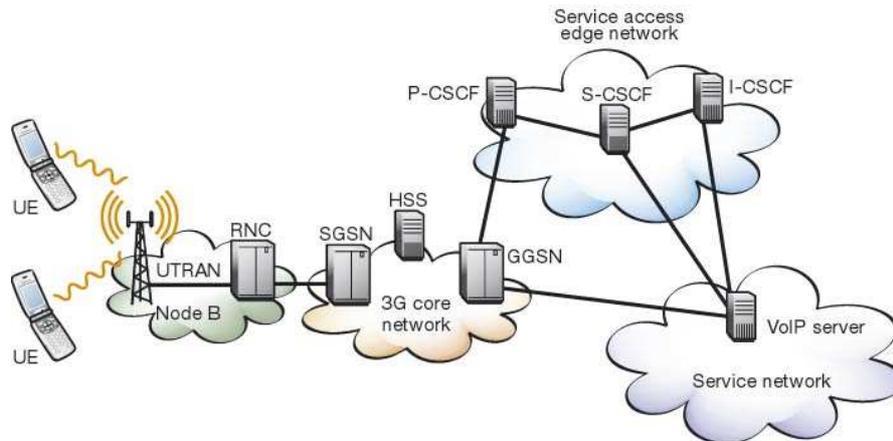


Figura 2.1: Arquitetura UTRAN.

superiores, chamados de RLC *Service Data Units* (SDUs), processá-los na forma de pacotes de tamanhos iguais através da segmentação/concatenação e adição de cabeçalho e, por fim, enviá-los a camada inferior, agora denominados RLC *Protocol Data Units* (PDUs). Ela pode ser configurada basicamente de três modos: transparente, sem confirmação e com confirmação. Em inglês, estes são chamados *Transparent Mode* (TM), *Unacknowledged Mode* (UM) e *Acknowledged Mode* (AM), respectivamente.

O TM não aplica nenhuma modificação aos pacotes recebidos, enviando-os da mesma maneira que recebe. Ele é utilizado para serviços de voz por chaveamento de circuitos. Ambos AM e UM operam de maneira semelhante no processamento dos pacotes recebidos da camada superior. A diferença básica reside no fato de o primeiro apresentar o protocolo de retransmissão chamado *Automatic Repeat Request* (ARQ), o que não ocorre com o UM [10].

Já a MAC é a camada logo abaixo da RLC que recebe os pacotes da mesma, agora denominados MAC SDUs. Ela é subdividida em várias subcamadas, mas as mais importantes de serem consideradas são a MAC-d, situada também na RNC e responsável pelo tratamento de cada fluxo de um usuário, e a MAC-hs, situada no *Node B* e responsável pelas funcionalidades referentes ao HSDPA.

A MAC-hs tem como principais finalidades o recebimento de pacotes da MAC-d (MAC-hs SDUs) e concatená-los em um bloco de transporte (MAC-hs PDU) com tamanho variado e especificado pelo esquema de adaptação de enlace [11].

Controle de potência

O controle de potência é peça essencial para a operação de sistemas limitados por interferência. No WCDMA, existem três esquemas de controle de potência: os de laços aberto, interno e externo.

O controle de potência de laço aberto tem a finalidade de estabelecer a potência inicial para cada enlace de rádio no momento de seu estabelecimento. Esta potência inicial pode ter um valor fixo, ou mesmo um valor estimado a partir da estimação da qualidade do canal a partir da recepção de outro canal existente como, por exemplo, o piloto que no UMTS é denominado *Common Pilot Channel* (CPICH).

Já o controle de potência em laço interno, tradução de *Inner-loop Power Control* (ILPC), é responsável pelo ajuste da potência de transmissão de acordo com as medidas de *Carrier-to-Interference Ratio* (CIR) efetuadas a cada *slot* e um valor alvo. Ele é modelado de acordo com [12]. Com o intuito de garantir que os valores absolutos da potência do canal não exceda determinados limiares, a norma [12] estabelece que o controle de potência deve ter um teto e um piso de potência com valores importados de [13].

Em um passo de tempo maior opera o controle de potência de laço externo, em inglês chamado de *Outer-loop Power Control* (OLPC). Ele é responsável pelo ajuste da CIR alvo utilizada pelo ILPC baseado em medidas de *Block Error Rate* (BLER).

Soft Handover (SHO)

O procedimento de SHO objetiva uma mudança gradual na troca de bases e também proporciona um ganho de diversidade do fato de um único UE poder ser servido por mais de uma célula. O conjunto de células transmitindo a um único UE é chamado de *Active Set* (AS). A implementação do algoritmo é baseado em [14]. Os três principais eventos do SHO são modelados: 1A, 1B e 1C, os quais são descritos a seguir. Existe também um intervalo de filtragem, quando a métrica é reportada do UE para o sistema. Depois de alguns intervalos de filtragem, vêm o tempo de disparo quando, de acordo com a média das métricas reportadas, alguns eventos devem ser disparados. Assim, podemos definir:

- **Evento 1A** - quando a métrica reportada está maior que o limiar, tomando como referência a métrica reportada da melhor célula dentro do AS. O novo enlace (conexão a uma nova célula) é adicionado ao AS se o controle de admissão para este enlace for autorizado.

- ▶ **Evento 1B** - quando a métrica reportada para a pior célula dentro do AS está abaixo do limiar em relação à melhor célula, o enlace é removido do AS.
- ▶ **Evento 1C** - quando uma célula não pertencente ao AS tem uma métrica reportada melhor que a pior célula no AS a partir de um determinado limiar, o evento de substituição ocorre.

Como apenas uma célula do AS é permitida de transmitir dados utilizando o HSDPA para o usuário de interesse, esta é denominada *célula servidora*. O procedimento de troca da *célula servidora* é efetuado a cada tempo de disparo ao se escolher a célula com a melhor métrica dentro do AS.

2.1.1 *High Speed Downlink Packet Access (HSDPA): Evolução do Padrão*

O HSDPA é um conceito que emergiu como uma atualização do WCDMA com uma ótima relação custo-benefício e está presente a partir do *Release 4 (R4)* [15,16]. O principal apelo é o aumento das taxas de pico, levando a uma melhoria no QoS e da eficiência espectral para serviços assimétricos, isto é, com uma grande demanda de tráfego no enlace direto.

Os corpos de padronização prevêm que o HSDPA deve coexistir com a primeira versão do WCDMA *Release '99 (R99)*, permitindo uma atualização suave e com o intuito de comportar a crescente demanda por serviços de internet sem fio com altas taxas.

O HSDPA se caracteriza pela adição de algumas novas funcionalidades que são listadas a seguir:

Canal compartilhado

O *High Speed Downlink Shared Channel (HS-DSCH)* é o mecanismo lógico de transporte para a comunicação em pacotes. Ele é baseado em uma transmissão compartilhada através de um algoritmo definido de escalonamento com uma alocação variável de recursos.

O HS-DSCH tem um fator de espalhamento fixo em 16 e pode fazer uso simultâneo de até 15 códigos de canalização. A duração do quadro de transmissão, também chamado de *Transmission Time Interval (TTI)*, é estabelecida como igual a 2 ms.

Vale ressaltar que cada usuário do HS-DSCH aloca um canal dedicado, denominado *Associated Dedicated Physical Channel (A-DPCH)*, para troca de informações de controle.

Escalonamento Rápido

O escalonamento também é uma responsabilidade da MAC-hs. Ele opera a cada TTI do HSDPA, que é igual a 2 ms. O número de usuários escalonados em cada TTI por célula depende de um parâmetro de configuração da célula que dita o número máximo de usuários multiplexados em código. Esta funcionalidade é extremamente importante quando serviços como VoIP são oferecidos, porque são muito sensíveis ao atraso e não demandam altas taxas, i.e. não existe uma demanda por configurações de transmissão com alta eficiência espectral. Para prover a multiplexação de código, a célula deve ter um *High Speed Shared Control Channel* (HS-SCCH) para cada usuário multiplexado em código. Este parâmetro deve ser cuidadosamente configurado já que cada HS-SCCH demanda códigos de canalização e potência, o que tem influência direta na capacidade do sistema.

Na modelagem assumida, depois de definida a prioridade de cada usuário baseado no algoritmo de escalonamento empregado, o escalonador configura a transmissão para o usuário de maior prioridade em cada célula. Dependendo da disponibilidade de códigos restantes, o escalonador pode escolher o usuário com a segunda maior prioridade na fila de transmissão, e assim por diante até que o número máximo de usuários multiplexados em código seja atingido.

O escalonador tem também a função de descartar MAC-hs SDUs que tenham expirado seus tempos de espera [11,17]. Este procedimento é de suma importância para serviços VoIP, já que pacotes muito antigos serão descartados nas camadas superiores. O UE também possui um *buffer* de reordenamento [11], dessa maneira, um pacote muito atrasado acarreta em um retardo excessivo nos pacotes subsequentes para atingir a camada de aplicação. Assim, o mecanismo de descarte da MAC previne as filas do sistema de serem carregadas com pacotes inúteis.

Adaptação de enlace

A adaptação de enlace é a principal funcionalidade do HSDPA. Através desta técnica, o sistema é capaz de mudar os parâmetros de transmissão com o objetivo de se obter uma maior vazão ou mesmo se tornar mais robusto contra erros na transmissão de acordo com a qualidade do enlace. Para o HSDPA, dois parâmetros podem ser ajustados:

- **Modulação** - que pode ser escolhida entre o *Quadrature Phase Shift Key* (QPSK) e o *16-Quadrature Amplitude Modulation* (16QAM).

- **Taxa de codificação de canal** - através de esquemas de puncionamento e repetição, haja vista que um codificador turbo com taxa fixa de 1/3 é utilizado [18].

A seleção do esquema de transmissão mais indicado para as atuais condições de canal é decidido através do envio da informação de *Channel Quality Indicator* (CQI), que também define o tamanho do bloco de transporte e o número de códigos de canalização alocados para determinada transmissão [12].

Retransmissões utilizando ARQ híbrido

Assim que o UE recebe o bloco de dados previamente transmitido, ele detecta se ele foi recebido corretamente ou não. Logo em seguida, é sinalizado para o *Node B* o sucesso ou insucesso na transmissão.

Em caso de recepção incorreta, o bloco de dados deve ser retransmitido. No esquema de retransmissão híbrido, os blocos de dados recebidos incorretamente não são descartados, mas sim combinados com as próximas retransmissões. Este procedimento é chamado de *soft combining*.

Existem basicamente dois esquemas de combinação:

- **Chase Combining** - neste esquema, o transmissor envia réplicas dos blocos de transporte já transmitidos e o receptor aplica o *Maximal Ratio Combining* (MRC) na combinação.
- **Redundância Incremental** - onde bits adicionais de redundância são enviados nas retransmissões. Ele proporciona um maior ganho de codificação mas exige maiores tamanhos de *buffers* no UE [19].

Também é empregado o protocolo *Stop-And-Wait* (SAW) onde um novo dado é transmitido apenas quando o último dado enviado recebe uma confirmação de recepção correta. Para se evitar atrasos devido à espera pela confirmação, vários processos SAW são empregados paralelamente e acessados de forma cíclica [15, 16].

2.2 Voz sobre IP

O lançamento das redes 3G possibilitam uma grande variedade de serviços que não eram possíveis de serem providos previamente. Dentre estas novas ofertas, alguns têm requerimentos bastante especiais além das altas taxas de transmissão. São os chamados serviços de tempo-real.

Atualmente, um serviço com bastante destaque é o de voz sobre IP (VoIP), que transporta dados de voz sobre a internet, que pode ser provida por sistemas sem fio de pacotes, como o HSDPA. O interesse pelo VoIP pode ser vantajoso por sinalizar possíveis reduções de custos para as operadoras, no sentido de que a parte da rede que opera por chaveamento de circuitos poder ser removida, necessitando as operadoras de apenas um único tipo de rede para todos os tipos de comunicação, quer seja voz ou dados [20]. Este é um dos degraus no caminho do ambiente *all-IP* [3].

Uma potencial vantagem de se utilizar canais compartilhados no tempo para serviços de voz é a possibilidade de se otimizar a utilização de recursos [21], já que potência e códigos de canalização são alocados dinamicamente e podem ser liberados durante os períodos de silêncio. A introdução de novos serviços de multimídia que agregam diferentes tipos de tráfego (audio, vídeo e texto) também se torna mais simples. Por outro lado, algum *overhead* de protocolo é necessário, para que os quadros de voz se utilizem da camada IP. Isto leva a um aumento na banda requerida para o serviço de voz, quando comparado à necessária para a transmissão por chaveamento de circuitos [22].

2.2.1 Qualidade para VoIP

O QoS para o serviço VoIP é basicamente influenciado por três fatores: atraso de pacote, *jitter* de atraso e perda de pacote.

O atraso é o montante de tempo que a voz leva para atingir seu destino. O atraso e suas conseqüências são alguns dos mais importantes aspectos que afetam o serviço VoIP. Dois problemas que resultam em um grande atraso fim-a-fim em uma comunicação de voz são o eco e a sobreposição de quadro [23].

O *jitter* de atraso é a diferença entre o tempo esperado para a chegada do pacote para seu tempo real de chegada. Já que pacotes de voz são gerados a uma taxa constante, todo pacote é esperado chegar no destino a cada 20 ms, tomando-se como referência a operação do CODEC *Adaptive Multirate* (AMR) em questão. No entanto, isto nem sempre ocorre já que os dispositivos da rede causam um montante de atraso imprevisível nos pacotes [23]. Além disso, o escalonador localizado no *Node B* adiciona *jitter* porque ele distribui os recursos entre os usuários de acordo com um determinado critério (condições de canal, justiça, etc.). O *jitter* de atraso surge do fato de alguns pacotes esperarem para serem escalonados. Adiciona-se a isto o fato de pacotes que não são recebidos corretamente devem ser retransmitidos pela funcionalidade de HARQ até que sejam corretamente recebidos ou que o máximo

número de retransmissões seja atingido. As retransmissões do HARQ adicionam ainda mais *jitter* de atraso ao fluxo de pacotes [8].

Para compensar as variações de atraso, o receptor emprega um *buffer* de *play-out*. Este é projetado para suavizar a variação de atraso na qualidade da voz recebida, de maneira que a voz reconstruída não seja afetada pelo *jitter*. É relevante dizer que estes *buffers* resultam em um atraso adicional e podem descartar pacotes que chegam muito atrasados. É melhor descartar um pacote do que aumentar a latência de todos os outros pacotes. *Buffers* com grandes tamanhos diminuem a perda de pacotes mas comprometem a interatividade [8, 23].

É assumido neste trabalho que o *buffer* de *play-out* do UE é configurado de tal maneira que virtualmente elimina o *jitter* de atraso dos pacotes VoIP. Assim, este trabalho não avalia o comportamento do *jitter* de atraso.

O desempenho do serviço VoIP é medido em termos de taxa de perda de pacotes, i.e. o percentual dos pacotes que não são entregues ao receptor. Apesar de a perda de pacotes de qualquer maneira ser indesejável, alguma perda pode ser tolerada. CODECs utilizam sinais de áudio digitalmente amostrados que podem mascarar uma perda de pacote ao se aplicar técnicas de recuperação de quadro. Com isso, uma boa inteligibilidade de voz para percepção humana pode ser garantida. Estas técnicas são efetivas quando pacotes são perdidos individualmente ou de maneira aleatória, mas se mostram ineficientes quando lidam com perdas em rajadas [23], tornando a conversação impossível [2].

2.3 Controle de Congestionamento

Mecanismos de Controle de Congestionamento (CC) são parte de qualquer estratégia de gerenciamento em redes de comunicação. Existem duas classes básicas de técnicas de CC: controle reativo e preventivo [24]. A primeira classe tenta detectar o congestionamento o mais rápido possível e então toma as medidas corretivas apropriadas. No entanto, a detecção prévia é extremamente complexa e possivelmente não confiável [25]. A maioria dos esquemas de CC tentam prevenir sua ocorrência ou ao menos minimizar seus efeitos.

Dessa maneira, em sistemas de comunicação com multiplexação temporal, o escalonamento é um dos mecanismos de CC que tenta evitar que o congestionamento ocorra através da distribuição otimizada de janelas de transmissão (recursos). Este distribui recursos de modo que o QoS dos usuários na rede possa ser garantido baseado em determinadas métricas de desempenho. Algoritmos de escalonamento

mais apurados se fazem necessário principalmente devido às flutuações estatísticas das diferentes fontes de tráfego existentes [26]. Em [4], o escalonador de atraso é desenvolvido com o intuito de prover QoS para usuários VoIP, tendo em mente que o principal requerimento deste serviço é o atraso.

Já o controle de admissão, que também pode ser considerado um esquema preventivo, evita que o sistema atinja estados de sobrecarga através do bloqueio de novos usuários. A operação básica do controle de admissão é limitar o acesso de usuários ao sistema se este não tem meios de prover uma boa qualidade para o serviço requerido, e também para proteger os usuários já conectados contra grandes variações na qualidade.

2.4 Medidas de Desempenho

Ao simular o comportamento de uma rede celular real, as avaliações são feitas baseadas em métricas retiradas das simulações que indicam o desempenho do sistema. Tais medidas são listadas e detalhadas nesta seção.

Bloqueio

A taxa de bloqueio é a razão entre o número de chamadas negadas pelo controle de admissão e o número total de chamadas inicializadas. É possível discriminar as métricas de bloqueio para cada classe de serviço, em caso de algum esquema de priorização de alguma delas na admissão.

Esta métrica é apenas avaliada quando o controle de admissão é aplicado, o que será visto apenas no Capítulo 4.

Carga

Duas diferentes representações de carga são consideradas neste trabalho. A principal diferença entre elas é a influência do bloqueio, no caso de o algoritmo de controle de admissão ser empregado. No caso de carga oferecida (em número médio de usuários por célula), quando assumida, nenhuma consideração é feita se os usuários que tentam se conectar são aceitos ou não. No caso da carga servida, é considerada a vazão total de dados no sistema normalizada pelo número de células. É uma métrica bastante importante do ponto de vista do provedor de serviço, pois os serviços de dados devem ser taxados pelo volume de dados transmitidos.

Uma ilustração interessante é a de como se comporta a carga servida no sistema com uma crescente carga oferecida. Isto dá uma boa idéia de como a carga oferecida está sendo servida no sistema.

No Capítulo 4, a representação da carga no sistema é também normalizada por um valor de referência, que é assumido como sendo a configuração do sistema utilizando o escalonador *Proportional Fair* (PF) e também sem o controle de admissão.

Taxa de Apagamento de Quadro

Neste trabalho, a única real medida de desempenho de um usuário VoIP é a taxa de apagamento de quadro, ou *Frame Erasure Rate* (FER). As únicas possíveis fontes de um pacote ser perdido são:

- ▶ Perda devido a erros no canal - quando o máximo número de tentativas de transmissão na MAC-hs é atingido sem sucesso, o pacote é perdido pois a MAC reporta uma confirmação negativa de envio do pacote e, como a RLC trabalha no UM na configuração deste serviço, o quadro de voz acaba por ser perdido.
- ▶ Perda devido ao descarte na MAC-hs - quando um pacote permanece na MAC-hs por mais tempo que um determinado limiar de atraso do pacote, ele é descartado antes mesmo de ser transmitido, e então entra na contabilização de pacotes perdidos.
- ▶ Perda devido ao descarte no *buffer* de *play-out* - mesmo após ser transmitido, o pacote ainda pode chegar no receptor com um atraso indevido. Neste caso, ele é descartado na aplicação. Vale ressaltar que essa métrica é apenas abordada no estudo desenvolvido no Capítulo 3.

Vazão Média por Pacote

Para o serviço *Web*, a medida de desempenho assumida é a vazão média de pacotes durante a sessão. Ela é calculada através da média aritmética de todos os valores de vazão de cada pacote. Ao ser recebido, um pacote pode ter seu valor de vazão calculado através da Equação (2.1):

$$R_p = \frac{S_p}{T_{fp} - T_{ap}}, \quad (2.1)$$

em que R_p é a vazão do pacote p , S_p é o tamanho de p , T_{fp} e T_{ap} são os tempos de recepção e de criação do pacote p , respectivamente.

É importante ressaltar que a métrica de vazão média por pacote não considera os períodos de inatividade da sessão como os tempos de leitura.

Satisfação de Usuário

A satisfação de um usuário depende diretamente de qual classe de serviço ele se utiliza. Com relação ao serviço de voz, a satisfação é atingida quando o mesmo tem uma FER menor que 1%, já que, para o CODEC AMR, uma FER de 2% deve ser garantida (1% para cada direção do enlace) para uma boa qualidade de voz [27]. Para usuários que se utilizam do serviço *Web*, a satisfação é garantida quando a vazão média de pacote é, no mínimo, 64 kbps.

O percentual de usuários satisfeitos é calculado para cada carga oferecida e a capacidade do sistema é atingida quando o serviço atinge seu limite de QoS, que é definido como sendo uma taxa de satisfação mínima. Para os serviços de VoIP e *Web*, os limites de QoS considerados foram de 90% [28]. Quando um algoritmo de controle de admissão é empregado, os usuários bloqueados são considerados como insatisfeitos.

Uma outra maneira de se avaliar o índice de satisfação é restringindo a análise a apenas os usuários conectados ao sistema (não bloqueados). Para uma comparação mais justa entre as várias configurações do sistema, a taxa de satisfação de usuários conectados é relacionada com a carga servida, dando uma ideia entre a relação custo-benefício entre QoS e vazão no sistema (lucro para a operadora).

Capítulo 3

Escalonamento de Pacotes para VoIP em HSDPA

O problema de escalonamento em sistemas sem fio tem uma característica básica que o distingue dos sistemas cabeados: a alta variabilidade do enlace. Esta é a razão pela qual esta técnica tem um alto fator de importância em prover serviços de dados em chaveamento de pacotes.

Os algoritmos empregados devem estar atentos a diversos quesitos tais como: o enlace de rádio, para se evitar o desperdício de recursos durante más condições de canal; a justiça, ao se tentar garantir uma certa taxa de serviço a um determinado fluxo [29]; o QoS por meio da diferenciação de serviços, que determina diferentes prioridades; a complexidade, que deve permitir a implementação em sistemas reais de comunicações.

O fluxograma apresentado na Figura 3.1 ilustra o funcionamento genérico dos algoritmos de escalonamento. A função de prioridade de cada usuário é calculada se utilizando de informações do canal e de métricas de qualidade. Após este cálculo, os usuários são ordenados de acordo com o resultado deste cálculo. Logo após, o escalonamento propriamente dito é efetuado através da alocação de recursos (códigos e potência) de acordo com a ordem de prioridade estabelecida e à limitação do número máximo de usuários escalonados simultaneamente.

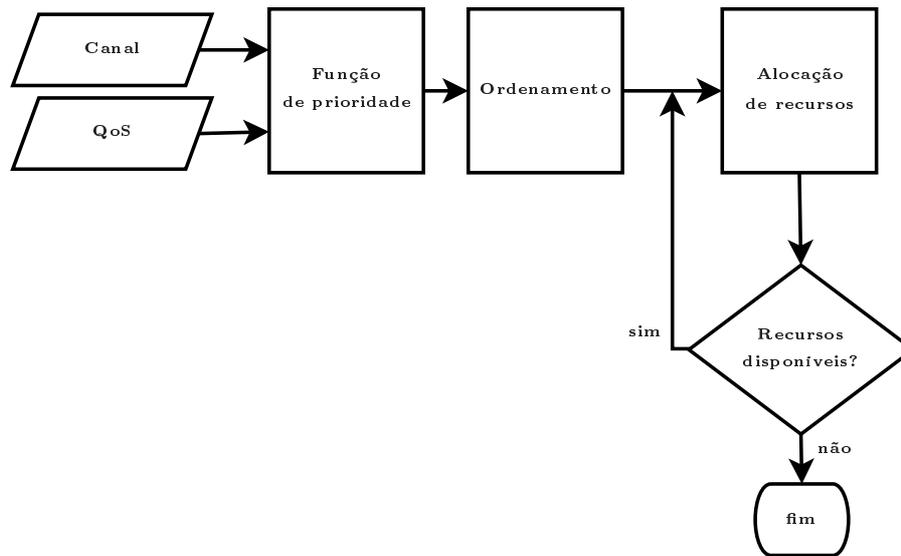


Figura 3.1: Funcionamento do esquema de escalonamento.

Este capítulo propõe algumas alternativas para o escalonamento para o serviço VoIP no HSDPA e faz uma extensa avaliação comparativa com outros algoritmos clássicos. Primeiramente os algoritmos são formulados na Seção 3.1. Em seguida, a Seção 3.2 descreve a ferramenta de simulação utilizada para a obtenção dos resultados. Estes são apresentados na Seção 3.3. Ao final do capítulo, na Seção 3.4, algumas considerações parciais são realizadas.

3.1 Algoritmos de Escalonamento

Esta seção descreve cada algoritmo utilizado neste estudo. Eles foram agrupados de duas maneiras: QoS-diferenciado e QoS-não-diferenciado. Este último não é capaz de diferenciar entre serviços ou demandas de QoS para cada usuário específico. As capacidades de todos os escalonadores apresentados em prover QoS a ambos os serviços VoIP e *Web* são avaliadas neste capítulo.

A descrição de cada algoritmo consiste no cálculo da função de prioridade para cada usuário. Deve-se ter em mente que, para todos os algoritmos, usuários com retransmissões têm maior prioridade sobre aqueles com novos pacotes. Esta prioridade é expressa através de um fator multiplicativo aplicado no cálculo da prioridade.

3.1.1 Algoritmos de QoS-não-diferenciado

Round Robin (RR)

Os usuários são servidos de forma cíclica ignorando-se as condições de canal [30]. O cálculo de prioridade é baseado no tempo de enfileiramento de cada usuário. É importante salientar que o tempo de enfileiramento é atualizado apenas se o *buffer* de transmissão não se encontra vazio. Este algoritmo proporciona uma distribuição igual de recursos dentre todos os usuários na fila.

Proportional Fair (PF)

Este algoritmo tem o intuito de servir usuários com condições favoráveis de canal em prover uma alta taxa de transmissão instantânea relativa a sua vazão média [30]. O valor de prioridade p_i para o usuário i é calculado baseado na Equação (3.1):

$$p_i = \frac{R_i}{T_i}, \quad (3.1)$$

em que R_i é a taxa estimada para a próxima tentativa de transmissão e T_i é a vazão média suportada pelo usuário.

Queue Based Max CIR (QBMC)

Este algoritmo incorpora ao algoritmo clássico Max CIR uma informação referente ao tamanho da fila no *buffer* de transmissão de cada usuário, em número de bits. Alguns outros algoritmos propostos em [30] e [31] também levam em consideração esta informação. Através dela, o escalonador é capaz de priorizar filas que não vêm sendo servidas devido a suas condições de canal, resultando em uma maior justiça entre os usuários. De fato, é utilizada uma informação indireta de atraso. O cálculo da prioridade p_i para o usuário i é apresentado na Equação (3.2):

$$p_i = C_i \cdot B_i, \quad (3.2)$$

em que C_i é a CIR instantânea medida e B_i é o tamanho do *buffer* em número de bits ainda não transmitidos para o usuário i .

3.1.2 Algoritmos de QoS-diferenciado

Linear Delay Scheduler (LDS) e Asymptotic Delay Scheduler (ADS)

Originalmente, o ADS foi proposto em [4] exclusivamente para o serviço VoIP utilizando uma Função Barreira (FB) cujo valor era multiplicado pela mesma função de prioridade do escalonador PF. Isto resulta em uma função de prioridade que

está ciente dos requerimentos de atraso de cada usuário específico, assim como a razão entre taxa de bits instantânea e média. Equação (3.3) apresenta o cálculo de prioridade p_i para cada usuário i :

$$p_i = \frac{R_i}{T_i} \cdot F_i, \quad (3.3)$$

em que F_i é a FB do usuário i , que pode ser calculada utilizando Equações (3.4) ou (3.5):

$$F_i^{ADS} = 1 + \frac{1}{D_{th_{s_i}} - D_i}, \quad (3.4)$$

$$F_i^{LDS} = \frac{99}{D_{th_{s_i}}} \cdot D_i + 1, \quad (3.5)$$

nas quais $D_{th_{s_i}}$ é o limiar de atraso do serviço s_i e D_i o valor de atraso do pacote mais antigo na fila de transmissão do usuário i . A primeira é uma variação da FB proposta em [4] para a incorporação de serviços de tempo não-real. Já a outra equação representa uma forma alternativa de cálculo da FB considerando uma função linear. Quando D_i é igual ou maior que $D_{th_{s_i}}$, um valor constante igual a 100 é assumido para a FB.

O algoritmo LDS também é baseado em alguns outros trabalhos [30, 32] que se utilizam de uma componente linear do atraso para ser incorporada à função de prioridade.

Esta formulação proposta para a FB tem a capacidade de proporcionar a diferenciação de serviços de maneira que, os valores assumidos para $D_{th_{s_i}}$ são baseados nos requerimento da classe de serviço do usuário. A configuração desse parâmetro o torna um algoritmo capaz de buscar o requerimento de atraso de cada serviço. O cálculo da FB apresentado tem a vantagem de, quando o serviço não demandar nenhum requerimento de atraso, $D_{th_{s_i}}$ pode ser configurado com um valor bastante alto de forma que a FB tem valor sempre próximo a 1, tornando o cálculo de prioridade para o serviço igual ao PF.

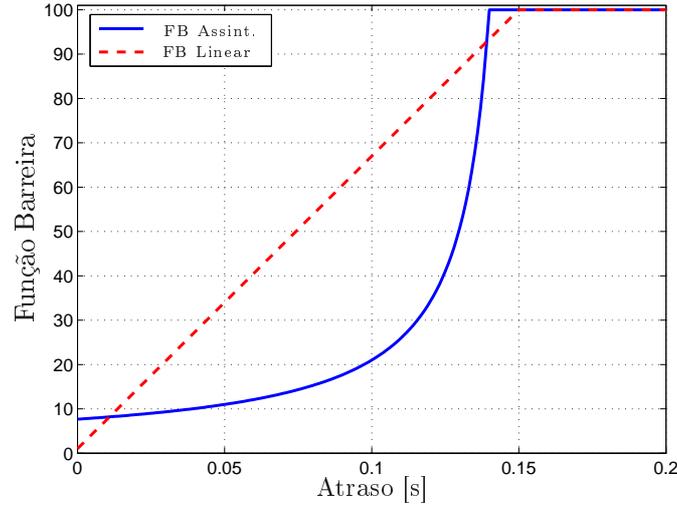


Figura 3.2: Comportamento da função barreira.

A Figura 3.2 ilustra o comportamento de ambas as FBs considerando $D_{th_{s_i}} = 150 \text{ ms}$. Pode-se perceber que a função linear apresenta maiores valores para baixos atrasos enquanto que a função assintótica realmente cresce apenas quando o atraso atinge valores próximos ao limiar.

Neste trabalho, o limiar de atraso assumido pelo escalonador é o mesmo adotado pelo mecanismo de descarte na MAC-hs. Para serviços VoIP, o valor depende da configuração do atraso requerido pelo serviço. Quando usuários *Web* são considerados, o limiar considerado tem valor igual a infinito.

Weighted Proportional Fair (WPF)

Alguns algoritmos desenvolvidos em [32] e [33] estabelecem algum tipo de prioridade fixa para determinadas classes de serviço. O algoritmo WPF trabalha praticamente da mesma maneira que o PF clássico. A única diferença é um peso multiplicativo, W_{s_i} , que se traduz como um fator de diferenciação de QoS a ser considerado para o serviço s_i do usuário i , como pode ser visto na Equação (3.6).

$$p_i = W_{s_i} \frac{R_i}{T_i}. \quad (3.6)$$

Esta é uma maneira simples de se estabelecer uma hierarquia de prioridade entre diferentes classes de serviço e se utilizando de um algoritmo clássico. Os valores de peso arbitrariamente assumidos para VoIP e *Web* são, respectivamente, 2, 0 e 1, 0. Tais valores tentam refletir uma tendência de priorização para os serviços de voz, sendo considerado o mais importante.

3.2 Modelos de Simulação

O presente estudo fez uso de um simulador dinâmico construído utilizando a linguagem de programação C++ que modela o enlace direto do sistema WCDMA com suporte ao HSDPA. Esta seção consiste nos modelos computacionais utilizados nesta ferramenta de software.

3.2.1 Grid Celular

A área de cobertura é composta por sítios macrocelulares, onde cada sítio é formado por três setores/células hexagonais. Uma distância entre sítios de 1,5 km é considerada, i.e. um raio de célula igual a 500 m. A Tabela 3.1 apresenta os principais parâmetros referentes à modelagem do *grid* celular.

Parâmetro	Valor	Unidade
Raio de Célula	500	m
Distancia entre Sítios	1500	m
Número de Sítios	9	-
Células por Sítio	3	-

Tabela 3.1: Parâmetros do *grid*.

O conjunto de células assumido é formado por 27 células, onde todas geram resultados estatísticos no decorrer das simulações. Com o intuito de se evitar efeitos de borda no cálculo de interferência, a técnica de *wrap around* é utilizada [34,35]. Dessa maneira, pode-se dizer que os resultados de simulação são confiáveis de um ponto de vista estatístico e o modelo pode representar sistemas celulares reais.

3.2.2 Propagação

Para o ambiente macrocelular, a perda média de propagação segue as considerações do ambiente de teste veicular, que é um dos ambientes operacionais do UMTS descritos em [1]. Este modelo de implantação é caracterizado por células grandes e altas potências de transmissão. Todos os cenários são construídos utilizando-se da descrição e caracterização UMTS encontrado em [1]. Um resumo dos parâmetros considerados para a modelagem dos efeitos de propagação é mostrado na Tabela 3.2.

Os efeitos de propagação são divididos em três tipos distintos que modelam os diferentes processos de mitigação do sinal. Eles são:

- ▶ Desvanecimento de larga escala - a área de cobertura do sistema é mapeada

Parâmetro	Valor	Unidade
Perfil de Canal	AWGN	-
Velocidade do Móvel	3	km/h
Desvio Padrão do Sombreamento	8,0	dB
Correlação do Sombreamento entre Sítios	0,5	-
Distância de Descorrelação do Sombreamento	50	m
Ruído no Receptor	-129,1567	dBW

Tabela 3.2: Principais modelos de propagação.

com quadrados cujos vértices possuem valores descorrelacionados seguindo uma distribuição log-normal. O valor de desvanecimento é obtido através da interpolação linear dos vertices pertencentes ao quadrado que circunda o ponto de interesse [35,36].

- ▶ Desvanecimento de pequena escala - um vetor com amostras correlacionadas no tempo através do modelo de Jakes [37] é empregado.
- ▶ Perda de percurso - os modelos utilizados foram baseados em [1,15]. O cálculo da perda de percurso é calculado através da Equação (3.7):

$$P = 128,0 + 37,6 \cdot \log_{10}(d_{km}), \quad (3.7)$$

na qual d_{km} é a distância em quilômetros.

- ▶ Perfil de antena setorizada - foi considerado o mesmo perfil de ganho horizontal adotado em [1].

O cálculo do desvanecimento de larga escala, perda de percurso e ganho de antena é efetuado para todos os pontos da área do grid com uma resolução de 10 m e armazenado em mapas de propagação.

3.2.3 Mobilidade

Os perfis de mobilidade descrevem como os usuários se movem sobre a área do sistema. De acordo com o cenário estudado, os usuários podem ser estáticos, e.g., quando se utiliza um computador, ou móveis. Com relação ao último grupo de usuários, eles se movem de acordo com modelos apresentados em [1].

A movimentação dos usuários segue um modelo de mobilidade aleatório [38,39]. A velocidade é distribuída uniformemente entre 1 e 5 km/h, resultando em uma

média de 3 km/h. A direção também é distribuída uniformemente no intervalo de 0 a 2π radianos. A velocidade e a direção de cada móvel são atualizados a cada 5 m de distância percorrida.

3.2.4 Tráfego

A modelagem da camada de aplicação é efetuada ao se considerar diferentes modelos de tráfego para cada classe de serviço: interativa (navegação *Web*) e conversacional (VoIP).

Navegação *Web*

A aplicação *Web* considerada neste estudo segue a modelagem existente em [1]. Ela é basicamente modelada como uma seqüência de conjuntos de pacotes (*packet calls*), que representam uma seqüência, em rajadas, de pacotes. Os *packet calls* são separados por tempos de leituras.

Cada sessão apresenta como propriedades um número médio de conjuntos de pacotes, tempos de leitura, número de pacotes por conjunto e o tempo médio entre pacotes, todos com distribuição de probabilidade geométrica. Os principais parâmetros de configuração do tráfego *Web* são mostrados na Tabela 3.3.

Parâmetro	Valor	Unidade
Tempo médio de leitura	10,0	s
Tamanho máximo de pacote	66666,0	kbytes
Tamanho médio de pacote	81,5	bytes
Desvio padrão do tamanho de pacote	1,1	bytes
Número médio de <i>packet calls</i>	5	-
Número médio de pacotes em um <i>packet calls</i>	25	-
Tempo médio entre pacotes	0,00195	s

Tabela 3.3: Parâmetros do tráfego *Web*.

VoIP

O padrão de tráfego VoIP é baseado no modelo de dois estados [1], com um tempo médio de sessão de 60 s. A transmissão descontínua (DTX, do inglês *Discontinuous Transmission*) é modelada com 50% de atividade.

O CODEC de voz AMR é mandatário para serviços de voz nos sistemas *Global System for Mobile communication* (GSM) e WCDMA. O 3GPP também o escolheu para ser utilizado para o serviço VoIP [40]. Durante os períodos de atividade, com

o AMR 12.2 kbps, a aplicação VoIP gera 32 bytes de carga útil de voz em intervalos de 20 ms. Na Tabela 3.4 é apresentada a configuração resumida do tráfego de VoIP.

Parâmetro	Valor	Unidade
Taxa do Codec AMR	12,2	kbps
Atividade de voz	50	%
Compressão de cabeçalho	sim	-
Tamanho do pacote (cabeçalhos incluídos)	304	bits
Tamanho do cabeçalho (média assumida)	48	bits
Duração média da chamada	60	s

Tabela 3.4: Parâmetros do tráfego de voz.

Em geral, o serviço VoIP se utiliza três camadas de protocolos: *Real-Time Transport Protocol* (RTP) [41], *User Datagram Protocol* (UDP) [42] e IP [43], criando um massivo *overhead* de 40 bytes, para o caso do *Internet Protocol Version 4* (IPv4). Isto é inaceitável para um ambiente sem-fio, ao se comparar o *overhead* com a carga útil de voz. Para se resolver este problema, técnicas de compressão de cabeçalho são empregadas. Como exemplo, pode-se citar *Robust Header Compression* (ROHC) [8], que pode reduzir o cabeçalho RTP/UDP/IP de 40 para 2 bytes em média [44].

3.2.5 Camadas RLC/MAC

No presente estudo, a operação AM da RLC é utilizada pela classe de serviços interativa. Retransmissões da RLC com um tempo de *polling* de 100 ms são modeladas. O serviço conversacional provido sobre o modo de chaveamento de pacotes (VoIP) utiliza a operação UM, devido a sua restrição de atraso [10]. Os tamanhos dos cabeçalhos RLC são assumidos como 16 e 8 bits para o AM e UM, respectivamente.

O componente MAC-hs da camada MAC modelado neste estudo provê HARQ com 6 processos paralelos SAW, 5 como número máximo de tentativas de transmissão e *Chase Combining*; multiplexação de código e mecanismo de descarte de MAC-hs SDU [11].

Todo novo RLC PDU (MAC-hs SDU) que chega no *buffer* do MAC-hs é associado a um temporizador de descarte. A cada TTI, a camada MAC-hs checa seu *buffer* procurando por MAC-hs SDUs cujos temporizadores tenham expirado. Estes são descartados antes de serem concatenados em um MAC-hs PDU, ou seja, antes da transmissão. Isto é feito com o objetivo de não desperdiçar recursos, já que estes pacotes seriam provavelmente descartados pelo *buffer* de *play-out* (o limiar

de atraso da MAC-hs é tipicamente um pouco menor que o limiar do *buffer* de *play-out*). Sempre que um MAC-hs PDU é transmitido, é movido para o gerenciador do HARQ e os temporizadores de descarte são desabilitados. Assim, quando um pacote é transmitido, pode ser perdido apenas se não for corretamente recebido após o máximo número de retransmissões permitidas ou que tenha chegado atrasado demais para o *buffer* de *play-out*.

3.2.6 Tempo de simulação

O simulador funciona em uma base de tempo de *slot* (0,667 ms). O critério de parada das simulações é baseado no mínimo número de sessões finalizadas de cada classe de serviço (VoIP e *Web*). Para o cenário de serviços simples é assumido que 10000 sessões finalizadas é suficiente para garantir uma boa confiabilidade estatística. No caso de serviços mistos, a simulação para quando ambos os serviços atingem pelo menos 5000 sessões finalizadas. Um menor número de amostras é considerado com intuito de reduzir o tempo de simulação.

Uma técnica de supressão de transiente é considerada nas simulações objetivando a diminuição do período de transiente necessário para a convergência do número de usuários no sistema. Para o cenário de serviços simples, o período de *warm-up* de 5000 iterações (3,33 s) é considerado. No caso de serviços mistos, um tempo maior precisa ser considerado, já que a técnica de supressão de transiente é menos confiável quando o tráfego *Web* está presente, pois o modelo de tráfego utilizado [1] não possui tempo médio de sessão. Neste caso, um período de *warm-up* de 200000 iterações (133,33 ms) é considerado.

3.3 Resultados

Os resultados são compostos da avaliação de algoritmos de escalonamento em ambos os cenários de serviços simples, onde apenas VoIP é oferecido ao sistema, e mistos, onde usuários VoIP e *Web* dividem os recursos da rede. Na Tabela 3.5, os principais parâmetros de configuração da operação do sistema são mostrados.

3.3.1 Cenário de serviço simples (VoIP)

Este primeiro cenário consiste exclusivamente de usuários VoIP. Neste cenário, o desempenho dos algoritmos de escalonamento é analisado com foco em suas habilidades de distribuir os recursos e proporcionar QoS entre os usuários VoIP.

Primeiramente, os algoritmos são avaliados por meio da taxa de satisfação e considerando três diferentes requerimentos de atraso na camada de aplicação: 150,

Parâmetro	Valor	Unidade
Processos HARQ paralelos	6	-
Máximo número de retransmissões na MAC	5	-
Combinação HARQ	<i>Chase Combining</i>	-
Max. # de multiplexação de código	4	-
# de códigos para HSDPA	5	-
Máxima potência de transmissão da base	20	W
Potência dos canais comuns	1,0	W
Potência do CPICH	2,0	W
Potência de cada HS-SCCH	0,5	W
Limite de QoS	90	%

Tabela 3.5: Principais parâmetros de simulação.

130 e 110 ms. É importante mencionar que o limiar de descarte na MAC-hs também é ajustado em cada uma dessas configurações. Ele é assumido como sempre 10 ms menos que o requerimento de atraso, i.e. 140, 120 e 100 ms, respectivamente. Isto se deve ao fato de, quando MAC-hs SDU é concatenado em um MAC-hs PDU, ele não pode mais ser descartado [22]. Esta diferença tenta dar 10 ms ao sistema para finalizar a transmissão do pacote antes que o limite de atraso na aplicação seja atingido.

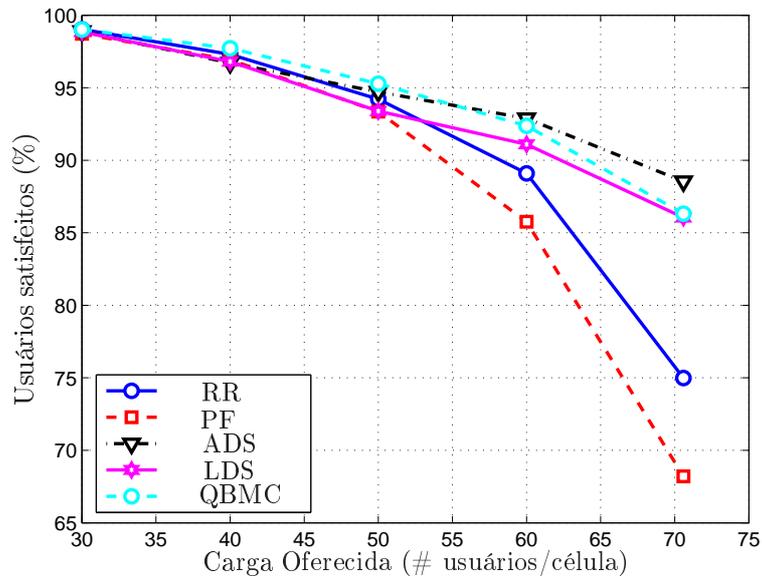


Figura 3.3: Usuários VoIP satisfeitos com um requerimento de atraso de 150 ms.

Na Figura 3.3 o percentual de usuários satisfeitos é apresentado e os algoritmos de

escalonamento podem ser comparados considerando-se um requerimento de atraso de 150 ms. Pode-se perceber que o escalonador PF proporciona o pior desempenho. Como ele apenas se utiliza de informação de taxa para ordenar os usuários em prioridade, ele não pode prover garantias de QoS para VoIP. O RR consegue prover melhores resultados em relação ao PF devido a sua propriedade de distribuição igual dos *slots* de transmissão (TTIs) entre todos os usuários na fila.

Os algoritmos não-clássicos avaliados proporcionaram resultados bastante similares entre eles, e um melhor desempenho que o RR, principalmente sob cargas além do limite de capacidade (90% de satisfação). Eles são mais capazes de prover garantias de QoS em situações de sobrecarga do que os algoritmos clássicos. O ADS proporciona o melhor desempenho seguido pelo QBMC. Para a carga mais alta avaliada, o LDS se aproxima do QBMC.

No caso de ambos ADS e LDS, o ganho de desempenho vem do fator de priorização (FB) estabelecido pelo valor do atraso do pacote mais antigo. A medida que o valor do atraso se aproxima do limiar de descarte, a prioridade do usuário aumenta. O ADS leva uma vantagem por somente considerar esse aumento de prioridade quando o pacote está bem próximo do descarte, através de um crescimento assintótico da FB. Já o LDS aumenta a prioridade linearmente com o aumento do atraso. Isto provoca uma maior influência do fator referente ao atraso no escalonamento, tornando o ganho de diversidade multi-usuário devido ao canal menos eficiente.

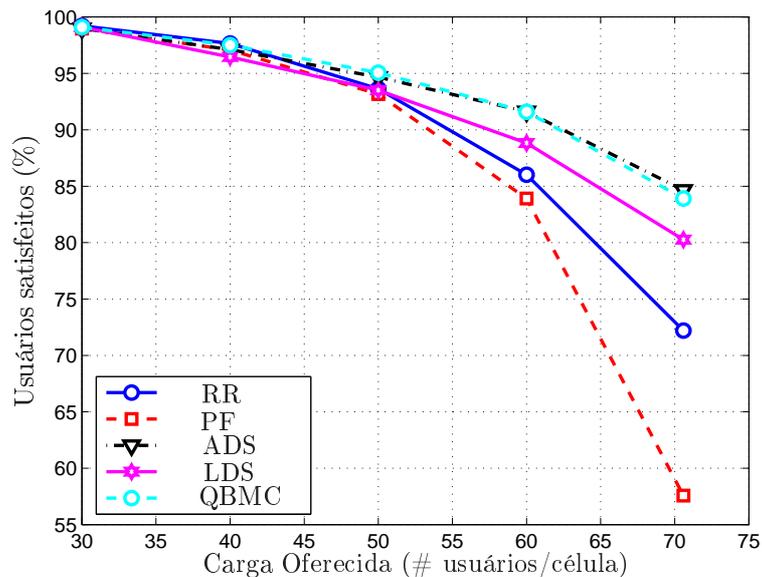


Figura 3.4: Usuários VoIP satisfeitos com um requerimento de atraso de 130 ms.

O requerimento de atraso de 130 ms é avaliado na Figura 3.4. Comparando-se a ambos ADS e QBMC, o desempenho do LDS é pior. Comparando com o caso em quem o requerimento de atraso é configurado para 150 ms, a diferença de desempenho entre o ADS e o QBMC é menor. É também fácil de se perceber que o comportamento relativo entre o RR e o PF não muda significativamente.

Mas quando o requerimento de atraso é ainda menor, como mostra a Figura 3.5, o QBMC se torna o melhor algoritmo para a carga mais alta simulada. Fica então mais evidente que ele proporciona uma melhor relação custo-benefício entre utilização eficiente do canal, pela componente Max CIR, e justiça pelo fator multiplicativo do tamanho da fila. Isto significa que, para este cenário, o QBMC proporciona um melhor desempenho relativo em períodos de sobrecarga, quando o serviço VoIP a ser oferecido tem um requerimento de atraso mais exigente.

Comparando-se as Figuras 3.3 e 3.5 para a carga mais alta simulada, o LDS tem cerca de 5% de aumento na satisfação comparado ao RR assumindo um requerimento de 110 ms para o atraso, enquanto que mais de 10% é obtido ao se considerar um maior valor de tolerância ao atraso. Como o sistema opera mais freqüentemente com atrasos de pacote mais próximo do valor de limiar, o escalonador força mais comumente um usuário a ser escalonado sob condições não ideais de canal.

Como esperado, o desempenho geral do sistema sofreu uma degradação a medida que o valor do requerimento de atraso é diminuído.

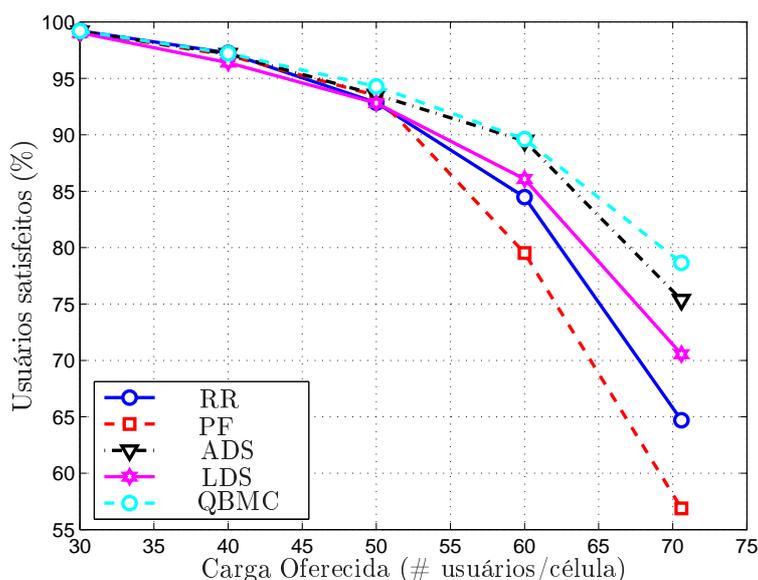


Figura 3.5: Usuários VoIP satisfeitos com um requerimento de atraso de 110 ms.

Algumas métricas de desempenho podem também ser avaliadas objetivando uma explanação mais detalhada das características relacionadas a cada algoritmo. O 98º percentil da taxa de perda de pacotes devido a erros no canal para cada requerimento de atraso (150, 130 e 110 ms) é apresentado nas Figuras 3.6(a), 3.7(a) e 3.8(a). Esta perda de pacote ocorre no HSDPA quando o máximo número de tentativas de transmissão é atingido sem sucesso. Isto mostra como cada algoritmo é capaz de dar acesso a transmissões usuários que experimentam melhores condições de canal. As figuras mostram que todos os algoritmos proporcionam taxas aproximadamente entre 1,0% e 2,5% para a carga mais alta simulada.

Um comportamento que, a princípio, pode ser considerado estranho é que, a partir de uma certa carga, os algoritmos RR, PF e QBMC deixam de apresentar um acréscimo nessa taxa de perda de pacotes. Isto pode ser facilmente explicado pelo fato de estes algoritmos apresentarem taxas de descarte na MAC-hs bastante altas para estes valores de carga, o que acaba por diminuir a fração de pacotes perdidos devido ao canal. Os resultados de descarte na MAC-hs serão apresentados logo em seguida.

Algumas vezes o 98º percentil pode não ser totalmente claro sobre o desempenho de determinada métrica. Por este motivo, nas Figuras 3.6(b), 3.7(b) e 3.8(b), a *Cummulative Distribution Function* (CDF) desta taxa de perda de pacotes com relação à carga mais alta avaliada é mostrada com o intuito de se tecer uma conclusão mais confiável.

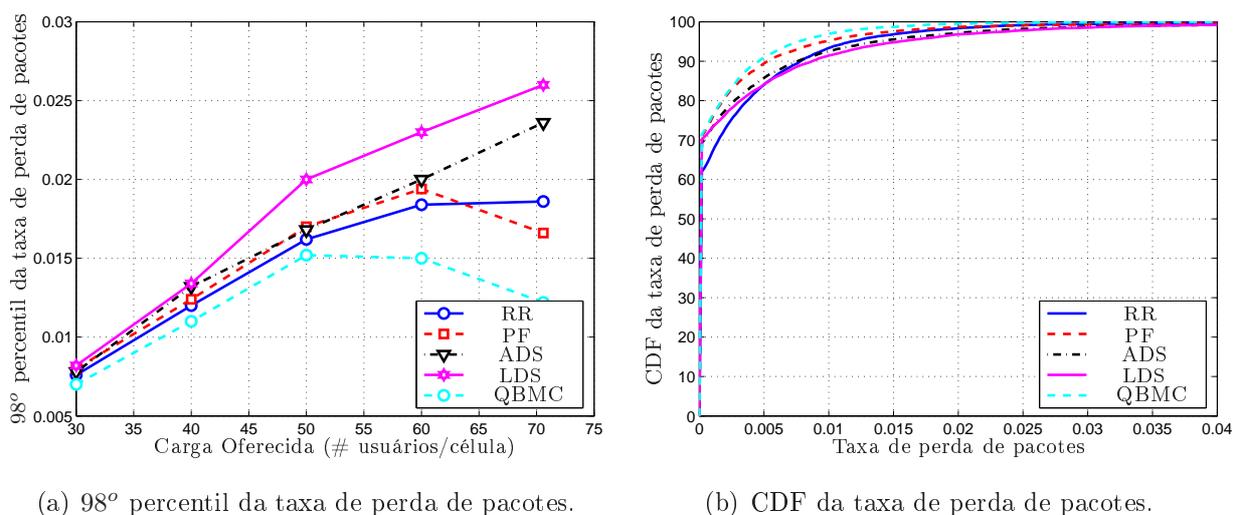


Figura 3.6: Taxa de perda de pacotes devido a erros no canal para um requerimento de atraso de 150 ms.

Apesar de o RR proporcionar um valor de 98^o percentil melhor, a CDF mostra algo bem diferente. Focando na taxa de 1% da perda de pacotes, os melhores algoritmos são o PF e o QBMC. O RR tem desempenho próximo aos ADS e LDS. Isto indica que os primeiros algoritmos mencionados são mais capazes de dar acesso ao canal a usuários com melhores condições de canal e ambos os ADS e LDS dão maior importância ao atributo de atraso durante situações de sobrecarga, tornando a condição de canal menos relevante para o processo de escalonamento.

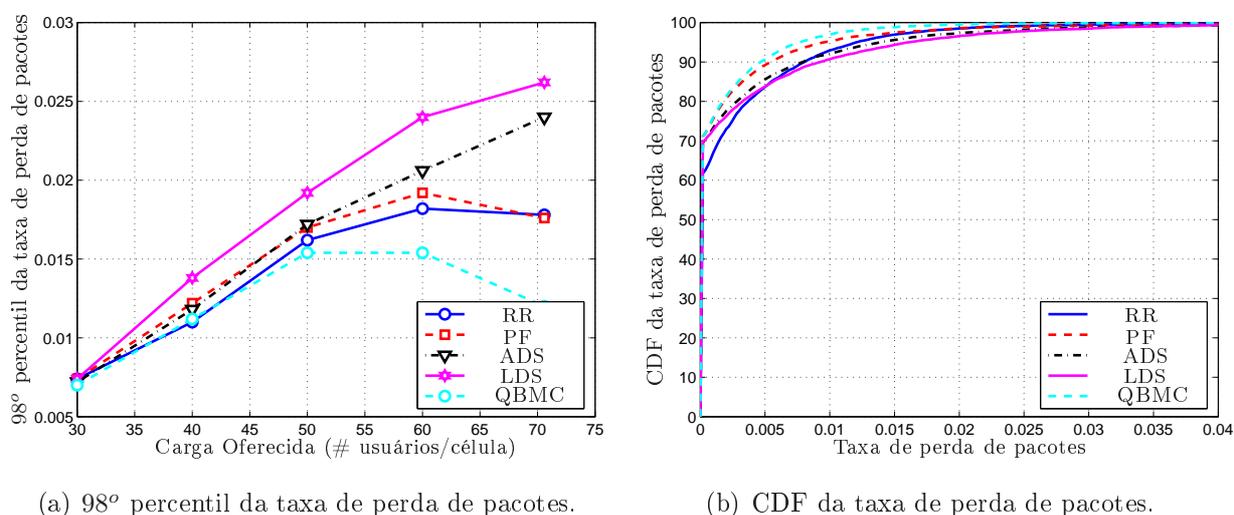
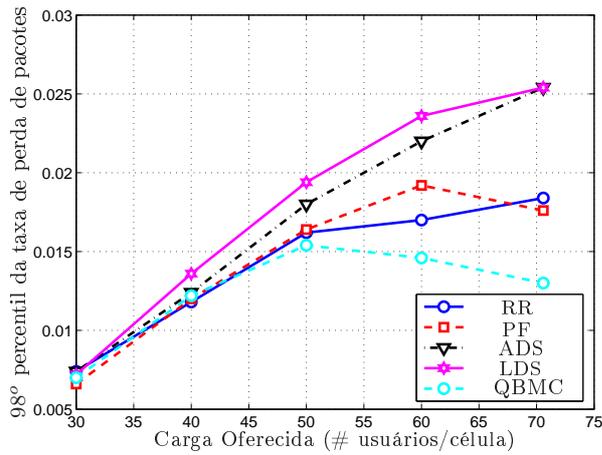
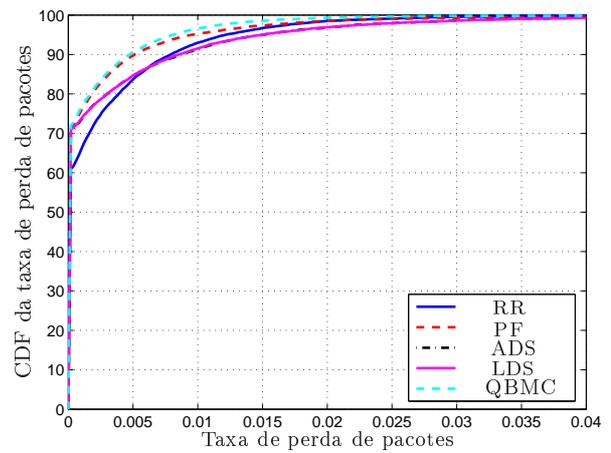


Figura 3.7: Taxa de perda de pacotes devido a erros no canal para um requerimento de atraso de 130 ms.

Quando o requerimento de atraso é decrescido para 130 e 110 ms, como pode ser visto nas Figuras 3.7(a) e 3.8(a), respectivamente, os resultados obtidos são bastante similares. A única diferença interessante é que a taxa de pacotes perdidos do algoritmo ADS é mais próxima do LDS para a carga mais alta simulada em ambos 98^o percentil e CDF. Isto indica que existem mais usuários com pacotes na iminência de serem descartados na MAC quando o requerimento de atraso é mais severo, levando a um número maior de usuários serem escalonados ineficientemente do ponto de vista da qualidade do canal.

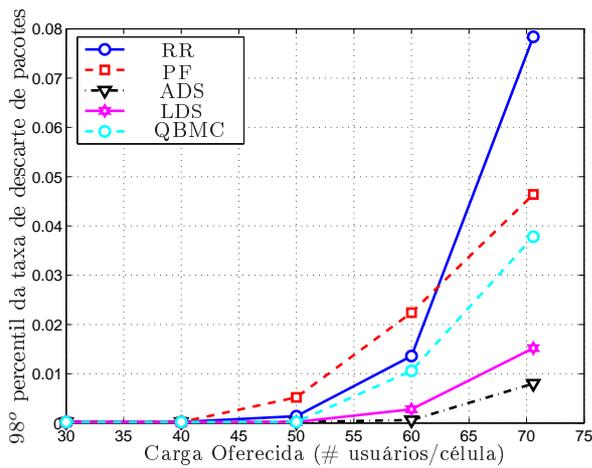


(a) 98º percentil da taxa de perda de pacotes.

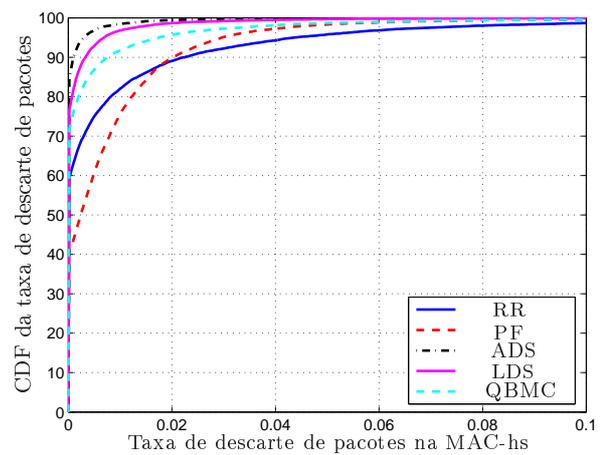


(b) CDF da taxa de perda de pacotes.

Figura 3.8: Taxa de perda de pacotes devido a erros no canal para um requerimento de atraso de 110 ms.



(a) 98º percentil da taxa de descarte na MAC-hs.



(b) CDF da taxa de descarte na MAC-hs.

Figura 3.9: Taxa de descarte de pacotes na MAC-hs para um requerimento de atraso de 150 ms.

É importante ressaltar que o desempenho dos algoritmos RR, PF e QBMC na taxa de pacotes perdidos permanece aproximadamente o constante, mesmo quando o requerimento de atraso é modificado. Isto pode ser explicado pelo desconhecimento de qualquer informação referente ao atraso nestes algoritmos.

A taxa de descarte na MAC-hs, para cada requerimento de atraso, é analisada nas Figuras 3.9(a), 3.10(a) e 3.11(a) através do 98º percentil. Pode-se perceber que ambos ADS e LDS proporcionam o melhor desempenho. O primeiro apresenta as

menores taxas de descarte em todos os requerimentos de atraso avaliados. O RR é superado pelo PF apenas na carga mais alta avaliada. Isto indica que os tempos de fila no *buffer* de HARQ ocasionados pela inexistência de conhecimento do canal no processo de escalonamento leva a um maior número de retransmissões necessárias. Isto acarreta em um maior tempo de enfileiramento na MAC-hs devido à ocupação do *buffer* de HARQ, acarretando em um maior atraso na transmissão de novos pacotes.

É interessante notar que, na Figura 3.11(a), o QBMC tem um desempenho semelhante ao PF sob altas cargas. Outra conclusão é que algoritmos que consideram diretamente o requerimento de atraso proporcionam melhores desempenhos com relação à esta métrica comparados a algoritmos que não consideram (PF e RR) ou mesmo a consideram de maneira indireta (QBMC).

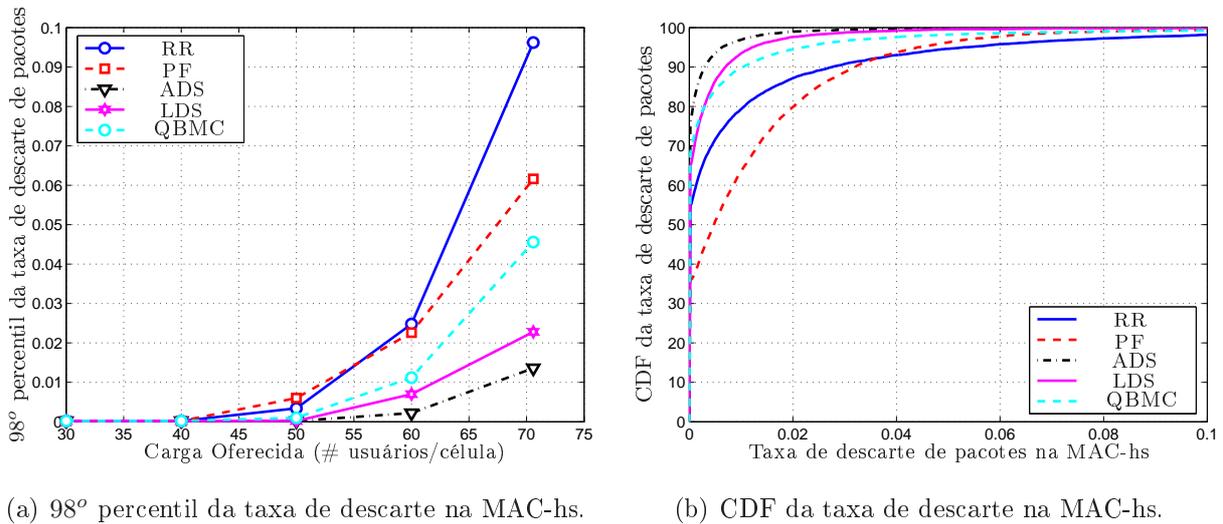


Figura 3.10: Taxa de descarte de pacotes na MAC-hs para um requerimento de atraso de 130 ms.

A CDF da taxa de descarte na MAC-hs também é apresentada pelas Figuras 3.9(b), 3.10(b) e 3.11(b). Pode-se ver que, em todas as figuras, os algoritmos RR e PF são os piores. Apesar de que o QBMC também não está a par do requerimento de QoS de VoIP, seu desempenho é bastante próximo a ambos ADS e LDS, que monitoram a métrica de atraso e o requerimento deste serviço. De fato, o QBMC se aproxima do ADS quando o atraso tolerado é diminuído (supera o LDS). Isto se deve ao fato de ambos ADS e LDS se tornarem menos eficientes em tomar vantagem da diversidade multi-usuário quando eles operam próximo ao limiar de atraso, levando a um maior número de retransmissões necessárias e, conseqüentemente, a uma maior

ocupação no *buffer* de HARQ prolongando o tempo de espera de pacotes para serem concatenados em um bloco de transporte.

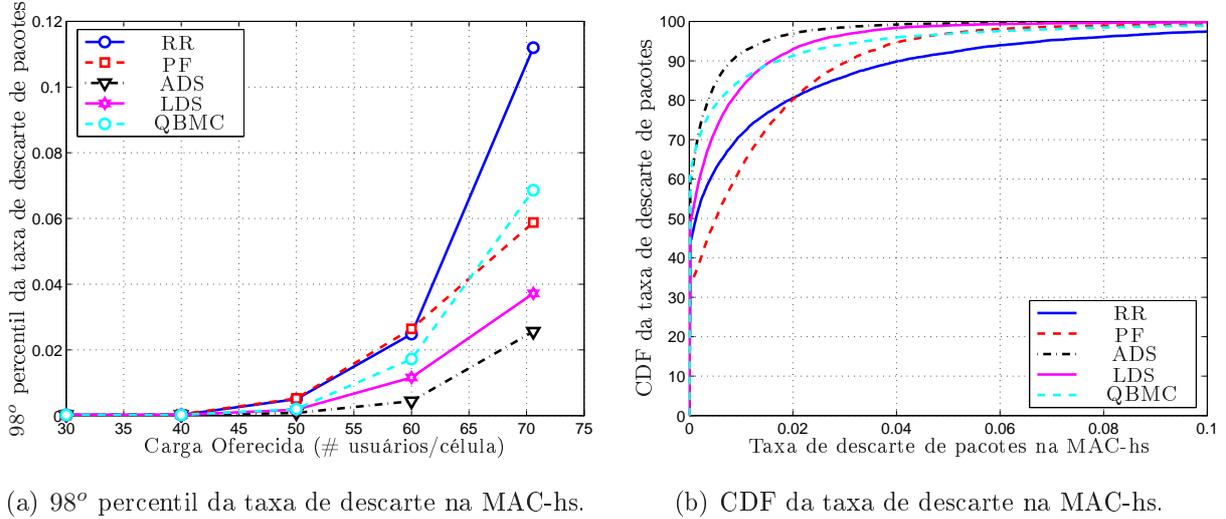


Figura 3.11: Taxa de descarte de pacotes na MAC-hs para um requerimento de atraso de 110 ms.

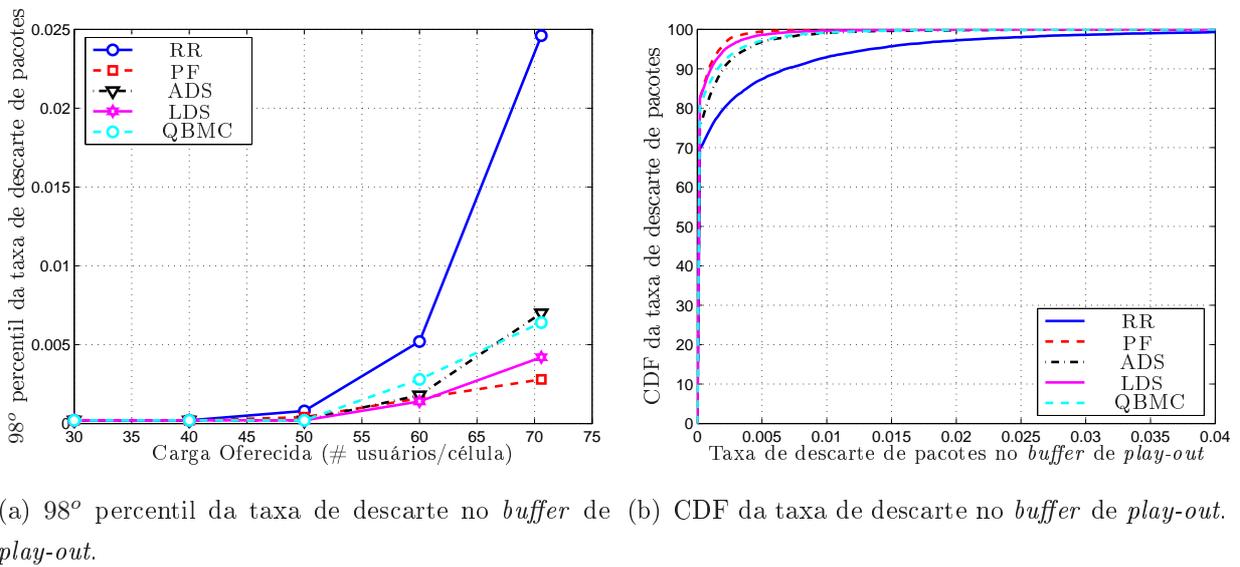


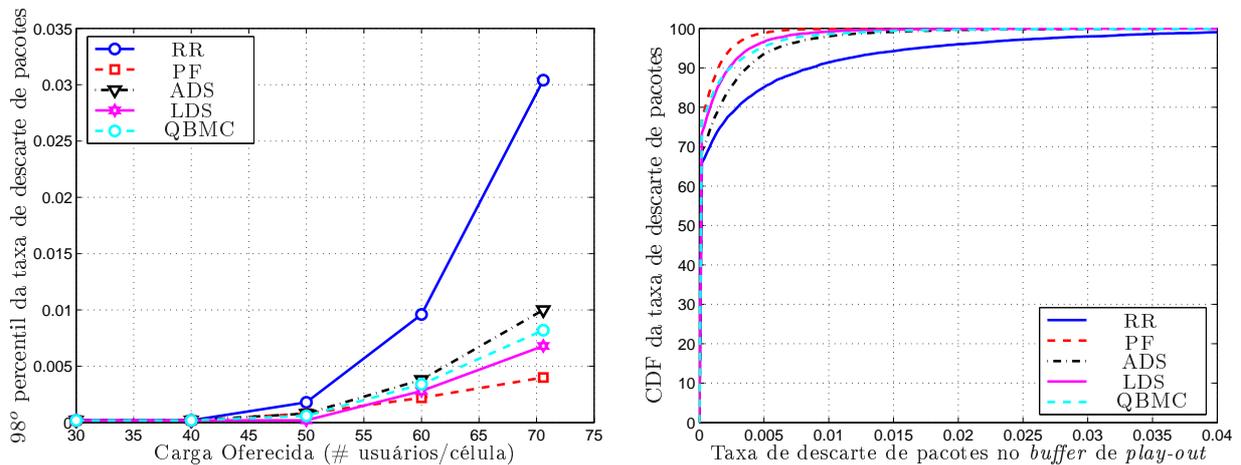
Figura 3.12: Taxa de descarte de pacotes no *buffer* de *play-out* para um requerimento de atraso de 150 ms.

Agora, a taxa de descarte no *buffer* de *play-out* na camada de aplicação é avaliada nas Figuras 3.12(a), 3.13(a) and 3.14(a). O RR proporciona sempre o pior desempenho com uma taxa de descarte bem superior aos outros algoritmos. É interessante perceber que, quando a tolerância de atraso é decrescida, i.e. o

requerimento se torna maior, o ADS se torna pior comparado ao QBMC. O LDS também sofre uma degradação devido à diminuição da tolerância ao atraso.

Um requerimento de atraso mais exigente torna o fator referente ao atraso no cálculo da função de prioridade mais forte que a condição de canal, no caso dos algoritmos ADS e LDS, levando a um maior número de retransmissões devido à utilização menos eficiente das variações no canal. Dessa maneira, o atraso final atinge valores superiores ao mínimo na camada de aplicação. O PF mostra-se como o melhor algoritmo com relação a esta métrica. Isto se deve principalmente ao fato de apenas usuários com razoáveis condições de canal serem escalonados, levando a um menor número de retransmissões para se concretizar o recebimento correto de um pacote.

Vale mencionar que o algoritmo de escalonamento tem a capacidade de controlar a taxa de descarte na MAC-hs, porque ele pode forçar um usuário a ser escalonado, evitando-o de ter pacotes descartados. Por outro lado, se o usuário é forçado a transmitir durante más condições de canal, pode levar um tempo considerável para concretizar a transmissão do pacote devido a um grande número de retransmissões, ou mesmo expirar o número máximo de tentativas, levando à perda do pacote. Um maior montante de retransmissões pode também levar a uma maior taxa de descarte na camada de aplicação (*buffer de play-out*).

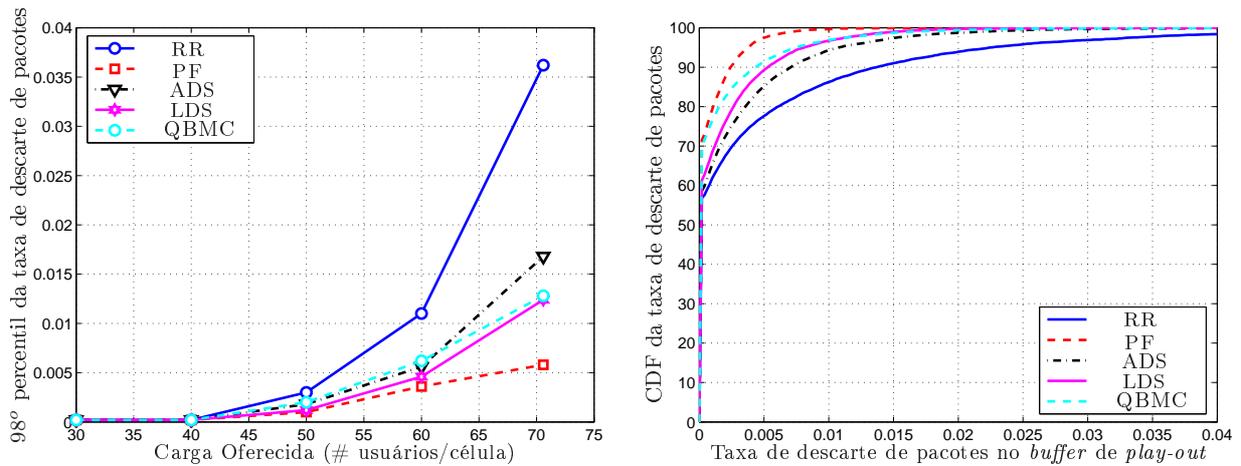


(a) 98^o percentil da taxa de descarte no *buffer de play-out*. (b) CDF da taxa de descarte no *buffer de play-out*.

Figura 3.13: Taxa de descarte de pacotes no *buffer de play-out* para um requerimento de atraso de 130 ms.

A CDF da taxa de descarte no *buffer de play-out* também é mostrada, agora nas

Figuras 3.12(b), 3.13(b) and 3.14(b). A CDF mostra que o desempenho do RR é pior, como esperado de algoritmos que ignoram as condições de canal, resultando em maiores tempos para a concretização da recepção do pacote. Os outros algoritmos proporcionam um desempenho bastante similar. É interessante notar que o PF sofre uma pequena degradação ao se diminuir a tolerância ao atraso, já que ele não força de maneira alguma a transmissão devido ao atraso ou mesmo tamanho de *buffer*, como ADS, LDS e QBMC.



(a) 98^o percentil da taxa de descarte no *buffer* de *play-out*. (b) CDF da taxa de descarte no *buffer* de *play-out*.

Figura 3.14: Taxa de descarte de pacotes no *buffer* de *play-out* para um requerimento de atraso de 110 ms.

Com toda esta análise em um cenário simples composto por apenas usuários do serviço VoIP, pode-se obter uma caracterização do comportamento deste serviço frente aos algoritmos utilizados. A próxima seção tem como objetivo a avaliação do desempenho destes algoritmos ao se injetar um tráfego composto por dois serviços bastante antagônicos como VoIP e *Web*.

3.3.2 Cenário misto de serviços (VoIP e *Web*)

Esta seção consiste na avaliação do cenário onde usuários VoIP e *Web* compartilham dos recursos de rádio disponíveis. Os resultados apresentados se resumem à satisfação em ambos os serviços, onde o desempenho dos algoritmos de escalonamento é comparado.

Os resultados são separados referentes à proporção de cada serviço. A primeira parte das avaliações se preocupa com o cenário onde existe uma dominância de

usuários VoIP em número de usuários (75% VoIP / 25% *Web*). Os outros resultados consistem em ambos os serviços em iguais proporções (50% VoIP / 50% *Web*) e a predominância do serviço *Web* (25% VoIP / 75% *Web*). O último conjunto de avaliações mostra as regiões de capacidade conjunta para todos os algoritmos avaliados.

75% VoIP + 25% *Web*

Esta seção apresenta os resultados no cenário misto de serviços em que o serviço VoIP tem uma maior proporção na carga oferecida.

As Figuras 3.15(a) e 3.15(b) apresentam o desempenho da rede para usuários VoIP e *Web*, respectivamente, assumindo um requerimento de atraso de 150 ms. O desempenho do serviço VoIP na Figura 3.15(a) mostra uma degradação considerável no QBMC comparado ao cenário com um único serviço (Figura 3.3). Isto se deve principalmente aos maiores tamanhos de *buffer* apresentados por usuários *Web*, resultando em uma maior prioridade a eles de acordo com a formulação do algoritmo na Seção 3.1. Ambos ADS e LDS apresentam comportamento similar, ou seja, com desempenho bastante próximo um do outro e superior aos demais. O RR e WPF apresentam um desempenho levemente pior que estes últimos mencionados, mas bem superior ao clássico PF (semelhante ao cenário de serviço simples).

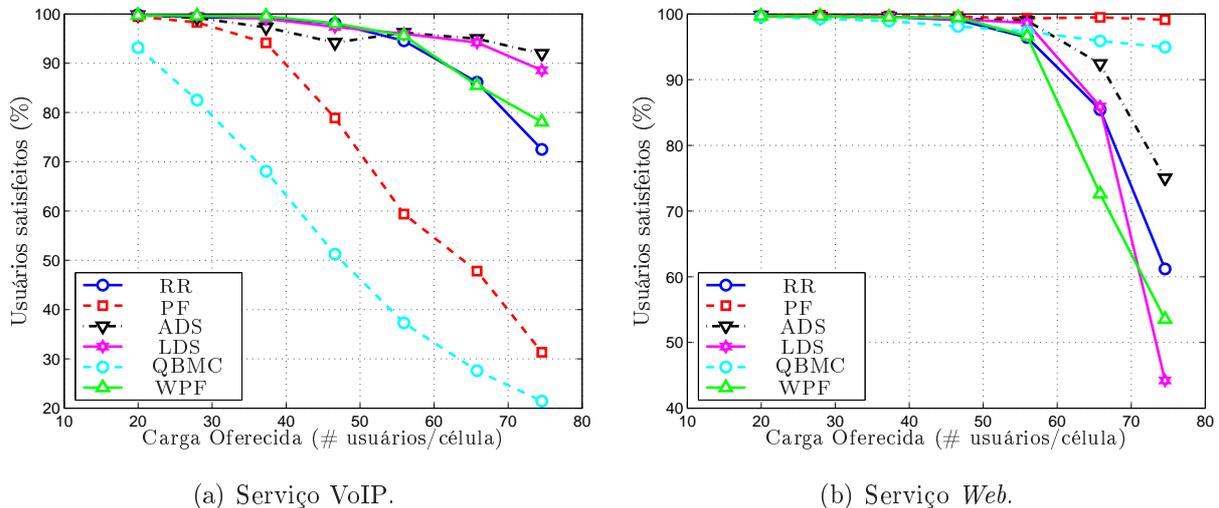


Figura 3.15: Percentual de usuários satisfeitos considerando um requerimento de atraso de 150 ms.

Com relação ao serviço *Web*, a Figura 3.15(b) apresenta o PF como o melhor algoritmo seguido do QBMC. O maior peso dado ao serviço VoIP considerado pelo

WPF resulta em um desempenho pior para usuários *Web*. O ADS proporciona resultados melhores que ambos LDS (menor interferência para valores pequenos de atraso) e RR. Também apresenta um desempenho levemente pior comparado a ambos PF e QBMC.

É interessante notar que, em resumo, os algoritmos que proporcionam o melhor desempenho para um serviço obtém os piores resultados no outro, devido à priorização de serviço (ADS, LDS, WPF e QBMC) ou mesmo à distribuição de recursos mais apropriada a determinado serviço (RR e PF).

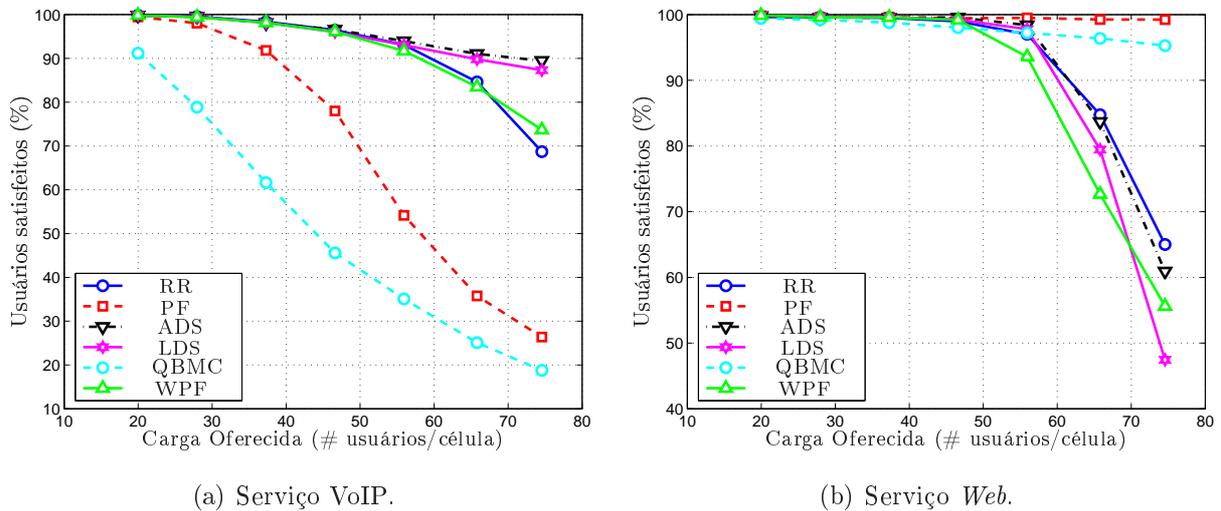


Figura 3.16: Percentual de usuários satisfeitos considerando um requerimento de atraso de 130 ms.

As Figuras 3.16(a) e 3.16(b) apresentam o desempenho do sistema com a adoção de 130 ms como requerimento de atraso do serviço VoIP. Fica claro que o desempenho relativo dos algoritmos com relação ao serviço VoIP (Figure 3.16(a)) é bastante semelhante ao obtido na Figura 3.15(a) (requerimento VoIP de 150 ms). Os resultados para o caso 130 ms apresentam uma degradação geral devido ao maior requerimento.

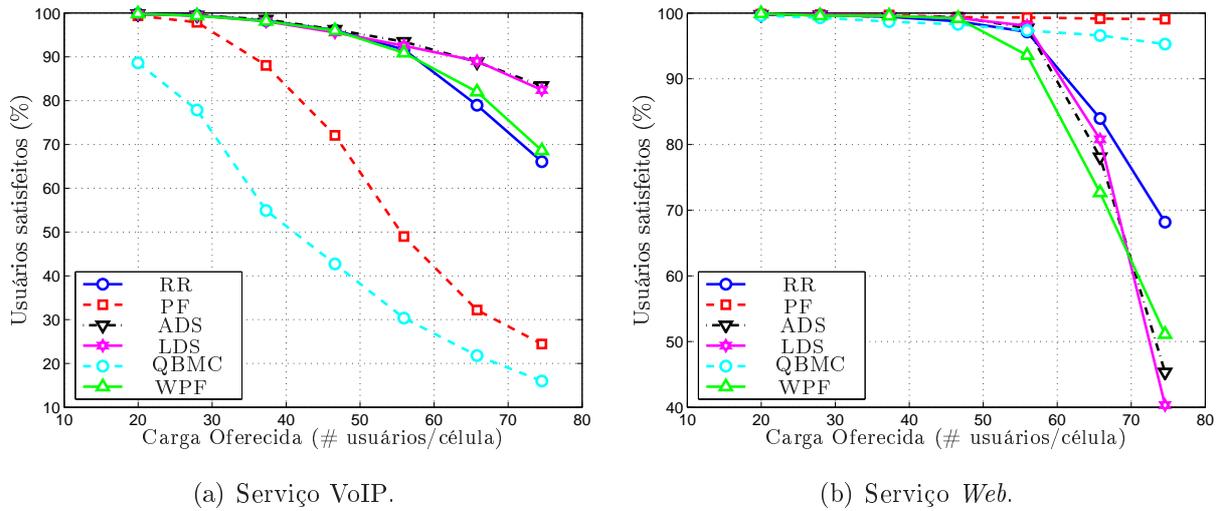


Figura 3.17: Percentual de usuários satisfeitos considerando um requerimento de atraso de 110 ms.

Com relação ao serviço *Web*, na Figura 3.16(b), o desempenho do ADS e do LDS é consideravelmente degradado pelo decréscimo da tolerância ao atraso. O primeiro apresenta um desempenho semelhante ao RR. Ao se dar maior prioridade a VoIP, o serviço *Web* acaba sendo degradado, já que uma menor tolerância ao atraso impõe uma restrição ainda maior ao QoS de VoIP, que se torna um serviço ainda mais sensível.

O desempenho dos algoritmos RR, PF, QBMC and WPF são aproximadamente os mesmos aos obtidos na Figura 3.15(b).

O efeito de uma nova redução na tolerância ao atraso é apresentado nas Figuras 3.17(a) and 3.17(b). O algoritmo RR apresenta uma leve degradação na satisfação quando usuários VoIP são analisados e comparando-se aos algoritmos ADS, LDS e WPF. Por outro lado, o desempenho do ADS entre usuários *Web* (Figura 3.17(b)) sofre uma degradação ainda maior. Sob cargas mais altas, é superado pelo RR.

Para o serviço *Web*, com excessão ao ADS e LDS, o desempenho dos algoritmos permanece aproximadamente constante com o decréscimo da tolerância ao atraso. Isto é causado pelo impacto direto dessa mudança de requerimento no cálculo de prioridades dos primeiros algoritmos citados.

50% VoIP + 50% *Web*

Agora, o sistema é avaliado no cenário caracterizado pela mesma proporção entre usuários VoIP e *Web*.

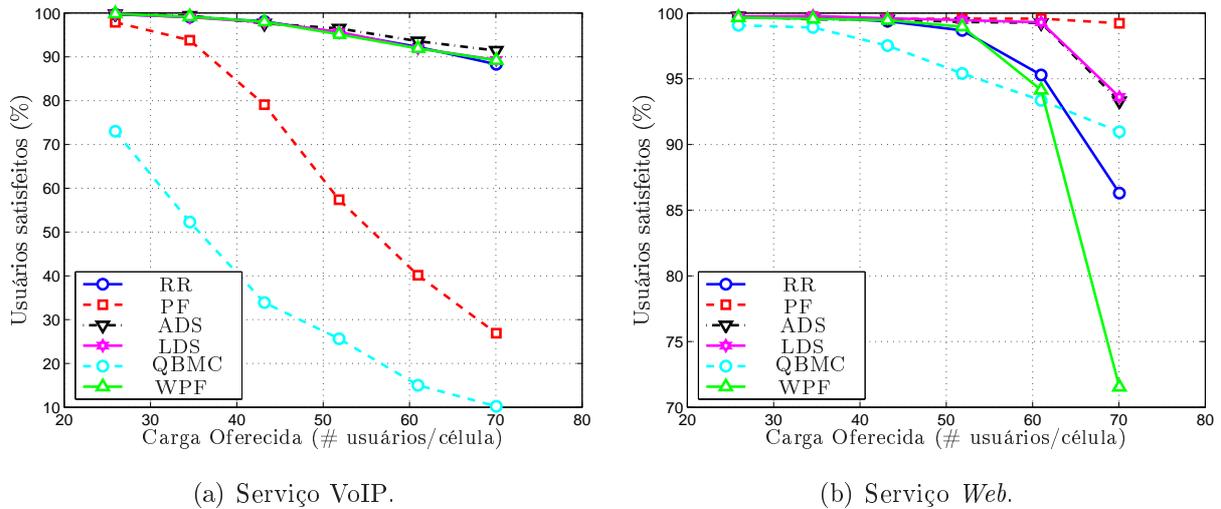


Figura 3.18: Percentual de usuários satisfeitos considerando um requerimento de atraso de 150 ms.

O desempenho de usuários VoIP considerando uma tolerância ao atraso de 150 ms é apresentado na Figura 3.18(a). Fica claro que o algoritmo QBMC tem a pior qualidade, seguido pelo PF. O comportamento relativo entre ambos os algoritmos permanece o mesmo comparando-se ao cenário anterior (Figura 3.15(a)) mas em números absolutos é pior. Por outro lado, o RR e o WPF melhoraram. O primeiro deve-se ao proveito que os usuários VoIP tiram da natureza mais incostante do tráfego *Web*. Já o WPF tira proveito da diminuição da proporção de usuários de alta prioridade, ou seja, menor concorrência por recursos. Ambos ADS e LDS obtiveram os mesmos bons resultados.

A taxa de satisfação com relação a usuários *Web* é apresentada na Figura 3.18(b). A superioridade do PF não fica tão evidente à obtida com uma menor proporção do serviço *Web*. Ambos ADS e LDS proporcionam um desempenho bastante semelhante. O RR e o WPF proporcionam desempenhos bastante similares, com uma vantagem (cerca de 10% a mais em satisfação para altas cargas) para o RR em situações de sobrecarga.

Fica claro que, com excessão ao QBMC que é pior e o PF que permanece o mesmo, existe uma significativa melhora em todos os algoritmos no desempenho do serviço *Web* comparado ao cenário anterior (Figura 3.15(b)). Isto pode ser explicado pelo fato de estes algoritmos proporcionarem uma priorização de serviço para o VoIP. Como existe um menor número de usuários VoIP, existem mais recursos sobrando para o serviço *Web*. A razão pela qual o QBMC fica pior vem do fato de que este

proporcionar uma maior prioridade a usuários *Web* por causa dos maiores tamanhos de *buffer* experimentados por estes usuários. Dessa maneira, a concorrência entre usuários de maior prioridade no sistema aumenta.

Quando a tolerância ao atraso é diminuída a 130 ms, o desempenho de usuários VoIP permanece praticamente constante, como pode ser visto na Figura 3.19(a), comparando-se à Figura 3.18(a). Apenas o PF e o QBMC sofrem uma certa degradação no desempenho.

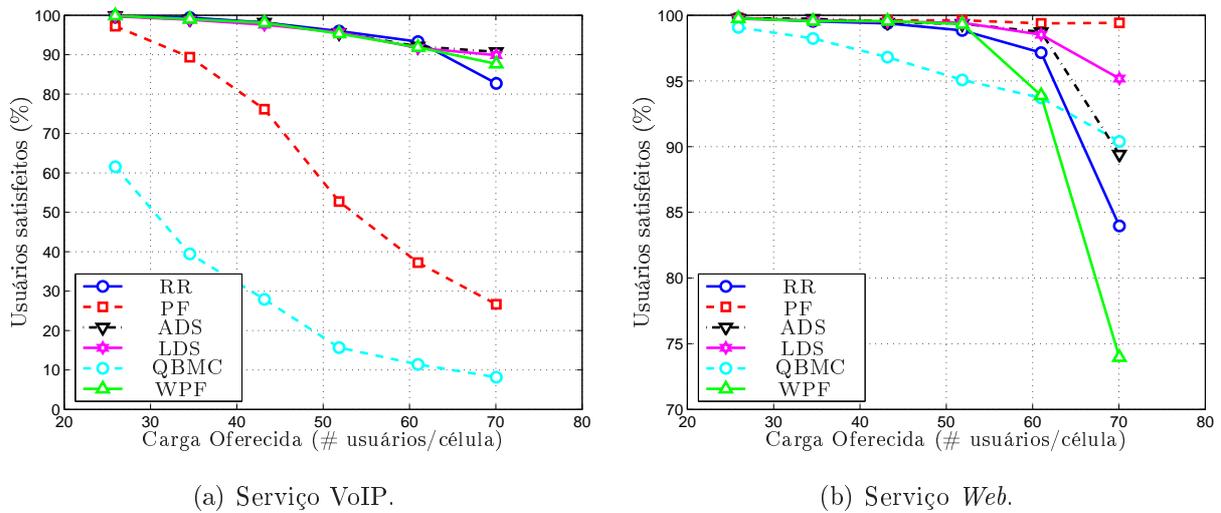


Figura 3.19: Percentual de usuários satisfeitos considerando um requerimento de atraso de 130 ms.

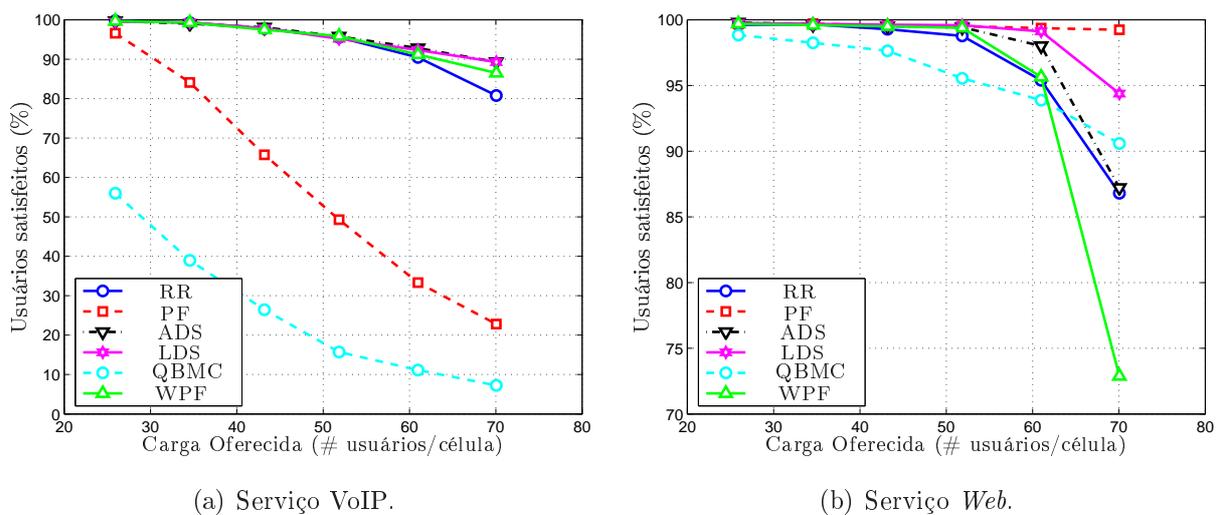


Figura 3.20: Percentual de usuários satisfeitos considerando um requerimento de atraso de 110 ms.

O desempenho do ADS para usuários *Web*, apresentado na Figura 3.19(b), sofre uma leve degradação comparado a ambos PF e LDS.

Quando o requerimento de atraso é configurado para 110 ms, o desempenho relativo de todos os algoritmos com relação a ambos os serviços não se diferencia significativamente, como pode ser visto nas Figuras 3.20(a) e 3.20(b).

A relação de desempenho entre ambos os algoritmos ADS e LDS é invertida. Para valores de tolerância ao atraso maiores o ADS é melhor, mas quando o requerimento de atraso se torna mais restrito, o LDS supera o ADS. Isto pode ser explicado pelo fato de que a FB do LDS aumenta linearmente a média que o atraso de pacote cresce, proporcionando um aumento gradual na prioridade e evitando o limiar de atraso de ser atingido, enquanto que o ADS aumenta a prioridade drasticamente quando o atraso atinge valores próximos ao limiar.

25% VoIP + 75% *Web*

Este último cenário de serviços mistos consiste em uma dominância de usuários *Web* na carga oferecida.

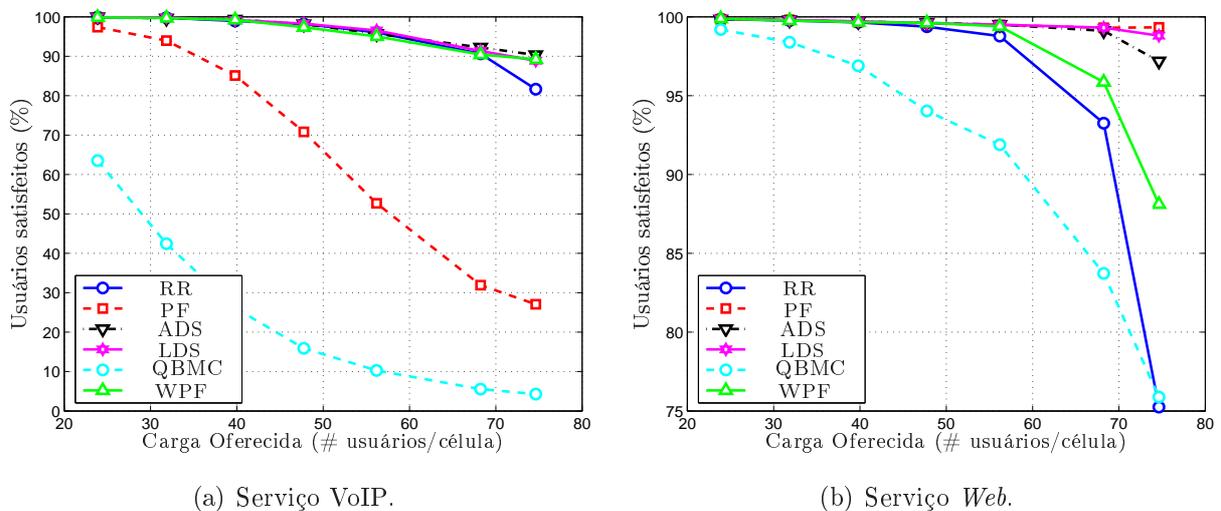
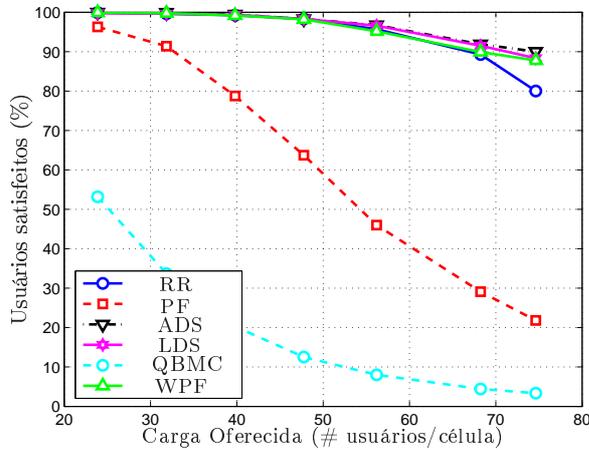


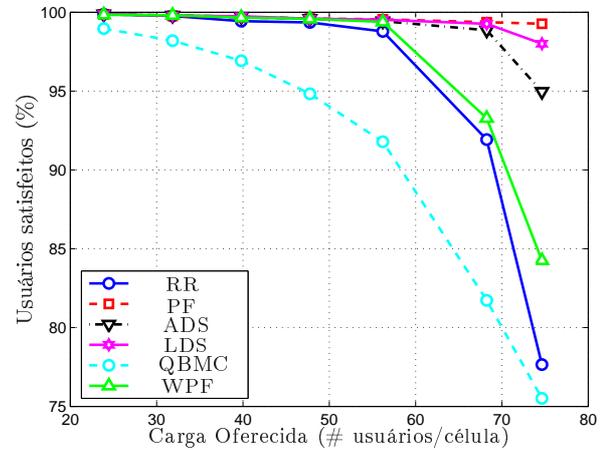
Figura 3.21: Percentual de usuários satisfeitos considerando um requerimento de atraso de 150 ms.

O desempenho dos usuários VoIP é mostrado na Figura 3.21(a). Existe um comportamento semelhante comparando-se ao cenário com um número igual de usuários VoIP e *Web* (Figura 3.18(a)). O algoritmos ADS, LDS e WPF proporcionam desempenho bastante similar, enquanto que o QBMC é bem fraco em garantir um bom QoS. O RR é bem melhor que o QBMC mas ainda bastante

inferior aos demais algoritmos. Ambos apresentam desempenho pior comparando-se ao cenário anterior.

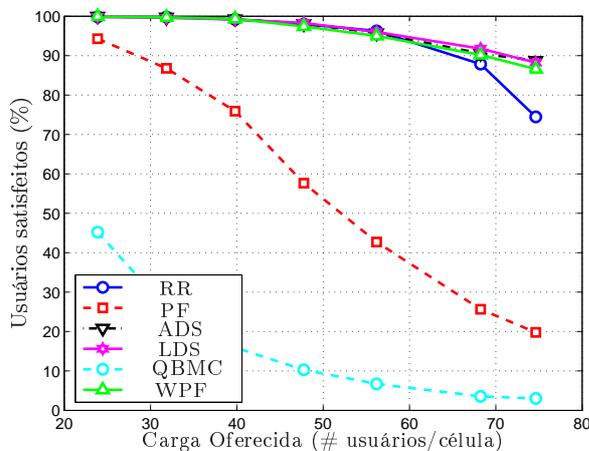


(a) Serviço VoIP.

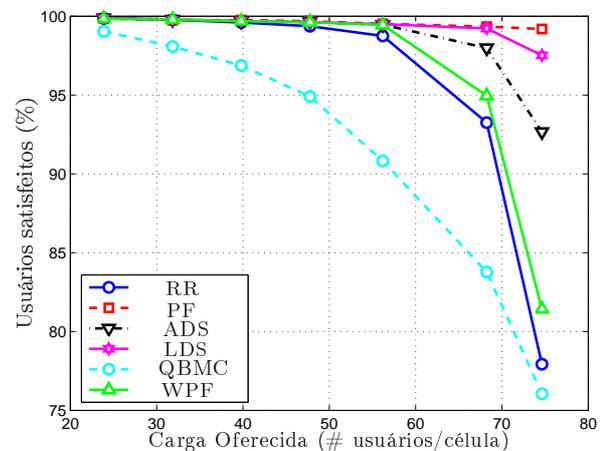


(b) Serviço Web.

Figura 3.22: Percentual de usuários satisfeitos considerando um requerimento de atraso de 130 ms.



(a) Serviço VoIP.



(b) Serviço Web.

Figura 3.23: Percentual de usuários satisfeitos considerando um requerimento de atraso de 110 ms.

Um fato interessante com relação a este cenário relaciona-se ao escalonador WPF. Ele é capaz de prover um desempenho razoável para ambos os serviços. Isto pode ser explicado pelo fato de, como existe um menor número de usuários com maior prioridade no sistema, o escalonador é capaz de efetuar uma reserva de recursos e, ao mesmo tempo, proporcionar um bom QoS ao outro serviço, no caso, *Web*.

O desempenho do WPF para usuários *Web* supera agora o RR, como pode ser verificado na Figura 3.21(b).

Um desempenho bastante semelhante para o serviço VoIP é obtido ao se diminuir a tolerância ao atraso para 130 ms, na Figura 3.22(a). A diferença surge na Figura 3.22(b), onde, para altas cargas, o ADS sofre uma degradação comparada a ambos PF e LDS. Também com relação às mesmas cargas, existe uma vantagem do WPF sobre o RR. Nenhuma alteração significativa é constatada ao se analisar o desempenho do sistema considerando um requerimento de atraso de 110 ms na aplicação, como pode ser visto nas Figuras 3.23(a) e 3.23(b).

3.3.3 Regiões de capacidade conjunta

As regiões de capacidade podem ser definidas como o número de usuários esperados no sistema em que uma qualidade aceitável é suportada por todos os tipos de serviços [45].

Esta seção apresenta algumas figuras de capacidade representativas de todos os algoritmos de escalonamento e assumindo os três requerimentos de atraso estudados. A capacidade é baseada nos limites de QoS com relação a todas as proporções entre os serviços, com o intuito de uma definição mais detalhada de como cada algoritmo é melhor em cada configuração de proporção de carga entre os serviços.

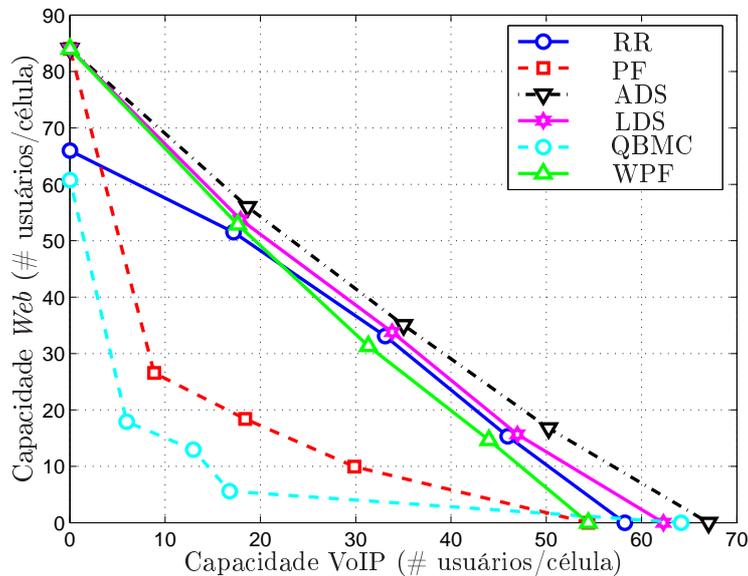


Figura 3.24: Regiões de capacidade considerando um requerimento de atraso para VoIP de 150 ms.

A Figura 3.24 apresenta a capacidade mista considerando a tolerância ao atraso

do VoIP de 150 ms. Pode-se perceber que o ADS proporciona a maior capacidade devido a sua capacidade de aliar o desempenho em termos de vazão na rede e justiça para ambos os serviços VoIP e *Web*. A vantagem é maior a medida que se aumenta a proporção do número de usuários VoIP. Para o cenário com serviços mistos de usuários *Web*, o desempenho do ADS se iguala ao PF, como esperado. Ele é então capaz de maximizar a capacidade para todos os mixes.

O desempenho do LDS segue o mesmo comportamento do ADS, mas não com a mesma eficiência. O RR tem um bom desempenho quando o tráfego VoIP está presente, mas uma capacidade bem restrita quando existem apenas usuários *Web*. O PF tem o comportamento inverso. Um desempenho muito bom quando apenas usuários *Web* existem, mas um desempenho muito pobre para qualquer outro cenário com a presença do serviço VoIP. O WPF tem um desempenho consideravelmente melhor que o PF, apesar de não tão bom como o ADS.

O QBMC tem um bom desempenho quando apenas o serviço VoIP é oferecido. Para todos os outros cenários, ele apresenta o pior desempenho.

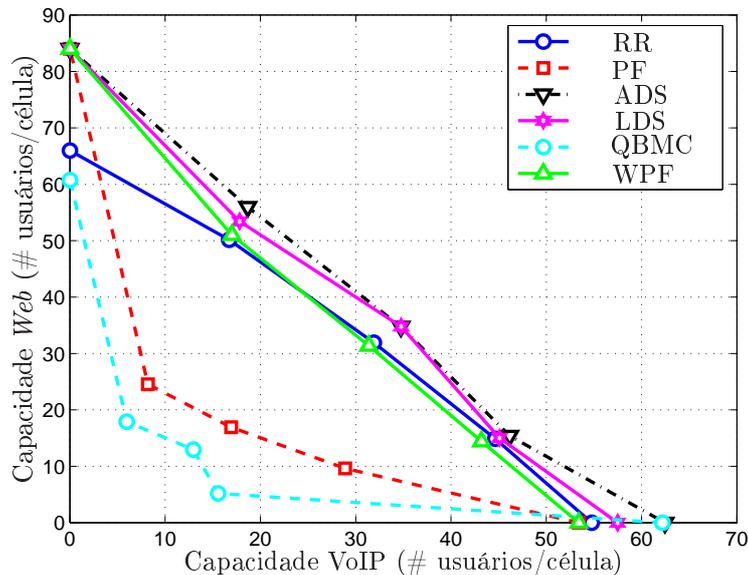


Figura 3.25: Regiões de capacidade considerando um requerimento de atraso para VoIP de 130 ms.

Quando o requerimento de atraso de VoIP é configurado para 130 ms, existe naturalmente uma degradação na capacidade quando o serviço VoIP é oferecido na Figura 3.25. O comportamento relativo das regiões de capacidade é bastante semelhante. A única diferença é que a melhora obtida pelo ADS é menor.

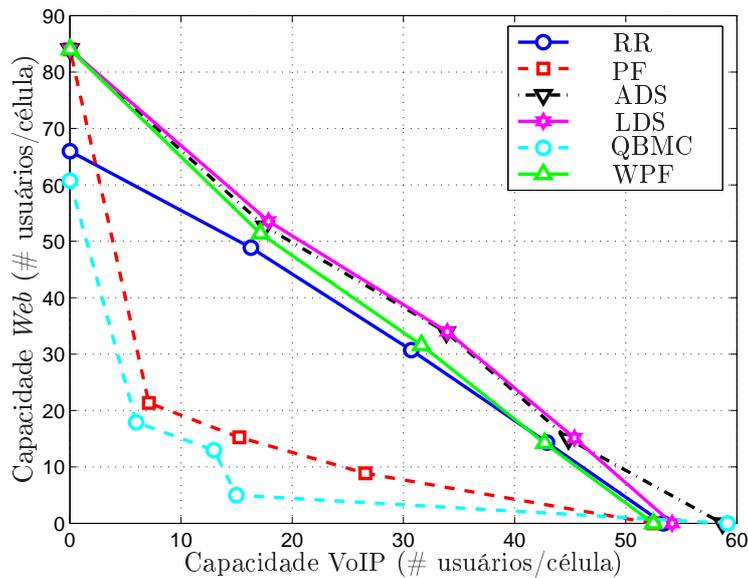


Figura 3.26: Regiões de capacidade considerando um requerimento de atraso para VoIP de 110 ms.

A Figura 3.26 apresenta a capacidade conjunta considerando uma tolerância ao atraso de 110 ms. Uma característica que distingue-se consideravelmente das outras curvas é que a vantagem do ADS sobre o LDS não é mais tão evidente, chegando a ser levemente pior em alguns casos.

Outra conclusão é que, a medida que o requerimento de atraso torna-se mais restrito, a capacidade do algoritmo WPF torna-se comparável ao RR quando se tem uma maior proporção de usuários VoIP. Isto pode ser explicado pelo fato de, considerando um requerimento de atraso menor, a limitação na capacidade do sistema é estabelecida pelo serviço *Web* para ambos os escalonadores. Aumentando-se a exigência de atraso leva a capacidade do sistema a ser limitada pelo serviço VoIP.

3.4 Sumário

Neste capítulo, uma avaliação bastante completa de alguns algoritmos de escalonamento foi efetuada considerando cenários com múltiplos serviços.

Basicamente, variações de algoritmos já antes propostos foram apresentadas de maneira tal que seja possível atender aos requerimentos de ambos os serviços.

Nas Figuras 3.27, 3.28 e 3.29 os gráficos em barras representam as capacidades absolutas obtidas por cada algoritmo e considerando diferentes requerimentos de atraso.

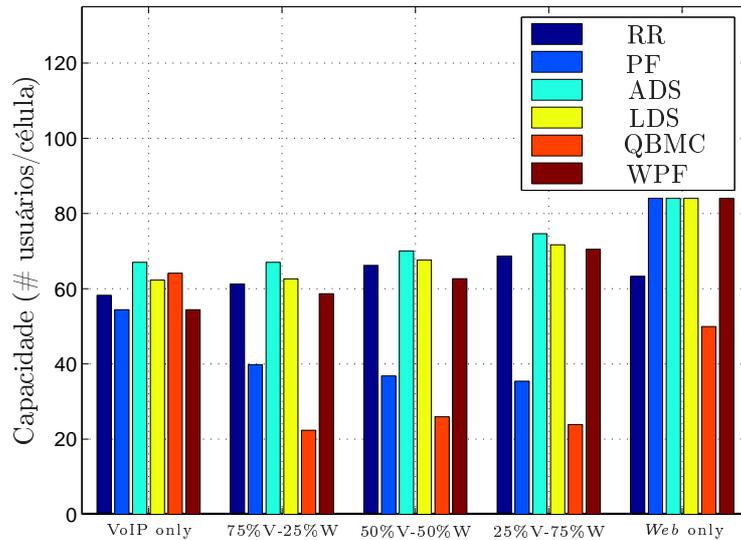


Figura 3.27: Capacidade absoluta considerando um requerimento de atraso para VoIP de 150 ms.

Os algoritmos ADS e LDS obtiveram os melhores desempenhos ao se analisar todas as proporções entre os serviços, com uma leve vantagem para o primeiro. Tais algoritmos consistem em uma variação do PF, com uma adição de um termo multiplicativo regido por uma função, que assume uma forma assintótica ou linear.

Pode-se perceber que o LDS obtém uma melhora no seu desempenho em relação ao ADS quando o requerimento de atraso aumenta. Isto pode ser explicado pelo fato de, como o sistema tende a operar bem mais próximo ao requerimento de atraso, a função linear do LDS age com mais intensidade quando o requerimento não é atingido, evitando que o atraso do usuário alcance os limiares de descarte. O ADS tende a agir somente quando o atraso do pacote mais antigo está bem próximo do descarte, tendo portanto menos tempo para a recuperação em situações em que vários usuários estejam próximo de terem pacotes descartados.

O QBMC, que leva em consideração o tamanho do *buffer* de transmissão, saiu-se bem para o cenário onde apenas usuários VoIP existiam. Com destaque para sua vantagem ao se aumentar o requerimento de atraso, como pode ser percebido na Figura 3.29. A capacidade para este cenário chega a superar a marca dos 60 usuários, assim como a obtida ao se empregar ambos ADS e LDS. No entanto, este algoritmo saiu-se muito ruim quando o serviço *Web* é adicionado ao sistema.

Os algoritmos clássicos comportaram-se de forma antagônica, como esperado. O RR proporcionou um melhor desempenho para o serviço VoIP, com uma degradação

na qualidade do serviço *Web*. Já o PF proporciona um ótimo desempenho para o serviço *Web*, com um desempenho muito ruim proporcionado aos usuários do VoIP.

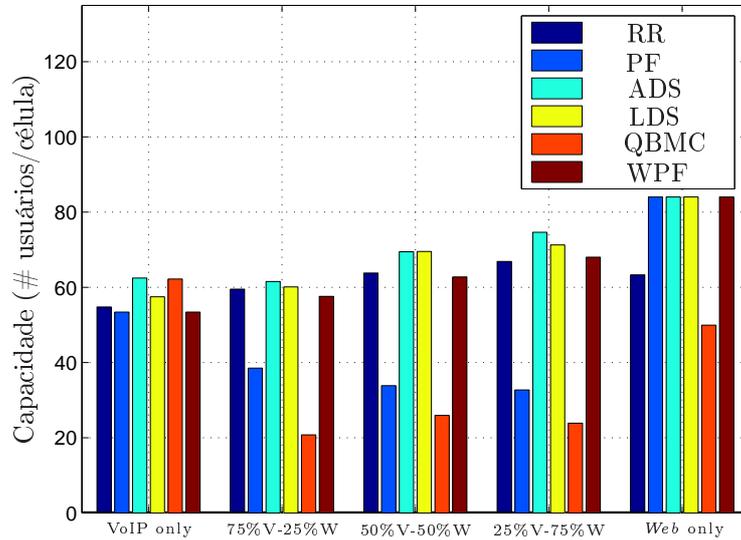


Figura 3.28: Capacidade absoluta considerando um requerimento de atraso para VoIP de 130 ms.

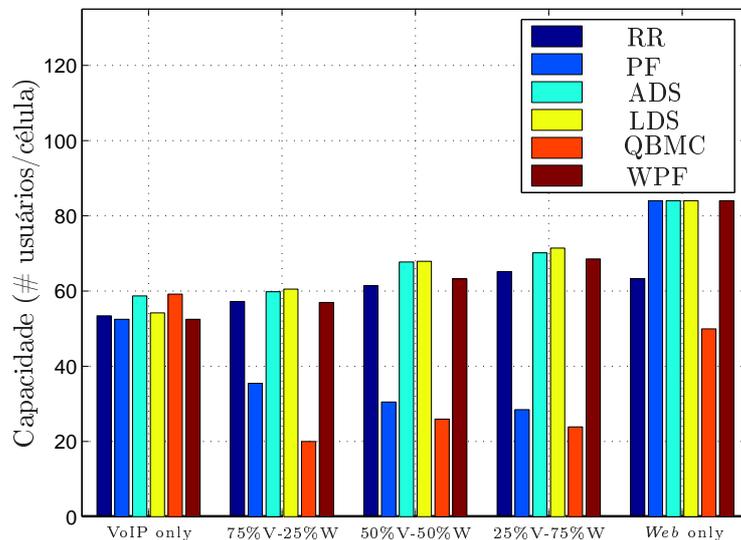


Figura 3.29: Capacidade absoluta considerando um requerimento de atraso para VoIP de 110 ms.

De maneira geral, os algoritmos com os melhores resultados podem ser definidos como variações do PF em que o requisito de atraso é levado em consideração através

da FB. Isto proporciona uma flexibilidade na garantia de QoS para serviços com e sem restrição de atraso e em qualquer proporção de distribuição de carga.

Outra conclusão bastante interessante, mas também esperado, é que a capacidade absoluta em número de usuários cresce a medida que a proporção de usuários *Web* aumenta. Isto reflete a própria natureza do HSDPA que foi concebido visando serviços de melhor esforço.

Controle de Admissão para VoIP em HSDPA

O CAC tem o objetivo de evitar que o sistema entre em sobrecarga, quando a taxa de chegada de usuários é muito alta. Se novas chamadas são admitidas durante situações de sobrecarga, a qualidade do serviço degrada devido a altos valores de atraso e perda de pacotes, se o VoIP é o serviço a ser provido.

No UTRAN, a RNC tem papel de executar o algoritmo de CAC para novos usuários, assim como aqueles em processo de *handover*. Os recursos em cada célula controlada pela RNC são monitorados por meio de medidas reportadas do *Node B* para a RNC através da sinalização *Node B Application Part* (NBAP) [46].

Na operação básica do UMTS, especialmente para canais dedicados, os recursos monitorados são: potência de transmissão e códigos de canalização. i.e. apenas limitações físicas são consideradas. De fato, se serviços com QoS bastante sensível devem ser providos, alguns outros recursos ou métricas devem ser monitorados.

A Figura 4.1 apresenta um fluxograma representativo do esquema de CAC. A cada novo usuário no sistema (sessão), o algoritmo é responsável por, baseado em métricas de utilização de recursos, decidir pela aceitação ou bloqueio desta nova sessão.

Este capítulo consiste na proposta de um arcabouço de CAC baseado no QoS para o serviço VoIP. A classificação dos esquemas de CAC assim como a descrição do mecanismo de CAC proposto são feitos na Seção 4.1. Em seguida, os modelos utilizados na ferramenta de simulação são apresentados na Seção 4.2. Os resultados são mostrados na Seção 4.3. Por fim, alguns comentários são realizados na Seção 4.4.

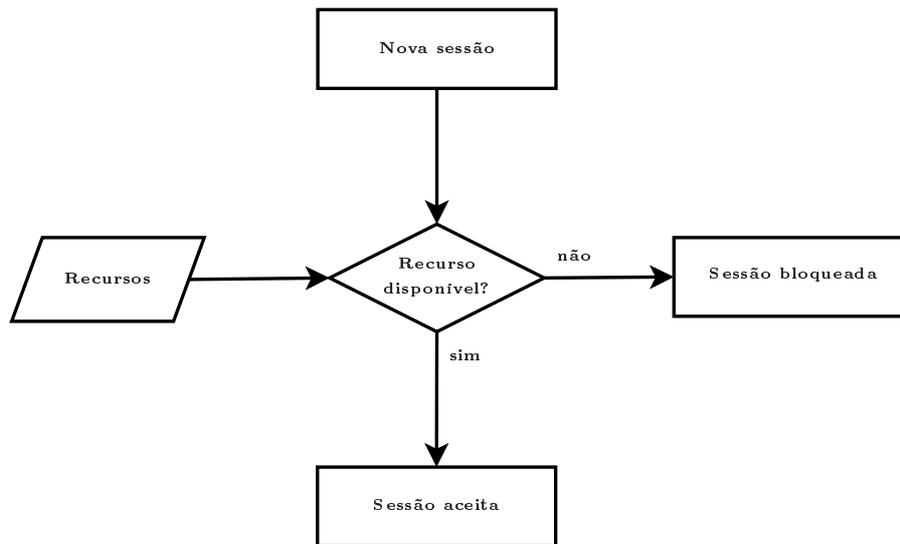


Figura 4.1: Ilustração do funcionamento do controle de admissão.

4.1 Algoritmos de Controle de Admissão

4.1.1 Controle de Admissão de Enlace

Basicamente, o controle de admissão de enlace deve se preocupar com os recursos de rádio e distribuí-los dentre as diferentes classes de serviço de modo a atingir um desejado padrão no QoS [47].

No enlace direto do WCDMA, os recursos limitantes são potência e códigos de canalização. Tipicamente, a potência limita o sistema antes que o recurso de código seja exaurido [21]. Com relação à potência, o controle de admissão de enlace deve ser baseado nas métricas de potência de transmissão e reports em cada célula e na estimativa da demanda de potência de cada novo enlace (novo usuário ou *handover*). Se a adição de um novo enlace não viola os limites de transmissão de potência da célula, este é criado, caso contrário é negado.

Quando o HS-DSCH também é suportado pelo sistema, o controle de admissão de enlace fica um pouco mais complexo. O HS-DSCH utiliza-se da potência restante da base deixada pelos canais dedicados e comuns. Dessa maneira, um sistema em operação normal e utilizando-se do HS-DSCH está quase sempre transmitindo com a capacidade de potência máxima da base.

O controle de admissão de enlace deve então considerar apenas a potência utilizada pelos canais non-HS em cada célula, que são os canais comuns com alocação de potência fixa e os canais dedicados regidos pelo algoritmo de controle de potência,

que transportam dados de tráfego e/ou controle. O trabalho em [47] avaliou o desempenho do controle de admissão de enlace utilizando diferentes limiares. Utilizando este método, não existe qualquer reserva de potência para o HS-DSCH e, conseqüentemente, nenhuma garantia de QoS no caso de altas cargas suportadas pela rede. Em [28], a taxa de bits é garantida para usuários já conectados ao sistema em um cenário de serviços mistos consistindo de voz no DCH e tráfego *Web* e *Streaming* no HS-DSCH, e é baseado em um envio de métrica existente no padrão [48].

4.1.2 Controle de Admissão de Sessão

O controle de admissão de sessão é responsável por, baseado em medidas de desempenho recebidas do *Node B*, decidir como o sistema será capaz de servir uma nova sessão sem degradar o desempenho de serviços de mesma ou mais prioritária classe.

A maneira mais simples de se verificar o estado do QoS de um sistema é, naturalmente, monitorar algumas métricas específicas que caracterizam uma boa entrega do serviço oferecido. Assumindo que o serviço de voz é provido utilizando uma interface compartilhada de pacotes, o controle do atraso torna-se algo crucial.

O esquema de controle de admissão de sessão proposto neste trabalho considera o atraso como um recurso a ser compartilhado dentre os usuários no sistema. A Figura 4.2 representa o controle de admissão de sessão proposto considerando ambos tráfegos VoIP e *Web*. Ele é baseado em um indicador de utilização do recurso e um limiar específico. As novas sessões são aceitas ou bloqueadas dependendo do valor corrente da utilização do atraso mais uma estimativa da demanda de recurso para cada usuário. Se a soma atinge valores acima do limiar pré-determinado, o usuário é bloqueado. Deve-se ter em mente que a utilização dos recursos apenas considera as métricas de qualidade dos usuários VoIP. Dessa maneira, em um cenário de serviços mistos, os outros serviços devem respeitar as limitações de recurso de atraso de VoIP, considerando-se este como o serviço com a maior prioridade. Também é possível considerar uma margem de prioridade para os usuários VoIP, i.e. o acesso ao sistema pode ser ainda mais restrito às outras classes de serviço.

A seguir, os dois esquemas utilizados para se estimar a utilização do recurso de atraso são descritos.

Controle de admissão baseado em atraso estimado

O envio de informações de medidas do *Node B* para a RNC pode não ser factível devido a limitações na arquitetura UTRAN. Conseqüentemente, um método de,

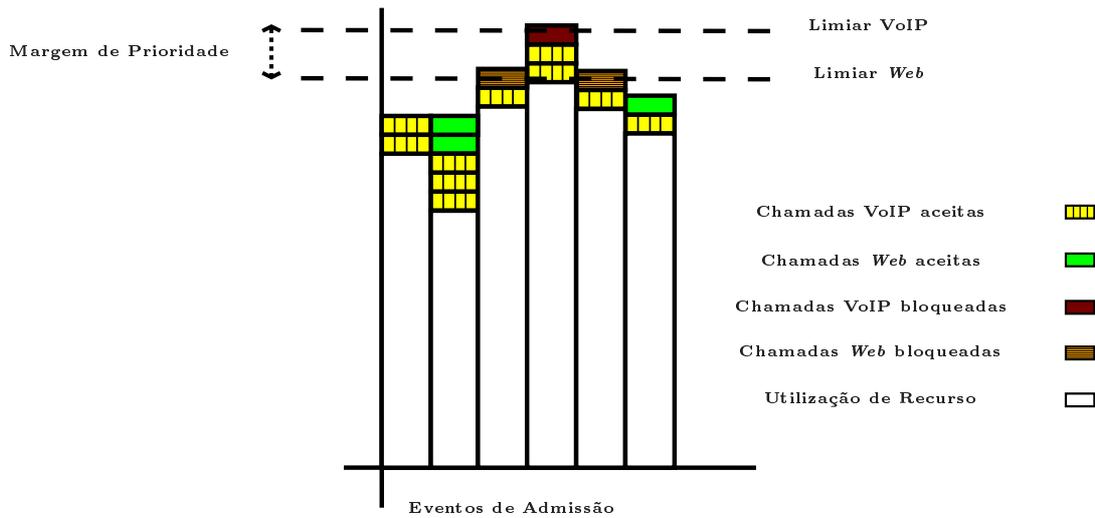


Figura 4.2: Esquema de CAC proposto.

indiretamente, se estimar o atraso em uma determinada fila baseado em outras medidas é desejado. Em [5], um esquema de CAC é proposto combinando o controle de admissão de enlace baseado em potência e uma admissão de sessão baseada em medidas de tráfego, como o número de blocos de transporte na fila e a taxa com que os mesmos são servidos. Ao se utilizar estas métricas, uma estimativa de atraso pode ser calculada e comparada a um limiar no processo de admissão.

Em [17], uma nova medida é proposta visando capacitar a RNC a garantir requerimentos de atraso, baseado na taxa de descarte de pacotes devido a expiração do temporizador. Uma outra medida que pode ser facilmente implementada é o número médio de transmissões (incluindo retransmissões) para se concretizar a recepção de um certo pacote.

Em [11], o cálculo da taxa de bits provida para cada classe de serviço é padronizada. Ela é calculada utilizando o número total de bits que foram transmitidos com sucesso pela MAC-hs no *Node B* durante o último período de medida. Este intervalo também é padronizado como sendo igual a 100 ms. Utilizando este método, o *Node B* reporta para a RNC a taxa de bits provida para cada classe de serviço utilizando a sinalização NBAP através da interface *Iub* a cada intervalo de envio.

Esta informação pode ser utilizada, juntamente com o estado atual do *buffer*, para prover uma aproximação de quanto tempo decorrerá para se transmitir os bits remanescentes no *buffer*. Isto dá uma estimativa do atraso experimentado pelo último pacote transmitido da fila. O atraso para cada classe de serviço p pode ser

estimado utilizando a Equação (4.1):

$$\hat{d}_p = \frac{B_p}{R_p}, \quad (4.1)$$

na qual B_p e R_p são o número de bits no *buffer* para todos os usuários e a taxa de bits provida (ambos com relação à mesma classe de serviço), respectivamente.

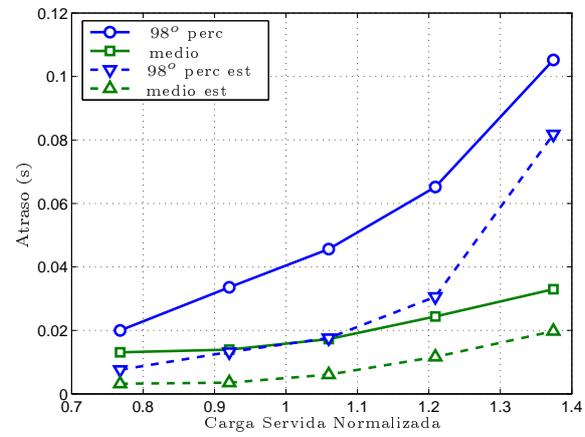
Nas Figuras 4.3(a), 4.3(b) e 4.3(c), este método de estimação é avaliado ao ser comparado às medidas de atraso percebido na MAC para os pacotes transmitidos (média e 98º percentil) para diferentes cargas (normalizadas tomando como referência o ponto de satisfação de 95% utilizando o escalonador PF), utilizando diferentes algoritmos de escalonamento (RR, PF e ADS) e em um cenário com apenas usuários VoIP. Pode-se perceber que a estimação do atraso é bem otimista comparando-se ao atraso medido, mas mantém o mesmo comportamento em média e 98º percentil com o aumento da carga. Isto se dá pelo fato de a estimação não levar em consideração o tempo em que os pacotes já estão esperando na fila no momento da estimação. Também pode-se concluir que a estimação é mais precisa quando se utiliza o escalonamento RR, isto devido a sua capacidade de distribuir igualmente os recursos de janela de transmissão dentre todos os usuários, tornando a taxa de bits provida da classe de serviço a ser dividida igualmente.

Controle de admissão baseado em atraso medido

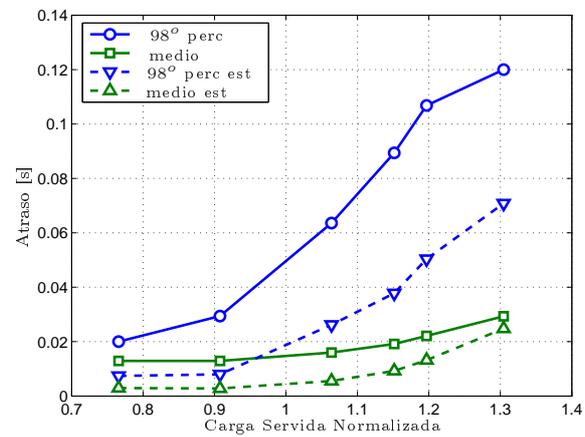
Uma maneira natural de se monitorar um determinado comportamento do sistema é efetuar medidas. Dessa maneira, o atraso deve ser medido de uma forma que apenas uma métrica deve ser capaz de representar com precisão o desempenho do sistema com relação ao atraso percebido por todos os pacotes de todos os usuários na célula.

Pela arquitetura do UTRAN as medidas devem ser efetuadas no *Node B*, que está mais próximo da interface de rádio através da MAC-hs, e reportados à RNC. Estas sinalizações de medidas já são padronizados e se utilizam de uma filtragem exponencial baseado em filtros *Autoregressive Moving Average* (ARMA) para medidas da camada física [46]. O problema com este método é que o processo de filtragem rastreia a média e, o atraso médio para o tráfego VoIP no HSDPA não aumenta significativamente com carga crescente, enquanto que a satisfação é consideravelmente degradada com o aumento da carga, como pode ser visto nas Figuras 4.3(a), 4.3(b) e 4.3(c).

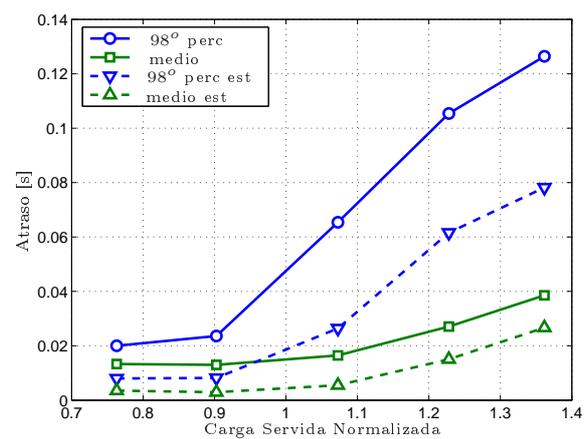
Em [28], uma variação da filtragem exponencial é considerada para ser



(a) Round Robin.



(b) Proportional Fair.



(c) Delay Scheduler.

Figura 4.3: Média e 98^o percentil do atraso medido e estimado.

implementada na RNC para a filtragem de medidas enviadas do *Node B*. Ele é chamado de filtro *Attack-Decay* (AD). A atualização das medidas filtradas é efetuada de acordo com as seguintes expressões:

$$D_n = M_n - F_{n-1}, \quad (4.2)$$

$$\text{if } D_n > 0 \text{ then } F_n = F_{n-1} + D_n \cdot f_{up}, \quad (4.3)$$

$$\text{else if } D_n < 0 \text{ then } F_n = F_{n-1} + D_n \cdot f_{down}, \quad (4.4)$$

onde F_{n-1} é o antigo resultado do filtro, F_n é a medida filtrada atualizada e M_n é a nova amostra que alimenta o filtro. Este filtro tem a vantagem de ser capaz de rastrear qualquer percentil. Para isso, deve-se apenas ajustar os valores de f_{up} e f_{down} , fatores de ataque e decaimento, respectivamente.

Esta é a motivação de se utilizar o filtro AD para se avaliar as medidas de atraso dos pacotes VoIP. Cada pacote transmitido ou descartado provê uma amostra de atraso que alimenta o filtro AD. O valor das filtrações atualizadas de cada célula é reportado à RNC a cada intervalo de envio de medidas, para ser utilizado pelo CAC. Os valores de f_{up} and f_{down} assumidos neste trabalho foram 0,5 e 0,05 respectivamente.

4.2 Modelos de Simulação

Com o intuito de se avaliar as propostas, as simulações computacionais se utilizaram de um simulador de interface de rádio WCDMA com suporte ao HS-DSCH, baseado em um *toolbox* para simulações de redes celulares no MATLAB[®].

Esta seção consiste na descrição dos principais modelos e considerações utilizados neste trabalho.

4.2.1 Grid Celular

Nas simulações, um *grid* com 3 células hexagonais por cada sítio foi considerado. No total, 7 sítios celulares existem na área de cobertura do sistema. A técnica de *wrap around* foi utilizada para se evitar efeitos de borda [34,35]. A Tabela 4.1 mostra as principais considerações com relação ao modelo utilizado para o *grid* celular.

Parâmetro	Valor	Unidade
Raio de Célula	500	m
Distancia entre Sítios	1500	m
Número de Sítios	7	-
Células por Sítio	3	-

Tabela 4.1: Parâmetros do *grid*.

4.2.2 Propagação

A modelagem do enlace de rádio no simulador é efetuada considerando os seguintes fenômenos:

- ▶ Desvanecimento de Larga Escala - também chamado de sombreamento, é modelado como um mapa sobre o grid contendo variáveis aleatórias com distribuição log-normal a cada distância de descorrelação. Também é considerado um certo grau de correlação entre os diferentes sítios.
- ▶ Desvanecimento de Pequena Escala - também é utilizado um vetor para emular as variações rápidas de canal assim como o efeito do fator de ortogonalidade variável devido à presença de multipercursos. O perfil de canal multipercurso está embutido no mapa, assim como a frequência da portadora e a banda de coerência. Sua resolução e tamanho também podem ser ajustados.
- ▶ Perda de Percursso - as simulações utilizaram do modelo de Okumura-Hata [49] para o modelo da perda média. A perda de percurso pode ser expressa através da Equação (4.5):

$$P = 28,6 + 3,5 \cdot \log_{10}(d_m), \quad (4.5)$$

onde d_m é a distância em metros.

- ▶ Perfil de antena setorizada - considera um perfil de antena setorizada.

A Tabela 4.2 mostra as principais considerações em relação aos parâmetros do enlace de rádio.

4.2.3 Mobilidade

O perfil de mobilidade considera uma velocidade absoluta com distribuição de Rayleigh e aceleração Gaussiana. A distribuição de velocidade também inclui correlação de aceleração e velocidade (por exemplo devido à resistência do ar).

Parâmetro	Valor	Unidade
Perfil de Canal	<i>3GPP Typical Urban</i>	-
Velocidade do Móvel	3	km/h
Desvio Padrão do Sombreamento	8,9	dB
Correlação do Sombreamento entre Sítios	0,5	-
Distância de Descorrelação do Sombreamento	100	m
Mínima Perda de Percursso	70	dB
Máxima Perda de Percursso	146	dB
Ruído no Receptor	-130,1567	dBW

Tabela 4.2: Principais modelos de propagação.

4.2.4 Tráfego

Dois modelos de tráfego distintos foram considerados neste trabalho. O primeiro é o modelo de voz, com geração de pacotes com taxa constante durante os períodos de atividade. O outro é o serviço interativo de navegação *Web*, que tem um comportamento de geração de pacotes em rajadas e com taxa variável. Nesta seção, ambos os serviços são descritos à maneira em que são utilizados no simulador.

Voz sobre IP

A geração de tráfego de voz é baseada em um modelo de transmissão descontínua para serviços de chaveamento de circuito, apesar de algumas novas características serem adicionadas com o intuito da modelagem dos cabeçalhos inseridos pela camada IP, como:

- ▶ Número variável de quadros AMR por pacote IP - um número maior de quadros concatenados em um pacote IP aumenta a eficiência devido à redução relativa dos cabeçalhos, no entanto ocasiona um aumento do atraso.
- ▶ Compressão de Cabeçalho - muito importante para serviços VoIP devido a sua capacidade de redução do tamanho do cabeçalho. Um comportamento constante do algoritmo é assumido, resultando em um cabeçalho de tamanho constante. Em um sistema real, o cabeçalho tem tamanho variável [8].

Durante os períodos de atividade, a sessão de voz (CODEC) gera um quadro de voz (pacote) que é entregue às camadas inferiores. O estado inicial de uma sessão (ativo ou inativo) é definido quando a mesma é criada baseado na atividade de voz definida. O estado é atualizado a cada 20 ms, e a troca de estados é baseada em uma probabilidade, que é exponencialmente distribuída. Na Tabela 4.3, as principais

considerações com relação ao tráfego de voz são apresentadas. O tamanho médio dos pacotes já inclui os cabeçalhos IP comprimidos, assim como os cabeçalhos da camada RLC. Os pacotes VoIP de tamanho igual a 183 bits chegando na camada MAC-hs equivalem a uma taxa de geração de 9.15 kbps.

Parâmetro	Valor	Unidade
Taxa do Codec AMR	7,95	kbps
Atividade de voz	40	%
Compressão de cabeçalho	sim	-
Tamanho do pacote (cabeçalhos incluídos)	183	bits
Tamanho do cabeçalho (média assumida)	24	bits
Duração média da chamada	30	s
Quadros AMR por pacote IP	1	-
Tempo médio de atividade	2	s
Tempo médio de inatividade	3	s

Tabela 4.3: Parâmetros do tráfego de voz.

Navegação *Web*

A sessão *Web* é modelada utilizando-se de duração e tempos de leitura exponencialmente distribuídos. O tamanho do pacote segue uma distribuição log-normal truncada. A Tabela 4.4 mostra os principais parâmetros de modelagem do tráfego *Web*. Este modelo provê uma geração de dados de 27.5 kbps por um único usuário.

Parâmetro	Valor	Unidade
Tempo médio de sessão	60	s
Tempo médio de leitura	5	s
Tamanho mínimo de pacote	1	kbytes
Tamanho máximo de pacote	200	kbytes
Tamanho médio de pacote	3,605	bytes
Desvio padrão do tamanho de pacote	0,7839	bytes

Tabela 4.4: Parâmetros do tráfego *Web*.

4.2.5 Camadas RLC/MAC

Como esta avaliação está focada em otimizações com relação ao gerenciamento de diferentes tráfegos, modelos simplificados de camadas superiores, como *Transport Control Protocol* (TCP) e RTP, são considerados. Algumas das influências das

mesmas já são incluídas nos modelos de tráfego descritos na sessão 4.2.4. Dessa maneira, apenas as entidades de protocolo dentro do UTRAN são consideradas e descritas nesta seção.

Na RLC, o modo UM é considerado para o tráfego VoIP. Neste modo de operação, segmentação de pacotes pode ser realizada mas pacotes recebidos incorretamente não são retransmitidos, sendo apenas descartados. Isto se deve ao estrito requerimento de atraso para serviços de voz, em que não existe tempo para a espera de confirmações e requerimentos de retransmissões na camada RLC. Para serviços em tempo não-real, como navegação *Web*, o modo AM é utilizado. Ele se utiliza do protocolo ARQ oferecido pela camada RLC.

As principais atribuições da camada MAC na modelagem de simulação dizem respeito à sub-camada MAC-hs, que situa-se no *Node B*. Os pacotes provenientes da RLC são concatenados de acordo com o formato de transporte selecionado, e então transmitidos. O protocolo HARQ também é gerenciado pela MAC, que considera processos paralelos no modelo SAW [11]. Quando o número de retransmissões atinge o limite, o pacote MAC é descartado e, dependendo do modo de operação da camada RLC, os bits são perdidos ou retransmitidos pela RLC.

4.2.6 Tempo de simulação

As simulações são baseadas no tempo de quadro de rádio (10 ms). Em cada iteração, laços de TTI (2 ms) e slot (0,667 ms) são também considerados.

A taxa de chegada de usuários é baseada em Poisson e, já que poucos usuários existem no sistema no início da simulação, um período de transiente é considerado antes de as estatísticas começarem a ser coletadas.

A simulação roda por um certo tempo, no entanto, alguns *snapshots* com diferentes sementes aleatórias são considerados para cada carga e configuração. Na análise, médias sobre os *snapshots* são calculadas para cada ponto. Na Tabela 4.5, os parâmetros com relação aos tempos de simulação são apresentados. Pode-se observar que alguns parâmetros se diferenciam nos cenários com serviços simples e mistos. Isto se deve ao fato de, como o tráfego *Web* ter características em rajadas, um maior tempo de simulação é necessário para se obter estatísticas mais confiáveis.

4.3 Resultados

Os algoritmos de CAC foram avaliados em dois diferentes cenários:

- ▶ Serviços Simples - onde apenas o serviço VoIP é oferecido. É um cenário bem

Parâmetro	Valor	Unidade
Tempo de simulação (serviços simples)	100	s
Tempo de simulação (serviços mistos)	200	s
Número de Snapshots	5	-
Período de transiente	5	s
Período de Slow Start	200	ms

Tabela 4.5: Tempos de simulação.

simplificado onde não é possível de se efetuar gerenciamento de prioridade entre serviços.

- **Serviços Mistos** - onde os serviços de VoIP e *Web* compartilham os recursos da rede. Neste cenário é possível se gerenciar os serviços baseado em prioridade tanto durante a admissão como no escalonamento. No entanto, apenas o primeiro é avaliado.

Alguns parâmetros de simulação são mostrados na Tabela 4.6. No cenário com serviços mistos, o ADS não foi empregado porque uma versão simplificada é utilizada que não suporta diferentes requerimentos de atraso. Dessa maneira, apenas os algoritmos PF e RR foram considerados. O primeiro oferece uma atrativa relação custo-benefício entre vazão de dados no sistema e justiça, explorando a diversidade multi-usuário [30]. Já o último tem a propriedade de dividir igualmente os recursos de janela para transmissão.

Vale ressaltar que os valores de carga apresentados são normalizados arbitrariamente com relação ao ponto onde o algoritmo PF atinge uma satisfação de 95%, em ambos os cenários.

4.3.1 Cenário de serviço simples (VoIP)

Cenário

Três algoritmos de escalonamento foram considerados neste estudo: os clássicos RR e PF, e o ADS. Na Figura 4.4 é possível se comparar o desempenho do sistema para os diferentes esquemas de escalonamento, sem nenhum CAC, através da taxa de satisfação de usuários para o serviço VoIP. Está bem nítido que o RR supera o PF em desempenho, principalmente devido à distribuição igual de janelas de transmissão entre os usuários, o que é o mais importante especificamente para o serviço VoIP, devido a seus requerimentos de baixa taxa de transmissão e atraso. O ADS se comporta ainda melhor que o RR, porque tem uma vantagem de priorizar

Parâmetro	Valor	Unidade
Escalonamento (serviço simples)	RR, PF e ADS	-
Escalonamento (serviços mistos)	PF e RR	-
Processos HARQ paralelos	6	-
Máximo número de retransmissões na MAC	5	-
Combinação HARQ	<i>Chase Combining</i>	-
Max. # de multiplexação de código	4	-
# de códigos para HSDPA	8	-
Máxima potência de transmissão da base	20	W
Potência dos canais comuns	3	W
Potência do SCH	0,2	W
Potência do CPICH	1,2619	W
Potência de cada HS-SCCH	0,8	W
Taxa alvo de erro de TTI	0,01	-
Limiar de descarte do VoIP	130	ms
Limite de QoS	90	%

Tabela 4.6: Principais parâmetros de simulação.

usuários com melhores condições de canal e, ao mesmo tempo, tentando garantir os requerimentos de atraso através da FB.

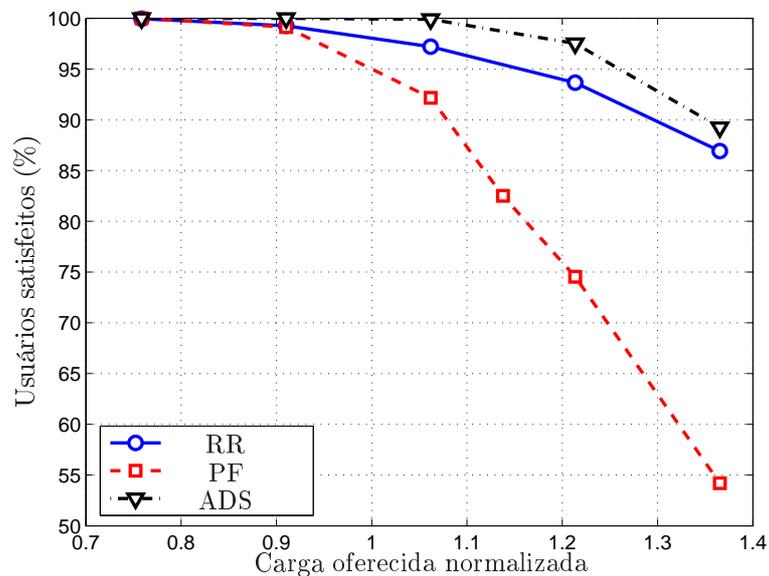


Figura 4.4: Desempenho dos escalonadores para tráfego VoIP.

Mas quando algumas outras medidas de desempenho são analisadas, é interessante perceber que, quando se observa os bits perdidos por erros de

transmissão na Figura 4.5, o desempenho do PF é melhor que o RR, porque o primeiro considera as condições de canal para prover acesso à transmissão. Dessa maneira, usuários com mais alta CIR são postos a transmitir, significando que as probabilidades de perda por erro na transmissão são menores, sendo ainda menos freqüente que o número máximo de tentativas de transmissão na MAC seja atingido e o pacote seja perdido, já que a RLC opera no modo UM.

O desempenho do ADS para os bits perdidos é ainda melhor que o PF. Isto pode ser explicado pelo fato de este prover acesso à usuários com melhores condições de canal, assim como o PF. Este quesito se torna menos relevante apenas quando o atraso de pacote de determinado usuário cresce perto do valor de descarte. Sobre baixas cargas, onde o tempo de espera não é significante, o ADS trabalha praticamente igual ao PF.

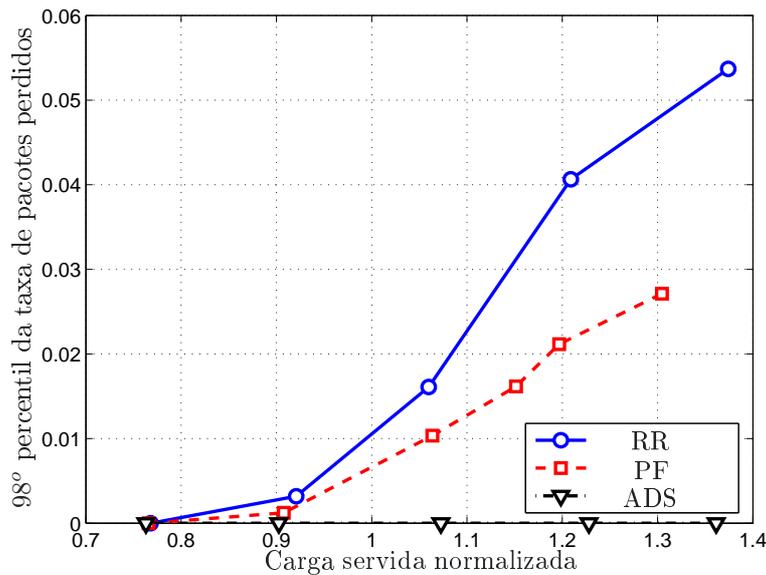


Figura 4.5: 98º percentil da taxa de pacotes perdidos.

Na Figura 4.6, o melhor resultado geral em favor do RR, comparado ao PF, é explicado. A taxa de descarte para o RR é consideravelmente menor. Isto se deve ao fato de o PF não considerar o atraso de pacote quando ordenando os usuários para escalonamento. Dessa maneira, usuários com altos valores de atraso de pacote podem não ser escalonados até o temporizador de descarte expirar e, então, o pacote ser descartado.

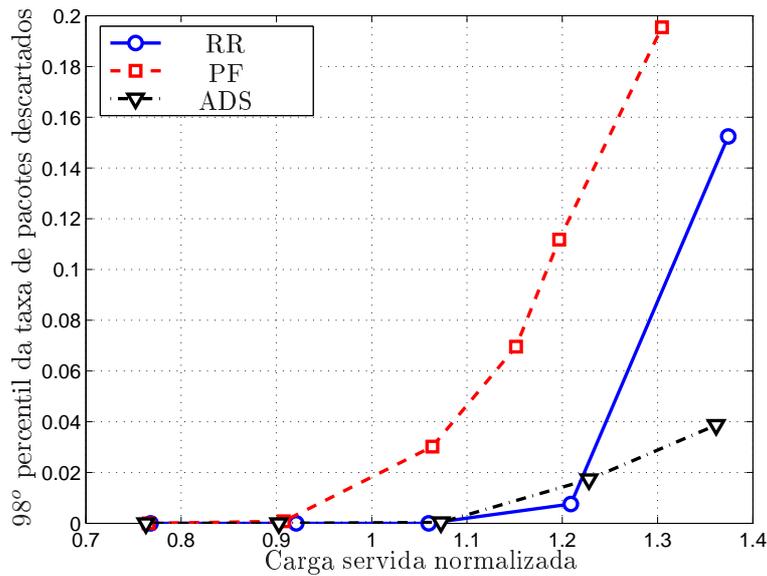


Figura 4.6: 98º percentil da taxa de descarte de pacotes.

Já o ADS tem um comportamento um pouco pior que o RR para cargas baixas, mas melhor para cargas altas. E ainda é bem melhor que o PF, por causa de seu esquema de priorização de pacotes na eminência de serem descartados.

Através desta rápida análise, pode-se chegar facilmente à conclusão de que, quando se considera o escalonador RR, a limitação de QoS para VoIP tem uma forte influência de erros de transmissão do canal rádio móvel, o qual não é utilizado de forma adequada pelo escalonador. O atraso se torna uma forte causa de degradação de qualidade apenas em cargas bastante altas. Por outro lado, quando o PF é considerado, o escalonador otimiza o acesso ao meio de maneira que este seja efetuado durante melhores condições de canal, minimizando a perda de pacotes devido a erros na transmissão. O atraso de pacote é então o fator crítico para o PF, já que ele não leva nenhuma consideração no ato do escalonamento. A limitação de QoS é atingida devido aos descartes de pacote na MAC, enquanto os pacotes perdidos devido a erros de transmissão são bem menos significantes.

Ao se analisar o ADS, é evidente que a única causa de perda de pacotes é o atraso. Pode-se concluir que praticamente não existe perda de pacotes devido ao canal, mesmo para altas cargas. Isto o leva a ter o melhor desempenho geral. Dessa maneira, fica provado que o ADS consegue combinar a justiça do RR com o ganho de diversidade multi-usuário de maneira melhor que o PF puro para o serviço de VoIP.

Uma outra consideração neste trabalho é de que nenhum CAC de enlace é

utilizado. Isto significa que nenhuma limitação de potência ou códigos é imposta no momento da admissão. Isto se deve ao fato de o controle de admissão de enlace ser baseado na potência não utilizada pelo HS-DSCH, a potência non-HS. Já que não existe nenhum tráfego oferecido no sistema de chaveamento de circuito, a demanda de potência para os canais non-HS é bastante limitada. Isto porque apenas os canais comuns de controle e o canal dedicado associado demandam potência com maior prioridade, e os mesmos têm requerimentos bastante baixos de potência.

A Figura 4.7 mostra o comportamento da potência non-HS para uma carga crescente. Para as maiores cargas simuladas, o 98º percentil atinge cerca de 9 W. Em [47] e [28] os valores típicos assumidos para os limiares para o controle de admissão de enlace são entre 15 e 20 W. Utilizando-se desse dado, o controle de admissão de enlace não tem qualquer efeito no desempenho do sistema.

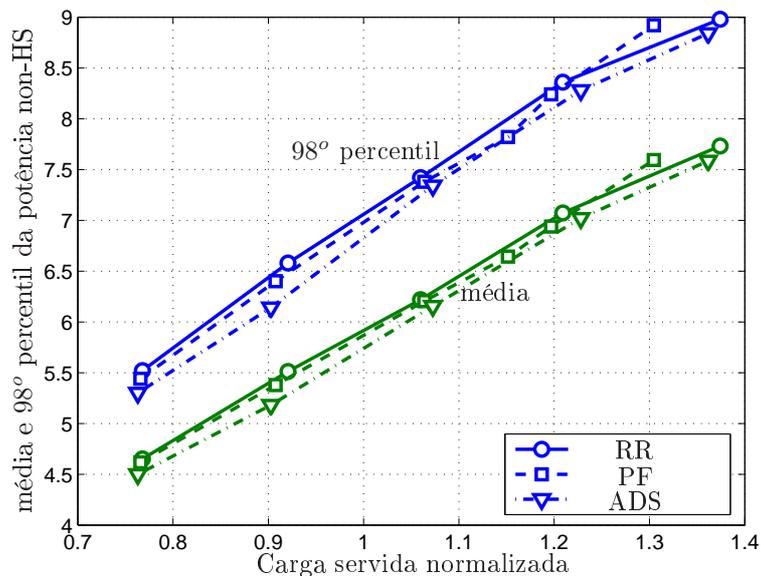


Figura 4.7: 98º percentil e média da potência non-HS.

Análise de Desempenho

Nesta avaliação onde apenas usuários VoIP existem no sistema, um limiar de admissão de 55 ms foi considerado pelo fato deste valor ser aproximadamente o 98º percentil para o limite de QoS de referência (PF sem CAC na Figura 4.3(b)). É também cerca de metade do valor do tempo para o descarte do pacote, na medida que ele tenta melhorar o desempenho através da minimização da taxa de descarte de pacote e, conseqüentemente a taxa de apagamento de quadro. A estimativa para o recurso de atraso necessária para cada nova sessão é de 2 ms, assumindo que cada

novo usuário provocará um aumento no atraso para os outros de um TTI do HSDPA.

Em ambos os esquemas de CAC, a atualização do recurso de atraso na RNC é efetuada a cada 100 ms.

O desempenho de satisfação de usuário é apresentada na Figura 4.8 para o caso do escalonador RR. Pode-se perceber que o CAC baseado em atraso estimado resulta em uma maior melhoria de desempenho. As melhorias podem ser percebidas para cargas próximas e maiores do que o limite de QoS, trazendo um ligeiro ou nenhum aumento de capacidade de até cerca de 5%. As melhorias do CAC são também destacadas em situações de sobrecarga, onde até 4% de aumento no índice de satisfação pode ser atingido.

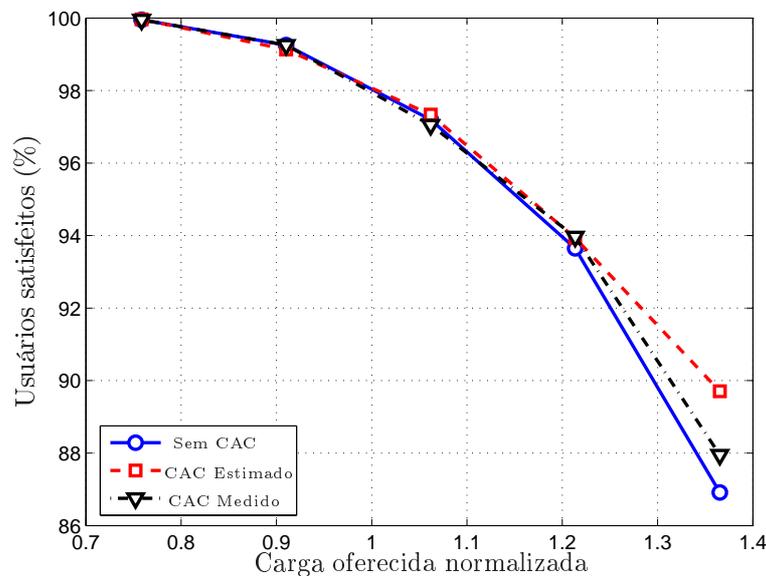


Figura 4.8: Satisfação de usuários utilizando RR.

Quando apenas os usuários que foram aceitos pelo CAC são analisados, como é mostrado na Figura 4.9, os melhoramentos de desempenho do CAC baseado em atraso medido são maiores para taxas de satisfação maiores. Isto se deve ao bloqueio imposto, que estabelece uma proteção maior aos usuários já conectados ao sistema. Considerando o mesmo limite de QoS de 90% de satisfação, o CAC prevê um crescimento de vazão no sistema (até 10% de capacidade de carga servida), significando uma maior eficiência espectral.

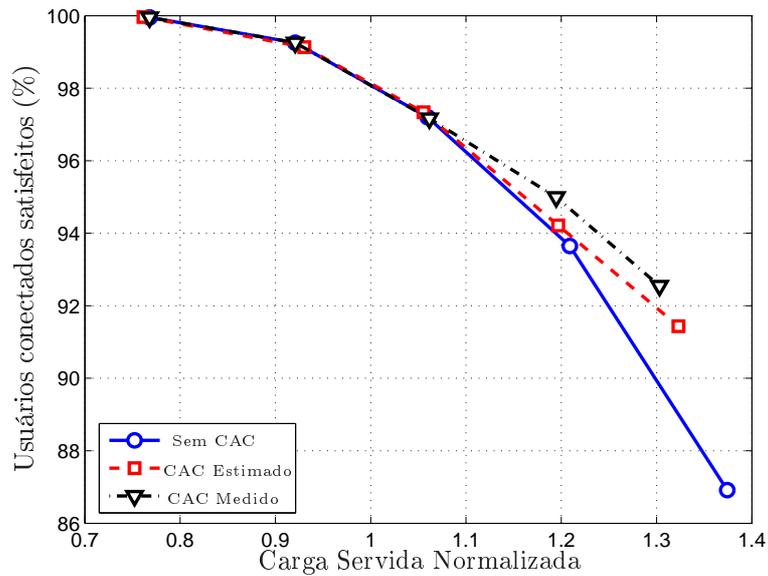


Figura 4.9: Satisfação de usuários conectados utilizando RR.

A Figura 4.10 mostra as taxas de bloqueio para cargas crescentes. O CAC baseado em atraso medido ocasiona taxas de bloqueio de até 5% para a carga máxima simulada, enquanto que o algoritmo que utilizam o atraso estimado ocasiona taxas de até 2%.

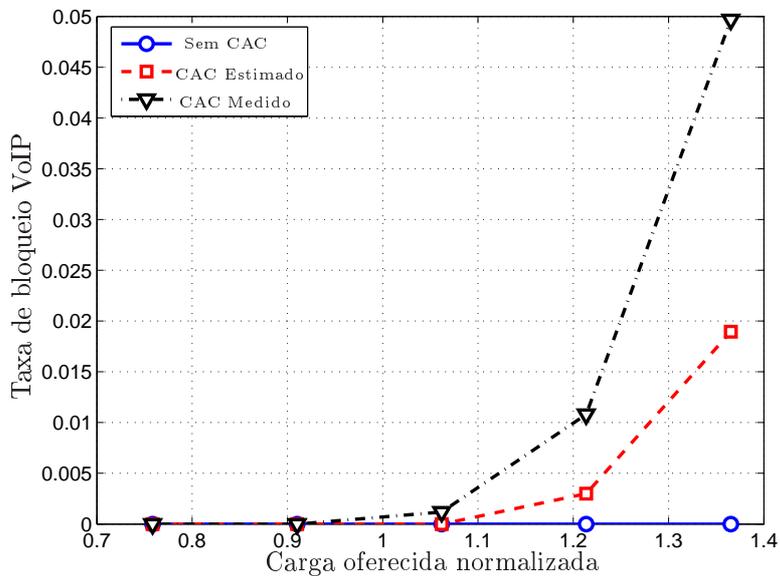


Figura 4.10: Taxa de bloqueio utilizando RR.

A razão pela qual o algoritmo com atraso estimado resulta em uma melhor satisfação de usuários é porque sua estimativa de atraso é bastante precisa quando

se utiliza o escalonador RR. Isto pelo fato da estimação ser feita baseada na taxa provida para a classe de serviço. Como um algoritmo igualitário, a taxa acaba sendo dividida igualmente entre todos os usuários, resultando em uma estimação bem próxima do tempo para se transmitir todos os bits no *buffer*.

Na Figura 4.11, o 98° percentil da taxa de perda de bits devido a erros de transmissão é apresentada. Não existe diferença efetiva em se utilizar o CAC. Isto se deve ao fato de o CAC se concentrar apenas no atraso como uma indicativa de utilização de recursos. O bloqueio poderia melhorar as condições de canal apenas através da redução de interferência, o que não parece ocorrer.

A vantagem efetiva de se utilizar o CAC se reflete na taxa de descarte de pacote, mostrada na Figura 4.12. A redução na taxa de apagamento de quadro de voz é proveniente de uma considerável diminuição no atraso, que decresce as expirações do temporizador na camada MAC em até 8% no 98° percentil. O CAC, de fato, estabelece uma limitação no tamanho da fila de escalonamento baseado no desempenho do sistema, medido na forma do atraso. Limitando-se o tamanho da fila reduz imediatamente o atraso de pacote, no caso de se utilizar o escalonador RR.

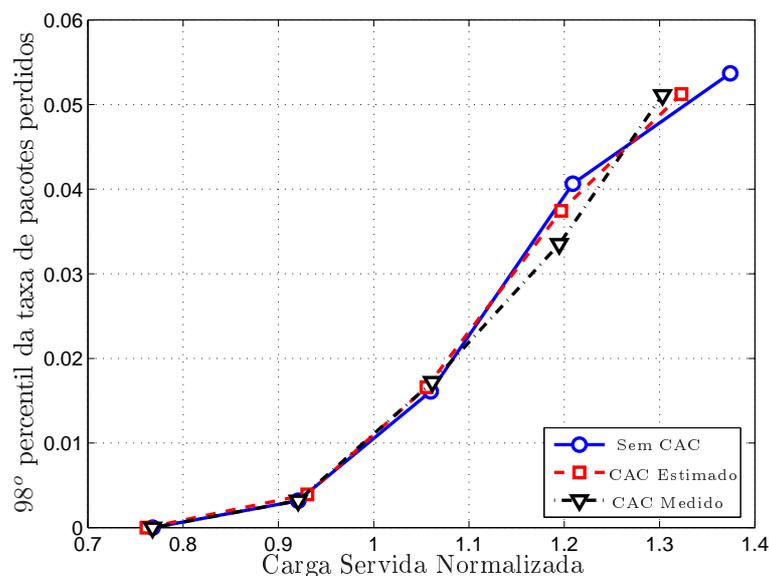


Figura 4.11: 98° percentil da taxa de pacotes perdidos utilizando RR.

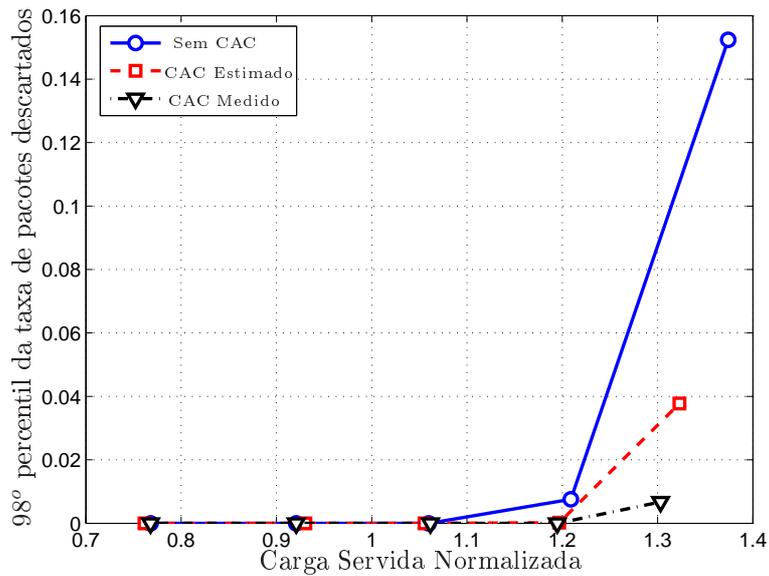


Figura 4.12: 98º percentil da taxa de descarte de pacotes utilizando RR.

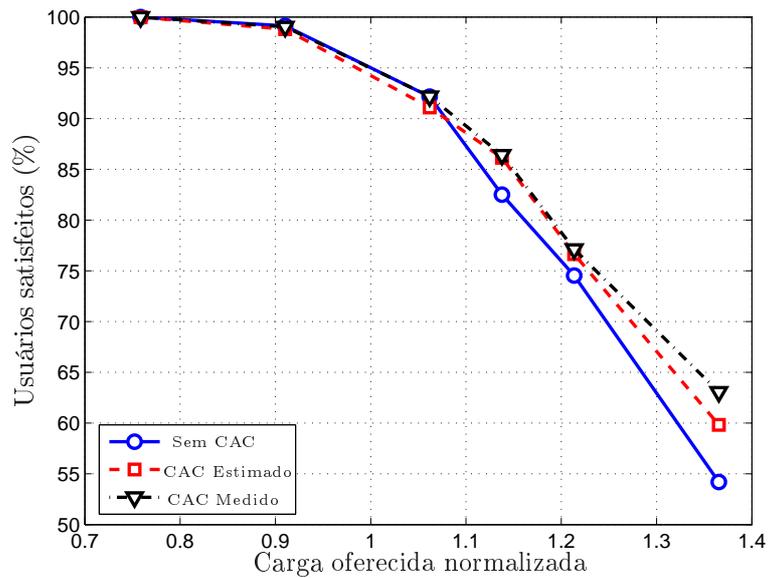


Figura 4.13: Satisfação de usuários utilizando PF.

Agora, o CAC será analisado se utilizando do escalonador PF. Na Figura 4.13, os ganhos do CAC são menores dos obtidos no caso RR, e para cargas além da capacidade estabelecida pelo limite de QoS, ou seja, onde a taxa de satisfação já é bastante baixa. Ambos os esquemas de CAC proporcionam ganhos similares para altas cargas, ocasionando um aumento de até 8% na taxa de satisfação.

Ao se observar a taxa de satisfação para os usuários conectados, na Figura 4.14,

as melhorias de se utilizar o CAC não são destacadas. O sistema não melhora em desempenho significativamente através do controle dos tamanhos de fila. Isto se deve ao fato de, apesar de a satisfação ser melhorada, a carga servida é significativamente reduzida, resultando em um desempenho similar.

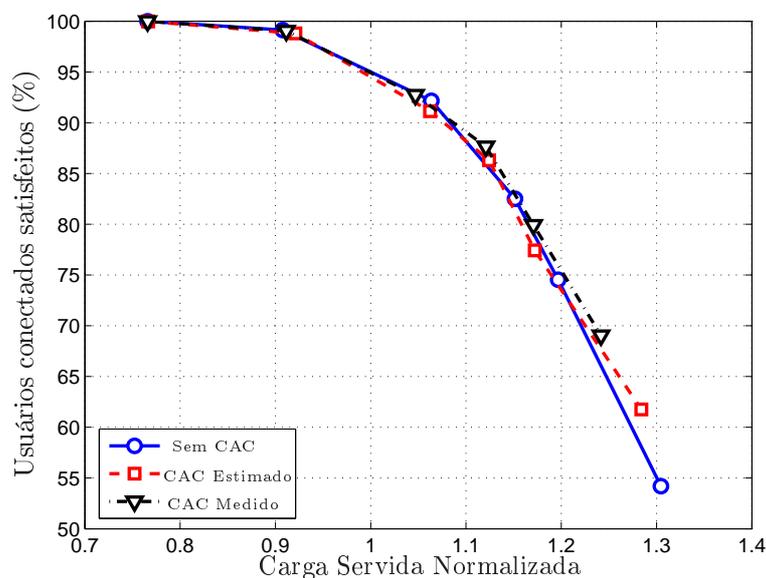


Figura 4.14: Satisfação de usuários conectados utilizando PF.

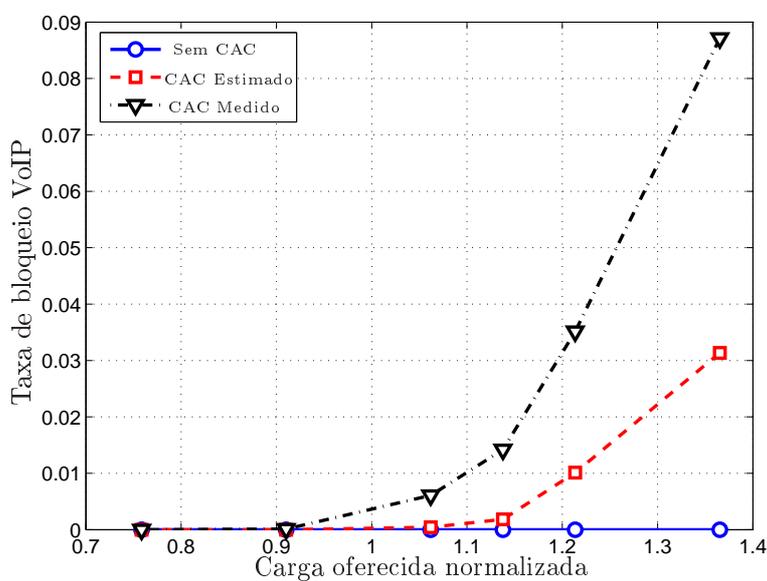


Figura 4.15: Taxa de bloqueio utilizando PF.

A Figura 4.15 mostra as taxas de bloqueio para o caso com o escalonador PF. Pode-se perceber que elas são maiores do que quando se utiliza do escalonador RR.

Isto se deve ao fato de que, quando um usuário é bloqueado, o atraso garantido pelo sistema não é igualmente distribuído entre os usuários conectados. A redução de atraso não se dá de forma tão direta quanto no caso RR, ocorrendo mais precisamente um controle de carga do que realmente um controle de atraso.

A Figura 4.16 apresenta o mesmo comportamento do caso com o escalonador RR. Já que o CAC não monitora os erros do canal, o mesmo acaba não tendo qualquer influência nesta métrica.

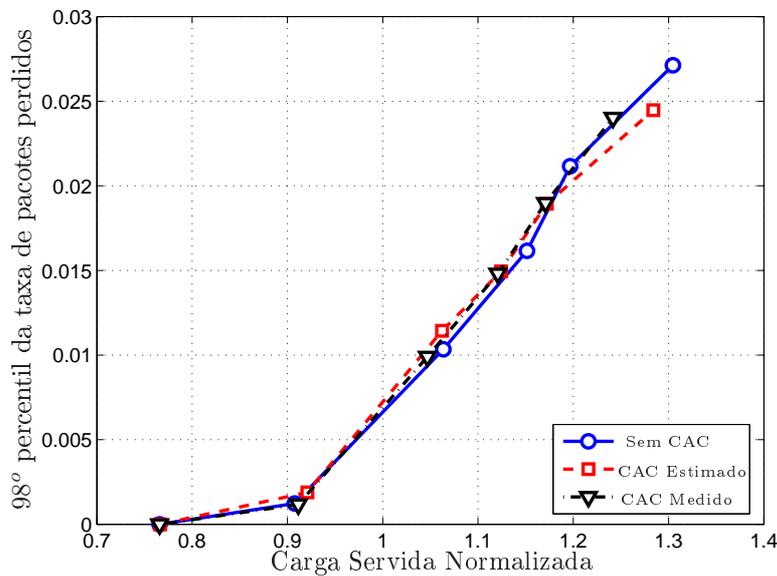


Figura 4.16: 98º percentil da taxa de perda de pacotes utilizando PF.

Os melhoramentos da taxa de descarte de pacote são apresentados na Figura 4.17. Apenas em cargas além do limite de QoS de 90% de usuários satisfeitos é que o CAC melhora o desempenho. Isto é basicamente devido ao fato de que, apesar de o tamanho de fila ser mais controlado e mais janelas de transmissão estão disponíveis, o compartilhamento de recursos não é igualitário devido ao escalonador PF. Já que os requerimentos de taxa para o VoIP é muito pequeno comparado à capacidade do HSDPA, usuários com boas taxas podem experimentar altos atrasos e, conseqüentemente altas taxas de descarte. O PF não é capaz de proporcionar justiça para usuários do serviço VoIP, pois a garantia de taxa de transmissão não é um problema para o sistema, mas sim a de atraso.

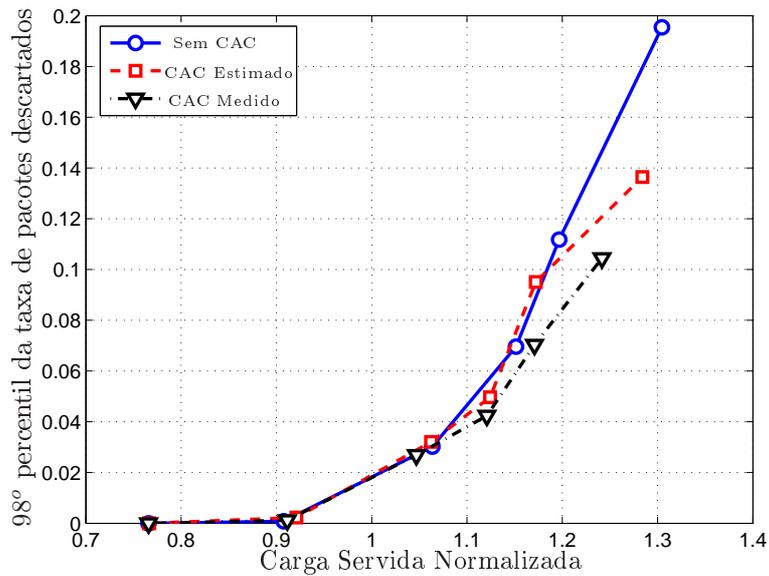


Figura 4.17: 98º percentil da taxa de descarte de pacotes utilizando PF.

Analisando-se o ADS, é possível de se observar as melhorias proporcionadas pelo CAC. A Figura 4.18 mostra a taxa de satisfação. Considerando apenas os usuários que tentam se conectar à rede, apenas o esquema com atraso estimado parece melhorar o sistema. É nele que o CAC proporciona o maior incremento de carga oferecida, de cerca de 10%.

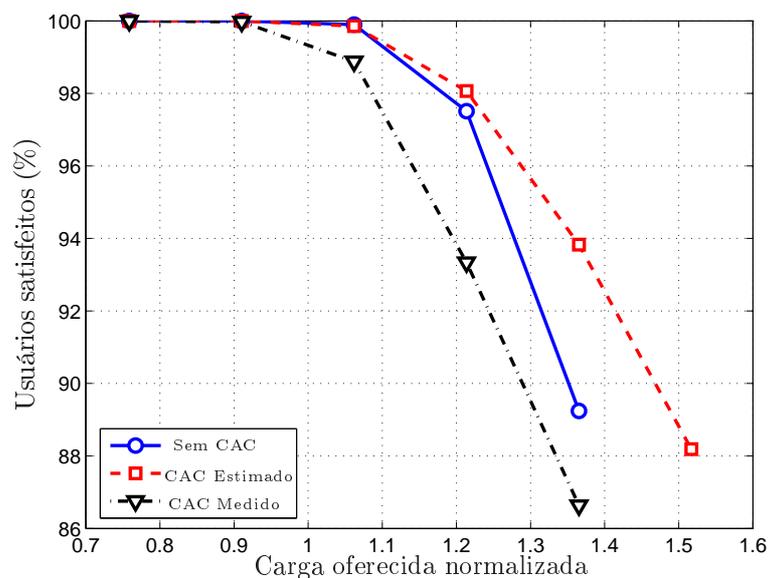


Figura 4.18: Satisfação de usuários utilizando ADS.

Agora considerando apenas os usuários conectados na Figura 4.19, o CAC com

atraso medido garante uma melhor qualidade. Isto porque sua métrica de atraso é mais confiável, mas não considera que o escalonador prioriza pacotes com maiores atrasos. Por isso que o CAC baseado em atraso medido tem uma taxa de bloqueio maior, o que pode ser observado na Figura 4.20, diminuindo a taxa de satisfação devido ao bloqueio, o que é refletido em um desempenho pior na Figura 4.18.

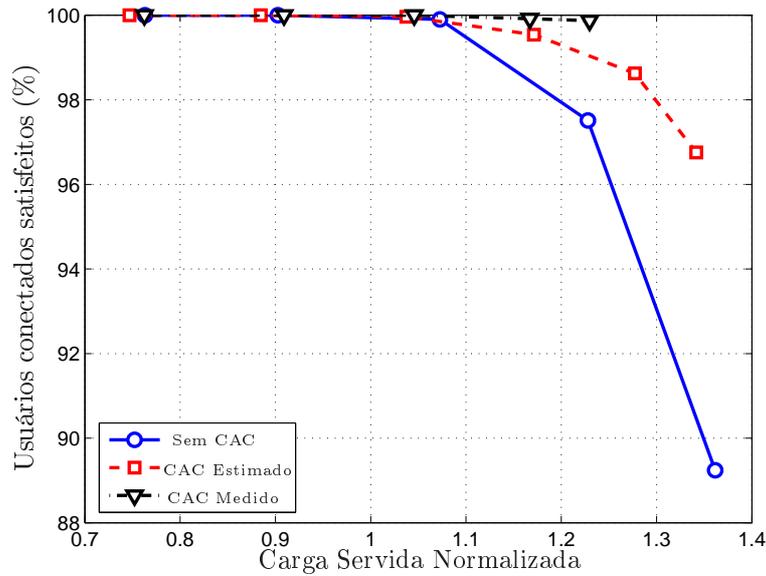


Figura 4.19: Satisfação de usuários conectados utilizando ADS.

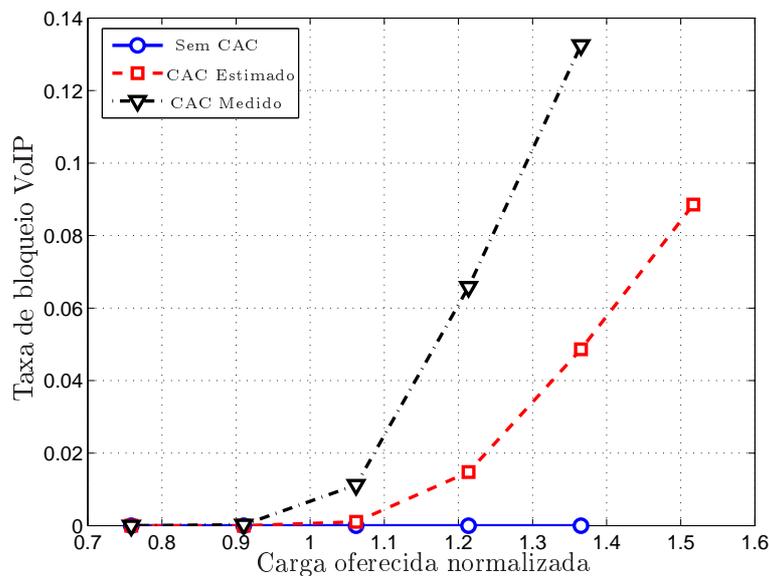


Figura 4.20: Taxa de bloqueio utilizando ADS.

Como as perdas de pacotes devido ao canal são desprezíveis ao se utilizar o

ADS, o CAC age exatamente na única causa de perda de pacotes, que é o atraso. Na Figura 4.21, o 98º percentil da taxa de descarte de pacotes mostra as melhorias de se utilizar o CAC. Isto também explica a maior taxa de satisfação obtida pelo CAC baseado em atraso medido para os usuários conectados, i.e. menos pacotes descartados ao se utilizar do CAC.

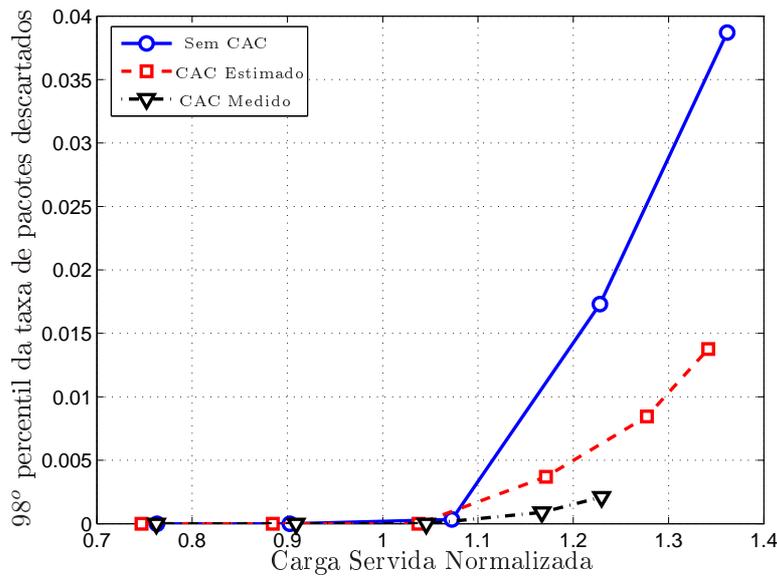


Figura 4.21: 98º percentil da taxa de descarte de pacotes utilizando ADS.

4.3.2 Cenário misto de serviços (VoIP e Web)

Cenário

No cenário composto de ambos serviços VoIP e *Web*, todo tráfego continua sendo transmitido utilizando-se do HS-DSCH. O cenário antecipa a evolução em direção ao estágio onde todos os tipos de tráfego convergiram para um ambiente *all-IP*.

A Figura 4.22 mostra um resultado ilustrativo com relação ao comportamento da capacidade do sistema em termos de carga servida para três diferentes proporções de carga oferecida entre VoIP e *Web*. É possível se observar que, quando a proporção de usuários VoIP é maior (75%V-25%W), a vazão do sistema é menor comparando a uma mesma carga oferecida (50%V-50%W) e um maior número de usuários *Web* (25%V-75%W). Isto se deve ao menor requerimento de banda dos usuários VoIP. Pode-se concluir que a inserção de VoIP ocasiona um decréscimo na carga servida porque eles ocupam janelas de transmissão com muito menos dados úteis. Tornam-se ainda mais importantes mecanismos que tentam aumentar a vazão do sistema, com o intuito de compensar pela diminuição de eficiência espectral ocasionada pela adição

de serviços VoIP, e ao mesmo tempo manter o QoS para ambos os serviços.

Como o tráfego de voz é o mais importante e popular nas redes de comunicação sem-fio, a distribuição considerada para a carga oferecida média nas avaliações do CAC é composta por 75% de usuários de voz em tempo-real (tráfego VoIP) e 25% de usuários interativos de tempo não-real (navegação *Web*).

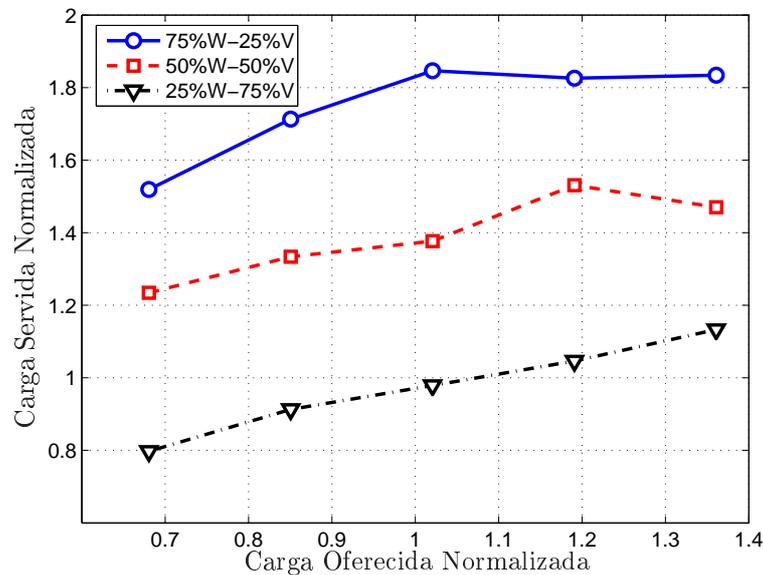
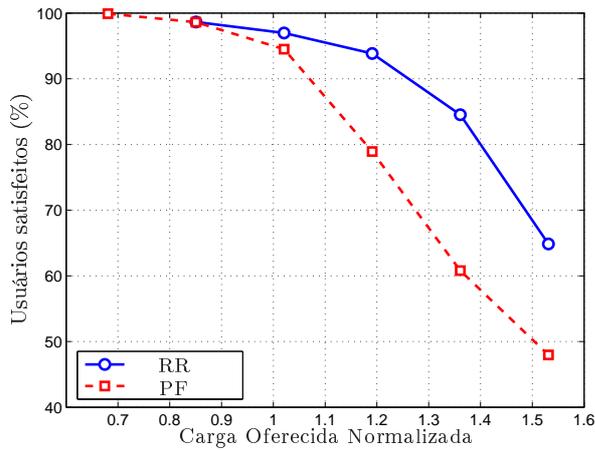
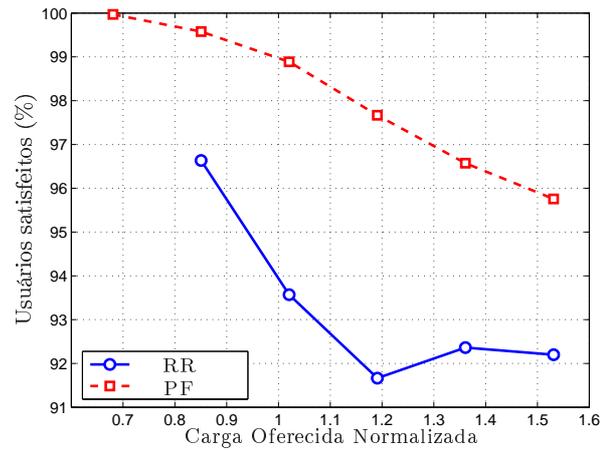


Figura 4.22: Carga servida em diferentes *mix* de serviços.

Nas Figuras 4.23(a) e 4.23(b), o desempenho do sistema é apresentado para o caso de referência (sem CAC) considerando ambos os escalonadores PF e RR. Do ponto de vista de QoS, o serviço VoIP limita a capacidade do sistema (90% de taxa de satisfação marca a limitação de carga total no sistema). O serviço *Web* não chega a atingir seu limite de QoS (90% de usuários satisfeitos) dentre os pontos simulados. Dessa maneira, parece bastante plausível o sacrifício da qualidade do serviço *Web* com o intuito de se melhorar o desempenho do serviço VoIP. Isto levaria a um aumento na capacidade total do sistema.



(a) Serviço VoIP.

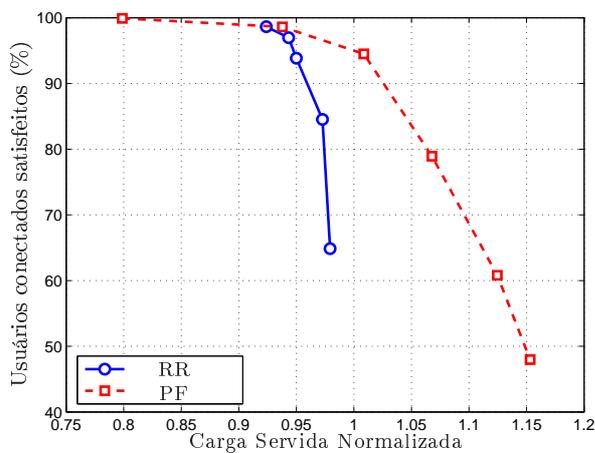


(b) Serviço Web.

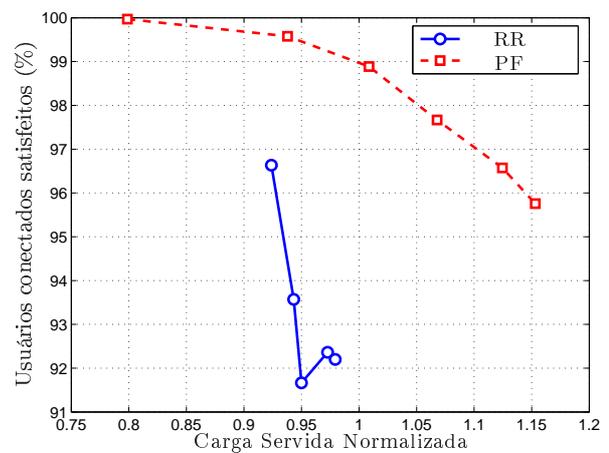
Figura 4.23: Satisfação de serviços VoIP e *Web* x Carga oferecida.

Ao se comparar os algoritmos de escalonamento, o RR resulta em um melhor desempenho para VoIP, mas é superado pelo PF quando os usuários *Web* são analisados. Assumindo que o serviço VoIP limita a capacidade, o RR proporciona uma maior capacidade em termos de carga oferecida.

Nas Figuras 4.24(a) e 4.24(b) é interessante observar o como a vazão no sistema é degradada pelo uso do escalonador RR. Utilizando-se deste ponto-de-vista, o PF proporciona um maior capacidade para a rede.



(a) Serviço VoIP.



(b) Serviço Web.

Figura 4.24: Satisfação de serviços VoIP e *Web* x Carga servida.

Análise de Desempenho

Os mesmos esquemas de CAC foram considerados, mas agora duas diferentes políticas são também avaliadas. Uma delas não considera nenhuma margem de prioridade no processo de admissão, i.e. ambos os serviços têm o mesmo limiar de admissão (55 ms). Dessa maneira, o CAC tenta proteger os usuários VoIP já conectados. Outra análise consistiu na adoção de um limiar de admissão menor para os usuários *Web*, utilizando de uma margem de prioridade de 10 ms. Assim, o limiar de admissão para usuários *Web* é igual a 45 ms. Todas as outras considerações deste cenário são as mesmas do caso com serviços simples descrito na Seção 4.3.1.

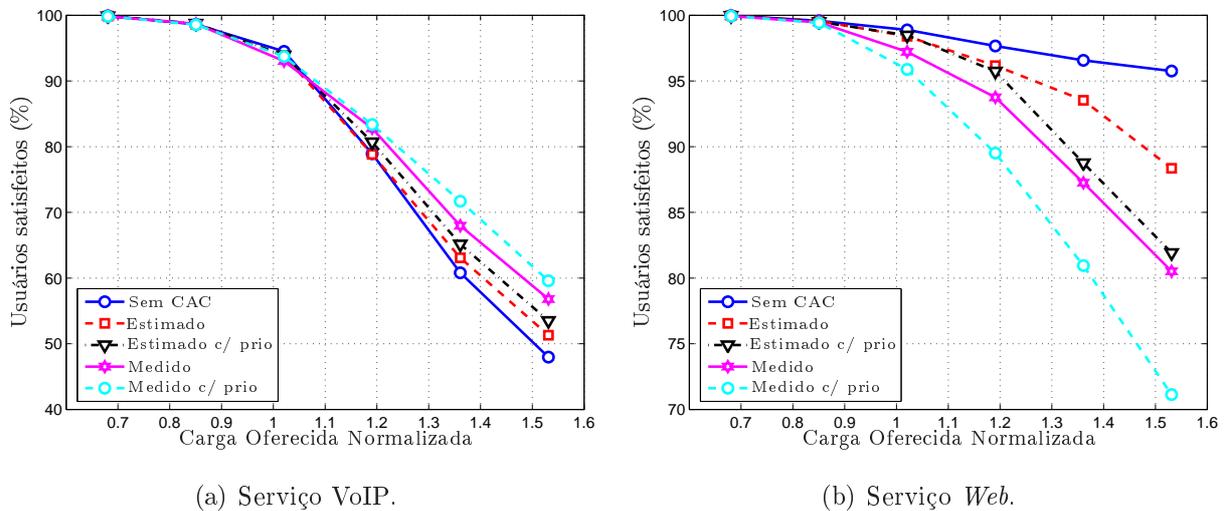


Figura 4.25: Satisfação de serviços VoIP e *Web* User utilizando PF.

O desempenho é apresentado nas Figuras 4.25(a) e 4.25(b) para os serviços VoIP e *Web* respectivamente, e utilizando o escalonador PF. É possível perceber, na Figura 4.25(a), que o CAC não aumenta a capacidade em termos de carga oferecida para VoIP, mas melhora a taxa da satisfação para altas cargas. Os melhoramentos são destacados quando a margem de prioridade é adotada. No entanto, é possível concluir que o CAC é capaz de mitigar os efeitos de situações de sobrecarga (até 10% de aumento na taxa de satisfação).

Quando o desempenho do tráfego *Web* é analisado na Figura 4.25(b), o bloqueio tem uma grande influência no aumento da insatisfação (até 25% de redução na taxa de satisfação). Quando a margem de prioridade é considerada, a satisfação dos usuários *Web* decresce ainda mais, devido às altas taxas de bloqueio impostas. De fato, ocorre um decréscimo na capacidade de carga oferecida para o serviço *Web*, mas considerando que o serviço VoIP é o limitante, não existe diminuição na

capacidade do sistema como um todo. De fato, a configuração do CAC provê um ajuste nas qualidades percebidas pelos serviços.

Quando apenas os usuários conectados são analisados, alguns melhoramentos podem ser melhor percebidos ao se utilizar do CAC. Na Figura 4.26(a), pode-se observar que um pequeno aumento na capacidade de cerca de 2% pode ser obtido ao se utilizar o CAC (considerando 90% de taxa de satisfação). No entanto, as melhorias na satisfação de usuário para altas cargas atinge cerca de 8% quando o CAC baseado em atraso medido e sem margem de prioridade é considerado.

O desempenho do serviço *Web*, no entanto, não é consideravelmente afetado pelo CAC. Pode-se observar na Figura 4.26(b) que, apesar do bloqueio imposto às novas chamadas, não existe uma redução considerável na carga servida do sistema, já que este serviço tem muito mais informação útil a ser servida pela rede. Na verdade, apenas sob altas cargas e utilizando os esquemas de CAC mais restritivos é que pode-se perceber uma diminuição na carga servida para um mesmo índice de satisfação.

A Figura 4.26(b) também mostra que, nas cargas simuladas utilizando o PF, a taxa de satisfação para os usuários de serviço *Web* admitidos permanece superior a 90%, significando que é possível prover algum tipo de balanceamento de carga para os usuários conectados através de algoritmos de escalonamento, que pode gerenciar o acesso ao meio. Isto é abordado no capítulo anterior.

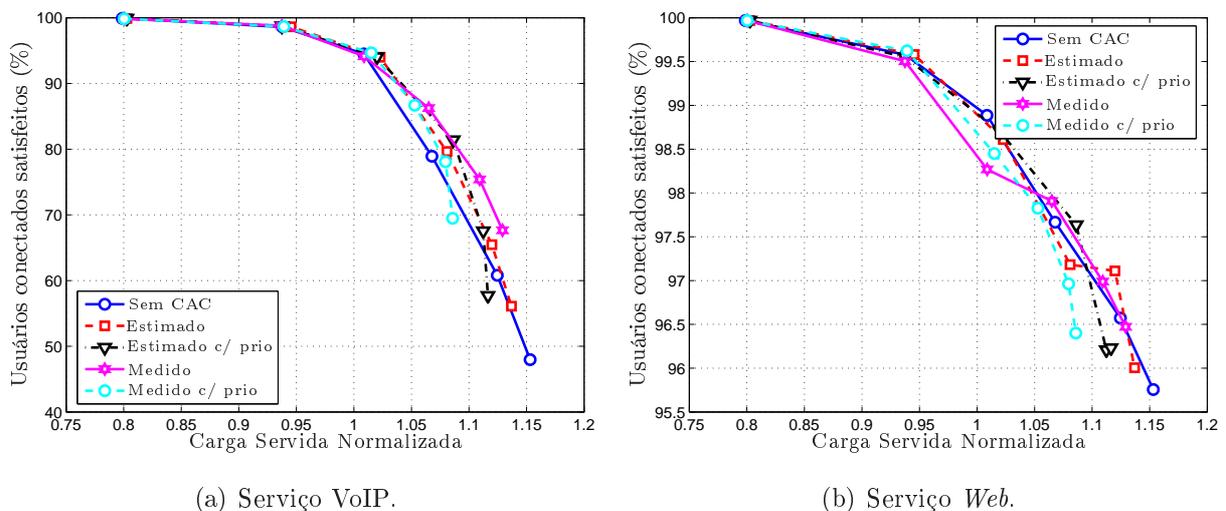


Figura 4.26: Satisfação de usuários VoIP e *Web* conectados utilizando PF.

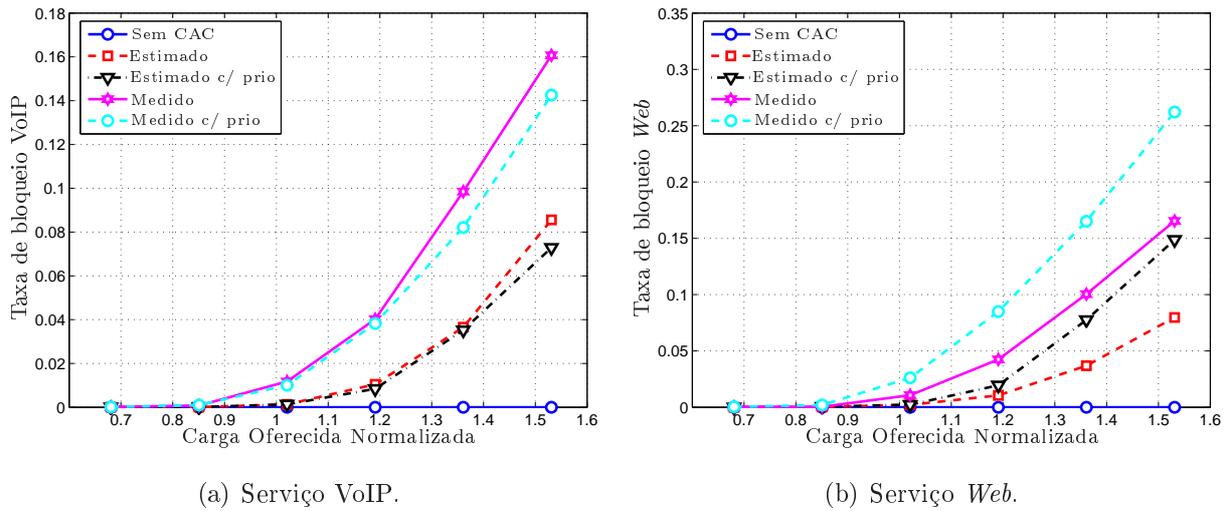


Figura 4.27: Taxas de bloqueio para serviços VoIP e *Web* utilizando PF.

Nas Figuras 4.27(a) e 4.27(b), as taxas de bloqueio para ambos os serviços são apresentadas. O CAC baseado em atraso medido se baseia em uma medida de uso de recurso de atraso mais pessimista, acabando por proteger mais os usuários VoIP. Quando nenhuma margem de prioridade é considerada para VoIP, as taxas de bloqueio para ambos os serviços são equivalentes. Mas quando uma margem é adotada, as taxas de bloqueio para o serviço *Web* são consideravelmente maiores. Esta limitação que melhora o desempenho para o serviço VoIP. A adoção da margem também leva a uma redução nas taxas de bloqueio para o serviço VoIP.

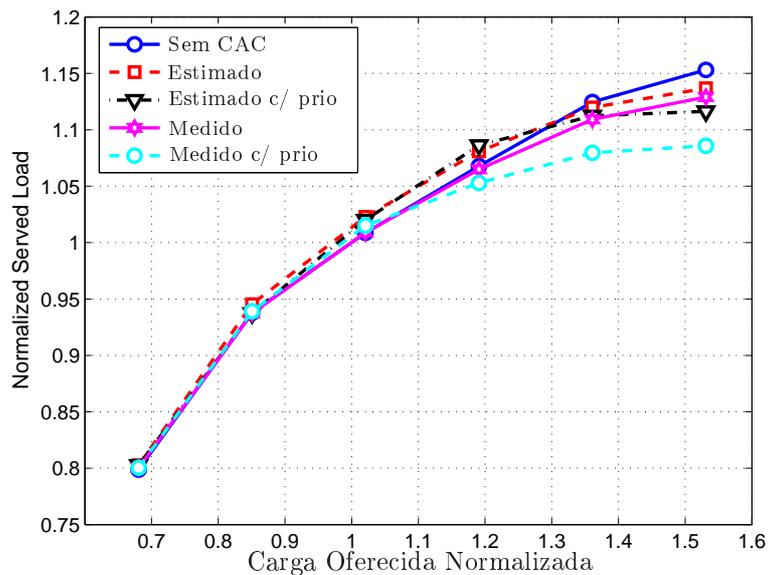


Figura 4.28: Efeito do controle de admissão na carga servida utilizando PF.

O efeito do CAC na vazão do sistema é apresentado na Figura 4.28. Isto mostra que, quanto mais prioridade é dada para o serviço VoIP, maior é a redução na vazão do sistema. Estabelece-se então uma relação custo-benefício entre a qualidade do serviço VoIP e a eficiência espectral desconsiderando os limites de QoS.

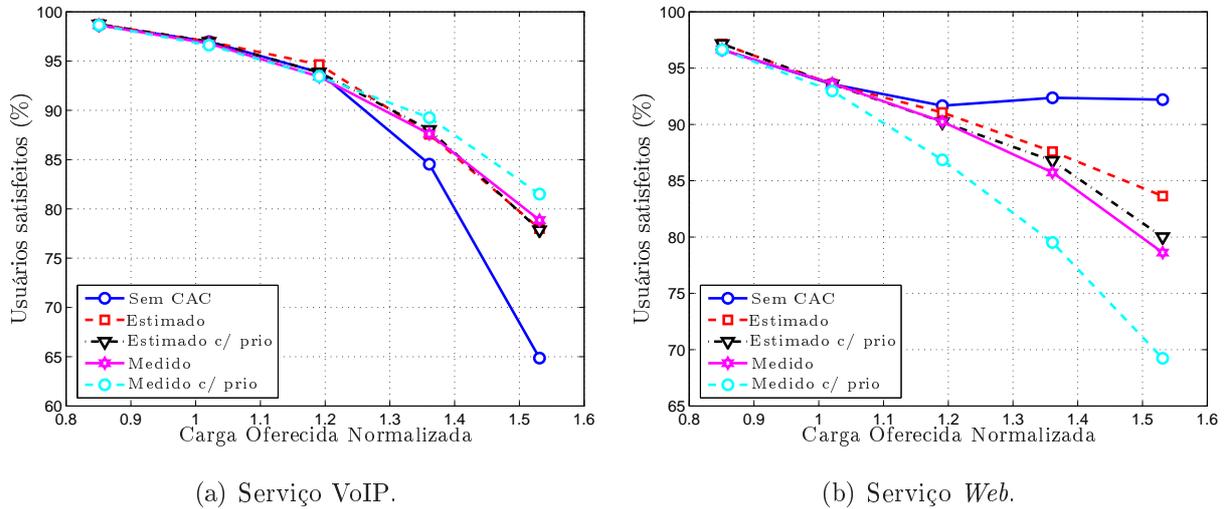


Figura 4.29: Satisfação de serviços VoIP e *Web* utilizando RR.

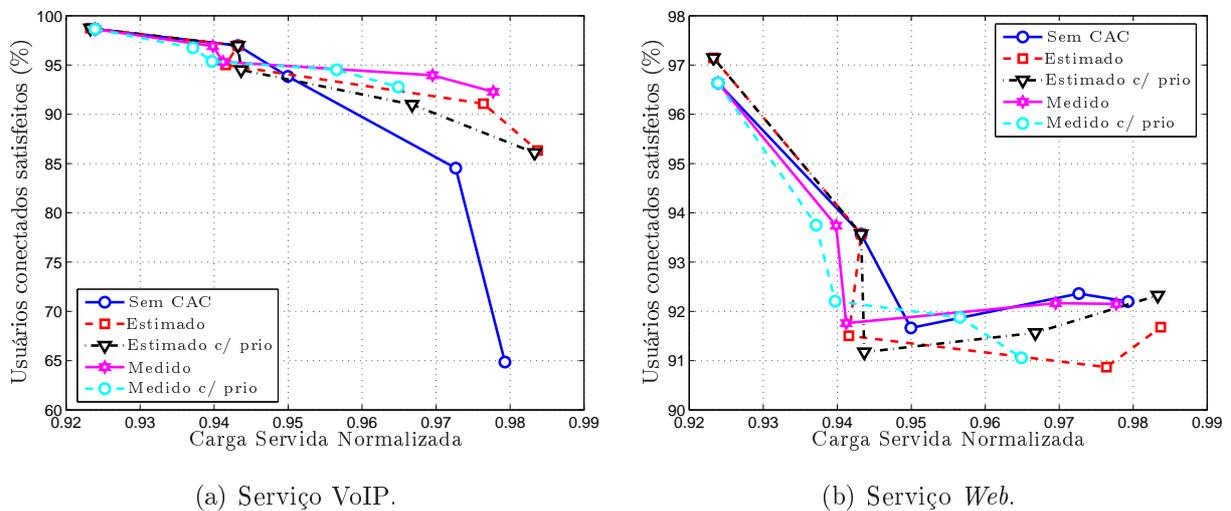


Figura 4.30: Satisfação de usuários VoIP e *Web* conectados utilizando RR.

Agora, considerando o RR, a Figura 4.29(a) mostra boas melhorias de desempenho para situações de sobrecarga (até cerca de 15% de aumento na taxa de satisfação é atingido) e um aumento de capacidade VoIP de cerca de 5%. O problema também recai no fato de que o desempenho de *Web* utilizando o RR não é satisfatório, e o bloqueio imposto degrada ainda mais o serviço, o que mostra a

Figura 4.29(b). O CAC acaba por tornar o serviço *Web* no limitador de capacidade do sistema causando, inclusive, uma pequena redução de capacidade total no sistema.

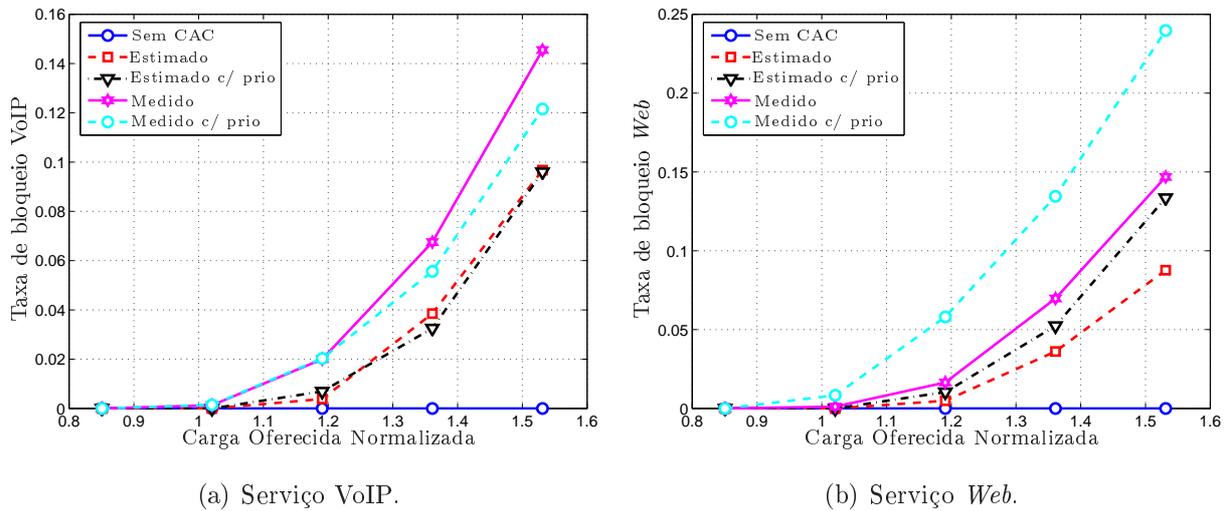


Figura 4.31: Taxas de bloqueio para serviços VoIP e *Web* utilizando RR.

As Figuras 4.30(a) e 4.30(b) mostram o desempenho de satisfação dos usuários conectados. Em termos de carga servida, o CAC é capaz de melhorar a taxa de satisfação de maneira mais eficiente. Pois cerca de 2% de aumento de capacidade pode ser obtido, sendo o serviço VoIP o limitador de capacidade. Isto se deve ao fato de o RR distribuir igualmente os recursos entre todos os usuários, com a desvantagem de diminuir a eficiência da rede, já que não considera as condições de canal.

Nas Figuras 4.31(a) e 4.31(b), as taxas de bloqueio são analisadas. Pode-se observar que o bloqueio, quando considerando a margem de prioridade, aumenta para o serviço *Web* e diminui para VoIP, porque a política de admissão se torna mais restrita para o serviço *Web*.

Os resultados na Figura 4.32 com relação ao efeito do CAC na carga servida têm o mesmo comportamento do caso com o PF. Como mais acesso é dado ao serviço VoIP, a eficiência espectral é diminuída devido ao menor requerimento de taxa de bits para este serviço. A redução na carga servida é ainda maior ao se comparar ao caso PF.

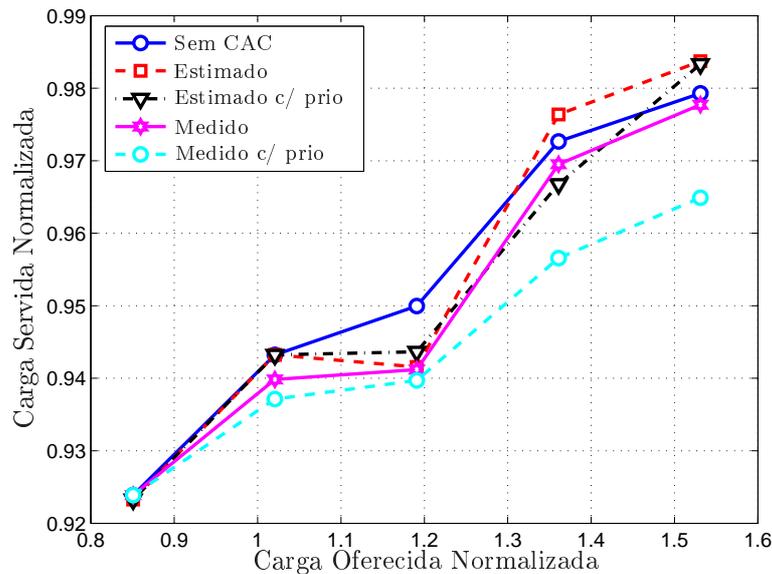


Figura 4.32: Efeito do controle de admissão na carga servida utilizando RR.

4.4 Sumário

Neste capítulo, foi proposto um esquema de controle de admissão baseado em uma métrica chamada de recurso de atraso. Dois métodos de indicação desta métrica foram também apresentados:

- ▶ Medição - utilizando um filtro capaz de rastrear altos percentis de amostras de pacotes entregues e descartados.
- ▶ Estimação - baseado em uma medida padronizada de taxa de bits provida e o tamanho do *buffer* de transmissão.

O esquema foi avaliado em dois cenários contendo o serviço VoIP. O primeiro não apresentava o serviço de navegação *Web*, enquanto que o segundo sim.

O gráfico de barras na Figura 4.33 apresenta os valores de capacidades, para efeito de comparação. O CAC estimado resulta nos maiores valores de capacidade, principalmente quando o ADS é empregado.

Quando a capacidade em termos de carga servida é analisada, como mostra a Figura 4.34, os ganhos do CAC não são evidentes. Isto se deve, em grande parte, de o CAC não permitir que o sistema atinja o limite de QoS, o que impede de a capacidade de carga servida ser estabelecida. Tal fato ocorre para os casos dos escalonadores RR e ADS, justamente os que apresentam os melhores desempenhos no cenário de serviços simples.

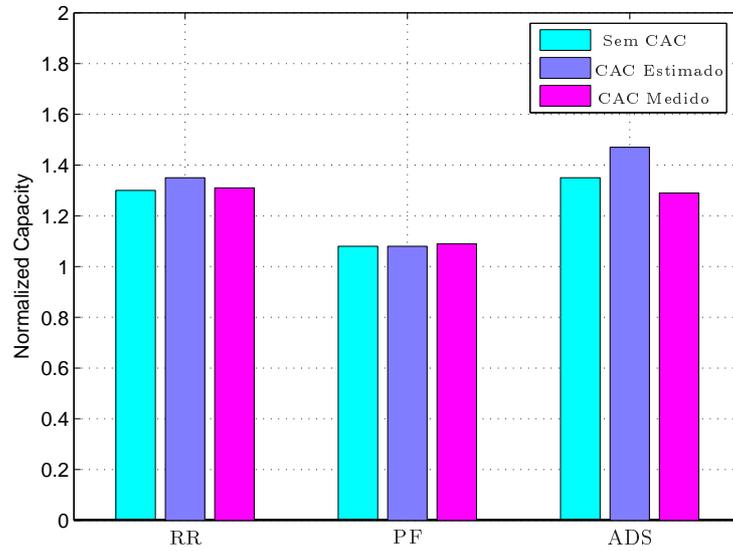


Figura 4.33: Capacidade relativa de carga oferecida empregando.

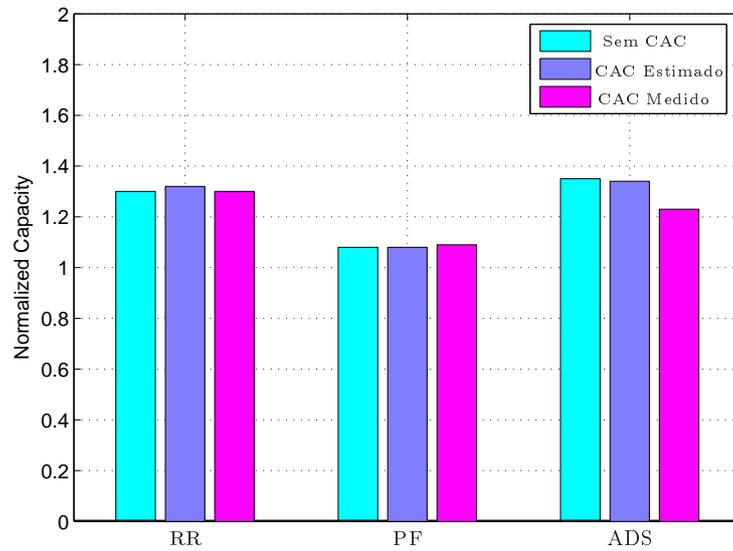


Figura 4.34: Capacidade relativa de carga servida empregando.

Na Figura 4.35, os resultados referentes ao cenário composto por ambos os serviços de VoIP e *Web* são mostrados em termos de capacidade de carga oferecida. Pode-se perceber que a utilização do CAC ocasiona uma diminuição de capacidade para o caso em que o RR é empregado. Isto se deve à ineficiência deste algoritmo em prover QoS para usuários *Web*, além da priorização estabelecida ao serviço VoIP através do CAC. Desta maneira, o serviço *Web* se torna o limitante de capacidade.

Com relação ao PF, o CAC não afeta de forma significativa a capacidade do sistema, apesar de que, como visto no decorrer deste capítulo, nesse cenário o CAC proporciona o melhor desempenho em termos de mitigação dos efeitos de sobrecarga no sistema.

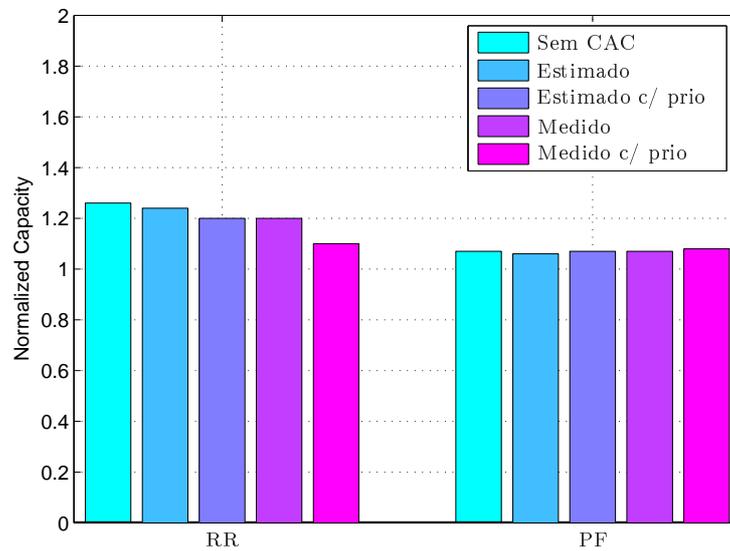


Figura 4.35: Capacidade relativa de carga oferecida empregando.

Quando a capacidade em termos de carga servida é analisada na Figura 4.36, nenhum efeito é percebido na capacidade em ambos os escalonadores.

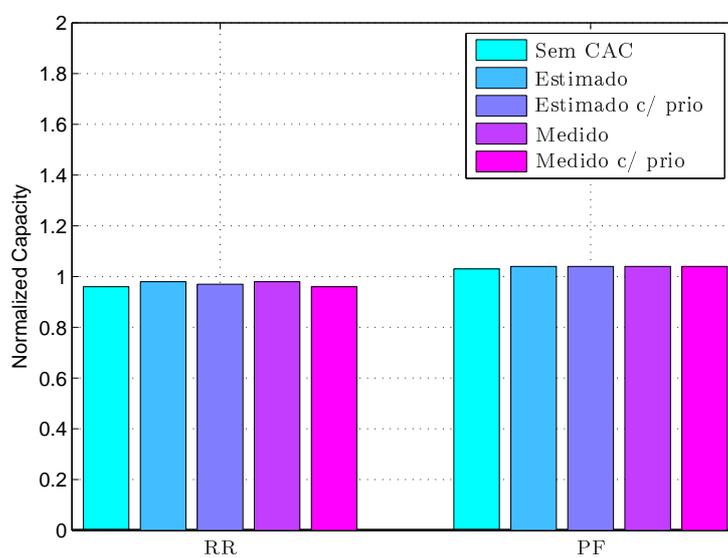


Figura 4.36: Capacidade relativa de carga servida empregando.

Conclusões e Perspectivas

5.1 Conclusões

A maior contribuição deste trabalho consiste em uma completa avaliação da capacidade de alguns algoritmos de controle de congestionamento em prover QoS para o serviço VoIP em diferentes cenários de tráfego onde os recursos são compartilhados com usuários *Web*. O QoS do serviço *Web* também foi considerado no cálculo dos limites de capacidade.

Foram avaliados dois tipos de algoritmos de controle de congestionamento: escalonamento de pacotes e controle de admissão.

Com relação ao estudo de escalonamento, foram propostos alguns algoritmos divididos em dois conjuntos: os com QoS-diferenciado (ADS, LDS e WPF) e QoS-não-diferenciado (RR, PF e QBMC). A avaliação foi efetuada considerando um cenário simples onde apenas o serviço VoIP existe como em cenários mistos com diferentes proporções de usuários VoIP e *Web*. Ao final, curvas de regiões de capacidade foram traçadas.

Pode-se concluir que, apesar de o algoritmo PF ter obtido resultados bastante ruins devido a limitações estabelecidas pelo serviço VoIP, os algoritmos que obtiveram os melhores resultados são, de alguma forma, algum tipo de variação deste algoritmo. O ADS e LDS proporcionaram os melhores resultados em todas as proporções entre os serviços. Quando apenas o serviço VoIP é oferecido ao sistema, eles proporcionaram grandes aumentos de capacidade comparados a algoritmos clássicos como PF e RR. Quando a proporção de tráfego *Web* é aumentada no sistema, eles ainda são capazes de prover um bom QoS para ambos os serviços, levando a uma maior capacidade conjunta.

O QBMC proporciona uma considerável melhora no desempenho quando apenas usuários VoIP são oferecidos ao sistema, mas um desempenho muito fraco quando usuários *Web* são inseridos na rede.

O WPF não resulta em um bom desempenho como o ADS ou LDS, principalmente se existe uma maior proporção de usuários VoIP, devido a uma limitação na capacidade imposta pelo serviço *Web*. No entanto, ele se sai bem melhor que o PF puro, tornando-se uma solução simples para a diferenciação de serviços.

O esquema de controle de admissão proposto fazia uso de uma métrica de utilização do recurso de atraso. Tal recurso pode ser estabelecido de duas maneiras: através de medidas de atraso ou através de uma estimativa do atraso. Ambos os esquemas foram comparados com uma configuração sem controle de admissão. Assim como os algoritmos de escalonamento, o controle de admissão foi avaliado em um cenário composto por apenas usuários VoIP e se utilizando de três algoritmos de escalonamento (RR, PF e ADS). Além de um cenário misto de serviço com uma proporção fixa entre usuários VoIP e *Web*, e empregando dois tipos de esquemas de escalonamento (RR e PF).

O cenário simples mostrou que o controle de admissão proporcionou melhorias tímidas de capacidade quando os algoritmos RR e PF foram empregados. Quando o ADS é utilizado, as melhorias de capacidade foram notáveis.

Já com o cenário de serviços mistos, pode-se concluir que o controle de admissão é capaz de prover uma priorização de serviço bastante efetiva, já que o serviço *Web* é bastante robusto.

O desempenho dos esquemas de admissão por atraso medido ou estimado forma bastante próximos. O primeiro proporciona uma maior restrição à entrada de novos usuários no sistema, ou seja, uma maior proteção aos usuários conectados e, conseqüentemente, um maior bloqueio. Com relação ao esquema de admissão por atraso estimado, ele se saiu, em alguns casos inclusive, melhor devido a um melhor compromisso entre QoS e bloqueio. Com a vantagem de não requerer nenhuma sinalização adicional na interface *Iub*.

5.2 Perspectivas

As principais perspectivas para a continuidade deste trabalho são:

- ▶ a proposta e avaliação de outros algoritmos de escalonamento;

- ▶ avaliação desses algoritmos em cenários com existência de tráfego na parte com chaveamento por circuito da rede UTRAN. Este cenário mostra-se bastante interessante e realístico do ponto de vista do mercado atual 3G;
- ▶ uma continuação lógica deste trabalho é uma avaliação mais extensiva do algoritmo de controle de admissão combinado com escalonadores avançados e com diversos modelos de tráfego.

Referências Bibliográficas

- [1] UMTS, “Selection Procedures for the Choice of Radio Transmission Technologies of the UMTS,” ETSI, UMTS TR 101.112 v.3.2.0, Relatório técnico, Abril 1998.
- [2] Y.-T. Wang, D. J. Houck e H. Uzunalioglu, “Call Admission Control (CAC) for Quality of Service in Voice over Internet (VoIP) Networks,” *Lucent Technologies White Paper*, Setembro 2005.
- [3] U. Olsson, “Toward the all-IP Vision,” *Ericsson Review*, vol. 82, no. 1, pp. 44–53, 2005.
- [4] P. Hosein, “Scheduling of VoIP Traffic over a Time-shared Wireless Packet Data Channel,” in *IEEE International Conference on Personal Wireless Communications (ICPWC)*, Janeiro 2005, pp. 38–41.
- [5] M. Kazmi, P. Godlewski e C. Cordier, “Admission Control Strategy and Scheduling Algorithms for Downlink Packet Transmission in WCDMA,” in *IEEE Vehicular Technology Conference, Fall*, vol. 2, Setembro 2000, pp. 674–680.
- [6] P. Hosein, “A Class-based Admission Control Algorithm for Shared Wireless Channels Supporting QoS Services,” in *5th IFIP TC6 International Conference on Mobile and Wireless Communications Networks, Singapore*, Outubro 2003.
- [7] H. Holma e A. Toskala, *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*, 3ª edição. John Wiley & Sons, Ltd, 2004.
- [8] M. Ericson, L. Voigt e S. Wänstedt, “Providing Reliable and Efficient VoIP over WCDMA,” *Ericsson Review*, vol. 82, no. 2, pp. 110–117, 2005.

- [9] 3GPP, “High Speed Downlink Packet Access (HSDPA): Overall Description,” 3rd Generation Partnership Project, Sophia Antipolis, France, Relatório técnico TS 25.308 v6.3.0 - Release 6, Dezembro 2004. [Online]. Disponível: <http://www.3gpp.org>
- [10] —, “Radio Link Control (RLC) Protocol Specification,” 3rd Generation Partnership Project, Sophia Antipolis, France, Relatório técnico TS 25.322 v6.2.0 - Release 6, Dezembro 2004. [Online]. Disponível: <http://www.3gpp.org>
- [11] —, “Medium Access Control (MAC) Protocol Specification,” 3rd Generation Partnership Project, Sophia Antipolis, France, Relatório técnico TS 25.321 v6.3.0 - Release 6, Janeiro 2005. [Online]. Disponível: <http://www.3gpp.org>
- [12] —, “Physical Layer Procedures (FDD),” 3rd Generation Partnership Project, Sophia Antipolis, France, Relatório técnico TS 25.214 v6.4.0 - Release 6, Janeiro 2005. [Online]. Disponível: <http://www.3gpp.org>
- [13] —, “User Equipment (UE) Radio Transmission and Reception (FDD),” 3rd Generation Partnership Project, Sophia Antipolis, France, Relatório técnico TS 25.101 v7.2.0 - Release 7, Dezembro 2005. [Online]. Disponível: <http://www.3gpp.org>
- [14] —, “Radio Resource Management Strategies,” 3rd Generation Partnership Project, Sophia Antipolis, França, Relatório técnico TR 25.922 v6.2.0 - Release 6, Outubro 2005. [Online]. Disponível: <http://www.3gpp.org>
- [15] 3GPP, “Physical Layer Aspects of UTRA High Speed Downlink Packet Access,” 3GPP, Sophia Antipolis, France, Relatório técnico TS 25.848 v4.0.0 - Release 4, Março 2001. [Online]. Disponível: <http://www.3gpp.org>
- [16] —, “UTRAN High Speed Downlink Packet Access,” 3GPP, Sophia Antipolis, France, Relatório técnico TS 25.950 v4.0.0 - Release 4, Março 2001. [Online]. Disponível: <http://www.3gpp.org>
- [17] 3GPP, “Deviation from Delay Requirements,” 3rd Generation Partnership Project, Sophia Antipolis, France, Relatório técnico TSG-RAN WG3 meeting 37 R3-031011, Agosto 2003. [Online]. Disponível: <http://www.3gpp.org>

- [18] —, “Multiplexing and channel coding (FDD),” 3rd Generation Partnership Project, Sophia Antipolis, France, Relatório técnico TS 25.212 V6.7.0 - Release 6, Dezembro 2005. [Online]. Disponível: <http://www.3gpp.org>
- [19] 3GPP, “Proposal of Bit Mapping for Type-III HARQ,” 3GPP, 3GPP TSG RAN 18, Relatório técnico, Janeiro 2001. [Online]. Disponível: <http://www.3gpp.org>
- [20] S. Wänstedt, F. Rui, M. Ericson e M. Nordberg, “Providing Reliable and Efficient VoIP over Cellular Networks,” in *Future Telecommunications Conference*, Outubro 2005.
- [21] N. Lithhammer, “Speech on HS-DSCH,” Dissertação de mestrado, Kungliga Tekniska Högskolan (KTH), Stockholm, Sweden, Dezembro 2003.
- [22] B. Wang, K. I. Pedersen, T. E. Kolding e P. E. Mogesen, “Performance of VoIP on HSDPA,” in *IEEE Vehicular Technology Conference, Spring*, Maio 2005.
- [23] H. Zheng, G. Rittenhouse e M. Recchione, “The Performance of Voice over IP over 3G Downlink Shared Packet Channels under Different Delay Budgets,” *IEEE Global Telecommunications Conference - GLOBECOM*, pp. 2501–2505, Outubro 2003.
- [24] W. Sun e D. Liu, “Call Admission Control: A Survey for Effective Bandwidth-Based Approach and Measurement-Based Approach,” Novembro 1998. [Online]. Disponível: <http://citeseer.ist.psu.edu/108954.html>
- [25] K. H. Yum, E. J. Kim, C. R. Das e M. Yousif, “Integrated Admission and Congestion Control for QoS Control in Clusters,” in *IEEE International Conference on Cluster Computing*, Setembro 2002, pp. 325–332.
- [26] A. M. Jr. e J. M. A. Coello, “Packet Scheduling Based on Learning in the Next Generation Internet Architectures,” in *5th IEEE Symposium on Computers and Communications*, Julho 2000, pp. 773–778.
- [27] 3GPP, “Performance Characterization of the Adaptive Multi-Rate (AMR) Speech Codec,” 3rd Generation Partnership Project, Sophia Antipolis, France, Relatório técnico TR 26.975 v6.0.0 - Release 6, Dezembro 2004. [Online]. Disponível: <http://www.3gpp.org>

- [28] E. B. Rodrigues, “Admission Control for Streaming Services over HSDPA,” Ericsson Research / GTEL, Relatório técnico, 2004.
- [29] Y. Cao e V. O. K. Li, “Scheduling Algorithms in Broad-Band Wireless Networks,” *Proceedings of the IEEE*, pp. 76–87, Janeiro 2001.
- [30] P. J. A. Gutiérrez, “Packet Scheduling and Quality of Service in HSDPA,” Tese de doutorado, Aalborg University, 2003.
- [31] P. Falconio e P. Dini, “Design and Performance Evaluation of Packet Scheduling Algorithms for Video Traffic in the High Speed Downlink Packet Access,” in *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2004. PIMRC 2004*, vol. 3, Setembro 2004, pp. 2202–2206.
- [32] W. Jeon, D. G. Jeong e B. Kim, “Packet Scheduler for Mobile Internet Services using High Speed Downlink Packet Access,” *IEEE Transactions on Wireless Communications*, vol. 3, pp. 1789–1801, Setembro 2004.
- [33] P. Hosein, “QoS Control for WCDMA High Speed Packet Data,” in *4th IEEE Conference on Mobile and Wireless Communications*, Setembro 2002.
- [34] J. Zander, S.-L. Kim, M. Almgren e O. Queseth, *Radio Resource Management for Wireless Networks*, ser. Mobile Communications Series. London: Artech House Publishers, 2001.
- [35] P. Cardieri, “Resource allocation and adaptive antennas in cellular communications,” Tese de doutorado, Virginia Polytechnic Institute and State University, Blacksburg, EUA, Setembro 2000.
- [36] H. W. Arnold, D. C. Cox e R. R. Murray, “Macroscopic diversity performance measured in the 800-MHz portable radio communications environment,” *IEEE Transactions Antennas and Propagation*, pp. 277–280, 1988.
- [37] W. C. Jakes, *Microwave Mobile Communications*. John Wiley & Sons, Ltd, 1994.
- [38] P. I. Bratanov, “User mobility modeling in cellular communications networks,” Tese de doutorado, Technischen Universität Wien, Fevereiro 1999.

- [39] T. Camp, J. Boleng e V. Davies, “A survey of mobility models for ad hoc network research,” *Wireless Communications and Mobile Computing (WCMC)*, vol. 2, no. 5, pp. 483–502, 2002.
- [40] 3GPP, “Packet Switched Conversational Multimedia Applications: Default Codecs,” 3rd Generation Partnership Project, Sophia Antipolis, France, Relatório técnico TS 26.235 v7.0.0 - Release 7, Dezembro 2005. [Online]. Disponível: <http://www.3gpp.org>
- [41] IETF, “RTP: A transport protocol for real-time applications,” The Internet Engineering Task Force, Relatório técnico RFC 3550, 2003.
- [42] —, “User datagram protocol,” The Internet Engineering Task Force, Relatório técnico RFC 768, 1980.
- [43] —, “Internet protocol,” The Internet Engineering Task Force, Relatório técnico RFC 791, 1981.
- [44] —, “Robust Header Compression (ROHC),” The Internet Engineering Task Force, Relatório técnico RFC 3095, 2001.
- [45] A. Furuskär, “Radio Resource Sharing and Bearer Service Allocation for Multi-Bearer Service, Multi-Access Wireless Networks,” Tese de doutorado, Kungliga Tekniska Högskolan (KTH), 2003.
- [46] 3GPP, “UTRAN Iub Interface NBAP Signalling,” 3rd Generation Partnership Project, Sophia Antipolis, France, Relatório técnico TS 25.433 v6.4.0 - Release 6, Janeiro 2005. [Online]. Disponível: <http://www.3gpp.org>
- [47] K. Hiltunen, M. Lundevall e S. Magnusson, “Performance of Link Admission Control in a WCDMA System with HS-DSCH and Mixed Services,” in *15th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications - PIMRC*, vol. 2, Setembro 2004, pp. 1178–1182.
- [48] 3GPP, “Information to be Signalled for HSDPA Call Admission Control and Congestion Control,” 3rd Generation Partnership Project, Sophia Antipolis, France, Relatório técnico TSG-RAN WG3 meeting 35 R3-030485, Abril 2003. [Online]. Disponível: <http://www.3gpp.org>

- [49] M. Hata, “Empirical Formula for Propagation Loss in Land Mobile Radio Services,” in *IEEE Vehicular Technology Conference, Spring*, vol. VT-29, Setembro 1980, pp. 317–325.