

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

EDUARDO MAROÑAS MONKS

**Planejamento de Capacidade em Redes
Corporativas para Implementação de
Serviços VoIP**

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

Prof. Dr. Antônio Carlos da Rocha Costa
Orientador

Prof. Dr. Juergen Rochol
Co-orientador

Porto Alegre, maio de 2006

CIP – CATALOGAÇÃO NA PUBLICAÇÃO

Monks, Eduardo Maroñas

Planejamento de Capacidade em Redes Corporativas para Implementação de Serviços VoIP / Eduardo Maroñas Monks. – Porto Alegre: PPGC da UFRGS, 2006.

115 f.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2006. Orientador: Antônio Carlos da Rocha Costa; Co-orientador: Juergen Rochol.

1. VoIP. 2. QoS. 3. Planejamento de Capacidade. I. Rocha Costa, Antônio Carlos da. II. Rochol, Juergen. III. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. José Carlos Ferraz Hennemann

Vice-Reitor: Prof. Pedro Cezar Dutra Fonseca

Pró-Reitora de Pós-Graduação: Prof^a. Valquíria Linck Bassani

Diretor do Instituto de Informática: Prof. Philippe Olivier Alexandre Navaux

Coordenador do PPGC: Prof. Flávio Rech Wagner

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

*“ I have always wished that my computer would be as easy to use as my
telephone.*

My wish has come true.

I no longer know how to use my telephone.

— PROFESSOR BJARNE STROUSTRUP

[HTTP://WWW.RESEARCH.ATT.COM/ BS/HOMEPAGE.HTML](http://www.research.att.com/bs/homepage.html)

AGRADECIMENTOS

Este trabalho só se tornou possível graças a oportunidade a mim dada pelo professor Antônio Carlos da Rocha Costa. Muito obrigado e jamais esquecerei, professor Rocha.

Agradeço ao professor Juergen Rochol, pela amizade, convívio e pela orientação. Ao professor Lisandro Granville, pela força e a oportunidade de utilizar o Labcom. E a todos os professores com os quais tive contato e aos funcionários do instituto que sempre se mostraram prestativos e atenciosos.

Aos colegas Weldson, Evandro, Michele, Oscar Mori, José Antônio, Oscar, Aurélio, Rodrigo Sanger e Jamile pelo convívio, amizade e força em todos os momentos.

A UCPel (Universidade Católica de Pelotas), em especial, a Paula Pruski Yamim, minha chefe e segunda mãe.

A minha família, especialmente a meus pais, Pedro e Maria, que não mediram esforços para eu conseguir terminar este trabalho. A eles, dedico esta dissertação.

Um agradecimento especial a minha mulher, Vanessa, que sempre esteve do meu lado, desde as primeiras tentativas de contato com o instituto. Uma pessoa que foi compreensiva e companheira, que me ajudou nos momentos mais difíceis. A ela, a qual soube entender a minha ausência, as minhas preocupações e frustrações. A ela, dedico o meu título.

SUMÁRIO

LISTA DE ABREVIATURAS E SIGLAS	8
LISTA DE FIGURAS	10
LISTA DE TABELAS	12
RESUMO	13
ABSTRACT	14
1 INTRODUÇÃO	15
1.1 Objetivos	16
1.2 Organização	16
2 VOZ SOBRE IP (VOIP)	17
2.1 Telefonia Convencional	17
2.1.1 Comunicação Telefônica	18
2.2 Breve Histórico de VoIP	21
2.3 Fundamentos	22
2.3.1 Tipos de Redes	23
2.4 Padrões e protocolos	25
2.4.1 Protocolos de Sinalização	26
2.4.2 Protocolos de Mídia	35
2.5 Codificadores (Codecs)	37
2.6 Arquiteturas	39
2.6.1 Ponto a Ponto	40
2.6.2 Com gateway	40
2.6.3 Híbrida	40
2.7 Equipamentos	40
2.7.1 Terminais	40
2.7.2 Centrais	42
2.7.3 Gateways	42
2.8 Numeração	42
2.9 Exemplos de Aplicação	44
2.9.1 Skype	44
2.9.2 Gizmo Project	45

3	QUALIDADE DE SERVIÇO (QOS)	47
3.1	QoS em Telefonia Convencional	47
3.1.1	Taxa de bloqueio (GOS)	47
3.1.2	Disponibilidade	48
3.1.3	Metas de Qualidade - ANATEL (Agência Nacional de Telecomunicações)	48
3.2	Fatores de QoS	51
3.2.1	Atraso	51
3.2.2	Variação do Atraso (<i>Jitter</i>)	54
3.2.3	Perda de Pacotes	54
3.3	Fatores que afetam QoS	56
3.3.1	Eco	56
3.3.2	Tráfego Concorrente	57
3.3.3	Convergência de Rotas	57
3.4	Garantias de QoS	57
3.4.1	Equipamentos	57
3.4.2	Mecanismos de QoS Nível 2 e 3	57
3.4.3	Buffer de <i>Jitter</i> (Equalizador de <i>Jitter</i>)	62
3.4.4	Redundância	62
3.4.5	Fragmentação de pacotes	63
3.4.6	Compressão de cabeçalhos	63
3.4.7	Alimentação de Energia	64
3.4.8	Planejamento de Capacidade da Rede	64
3.5	Formas de Medição da Qualidade da Voz	64
3.5.1	MOS - (<i>Mean Opinion Score</i>)	65
3.5.2	Modelo E	66
4	PLANEJAMENTO DE CAPACIDADE EM REDES PARA SERVIÇOS VOIP	68
4.1	Planejamento de Capacidade em Redes	68
4.1.1	Fases do Planejamento de Capacidade	71
4.2	Métodos de Planejamento de Capacidade em Telefonia Convencional	72
4.2.1	Breve Histórico	72
4.2.2	Métodos de Erlang	73
4.2.3	Fórmulas de Teletráfego	75
4.2.4	Fórmula de Erlang B	77
4.2.5	Metodologia	78
4.3	Aplicação da Metodologia	79
4.3.1	Coleta de Tráfego Telefônico	79
4.3.2	Cálculo da Hora de Maior Movimento	80
4.3.3	Definição do GoS	81
4.3.4	Aplicação da fórmula de Erlang B	82
4.4	Metodologia Adaptada para VoIP	82
4.4.1	Tipo de <i>Codec</i>	83
4.4.2	Amostras por Pacote	84
4.4.3	Supressão de Silêncio	84
4.4.4	Compressão de Cabeçalhos	85
4.5	Recursos de Rede	85
4.5.1	Largura de Banda	86

4.5.2	Mecanismos de QoS	87
4.5.3	Metodologia Adaptada para VoIP	87
5	ESTUDO DE CASO	88
5.1	Cenário	88
5.2	Rede de Telefonia	88
5.3	Rede de Dados	88
5.3.1	Tráfego de Rede	90
5.3.2	Mecanismos de QoS	90
5.4	Planejamento de Capacidade da Rede	91
5.4.1	Cálculo de HMM	91
5.4.2	Definição do GoS	92
5.4.3	Aplicação da Fórmula de Erlang B	92
5.4.4	Tipo de <i>Codec</i>	93
5.4.5	Número de Amostras por pacote	93
5.4.6	Supressão de Silêncio	93
5.4.7	Compressão de Cabeçalhos	93
5.4.8	Cálculo da Largura de Banda	94
5.5	Medições	95
5.5.1	Ferramentas	95
5.5.2	Procedimentos	96
5.5.3	Ambiente de Testes	98
5.6	Resultados	98
5.6.1	Análise dos Resultados	100
6	CONCLUSÃO	104
	REFERÊNCIAS	106
	APÊNDICE A GRÁFICOS POR SEGMENTO	112

LISTA DE ABREVIATURAS E SIGLAS

ACELP	<i>Algebraic Code Excited Linear Prediction</i>
ADSL	<i>Assymetric Digital Subscriber Line</i>
ANATEL	Agência Nacional de Telecomunicações
ATA	<i>Analog Telephone Adapter</i>
ATM	<i>Asynchronous Transfer Mode</i>
BGP	<i>Border Gateway Protocol</i>
CELP	<i>Code Excited Linear Prediction</i>
CRC	<i>Cyclic Redundancy Check</i>
CoS	<i>Class of Services</i>
DHCP	<i>Dynamic Host Control Protocol</i>
DNS	<i>Domain Naming System</i>
DSL	<i>Digital Subscriber Line</i>
DSP	<i>Digital Signal Processor</i>
DTMF	<i>Dual-Tone Multiple Frequency</i>
DiffServ	<i>Differentiated Services</i>
ENUM	<i>Eletronic Number</i>
GOS	<i>Grade of Service</i>
HTTP	<i>Hyper Text Transfer Protocol</i>
IETF	<i>Internet Engineering Task Force</i>
ITU	<i>International Telecommunication Union</i>
LAN	<i>Local Area Network</i>
MEGACO	<i>Media Gateway Control</i>
MGCP	<i>Media Gateway Control Protocol</i>
MIME	<i>Multipurpose Internet Mail Extensions</i>
MRTG	<i>Multi Router Traffic Grapher</i>
NAPTR	<i>Naming Authority Pointer</i>

OSI	<i>Open Systems Interconnection</i>
P2P	<i>Peer to Peer</i>
PABX	<i>Private Automatic Branch Exchange</i>
PBX	<i>Private Branch Exchange</i>
PCM	<i>Pulse Code Modulation</i>
PGQM	Plano Geral de Metas de Qualidade
PPP	<i>Point to Point Protocol</i>
PSTN	<i>Public Switched Telephone Network</i>
QoS	<i>Quality of Service</i>
RFC	<i>Request for Comments</i>
RNP	Rede Nacional de Pesquisa
RPTC	Rede Pública de Telefonia Comutada
RTCP	<i>Real Time Transport Control Protocol</i>
RTP	<i>Real Time Transport Protocol</i>
SCTP	<i>Streaming Control Transmission Protocol</i>
SDP	<i>Session Description Protocol</i>
SIP	<i>Session Initiation Protocol</i>
SMTP	<i>Simple Mail Transfer Protocol</i>
SNMP	<i>Simple Network Management Protocol</i>
TCP/IP	<i>Transmission Control Protocol/Internet Protocol</i>
TDM	<i>Time Division Multiplexing</i>
TLS	<i>Transport Layer Security</i>
UAC	<i>User Agent Client</i>
UAS	<i>User Agent Server</i>
UDP	<i>User Datagram Protocol</i>
URA	Unidade de Resposta Audível
URI	<i>Uniform Resource Identifier</i>
VAD	<i>Voice Activity Detection</i>
VLAN	<i>Virtual Local Area Network</i>
VPN	<i>Virtual Private Network</i>
VoDSL	<i>Voice over DSL</i>
cRTP	<i>Compressed Real Time Transport Protocol</i>
VOIP	<i>Voice over IP</i>

LISTA DE FIGURAS

Figura 2.1:	Representação do experimento de Graham Bell (MORTON, 1999).	18
Figura 2.2:	Operadora de mesa de chaveamento (1936) (AT&T, 2005).	19
Figura 2.3:	Processo PCM de digitalização da voz.	20
Figura 2.4:	Gráfico comparativo entre tráfego de voz e dados.	22
Figura 2.5:	Transmissão da voz através de rede de pacotes.	23
Figura 2.6:	Comutação de Circuitos (TDM).	24
Figura 2.7:	Comutação de Pacotes.	25
Figura 2.8:	Esquema de alto nível dos protocolos VoIP.	26
Figura 2.9:	Ligação com SIP usando um servidor <i>proxy</i> .	28
Figura 2.10:	Funções dos servidores SIP.	29
Figura 2.11:	Componentes da recomendação H.323.	33
Figura 2.12:	Pilha de protocolos H.323.	34
Figura 2.13:	Cabeçalho do protocolo RTP.	36
Figura 2.14:	Cabeçalho de um pacote RTP com voz, capturado com o software Ethereal.	36
Figura 2.15:	Tempos de conversão entre <i>codecs</i> .	37
Figura 2.16:	Arquiteturas de VoIP mais comuns.	41
Figura 2.17:	Telefone IP Cisco modelo 7970.	41
Figura 2.18:	Tela do <i>softphone</i> Xlite.	42
Figura 2.19:	Conexões do Cisco 186 ATA (<i>Analog Telephone Adapter</i>).	43
Figura 2.20:	Tela do cliente do serviço Skype.	45
Figura 2.21:	Tela do cliente do Gizmo.	46
Figura 3.1:	Efeito do atraso no sinal de voz.	52
Figura 3.2:	Diagrama de inserção de atrasos na transmissão.	53
Figura 3.3:	Efeito do <i>jitter</i> no sinal de voz.	55
Figura 3.4:	Esquema de uso dos mecanismos de Qos e Cos.	61
Figura 3.5:	Posicionamento do <i>buffer</i> de <i>jitter</i> .	63
Figura 3.6:	Cabeçalho com compressão do protocolo RTP.	64
Figura 4.1:	Relação entre capacidade, tráfego e desempenho.	70
Figura 4.2:	Fases do Planejamento de Capacidade de Rede.	72
Figura 4.3:	A.K. Erlang (MILLENNIUM MATHEMATICS PROJECT, 1997).	73
Figura 4.4:	Parte de uma tabela de Erlang.	77
Figura 4.5:	Metodologia para o planejamento de capacidade da rede telefônica convencional.	79
Figura 4.6:	Tabela de Erlang para GoS 0,01 e HMM 2,7e.	83
Figura 4.7:	Diagrama da rede da captura com o Ethereal.	83

Figura 4.8:	Cálculo da largura de banda do <i>codec</i> G.711.	84
Figura 4.9:	<i>Overhead</i> dos pacotes de voz (<i>codec</i> G.711).	85
Figura 4.10:	Exemplo de medição de tráfego com o MRTG.	86
Figura 4.11:	Metodologia para o planejamento de capacidade da rede para serviços VoIP.	87
Figura 5.1:	Diagrama da rede com a distribuição dos ramais	89
Figura 5.2:	Topologia de rede simplificada.	89
Figura 5.3:	Número de servidores para tráfego de 0,67e e GoS 1%	93
Figura 5.4:	Diagrama da distribuição de ramais e a largura de banda	94
Figura 5.5:	Tela da ferramenta Ethereal para análise de fluxos RTP	96
Figura 5.6:	Ambiente de testes.	99
Figura 5.7:	Análise de fluxos do segmento da Reitoria.	99
Figura 5.8:	Gráfico das ligações com volume de tráfego e o número de ramais atuais.	102
Figura 5.9:	Gráfico das ligações com 50% a mais de volume de tráfego e o número de ramais atuais.	102
Figura 5.10:	Gráfico das ligações com 50% a mais de volume de tráfego e de número de ramais.	103

LISTA DE TABELAS

Tabela 2.1:	Características dos <i>Codecs</i>	39
Tabela 3.1:	Períodos de Maior Movimento (PMM).	49
Tabela 3.2:	Origem dos atrasos usando o <i>codec</i> G.729 (GOODE, 2002).	54
Tabela 3.3:	Classes de serviços do IEEE 802.1p.	59
Tabela 3.4:	Classes de serviços dos <i>switches</i> HP.	60
Tabela 3.5:	Mapeamento entre valores de CoS (802.1p) e DSCP (DiffServ).	60
Tabela 4.1:	Modelos de Tráfego Telefônico.	76
Tabela 4.2:	Levantamento de tráfego telefônico UCPel e UFRGS (Campus do Vale).	80
Tabela 4.3:	Intensidade do tráfego telefônico na UCPel e na UFRGS.	81
Tabela 4.4:	Valores medidos e calculados dos <i>codecs</i>	84
Tabela 4.5:	Largura de banda para quantidade de amostras de voz por pacote (<i>codec</i> G.711).	85
Tabela 5.1:	Levantamento do tráfego de dados.	90
Tabela 5.2:	Intensidade de tráfego por segmento.	92
Tabela 5.3:	Número de linhas por segmento.	93
Tabela 5.4:	Largura de banda por segmento.	95
Tabela 5.5:	Total de chamadas por segmento.	96
Tabela 5.6:	Número de ramais e largura de banda por segmento, com 50% de acréscimo.	100
Tabela 5.7:	Resultados das medições para intensidade de tráfego atual.	100
Tabela 5.8:	Resultados das medições para intensidade de tráfego de rede com 50% acréscimo.	100
Tabela 5.9:	Resultados das medições para intensidade de tráfego e número de ramais com 50% de acréscimo.	101

RESUMO

Este trabalho tem como objetivo o estudo da tecnologia VoIP (Voz sobre IP) e a sua aplicação em redes corporativas, enfocando o planejamento de capacidade da rede de dados para absorver serviços VoIP. Serão apresentados tópicos sobre a fundamentação teórica de VoIP (Voz sobre IP), os requisitos de arquitetura de rede e QoS (Qualidade de Serviço) exigidos pelo serviço. Mostra-se também como a metodologia para planejamento de capacidade usado em telefonia convencional pode ser adaptada aos serviços VoIP em uma rede corporativa. Foi aplicada a metodologia adaptada através de um estudo de caso em uma rede corporativa real.

Palavras-chave: VoIP, QoS, Planejamento de Capacidade.

Corporate Network Capacity Planning to VoIP Services Implementation

ABSTRACT

This work has as objective the study of capacity planning in corporate networks for the implementation of VoIP (Voice over IP) services. We will presents topics about the theoretical background of VoIP, the requirements of architecture of network and QoS (Quality of Service) demanded by the service. It will also reveal how the methodology used for planning capacity in conventional telephony, could be adjusted to the VoIP services in a corporate network. The adjusted methodology was applied in a real corporate network.

Keywords: VoIP, QoS, Capacity Planning.

1 INTRODUÇÃO

As telecomunicações experimentaram uma grande revolução nos últimos anos. A antiga promessa de uma rede convergente (voz, dados e vídeo) começou a acontecer. A convergência se deu em redes baseadas na pilha de protocolos TCP/IP, que possibilitou o oferecimento de serviços avançados, a um custo baixo, para todos os tipos de usuários sejam eles residenciais, corporativos ou provedores de serviços. Uma das tecnologias para a convergência está sendo a voz sobre IP (VoIP), (GOODE, 2002).

Os serviços de VoIP são suportados por redes baseadas em TCP/IP, tanto para a sinalização de chamadas telefônicas como para a transmissão de tráfego de voz, com possibilidade de integração com a rede de telefonia pública. Serviços de VoIP, tais como o Skype (SKYPE, 2005) e o FWD (*Free World Dialup*) (FWD: FREE WORLD DIALUP, 2005), que oferecem comunicação de voz ponto a ponto através da Internet de forma gratuita, estão se tornando cada vez mais populares entre usuários residenciais.

O mercado para serviços VoIP tem crescido de forma considerável nos últimos anos (VARSHNEY et al., 2002). Este crescimento é motivado pela redução de custos e ao maior aproveitamento da infra-estrutura existente. Na área acadêmica, projetos tais como o fone@RNP (2005) no Brasil e SIP@Edu (SIP.EDU: VOICE OVER IP WORKGROUP, 2005) nos EUA, estudam os detalhes da tecnologia VoIP fazendo uso de aplicações avançadas na infra-estrutura da rede Internet2 (INTERNET2, 2005).

Para os serviços de VoIP obterem maior aceitação junto aos usuários, alguns aspectos devem ser levados em consideração. O mais importante aspecto é a comparação com a telefonia convencional em relação à qualidade de serviço das chamadas. Para garantir a qualidade da telefonia convencional em redes TCP/IP deve-se fazer uso de planejamento de capacidade, de configurações de QoS (*Quality of Service*) e de uma adequada arquitetura de rede com redundância de equipamentos, aplicações e serviços. Para desta forma, tentar chegar à disponibilidade de 99,999% tipicamente oferecida pela rede de telefonia convencional (VARSHNEY et al., 2002). Além disto, o quesito de GoS (*Grade of Service*), que define a probabilidade de bloqueio de chamadas, deve seguir os mesmos parâmetros da rede de telefonia convencional, ou seja, cerca de 1% probabilidade de bloqueios (CONWAY, 2000).

A garantia da qualidade do serviço de VoIP é de vital importância para a aceitação da tecnologia. Embora em serviços gratuitos como o Skype e o FWD não exista qualquer garantia de qualidade da chamada, em redes corporativas isto se faz necessário. No caso dos serviços gratuitos, grande parte dos usuários são domésticos fazendo uso da Internet para as ligações. Para este tipo de usuário a economia nas ligações acaba compensando a qualidade do serviço recebido. Entretanto, mesmo na Internet pública sem QoS, principalmente nos EUA, a qualidade das chamadas não é tão prejudicada conforme estudos publicados em (MARKOPOULOU; TOBAGI; KARAM, 2003) e (JI; SCHULZRINNE,

2003).

No ambiente corporativo, as empresas dependem do bom funcionamento dos seus serviços de telefonia para o encaminhamento dos seus negócios. No caso do uso de serviços VoIP em redes corporativas, não é aceitável que as ligações sejam prejudicadas por tráfego de outros serviços, tais como acesso a páginas ou *download* de arquivos. Para isto, devem haver garantias de QoS na rede corporativa e, principalmente, um planejamento de capacidade da rede para absorver os serviços de VoIP com qualidade e sem prejuízo para os outros serviços já existentes.

1.1 Objetivos

O uso de serviços VoIP em redes corporativas demandam estratégias peculiares de implantação. Devido à grande diversidade de equipamentos e aplicações existentes nestas redes e à dependência, cada vez maior, do negócio das corporações na disponibilidade dos serviços de comunicação faz com que a agregação de mais um serviço deva ser cuidadosamente analisado. E sendo o serviço de VoIP dependente de requisitos críticos de QoS, os cuidados devem ser ainda maiores. A implantação de serviços VoIP necessita de um planejamento de capacidade da rede e a determinação do impacto nos recursos de rede disponíveis. Para isto, deve ser feito o levantamento de requisitos do serviço. Um dos requisitos mais importantes é a intensidade do tráfego telefônico convencional, que deve ser considerado quando da sua utilização em redes TCP/IP. Portanto, é necessário obedecer a uma metodologia que determine os requisitos necessários para implantar o serviço de VoIP em redes corporativas com qualidade equivalente à telefonia convencional.

Pretende-se, através deste estudo, definir uma metodologia para o planejamento da capacidade de rede, levando-se em conta os parâmetros de QoS (*Quality of Service*) próprios dos serviços VoIP.

1.2 Organização

Este trabalho divide-se em 6 capítulos, sendo este, a introdução, o primeiro deles.

Em seguida, no capítulo 2, é mostrado as especificidades da tecnologia VoIP, os protocolos, arquiteturas e aplicações.

No capítulo 3, é feita uma revisão de qualidade de serviço (QoS) em telefonia convencional e em serviços VoIP.

O capítulo 4, trata da fundamentação teórica sobre planejamento de capacidade de rede, métodos de planejamento de capacidade de rede em telefonia convencional e a apresentação da metodologia para planejamento de capacidade de rede para serviços VoIP.

No capítulo 5 é apresentado o estudo de caso para a aplicação da metodologia desenvolvida no capítulo 4, juntamente com os resultados do experimento.

Este trabalho finda no capítulo 6, com as sugestões de trabalhos futuros e as conclusões do autor.

2 VOZ SOBRE IP (VOIP)

A tecnologia de VoIP é baseada no funcionamento da telefonia convencional. Embora o objetivo final, a comunicação entre os usuários, seja o mesmo, algumas técnicas são particulares para cada uma das tecnologias. Portanto, serão revistos alguns conceitos e apresentados os mecanismos de funcionamento da telefonia convencional e na sequência do capítulo, os conceitos e funcionamento da tecnologia de VoIP.

2.1 Telefonia Convencional

Em 10 de Março de 1876, em Boston, Massachussets, nos Estados Unidos, Alexander Graham Bell inventou o telefone (figura 2.1). Desde então, a forma de se comunicar mudou. O advento do telefone tornou as comunicações mais simples e de acesso ao cidadão comum, algo que não tinha acontecido até então. Mas o serviço de telefonia obteve sua preeminente posição entre os serviços de comunicação de forma gradativa. Por exemplo, mesmo nos Estados Unidos não foi antes de 1910 que os ganhos da indústria de telefonia excederam os ganhos do sistema postal. Na maioria dos países, este tempo foi ainda maior (FARLEY, 2005).

A principal razão para esta demora foi que o telefone teve que criar sua própria infraestrutura. E além disto, teve que competir com o telégrafo, que era o serviço mais utilizado na época. Em muitos países, a telefonia foi boicotada em prol dos serviços de correio e telégrafos estatais. No caso de empresas, por exemplo, a Western Union, após um curto período de análise, decidiu a favor de continuar usando o telégrafo para as suas comunicações, ao invés da telefonia (FARLEY, 2005). Esta decisão foi tomada baseada nos problemas econômicos e tecnológicos os quais a telefonia sofria na época.

Em relação a custo, por exemplo, em Nova Iorque, no ano de 1900, uma assinatura mensal para os serviços de telefonia residencial era de U\$ 20,00 (U\$ 2000 nos valores de hoje) e U\$ 40,00 (U\$ 4000 em valores atuais) para uso comercial, restringindo o uso apenas para usuários corporativos ou usuários domésticos com grande poder aquisitivo (ODLYZKO, 2000). O custo do serviço começou a baixar de forma mais acentuada a partir da automatização das centrais telefônicas entre 1945 e 1969, e do uso de sistemas digitais na década de 80. A partir da década de 90, com o aumento da concorrência entre as empresas de telefonia e a ampliação do uso das redes de dados, principalmente a Internet, as empresas de telefonia passaram a prover um grande volume de serviços de dados. O volume de dados passou o de ligações telefônicas e forçou as empresas de telecomunicações, não mais chamadas de empresas de telefonia como antes, a adaptarem suas infra-estruturas. Com isto, a convergência das comunicações passou a ser possível, incentivando o uso de serviços de VoIP no aproveitamento dos recursos já existentes e a conseqüente redução de custos para as empresas de telecomunicações.

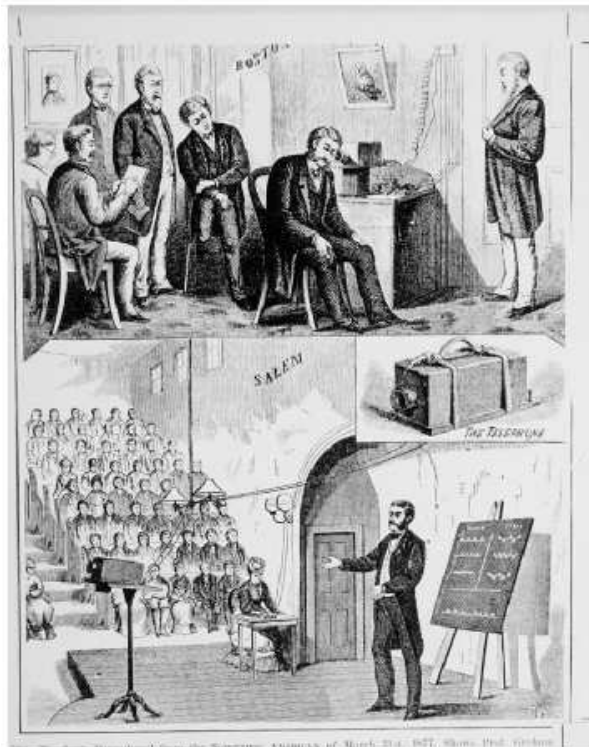


Figura 2.1: Representação do experimento de Graham Bell (MORTON, 1999).

2.1.1 Comunicação Telefônica

A primeira transmissão de voz, feita por Alexander Graham Bell, se deu através de um fio interconectando dois dispositivos. Não havia sinal de chamada e muito menos discagem, era basicamente uma ligação ponto a ponto do tipo *half-duplex*. No decorrer do tempo, o desenvolvimento da tecnologia permitiu a transmissão *full-duplex* tornando possível a conversação telefônica. Na forma mais básica, a transmissão de voz através dos fios necessita de um microfone de carvão, uma bateria, um eletroímã e um diafragma de ferro. Além disto, requer um meio físico entre as duas localidades que desejem se comunicar (o conceito de discagem, inicialmente, não existia). A pouca escalabilidade desta estratégia, e o alto custo, impossibilitaram a expansão do sistema telefônico. A expansão do sistema foi avançando com o uso de comutadores (*switches*). Com o uso de comutadores, não foi mais necessário a ligação física direta entre os telefones dos usuários.

No começo, o mecanismo de comutação era exercido por operadores humanos (figura 2.2). Desta forma, o telefone do usuário necessitava apenas de uma ligação com o comutador, ao invés de uma ligação direta com outro telefone. O primeiro comutador automático foi inventado por um usuário de telefone, por sua própria necessidade. Por achar que estava sendo prejudicado no seu negócio de pompas fúnebres, o americano Almon Strowger criou um protótipo que eliminava a necessidade de usar operadores humanos na distribuição das chamadas. Ele foi motivado a inventar o comutador automático porque um competidor da sua cidade estava recebendo todas as chamadas de serviço de pompas fúnebres e ele nenhuma. O motivo era que a filha e a mulher do competidor faziam parte de grupo de operadoras do comutador (FARLEY, 2005).

A discagem para encontrar o destino das ligações começou a ser utilizada quando



Figura 2.2: Operadora de mesa de chaveamento (1936) (AT&T, 2005).

os comutadores se tornaram automáticos, e já havia um grande volume de ligações, o que tornava a operação manual impraticável. Estes comutadores evoluíram passando de eletromecânicos para eletrônicos com grande capacidade de processamento. A automação, a digitalização e as técnicas de amplificação de sinal permitiram a expansão dos serviços de telefonia e a escalabilidade necessária para prover, com qualidade, a comunicação ao redor do mundo.

Existem duas formas de transmissão do sinal de voz em telefonia, a forma analógica e a forma digital.

Na origem do telefone, toda a comunicação era analógica. Por este motivo, a qualidade da voz era prejudicada na transmissão em grandes distâncias e era necessário amplificar o sinal. O problema era que a amplificação também era feita no ruído da linha de transmissão. Isto acarretava em uma qualidade péssima da voz, inutilizando, na maioria das vezes, a chamada. Atualmente, a maioria dos terminais residenciais são analógicos. Portanto, a transmissão é feita de forma analógica até a central telefônica. Em casos de uso de PABX (*Private Automatic Branch Exchange*), os ramais são na maioria analógicos existindo alguns ramais digitais dependendo dos recursos da central.

A transmissão digital foi introduzida para reduzir o problema da amplificação sofrida na transmissão analógica. Além disto, possibilitou que a voz fosse transferida em redes de pacotes. A digitalização da voz é feita através de *codecs* (codificadores/decodificadores) de voz. As transmissões digitais são feitas entre as centrais telefônicas, ou seja, no *backbone* das empresas provedoras de serviços. Os comutadores devem fazer a conversão digital/analógico quando o sinal for em direção ao assinante. O *codec* mais utilizado é o PCM (*Pulse Code Modulation*). As técnicas de amostragem são baseadas no teorema de Nyquist (SOARES; LEMOS; COLCHER, 1995), onde é definido que o número de amostras devem ser duas vezes maior, do que a maior frequência do sinal, o que no caso da voz é de 8000 vezes por segundo (a maior frequência é 4KHz numa linha de voz), para representar com maior fidelidade o sinal original. O processo de digitalização da voz usando PCM é o seguinte (figura 2.3):

- As formas de onda analógicas são passadas por um filtro de voz para eliminar qualquer frequência acima de 4000 Hz. Estas frequências são filtradas até 4000 Hz para limitar a quantidade de *crosstalk* (interferência entre linhas) na rede de voz.

- É feita a amostragem do sinal analógico na taxa de 8000 vezes por segundo.
- Após a amostragem, o sinal é convertido em um formato digital discreto. A amostra é representada por um código que indica a amplitude da forma de onda no instante da amostragem. O PCM utilizado em telefonia usa 8 bits em cada amostra.

A largura de banda utilizada com o PCM nas chamadas é de 64kbits, o que corresponde a *frames* de 8 bits sendo amostrados 8000 vezes por segundo. Os 8 bits de cada amostra representam, um bit para sinal, 3 bits para uma faixa logarítmica e 4 bits para um intervalo dentro da faixa logarítmica.

Existem duas variações da codificação PCM comumente usadas, a *mu-law* padrão na América do Norte e Japão e a *a-law* que é o padrão usado na Europa e em outros países. No Brasil o padrão adotado é o *a-law*. Estes métodos são similares no processo de compressão logarítmica para conseguir colocar de 12 a 13 bits em 8 bits, mas diferem em alguns detalhes. Por exemplo, o método *mu-law* possui ligeira vantagem sobre o *a-law* em termos de relação ao desempenho sinal/ruído (DAVIDSON; JAMES, 2000). Quando são feitas ligações de longa distância entre países com padrões diferentes, quem possuir o método *mu-law* ficará responsável pela conversão do sinal.

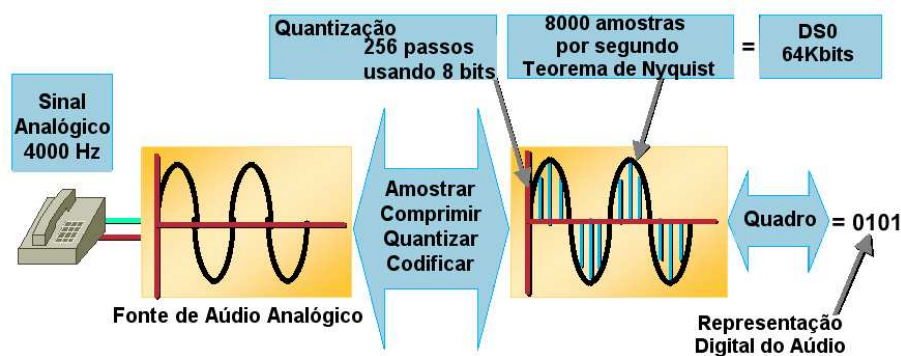


Figura 2.3: Processo PCM de digitalização da voz.

2.1.1.1 Infra-estrutura

A infra-estrutura de telefonia começa com os pares de fios conectados ao terminal telefônico, chamado *local loop*. Estes pares de fios são conectados a uma central telefônica, seja ela do provedor de serviços de telefonia ou uma central particular, tal como um PABX. A comunicação entre comutadores centrais ou entre PABXs é chamado de *trunk* (tronco).

A sinalização é feita em dois sentidos, usuário-rede telefônica e rede telefônica-rede telefônica. No primeiro caso, é a forma de como o usuário final se comunica com a rede telefônica, normalmente através de um terminal analógico. O método de sinalização mais comum é o DTMF (*Dual Tone Multi-Frequency*) que possibilita através da variação da frequência representar dígitos e comandos no terminal telefônico. É chamado de sinalização *inband*, pois sua transmissão ocorre no mesmo caminho da voz. No caso do sentido rede telefônica-rede telefônica, a sinalização ocorre entre centrais telefônicas.

Existem protocolos de sinalização *inband* e *outband*. Os protocolos *inband* entre centrais foram abolidos nos anos 70, por limitações técnicas e por serem susceptíveis a fraudes, tais como o uso do apito de cereais do *Captain Crunch* (DRAPER, 2005). O protocolo *Signaling System 7* (SS7) é um caso de protocolo de sinalização *outband*, pois

possui um caminho independente da voz. Este protocolo reserva um canal de 64kbps para controle das chamadas e através de pacotes controla as conexões na central e entre centrais. Funções como identificação do telefone chamador e diminuição no tempo de complemento de chamadas são conseguidas usando este protocolo. É o padrão usado atualmente.

2.2 Breve Histórico de VoIP

A transmissão de voz digitalizada através de redes vem sendo arquitetada há bastante tempo. Em 1962, através do artigo (BARAN, 1962) foi comentado o uso de redes para tráfego multimídia. Em 1977, foi proposta pelo IETF (*Internet Engineering Task Force*) através de uma RFC (*Request for Comments*) (COHEN, 1977) a criação de um protocolo para transmissão de voz em forma de pacotes para uso militar. Em 1995, a empresa Net2Phone (NET2PHONE, 2005) anunciou planos para o lançamento do primeiro serviço de comunicação entre computadores pessoais e a telefonia convencional. Pouco depois, em 1996, o ITU (*International Telecommunication Union*) disponibilizou a primeira versão da recomendação H.323 (ITU-T, 1997) sendo considerado o primeiro padrão para VoIP. O protocolo SIP (*Session Initiation Protocol*) foi apresentado em 1999 pelo IETF através da RFC 2543.

Com o aumento de largura de banda e custo cada vez menor dos enlaces de dados propiciou-se a retomada do interesse de voz sobre redes de pacotes. Outro fator importante foi a padronização dos protocolos de VoIP, sendo o pioneiro a recomendação H.323 do ITU-T. Os serviços de VoIP podem ser utilizados por provedores de serviços, corporações e usuários domésticos.

Os provedores de serviços de telefonia convencional passaram a usar a tecnologia de VoIP para diminuir os custos das ligações entre as centrais. Fazendo uso de pacotes ao invés de circuitos, os provedores conseguiram utilizar de forma mais eficiente os recursos das suas redes. Portanto, o uso de voz em redes de pacotes é bastante utilizado em provedores de serviços, e para aproveitar a estrutura dos *backbones* de dados existentes, que trafegam TCP/IP (figura 2.4).

A tecnologia de VoIP não é novidade no ambiente corporativo. As empresas estão usando serviços VoIP para aproveitar os enlaces de dados com pouco uso, eliminar os altos custos de centrais telefônicas convencionais e, principalmente, diminuir os custos de ligações entre filiais e matrizes. Embora os custos das ligações sejam baixos usando VoIP, o custo do investimento inicial é alto. O custo de um telefone IP, por exemplo, é cerca de dez vezes maior do que um telefone convencional. Para reduzir o custo com os telefones podem ser usados *softphones*, que são telefones por software, juntamente com *headsets* (conjunto de fone de ouvido e microfone). O problema deste tipo de abordagem é a inconveniência criada para o usuário, que deverá mudar a sua forma de interagir com o telefone. Os *headsets* de boa qualidade possuem um custo maior do que um telefone convencional, mas menor do que um telefone IP. Além do custo, alguns problemas tais como a disponibilidade e a qualidade do serviço, fizeram com que muitas empresas não aderissem em massa aos serviços VoIP. A disponibilidade do serviço de telefonia convencional é o "cinco 9s" ou seja 99.999% de tempo do serviço em funcionamento, cerca de 5 minutos de parada em um ano, o que é bastante complicado de atingir em redes de dados (VARSHNEY et al., 2002). Em relação a qualidade do serviço, as redes de pacotes são mais susceptíveis a fatores tais como atraso, *jitter* e perdas, que afetam diretamente a qualidade da voz. Existem medidas para a qualidade da voz em VoIP e formas de garantia

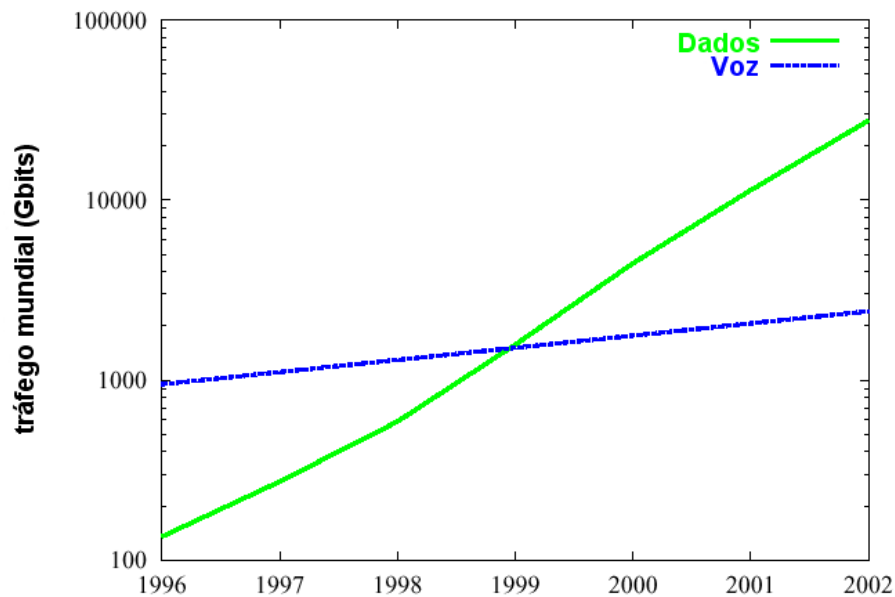


Figura 2.4: Gráfico comparativo entre tráfego de voz e dados.

que serão mostradas com mais detalhes no capítulo 3.

Em se tratando de usuários domésticos, o uso dos serviços de VoIP foram alavancados pela popularização dos serviços de acesso por banda larga tais como ADSL (*Assymmetric Digital Subscriber Line*), acessos via rádio e cabo. Com mais banda para acesso aos usuários, novos serviços e aplicações surgiram. Um dos serviços de maior sucesso é o de compartilhamento de arquivos. Este tipo de serviço permite que usuários troquem informações de forma ponto a ponto (P2P - *Peer to Peer*), bastando apenas alguma forma de indexação para encontrar o arquivo desejado. Um dos serviços deste tipo de maior sucesso é o Kazaa (KAZAA, 2005). Baseado nesta estrutura, os mesmos criadores do Kazaa criaram o serviço Skype (SKYPE, 2005). Com o mesmo princípio, de não ter um ponto central, os usuários poderiam achar uns aos outros e conversar através da Internet. Embora já existissem pioneiros neste tipo de serviço tais como Net2Phone e Microsoft Messenger (MESSENGER, 2005), o Skype aproveitou a popularização dos serviços de banda larga e a similaridade de uso das aplicações de compartilhamento de arquivos, para atingir mais de 100 milhões de usuários (SKYPE, 2005). Devido a esta popularidade, vários outros serviços surgiram tais como FWD, Gizmo Project (GIZMO PROJECT, 2005) e Sipphone (SIPPHONE, 2005a), fornecendo qualidade e funcionalidade iguais ou superiores ao Skype, com facilidade semelhante de uso. Os outros serviços fazem uso de padrões e protocolos abertos, tais como SIP (*Session Initiation Protocol*), para a sinalização e transporte dos fluxos de áudio, ao contrário do Skype que faz uso de protocolos proprietários. Com isto, os serviços abertos podem fazer uso de um grande número de equipamentos e softwares já existentes. O autor acredita que o Skype perderá em breve sua posição de liderança incontestável nos serviços VoIP para usuários domésticos, sendo os usuários distribuídos nos serviços que fazem uso de protocolos abertos.

2.3 Fundamentos

A transmissão da voz sobre redes de pacotes requer uma série de técnicas e protocolos. Para transformar as ondas sonoras de voz em códigos binários são necessários o uso

de codificadores/decodificadores (*codecs*). Após a captação do sinal da voz por meio de um microfone, as amostras do sinal devem passar por processos de quantização, codificação e compressão para poderem ser transportados de forma eficiente em redes de pacotes (figura 2.5). Os processos são dependentes das técnicas usadas pelos diversos tipos de codificadores de voz existentes. Na figura 2.3 é demonstrado um esquema da transformação da voz para código binário. A técnica de digitalização da voz através de codificadores de áudio é usada em telefonia pública há bastante tempo, usando o método PCM. A voz codificada é repassada para as camadas responsáveis pelo transporte na pilha de protocolos TCP/IP, os protocolos RTP (*Real-Time Transport Protocol*) e UDP (*User Datagram Protocol*). Desta forma, a voz codificada assume o papel dos dados fazendo analogia com o modelo de referência OSI-RM (*Open Systems Interconnection - Reference Model*). O processo de decodificação da voz acontece no receptor que deve possuir o *codec* apropriado. Para a chamada ser realizada com sucesso deve haver um mecanismo de sinalização, ou seja, uma forma de controlar o fluxo de pacotes. Os principais mecanismos de sinalização estão representados na recomendação H.323 e o protocolo SIP. Além disto, os problemas observados em redes TCP/IP, tais como atraso, perdas e *jitter* são fatores determinantes na qualidade da chamada.

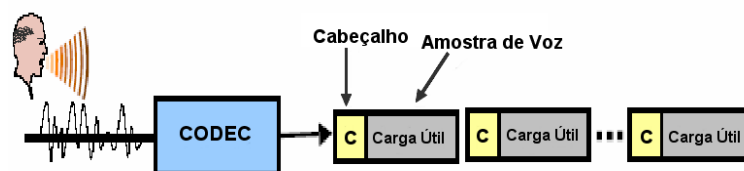


Figura 2.5: Transmissão da voz através de rede de pacotes.

2.3.1 Tipos de Redes

Os tipos de redes mais comuns são as baseadas em circuitos e as baseadas em pacotes.

2.3.1.1 Redes Baseadas em Circuitos

A tecnologia tradicional de rede empregada para carregar voz é a comutação de circuitos. Por ser uma tecnologia orientada a conexão, utiliza técnicas de multiplexação para concentrar muitas conversações em um único tronco físico. A técnica de multiplexação TDM (*Time Division Multiplexing*) não requer funções complexas de controle de tráfego. Os usuários são assinalados a intervalos fixos de tempo, tornando a qualidade do serviço determinística e previsível (exceto por erros de transmissão e bloqueio de conexão) e os usuários não podem afetar a qualidade de serviço de outro usuário. Não há contenção por banda ou recurso, a não ser pelo número de intervalos de tempo disponíveis para estabelecer conexões. Entretanto, o esquema TDM não permite atingir o importante objetivo de maximizar a utilização da banda utilizada. Muitas aplicações, como transferência de arquivo e vídeo conferência, requerem banda variável durante suas transmissões. Alocação de banda acima do necessário leva a uma severa ineficiência, pois a banda não utilizada por transmissões lentas ou intermitentes poderia ser utilizada por outra conexão. Em TDM, as fatias de tempo são alocadas para uma conexão independentemente se o usuário está ou não transmitindo algo (figura 2.6). A forma de alocação garante a qualidade, mas propicia o desperdício de recursos. Por causa de sua estática alocação dos intervalos de tempo, o TDM não é eficiente para suportar tráfego com características de rajada. Por

outro lado é eficiente para suportar tráfego de voz (GONCALVES, 2000).

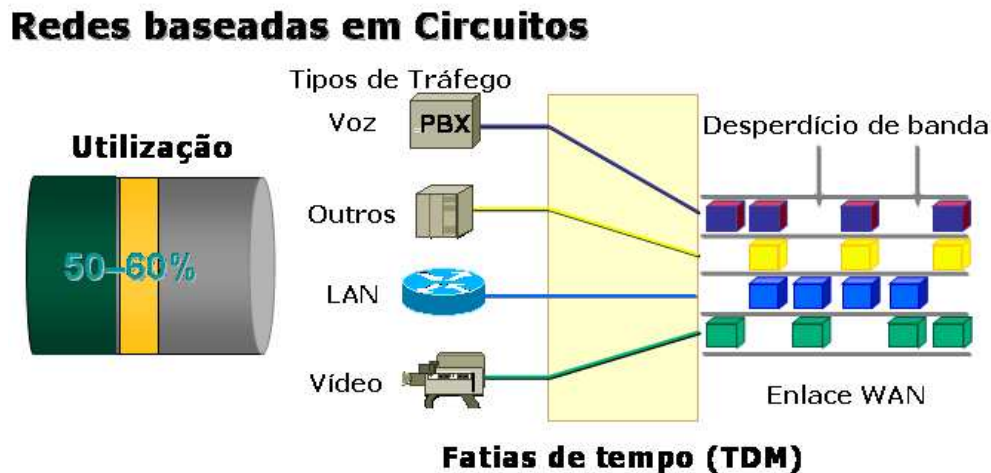


Figura 2.6: Comutação de Circuitos (TDM).

2.3.1.2 Redes Baseadas em Pacotes

A finalidade da rede de comunicação de dados é conectar e transportar informações entre computadores remotos, chamados usuários da rede. A eficiência está em transmitir o mais rápido possível quantidades de informações válidas entre o maior número de usuários. O parâmetro de desempenho básico utilizado em redes de computadores é a vazão (*throughput*) medida em bits por segundo ou octetos por segundo. A vazão da rede é afetada pelo congestionamento, ou seja, pela sobrecarga temporária de recursos. São os congestionamentos que geralmente causam o aumento de retransmissão das informações e as perdas de pacotes.

Estas variações de sobrecarga sobre a rede são percebidas pelos usuários através do aumento do tempo de resposta, gerando insatisfações quanto à qualidade de serviço prestada. O ideal seria a rede ter um comportamento transparente, como se os usuários estivessem interligados sem o compartilhamento de recursos. Dessa forma não haveria nenhum tipo de contenção de tráfego causado pela rede. Aproximar a realidade das redes de dados a este ideal é uma meta a ser perseguida. A tecnologia tradicional de rede projetada para carregar dados é a comutação de pacotes. Os pacotes podem ser comutados sem estabelecer uma conexão prévia. É o que acontece com o protocolo IP. Os pacotes são roteados em cada nó da rede e nenhuma banda é pré-alocada para a transferência de dados. Esta tecnologia tem a capacidade de maximizar o uso da rede, mas não pode garantir a qualidade de serviço porque não há qualquer tipo de alocação de recurso (figura 2.7). Esta situação caracteriza uma rede de melhor esforço (*best effort*) onde a largura de banda ocupada pelo usuário varia de acordo com a carga da rede a cada instante.

A Internet, foi projetada como rede de melhor esforço, e originalmente não previa qualquer mecanismo para garantia de qualidade. A filosofia de melhor esforço torna-se inaceitável, pois a qualidade degrada rapidamente quando a demanda da rede cresce. Simplesmente aumentar a capacidade dos enlaces e dos equipamentos não garante a qualidade indefinidamente. Propostas de mecanismos de identificação de prioridades e alocação de recursos têm sido apresentadas para suprir QoS em redes TCP/IP. Pode-se citar os protocolos DIFFerentiated SERVices (DiffServ) e MPLS (*Multiprotocol Label Switching*)

como alternativas para implementar QoS.

Redes de Pacotes

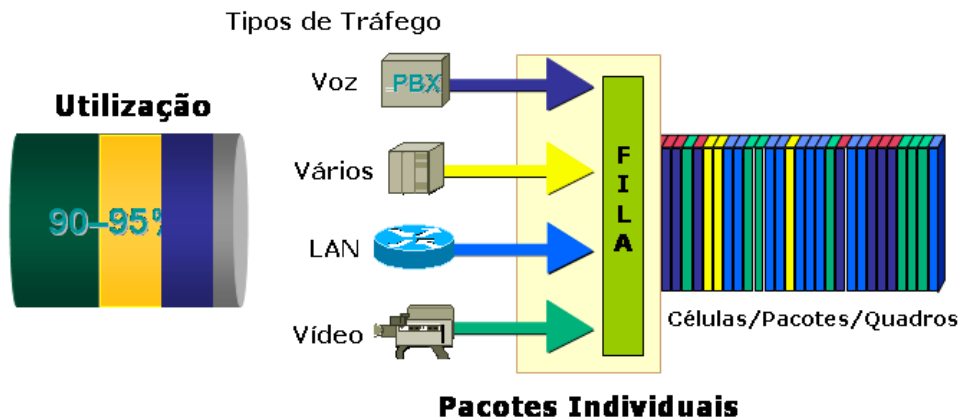


Figura 2.7: Comutação de Pacotes.

2.4 Padrões e protocolos

Existem basicamente dois tipos de protocolos em VoIP: os proprietários e os abertos. Dentre os proprietários estão o Skinny da empresa Cisco, o usado no serviço Skype e as variações aos protocolos abertos, como no caso do SIP no Microsoft Messenger. Existem ainda protocolos utilizados por fabricantes de centrais telefônicas que fazem uso de *software* e *hardware* proprietários. Embora exista ainda uma grande base destes protocolos proprietários, cada vez mais os protocolos e padrões abertos estão sendo utilizados e adotados pelos fabricantes de equipamentos, principalmente o protocolo SIP.

Os protocolos abertos mais utilizados e difundidos são o SIP e a recomendação H.323. O SIP foi concebido pelo IETF e possui características similares ao protocolo HTTP. É um protocolo simples e flexível, usado em VoIP para o controle (sinalização) das chamadas. O controle das características da mídia é feito pelo protocolo SDP (*Session Description Protocol*), que funciona em conjunto com o protocolo SIP.

A recomendação H.323 é uma recomendação “guarda-chuva”, pois abrange sinalização, codificação, segurança, áudio, vídeo e gerenciamento das chamadas. Criada pelo ITU-T em 1996 é fortemente baseada na arquitetura de telefonia convencional o que a torna bastante completa. Em quesitos técnicos de recursos, o H.323 é mais completo do que o SIP. Devido à complexidade do H.323, o protocolo SIP, que é mais simples e baseado no paradigma da Internet, tem recebido a preferência de usuários e fabricantes. Por ser ligado a um órgão de padronização não tão aberto e dinâmico quanto o IETF, o H.323 acaba levando mais tempo para ser atualizado e revisado. Existem também os protocolos de controle de *gateways*, como o MGCP (*Media Gateway Controller Protocol*).

Os protocolos responsáveis pelo tráfego da mídia são o RTP e o protocolo auxiliar RTCP (*Real-Time Transport Control Protocol*). Ambos os protocolos são transportados sobre os protocolos UDP e IP. Na figura 2.8, um esquema de alto nível dos protocolos mais comuns de VoIP.

Os protocolos VoIP podem ser categorizados em protocolos de sinalização e em protocolos de mídia.

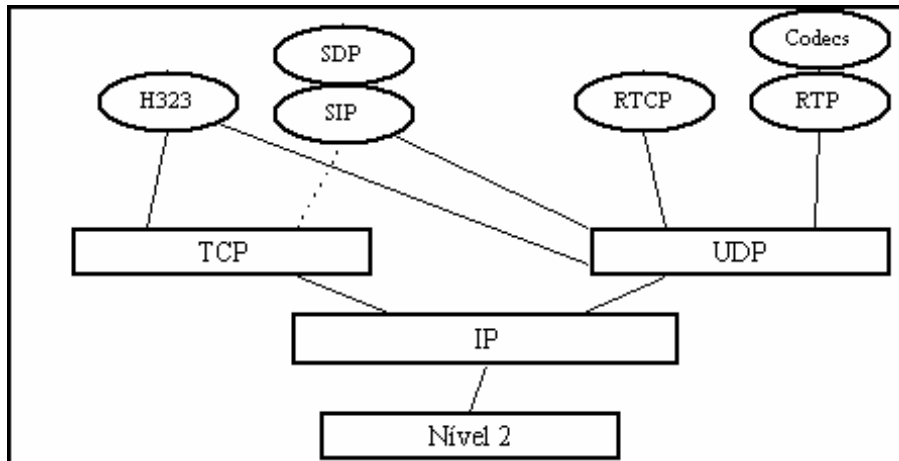


Figura 2.8: Esquema de alto nível dos protocolos VoIP.

2.4.1 Protocolos de Sinalização

2.4.1.1 SIP (*Session Initiation Protocol*)

O SIP (*Session Initiation Protocol*), definido em (ROSENBERG et al., 2002), é um protocolo de sinalização do nível de aplicação que tem funções de iniciar, modificar e terminar, em tempo real, sessões entre participantes sobre uma rede IP. O SIP pode suportar qualquer tipo de mídia. O protocolo é um dos componentes do conjunto de protocolos e serviços necessários no suporte à troca de dados multimídia na Internet. A sinalização do SIP possibilita que sejam feitas chamadas entre participantes, chamadas telefônicas por exemplo, e a negociação dos parâmetros da sessão. O conteúdo das sessões, sejam eles áudio, vídeo ou outro conteúdo multimídia, são trafegados entre os participantes da sessão usando um protocolo de transporte. Na maioria dos casos, o protocolo de transporte usado é o RTP (*Real-Time Transport Control*). Protocolos de acesso a diretórios e procura de endereços também são necessários, mas são independentes do protocolo SIP. A grande aplicação do protocolo SIP é no serviço de VoIP, embora possa ser usado em outros serviços tais como troca de mensagens (*Instant Messaging*) e videoconferência. Existe a tendência na indústria de que o SIP será o mecanismo de sinalização para serviços de voz, tanto nos provedores de serviços quanto nos usuários corporativos e domésticos (STALLINGS, 2003).

O SIP possui cinco passos básicos em relação ao estabelecimento e o término da comunicação:

- **Localização do Usuário:** determinar a localização do usuário na rede. Os usuários podem se mover para outros locais e acessar o seu telefone e funcionalidades de aplicações remotamente. Por exemplo, em uma rede corporativa o usuário receberá chamadas para o seu ramal não importando onde esteja dentro da empresa.
- **Disponibilidade do Usuário:** verifica se o usuário deseja ou está, disponível para comunicação.
- **Capacidade do Usuário:** determina os parâmetros e o formato da mídia a serem usados na sessão. A negociação dependerá da capacidade dos equipamentos envolvidos.

- Configuração da Sessão: chamadas ponto a ponto ou em conferência devem ser estabelecidas, com os parâmetros negociados.
- Gerenciamento da Sessão: incluem-se neste a transferência e terminação de sessões, a modificação de parâmetros da sessão e o carregamento de novos serviços.

O SIP emprega em seu projeto elementos desenvolvidos em protocolos anteriores. É baseado no modelo de transação requisição/resposta do HTTP (*Hyper Text Transfer Protocol*). Cada transação consiste na requisição de um cliente, que invoca um método particular ou função no servidor e recebe ao menos uma resposta. O SIP usa grande parte dos campos de cabeçalho, regras de codificação e códigos de *status* do HTTP. Com isto, as trocas de mensagens do SIP são em texto legível. Para determinar o conteúdo da sessão, é utilizado o protocolo SDP, que será descrito posteriormente.

Um sistema SIP pode ser visto como componentes distintos: cliente/servidor e elementos individuais de rede. Em (ROSENBERG; SCHULZRINNE, 2002a), são definidos cliente e servidor desta forma:

- Cliente: um cliente é qualquer elemento de rede que envia requisições SIP e recebe respostas SIP. Os clientes podem ou não interagir diretamente com o usuário. O *User Agent* e os *Proxies* são considerados clientes.
- Servidor: um servidor é um elemento de rede que recebe requisições e as responde. Exemplos de servidores são os *Proxies*, os servidores *User Agent*, servidores de redirecionamento (*Redirect*) e de registro (*Registrar*).

Os elementos individuais de uma configuração SIP padrão incluem (figura 2.9):

- *User Agent* (Agente do Usuário): deve existir em qualquer estação de terminação SIP. Atua de duas formas:

User Agent Client (UAC) (Agente do Usuário modo Cliente): trata das requisições SIP.

User Agent Server (UAS)(Agente do Usuário modo Servidor): recebe as requisições SIP e gera uma resposta que pode ser de aceitação, rejeição ou redirecionamento da requisição.

- *Redirect Server* (Servidor de Redirecionamento): o servidor de redirecionamento é usado durante a iniciação da sessão para determinar o endereço do dispositivo que está sendo chamado. Esta informação é retornada para o dispositivo que originou a chamada, direcionando o UAC para contatar um endereço alternativo no formato URI (*Universal Resource Identifier*) (BERNERS-LEE; FIELDING; MASINTER, 1998). Um URI é um identificador genérico usado para nomear qualquer recurso na Internet. Por exemplo, uma URL (*Uniform Resource Location*) utilizada para endereço de páginas é um tipo de URI.
- *Proxy Server* (Servidor de Procuração): o servidor de procuração é um entidade intermediária que atua tanto como servidor ou cliente. No modo cliente, tem o propósito de fazer as requisições por outros clientes. No modo servidor, um *proxy* tem como função principal o roteamento, fazendo com que a requisição seja enviada para uma outra entidade mais próxima do dispositivo final. Os *proxies* também são úteis para reforçar políticas de uso, como por exemplo verificar se determinado

usuário tem direitos de estabelecer uma chamada. Além disto, o *proxy* interpreta e, se for necessário, reescreve partes específicas da mensagem de requisição antes de repassá-la adiante. Na figura 2.10, um exemplo de chamada SIP utilizando um servidor *proxy*.

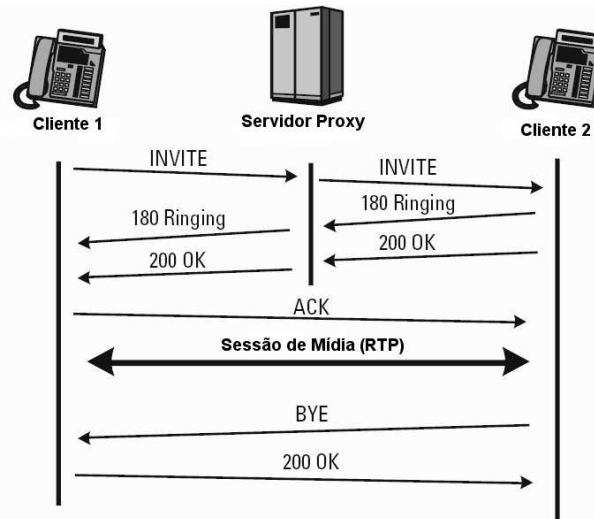


Figura 2.9: Ligação com SIP usando um servidor *proxy*.

- **Registrar** (Servidor de Registro): este servidor tem a função de aceitar requisições do tipo *REGISTER* e repassar as informações recebidas, que são o endereço SIP e o endereço IP associado, para o serviço de localização (*Location Service*) do domínio.
- **Location Service** (Serviço de Localização): este serviço é utilizado através de mensagens de redirecionamento (*redirect*) ou através de um servidor de procuração para obter informações sobre a possível localização do destino sendo chamado. Para este propósito, o serviço de localização mantém um banco de dados como mapeamento do endereço SIP e o endereço IP. Estas informações são fornecidas pelo servidor de registro. Normalmente, o servidor de registro e o serviço de localização se encontram no mesmo *host*.

Os vários servidores e serviços são definidos na RFC 3261 do protocolo como dispositivos lógicos. Eles podem ser implementados separadamente ou em conjunto, por exemplo, um mesmo *host* pode servir como *proxy* e redirecionador de chamadas. Servidores *proxies* podem atuar como servidores redirecionadores quando preciso. Se um redirecionamento é feito, o *proxy* necessita consultar o banco de dados do serviço de localização, que pode estar junto com o servidor de *proxy* ou não.

O serviço de DNS (*Domain Name System*) é uma importante parte da operação do SIP. Tipicamente, o UAC requisita através do DNS, o nome do UAS, e não o endereço IP. O servidor de procuração necessita consultar o servidor de DNS para achar o servidor de procuração do domínio destino. Por exemplo, a URI sip:emmonks@voip.com.br será traduzida pelo serviço de DNS para o endereço do servidor de *proxy* do domínio voip.com.br, algo como sip.atlanta.com e o IP correspondente.

O SIP, freqüentemente, é utilizado sobre UDP por razões de desempenho e deve prover seus próprios mecanismos de garantia de entrega, embora o protocolo possa ser utilizado sobre TCP (*Transmission Control Protocol*). Se for necessário um mecanismo de

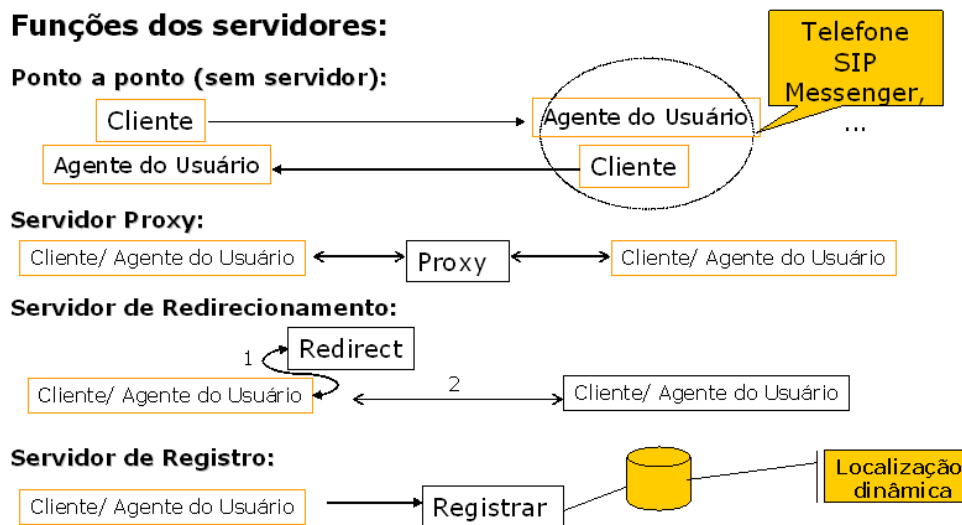


Figura 2.10: Funções dos servidores SIP.

transporte seguro, as mensagens SIP podem ser utilizadas de forma criptografada sobre o protocolo TLS (*Transport Layer Security*) ou através de tunelamento, com VPN (*Virtual Private Network*), por exemplo.

Existem dois tipos diferentes de mensagens SIP, a tipo requisição (*request*) e a tipo resposta (*response*). A diferença de formato entre os dois tipos de mensagens é identificada na primeira linha da mensagem. Na primeira linha de uma requisição tem um método definindo a natureza da requisição e a URI, indicando para onde a requisição deve ser enviada.

Já na primeira linha da resposta está contido um código. Todas as mensagens incluem um cabeçalho, consistindo em linhas começando com uma identificação de cabeçalho. Uma mensagem pode conter também um corpo, descrevendo o formato da mídia transportada.

Os métodos das requisições SIP definidos em (ROSENBERG; SCHULZRINNE, 2002a) são os seguintes:

- REGISTER: usado pelo agente usuário para notificar o endereço IP e a URL a qual deseja receber chamadas.
- INVITE: usado para estabelecer sessões de mídia entre os agentes.
- ACK: usado para confirmar a entrega das mensagens.
- CANCEL: usado para terminar requisições pendentes.
- BYE: usado para terminar sessões.
- OPTIONS: usado para solicitar informações sobre as capacidades do destino da chamada.

Exemplo de uma mensagem SIP do tipo requisição:

```
INVITE sip:bob@biloxi.com SIP/2.0
Via: SIP/2.0/UDP 12.26.17.91:5060
Max-Forwards: 70
To: Bob <sip:bob@biloxi.com>
```

```

From: Emmonks <sip:emmonks@voip.com.br;tag=1928301774
Call-ID: a84b4c76e66710@12.26.17.91
CSeq: 314159 INVITE
Contact: <sip:emmonks@voip.com.br>
Content-Type: application/sdp
Content-Length: 142

```

A primeira linha contém um método (**INVITE**), uma URI (**sip:bob@biloxi.com**) e o número da versão do SIP (**2.0**). As linhas seguintes são campos do cabeçalho. O cabeçalho **Via** mostra o caminho que a requisição percorreu, e é usado para rotear as respostas pelo mesmo caminho. Além disso, contém o endereço IP (**12.26.17.91**), o protocolo de transporte (**UDP**) e a porta (**5060**) a qual o receptor deverá enviar a resposta. O cabeçalho **Max-Forwards** limita o número de *hops* que uma requisição pode fazer no caminho até o destino. Este valor é decrementado a cada passagem de *hop*. O cabeçalho **To** contém o destino da mensagem e o campo **From** o originador da mensagem. O campo **From** também possui uma *tag* (etiqueta) que é um valor randômico identificador da sessão, que é gerado pelo UAC.

O campo **Call-ID** contém um identificador global único para esta chamada, gerado pela combinação de uma *string* randômica e o nome ou o endereço IP do *host*. A combinação dos campos **To**, **From** e **Call-ID** definem completamente uma relação SIP ponto a ponto entre Emmonks e Bob, como no exemplo acima. A relação entre dois agentes é definida como sendo um **diálogo**. O campo **CSeq** (*Command Sequence*) contém um inteiro e um nome de método. O número do **CSeq** é inicializado no começo da chamada e é incrementado para cada nova requisição dentro do diálogo, sendo um contador sequencial. Este campo é usado para distinguir uma retransmissão de uma nova requisição. O campo **Contact** contém a URI para comunicação direta entre os agentes. O campo **Content-Type** indica o tipo de formatação do corpo da mensagem. O campo **Content-Length** mostra o tamanho, em octetos, do corpo da mensagem.

Os códigos das respostas SIP definidas em (ROSENBERG; SCHULZRINNE, 2002a) são as seguintes:

- *Provisional* (1xx): a requisição foi recebida e está sendo processada.
Exemplos: 100 *Trying* (tentando a conexão), 180 *Ringin* (tocando no destino).
- *Success* (2xx): a ação foi recebida, entendida e aceita com sucesso.
Exemplo: 200 *OK* (procedimento executado com sucesso).
- *Redirection* (3xx): outras ações são necessárias para a requisição ser completada.
Exemplos: 301 *Moved Permanently* (Movido permanentemente), 302 *Moved Temporarily* (movido temporariamente).
- *Client Error* (4xx): a requisição contém sintaxe errada ou não pode ser completada pelo servidor.
Exemplos: 403 *Forbidden* (não permitido), 404 *Not Found* (não encontrado).
- *Server Error* (5xx): o servidor falhou em processar uma requisição aparentemente válida.
Exemplos: 500 *Internal Server Error* (erro interno do servidor), 505 *SIP Version Not Supported* (versão do protocolo SIP não suportada).
- *Global Failure* (6xx): a requisição não foi completada em nenhum servidor.
Exemplos: 600 *Busy Everywhere* (todas a localidades ocupadas), 604 *Does Not Exist Anywhere* (não existe a requisição em nenhum lugar conhecido)

Exemplo de uma mensagem SIP do tipo resposta:

```
SIP/2.0 200 OK
Via: SIP/2.0/UDP server10.biloxi.com
Via: SIP/2.0/UDP bigbox3.site3.atlanta.com
Via: SIP/2.0/UDP 12.26.17.91:5060
To: Bob <sip:bob@biloxi.com;tag=a6c85cf>
From: Emmonks <sip:emmonks@voip.com.br;tag=1928301774>
Call-ID: a84b4c76e66710@12.26.17.91
CSeq: 314159 INVITE
Contact: <sip:bob@biloxi.com>
Content-Type: application/sdp
Content-Length: 131
```

No cabeçalho de uma resposta SIP, a primeira linha contém o número da versão do protocolo SIP, o código e o nome da resposta. As linhas seguintes correspondem ao campos do cabeçalho. Os campos **Via**, **To**, **From**, **Call-ID** e **CSeq** são copiados da requisição **INVITE**. No exemplo, o campo **Via**, foi modificado porque foram adicionados ao caminho o UAC de *Emmonks*, o *proxy* do domínio *atlanta.com* e o *proxy* do domínio *biloxi.com*. O agente *Bob* adicionou uma etiqueta (*tag*) no campo **To**. Esta etiqueta é incorporada por ambos os participantes no diálogo e é incluída em todas as requisições e respostas futuras da chamada.

2.4.1.2 SDP (*Session Description Protocol*)

O protocolo SDP definido em (HANDLEY; JACOBSON, 1998) tem como função principal descrever o conteúdo das sessões. No caso de VoIP, é usado em conjunto com o protocolo de sinalização SIP para descrever a formatação das chamadas telefônicas.

O SDP possui indicadores que descrevem informações sobre:

- Fluxo de mídias: uma sessão pode incluir múltiplos fluxos de mídia com conteúdos diferentes. O SDP atualmente define áudio, vídeo, dados, controle e aplicações como tipos de fluxos, similar aos tipos MIME (*Multipurpose Internet Mail Extensions*) utilizados no serviço de *e-mail*.
- Endereçamento: indica os endereços de destino, os quais podem ser endereços de *multicast*, para o fluxo de mídia.
- Portas: para cada fluxo, as portas UDP para envio e recebimento são especificadas. Diferente dos protocolos de sinalização SIP ou H.323, que possuem portas fixas para a comunicação, os fluxos de mídias usam faixas de portas altas UDP. Devido ao uso de firewalls e NAT (*Network Address Translation*) esta característica dificulta a comunicação entre os terminais, principalmente em redes corporativas.
- Tipos de Conteúdo: para cada tipo de fluxo, o tipo de conteúdo indica quais os formatos que podem ser utilizados durante a sessão. No caso de SIP, são indicados os tipos de codificadores que podem ser utilizados pelos agentes envolvidos no diálogo.
- Tempos de Início e Parada: são aplicados para sessões em *broadcast*, por exemplo, um programa de rádio ou televisão, para indicar os tempos de início, parada e número de repetições.
- Originador: para sessões do tipo *broadcast*, o originador é especificado, com informações para contato. Isto pode ser importante para o receptor reportar problemas técnicos a fonte da transmissão.

Embora na primeira versão do protocolo SDP fosse possível descrever o conteúdo multimídia, faltavam mecanismos de negociação das capacidades entre os participantes da sessão, como ocorre na recomendação H.323. Este problema foi resolvido em (ROSENBERG; SCHULZRINNE, 2002b), onde foi definido um modelo de pergunta/resposta, no qual as partes da sessão trocam mensagens SDP para negociar a natureza do conteúdo multimídia a ser transmitido. Este mecanismo permite que haja a negociação do tipo de *codec* e capacidades avançadas entre os agentes envolvidos na sessão.

Abaixo, uma captura do corpo de uma mensagem SDP, apresentando os *codecs* disponíveis para a conversação entre os usuários dos ramais 18 e 1003:

```
INVITE sip:18@200.17.170.37 SIP/2.0
Via: SIP/2.0/UDP 192.168.200.200:5060
From: Eduardo Monks <sip:1003@200.17.170.37>;tag=2786616179
To: <sip:18@200.17.170.37>
Contact: <sip:1003@192.168.200.200:5060>
Call-ID: 519C8A6B-4ADB-44F6-AF0A-BD579B0A4F36@192.168.200.200
CSeq: 12705 INVITE
Max-Forwards: 70
Content-Type: application/sdp
User-Agent: X-Lite release 1105x
Content-Length: 313

v=0
o=1003 27735444 27735459 IN IP4 192.168.200.200
s=X-Lite
c=IN IP4 192.168.200.200
t=0 0
m=audio 8000 RTP/AVP 0 8 3 98 110 101
a=rtpmap:0 pcmu/8000
a=rtpmap:8 pcma/8000
a=rtpmap:3 gsm/8000
a=rtpmap:98 iLBC/8000
a=rtpmap:110 speex/8000
a=rtpmap:101 telephone-event/8000
a=fmtp:101 0-15
a=sendrecv
```

2.4.1.3 H.323

A recomendação H.323 do ITU é composta por protocolos de conferência multimídia, que incluem voz, vídeo e dados para uso sobre rede de pacotes (ITU-T, 1997). Sua primeira versão foi aprovada em fevereiro de 1996, no mesmo mês do primeiro *draft* do protocolo SIP. Projetada para operar sobre redes complexas, teve como foco inicial o uso em LANs (*Local Area Network*) e foi estendida para redes WANs (*Wide Area Network*) nas versões seguintes. Embora na configuração básica seja requerida somente a voz, o vídeo dominou as implementações iniciais e continua sendo o ponto forte do H.323. Foi o primeiro padrão para VoIP, sendo amplamente adotado pela indústria.

A recomendação é composta de uma série de documentos que descrevem as funcionalidades dos componentes e mecanismos que são utilizados para:

- sinalização;
- estabelecimento de sessões;
- controle de chamadas;
- gerenciamento de largura de banda;
- controle de admissão de chamadas;
- determinação dos *codecs* para transferência de áudio e vídeo;
- uso de protocolos de transferência de dados.

Os componentes de um sistema H.323 são os seguintes (figura 2.11):

- *Terminais*: são entidades do H.323 das extremidades de uma rede de transmissão multimídia. A comunicação é feita em duplo sentido e em tempo real com outros terminais H.323, através da transmissão e recepção de sinais de controle, áudio, vídeo e dados (isoladamente ou em conjunto).
- *Multipoint Control Units (MCUs)*: são componentes H.323 que fornecem a capacidade de suportar conferência multiponto com mais de dois terminais e/ou *gateways*.
- *Gateways*: são componentes opcionais em um sistema de conferência H.323. Eles fornecem muitos serviços, sendo que o mais comum é a tradução de formatos de sinais entre terminais H.323 e terminais de outros tipos ou tecnologias. Normalmente, tem a função de fazer a interconexão do serviço de VoIP com a rede pública de telefonia.
- *Gatekeeper*: fornece serviços de controle de chamadas para os terminais H.323 presentes na zona (*zone*). A zona é um conjunto de terminais, *gateways* e MCUs gerenciados por um único *gatekeeper* ou domínio administrativo. Suas funções são:
 - Controle de acesso (autorização de chamadas);
 - Tradução de endereços;
 - Gerenciamento de largura de banda;
 - Localização de *gateways*;
 - Gerenciamento de zona;
 - Sinalização de controle de chamada (no modo de conferência roteado);
 - Gerenciamento de chamadas.

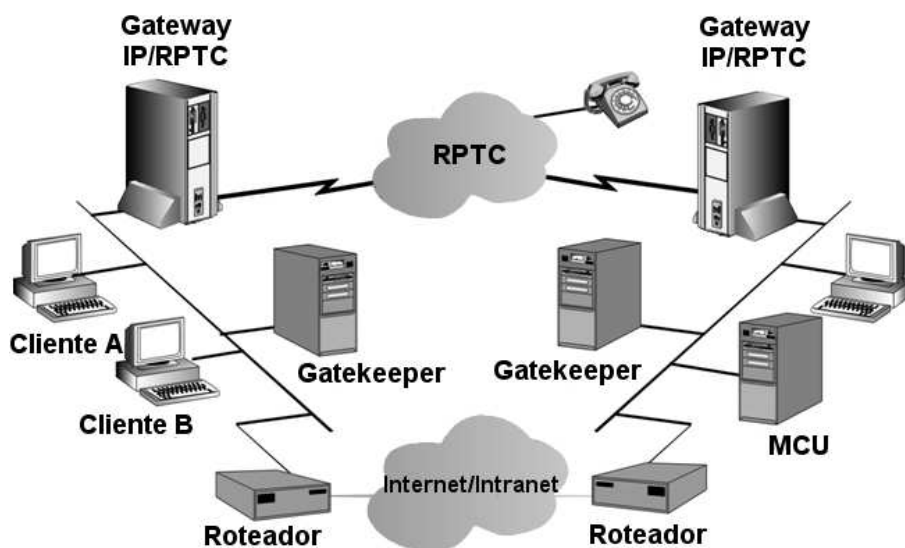


Figura 2.11: Componentes da recomendação H.323.

Os vários padrões e protocolos definidos na recomendação H.323 abrangem desde os *codecs* utilizados até a segurança da ligação, sendo os mais importantes para VoIP relacionados abaixo (figura 2.12):

- H.323 - é um documento que representa como os vários componentes se encaixam no sistema.
- H.225.0 - define a sinalização entre os terminais e o *gatekeeper*.
- RTP/RTCP - são usados para transmitir a mídia, áudio e vídeo, em redes IP. Estes protocolos são definidos pelo IETF, mas fazem parte da recomendação H.323.
- H.245 - é o protocolo usado para o controle do estabelecimento e fechamento dos canais de transferência de mídia e para o controle de conferências.
- H.450.x - é uma série de protocolos de serviço suplementares (chamadas em espera, transferências de chamadas).
- T.120 - especifica como tratar dados em conferências.
- T.38 - define como tratar o sinal de *fax*.
- H.235 - define a segurança do sistema.

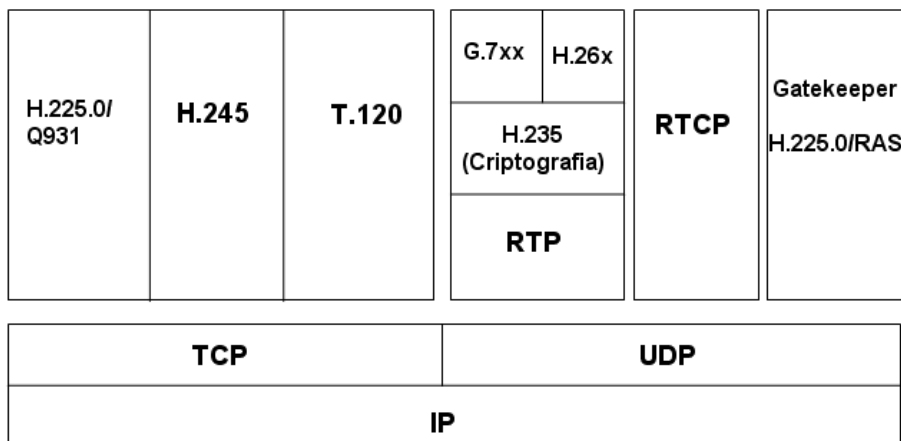


Figura 2.12: Pilha de protocolos H.323.

2.4.1.4 MGCP e Megaco/H.248

Os protocolos MGCP (*Media Gateway Control Protocol*) do IETF e a recomendação H.248 do ITU-T, também conhecida como Megaco (*Media Gateway Control*), são utilizados nos *gateways* para controle da mídia e sinalização. O Megaco é um protocolo desenvolvido em conjunto entre o IETF e o ITU-T, sendo que na nomenclatura do ITU é chamado de H.248.

O *gateway* tem a função de interconectar a rede de telefonia pública com os serviços de VoIP. Para isto são necessárias conversões para que os protocolos de sinalização da telefonia convencional sejam adaptados aos de VoIP e vice-versa. Além disto, existe a necessidade de converter os diversos tipos de codificadores que podem ser utilizados em VoIP, para o codificador padrão PCM (*mu-law* ou *a-law*) da telefonia convencional. O *gateway* pode ser um simples ATA (*Analog Telephone Adapter*) ou um equipamento que permita a interconexão de chamadas VoIP através de vários troncos de linhas integradas com a rede pública.

Originalmente, os *gateways* eram vistos como dispositivos monolíticos que faziam o controle das chamadas através de H.323 ou SIP e necessitavam de um hardware para interfacear com a rede pública de telefonia. Em 1998, a idéia de separar o *gateway* em duas partes lógicas foi proposta (DAVIDSON; JAMES, 2000). A parte que continha o controle da lógica da chamada passou a ser denominada de MGC (*Media Gateway Controller*), controlador de mídia do *gateway*, ou agente de chamadas. A outra parte, que faz a conexão com a rede pública, ficou denominada MG (*Media Gateway*), ou *gateway* de mídia. Com esta divisão, surgiu uma nova interface entre o MG e MGC. Para resolver as interações entre as duas partes do *gateway* foram definidos o protocolo MGCP e a recomendação H.248/Megaco. Os *Media Gateways* fornecem as seguintes funções de alto nível (DAVIDSON; JAMES, 2000):

- Terminação de troncos TDM oriundos da rede de telefonia pública ou de PABXs.
- Cancelamento de eco.
- Balanceamento dos *buffers* de *jitter*.
- *Voice Activity Detection* (VAD), supressão de silêncio e geração de ruído de conforto.
- Compressão da voz usando *codecs* tais como: G.711, G.723.1, e G.729.
- Geração de tons, que originam os tons de discagem, ocupado e congestionamento.
- Transporte de DTMF (*Dual-Tone Multi-Frequency*), que habilitam o uso de tons de toque para aplicações, tais como caixa postal de voz com *codecs* que suportam a detecção/transporte de DTMF.
- Conversão entre os *codecs mu-law* e *a-law* quando necessário.
- Suporte para *Quality of Service* (QoS).

2.4.2 Protocolos de Mídia

2.4.2.1 RTP/RTCP (*Real-Time Transport Protocol/Real-Time Transport Control Protocol*)

O RTP (*Real-Time Transport Protocol*) definido em (SCHULZRINNE et al., 1996) fornece funções de transporte de rede fim-a-fim adequado a aplicações que transmitem conteúdos de tempo real, tais como áudio e vídeo, sobre serviços de rede *unicast* ou *multicast*. O RTP não fornece reserva de recursos e não garante QoS para serviços de tempo real. O monitoramento da transmissão dos dados através de *feedback* do andamento da transmissão pode ser realizada pelo RTCP (*Real-Time Transport Control Protocol*) que funciona em conjunto ao RTP, embora não seja obrigatório a sua implementação. Ambos os protocolos, RTP e RTCP, usam o UDP como transporte. No caso de VoIP, o RTP é responsável pela transmissão dos pacotes de voz. O protocolo RTCP tem a função de passar estatísticas ao emissor e ao receptor da chamada, tais como atraso, *jitter* (variação do atraso) e perdas de pacotes. Tais estatísticas ajudam a determinar a qualidade da chamada.

O cabeçalho do RTP é ilustrado na figura 2.13. Ele consiste em três palavras de 32 bits e, potencialmente, algumas extensões. A primeira palavra contém o campo **Version**,

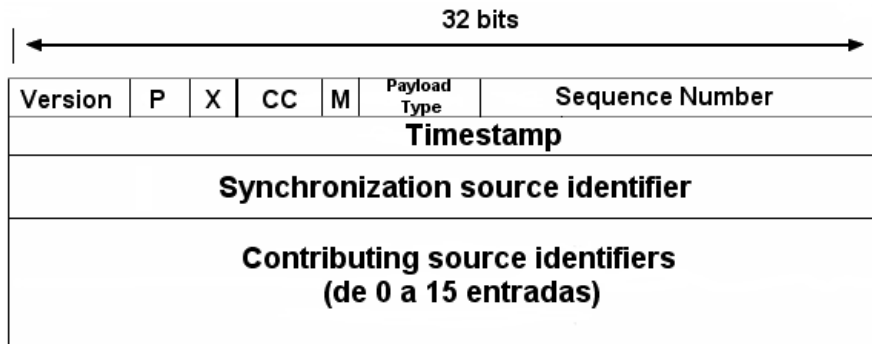


Figura 2.13: Cabeçalho do protocolo RTP.

atualmente na versão 2. O bit **P** indica que o pacote foi completado até chegar a um múltiplo de 4 bytes. O último byte de preenchimento informa quantos bytes foram acrescentados. O bit **X** indica que um cabeçalho de extensão está presente. O formato e o significado do cabeçalho de extensão não são definidos na RFC, deixando o cabeçalho flexível para expansões. O campo **CC** informa quantas origens de contribuição estão presentes, de 0 a 15. O bit **M** é um bit marcador específico da aplicação, pode servir para marcar o início de um quadro de vídeo ou o começo de uma palavra em um canal de áudio. O campo **Payload Type** informa o tipo de *codec* que está sendo utilizado na transmissão. O campo **Sequence Number** é um contador incrementado a cada pacote RTP enviado e tem a função de detectar a perda de pacotes. O timbre de hora, representado pelo campo **timestamp**, é produzido pela origem do fluxo para anotar quando a primeira amostra do pacote foi realizada. Através da diferença entre os valores deste campo na origem e no destino pode ser calculado o *jitter* e ajustado o *buffer de jitter*. O campo **Synchronization source identifier** informa a que fluxo o pacote pertence. Este método é usado para multiplexar e demultiplexar vários fluxos de dados em um único fluxo de pacotes UDP. Os campos **Contributing source identifiers**, se estiverem presentes, serão usados quando houverem misturadores (*mixers*) de áudio. Neste caso, o misturador será a origem de sincronização, e os fluxos que estão sendo mixados serão listados nesse campo. Na figura 2.14, um exemplo de um pacote RTP capturado com o analisador de pacotes Ethereal (ETHEREAL: A NETWORK PROTOCOL ANALYZER, 2005).

```

No.      Time      Source      Destination      Protocol Info
  8 0.158560 200.17.170.37 192.168.200.200  RTP      Payload type=I
TU-T G.711 PCMA, SSRC=47780547, Seq=31848, Time=1280

Frame 8 (214 bytes on wire, 96 bytes captured)
Ethernet II, Src: TflLan_06:14:32 (00:a0:4b:06:14:32), Dst: 3com_97:f8:b4 (00:01:02:97:f8:b4)
Internet Protocol, Src: 200.17.170.37 (200.17.170.37), Dst: 192.168.200.200 (192.168.200.200)
User Datagram Protocol, Src Port: 12054 (12054), Dst Port: 8000 (8000)
Real-Time Transport Protocol
 10.. .... = Version: RFC 1889 Version (2)
 ..0. .... = Padding: False
 ...0 .... = Extension: False
 .... 0000 = Contributing source identifiers count: 0
 0... .... = Marker: False
Payload type: ITU-T G.711 PCMA (8)
Sequence number: 31848
Timestamp: 1280
Synchronization source identifier: 47780547
Payload: 5DFAEF9F98878082828C8F8D8D828186859996EAF9DC716C...

```

Figura 2.14: Cabeçalho de um pacote RTP com voz, capturado com o software Ethereal.

2.5 Codificadores (*Codecs*)

O termo *codec* é um acrônimo de codificador/decodificador (*coder/decoder*). Basicamente, um *codec* é um algoritmo que faz a digitalização e a quantização do sinal de vídeo ou áudio reduzindo o número de bytes gerados e, conseqüentemente, diminuindo a largura de banda necessária para a transmissão ou o espaço para o armazenamento. No caso de VoIP, são usados *vocodecs* (*voice codecs*), que são *codecs* específicos para o espectro de frequência do sinal de voz.

O processo de digitalização da voz tem como principais vantagens a transmissão em longas distâncias e a multiplexação entre voz e dados. Em transmissões de longa distância, o sinal sofre atenuação e perdas inerentes ao meio de propagação. Com isto, o sinal sofre a ação de ruído. O processo de remoção dos ruídos é mais complexo em sinais analógicos do que em sinais digitais (MINOLI; MINOLI, 1998). Além disto, a amplificação que o sinal analógico sofre no decorrer da transmissão não diferencia o ruído do sinal de voz. No sinal digital, a filtragem do ruído é mais simples, pois antes da amplificação o sinal é regenerado pelos repetidores. Existem dois métodos básicos de digitalização e quantização: um baseado em amplitude do sinal no decorrer do tempo (*waveform*), e outro na análise do espectro de frequências (*source codecs*).

A qualidade da voz obtida através dos *codecs* é função do nível de compressão, da complexidade do algoritmo e do atraso no processamento. Usualmente, existe uma forte interdependência entre os fatores e devem ser balanceados para não prejudicar a qualidade da voz. No caso de VoIP, é comum a conversão entre diferentes tipos de *codecs*. Isto é ocasionado pelos diferentes tipos de terminais e pela integração com a RTPC. A conversão adiciona atraso na conversação e perda de qualidade da voz, devido às perdas de características do sinal no processo de codificação. Na figura 2.15 estão representados os tempos de conversão entre *codecs* em um servidor *Asterisk* (DIGIUM, 2005). Neste caso, o maior tempo de conversão entre *codecs* acontece entre o iLBC e o Speex, cerca de 58ms.

Tempos de conversão entre codecs (em milisegundos)											
	Formato Original (Linhas)						Formato Destino (Colunas)				
	g723	gsm	ulaw	alaw	g726	adpcm	slin	lpc10	g729	speex	ilbc
g723	-	-	-	-	-	-	-	-	-	-	-
gsm	-	-	2	2	4	2	1	4	-	45	21
ulaw	-	5	-	1	4	2	1	4	-	45	21
alaw	-	5	1	-	4	2	1	4	-	45	21
g726	-	6	3	3	-	3	2	5	-	46	22
adpcm	-	5	2	2	4	-	1	4	-	45	21
slin	-	4	1	1	3	1	-	3	-	44	20
lpc10	-	7	4	4	6	4	3	-	-	47	23
g729	-	-	-	-	-	-	-	-	-	-	-
speex	-	7	4	4	6	4	3	6	-	-	23
ilbc	-	18	15	15	17	15	14	17	-	58	-

Figura 2.15: Tempos de conversão entre *codecs*.

Atualmente, a maioria dos *codecs* de voz trabalham com largura de banda fixa, não dependendo das características do sinal de entrada. Entretanto, existem *codecs* que fazem uso de largura de banda variável, usando mecanismos de supressão de silêncio (*silence supression*) ou VAD (*Voice Activity Detection*) que não geram tráfego quando não há atividade de conversação. O desempenho do algoritmo de supressão de silêncio é crítico para

a qualidade da voz. Se existirem poucos intervalos de silêncio na conversação, os ganhos com a supressão de silêncio não se concretizam. Normalmente, a porcentagem média de atividade de conversação é de aproximadamente 40% do tempo da chamada (JIANG; SCHULZRINNE, 2000). Os problemas da técnica de supressão de silêncio são os ruídos do ambiente que confundem o algoritmo e a retomada da conversação, para resolver estes problemas podem ser utilizados *codecs* diferentes para envio da voz e para envio do ruído. Por exemplo, o ruído vai ser enviado com um *codec* que consuma menos largura de banda e com qualidade pior do que o *codec* usado para a voz. Se houver demora na detecção da retomada da conversação, o início da conversa pode ser prejudicado. Dependendo do *codec* utilizado e do terminal, o recurso pode ser ativado ou não. Outro recurso utilizado para melhoria da qualidade da chamada é a geração de ruído de conforto (*Comfort Noise Generation*). O ruído de conforto é gerado no receptor da chamada para reconstruir o ruído de fundo natural da telefonia, ao invés do silêncio total. Este mecanismo tem a função principal de indicar ao receptor que a chamada continua ativa, e é uma herança do sistema telefônico convencional.

A ocultação de pacotes perdidos (*Packet Loss Concealment*) é um mecanismo utilizado pelos *codecs* visando a diminuição do efeito da perda de pacotes. A perda de pacotes pode se dar por um atraso excessivo ou por problemas no meio físico. As perdas representam queda de qualidade na conversação, tornando-a muitas vezes inteligível. O mecanismo de ocultação de pacotes perdidos é utilizado no *codec* do destino, o qual preencherá as falhas no sinal de voz. O preenchimento se dá através da repetição do último *frame* recebido ou a geração de um trecho de voz artificial baseado nos *frames* recentemente recebidos. Tal compensação de falhas no sinal de voz podem reduzir consideravelmente os efeitos da perda de pacote para o usuário (HARDY, 2003).

Uma das principais características dos *codecs* é o fator MOS (*Mean Objective Score*). O MOS é um processo subjetivo de classificação da qualidade da voz realizado pelo ITU. A classificação vai de 1 a 5, sendo o valor 5 a melhor qualidade. Os *codecs* são classificados pelo fator MOS. O fator MOS será abordado com maiores detalhes no capítulo 3.

Diferente do que acontece na rede de telefonia pública, vários tipos diferentes de *codecs* podem ser utilizados em VoIP. Dependendo do método de compressão utilizado, a qualidade da voz e a largura de banda utilizada na chamada podem variar. Os métodos de compressão mais comuns são os seguintes:

- PCM (*Pulse Code Modulation*):
 - Utilizado na rede telefônica convencional;
 - Possui a melhor qualidade entre os *vocodecs* por não usar compressão.
- ADPCM (*Adaptative Differential Pulse Code Modulation*):
 - Versão com compressão do PCM;
 - Faz o cálculo da diferença entre duas amostras, reduzindo a banda para até 16kbts.
- CELP (*Code Excited Linear Prediction*):
 - Faz uso de *codebooks* que geram índices do padrão da voz;
 - São transmitidos apenas os índices das amostras;
 - Usa mais CPU, além do aumento no atraso fim a fim entre 50 e 100ms.

- LD-CELP (*Low Delay Code Excited Linear Prediction*):
 - Variação do CELP;
 - Atrasos bem menores, na ordem de 2ms.
- CS-ACELP (*Conjugate Structure Algebraic Code Excited Linear Prediction*):
 - Um dos métodos mais utilizados;
 - Possui ótima qualidade de voz e baixo consumo de banda.
- ACELP (*Algebraic Code Excited Linear Prediction*):
 - Possui um algoritmo de *look-ahead*, de 7,5 ms;
 - Adiciona atraso de 37,5 ms pois analisa 3 amostras de 10ms antes de formar um pacote;

Diversos codecs são padronizados pelo ITU e pelo IETF, mas existem *codecs* padronizados pela indústria e outros de código-fonte aberto. Os mais comumente encontrados nas aplicações de VoIP são o G.711 (*mu-law* e *a-law*), GSM, G.729 e iLBC. Na tabela 2.1, são elencadas as características destes e de outros *codecs*.

Tabela 2.1: Características dos *Codecs*.

<i>Codec</i>	Método de Codificação	Tamanho da Amostra (bytes)	Intervalo da Amostra (ms)	Banda (kbits)	MOS
GSM	RPE/LT	33	20	13,2	3,54
G.729a	CS-ACELP	10	10	8	3,92
G.711	PCM	80	10	64	4,1
G.723.1	MP MLQ	24	30	6,3	3,9
G.723.1	ACELP	20	30	5,3	3,8
G.726	ADPCM	15	5	24	não disponível
G.728	LD-CELP	10	5	16	3,61
iLBC	LPC	50 ou 38	20 ou 30	13,33 e 15	3,92
Speex-8k	CELP	20	20	2,15 a 24,6	não disponível

2.6 Arquiteturas

Existem três arquiteturas básicas para implementação de serviços VoIP. São elas:

- Ponto a Ponto
- Com *gateway*
- Híbrida

2.6.1 Ponto a Ponto

Nesta arquitetura, os pontos terminais fazendo uso de *softphones*, telefones IP ou clientes de serviços (Skype, FWD) conectam-se diretamente entre si (figura 2.16c). Para encontrar os pares da comunicação, são usados contatos externos ao serviço, tais como *e-mail* ou páginas *web*, ou pode ser usado algum recurso de procura. O recurso de procura é usado em serviços como o Skype de forma distribuída ou no caso do FWD de forma centralizada. O usuário faz a busca pelo nome ou número do par com o qual deseja falar. A numeração é gerada em cada serviço e, não necessariamente, corresponde a alguma forma de endereçamento similar ao encontrado na rede de telefonia pública. De posse do endereço do par, pode-se fazer a conexão com o destino. Normalmente, este tipo de arquitetura não possui tarifação.

2.6.2 Com gateway

O *gateway* tem a função de integrar os serviços VoIP com a rede de telefonia pública. Esta arquitetura é utilizada para trafegar as ligações através de serviços VoIP sem modificar o serviço de telefonia convencional nos terminais (figura 2.16b). Com isto, o tráfego de ligações entre as localidades conectadas com *gateways*, não será roteada pela rede de telefonia pública. Esta arquitetura é comumente encontrada em empresas que possuem unidades separadas geograficamente e interconectadas com enlaces de dados. Aproveitando a estrutura de enlaces de dados, é possível fazer o roteamento das chamadas entre as unidades sem passar pela rede pública de telefonia. Desta forma, economiza-se no custo das ligações telefônicas.

2.6.3 Híbrida

Neste caso, os pares da comunicação podem estar em uma rede de dados, o originador na rede de dados e o receptor na rede pública de telefonia e vice-versa, ou ambos na rede de telefonia pública sendo o tráfego transferido através de uma rede de dados (figura 2.16a). Portanto, devem haver conversões de codificadores e a numeração deve corresponder a numeração utilizada na rede de telefonia pública, pois os usuários de serviços VoIP devem poder enviar e receber ligações da rede pública.

2.7 Equipamentos

Os equipamentos utilizados nas arquiteturas de serviços VoIP são basicamente terminais e sistemas de controle. As funções realizadas pelos sistemas de controle recebem nomenclaturas diferentes e diferem em funcionalidades quando utilizadas com o protocolo SIP ou a recomendação H.323. Os equipamentos foram divididos de acordo com as suas funcionalidades. São eles Terminais, Centrais e *Gateways*.

2.7.1 Terminais

Os terminais têm como funções principais coletar a voz, codificá-la e transmiti-la. Devem obedecer aos protocolos e recomendações para a sinalização de chamadas. Os terminais podem ser em *hardware* ou *software*. Nos terminais podem ser configuradas as características dos *codecs*, tais como a número de amostras de áudio por pacote, pode ser ativado o recurso de VAD (*Voice Activity Detection*) e tamanho do *buffer* de *jitter*.

Em *hardware*, os telefones IP são os equipamentos mais comuns (figura 2.17). Existem diversos tipos de telefones IP, que são compatíveis com SIP, H.323 e protocolos pro-

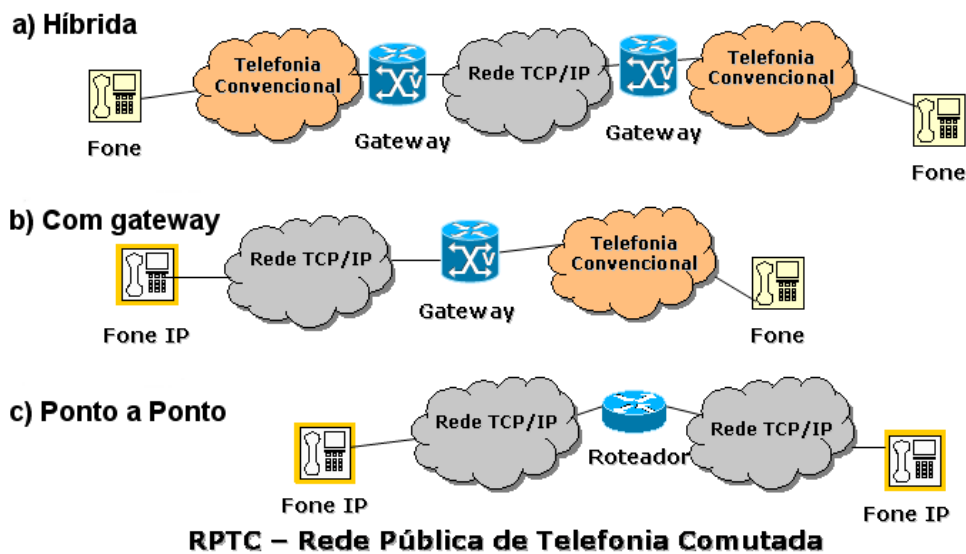


Figura 2.16: Arquiteturas de VoIP mais comuns.

prietários. O custo destes telefones é bem maior do que os telefones convencionais, além de possuírem a necessidade de alimentação externa. Isto pode causar problemas em caso de interrupção de energia elétrica. Alguns telefones IP permitem a configuração de *QoS*, sendo possível configurar a marcação de quadros ethernet no padrão IEEE 802.1p e do campo *ToS (Type of Service)* do protocolo IP.



Figura 2.17: Telefone IP Cisco modelo 7970.

Em *software*, são chamados *softphones* (figura 2.18) e possuem, normalmente, mais funcionalidades do que os telefones IP, tais como maior número de tipos de *codecs*, capacidade maior de agenda e integração com *browsers*. Além da compatibilidade com os protocolos padrão, existem *softphones* que obedecem a protocolos proprietários.

Para a interação com os usuários são necessários microfones, fones de ouvido e placas de som com capacidade *full-duplex*. Por ser executado em uma estação, todos os problemas que podem ocorrer na estação afetam o funcionamento dos *softphones*, tais como incidência de vírus, uso excessivo de processamento e reinicializações inesperadas do sistema operacional. Seu custo é muito menor do que um telefone IP, inclusive existindo vários *softphones* gratuitos.



Figura 2.18: Tela do *softphone* Xlite.

2.7.2 Centrais

As centrais têm a função de controlar e gerenciar as ligações entre os terminais. Podem ser equipamentos específicos, tais como PABXs IP, ou softwares que podem ser instalados em um PC comum de forma gratuita, tal como o Asterisk (DIGIUM, 2005). As funções de tarifação, políticas de uso, redirecionamento de chamadas, caixa postal, menu interativo e etc. são executadas nas centrais. No caso da recomendação H.323, o *Gatekeeper* é o responsável por estas funções. No caso do protocolo SIP, o *Proxy Server* tem estas funções.

2.7.3 Gateways

A função principal dos *gateways* é converter a sinalização de controle e o formato da mídia das chamadas telefônicas. Normalmente, tem a função de converter a sinalização e a mídia das chamadas entre a rede pública de telefonia e o serviço de VoIP. É usado também para adaptar a conversação entre SIP e H.323, convertendo a sinalização e se necessário, o formato da mídia através do processo de transcodificação. Os adaptadores para telefones convencionais chamados ATA (*Analog Telephone Adapter*) são *gateways* com a função de adaptar os telefones convencionais para uso com serviços VoIP. A inteligência da comunicação está embutida no ATA, tendo o telefone como a interface com o usuário (Figura 2.19).

2.8 Numeração

Cada terminal do sistema telefônico convencional, seja ele fixo ou celular, tem associado um conjunto de números ou códigos de acesso que permitem que ele seja identificado de forma exclusiva em todo o mundo. Para que isto fosse possível, o ITU-T estabeleceu a recomendação E.164 (ITU-T, 2005) que define a estrutura e a distribuição dos códigos telefônicos no mundo. A numeração é composta de 15 dígitos decimais, precedidos por um sinal de "+", que identificam unicamente cada terminal telefônico, seja fixo ou móvel. Estes números são gerenciados dentro dos diferentes planos de numeração nacionais, regionais ou globais. O ITU é responsável pela distribuição dos números e códigos de cada país, por exemplo 55 para o Brasil ou 1 para os EUA, para as regiões, por exemplo

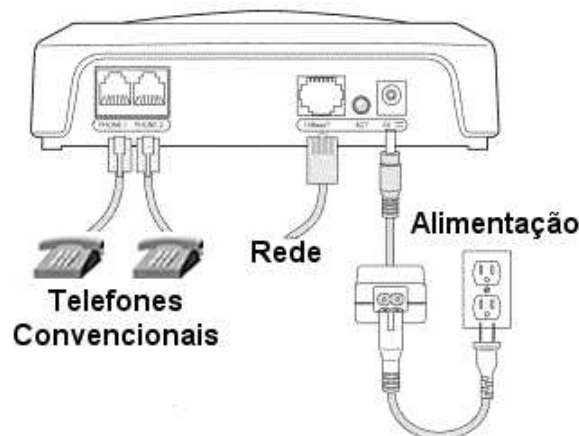


Figura 2.19: Conexões do Cisco 186 ATA (*Analog Telephone Adapter*).

3883 para Europa e números globais, por exemplo 00800 que define um número de ligação gratuita em todo o mundo.

Para integrar serviços VoIP de forma transparente no plano de numeração existente, foi criado o protocolo ENUM (FALTSTROM; MEALLING, 2004). O ENUM foi desenvolvido como solução para as questões de como elementos de rede poderiam usar serviços na Internet usando apenas números telefônicos e como telefones convencionais poderiam acessar serviços na Internet usando um teclado de 20 teclas. O ENUM, basicamente, tem a função de tornar transparente a convergência entre redes IP e a rede de telefonia pública fazendo o mapeamento de números telefônicos convencionais para as funcionalidades da Internet. O ENUM faz uso do serviço de DNS para transformar números no padrão E.164 em formato URI (HUSTON, 2002). Desta forma, o uso de endereços numéricos podem ser usados para representar qualquer tipo de serviço, através do registro de serviços ENUM, tais como e-mail ou endereços de páginas (IANA, 2005). Quando for digitado um endereço telefônico, ele irá ser mapeado para uma URI da seguinte forma (ENUM.ORG, 2005):

1. O endereço telefônico é traduzido para um número completo E.164, adicionado os códigos de área e do país.
Exemplo: O número 3316-6835 sendo discado em Porto Alegre, torna-se +55-51-33166835. Onde o "55" representa o código do país (Brasil), o "51" representa o código que representa uma área do estado do Rio Grande do Sul e o símbolo de "+" identifica um número E.164.
2. Todos os caracteres são removidos, exceto os dígitos.
Exemplo: 555133166835.
3. A ordem dos dígitos é invertida.
Exemplo: 538661331555
4. Pontos são colocados entre cada dígito.
Exemplo: 5.3.8.6.6.1.3.3.1.5.5.5
5. O domínio e164.arpa é acrescentado como sufixo.
Exemplo: 5.3.8.6.6.1.3.3.1.5.5.5.e164.arpa

Após esta transformação, o ENUM faz uma consulta ao servidor de DNS do domínio. Quando o servidor de nomes responsável (*authoritative name server*) é encontrado, o ENUM recupera informações relevantes através da entradas nos registros NAPTR (*Naming Authority Pointer*). Os registros NAPTR apontam para os serviços que estão configurados para o número telefônico cadastrado na base de dados do DNS. Abaixo um exemplo de entrada NAPTR:

```

$ORIGIN 5.3.8.6.6.1.3.3.1.5.5.5.e164.arpa.
NAPTR 10 100 "u" "E2U+sip" "!^.*$!sip:emmonks@inf.ufrgs.br!".
NAPTR 10 101 "u" "E2U+h323" "!^.*$!h323:info@inf.ufrgs.br".
NAPTR 10 102 "u" "E2U+msg" "!^.*$!mailto:emmonks@inf.ufrgs.br!".

```

Este registro tem a função de descrever que o domínio 5.3.8.6.6.1.3.3.1.5.5.5.e164.arpa. é preferencialmente contatado por voz primeiramente através de SIP, e após por H.323. Em caso de não haver sucesso em nenhum destes contatos, será tentado o contato por e-mail através do protocolo SMTP (*Simple Mail Transfer Protocol*). O código *E2U* determina que a entrada é uma regra de reescrita ENUM e as diretivas *sip*, *h323* e *msg* são tipos de serviços registrado pelo IANA para determinar o serviço a ser acessado (IANA, 2005). São usados mecanismos de expressões regulares para reescrever as regras. No exemplo acima, o trecho "*!^.*\$!sip:emmonks@inf.ufrgs.br!*" faz com que seja substituído o número telefônico 5.3.8.6.6.1.3.3.1.5.5.5.e164.arpa, pelo endereço *sip:emmonks@inf.ufrgs.br*.

Sendo assim, é possível utilizar a estrutura existente da Internet atual para integrar o plano de numeração da rede de telefonia pública com o serviços de VoIP de forma consistente.

2.9 Exemplos de Aplicação

2.9.1 Skype

Com quase 250 milhões de *downloads* do cliente, o Skype é um serviço de VoIP bastante utilizado atualmente (SKYPE, 2005). Baseado em protocolos P2P (*Peer to Peer*), o Skype é um serviço gratuito para comunicação de voz através da Internet entre os usuários. Possui integração com a rede de telefonia pública, através dos serviços *SkypeOut* e *SkypeIn*. O primeiro é um serviço que possibilita aos usuários do Skype fazerem ligações para usuários localizados na rede de telefonia pública. O segundo serviço tem a função de possibilitar que usuários localizados na rede pública de telefonia façam ligações para usuários no Skype. Ambos os serviços são tarifados, embora o custo seja bem abaixo dos praticados pelas empresas de telefonia convencional

Para usar o serviço, o usuário deve baixar o cliente e fazer um cadastro. O endereço para a localização do usuário no sistema não possui relação com a numeração em telefonia convencional. O usuário é encontrado por um endereço alfanumérico, tal como *emmonks_123*. No caso do uso do serviço *SkypeIn*, é fornecido um número para o usuário ser contatado.

Os protocolos e os *codecs* usados pelo Skype são proprietários e não são divulgados. Oficialmente, as informações públicas a respeito dos *codecs* utilizados reportam que o consumo de largura de banda fica entre 24 e 64 kbits, dependendo das condições da rede e processamento. Em (BASET; SCHULZRINNE, 2004) foi feito um levantamento do funcionamento do protocolo do Skype, onde foi relatado que os *codecs* utilizados são o iLBC e iSAC. Neste estudo, não foi possível determinar as trocas de mensagens porque são utilizados mecanismos de criptografia tanto na sinalização quanto nos fluxos de mídia.



Figura 2.20: Tela do cliente do serviço Skype.

2.9.2 Gizmo Project

O projeto Gizmo é baseado em padrões abertos, os quais são críticos para desenvolver o potencial de VoIP. Segundo a empresa SIPphone, responsável pelo projeto, as chamadas telefônicas através da Internet devem ser gratuitas tais como as páginas *web*, *e-mail* e mensagens instantâneas. E para atingir o maior número de usuários foi adotado o protocolo SIP para a sinalização (GIZMO PROJECT, 2005).

Para usar o serviço é necessário o *download* do cliente (*softphone*) e o cadastro do usuário. Após o cadastro, o usuário receberá um número de identificação no padrão 1-747-XXX-XXXX. O número é determinante para a integração com a rede de telefonia pública. Os serviços adicionais *CallOut* e *CallIn*, permitem fazer e receber ligações da rede pública de telefonia, respectivamente. A figura 2.21 apresenta o balanço dos créditos para as ligações com o serviço de *CallOut*. Estes serviços adicionais são pagos, mas as taxas são bem menores dos que as praticadas pelas empresas de telefonia convencionais.

Para os usuários domésticos obterem uma boa qualidade nas chamadas, é necessário possuir uma conexão de banda larga, embora um acesso discado possa ser usado. A largura de banda utilizada vai depender dos *codecs* utilizados. Abaixo o *codecs* usados pelo Gizmo:

- GSM
- iLBC
- iSAC
- G.711 (*mulaw e alaw*)
- EG711
- iPCM

Uma opção interessante para redes corporativas é o recurso de *peering*. Desta forma, usando um servidor SIP pode-se habilitar que os números internos da empresa sejam



Figura 2.21: Tela do cliente do Gizmo.

acessíveis externamente. Se for desejável, podem-se colocar placas para acesso a rede de telefonia convencional. Com isto, um ramal convencional teria a capacidade de fazer uma ligação para outro ramal convencional em outra empresa ou instituição. Todo o tráfego da chamada seria repassado pela Internet. Esta solução é de baixo custo e de fácil implementação (SIPPHONE, 2005b).

3 QUALIDADE DE SERVIÇO (QOS)

O QoS (Qualidade de Serviço) é uma medida coletiva de nível de serviço apresentado ao usuário. Pode ser considerado como sendo o nível de confiança na rede por determinada aplicação para atingir os requisitos necessários para o seu funcionamento (COLLINS, 2003). O QoS pode ser caracterizado por vários critérios de desempenho tais como disponibilidade, vazão, tempo de conexão, porcentagem de transmissões com sucesso e a velocidade da detecção e correção de falhas. O uso de QoS para serviços VoIP é um fator fundamental para garantia de qualidade da voz. A falta de mecanismos de QoS é uma das razões para que muitos dos serviços de VoIP pela Internet sejam, atualmente, gratuitos. A filosofia que se baseia esta estratégia é que dificilmente os usuários vão reclamar da qualidade das chamadas, se elas forem gratuitas. Entretanto, a qualidade das chamadas realizadas através de serviços como o Skype e o Gizmo Project são muito boas na maior parte das vezes, embora não exista garantia nenhuma de qualidade. Em uma rede corporativa, é possível e desejável garantir a qualidade das chamadas através de técnicas de QoS nos equipamentos de rede.

As medições da qualidade de serviço em telefonia convencional possuem características particulares, que serão descritas neste capítulo e a seguir as características de QoS para serviços VoIP.

3.1 QoS em Telefonia Convencional

3.1.1 Taxa de bloqueio (GOS)

A taxa de bloqueio (*Grade of Service ou GoS*) é a probabilidade de que uma nova tentativa de chamada seja bloqueada ou perdida. Na rede pública de telefonia a relação do número de usuários não é igual ao número de linhas disponíveis, portanto os bloqueios irão acontecer inevitavelmente. Em geral, é esperado que cerca de 5 a 8% dos usuários de uma central façam ligações simultâneas nos períodos de maior movimento, sendo que cada telefone está ocupado de 10 a 16% do tempo (IVERSEN et al., 2005).

Em VoIP, o GoS é um fator de QoS quando existe um *gateway* envolvido na chamada VoIP ou está sendo usada alguma forma de controle de admissão de chamadas (*call admission control*). O controle de admissão de chamadas tem a função de limitar o número de ligações simultâneas na rede. Desta forma, o planejamento de capacidade da rede é garantido e os recursos da rede não são esgotados. No caso da arquitetura SIP, o controle é feito nos *proxies*, embora não exista nenhum mecanismo padronizado para tal. No caso do H.323, o controle de admissão de chamadas é definido na recomendação H.323 e está presente nas funcionalidades do *Gatekeeper*. O GoS é representado na forma de um fator de bloqueio P.xx, onde xx é a porcentagem das chamadas que serão bloqueadas ou per-

didadas em determinado sistema. Por exemplo, um fator de P.01 define a probabilidade de que 1% das chamadas dos usuários possam ser bloqueadas.

Um fator de P.00 em telefonia convencional é raramente requisitado e raramente acontecerá porque 100 % de não existência de bloqueio, significa dizer que o projeto da rede deve manter a razão de 1:1 entre usuários e linhas. Já em redes de pacotes, esta relação passa a ser possível. Devido ao compartilhamento dos canais de transmissão entre os usuários, apenas a limitação dos recursos da rede, principalmente a largura de banda, são os fatores limitantes para as chamadas serem realizadas.

A contabilidade do GoS não leva em conta as ligações que não foram completadas em razão do usuário estar ocupado ou não estar disponível. São contabilizados somente os bloqueios originados por falhas ou congestionamento na rede.

O fator GoS tem papel fundamental no planejamento de capacidade de um sistema telefônico.

3.1.2 Disponibilidade

O sistema de telefonia pública convencional é considerado robusto e confiável. Os usuários de telefonia estão habituados a retirar o fone do gancho e quase invariavelmente escutar o tom de discagem. A real porcentagem de disponibilidade do serviço de telefonia é de 99,999 %. Isto significa que o serviço fica indisponível para uso por apenas 5 minutos ao ano. Pode-se creditar estas taxas de disponibilidade, a simplicidade do sistema e a preocupação dos fabricantes em produzirem equipamentos confiáveis.

Já as redes de dados obtêm valores de disponibilidade bem piores. Devido à complexidade do uso de múltiplos protocolos, e à diversidade de fabricantes, provedores de serviços, sistemas operacionais, sistemas de gerenciamento de redes e outros, é quase impossível para as redes de dados alcançarem o nível de confiabilidade e disponibilidade da telefonia convencional. Além disto, problemas de segurança tais como vírus e ataques de negação de serviço afetam a disponibilidade de forma negativa. As melhores redes privadas estão disponíveis cerca de 94 % do tempo, em média. Isto significa que os usuários ficam sem acesso aos serviços na rede de dados 22 dias durante o ano (CHONG; MATTHEWS, 2004).

Em se tratando da Internet pública com um todo, a disponibilidade chega a apenas 61% do tempo, o que em um ano representa 142 dias indisponíveis (CHONG; MATTHEWS, 2004).

3.1.3 Metas de Qualidade - ANATEL (Agência Nacional de Telecomunicações)

O Plano Geral de Metas de Qualidade (PGMQ) (ANATEL, 1998) foi criado pela ANATEL para garantir a qualidade dos serviços de telefonia fixa e é aplicável às operadoras de telefonia fixa no Brasil. O PGQM estabelece que 65% das chamadas originadas por usuário têm que ser completadas. A chamada é considerada completada quando, após a discagem, a chamada resulta em comunicação com o destino desejado. As razões para não completar uma chamada podem ser:

- O terminal chamado não atende a chamada.
- O terminal chamado está ocupado.
- O número discado não existe ou foi discado incorretamente.
- Congestionamento na rede.

O PGMQ estabelece como meta que o número de chamadas não completadas por congestionamento na rede seja menor que 4%, GOS de P.04, das chamadas em cada um dos seguintes Períodos de Maior Movimento (PMM) conforme a tabela 3.1:

Tabela 3.1: Períodos de Maior Movimento (PMM).

PMM	Intervalo
Matutino	9hs às 11hs
Vespertino	14hs às 16hs
Noturno	20hs às 22hs

Além do monitoramento do complemento de chamadas, o PGMQ define os seguintes indicadores de qualidade (ANATEL, 1998):

- Indicadores de Complemento de Chamadas Locais e Longa Distância
 - Devem ser completadas pelo menos 60% das chamadas locais e de longa distância, nos períodos matutino, vespertino e noturno.
 - Não pode ultrapassar 4% o número de chamadas (locais e de longa distância) não completadas por congestionamento na rede de telecomunicações. Isso vale para os períodos matutino, vespertino e noturno.
- Indicadores de Mudança de Endereço
 - As solicitações de serviço de mudança de endereço dentro de uma mesma área local devem obedecer às seguintes metas de cumprimento de prazo:
 - Usuários residenciais: pelo menos 95% das solicitações atendidas em até 3 dias úteis;
 - Usuários não-residenciais: pelo menos 95% das solicitações atendidas em até 24 horas;
 - Usuários prestadores de serviços de utilidade pública: pelo menos 98% das solicitações atendidas em até 6 horas.
- Indicadores de Atendimento ao Cliente
 - Pelo menos 92% das chamadas de usuários deverão ser atendidas em até 10 segundos pelos serviços de auto-atendimento ou pela intervenção de telefonistas. Isso vale para os períodos matutino, vespertino e noturno.
 - Deverão ser respondidos em até 30 segundos pelo menos 95% dos pedidos de informação de números de telefone. Essa regra vale para as solicitações feitas através do “auxílio à lista”, também conhecido como “102”.
 - Toda correspondência dos usuários deve ser atendida em até cinco dias úteis após seu registro de entrada na operadora.
 - Pelo menos 95% dos clientes que se dirigirem aos postos de atendimento devem ser atendidos no prazo máximo de 10 minutos.
- Indicadores de Reparo de Telefone Convencional
 - Para cada 100 linhas em serviço numa localidade, não pode haver mais do que três solicitações de reparos no mês.

- Os prazos para atendimento das solicitações de reparos são os seguintes:
 - Usuários residenciais: pelo menos 95% das solicitações atendidas em até 24 horas.
 - Usuários não-residenciais: pelo menos 95% das solicitações atendidas em até oito horas;
 - Usuários prestadores de serviços de utilidade pública: pelo menos 98% das solicitações atendidas em até duas horas.
- Indicadores de Reparo de Telefone de Uso Público
 - Para cada 100 telefones de uso público em serviço numa localidade, não pode haver mais do que 15 solicitações de reparos no mês.
 - Os prazos para atendimento das solicitações de reparos de telefone de uso público são os seguintes: pelo menos 95% das solicitações atendidas em até oito horas.
- Indicadores de Erro em Conta
 - A cada mil contas emitidas (local e longa distância), no máximo quatro podem apresentar reclamação de erro.
 - Pelo menos 95% das contas contestadas (local e longa distância) que tenham direito a crédito deverão ter ressarcimento antes da emissão da próxima conta, em forma de crédito.
- Indicador de Sinal de Discar
 - O tempo máximo de espera pelo sinal de discar deve ser de apenas três segundos em pelo menos 98% das ligações. Isso vale para os períodos matutino, vespertino e noturno.
- Indicador de Modernização da Rede
 - Pelo menos 75% da rede local deverá estar digitalizada.

Estes indicadores e os resultados atuais das operadoras de telefonia fixa podem ser visualizados em (ANATEL, 2005). No caso de serviços VoIP, os seguintes indicadores da PGMQ seriam aplicáveis:

- Indicadores de Complemento de Chamadas Locais e Longa Distância
- Indicador de Sinal de Discar

Não existe legislação específica da ANATEL para os serviços VoIP em relação a qualidade de serviço. Existe um anexo (ANATEL, 2001), que trata de serviços de comunicação multimídia, mas não especifica nenhum indicador de qualidade de serviço. Portanto, os indicadores de qualidade usados na telefonia fixa comutada podem ser usados como parâmetros para o planejamento da capacidade da rede com serviços VoIP.

3.2 Fatores de QoS

Em redes de pacotes, os fatores de QoS são diferentes dos encontrados na telefonia convencional. Devido à multiplexação do tráfego, a concorrência por recursos de rede poderá ocorrer de forma imprevisível. Diferente do que acontece na telefonia convencional, onde são alocados recursos exclusivos para cada chamada.

A QoS oferecida pelas redes de pacotes pode ser representada através de parâmetros que indicam o comportamento do tráfego. Portanto, a garantia de serviços oferecida por redes de pacotes é medida através de parâmetros de qualidade de serviço. Os principais parâmetros de QoS do ponto de vista da rede são:

- Atraso
- Variação de Atraso (*Jitter*)
- Perdas de Pacotes

3.2.1 Atraso

O atraso em VoIP é caracterizado como sendo a quantidade de tempo que leva para a fala sair da boca de um interlocutor e chegar ao ouvido do receptor (DAVIDSON; JAMES, 2000). O ITU através da recomendação G.114 especifica que para se obter boa qualidade de voz, não mais de 150ms deve ser o atraso fim a fim de um participante da conversação. Na especificação também é definido que não pode ser ultrapassado os 400ms de atraso fim a fim de um participante. Se ultrapassado este limite, a conversação será excessivamente prejudicada. A recomendação G.114 do ITU-T define três faixas de aceitação para o atraso:

- 0-150 ms: aceitável para a maioria das aplicações.
- 150-400 ms: aceitável, desde que o administrador tenha conhecimento do tempo de transmissão e seu impacto na qualidade da transmissão para aplicações do usuário.
- Acima de 400 ms: inaceitável para propósitos gerais, entretanto, em alguns casos excepcionais este limite pode ser excedido.

Um pacote de voz é movido através da rede a um ponto destino. Os tempos dos processos de codificação e empacotamento na origem e os tempos dos processos de eliminação de *jitter* e decodificação no destino, somados, formam o atraso fim a fim ou atraso boca ao ouvido (*mouth-to-ear*). São contabilizados no atraso fim a fim os tempos de propagação do sinal na rede. Na figura 3.1, uma comparação entre o sinal original da voz e o sinal resultante da chegada ao destino.

Existem dois tipos de atraso, o fixo e o variável (CISCO, 2005). Os componentes do atraso fixo são adicionados diretamente ao atraso total da conexão (figura 3.2). Os atrasos variáveis são originados nos atrasos das filas dos roteadores, *switches* e *gateways* e no tempo de propagação dos pacotes através da rede.

Tipos de atrasos fixos:

- Atraso de processamento do *codec* - é o tempo gasto pelo processador de sinais digitais (DSP - *Digital Signal Processor*) para comprimir um bloco de amostras PCM. O tempo varia de acordo com o tipo de *codec* e o processador utilizado.

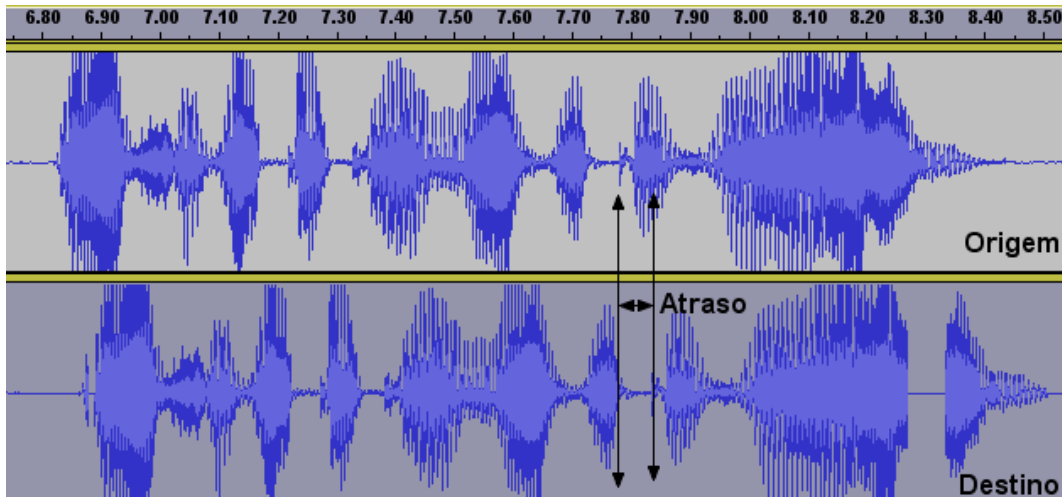


Figura 3.1: Efeito do atraso no sinal de voz.

Por exemplo, algoritmos do tipo ACELP (*Algebraic Code Excited Linear Prediction*) analisam blocos de 10ms de amostras PCM, e após é realizada a compressão. O tempo de compressão usado no algoritmo CS-ACELP (*Conjugate Structure Algebraic Code Excited Linear Prediction*) varia de 2,5ms a 10ms dependendo da carga no processador DSP. Normalmente, é utilizado o pior caso nos cálculos de atraso. O tempo de descompressão corresponde aproximadamente a 10% do tempo de compressão para cada bloco de amostras. Entretanto, o tempo de descompressão é proporcional ao número de amostras por *frames*. Conseqüentemente, o pior caso de descompressão para um *frame* com três amostras, com o algoritmo CS-ACELP, é 3ms (3 x 1ms). Usualmente, de dois ou três blocos comprimidos com o *codec* G.729 são colocados em um *frame* enquanto no caso do *codec* G.723.1 é colocada apenas uma amostra por *frame*.

- Atraso de Algoritmo - em *codecs* do tipo *vocoders*, os algoritmos de compressão baseiam-se em características conhecidas da voz para processar corretamente as amostras de um determinado bloco. O algoritmo deve ter algum conhecimento do que está contido no bloco seguinte para reproduzir com fidelidade o bloco atual. O processo de conhecer o bloco posterior antes do envio do bloco atual é denominado *look-ahead*. Este processo, que acontece repetidamente, gera atraso adicional chamado atraso de algoritmo. Por exemplo, no caso do *codec* G.729 é de 5ms e no caso do G.723.1 é de 7,5ms.
- Atraso de Empacotamento - é o tempo gasto para preencher os dados (*payload*) dos pacotes com voz codificada e comprimida. Este atraso é função do tamanho do bloco de amostra requerido pelo *codec* e do número de blocos alocados em um único *frame*. O atraso de empacotamento também é chamado de atraso de acumulação, pois as amostras de voz são acumuladas em um *buffer* antes de serem enviadas. Segundo (CISCO, 2005), como regra geral o atraso de empacotamento não deve ultrapassar os 30ms.
- Atraso de Serialização - é o tempo utilizado para o *frame* contendo voz ou dados ser repassado para a interface de rede. É diretamente relacionado com a velocidade do enlace. Devido às velocidades dos enlaces utilizados atualmente, o impacto deste

atraso é mínimo. Para enlaces superiores a 768kbits o atraso é desprezível.

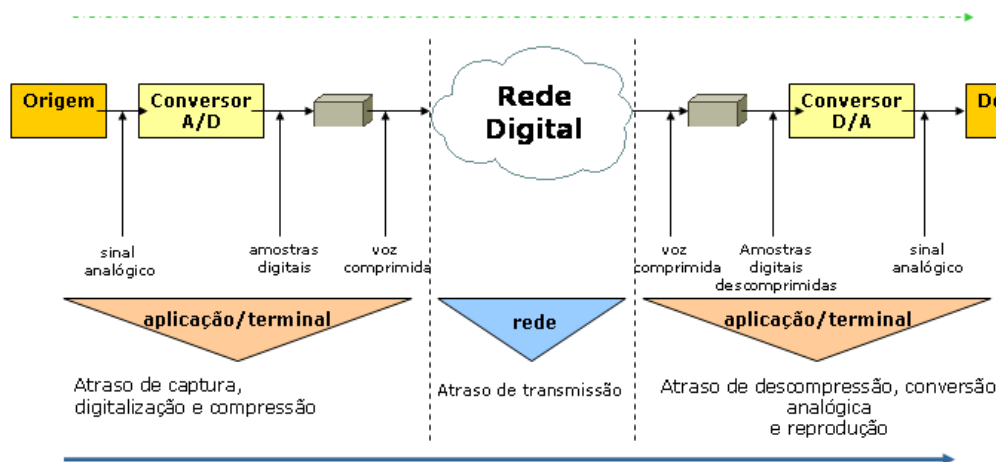


Figura 3.2: Diagrama de inserção de atrasos na transmissão.

- **Atraso de Enfileiramento/Acumulação** - após a voz ser comprimida, um cabeçalho é adicionado e o *frame* vai para fila de transmissão na conexão de rede. O tráfego de voz deve ter prioridade absoluta nas filas dos roteadores, *gateways* e *switches*. Por isto, um *frame* de voz só deve esperar na fila em casos em que um *frame* de dados já estar sendo transmitido ou aguardar que outros *frames* de voz sejam transmitidos. Essencialmente, o *frame* de voz espera o tempo de serialização de outros *frames* prévios na fila de saída. O atraso de enfileiramento é um atraso variável e é dependente da velocidade do enlace e do estado da fila. Um *frame* de voz deve ter a garantia (prioridade) de não esperar mais de um *frame* de dados para ser enviado ao meio físico.
- **Atraso de eliminação de Jitter** - o *buffer de jitter* tem a função de transformar o atraso variável em um atraso fixo. É um *buffer* que armazena a primeira amostra recebida por um certo período, até que um bloco de amostras seja formado e só após será liberada a reprodução. O período de armazenamento é chamado de atraso inicial de reprodução. Este tempo é essencial para o funcionamento eficiente e para eliminação do *jitter*. Se as amostras forem armazenadas por um curto período de tempo, as variações no atraso podem causar falhas na fala porque o *buffer* não será preenchido. Se as amostras forem armazenadas por um período muito longo, o *buffer* pode esgotar, e as amostras serão descartadas causando falhas na fala. Quando o período de armazenamento for muito longo, o tempo de atraso total fim a fim poderá atingir níveis inaceitáveis e tornar a qualidade da chamada ruim. O valor ótimo de tempo que um *buffer* deve armazenar é o tempo total do atraso variável na conexão. O valor do tempo de *buffer* pode ser adaptativo, mas o atraso máximo é fixo. Quando são configurados *buffers* adaptativos, o atraso torna-se variável. Entretanto, o atraso máximo pode ser usado como pior caso para propósito de projeto.

Na tabela 3.2, estão tabulados os atrasos resultantes de uma transmissão utilizando o codec G.729. O atraso denominado **Rede** corresponde à latência inserida no tráfego do pacote através da rede, que é variável.

Tabela 3.2: Origem dos atrasos usando o *codec* G.729 (GOODE, 2002).

Origem	Atraso (ms)
Captura da amostra pelo dispositivo	0,1
Codificação	17,5
Empacotamento	20
Enfileiramento (Emissor)	0,5
Acesso ao meio (Emissor)	10
Transmissão	<i>Rede</i>
Acesso ao meio (Receptor)	10
Enfileiramento (Receptor)	0,5
Equalização de <i>Jitter</i>	60
Decodificação	2
Reprodução	0,5
Total	121,1+ <i>Rede</i>
Rede Local	121,1+<10

3.2.2 Variação do Atraso (*Jitter*)

O atraso (latência) é representado como um valor de média. Entretanto, o serviço de VoIP é sensível a picos de atraso, o que torna o valor médio não muito apropriado para medir o efeito do atraso na qualidade da chamada.

Quando um fluxo de pacotes é transmitido através de uma rede IP, não existe garantia de que cada pacote realizará o mesmo trajeto pela rede. Isto ocorre devido às rotas alternativas que podem ser dinâmicas e variar de acordo com as condições da rede. Por não seguirem o mesmo caminho, os intervalos entre as chegadas dos pacotes podem variar. Quando um pacote percorrer mais passagens por roteadores (*hops*) do que os outros do mesmo fluxo, pode ser ocasionado um atraso considerável e uma latência maior. Além disto, partes da rede podem estar momentaneamente congestionadas, por exemplo, um roteador pode estar sobrecarregado devido ao tráfego intenso.

Caminhos alternativos e sobrecarga na rede, tanto em tráfego quanto em equipamentos, fazem com que a latência torne-se irregular, e estes atrasos irregulares são chamados de *jitter*. O efeito do *jitter* em VoIP corresponde à diminuição da qualidade da voz, uma vez que os intervalos irregulares tornam a conversação prejudicada. O *jitter* torna-se mais significativo quando é utilizada a Internet para trafegar serviços de VoIP, principalmente em horários de maior tráfego. Para minimizar este efeito é usado o recurso de *buffer* de *jitter*, que será abordado na seção 3.4.3.

O máximo *jitter* para os pacotes serem entregues de forma apropriada é de 75ms (CHONG; MATTHEWS, 2004). Na figura 3.3, o trecho de sinal de voz analisado está com 10ms de *jitter*, pois o *Atraso1* é de aproximadamente 58ms e o *Atraso2* é de 68ms.

3.2.3 Perda de Pacotes

As redes de pacotes sem recursos de QoS não tratam os pacotes de voz de forma diferenciada. Portanto, em casos de congestionamento e tráfego intenso os pacotes de voz serão descartados sem distinção dos pacotes de dados. No caso dos pacotes de dados, a retransmissão resolve o problema das perdas sem maiores prejuízos para a aplicação. Mas, quando há perda de pacotes de voz a retransmissão não é viável. Cada pacote de voz pode carregar de 10 a 80ms de voz, e dependendo da parte da conversação onde ocor-

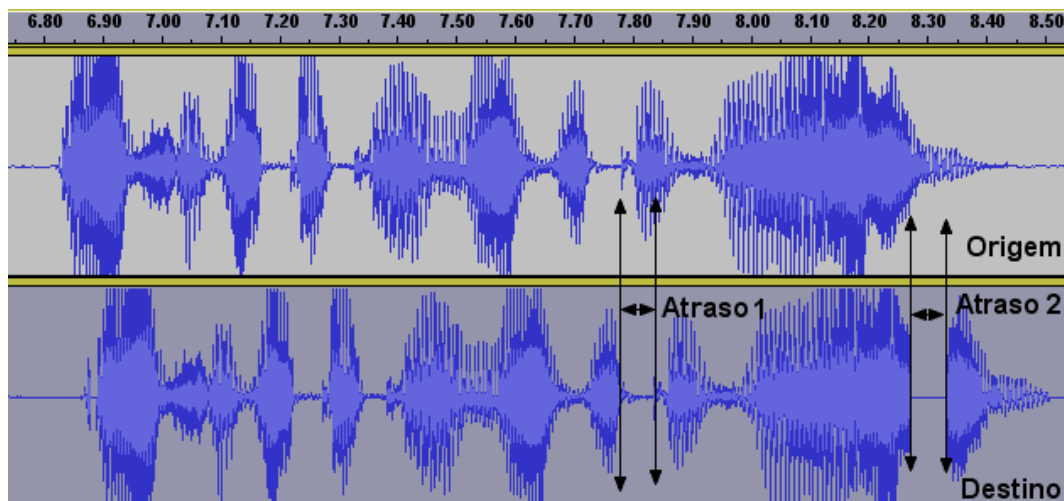


Figura 3.3: Efeito do *jitter* no sinal de voz.

rem as perdas, pode inutilizar o diálogo entre os usuários. Até mesmo 1% de perdas de pacotes, no caso de uso do *codec* G.711, podem significativamente degradar a qualidade da conversação (COLE; ROSENBLUTH, 2001). Outros *codecs* podem causar maior degradação da conversação devido à compressão por eles imposta. A compressão deixa os pacotes com maior tempo de sinal de voz e cada pacote perdido corresponderá a um trecho maior da conversação.

Os pacotes perdidos são ignorados no cálculo do valor de *jitter*, já que podem ser considerados pacotes com atraso infinito e prejudicariam o resultado. A taxa máxima de perdas de pacotes aceitável é de 3% (CHONG; MATTHEWS, 2004), mas *codecs*, tais como iLBC (ILBCFREEWARE.ORG, 2005) e Speex (SPEEX.ORG, 2005) possuem maior tolerância a perdas.

A perda de pacotes na rede ocorre quando os pacotes são enviados, mas não são recebidos no destino devido à problemas ocorridos na rede. Detectar exatamente os problemas de perdas de pacotes é difícil de fazer porque cada *codec* possui o seu método de ocultação de erros. Portanto, é possível que a qualidade da voz seja melhor usando um *codec* com compressão, tal como o G.729A, do que um *codec* que não use, como no caso do G.711 na mesma taxa de perdas.

Diversos fatores fazem com que os efeitos dos pacotes perdidos nas chamadas sejam variáveis. Entre eles estão (AVAYA, 2002):

- Os efeitos são mais sentidos em pequenas rajadas de perdas do que em perdas randômicas no decorrer do tempo. A perda de 10 pacotes contíguos causa maior prejuízo a conversação do que a perda de 10 pacotes distribuídos durante certo período de tempo.
- A perda de pacotes é sentida com maior intensidade para pacotes com áreas de dados maiores do que em pacotes com áreas menores. Isto ocorre porque mais trechos da fala são perdidos com áreas de dados maiores. O tamanho das áreas de dados são definidos nas configurações dos *codecs*.
- O efeito da perda de pacotes varia de acordo com a tolerância de cada *codec*.

Segundo (WALLINGFORD, 2005), o máximo de perda de pacotes entre dois pontos devem ser menor de 1%, para garantir que as chamadas tenham qualidade similar à telefo-

nia convencional. Entre 1% e 3% a qualidade da voz será similar à qualidade da telefonia celular. Mais de 3% de perdas pode até ser aceitável para voz, mas poderá interferir na sinalização da chamadas.

3.3 Fatores que afetam QoS

Existem fatores que afetam o QoS de serviços VoIP. Estes fatores são oriundos das características e dos equipamentos utilizados na estrutura da rede onde os serviços de VoIP estão sendo utilizados. Os fatores mais importantes que afetam o QoS em serviços VoIP, além do atraso, *jitter* e perdas de pacotes, são:

3.3.1 Eco

O eco torna-se um incômodo à medida que o atraso aumenta. Ecos com pouco atraso são mascarados pela fala e são até mesmo agradáveis de se ouvir durante a conversação, dando a impressão de que o ouvinte está escutando bem o que está sendo dito e que a linha não está muda. Mas quando o atraso aumenta, de tal forma que ao terminar a fala o usuário percebe a própria voz, o eco passa a degradar a qualidade da ligação. Por exemplo, se o tempo de *round-trip delay* (atraso de ida e volta) da rede é de aproximadamente 250ms, o eco se torna bastante irritante mesmo com um nível bastante baixo (RACHID, 2004).

Os dois tipos principais de eco são o acústico e de impedância, embora existam outras fontes de eco (RACHID, 2004). Será sentido o eco quando uma chamada VoIP sair de uma LAN através de um tronco analógico com problemas e ir para a rede de telefonia pública. Outro fator causador de eco é o não casamento de impedâncias entre sistema telefônicos de 2 e 4 fios. Circuitos de longa distância necessitam de um ganho elevado no sinal, por este motivo, são utilizados circuitos de quatro fios que permitem o uso deste ganho sem aparecer ruídos na linha. Em casos de linhas de assinantes, são utilizados 2 fios devido às distâncias serem menores e para a diminuição do custo (RACHID, 2004). Além disto, o eco pode ser sentido quando o não casamento de impedâncias ocorre na conversão entre o barramento TDM (*Time Division Multiplexing*) e a LAN, ou entre o *headset* e seu adaptador (AVAYA, 2002).

O não casamento de impedâncias causa ineficiência na transferência de energia. O desbanlaceamento de energia deve ir para algum lugar e acaba sendo refletido de volta em forma de eco. Usualmente, o participante da conversa que está falando escuta o eco, mas o participante receptor não. Uma fonte bastante comum de eco é com o uso de microfones e caixas de som para ligações telefônicas. O som recebido nas caixas de som é captado pelo microfone, retornado o sinal da voz para o emissor. Por isto, é recomendável a utilização de *headsets*.

Para amenizar o problema do eco são utilizados canceladores de eco. Os canceladores de eco comparam a voz recebida com os padrões recentes da voz. Se o padrão combinar, o cancelador atenua o eco. Canceladores de eco não são perfeitos. Sob algumas circunstâncias, o eco consegue passar o cancelador. O problema é potencializado em sistemas de VoIP. Se o atraso de um sentido do caminho *one-way* entre dois pontos for maior do que a memória do cancelador, o cancelador de eco não irá nem achar o padrão para o cancelamento.

Na dissertação de mestrado de (RACHID, 2004) são discutidas várias formas de cancelar eco em redes com serviços VoIP.

3.3.2 Tráfego Concorrente

O tráfego concorrente ao tráfego da voz pode aumentar o *jitter* e o atraso consideravelmente. A solução para garantir a qualidade dos pacotes de voz é priorizá-los em relação aos demais tipos de pacotes. Além disso, deve ser reservada largura de banda para uso exclusivo do tráfego de voz. Isto nem sempre é possível devido ao custo, principalmente em linhas dedicadas alugadas de provedores de serviços de telecomunicações. Neste caso, pode-se usar *codecs* de menor qualidade e compressão dos cabeçalhos RTP, diminuindo a necessidade de reserva de largura de banda para a voz.

3.3.3 Convergência de Rotas

O tempo de convergência de rotas pode ser tornar um problema para redes TCP/IP usando serviços de VoIP. Por exemplo, o protocolo de roteamento dinâmico BGP (*Border Gateway Protocol*) possui tempos de convergência de rotas de vários minutos, levando cerca de 2 minutos para restabelecer as rotas e 30 minutos para anunciá-las. No período de tempo até a convergência de rotas, os sistemas autônomos envolvidos ficarão inalcançáveis impossibilitando a comunicação de rede com os mesmos.

3.4 Garantias de QoS

Para se garantir a qualidade dos serviços de VoIP, alguns fatores devem ser levados em conta. Estes fatores, tais como os equipamentos e mecanismos de QoS, são determinantes para a garantia do nível de serviço e da disponibilidade do serviço VoIP.

3.4.1 Equipamentos

A qualidade da conversação telefônica sofre influência direta dos equipamentos utilizados. No caso de telefones convencionais, a qualidade do aparelho telefônico exerce grande influência na qualidade da conversação. Entretanto em telefonia convencional, os problemas mais comuns são relativamente simples, tais como problemas com o alto-falante ou com o microfone. Os equipamentos intermediários são geralmente bastante robustos e planejados em estruturas que garantem os níveis de QoS desejados. Em serviços VoIP, tanto os equipamentos utilizados nos terminais (*softphones*, telefones IP, *headsets*, microfones, placas de som) quanto no equipamentos intermediários (*gateways*, servidores, *switches*, roteadores) estão sujeitos a vários tipos de problemas. No caso de terminais VoIP, o atraso inserido no tratamento do sinal de voz pode comprometer a qualidade da conversação, principalmente em *softphones* (JIANG; KOGUCHI; SCHULZ-RINNE, 2003). A qualidade dos equipamentos intermediários pode afetar os fatores de QoS negativamente, principalmente em relação ao atraso. Isto ocorre devido ao tempo de conversão entre *codecs* e tratamento dos pacotes de áudio em roteadores, *switches* e *gateways*. Além disso, os softwares embarcados nos equipamentos estão sujeitos a *bugs* e a incompatibilidades. Devido ao aumento dos pontos de falhas, torna-se necessário o uso de equipamentos redundantes. Atualmente, existem diversos fabricantes de equipamentos VoIP, existindo modelos com os mais diferentes tipos de funcionalidades, preços e qualidades.

3.4.2 Mecanismos de QoS Nível 2 e 3

Na estrutura da Internet atual, garantir os níveis de QoS desejáveis para VoIP é praticamente impossível. Embora alguns experimentos mostrem ser possível obter boa qualidade

em conversações usando VoIP na Internet (BIYANI et al., 2003) e (JI; SCHULZRINNE, 2003), as garantias de qualidade são inexistentes. A disponibilidade do serviço deve ser semelhante à encontrada na telefonia convencional, e para atingir esta meta o uso da Internet sem QoS deve ser descartada. Entretanto, em redes corporativas é possível montar uma estrutura que possibilite as garantias de QoS necessárias para a disponibilidade dos serviços VoIP. Os mecanismos de garantia de qualidade se dividem em dois tipos, os mecanismos de QoS (*Quality of Service*) e os mecanismos de CoS (*Class of Services*).

Os mecanismos de QoS garantem os níveis necessários de serviço fim a fim, tais como largura de banda e atraso, para a garantia de qualidade de uma aplicação. Para uma aplicação de missão crítica, o QoS significa largura de banda garantida sem nenhuma perda de pacotes. Existe uma dificuldade maior para a implantação de QoS na rede devido ao maior controle necessário para a configuração. Por exemplo, no caso do protocolo IntServ (*Integrated Services*) cada dispositivo deve possuir uma entrada em tabela registrando cada fluxo. Em uma grande rede corporativa, os dispositivos podem ficar saturados com milhares de fluxos simultâneos. As redes ATM, *Frame Relay* e MPLS (*Multi-Protocol Label Switching*) são exemplos de redes que fornecem nível de serviço por fluxo de aplicação. O RSVP (*Resource Reservation Protocol*) é um protocolo que possui mecanismos de QoS. Este protocolo provê QoS através da reserva de largura de banda e de recursos ao longo de todo o caminho do fluxo. O RVSP é implementado em alguns roteadores, *switches* de nível 3 e no Windows 2000 Server ou superior (INTEL, 1999).

Os mecanismos de CoS (*Class of Services*) fornecem a priorização de tráfego em um simples enlace de dados. Enquanto os mecanismos de QoS tratam de redes mais complexas, o CoS trata de apenas um enlace de dados. Portanto, um *switch* ethernet pode prover priorização de pacotes através de CoS para um único *host*, enquanto um grupo de roteadores pode participar de um esquema mais elaborado de QoS. Os sistemas de CoS definem um comportamento por *hop*, isto é, por dispositivo que trata o pacote. Desta forma, não é possível garantir o nível de serviço em termos de capacidade ou velocidade fim a fim. Os sistemas CoS utilizam os recursos para entregar os pacotes marcados como prioritários de acordo com o comportamento configurado em cada dispositivo. As soluções com sistemas CoS são eficientes onde menos de 30% do tráfego é composto por pacotes de voz, o que ocorre na maioria das redes corporativas atualmente (WALLINGFORD, 2005). Dois esquemas, que suportam CoS, podem ser utilizados para garantir a prioridade de VoIP em redes corporativas. São eles:

- Nível 2 - IEEE 802.1Q/802.1p/ToS
- Nível 3 - DiffServ

3.4.2.1 Nível 2 - IEEE 802.1Q e IEEE 802.1p

O padrão IEEE 802.1Q, definido em 1998, modificou o cabeçalho do quadro ethernet para disponibilizar o uso de VLANs e prioridades em redes locais. As VLANs provêm a criação por software de domínios de difusão (*broadcast*) limitados, através da criação de subredes virtualmente separadas. Os *broadcasts*, em se tratando de redes padrão ethernet, são normais. Este mecanismo é utilizado por protocolos para a realização de várias funções, tais como o anúncio de serviços. Ao se criar uma rede VLAN separada para o tráfego de voz, ocorrerá a diminuição da quantidade de tráfego de *broadcast* que um host receberá. O uso de VLANs separadas resulta em uma utilização mais efetiva da largura de banda e a redução do uso de processador no caso de telefones IP, pois não haverá a

necessidade de analisar tráfego *broadcast* irrelevante. Outra vantagem das VLANs é a segurança, pois o tráfego de nível 2 ficará contido dentro da VLAN.

Para usar efetivamente o conceito de VLAN, os *switches* e as placas de rede devem seguir o padrão IEEE 802.1Q. Normalmente, são os switches os responsáveis por modificar os quadros entrantes na portas, mas é cada vez mais comum as placas de rede reconhecerem o padrão IEEE 802.1Q (TANENBAUM, 2003).

Os *hosts* são identificados pelos endereços físicos e indexados a uma VLAN. Existe pelo menos uma VLAN, a VLAN 0, que é padrão na configuração de *switches*. Podem ser criadas VLANs de 0 a 4096 e os quadros ficam marcados com o número da VLAN à qual pertencem. Por exemplo, se um quadro marcado como VLAN 4 chegar na porta 3, então alguma máquina da VLAN 4 está na porta 3. Com isto é possível montar de forma transparente a tabela de endereços físicos indexada pela identificação da VLAN. Em redes com serviços VoIP, é recomendável que se tenha pelo menos duas VLANs, uma para o tráfego de voz e outra para o tráfego restante.

O padrão IEEE 802.1p foi definido juntamente com o 802.1Q e permitiu o uso de mecanismos de qualidade de serviço em redes ethernet. Este padrão utiliza 3 bits do cabeçalho ethernet para classificar cada quadro em um nível particular de precedência. Cada nível de precedência define o tratamento que o quadro vai obter em cada *switch*, pois a quantidade e os tipos de filas em cada *switch* podem variar. O campo ToS (*Type of Service*) do protocolo IP é utilizado para marcar o valor da precedência do pacote e é interpretado nos roteadores. Normalmente, existe um mapeamento do 802.1p para os campos ToS. Muitos fabricantes chamam esta solução de *ToS IP Precedence*. Já que são alocados 3 bits para classificação dos quadros, existem 8 tipos de classes de serviços e são apresentadas na tabela 3.3. Quanto maior o valor, maior a prioridade. O valor 5 é padronizado para o tráfego de voz. Podem ser utilizadas até 8 filas nos *switches*, mas o mais comum é o uso de 2 filas por porta, sendo uma para tráfego com prioridade, número de classe diferente de 0, e outra para o restante do tráfego.

Tabela 3.3: Classes de serviços do IEEE 802.1p.

Número da Classe	Tipo de Tráfego
0	Melhor esforço (padrão)
1	Mais baixa prioridade
2	Indefinido
3	Média prioridade
4	Carga Controlada
5	Crítico (padrão para voz)
6	Roteamento
7	Gerenciamento (mais alta prioridade)

As prioridades das classes de serviços podem ser modificadas pelos fabricantes. Por exemplo, nos *switches* da HP modelos *ProCurve Switch 5300xl*, *ProCurve Switch 3400cl* e *ProCurve Switch 2650* existem quatro filas de prioridade por porta e a definição das classes de serviço está representada na tabela 3.4 (HP, 2005).

O padrão 802.1p é um recurso cada vez mais comum em *switches*, até mesmo em *switches* de baixo custo tais como os modelos *Baseline* da empresa 3Com (3COM, 2005).

Tabela 3.4: Classes de serviços dos *switches* HP.

Números da Classe	Prioridade
1 e 2	Baixa
0 e 3	Normal
4 e 5	Média
6 e 7	Alta

3.4.2.2 Nível 3 - DiffServ

O esquema de priorização com DiffServ (*Differentiated Services*), (NICHOLS et al., 1998; BLAKE et al., 1998), redefine o *byte* do campo TOS (*Type of Service*) do cabeçalho do protocolo IP em uma maneira mais elaborada do que no 802.1p. Enquanto o 802.1p é usado em *switches* ethernet, o DiffServ é usado em enlaces ponto a ponto entre roteadores. Quando um pacote é recebido no roteador, de um nó da rede ou de outro roteador, o DiffServ marca o campo ToS do cabeçalho do pacote baseado na política associada à natureza do pacote em questão. Uma vez marcado o pacote, os demais roteadores do caminho devem obedecer à marcação. Sendo os pacotes marcados nos roteadores mais próximos dos terminais, os demais roteadores somente repassam os pacotes, não necessitando refazer a marcação. O DiffServ não garante reserva de recursos, mas sim a priorização dos pacotes baseado em políticas. Normalmente, as marcações definidas no 802.1p e no DiffServ podem ser relacionadas através de tabelas. Na tabela 3.5, um exemplo do mapeamento entre marcações 802.1p e DiffServ, em destaque os valores atribuídos para o tráfego de voz.

Tabela 3.5: Mapeamento entre valores de CoS (802.1p) e DSCP (DiffServ).

CoS (802.1p)	0	1	2	3	4	5	6	7
DSCP	0	8	16	24	32	40	48	56

Os DSCP (*DiffServ Code Points*) são 6 bits do campos ToS do cabeçalho do protocolo IP que definem os níveis de serviço. Embora a maioria das implementações suportem apenas 3 bits, substituindo os 3 bits do campo ToS original. Os outros 3 bits são reservados para extensões do padrão DiffServ (WALLINGFORD, 2005). O comportamento dos valores de DSCP são divididos em três grupos, chamados classes PHB (*Per-Hop Behavior*), classes de tráfego ou classes DSCP e são elas:

- AF (*Assured Forwarding*) - classe DSCP para rápida vazão dos pacotes, usada principalmente para a sinalização das chamadas VoIP.
- EF (*Expedited Forwarding*) - a classe DSCP voltada para a vazão mais rápida dos pacotes, usada para a marcação dos pacotes que contêm voz.
- BE (*Best Effort*) - a classe DSCP utilizada para tráfego de melhor esforço.

Quando os pacotes são analisados pelos roteadores com DiffServ ativo, o roteador decide que tipo de marcação será usada baseado em qual classe de DSCP o pacote estará inserido. Pacotes com alta prioridade receberão uma marcação AF ou EF, enquanto os pacotes com mais baixa prioridade receberão uma marcação BE ou nenhuma. As classes DSCP são compatíveis com as classes ToS, portanto o DiffServ pode fazer uso da priorização com 802.1p em redes locais distintas. Um exemplo de uso do DiffServ é com roteadores

utilizado o sistema operacional GNU/Linux com kernel 2.4 ou mais atual. Através do uso da ferramenta *iptables* é possível configurar a marcação de pacotes com classes DSCP, como no exemplo:

```
iptables -A PREROUTING -p udp -D 0.0.0.0/0.0.0.0 --dport 8000 \
-j DSCP --set-dscp-class EF
```

Com a linha de comando acima, os pacotes do protocolo UDP com destino a porta 8000 serão marcado no campo DSCP como sendo do tipo EF. A marcação ocorrerá antes do roteamento (PREROUTING). A porta 8000 UDP deverá ser configurada nos terminais para o envio dos fluxos de áudio RTP.

Na figura 3.4, estão representados os mecanismos de QoS e CoS em uma rede corporativa hipotética.

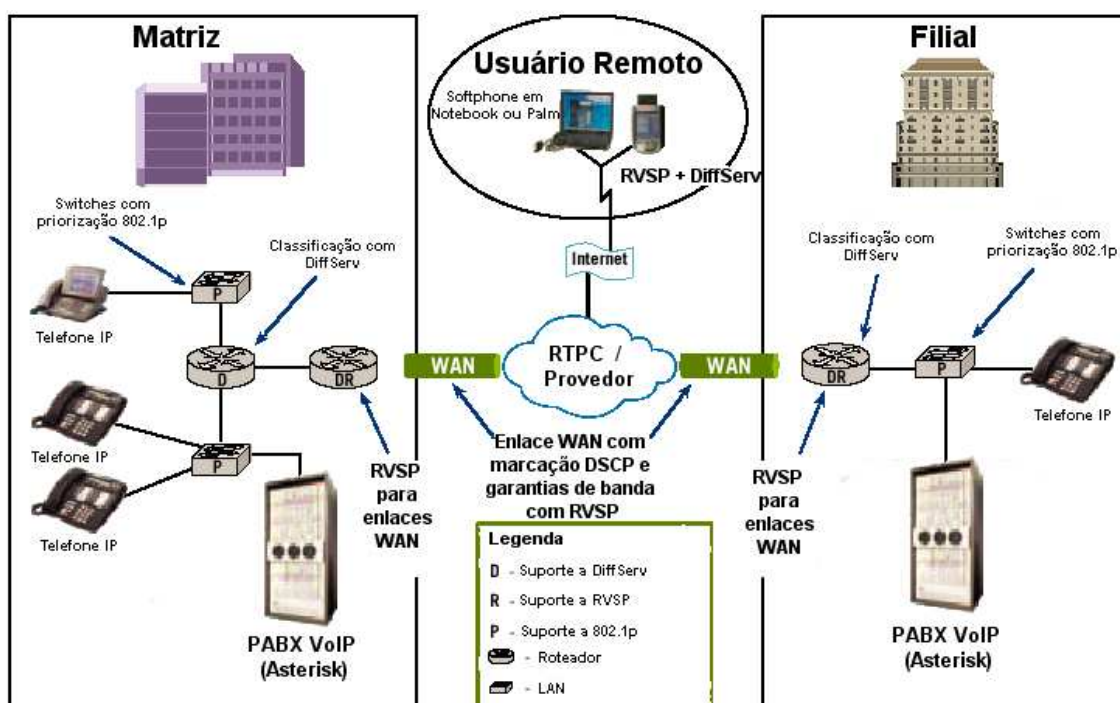


Figura 3.4: Esquema de uso dos mecanismos de QoS e Cos.

3.4.2.3 Melhores Práticas

Algumas práticas para garantir a qualidade de serviço na rede para serviços VoIP, são listadas a seguir (WALLINGFORD, 2005):

- Se o tráfego de rede para pacotes de voz for menor do que 30%, é eficiente o uso das técnicas de DiffServ e IEEE 802.1p;
- Em configurações com DiffServ, utilizar a classificação de fluxos RTP de áudio como EF (*Expedited Forwarding*);
- Em configurações com 802.1p, utilizar a classificação do tráfego de voz como marcação 5 (em alguns casos o valor é 160, pois são os 3 bits mais significantes de 8 bits. Exemplo 10100000 em binário corresponde a 160 em decimal).

- Utilizar, se possível, uma VLAN exclusiva para o tráfego de voz;
- Utilizar filas de baixa latência (LLQ - *Low Latency Queuing*) nos roteadores para os fluxos de áudio RTP;
- Quanto maior a largura de banda, menores as chances de ocorrerem problemas com atrasos, *jitter* e perdas no pacotes de voz. Embora não resolva o problema da garantia do serviço, o aumento da largura de banda diminui bastante os problemas atrelados ao VoIP;
- Utilizar o *codec* G.711 sempre que possível. É um dos *codecs* com melhor resistência a perdas e com o menor atraso de codificação;
- Para minimizar o atraso, pode-se diminuir o tamanho por pacote. Em redes locais, onde o atraso de propagação é baixo, o aumento do número do tamanho do pacote diminui o *overhead* e conseqüentemente a largura de banda. Em enlaces onde o atraso é maior, a diminuição de amostras por pacote torna o atraso fim a fim menor.
- Evitar ao máximo a transcodificação. Os atrasos e a perda de qualidade na conversão de *codecs* não são desejáveis.
- Manter a porcentagem de perdas de pacotes nos segmentos de rede abaixo de 1%;
- Quando existe limite de largura de banda, é recomendada a utilização de técnicas de QoS tais como RVSP e MPLS.

3.4.3 Buffer de Jitter (Equalizador de *Jitter*)

O *buffer* de *jitter* ou equalizador de *jitter* transforma uma atraso variável em um atraso fixo. Sua função é armazenar as amostras de áudio por um determinado tempo, com a função de sincronizá-las. Este tempo de armazenamento é chamado de atraso inicial de reprodução (*initial play out delay*). É essencial a duração do tempo de armazenamento das amostras no *buffer*. Se as amostras forem armazenadas por muito pouco tempo, variações no atraso podem causar falhas na conversação. Se as amostras forem armazenadas por muito tempo, o *buffer* pode estourar, e os pacotes descartados também podem causar falhas na conversação. Além disto, um tempo de armazenamento extenso pode elevar muito o atraso total da conversação. O valor ótimo para o tempo inicial de reprodução é igual ao tamanho do atraso variável ao longo da conexão (CISCO, 2005). O valor do *buffer* pode ser adaptativo, mas o valor máximo de atraso é fixado. Quando *buffers* adaptativos são utilizados, o atraso torna-se variável. Entretanto, o valor máximo de atraso pode ser usado com o pior caso para cálculos de projeto. Na figura 3.5, uma representação do mecanismo de equalização de *buffer*.

3.4.4 Redundância

Tradicionalmente os PABXs são sistemas extremamente confiáveis e muitos fabricantes garantem 99,999% por ano de disponibilidade. A arquitetura tradicional dos PABXs chega a este nível de disponibilidade devido a componentes confiáveis e à redundância prevista no projeto (BRANDL et al., 2004).

Um grande problema do serviço de VoIP é prover o mesmo nível de disponibilidade da telefonia convencional. Uma das formas de garantir a disponibilidade é fazer com que os *gateways* e os terminais possam se registrar em múltiplos servidores. Se um dos

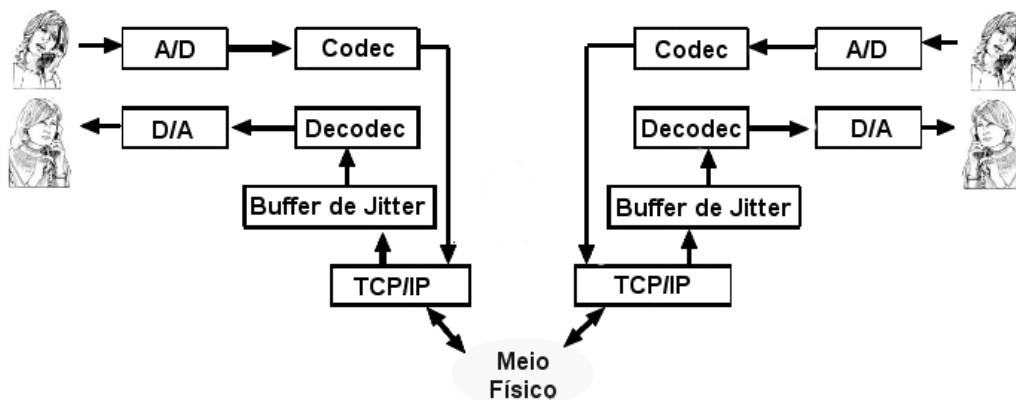


Figura 3.5: Posicionamento do *buffer de jitter*.

servidores ficar inoperante, o registro passa a ser feito por um servidor sobressalente. Dependendo de como isto é feito e da inteligência atrelada ao telefone e ao *gateway*, o novo registro pode ou não afetar a conexão ativa (o caminho do fluxo de voz). Se o servidor principal e o de *backup* compartilharem informações de controle de chamadas, então, após o novo registro, os usuários podem continuar as conversações. A continuidade da ligações pode ser crítica em *call centers* onde é importante não desconectar usuários que estão na fila para serem atendidos. Redes TCP/IP podem ser projetadas para serem altamente redundantes com múltiplos caminhos entre os dispositivos. Embora, em muitos casos, exista apenas um enlace entre o servidor de controle e o *gateway*. Se o enlace falhar, toda a área servida pelo *gateway* ficará sem o serviço. O problema pode ser resolvido utilizando enlaces alternativos e inteligência no controle dos servidores de controle e dos *gateways* (COFFMAN, 2004).

3.4.5 Fragmentação de pacotes

A fragmentação de pacotes em enlaces de baixa velocidade, abaixo de 1Mbits, pode ser usada quando o atraso de envio dos pacotes for significativo. Por exemplo, um pacote de 768 bytes leva 48ms para ser transferido em um enlace de 128Kbits. Se existir um pacote de voz esperando na fila para ser transmitido, será adicionado um atraso de 48ms na chamada. Sendo que os tamanhos dos pacotes podem variar, o *jitter* será aumentado. Para diminuir o problema, podem-se usar técnicas de fragmentação e intercalamento de pacotes entre os roteadores dos enlaces de baixa velocidade. Fragmentando os pacotes, o tempo de transferência é menor, conseqüentemente, os tempos de espera de transferência serão menores, e poderá haver maior vazão dos pacotes prioritários de voz.

3.4.6 Compressão de cabeçalhos

A largura de banda necessária para transmissão de voz pode ser reduzida se o *overhead* de cabeçalho for reduzido. Em (CASNER; JACOBSON, 1999) é descrito o protocolo cRTP (*Compressed RTP*) para compressão dos cabeçalhos dos protocolos RTP/UDP/IP. Normalmente, a soma dos cabeçalhos RTP/UDP/IP possuem 40 bytes (12 bytes RTP, 8 bytes UDP e 20 bytes IP). Utilizando a compressão, o cabeçalho poderá ser reduzido para 2 ou 4 bytes (figura 3.6). A compressão para 2 bytes não usa o campo *checksum* do protocolo UDP, enquanto a compressão com 4 bytes utiliza, diminuindo significativamente o *overhead*.

Embora seja reduzida a largura de banda consumida, é adicionado um atraso pelo uso

Compressão do cabeçalho RTP

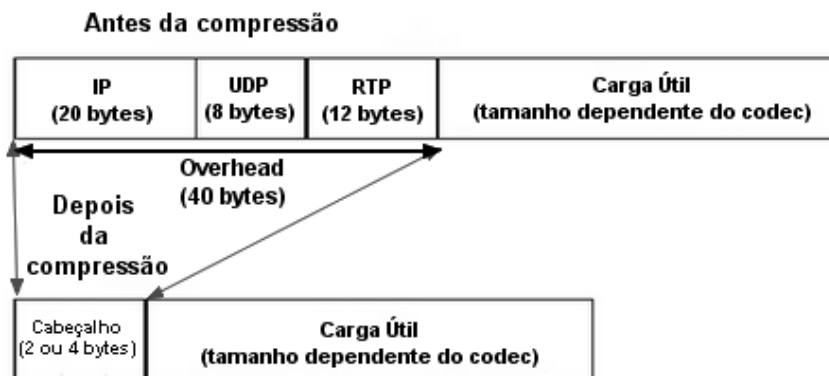


Figura 3.6: Cabeçalho com compressão do protocolo RTP.

da compressão dos cabeçalhos pelos roteadores. Portanto, deve haver um balanceamento destes fatores. Além disso, as compressões usadas com o RTP sofrem de problemas de incompatibilidade entre as implementações dos fabricantes e só podem ser usadas em enlaces ponto a ponto.

3.4.7 Alimentação de Energia

A alimentação de energia é uma fator importante para aumentar a disponibilidade dos serviços de VoIP. Na telefonia convencional, a alimentação dos terminais é fornecida pelas centrais telefônicas. Os equipamentos intermediários devem estar alocados em uma infra-estrutura que garanta o fornecimento de energia constante, através de *no-breaks* e geradores de energia. O problema maior é encontrado no fornecimento de energia para os terminais. Em caso de *softphones*, o computador deve estar conectado a um *no-break*, por exemplo. No caso de telefones IP, existe a possibilidade de usar o mecanismo de *Power over Ethernet* (PoE) (IEEE, 2003), no qual o fornecimento de energia é feito através do cabeamento de rede.

3.4.8 Planejamento de Capacidade da Rede

O planejamento de capacidade de rede é de suma importância para garantir a qualidade do serviço de VoIP, pois dimensiona o volume de serviços a quantidade de recursos da rede.

Será visto em maiores detalhes no capítulo 4.

3.5 Formas de Medição da Qualidade da Voz

A medição da qualidade da voz em sistemas de comunicação tende a ser subjetiva e de difícil mensuração. Ao longos dos anos, foram desenvolvidas várias técnicas com o objetivo de analisar a qualidade percebida pelo usuário. As técnicas são de dois tipos: objetivas e subjetivas.

O critério principal para definir a qualidade do áudio em conversações telefônicas é a percepção do usuário, portanto, um critério subjetivo. A qualidade pode ser medida através de métodos subjetivos. Embora os métodos subjetivos sejam os métodos mais confiáveis, eles também são de alto custo e demorados. Desta forma, métodos que possam

utilizar parâmetros físicos são desejáveis. Os métodos que podem analisar a qualidade do serviço de voz através de parâmetros de rede são chamados métodos objetivos.

Uma das primeiras abordagens para avaliação da qualidade da voz foi a utilização da Pontuação de Opinião Média, ou MOS (*Mean Opinion Score*), método subjetivo definido nas recomendações ITU-T P.800 e ITU-T P.830, pelo qual um conjunto de avaliadores ouvintes atribuem uma pontuação de 1 (pobre) a 5 (excelente) à qualidade da fala reproduzida pelo sistema de comunicação em teste. Como expressa diretamente a opinião média dos usuários, o MOS representa um índice de referência para avaliação da qualidade da fala em sistemas de comunicação. Dentre os métodos objetivos, destaca-se o Modelo E, originalmente proposto pelo ETSI (European Telecommunications Standards Institute), e posteriormente padronizado pelo ITU-T (International Telecommunications Union-Telecommunications Standard Sector), através da Recomendação G.107. Outros exemplos de métodos objetivos são o PSQM (*Perceptual Speech Quality Measure*), o PAMS (*Perceptual Analysis Measurement System*) e o PESQ (*Perceptual Evaluation of Speech Quality*). Estes métodos fazem uso do conhecimento do sistema auditivo humano para comparar um sinal de referência (trecho de voz previamente gravado) com um sinal degradado (sinal de referência submetido ao sistema de transmissão a ser avaliado), a fim de compor uma medida de distorção do sinal de voz (LUSTOSA, 2005).

Os métodos MOS e Modelo E são descritos a seguir.

3.5.1 MOS - (*Mean Opinion Score*)

O método subjetivo mais usado é o fator MOS (*Mean Opinion Score*) definido pelo ITU-T na recomendação P.800. O fator MOS é um valor entre 1 e 5 que define a qualidade da conversação. O valor 5 é a melhor qualidade, enquanto o valor 1 é a pior. Para determinar este número, uma série de orações faladas são transmitidas aos ouvintes por um sistema de compressão de voz (*codec*). Para cada oração, os ouvintes devem alocar um número de acordo com uma escala de valores, onde a média aritmética dos resultados de todos os ouvintes define a pontuação MOS para o sistema de compressão de voz em questão. Abaixo a escala de valores utilizada nos testes para obtenção do fator MOS:

- 5 - Excelente - o usuário não percebe degradações no sinal.
- 4 - Bom - o usuário percebe degradações mínimas no sinal.
- 3 - Razoável - o usuário percebe degradações no sinal, mas com esforço, consegue entender a mensagem.
- 2 - Ruim - o sinal possui interrupções devido as degradações e o usuário deve fazer esforço para entender trechos da mensagem.
- 1 - Péssimo - ininteligível, o usuário não entende nada da mensagem, mesmo com esforço.

O valor 4 do fator MOS define a qualidade da voz em telefonia fixa convencional, enquanto que valores entre 3,5 e 4 caracterizam a qualidade da voz em telefonia celular. Os valores dependem, principalmente, dos *codecs* utilizados. O *codec* com o melhor fator de MOS é o G.711 (tabela 2.1).

3.5.2 Modelo E

Os métodos subjetivos de medição de qualidade da voz são caros e demorados e não consideram os prejuízos ocasionados pela transmissão através de uma rede de pacotes. São necessários métodos para estimar a qualidade subjetiva da voz, medindo os parâmetros da rede e dos equipamentos envolvidos na comunicação. Todos os métodos que fazem estimativas da qualidade da voz através de parâmetros de rede e de equipamentos são chamados de métodos objetivos.

Os métodos objetivos podem ser categorizados em vários grupos do ponto de vista de objetivos, procedimentos de medição, entradas de dados e estimativa de MOS. Segundo (TAKAHASHI; YOSHINO; KITAWAKI, 2004), existem três grupos principais de métodos objetivos: os modelos de opinião (*opinion models*), os modelos de nível de voz (*speech-layer models*) e os modelos de nível de pacote (*packet-layer models*). Os modelos de opinião utilizam as características da rede e dos equipamentos para estimar o fator MOS. Os modelos de nível de voz utilizam sinais de voz como entrada e executam filtros para estimar o fator MOS. O método dos modelos de nível de pacote utilizam parâmetros dos pacotes IP para estimar o fator MOS.

O modelo E, definido na recomendação G.107 do ITU-T de 1998, e adotado pelo *European Telecommunications Standards Institute* (ETSI) e pelo *Telecommunications Industry Association* (TIA) como uma ferramenta de planejamento de rede é o modelo de opinião mais usado no mundo (TAKAHASHI; YOSHINO; KITAWAKI, 2004). O modelo E possui 20 parâmetros de entrada que representam o terminal, a rede e fatores de qualidade do ambiente. O resultado do cálculo dos 20 parâmetros é chamado de fator R. Inicialmente, os graus de degradação de qualidade devido a fatores tais como eco, atraso e distorção, são calculados e diminuídos do valor de referência. A degradação da qualidade introduzida pelos *codecs* e perdas de pacotes são tratadas como um fator de degradação de equipamento (*equipment impairment factor*).

A recomendação G.107 fornece um conjunto de valores padrão, os quais podem ser usados quando é assumido que os terminais e o uso do ambiente de rede são padrões (TAKAHASHI; YOSHINO; KITAWAKI, 2004).

O modelo E avalia o desempenho dos efeitos das variações de parâmetros de transmissão que afetam a qualidade de conversação em telefonia de banda estreita. O princípio do modelo E é baseado em pressupostos de que os prejuízos de uma transmissão podem ser transformados em elementos matemáticos. A principal saída do modelo E é o fator de transmissão R:

$$R = R_0 - I_s - I_d - I_e + A \quad (1)$$

Onde:

- R_0 representa a razão entre sinal-ruído;
- I_s representa os prejuízos simultâneos ocorridos com o sinal de voz;
- I_d representa os prejuízos causados por atrasos;
- I_e representa os prejuízos causados na compressão dos *codecs*;
- O fator de vantagem A pode ser usado para compensação quando existem outras vantagens de acesso para o usuário, por exemplo mobilidade.

O resultado final do cômputo dos fatores de perda é um fator escalar R , que varia de 0 (piores caso) a 100 (excelente). O fator R pode ser convertido para escala de pontuação MOS através da seguinte expressão de terceiro grau (ITU-T, 2003):

$$MOS = \begin{cases} 1 & R < 0 \\ 1 + 0.035R + R(R - 60)(100 - R) \cdot 7 \cdot 10^{-6} & 0 < R < 100 \\ 4.5 & R > 100. \end{cases} \quad (2)$$

Normalmente, o fator R é descrito em categorias de valores. Sistemas cujo a qualidade da fala seja avaliada em $R=60$ não são recomendáveis (LUSTOSA et al., 2004), sendo desejável obter valores acima de $R=70$.

As principais vantagens do Modelo E, em relação aos outros métodos de medição de qualidade de voz são a capacidade de medição em tempo real, já que não há necessidade de comparação entre os sinais de referência e degradado, e a contabilização em separado de cada um dos fatores responsáveis pela degradação da qualidade da voz, como perdas, atraso fim-a-fim, distorções inerentes a codificadores de alta compressão, entre outros. Estas funcionalidades permitem a avaliação da origem e do grau de influência de cada um destes fatores separadamente, tornando o diagnóstico de problemas de transmissão mais precisos. Em (COLE; ROSENBLUTH, 2001), foi proposta uma simplificação para a fórmula do modelo E usando valores padrão para o cálculo do fator R . Foram utilizados valores padrão para os parâmetros que não estão relacionados diretamente com o transporte através de pacotes, eliminando o fator de vantagem. Em (LUSTOSA et al., 2004) e (LUSTOSA, 2005) são discutidas extensões para o modelo E, tal como o efeito da perda de pacotes em rajadas.

4 PLANEJAMENTO DE CAPACIDADE EM REDES PARA SERVIÇOS VOIP

O planejamento de capacidade da rede é de suma importância para atingir os níveis de QoS necessários para as aplicações funcionarem de forma eficiente. Os recursos de rede são finitos e custosos, portanto deve-se fazer o planejamento de capacidade da rede para se obter os níveis de QoS necessários para as aplicações, com o menor custo possível. Em serviços VoIP, para se obter uma boa qualidade das ligações deve-se manter os fatores de QoS dentro dos valores esperados. Para alcançar este objetivo, deve-se otimizar ao máximo o uso dos recursos de rede, determinando a demanda dos serviços e a capacidade dos recursos existentes.

Este capítulo apresenta uma visão geral do planejamento de capacidade em redes, os métodos utilizados em redes telefônicas convencionais e uma metodologia para adaptação dos métodos tradicionais de planejamento de capacidade de rede para os serviços VoIP em uma rede corporativa.

4.1 Planejamento de Capacidade em Redes

Devido ao comprometimento com a qualidade dos serviços transportados e à limitação física dos recursos, como a capacidade de transmissão dos enlaces, as redes possuem um ponto flutuante de saturação dos recursos. Neste estado de saturação, a rede impede o ingresso de novos serviços ou começa a sofrer queda de desempenho devido a congestionamentos. O ponto de saturação é flutuante, pois depende de fatores externos à configuração dos recursos da rede. Os três fatores externos são a quantidade de tráfego, os requisitos de desempenho e a topologia de conexões. A combinação desses fatores especifica a carga de serviço que é submetida à rede.

A carga de serviços distinta aplicada sobre determinada rede resulta em diferentes quantidades possíveis de atendimento a serviços. Portanto, o projeto de uma rede deve considerar a carga de serviço fundamental para o dimensionamento dos recursos da rede. Procurando sempre identificar os fatores externos com o maior detalhamento possível para adequar a rede aos serviços previstos.

Outros fatores que influenciam na capacidade de atendimento da rede são os fatores internos referentes à configuração da rede. Parte desses são os diferentes recursos que podem ser reunidos na configuração física da rede, como capacidade dos enlaces, capacidade de comutação, quantidade de *buffers*, topologia de enlaces e outros. São intrínsecos aos equipamentos utilizados ou caracterizam uma solução de configuração adotada para a rede. A outra parte dos fatores internos são referentes à configuração lógica dos mecanismos adotados para o controle de tráfego. Novamente, combinando valores para estes

fatores, têm-se modificações na capacidade da rede. Enquanto os fatores externos consomem recursos, os fatores internos os disponibilizam. O levantamento desses fatores e a avaliação de equilíbrio entre eles é o início da garantia de sucesso do projeto de uma rede.

Superada a etapa de levantamento e qualificação dos tipos de tráfegos a serem suportados, resta avaliar o impacto desta carga sobre a estrutura de recursos da rede. Sendo os recursos limitados, a rede possui um ponto de saturação que depende de suas características e técnicas utilizadas para atender os serviços demandados. Partindo de uma determinada carga de serviço, o comportamento de desempenho da rede depende exclusivamente das características dos fatores internos físicos e de controle. O conjunto de características físicas envolve, entre outros elementos, topologia de rede, taxa de transmissão dos enlaces, capacidade de comutação e quantidade de memória. São recursos limitados geralmente pelo custo ou inviabilidade técnica, porém quanto maior for o número de recursos, melhor será o desempenho da rede. Os elementos de controle são representados por algoritmos que implementam os mecanismos de gerenciamento de tráfego, tais como admissão de conexões e tratamento de congestionamento. Os fatores de controle são fundamentais e as suas configurações determinam as características de utilização da rede. Aperfeiçoamentos dos elementos de controle otimizam a utilização dos recursos físicos e conseqüentemente aumentam a capacidade de serviço suportada pela rede.

Como qualquer outro sistema, as redes de comunicação têm medidas de desempenho para avaliar sua eficiência. As questões referentes a desempenho são importantes pois influenciam diretamente na quantidade e na qualidade dos serviços prestados. Tem-se como exemplo a largura de banda (vazão) que restringe a quantidade de bits que são transferidos por segundo em um meio físico. Se um usuário ou um grupo de usuários demandar um desempenho de largura de banda superior à disponível, ocorrerá uma contenção e enfileiramento para utilização do meio físico, impactando a vazão de informação pretendida pelo usuário. Os recursos de rede disponibilizados de forma compartilhada são alocados temporariamente pelos usuários. Após a utilização, o recurso é liberado, tornando-o disponível para outro usuário. A quantidade máxima de usuários atendidos simultaneamente é dada pela quantidade de recursos da rede. Ultrapassando este limite haverá um congestionamento e o usuário deverá aguardar a liberação do recurso para estabelecer a comunicação, ou seja, haverá perda de qualidade de serviço (disponibilidade). Como existe a alternância de usuários utilizando a rede, o compartilhamento disponibiliza o serviço de comunicação para uma quantidade maior de usuários. Desta forma, aproveita-se melhor a estrutura da rede reduzindo o custo por usuário.

Os recursos utilizados em uma rede de comunicação são limitados e, por questões de viabilidade, compartilhados pelos usuários. Como citado anteriormente, recursos limitados significam uma capacidade de atendimento restrita de usuários. A capacidade está relacionada com a quantidade de recursos alocados por usuário. As exigências de qualidade do serviço determinam a quantidade necessária e a forma de alocação dos recursos. Portanto, qualidade de serviço e capacidade de atendimento estão inter-relacionados. O aumento da exigência de qualidade impacta na capacidade de oferecer serviço, e vice-versa. O ideal é buscar o equilíbrio entre garantir a qualidade dos serviços e maximizar a utilização dos recursos da rede. A tecnologia adotada na rede define a relação entre qualidade de serviço e capacidade de atendimento. Nas redes telefônicas, o desempenho é avaliado utilizando parâmetros como: tempo para obter sinal para discar, tempo para estabelecer a ligação, disponibilidade de troncos, qualidade da voz e confiabilidade da ligação. Estes parâmetros determinam a qualidade de serviço (QoS) prestado, denominado como grade de serviço (GoS), já abordado no capítulo 2.

Em resumo, uma rede provê a habilidade de transmitir informações entre usuários, com o objetivo de prover um serviço efetivo a custos razoáveis. Não é eficiente manter recursos permanentemente dedicados aos usuários. Existe a necessidade de compartilhar e prover aos usuários meios de acessar estes recursos quando requisitados. Uma rede é composta por elementos físicos (*switches*, equipamentos terminais, enlaces de transmissão), cada qual com sua capacidade finita de transmissão de informações. Se a capacidade não é adequada para suportar a carga de serviços demandada pelos usuários, o desempenho da rede em transmitir as informações será prejudicado. Existe uma relação de dependência entre carga, capacidade e desempenho de uma rede (figura 4.1). A carga é a quantidade de tráfego a ser carregada na rede. A capacidade da rede é dada pelos recursos que a constituem. E o desempenho representa a qualidade de serviço oferecida pela rede aos usuários.

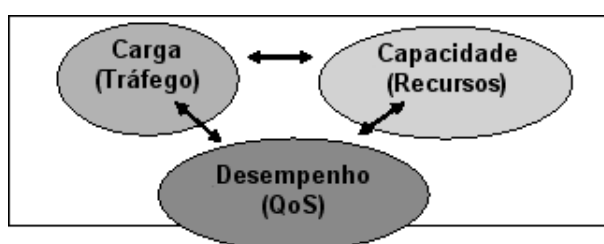


Figura 4.1: Relação entre capacidade, tráfego e desempenho.

Alterações em qualquer um destes três elementos influenciam o comportamento dos outros dois. Pode-se fixar um elemento em determinada quantidade para avaliar o comportamento dos outros dois elementos entre si. Por exemplo, pode-se fixar a capacidade (quantidade de recursos) de uma rede de comunicação. Aumentando continuamente a quantidade de carga (tráfego) sobre a rede, será atingido em determinado momento o esgotamento dos recursos ocasionando a redução no desempenho da rede percebido em questões de tempo de resposta, congestionamento e indisponibilidade do serviço. Os usuários percebem e avaliam os serviços de rede através dos níveis de serviço. Associado ao nível de serviço está o custo a ser pago pelo usuário. Portanto, de comum acordo com os usuários, a instalação deve definir quais são os níveis de serviço desejáveis e quais são intoleráveis. Exemplos de outros procedimentos que utilizam este relacionamento são:

- Dimensionamento de recursos: para especificar a capacidade requerida da rede é necessário avaliar a carga demandada e os níveis de desempenho esperados;
- Avaliação de desempenho: para determinar o desempenho da rede é medido o comportamento da carga sobre determinada capacidade de rede. Analisar o desempenho de um particular projeto de rede tem como ênfase variar o tráfego e medir o desempenho para dada capacidade de rede;
- Admissão de chamadas: para determinar a carga suportada pela rede é avaliado se a quantidade de recursos disponíveis permite alcançar o desempenho desejado. No caso de VoIP, o acréscimo de mais uma ligação na rede pode comprometer a qualidade de todas as outras ligações.

Os usuários, em geral, percebem os serviços de informática através do tempo de resposta, da disponibilidade do sistema e da facilidade de uso. A última está associada ao

software utilizado, enquanto as outras duas estão relacionadas ao dimensionamento de recursos, isto é, ao planejamento de capacidade da rede.

O planejamento de capacidade é baseado no relacionamento entre estes três elementos básicos: carga (tráfego), capacidade (recursos) e desempenho (QoS).

4.1.1 Fases do Planejamento de Capacidade

Um sistema de computação atende às solicitações dos usuários com um certo desempenho, que pode ser definido quantitativamente através de várias medidas, como por exemplo, tempo de resposta, taxa de processamento, índice de disponibilidade, e outros. O desempenho de um sistema de computação resulta da interação da carga de trabalho com os recursos que compõem o sistema. A capacidade de um sistema de computação é definida como sendo a carga de trabalho que o sistema pode processar, sem ultrapassar os limites de desempenho estabelecidos pelos níveis de serviço da instalação (GONCALVES, 2000). O planejamento de capacidade é um processo para determinar, no tempo preciso, a quantidade adequada de recursos para atender a carga de trabalho dentro de níveis de serviços propostos.

O processo de planejamento de capacidade para configuração de uma rede pode ser decomposto em quatro fases (figura 4.2):

- Caracterização da carga de trabalho;
- Definição dos níveis de serviço (QoS);
- Previsão de desempenho sobre os recursos disponíveis;
- Se necessário, adequação dos recursos.

O processo de dimensionar os recursos de uma rede consiste em utilizar técnicas que permitam prever o desempenho desse sistema frente a novas situações de carga e serviços. Uma vez de posse das estimativas de cargas e dos níveis de serviço desejáveis, deve-se estimar quando as demandas por recursos excederão a capacidade instalada da rede. O ponto de exaustão do sistema ocorrerá quando, em função das demandas previstas, o nível de serviço for considerado insatisfatório. A fase de previsão de desempenho deve relacionar o impacto da carga prevista sobre os níveis de serviços e a utilização corrente dos recursos. A questão final do planejamento de capacidade é estimar o ponto em que a demanda de recursos excederá a capacidade instalada de recursos. O processo requer uma ferramenta para a previsão de desempenho. A avaliação de um sistema pode ser a combinação de diversas medidas de desempenho alcançadas durante a realização de suas atividades. Métodos para a avaliação de desempenho estão divididos em duas categorias: técnicas de medição e técnicas de previsão.

As técnicas de medição são utilizadas quando é possível fazer medições sobre a rede nas condições de tráfego que se deseja. A técnica de medição requer que redes reais sejam avaliadas por experimentação. A vantagem da medição direta da rede é que nenhum detalhe da rede é excluído: a real utilização da rede está sendo monitorada e medida. Existem inconvenientes no caso de redes em produção, onde não é possível interromper os serviços para fazer experimentos de carga. Portanto, a rede não pode ser exercitada no seu limite de carga sem impactar em queda de desempenho para os usuários que a estão utilizando. Neste caso, o tipo de tráfego usual disponível restringe as condições de experimento da rede. Em situações de projeto de rede, onde se deseja avaliar o desempenho

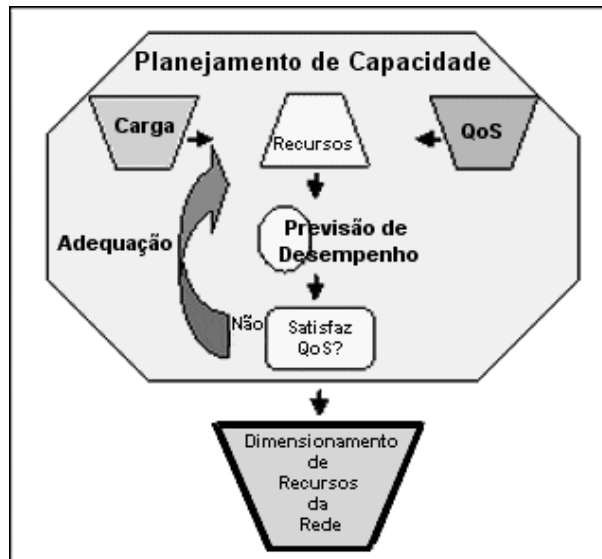


Figura 4.2: Fases do Planejamento de Capacidade de Rede.

da futura rede ou em situações futuras de tráfego, as técnicas de medição real não são possíveis. Isto acontece porque a rede não está implementada ou porque a carga ainda não existe para ser avaliada. Utilizam-se então técnicas de previsão que podem ser por modelo analíticos ou simulados. Comparando modelo analítico e simulação, os principais fatores a considerar são a precisão dos resultados, o tempo para obtê-los e o custo de utilizar o método. A vantagem das soluções analíticas é que elas podem ser usadas de forma razoavelmente rápida. Entretanto, a necessidade para resolver o modelo restringe a abrangência do sistema ou as características de tráfego que podem ser incluídas. Portanto a solução analítica é utilizada com frequência para produzir uma aproximação de um sistema, com resultados sendo produzidos de forma relativamente rápida e barata.

As redes de maior complexidade podem ser investigadas usando simulação. Sistemas podem ser modelados no nível de precisão desejado permitindo analisar as alternativas de configuração, levando-se em conta os aspectos de custos, as considerações de confiabilidade, a carga projetada e os níveis de serviço. Muitas vezes a simulação é o método mais adequado devido à natureza do problema e porque técnicas analíticas tornam-se muito difíceis de serem tratadas. Entretanto simulação pode ter um alto custo para desenvolver e executar. Simulações de eventos raros, tais como perda de pacotes em redes locais, podem consumir um tempo considerável de execução do simulador (GONCALVES, 2000).

4.2 Métodos de Planejamento de Capacidade em Telefonia Convencional

4.2.1 Breve Histórico

Os métodos modernos para otimização de redes telefônicas têm as suas raízes no trabalho feito por Agner Krarup Erlang (1878-1929), um cientista dinamarquês que ingressou na companhia telefônica *Copenhagen Telephone Company* em 1908 (figura 4.3). Ele iniciou a solução do problema chave no projeto de uma rede telefônica: quantos troncos são necessários para transportar um determinado volume de ligações?

Por exemplo, em um pequeno vilarejo onde cada casa tem um telefone conectado a uma central local. Quantos troncos a companhia telefônica deveria instalar entre a central



Figura 4.3: A.K. Erlang (MILLENNIUM MATHEMATICS PROJECT, 1997).

local e a central do próximo vilarejo? Erlang observou que não existia uma resposta única para o problema. Em vez disto, existia sempre um balanceamento entre serviço e custo. No caso do vilarejo, existiam duas opções extremas, sendo que nenhuma delas era aceitável:

- Disponibilizar somente um tronco, e deixar os usuários esperando até o tronco ficar disponível. O custo seria baixo, mas disponibilidade do serviço seria inaceitável.
- Disponibilizar um tronco para cada linha de usuário local, tornando a disponibilidade do serviço extremamente alta. O serviço seria excelente, mas o custo tornaria esta solução improvável.

O problema consistia em converter em números o balanceamento entre custo e serviço de forma a permitir que os projetistas de rede pudessem avaliar as melhores possibilidades. Para isto, Erlang conduziu os primeiros estudos sobre tráfego telefônico, e desenvolveu métodos matemáticos para avaliar o balanceamento de serviço e custo. Sua maior descoberta foi provar que a distribuição das chamadas seguia o modelo de distribuição de Poisson (BROCKMEYER; HALSTRØM; JENSEN, 1948). O trabalho teve repercussão mundial, ao ponto de um cientista dos laboratórios Bell Labs aprender o idioma dinamarquês para ler os artigos de Erlang no formato original (MILLENNIUM MATHEMATICS PROJECT, 1997). Em 1946, o CCITT (*International Consultative Committee on Telephones and Telegraphs*) adotou o nome de *erlang* para a unidade básica de tráfego telefônico em homenagem ao cientista.

4.2.2 Métodos de Erlang

4.2.2.1 Conceitos Básicos

Até recentemente, para se utilizar as técnicas de teletráfego era necessário ter um bom embasamento matemático ou aprender a usar livros de tabelas com valores previamente calculados. Com a disseminação dos computadores pessoais e à Internet, surgiram programas específicos para os cálculos e disponíveis para utilização, em páginas como (ERLANG.COM, 2005), (ICT, 2005) e (ERLANG.CO.UK, 2005), facilitando as aplicações

das fórmulas. Os parâmetros que serão aplicados às fórmulas definirão a eficiência do planejamento de capacidade da rede. Para poder aplicar os métodos de Erlang é necessário que alguns conceitos básicos sejam observados, tais como:

- *Erlang*: é a unidade básica de intensidade de teletráfego. A intensidade de tráfego é definida como o número de recursos ocupados em um conjunto, em um dado instante. O conjunto de recursos pode ser um grupo de servidores ou troncos.

Um *erlang* é uma unidade sem dimensão, representando o uso contínuo de um circuito (tronco). Entretanto, já que um único circuito usado continuamente transporta 60 minutos de chamadas em uma hora, um *erlang* é usualmente definido como sendo 60 minutos de tráfego. Por exemplo, se forem recebidas 300 ligações de 2 minutos de duração durante uma hora, foram recebidos 600 minutos, ou 10 *erlangs* de tráfego naquela hora. Nos EUA e Canadá a medição de tráfego é feita em CCS (*Centil Call Seconds*) ou 100 segundos de conversação telefônica. Para se converter os valores em CCS para *erlangs* basta dividir por 36.

- Tipos de tráfego:

Tráfego Efetivamente Utilizado (*Carried Traffic*): corresponde ao tráfego servido aos usuários do serviço. O Tráfego Oferecido (*Offered Traffic*) corresponde à real demanda de tráfego do sistema. Quanto maior for a porcentagem de bloqueio (GoS), maior será a diferença entre os dois tipos de tráfego. A diferença entre os dois tipos de tráfego está demonstrada na fórmula:

$$T_{eu} = \frac{T_o}{1 - GoS} \quad (3)$$

Onde:

- T_{eu} = tráfego efetivamente utilizado
- T_o = tráfego oferecido
- GoS = fator de bloqueio

Tráfego Oferecido - em modelos teóricos, o conceito de tráfego oferecido é usado. Este é o tráfego que seria transportado se nenhuma das ligações fosse rejeitada por falta de capacidade da rede (a capacidade da rede seria infinita). O tráfego oferecido é um valor hipotético e não pode ser medido (IVERSEN et al., 2005). Ele pode ser estimado através da fórmula 4, é equivalente à intensidade de tráfego e medido em *erlangs*.

$$T_o = \lambda \cdot s \quad (4)$$

Onde:

- λ : corresponde a média do número de chamadas
- s : corresponde ao tempo médio das chamadas

Exemplo:

Se um grupo de usuários faz 30 chamadas em uma hora, e cada chamada tem o

tempo de duração médio de 5 minutos, então a intensidade de tráfego é calculada da seguinte forma:

Minutos de tráfego em uma hora = número de chamadas x duração

Minutos de tráfego em uma hora = 30×5

Minutos de tráfego em uma hora = 150

Horas de tráfego em uma hora = $150 / 60$

Horas de tráfego em uma hora = 2.5

Intensidade de tráfego = 2.5 erlangs

A intensidade de tráfego é somente uma medida da utilização média durante um intervalo de tempo e não reflete o relacionamento entre a quantidade e a duração das chamadas. Portanto, muitas chamadas curtas podem produzir a mesma intensidade de tráfego do que poucas chamadas longas.

- Hora de Maior Movimento (*Busy Hour*): o tráfego medido na hora de maior movimento (HMM) representa a máxima carga de tráfego que a rede produz (CISCO, 2001). Quando não for possível fazer a medição, podem ser feitas estimativas de quantas chamadas são realizadas por dia. Em um ambiente padrão comercial, a hora de maior movimento do dia corresponde a entre 15 e 20% do total de tráfego diário. Geralmente, são utilizados nos cálculos o valor de 17% do total de tráfego diário para saber a HMM (CISCO, 2001). Para o tempo de chamada, quando não existe possibilidade de medição, podem ser utilizados valores entre 180 e 210 segundos como valor de referência (CISCO, 2001). Um telefone residencial, tipicamente, possui tráfego na HMM que está entre 0,05 e 0,1 erlangs. Sendo o tempo médio das ligações entre 3 e 4 minutos, então um telefone residencial realiza de uma a duas chamadas durante a HMM (CASTRO, 2002).
- Servidor: é algum dispositivo que manipula as chamadas (troncos, grupos de troncos, linhas). Por exemplo, no caso de um *call center* existem dois tipos de servidores, os troncos que transportam as chamadas e os agentes que atendem as chamadas. Com o uso de caixa postal de voz (*voice mail*) ou URA (Unidade de Resposta Audível) os servidores podem ser considerados como portas.
- GoS (*Grade of Service*) (Grade de Serviço): define a probabilidade de que todos os servidores estarão ocupados quando uma tentativa de chamada for feita. Por exemplo, em um grupo de troncos com GoS de P.02 significa que existe 2% de probabilidade de uma tentativa de chamada receber um sinal de ocupado (ser bloqueada). Em um *call center* este mesmo GoS significa dizer que existirá uma probabilidade de 2% de esperar para falar com um atendente.

4.2.3 Fórmulas de Teletráfego

Existem muitas fórmulas de teletráfego, apropriadas para as mais diferentes situações. Cada uma delas se adapta a uma modelo de tráfego específico. Os modelos de tráfego telefônico são definidos por (CISCO, 2001):

- Padrão de Chegada das Chamadas: a consideração mais fundamental da análise de tráfego clássica é que as solicitações de chamadas são independentes umas das outras. O fato de assumir a chegada das chamadas como aleatórias provê uma

formulação matemática que pode ser ajustada para produzir soluções aproximadas a problemas que seriam, de outra forma, matematicamente intratáveis (CASTRO, 2002). Os padrões de chegada de chamadas são:

Pouca variação: não existe grande variação na quantidade de chamadas no decorrer do tempo.

Picos: existem grandes variações de quantidade, sendo caracterizado por tráfego de exceção, como por exemplo em feriados.

Aleatório: conhecido como distribuição de Poisson, normalmente é observado em PABXs.

- Tipo de Bloqueio das Chamadas: uma chamada bloqueada é aquela que não é imediatamente servida. Os principais tipos são:

Lost Calls Held (LCH): as chamadas bloqueadas são perdidas, não retornando novamente.

Lost Calls Cleared (LCC): as chamadas bloqueadas são rejeitadas pela rede.

Lost Calls Delayed (LCD): as chamadas bloqueadas continuam na rede a espera da disponibilidade do sistema. Normalmente, são utilizadas filas de espera até as chamadas bloqueadas serem atendidas.

Lost Calls Retried (LCR): quando uma chamada é bloqueada, uma porcentagem dos usuários farão uma nova tentativa.

- Número de Origens das Chamadas: se o número de fontes de origem for grande e a sua atividade média é relativamente baixa, fontes ocupadas não reduzem de forma apreciável a taxa de chamadas. Por exemplo, considerando uma central local que serve a 10000 assinantes com 0,1 erlang de atividade cada um. Normalmente, haverá 1000 circuitos ativos e 9000 assinantes disponíveis para gerar novas chamadas. Se o número de usuários ativos aumentar por um fator de 50%, passando a 1500 circuitos ativos, o número de assinantes inativos é reduzido para 8500, uma mudança de apenas 5%. Portanto, pode-se considerar o número de origens como infinita, pois a taxa de chamadas é relativamente constante (CASTRO, 2002).
- Tempo de Duração das Chamadas: o tempo de duração das chamadas é exponencial. Geralmente, as chamadas possuem tempo de duração curtos, ao invés de tempos longos significando que a distribuição é exponencial negativa.

Na tabela 4.1 estão algumas características dos modelos de tráfego mais utilizados em telefonia (CISCO, 2001):

Tabela 4.1: Modelos de Tráfego Telefônico.

Modelo	Origem	Padrão de Chegada	Bloqueio	Duração
Poisson	Infinita	Aleatória	LCH	Exponencial
Erlang B	Infinita	Aleatória	LCC	Exponencial
Extended Erlang B	Infinita	Aleatória	LCR	Exponencial
Erlang C	Infinita	Aleatória	LCD	Exponencial
Engset	Finita	Pouca Variação	LCC	Exponencial

As fórmulas só funcionam, com eficiência, se existirem um grande número de origens independentes de tráfego. Por exemplo, se 10 usuários fizerem chamadas externas, sem recebimento de chamadas entrantes, nunca serão necessários mais do que 10 troncos, não importando o resultado do cálculo da fórmula. Matematicamente, estas fórmulas necessitam de fontes infinitas, mas na prática elas funcionam bem se existirem, pelo menos, 10 vezes mais origens (usuários) do que servidores (troncos e agentes) (ANGUS, 2002).

A fórmula mais amplamente adotada em cálculo de planejamento de teletráfego é a de Erlang B. Embora nos EUA, seja usada a fórmula de Poisson, que retorna valores semelhantes a fórmula de Erlang B (ANGUS, 2002);(CISCO, 2001).

4.2.4 Fórmula de Erlang B

A fórmula de Erlang B é utilizada quando uma chamada bloqueada é realmente bloqueada, por exemplo, quando alguém liga para uma linha telefônica e recebe o sinal de ocupado ou tenta acessar um tronco e o encontra em uso. A fórmula é composta de três variáveis: *servidores (linhas)*, *intensidade de tráfego* e *GoS*. Se duas variáveis forem conhecidas é possível calcular a outra variável. Para facilitar a busca dos resultados, existem tabelas prontas com a relação da taxa de bloqueio, intensidade de tráfego e número de linhas. Na figura 4.4, um exemplo da tabela (BROCKMEYER; HALSTRØM; JENSEN, 1948).

x	1 ‰	2 ‰	3 ‰	4 ‰
Servidores (Linhas)	Intensidade de Tráfego (Erlangs)			
1	0.001	0.002	0.003	0.004
2	0.046	0.065	0.081	0.094
3	0.19	0.25	0.29	0.32
4	0.44	0.53	0.60	0.66
5	0.76	0.90	0.99	1.07
6	1.15	1.33	1.45	1.54
7	1.58	1.80	1.95	2.06
8	2.05	2.31	2.48	2.62
9	2.56	2.85	3.05	3.21
10	3.09	3.43	3.65	3.82

Figura 4.4: Parte de uma tabela de Erlang.

A fórmula de Erlang B pode ser utilizada em grupos de troncos primários, onde não é levado em conta o número de repetições, porque os usuários são repassados para outro grupo de troncos, ou quando é esperado uma taxa muito baixa de bloqueios (CISCO, 2001). No caso de PABXs, normalmente, a taxa de bloqueio é muito baixa, sendo aplicável a fórmula de Erlang B.

A fórmula de Erlang B é a seguinte:

$$B(c, a) = \frac{\frac{a^c}{c!}}{\sum_{k=0}^c \frac{a^k}{k!}} \quad (5)$$

Onde:

- $B(c, a)$ é a probabilidade de bloqueio de uma chamada

- c é o número de circuitos
- a é a intensidade de tráfego (em *erlangs*)

São 4 etapas necessárias para a coleta de parâmetros a serem aplicados na fórmula de Erlang B. As etapas são as seguintes:

1. *Coletar o tráfego telefônico*: é necessário saber quanto tráfego será submetido para o grupo de troncos, a cada hora, por pelo menos 5 ou 10 dias úteis. É possível utilizar contas telefônicas, relatórios detalhados das chamadas gerados pelo PABX, estudos da empresa de telefonia e até mesmo contagem manual, ou fazer suposições baseadas em alguns fatores conhecidos da natureza do tráfego telefônico. O objetivo é produzir uma planilha contendo o número de chamadas e mostrando a quantidade de minutos utilizados em chamadas em cada hora. Com estes dados, será possível calcular o tráfego oferecido.
É importante ressaltar que o tráfego nos troncos pode ser bem maior do que o tempo real de conversação, pois o tempo de uso para chamadas externas e para as chamadas entrantes podem usar os mesmos troncos.
2. *Determinar a hora de maior movimento (HMM) (Average Busy Hour (ABH))*: analisar os registros de PABX ou estimar a intensidade de tráfego médio mais alto durante dias úteis. É importante utilizar um período que represente o tráfego médio anual de telefonia.
3. *Definir um fator de GoS*: na maioria dos casos, um fator de P.05 é aceitável, P.10 é ruim e P.001 é tão bom que a maioria dos usuários nunca receberiam um sinal de ocupado. O padrão de GoS definido pela Anatel é P.05 ou menor, para telefonia comercial fixa (ANATEL, 2005).
4. *Aplicar a fórmula*: calcular o número de troncos necessários para transportar a quantidade de tráfego dentro do fator de GoS determinado. Neste cálculo é preciso definir qual resultado é satisfatório. Normalmente, o resultado satisfatório vai depender do que se dispõe para investir em troncos. Se não existir condições financeira para arcar com o custo, pode-se diminuir o número de troncos aumentando a taxa de GoS havendo assim, diminuição da qualidade do serviço.

Para o entendimento de planejamento de capacidade em sistemas telefônicos é importante o conhecimento dos métodos de Erlang. Entretanto, o levantamento dos parâmetros a serem aplicados às fórmulas devem ser corretos para o resultado ser eficiente. Além disso, as fórmulas fazem simplificações da realidade, tal como a fórmula de Erlang B que assume que os usuários que receberem sinal de ocupado não tentarão imediatamente uma nova ligação. Todas as fórmulas de teletráfego calculam probabilidades, não valores absolutos. Os resultados prevêm o que irá acontecer, em média, durante horas de tráfego similar. As horas com tráfego de exceção, não serão representadas integralmente nos resultados das fórmulas (ANGUS, 2002). Estas exceções podem ocorrer em feriados, como no dia das mães, ou emergências em grande escala.

4.2.5 Metodologia

Na figura 4.5, está o fluxograma da metodologia necessária para o planejamento da capacidade de rede telefônica convencional.

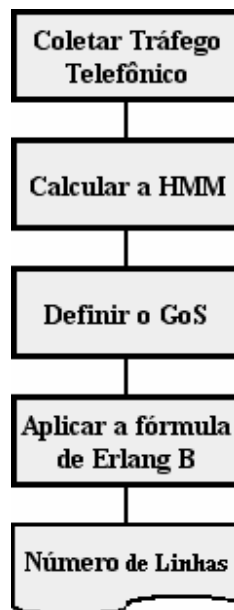


Figura 4.5: Metodologia para o planejamento de capacidade da rede telefônica convencional.

4.3 Aplicação da Metodologia

4.3.1 Coleta de Tráfego Telefônico

Para fazer o levantamento da demanda de tráfego telefônico é necessário conhecer o comportamento dos usuários e a natureza do negócio da instituição. Por exemplo, em caso de uma empresa com várias filiais espalhadas geograficamente, a tendência é que hajam chamadas entre a matriz e as filiais e filiais com filiais. Os dados do histórico de ligações pode ser obtido através dos registros do PABX local ou de um conta telefônica detalhada. Entretanto, para se obter o registro das ligações internas, ramal para ramal, somente os registros do PABX local podem mostrar isto. Este tipo de medição não é comumente monitorado em empresas por ter um alto volume de dados, o que dificulta o armazenamento e o processamento nos PABXs. A dificuldade se encontra no custo elevado de dispositivos de armazenamento para os PABXs, que possuem padrões fechados ou devem ser homologados pela empresa fabricante. Na falta de um registro mais detalhado, podem-se fazer estimativas baseadas nos valores padrão para chamadas telefônicas comerciais.

Para se obter o tempo de duração das chamadas, pode-se dividir o número de ligações pelo tempo total das chamadas. Caso não haja a possibilidade de obter esses dados, o valor usado como padrão em duração de chamadas comerciais locais é entre 180 e 210 segundos (CISCO, 2001). Em medições realizadas no Campus do Vale da UFRGS, os valores ficaram bem próximos do padrão, cerca de 204s em média, para ligações locais, nacionais e internacionais. Em medições de ligações entre ramais feitas na área administrativa da UCPel (Universidade Católica de Pelotas) o tempo de duração das chamadas foi de 54s em média.

Estes dados foram obtidos dos PABXs das instituições através de arquivos de registros (CDR - *Calls Details Records*) fornecidos pelo software tarifador Informatec da empresa STI (STI, 2005) em PABXs do fabricante Ericsson. A filtragem dos arquivos foi feita utilizando as ferramentas *grep*, *sort* e *uniq* rodando no ambiente *Cygwin* (CYGWIN, 2005) e o *Microsoft Excel*. Abaixo, um trecho do arquivo de registro do PABX da UCPel, repre-

sentando as chamadas feitas pelo ramal 8015:

```
Ramal:8015 Nome:NUCLEO CIDADANIA
8015 0 30/06/2005 10:29 INT 8220 Interna 00:00:12
8015 0 30/06/2005 10:31 INT 8206 Interna 00:00:30
8015 0 30/06/2005 10:34 INT 8245 Interna 00:01:06
8015 0 30/06/2005 11:15 INT 8249 Interna 00:00:42
8015 0 30/06/2005 13:43 INT 8258 Interna 00:01:06
8015 0 30/06/2005 13:50 INT 8217 Interna 00:01:00
```

Os registros apresentam o ramal originador, a data, a hora, a natureza da ligação (INT - Interna), o ramal destino e o tempo da ligação.

Na tabela 4.2, estão resumidos os registros coletados dos PABXs das instituições.

Tabela 4.2: Levantamento de tráfego telefônico UCPel e UFRGS (Campus do Vale).

Resumo	UCPel	UFRGS (Campus do Vale)
Período da Coleta	13/06/2005 a 13/07/2005	01/05/2005 a 31/05/2005
Origem	Interna (Ramal-Ramal)	Externa (Local, DDD e DDI)
Número de Ramais	149	2400
Períodos de Maior Movimento (PMM)	08hs às 11hs e 14hs às 17hs	09hs às 12hs e 14hs às 17hs
Hora de Maior Movimento (HMM)	14hs às 15hs	10hs às 11hs
Média de duração das chamadas	0:54s	03:24min
Chamadas total (período)	29826	187717
Chamadas total (período, sem fins de semana)	29762	183295
Chamadas total (PMM)	25650,36	150173,6
Chamadas total (HMM)	4130	9952,68
Dias úteis do período	23	21
Horas úteis do período	7hs às 23hs	7hs às 23hs
Chamadas por ramal total (período)	199,74	76,37
Chamadas por ramal por hora (período)	0,54	0,23
Chamadas por ramal por hora (PMM)	0,94	0,37
Chamadas por ramal por hora (HMM)	1,21	0,19
Concentração de chamadas (PMM)	86%	80%
Concentração de chamadas (HMM)	16,1%	6,63%

4.3.2 Cálculo da Hora de Maior Movimento

Para realizar o cálculo da hora de maior movimento é necessário identificar os períodos com o maior número de chamadas durante o dia. Devido à natureza do negócio das duas instituições (universidades) o período utilizado para o cálculo da HMM foi entre 7hs e 23hs. As ligações que foram geradas fora deste período representaram menos de 0,3% do total de ligações.

As horas de maior movimento são importantes para definir os valores máximos de tráfego. No caso da UCPel, os períodos de maior movimento foram na parte da manhã

das 8hs as 11hs e na parte da tarde das 14hs as 17hs, sendo a terça-feira o dia de maior movimento. Houve uma concentração de 86% das chamadas nesses períodos. A hora de maior movimento foi entre 14hs e 15hs e concentrou 16,1% das ligações realizadas no período. No Campus do Vale da UFRGS, houve apenas uma variação no período da manhã em relação a UCPel, ficando das 9hs as 12hs o intervalo de maior movimento. Na parte da tarde, foi o mesmo da UCPel. O dia de maior movimento foi a terça-feira e as horas de maior movimento concentraram 80% do total das chamadas do intervalo medido. A hora de maior movimento foi entre 10hs e 11hs e concentrou 6,63% das ligações.

Outro fator importante é a intensidade do tráfego. O número de ligações geradas por cada ramal é um dos fatores principais para o cálculo de intensidade de tráfego. No monitoramento feito no Campus do Vale da UFRGS a intensidade de tráfego por ramal ficou em torno de 0,011 erlangs por ramal na hora de maior movimento. No caso da UCPel, a intensidade de tráfego ficou em torno de 0,018 erlangs por ramal.

Para fazer o cálculo do número de linhas (circuitos) necessários para absorver o tráfego medido, deve-se totalizar a demanda de tráfego total. A fórmula 6 demonstra o cálculo de intensidade de tráfego na HMM da UCPel.

$$\begin{aligned} T_{HMM} &= ((n_{chamadas} \cdot t_{chamada}) / 3600) \cdot n_{ramais} \\ T_{HMM} &= ((1, 21.54s) / 3600) \cdot 149 \\ T_{HMM} &= 2,7 \text{ erlangs} \end{aligned} \quad (6)$$

Onde:

T_{HMM} - intensidade de tráfego na HMM

$n_{chamadas}$ - número de chamadas realizadas por cada ramal na HMM

$t_{chamada}$ - duração média das ligações

n_{ramais} - número de ramais

Na tabela 4.3, estão os resultados dos cálculos de intensidade de tráfego no período da coleta de informações, nos períodos de maior movimento e na hora de maior movimento.

Tabela 4.3: Intensidade do tráfego telefônico na UCPel e na UFRGS.

Resumo	UCPel	UFRGS (Campus do Vale)
Intensidade de tráfego por ramal (período)	0,0081e	0,013e
Intensidade de tráfego por ramal (PMM)	0,0141e	0,0209e
Intensidade de tráfego por ramal (HMM)	0,018e	0,011e
Intensidade de tráfego total (período)	2,69e	31,2e
Intensidade de tráfego total (PMM)	2,09e	120e
Intensidade de tráfego total (HMM)	2,7e	26,4e

4.3.3 Definição do GoS

A definição do GoS é baseada na qualidade desejada para o serviço de telefonia. O GoS pode ser medido através da porcentagem de ligações que sofreram bloqueios pela rede, devido a todas as linhas estarem ocupadas, em relação às ligações realizadas com sucesso. Os registros podem ser coletados dos arquivos de registro dos PABXs. Como

padrão, definiu-se o GoS com P.01, ou seja, 1% de probabilidade de bloqueio (COLLINS, 2003).

4.3.4 Aplicação da fórmula de Erlang B

A fórmula de Erlang B necessita de pelo menos dois parâmetros para determinar o terceiro. A intensidade de tráfego já foi determinada e o GoS também. Basta agora aplicar a fórmula para obter-se o número de linhas necessárias para o tráfego telefônico. A aplicação da fórmula é simplificada pelo uso de tabelas com valores já calculados. Por ser uma fórmula com recursividade a sua resolução é bastante complexa. Portanto, existem algoritmos que permitem a resolução do cálculo. Um dos algoritmos que está em formato de uma macro para o Microsoft Excel, **CircuitErlB**, está listado abaixo:

```
Public Function CircuitErlB(Aoffered, Pblock As Double) As Long

Dim Ion As Double
Dim m, Flag As Long

Flag = 0
Ion = 1
m = 1 ' Contador do número de linhas

' Se o tráfego ou a probabilidade for menor que zero, seta em 0
If (Aoffered <= 0) Or (Pblock <= 0) Then
Flag = 1
m = 0
End If

'Máximo de 20 milhões de linhas
While (m < 20000000) And (Flag = 0)

Ion = 1 / (1 + (m / (Ion * Aoffered)))
If (Ion <= Pblock) Then
Flag = 1
Else
m = m + 1
End If

Wend

CircuitErlB = m 'Retorna o número de circuitos

End Function
```

Os parâmetros da macro são o GoS e o tráfego da HMM em Erlangs, sendo o resultado o número de troncos (servidores). A utilização da macro é a seguinte:

```
CircuitErlB(tráfego HMM em erlangs;GOS)
CircuitErlB(2,7;0,01)
CircuitErlB=8 linhas
```

Na figura 4.6, está ilustrada a utilização da tabela de Erlang, para os mesmos parâmetros. Observa-se que a aproximação é feita para cima, utilizando o número de circuitos para uma intensidade de tráfego de 3,13e, maior que o valor calculado anteriormente de 2,7e.

Portanto, 8 linhas suportariam a intensidade de tráfego telefônico gerado por 149 ramais na hora de maior movimento na UCPel.

4.4 Metodologia Adaptada para VoIP

A partir do levantamento do tráfego telefônico, é possível fazer o cálculo da largura de banda necessária para absorver os serviços de VoIP. Para isto, devem ser definidos os tipos de *codecs* a serem utilizados, a quantidade de amostras por pacotes, utilização ou não de supressão de silêncio e compressão de cabeçalhos. Com isto, será possível elaborar uma metodologia para adaptar o planejamento de capacidade de rede da telefonia convencional para os serviços VoIP.

n	$B=0.001$	$B=0.002$	$B=0.003$	$B=0.004$	$B=0.005$	$B=0.010$
1	0.001	0.002	0.003	0.004	0.005	0.01
2	0.05	0.07	0.08	0.09	0.11	0.15
3	0.19	0.25	0.29	0.32	0.35	0.46
4	0.44	0.53	0.60	0.66	0.70	0.87
5	0.76	0.90	0.99	1.07	1.13	1.36
6	1.15	1.33	1.45	1.54	1.62	1.91
7	1.58	1.80	1.95	2.06	2.16	2.50
8	2.05	2.31	2.48	2.62	2.73	3.13

Figura 4.6: Tabela de Erlang para GoS 0,01 e HMM 2,7e.

4.4.1 Tipo de Codec

Para fazer o cálculo da largura de banda necessária para cada ligação, o tipo de *codec* a ser utilizado deve ser definido. Além do consumo de largura de banda, o tipo de *codec* influencia diretamente na qualidade da voz. Em uma rede corporativa, é aconselhável a utilização do *codec* G.711 (WALLINGFORD, 2005). Em casos de enlaces com pouca largura de banda disponível, é aconselhável a utilização de *codecs* com menos consumo de banda, tais como GSM ou Speex, ou a compressão de cabeçalhos RTP. Para fazer o cálculo, foram medidas as larguras de banda utilizadas pelos principais *codecs* gratuitos através do softphone Xlite da empresa Counterpath (COUNTERPATH, 2005) registrado em um *proxy* SIP Asterisk. Este *softphone* tem disponíveis os *codecs* G.711, GSM, iLBC e Speex. Foi medida a largura de banda de cada *codec*, fazendo a captura do tráfego utilizando o software Ethereal (ETHEREAL: A NETWORK PROTOCOL ANALYZER, 2005), em uma rede padrão ethernet conforme a figura 4.7.

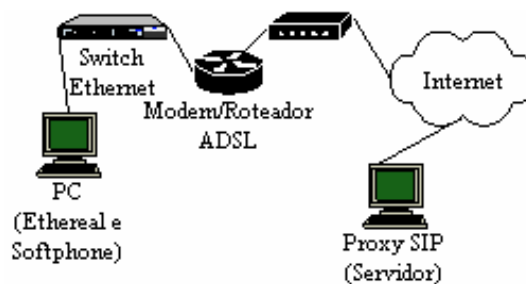


Figura 4.7: Diagrama da rede da captura com o Ethereal.

O consumo de largura de banda de cada *codec* pode ser calculado somando os cabeçalhos e a carga útil por pacote e multiplicando pelo número de pacotes por segundo. Na figura 4.8, o exemplo do cálculo para o *codec* G.711 com carga útil de 160 bytes (duas amostras de voz por pacote).

O cálculo de consumo de largura de banda e os resultados da medição e dos cálculos estão na tabela 4.4.

Os valores da tabela 4.4 correspondem a um sentido da chamada, por exemplo do

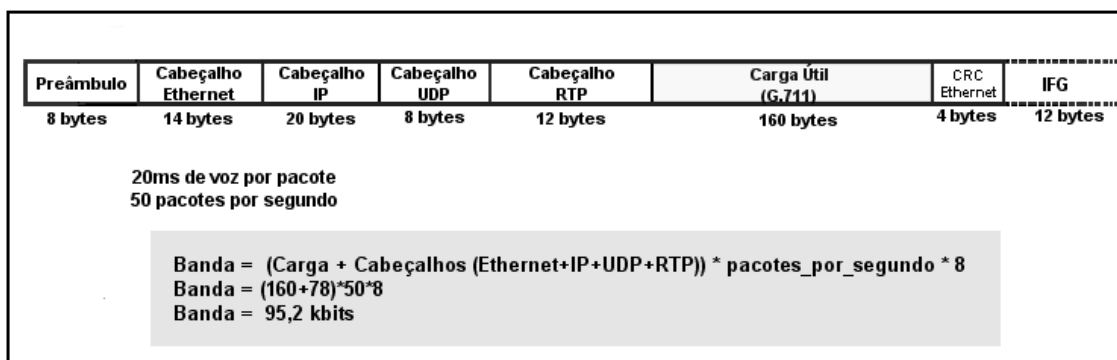


Figura 4.8: Cálculo da largura de banda do *codec* G.711.

Tabela 4.4: Valores medidos e calculados dos *codecs*.

Codec	Banda Medida (kbits)	Banda Calculada (kbits)
G.711	81,6	95,2
GSM	29,7	44,2
iLBC	24,4	34,13
Speex	35,9	39,2

usuário chamado para o chamador. Portanto, a largura de banda consumida é igual a duas vezes o valor medido. A variação ocorrida no cálculo é devido à forma como é feita a captura no software Ethereal em relação aos quadros ethernet. Na captura, são considerados apenas 14 bytes para os cabeçalhos dos quadros (ETHERREAL: WIKI, 2005). Não são considerados no cálculo os 4 bytes do campo FCS (*Frame Check Sequence*) que contêm o valor do CRC (*Cyclic Redundancy Check*), os 8 bytes do preâmbulo e os 12 bytes do IFG (*InterFrame Gap*) correspondente ao intervalo entre os quadros. O *overhead* do cabeçalho do quadro ethernet é de 38 bytes (SPURGEON, 2000). Para aplicação no planejamento de capacidade do cenário exemplo, será utilizado o pior caso, que é o valor calculado.

4.4.2 Amostras por Pacote

A largura de banda utilizada é diretamente afetada pelo número de amostras de voz que são transportadas em cada pacote. Devido ao excessivo *overhead*, aumentando o número de amostras por pacote, aumenta-se a carga útil dos pacotes fazendo com que o *overhead* diminua (figura 4.9). O aumento do número de amostras por pacote faz com que o atraso, fim a fim, aumente devido ao tempo de montagem dos pacotes que irá aumentar. O número de amostras por pacote é definido nas configurações dos *codecs* e pode ser feito nos terminais ou nos servidores.

A tabela 4.5, mostra a relação entre largura de banda, atraso, *overhead* e número de amostras de voz por pacote para o *codec* G.711 em uma rede ethernet.

O atraso e as perdas de pacotes em redes comutadas são mínimos em uma rede local corporativa, mesmo não utilizando mecanismos de QoS. Portanto pode-se usar a estratégia de várias amostras de voz em um mesmo pacote, em LANs.

4.4.3 Supressão de Silêncio

Uma conversação telefônica típica pode conter de 35 a 50% de períodos de silêncio (COLLINS, 2003). Todos os pacotes, inclusive os que contêm silêncio, são transmitidos

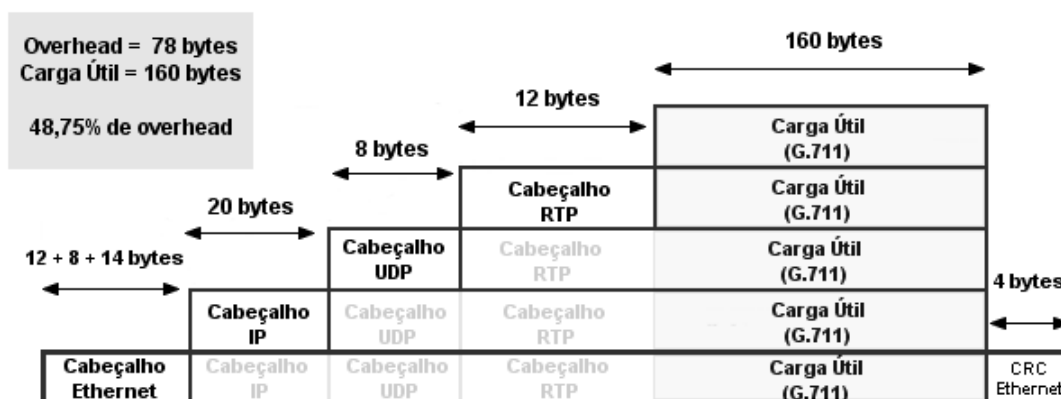


Figura 4.9: *Overhead* dos pacotes de voz (*codec* G.711).

Tabela 4.5: Largura de banda para quantidade de amostras de voz por pacote (*codec* G.711).

Amostras por pacote	Largura de Banda (kbits)	Atraso (ms)	<i>Overhead</i> (%)
1	126,4	10	97,5
2	95,2	20	48,75
3	84,8	30	32,5
4	79,6	40	24,38

consumindo maior largura de banda nas ligações. Com a utilização da técnica de supressão de silêncio, pode-se estimar ganhos de até 35% na banda consumida (CISCO, 2001). Alguns *codecs*, como por exemplo o G.729 Annex-B e o G.723 Annex-A, possuem mecanismos de supressão de silêncio embutidos no próprio *codec*. A supressão de silêncio pode ser uma funcionalidade dos terminais, sendo ativada opcionalmente, embora a qualidade das ligações possa ficar prejudicada, pois podem ocorrer perdas nos recomeços das conversações dos usuários que estavam em silêncio. Portanto, embora possa haver economia de largura de banda a qualidade da conversação pode ser prejudicada utilizando o mecanismo de supressão de silêncio. Em LANs, este mecanismo não é recomendado (WALLINGFORD, 2005).

4.4.4 Compressão de Cabeçalhos

Em enlaces ponto a ponto, é possível a utilização de compressão de cabeçalhos. Desta forma, a largura de banda necessária para as conversações diminui bastante, devido ao alto *overhead* dos pacotes de voz. Por exemplo, no caso de um enlace utilizando o protocolo PPP (*Point to Point Protocol*), sem compressão de cabeçalhos a largura de banda utilizando o *codec* G.711, com duas amostras por pacote, é de 82,4kbits. Com o uso de compressão, a largura de banda seria de 67,2kbits, portanto um ganho de cerca de 18%.

4.5 Recursos de Rede

Para fazer o levantamento de requisitos do planejamento de capacidade em redes corporativas para a absorção de serviços VoIP, é necessário seguir uma metodologia. Caso

contrário, podem surgir diversos problemas no processo de implantação do serviço. O levantamento destes requisitos devem ser feitos antes da compra de equipamentos e contratação de serviços. Em uma rede corporativa, a maior dificuldade é de como adaptar os serviços de VoIP nos recursos de rede existentes. Após fazer o levantamento do tráfego telefônico e a da largura de banda necessária, deve-se avaliar os recursos de rede existentes. Portanto, o tráfego da rede, sem os serviços de VoIP, deve ser analisado previamente, tanto para os valores médios quanto para os valores de pico de tráfego. Os períodos de maior tráfego devem ser identificados. Além disso, os mecanismos de QoS que estão disponíveis na rede devem ser configurados para garantir a qualidade do serviço de voz.

4.5.1 Largura de Banda

A largura de banda em redes corporativas, geralmente, é excedente para as aplicações e serviços de rede. Embora possam existir períodos de pico devido a *backups*, a incidência de vírus, leitura de *e-mail* no início da manhã ou relatórios de final de mês. Então, é importante mapear os picos e os horários onde o tráfego é mais intenso. O tráfego existente na rede, antes da implementação de serviços VoIP, pode ser medido nas interfaces dos *switches* ou roteadores que possuem o protocolo SNMP (*Simple Network Management Protocol*) ou alguma outra forma de gerenciamento. Podem ser usadas nas medições a ferramenta de código-fonte aberto MRTG (*Multi Router Traffic Grapher*) (MRTG: THE MULTI ROUTER TRAFFIC GRAPHER, 2005), que apresenta de forma gráfica o tráfego de entrada e saída das interfaces (figura 4.10) ou em formato texto através de arquivos de *logs*.

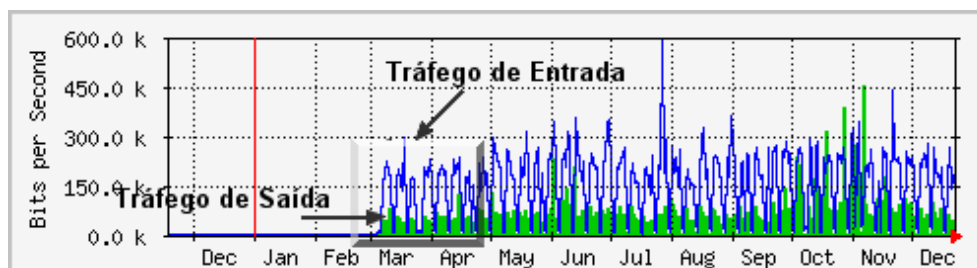


Figura 4.10: Exemplo de medição de tráfego com o MRTG.

Utilizando os arquivos de *logs*, pode-se filtrar os dias e os horários através de uma planilha de cálculos. Desta forma, torna-se possível achar os horários e os dias de maior movimento. Abaixo, um trecho de um arquivo de *log* do MRTG.

```
1138644930 52257 33277 52257 33277
1138644630 12942 27968 12942 27968
1138644600 13272 28617 16138 34253
1138644300 14875 32775 16138 34253
1138644000 5922 21025 23898 30665
```

O primeiro campo contém o número de segundos desde 01/01/1970 (*Unix Epoch Time*), o segundo campo mostra a média de octetos de entrada na interface, o terceiro campo mostra a média de octetos de saída, o quarto campo mostra o valor máximo (pico) de entrada nos últimos cinco minutos e o quinto campo mostra o valor máximo (pico) de saída nos últimos cinco minutos na interface monitorada.

4.5.2 Mecanismos de QoS

Os equipamentos de rede devem possuir mecanismos de QoS que garantam, ao menos, a priorização do tráfego de voz em relação ao restante do tráfego. O padrão 802.1p é uma funcionalidade comum nos *switches* atuais. Os *switches* mais simples possuem apenas duas filas para diferenciação do tráfego, enquanto os *switches* mais completos e de maior custo possuem mais filas. Os roteadores baseados em Linux e os roteadores dedicados, tais como os da empresa Cisco, possuem o protocolo DiffServ e mapeamento entre DSCPs e CoS. Algumas placas de rede atuais possuem *drivers* que permitem a marcação de quadros através do padrão 802.1p, mas não são o padrão em redes corporativas. Normalmente, os ATAs e os telefones IP possuem o padrão 802.1p para a marcação e priorização de quadros em todos os tipos de modelos atuais.

4.5.3 Metodologia Adaptada para VoIP

Na figura 4.11, está representado um diagrama da metodologia para o planejamento de capacidade de rede para implementação de serviços VoIP em redes corporativas.

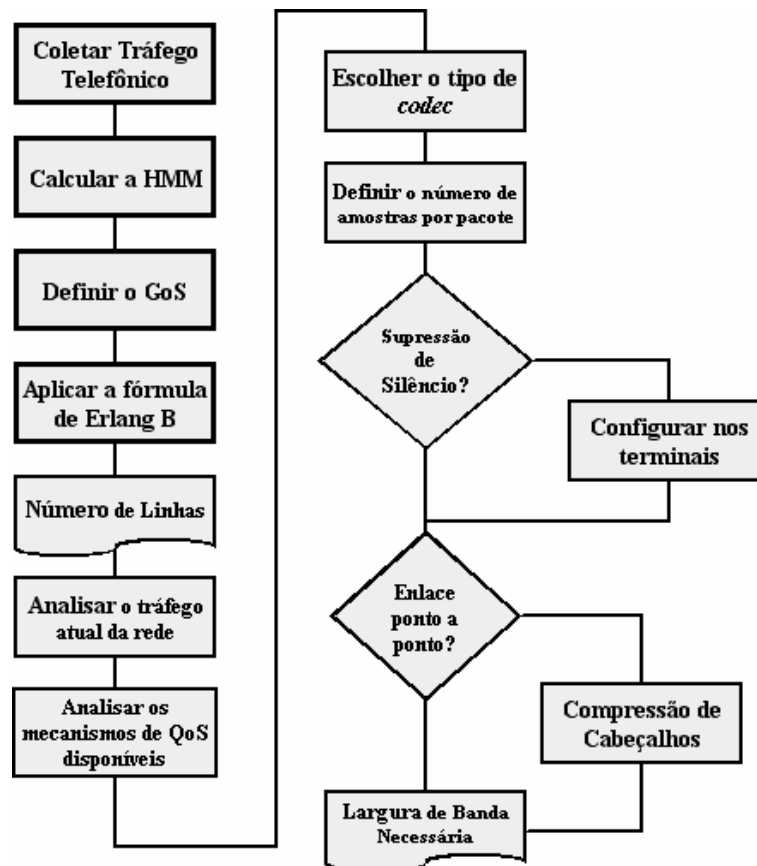


Figura 4.11: Metodologia para o planejamento de capacidade da rede para serviços VoIP.

5 ESTUDO DE CASO

A aplicação da metodologia definida no capítulo 4 será realizada como estudo de caso na rede administrativa da UCPel. Baseado nos levantamentos de tráfego telefônico convencional, já realizados neste trabalho, foram feitas simulações das chamadas VoIP utilizando a ferramenta Callgen323 (OPENH323, 2005). O objetivo das simulações foi medir a qualidade das ligações telefônicas, através de VoIP, em um ambiente real. Os parâmetros de QoS, atraso, *jitter* e perdas, das chamadas foram analisados com a ferramenta Ethereal. O tráfego de rede foi analisado para identificar a hora e o dia de maior movimento. Nas medições, para gerar o tráfego de rede da HMM foi utilizada a ferramenta Iperf (NLANR/DAST : IPERF 1.7.0 - THE TCP/UDP BANDWIDTH MEASUREMENT TOOL, 2005). A descrição do ambiente, os procedimentos e os resultados das medições serão apresentados e analisados a seguir.

5.1 Cenário

O cenário utilizado para a aplicação da metodologia de planejamento de capacidade foi a rede administrativa da Universidade Católica de Pelotas (UCPel). A partir do levantamento de intensidade de tráfego telefônico feito na instituição, foi possível determinar quantas ligações são realizadas na hora de maior movimento, e desta forma, os dados foram aplicados no cálculo para os serviços de VoIP. Neste cenário, o objetivo é eliminar a telefonia convencional, fazendo com que todos os serviços de telefonia internos, ramal para ramal, sejam em VoIP. Por exemplo, a utilização de um PABX IP em software, tal como o Asterisk, substituindo o PABX convencional.

5.2 Rede de Telefonia

Na UCPel, os serviços internos de telefonia são fornecidos por um PABX da empresa Ericsson modelo MD110, interligando cerca de 149 ramais. Conforme medições realizadas no capítulo 4, os dados resumidos nas tabelas 4.2 e 4.3 serão utilizados no estudo de caso. Na figura 5.1, é mostrada a distribuição da quantidade dos ramais nos segmentos da rede administrativa da UCPel.

5.3 Rede de Dados

A rede de dados da área administrativa da UCPel é composta de um *backbone* funcionando na velocidade de 100Mbits através de *switches* padrão ethernet, sendo segmentada por fibras óticas. Em cada ponta das fibras existem *switches* e *hubs*, na maioria a 10Mbits,

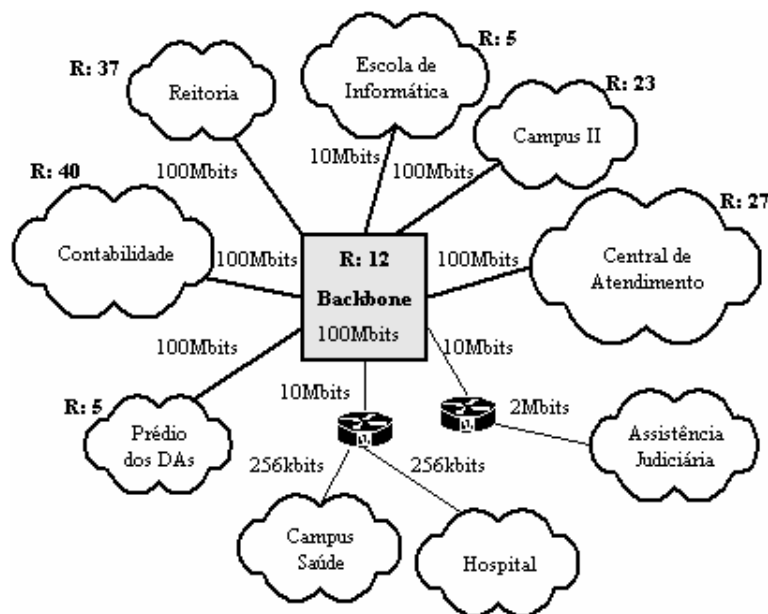


Figura 5.1: Diagrama da rede com a distribuição dos ramais

servindo de acesso para as estações. São cerca de 600 hosts, com mais de 5000 usuários entre alunos, funcionários e professores. Além disso, existem três localidades remotas que utilizam serviços de linhas dedicadas (Assistência Judiciária) e de *frame relay* (Hospital e Campus Saúde) para conexão com a rede administrativa. Os sites remotos possuem independência dos serviços telefônicos, e não serão considerados no estudo de caso.

Em um futuro próximo, serão disponibilizados *access points* para a comunicação sem fio. As particularidades das redes sem fios com o uso de serviços VoIP são discutidos em (ELAUD; FAMOLARI; GHOSH, 2005), onde é abordado o planejamento de capacidade de redes.

A figura 5.2, apresenta a ilustração da topologia simplificada da rede.

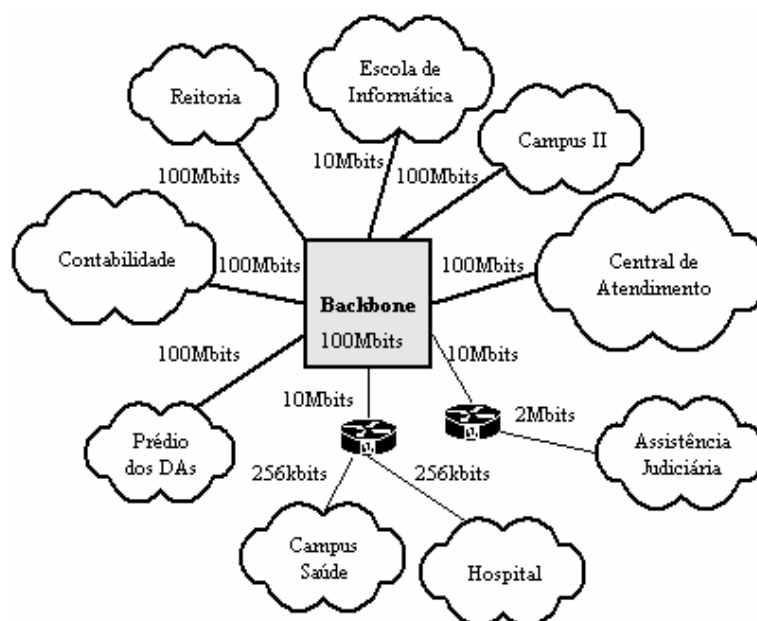


Figura 5.2: Topologia de rede simplificada.

5.3.1 Tráfego de Rede

Atualmente, o tráfego de rede é medido através da ferramenta MRTG, fazendo leitura das interfaces dos *switches* e roteadores. As estatísticas são coletadas de 5 e 5 minutos e são gerados gráficos com o tráfego de entrada e de saída. Portanto, pode-se saber o tráfego médio e os períodos de maior movimento durante o dia, a semana, o mês e o ano. Os picos de tráfego curtos são de difícil análise com a ferramenta MRTG. Embora, dependendo do tempo de amostragem, ainda seja possível identificar picos de tráfego, mesmo em coletas periódicas de 5 minutos.

Através de arquivos de *log* e utilizando planilhas de cálculos, é possível descobrir as horas e os dias de maior movimento e o tráfego médio da HMM, Na tabela 5.1, estão relacionados os valores de tráfego médio de cada segmento e a hora de maior tráfego coletados através dos *logs* da ferramenta MRTG.

Tabela 5.1: Levantamento do tráfego de dados.

Segmento	Média de Tráfego em HMM (kbits)	HMM	Dia de Maior Movimento
Reitoria	541,47	19hs	Terça
Campus II	486,56	19hs	Terça
Escola de Informática	1121,35	17hs	Quinta
Prédio dos DAs	284,6	17hs	Terça
Central de Atendimento	400,33	21hs	Quarta
Contabilidade	866,37	15hs	Terça
<i>Backbone</i>	9307,96	19hs	Terça

5.3.2 Mecanismos de QoS

Em relação as garantias de QoS, os *switches* do *backbone* possuem os padrões IEEE 802.1p e 802.1Q, o que torna possível separar o tráfego em VLANs e priorizar os pacotes de voz. Atualmente, não são usadas múltiplas VLANs e nenhum mecanismo de QoS para o tráfego de rede.

Os *switches* principais do backbone são da empresa 3COM modelo SuperStack III 4400SE, e possuem o padrão 802.1p ativado automaticamente, fazendo distinção entre os quadros marcados na origem. O *switch* possui duas filas para priorização de tráfego, uma de alta e outra de baixa prioridade, mas nenhuma forma de configuração das políticas das filas.

Entre os roteadores, são usados servidores com o sistema operacional Linux. Neste caso, podem ser usados mecanismos de *DiffServ* com auxílio da ferramenta IPTables (NETFILTER, 2005) para a priorização do tráfego de voz. Nos roteadores dos *sites* remotos, Hospital, Campus Saúde e Assistência Judiciária, são utilizados equipamentos da empresa Cisco com capacidade para *DiffServ* e *IP Precedence*.

As questões de configuração de QoS em *switches* e roteadores não serão abordados no cenário do estudo de caso, sendo o planejamento de capacidade da rede o foco principal. Nas simulações de chamadas, a ferramenta geradora de chamadas, Callgen323, não permite a marcação dos quadros com o padrão 802.1p. Portanto, os resultados dos testes correspondem a um rede sem mecanismos de QoS ou CoS.

5.4 Planejamento de Capacidade da Rede

Será aplicada a metodologia, definida no capítulo 4, para implantar serviços VoIP na rede administrativa da UCPel. Com os valores resultantes da aplicação da metodologia, será possível realizar a simulação das chamadas e a medição da qualidade das ligações.

5.4.1 Cálculo de HMM

A intensidade de tráfego medida através da coleta de registros do PABX, mostra que cada ramal telefônico na parte administrativa da UCPel possui 0,018 erlangs em hora de maior movimento. O cálculo é demonstrado abaixo:

$$\begin{aligned} A_{UCPel} &= N_{chamadas} \cdot T_{chamada} / N_{ramais} / 3600 \\ A_{UCPel} &= (179,57) \cdot 54 / 149 / 3600 \\ A_{UCPel} &= 0,018e \end{aligned} \quad (7)$$

Onde:

- $N_{chamadas}$ - corresponde a média do número de chamadas na hora de maior movimento. Conforme dados da tabela 4.2, foram feitas 4130 ligações em cada hora de maior movimento durante 23 dias. Portanto, o número médio de ligações é igual a $4130/23$, o que resulta em 179,57 ligações por HMM;
- $T_{chamada}$ - corresponde a duração média das chamadas (54 segundos);
- N_{ramais} - corresponde ao número total de ramais;
- 3600 - para obter o resultado em Erlangs.

Sabendo a intensidade de tráfego de cada ramal, é possível calcular o tráfego total oferecido na hora de maior movimento, da seguinte forma:

$$\begin{aligned} \text{Traf}_{\text{Oferecido}} &= \text{Traf}_{\text{Ramal}} \cdot \text{Ramais} \\ \text{Traf}_{\text{Oferecido}} &= 0,018 \cdot 149 \\ \text{Traf}_{\text{Oferecido}} &= 2,7e \end{aligned} \quad (8)$$

A intensidade de tráfego total de 2,7e, não corresponde a distribuição dos ramais dentro da rede administrativa da UCPel. O valor só seria válido se todos os ramais estivessem disputando o mesmo grupo de linhas, por exemplo um *gateway* com um grupo de linhas convencionais.

No caso de VoIP, a segmentação da rede tem a função de criar vários grupos de linhas. No caso específico da rede administrativa da UCPel, os segmentos conectados por fibras poderiam ser considerados como grupo de linhas (trancos). Sendo assim, a intensidade de tráfego deve ser calculada para cada um destes segmentos. Não está sendo considerado o tráfego entre ramais dentro do mesmo segmento. Observa-se que se um mesmo grupo de linhas respondesse por todo o tráfego, com GoS de 1%, seriam necessárias apenas 8 linhas. Fazendo o cálculo por segmento, o número de linhas necessárias passaria a ser de 22. Isto ocorre devido ao número de fontes originadoras de tráfego. Na aplicação da fórmula de Erlang B, quanto maior o número de fontes originadoras, maior será a taxa de ocupação das linhas (IVERSEN et al., 2005).

O cálculo da intensidade de tráfego para o segmento da Reitoria é mostrado na fórmula 9. O cálculo para os outros segmentos segue o mesmo procedimento e na tabela 5.2 são apresentados os resultados.

$$\begin{aligned} \text{Traf}_{\text{Oferecido}} &= \text{Traf}_{\text{Ramal}} \cdot \text{Ramais} \\ \text{Traf}_{\text{Oferecido}} &= 0,018.37 \\ \text{Traf}_{\text{Oferecido}} &= 0,67e \end{aligned} \quad (9)$$

Tabela 5.2: Intensidade de tráfego por segmento.

Segmento	Intensidade de Tráfego
Reitoria	0,67e
Campus II	0,41e
Escola de Informática	0,09e
Prédio dos DAs	0,09e
Central de Atendimento	0,49e
Contabilidade	0,72e
<i>Backbone</i>	0,22e

5.4.2 Definição do GoS

Foi definido que o GoS será de P0.01, ou seja, que 1% de probabilidade que alguma ligação seja bloqueada. Embora o conceito de bloqueio em uma rede de pacotes não seja o mesmo do que em uma rede baseada em circuitos, pode-se usar um mecanismo de controle de admissão de chamadas (*Call Admission Control*). Este mecanismo de controle deve estar localizado no SIP *proxy* ou no H.323 *gatekeeper*. Com isto, ao ser detectado o número máximo de ligações simultâneas, calculado previamente, não serão mais permitidas novas ligações. Por exemplo, o acréscimo de mais uma ligação pode tornar a qualidade de todas as outras ligações inaceitável, devido ao esgotamento dos recursos de rede.

Em arquiteturas que utilizam *gateways*, o GoS vai ser dependente do número de troncos disponíveis. Por exemplo, o mecanismo de controle de admissão de chamadas no *proxy* SIP Asterisk, é através da diretiva de configuração **call_limit**, que limita o número de ligações simultâneas.

5.4.3 Aplicação da Fórmula de Erlang B

A aplicação da fórmula de Erlang B torna possível descobrir quantas linhas são necessárias para comportar determinada intensidade de tráfego telefônico. Como exemplo, será usado o segmento da Reitoria. O tráfego oferecido é de 0,66e, com GoS de 1%. Portanto, aplicando a fórmula, através da macro do Excel **CircuitErlB** apresentada no capítulo 4, poderá ser obtido o número de linhas necessárias para o tráfego telefônico calculado, da seguinte forma:

```
CircuitErlB(tráfego HMM em erlangs;GOS)
CircuitErlB(0,67e;0,01)
CircuitErlB=4 linhas
```

Pode-se usar também a tabela de erlang para saber o número de linhas, como está demonstrado na figura 5.3.

Este procedimento foi aplicado em todos os segmentos da rede e os resultados estão listados na tabela 5.3.

n	$B=0.001$	$B=0.002$	$B=0.003$	$B=0.004$	$B=0.005$	$B=0.010$	$B=0.020$	$B=0.050$
1	0.001	0.002	0.003	0.004	0.005	0.01	0.02	0.05
2	0.05	0.07	0.08	0.09	0.11	0.15	0.22	0.38
3	0.19	0.25	0.29	0.32	0.35	0.46	0.60	0.90
4	0.44	0.53	0.60	0.66	0.70	0.87	1.09	1.52
5	0.76	0.90	0.99	1.07	1.13	1.36	1.66	2.22

Figura 5.3: Número de servidores para tráfego de 0,67e e GoS 1%

Tabela 5.3: Número de linhas por segmento.

Segmento	Intensidade de Tráfego	Linhas
Reitoria	0,67e	4
Campus II	0,41e	3
Escola de Informática	0,09e	2
Prédio dos DAs	0,09e	2
Central de Atendimento	0,49e	4
Contabilidade	0,72e	4
<i>Backbone</i>	0,22e	3

5.4.4 Tipo de Codec

O pior caso de consumo de largura de banda acontece quando da utilização do *codec* G.711. Em contrapartida, o G.711 possui a melhor qualidade de voz, sendo preferível em casos onde a largura de banda não é problema, como em redes locais. Além disto, o *codec* G.711 tem um bom desempenho em relação a perdas de pacotes e possui tempo de codificação e processamento muito baixo. Portanto, para efeito de projeto será utilizado o referido *codec*.

5.4.5 Número de Amostras por pacote

O número de amostras por pacote é dependente da incidência de erros e do atraso da rede. Como visto na capítulo 4, quanto maior o número de amostras de voz por pacote, maior o atraso e maior o prejuízo para a conversação em caso de perdas. Em uma rede local comutada, as perdas são desprezíveis e o atraso é normalmente baixo, podendo ser controlado com mecanismos de QoS. Portanto, serão utilizadas 3 amostras por pacote, possuindo cada pacote 240 bytes de carga útil (voz) com o uso do *codec* G.711. Com 3 amostras por pacote, cada ligação ocupará 169,6kbits.

5.4.6 Supressão de Silêncio

A supressão de silêncio pode ser ativada nos terminais para diminuir a largura de banda utilizada. A ferramenta utilizada na simulação das chamadas, Callgen323, não permite a utilização deste mecanismo e portanto não foi utilizado nos testes.

5.4.7 Compressão de Cabeçalhos

A compressão de cabeçalhos só pode ser utilizada em enlaces ponto a ponto, o que não foi aplicável nas medições efetuadas. No estudo de caso, poderia ser aplicada nos enlaces entre o *backbone* e as redes locais remotas da Assistência Judiciária, Campus Saúde e

Hospital.

5.4.8 Cálculo da Largura de Banda

Para realizar o cálculo do consumo de largura de banda necessária para absorver os serviços de VoIP na rede administrativa da UCPel, deve-se fazer o cálculo do número de linhas necessárias para absorver o tráfego telefônico na HMM, multiplicado pela largura de banda de cada chamada. Se fossem utilizados todos os ramais em um mesmo segmento, a largura de banda necessária para absorver todo o tráfego telefônico seria de 1,37 Mbits conforme a fórmula 10.

$$\begin{aligned}
 Banda_{Total} &= tron \cos . banda_{chamada} . 2 \\
 Banda_{Total} &= 8.84, 8kbits.2 \\
 Banda_{Total} &= 678, 4kbits.2 \\
 Banda_{Total} &= 1, 37Mbits
 \end{aligned}
 \tag{10}$$

Mas no estudo de caso, o cálculo deve ser aplicado para cada segmento em separado. O resultado do cálculo para o segmento da Reitoria está demonstrado na fórmula 11.

$$\begin{aligned}
 Banda_{Total} &= tron \cos . banda_{chamada} . 2 \\
 Banda_{Total} &= 4.84, 8kbits.2 \\
 Banda_{Total} &= 339, 2kbits.2 \\
 Banda_{Total} &= 678, 4kbits
 \end{aligned}
 \tag{11}$$

O resultado final para todos os segmentos está representado na tabela 5.4 e no diagrama da figura 5.4:

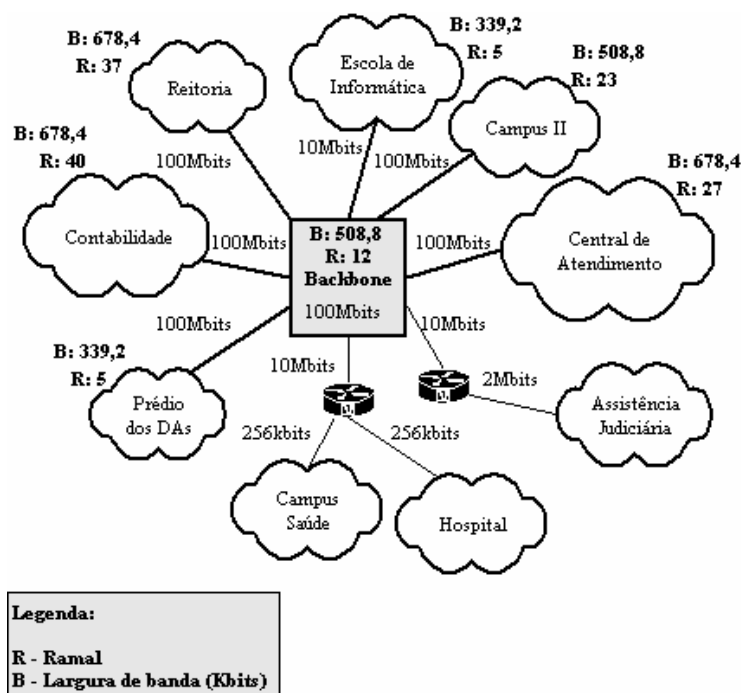


Figura 5.4: Diagrama da distribuição de ramais e a largura de banda

Tabela 5.4: Largura de banda por segmento.

Segmento	Ramais	Tráfego (erlangs)	Troncos	Banda (kbits)
Reitoria	37	0,67	4	678,4
Campus II	23	0,41	3	508,8
Escola de Informática	5	0,09	2	339,2
Prédio dos DAs	5	0,09	2	339,2
Central de Atendimento	27	0,49	4	678,4
Contabilidade	40	0,72	4	678,4
<i>Backbone</i>	12	0,22	3	508,8

5.5 Medições

Foram feitas medições utilizando a ferramenta Callgen323, para simulação de chamadas e os dados levantados e calculados conforme a metodologia aplicada na rede administrativa da UCPel. Os valores de atraso, perdas e *jitter* foram avaliados.

5.5.1 Ferramentas

As ferramentas **Callgen323**, **Tcpdump**, **Iperf** e **Ethereal** foram utilizadas para os testes.

5.5.1.1 Callgen323

A ferramenta CallGen323 faz parte do projeto OpenH323 (OPENH323, 2005) e tem como função gerar chamadas utilizando *codecs* de voz e os protocolos RTP e UDP. Originalmente, foi criada para testar a implementação de *gatekeepers* H.323, mas pode ser aplicada em qualquer ambiente que necessite testar fluxos de áudio RTP. É uma ferramenta de código-fonte aberto e pode ser utilizada em várias plataformas, tais como Linux e Windows. É possível, através dos arquivos de *log* gerados pela ferramenta, analisar os valores de atraso, *jitter* e perda de pacotes das chamadas. Também é possível salvar o arquivo de áudio recebidos para posterior comparação com o áudio original. A versão utilizada foi a 1.2.26.

5.5.1.2 Tcpdump

A ferramenta Tcpdump é a ferramenta padrão para a captura de tráfego de rede em sistemas operacionais Unix e Linux. Baseada na biblioteca *libpcap* (TCPDUMP, 2005), possui filtros que permitem a captura seletiva de pacotes provendo a gravação em arquivo para posterior análise.

5.5.1.3 Iperf

A ferramenta Iperf possibilita gerar tráfego UDP e medir a máxima largura de banda TCP (NLNR/DAST : IPERF 1.7.0 - THE TCP/UDP BANDWIDTH MEASUREMENT TOOL, 2005). Permite ajustar vários parâmetros do protocolo TCP, tal como tamanho de *buffer*, e várias características do protocolo UDP, tal como largura de banda, para gerar tráfego de rede. O Iperf pode gerar relatórios de largura de banda, atraso, *jitter* e perdas de pacotes. A versão utilizada foi a 2.0.2.

5.5.1.4 Ethereal

A ferramenta Ethereal permite capturar e analisar o tráfego de rede, reconhecendo mais de 400 protocolos (ETHERREAL: WIKI, 2005). Possui interpretadores, chamados (*dissectors*.) para os protocolos de VoIP, tais como SIP, H.323 e RTP. Para as medições realizadas foram utilizados recursos de análise de fluxos de áudio RTP, para medir atraso, perdas e *jitter* nos pacotes capturados (figura 5.5). O Ethereal é uma ferramenta de código-fonte aberto e está disponível para várias plataformas. A versão utilizada foi a 0.10.14.

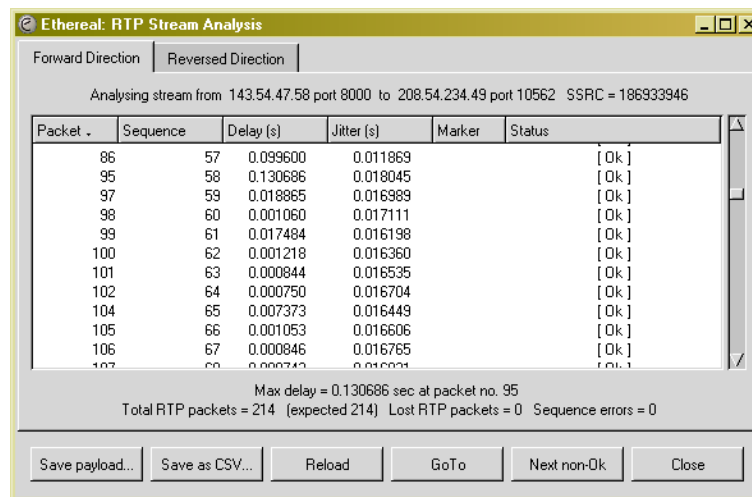


Figura 5.5: Tela da ferramenta Ethereal para análise de fluxos RTP

5.5.2 Procedimentos

Para realizar os testes é necessário saber o número de ligações que cada ramal realiza em HMM. De acordo com a tabela 4.2, são realizadas 1,21 chamadas por ramal em HMM. O número de linhas necessários para a intensidade de tráfego, corresponde a simultaneidade das chamadas. Por exemplo, no caso do segmento da Reitoria são 37 ramais, que geram um total de 44,77 chamadas (37 ramais x 1,21 chamadas por ramal). O número de linhas necessárias para absorver o tráfego, utilizando um GoS de 1%, é de 4 linhas. Portanto, existirão 4 chamadas simultâneas (4 linhas ocupadas) durante a HMM.

Na tabela 5.5, o total de chamadas por segmento em HMM.

Tabela 5.5: Total de chamadas por segmento.

Segmento	Ramais	Chamadas Simultâneas	Total de Chamadas
Reitoria	37	4	44,77
Campus II	23	3	27,83
Escola de Informática	5	2	6,05
Prédio dos DAs	5	2	6,05
Central de Atendimento	27	4	32,67
Contabilidade	40	4	48,4
<i>Backbone</i>	12	3	14,52

Em cada segmento, foram utilizados dois *hosts* para simular as chamadas e o tráfego. Um dos *hosts* executou a ferramenta Iperf, gerando tráfego, e o outro *host* gerou o número

de chamadas correspondentes ao volume de chamadas da HMM, utilizando a ferramenta CallGen323. Os dois *hosts* utilizaram as ferramentas na plataforma Windows.

Abaixo, os parâmetros utilizados nas ferramentas como clientes para o segmento da Reitoria:

Linha de comando:

```
callgen323 -n -m 4 -r 11 -O teste.wav --tmincall 54
--tmaxcall 54 --tminwait 1 --tmaxwait 4 200.17.170.37
```

```
-n: ignorar a presença de um gatekeeper H.323
-m: número de chamadas simultâneas
-r: número de repetições (4 simultâneas x 11)
-O: arquivo com áudio
--tmincall: tempo mínimo de duração da chamada
--tmaxcall: tempo máximo da duração da chamada
--tminwait: tempo mínimo de espera entre as chamadas
--tmaxwait: tempo máximo de espera entre as chamadas
```

Linha de comando:

```
iperf -u -c 200.17.170.152 -b 541,47K -d -t 2400
```

```
-u: utilizar o protocolo UDP
-c: endereço do ponto de conexão (servidor)
-b: valor da largura de banda (541,47K)
-d: bi-direcional
-t: tempo de duração do tráfego em segundos
```

O arquivo *teste.wav* foi criado através da ferramenta **Audacity** (AUDACITY: FREE AUDIO EDITOR AND RECORDER, 2005) e contém um trecho de conversação telefônica pré-gravada de 60 segundos. O formato do arquivo é 16 bits, 8000Hz e mono, correspondendo a voz humana. A ferramenta CallGen323 utiliza o arquivo para simular uma chamada real.

Os *hosts* receptores do tráfego e das chamadas utilizaram as ferramentas como servidores (modo de escuta). No caso do *host* que rodou a ferramenta CallGen323 como servidor, a ferramenta Tcpcdump ficou rodando no mesmo *host* para capturar o tráfego das chamadas. O tráfego capturado serviu para a posterior análise na ferramenta Ethereal.

Abaixo, as linhas de comando e os parâmetros utilizados para as ferramentas utilizadas no modo servidor:

Linha de comando:

```
callgen323 -l -n -ttt -o reitoria.trace
-n: ignorar a presença de um gatekeeper H.323
-l: ficar em estado de espera (recepção)
-ttt: nível de detalhamento do arquivo de log
-o: arquivo para gravação do log
```

Linha de comando:

```
tcpdump -i eth0 -w reitoria.dump udp
-i: interface para captura
-w: arquivo para gravação da captura
```

udp: filtro para capturar somente o pacotes com o protocolo UDP.

Linha de comando:

```
iperf -u -s -i 30
-u: utilizar o protocolo UDP
-s: ficar em estado de espera (recepção)
-i: geração de relatórios a cada 30s
```

5.5.3 Ambiente de Testes

Os testes foram realizados em horário fora do expediente de funcionamento da UC-Pel. Para gerar o tráfego correspondente ao da HMM, foi utilizada a ferramenta Iperf. Para gerar o tráfego, foi posicionado um gerador de tráfego Iperf em cada segmento e um servidor, localizado no segmento **Backbone**, para a terminação dos fluxos. O tráfego médio de cada segmento foi gerado de forma constante durante a duração dos testes com as chamadas. Desta forma, pode-se fazer uma aproximação do tráfego real de rede em HMM. A plataforma utilizada foi o Windows para os geradores de tráfego e Linux para as terminações. Para a recepção das chamadas, foi utilizado um servidor Linux rodando a versão 9 do RedHat, em um segmento de 100Mbits comutado. Este servidor fez o papel equivalente ao do PABX IP. A ferramenta CallGen323 ficou rodando em estado de espera, recebendo as chamadas e gerando estatísticas. No mesmo servidor, ficou rodando o utilitário *Tcpdump* para coletar o tráfego UDP na interface de rede. O tráfego foi armazenado em arquivo para posterior análise na ferramenta Ethereal. Nos clientes foi utilizada a ferramenta CallGen323 para Windows para gerar as ligações, nos períodos pré-determinados. Na figura 5.6, o ambiente de testes é representado.

Os arquivos capturados pelo tcpdump foram analisador através da ferramenta Ethereal. Os fluxos de RTP foram filtrados e as informações sobre *jitter*, atraso e perdas, foram exportados, através de arquivos texto, para planilhas de cálculo no Microsoft Excel. Então, os valores de *jitter* médio, *jitter* máximo, perdas máximas, atraso médio, atraso máximo, desvio padrão do *jitter* e do atraso foram calculados para cada conjunto de chamadas por segmento. Os atrasos correspondem somente a um caminho da ligação, sendo equivalente ao atraso entre o emissor e o receptor, sem o atraso de retorno da ligação. Para determinar o atraso total da ligação, basta multiplicar por dois o valor do atraso medido.

Na figura 5.7, a tela da ferramenta Ethereal com a análise dos fluxos RTP do segmento da Reitoria. Em destaque, a função para a exportação dos dados para arquivo texto.

Foram realizados testes com a intensidade de tráfego de rede e de chamadas atuais, com um volume adicional de 50% no tráfego de rede e com um aumento de 50% no tráfego de rede e de telefonia, projetando um crescimento da demanda de serviços, conforme a tabela 5.6. Os resultados das medições serão apresentados a seguir.

5.6 Resultados

Na tabela 5.6, são apresentados os resultados das medições feitas no ambiente de teste para a intensidade de tráfego atual. Na tabela 5.7, os resultados das medições com um acréscimo de 50% no tráfego de rede. Na tabela 5.8, os resultados das medições com um acréscimo de 50% no tráfego de rede e no número de ramais.

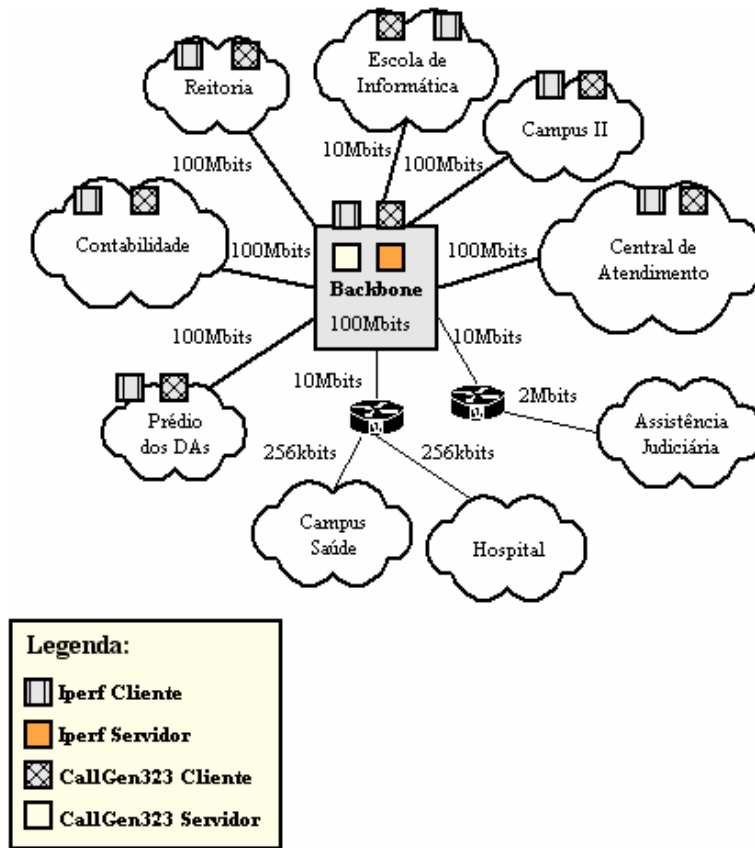


Figura 5.6: Ambiente de testes.

Ethereal: RTP Streams

Detected 43 RTP streams. Choose one for forward and reverse direction for analysis

Src IP addr	Src port	Dest IP addr	Dest port	SSRC	Payload	Packets	Lost	Max Delta (ms)	Max Jitter
200.17.170.160	5002	200.17.170.37	5002	3359016962	ITU-T G.711 PCMU	1331	0 (0.0%)	46.69	
200.17.170.160	5004	200.17.170.37	5004	3667696040	ITU-T G.711 PCMU	1324	0 (0.0%)	66.75	
200.17.170.160	5006	200.17.170.37	5006	6596593005	ITU-T G.711 PCMU	1326	0 (0.0%)	53.79	
200.17.170.160	5008	200.17.170.37	5008	3561101317	ITU-T G.711 PCMU	1325	0 (0.0%)	71.72	
200.17.170.160	5010	200.17.170.37	5010	1772039141	ITU-T G.711 PCMU	1331	0 (0.0%)	74.04	
200.17.170.160	5012	200.17.170.37	5012	1801072737	ITU-T G.711 PCMU	1330	0 (0.0%)	89.64	
200.17.170.160	5014	200.17.170.37	5014	529481101	ITU-T G.711 PCMU	1315	0 (0.0%)	71.71	
200.17.170.160	5016	200.17.170.37	5016	1486656504	ITU-T G.711 PCMU	1330	0 (0.0%)	100.87	
200.17.170.160	5018	200.17.170.37	5018	600814304	ITU-T G.711 PCMU	1331	0 (0.0%)	101.18	
200.17.170.160	5020	200.17.170.37	5020	2466305552	ITU-T G.711 PCMU	1317	0 (0.0%)	64.29	
200.17.170.160	5022	200.17.170.37	5022	2378790156	ITU-T G.711 PCMU	1330	0 (0.0%)	80.03	
200.17.170.160	5024	200.17.170.37	5024	2634353537	ITU-T G.711 PCMU	1292	0 (0.0%)	64.31	
200.17.170.160	5026	200.17.170.37	5026	2400263238	ITU-T G.711 PCMU	1312	0 (0.0%)	78.74	
200.17.170.160	5028	200.17.170.37	5028	3296446061	ITU-T G.711 PCMU	1314	0 (0.0%)	78.05	
200.17.170.160	5030	200.17.170.37	5030	1565061487	ITU-T G.711 PCMU	1324	0 (0.0%)	78.72	
200.17.170.160	5032	200.17.170.37	5032	1945685568	ITU-T G.711 PCMU	1290	0 (0.0%)	60.51	
200.17.170.160	5034	200.17.170.37	5034	632612419	ITU-T G.711 PCMU	1327	0 (0.0%)	61.54	
200.17.170.160	5036	200.17.170.37	5036	3853963606	ITU-T G.711 PCMU	1326	0 (0.0%)	61.55	
200.17.170.160	5038	200.17.170.37	5038	777779984	ITU-T G.711 PCMU	1302	0 (0.0%)	62.47	
200.17.170.160	5040	200.17.170.37	5040	2710719613	ITU-T G.711 PCMU	1276	0 (0.0%)	116.18	
200.17.170.160	5042	200.17.170.37	5042	1180611172	ITU-T G.711 PCMU	1310	0 (0.0%)	116.48	
200.17.170.160	5044	200.17.170.37	5044	3921300425	ITU-T G.711 PCMU	1318	0 (0.0%)	87.44	
200.17.170.160	5046	200.17.170.37	5046	1699644894	ITU-T G.711 PCMU	1324	0 (0.0%)	80.05	
200.17.170.160	5048	200.17.170.37	5048	2919278359	ITU-T G.711 PCMU	1313	0 (0.0%)	83.01	
200.17.170.160	5050	200.17.170.37	5050	3980365800	ITU-T G.711 PCMU	1331	0 (0.0%)	95.07	
200.17.170.160	5052	200.17.170.37	5052	2254914264	ITU-T G.711 PCMU	1331	0 (0.0%)	94.46	
200.17.170.160	5054	200.17.170.37	5054	2350400353	ITU-T G.711 PCMU	1311	0 (0.0%)	72.37	

Select a forward stream with left mouse button
Select a reverse stream with SHIFT + left mouse button

Unselect Find Reverse Save As Mark Packets Prepare Filter **Copy** Analyze Close

Exporta em arquivo texto

Figura 5.7: Análise de fluxos do segmento da Reitoria.

Tabela 5.6: Número de ramais e largura de banda por segmento, com 50% de acréscimo.

Segmento	Ramais	Tráfego (erlangs)	Troncos	Tráfego (kbits)
Reitoria	55	1	5	812,21
Campus II	34	0,62	4	729,84
Escola de Informática	7	0,14	2	1682,03
Prédio dos DAs	7	0,14	2	426,90
Central de Atendimento	40	0,73	4	600,50
Contabilidade	40	1,08	5	1299,56
<i>Backbone</i>	18	0,32	3	13961,94

Tabela 5.7: Resultados das medições para intensidade de tráfego atual.

Segmento	<i>Jitter</i> máx. (ms)	Perdas máx. (%)	Atraso máx. (ms)	<i>Jitter</i> méd. (ms)	Atraso méd. (ms)	Desv. Padrão <i>Jitter</i> (ms)	Desv. Padrão Atraso (ms)
Reitoria	10,73	0	67,29	2,47	59,28	0,06	6,28
Campus II	2,9	0	121,11	2,51	72,27	0,21	20,71
Escola de Informática	8,15	0	78,97	2,47	70,23	0,06	8,78
Prédio dos DAs	2,44	0	84,3	2,44	62,35	0,05	16,03
Central de Atendimento	2,69	0	99,27	2,54	70,79	0,07	14,56
Contabilidade	3,27	0	143,43	2,62	74,98	0,17	22,8
<i>Backbone</i>	14,95	0	116,48	2,62	75,17	0,19	15,71

Tabela 5.8: Resultados das medições para intensidade de tráfego de rede com 50% acréscimo.

Segmento	<i>Jitter</i> máx. (ms)	Perdas máx. (%)	Atraso máx. (ms)	<i>Jitter</i> méd. (ms)	Atraso méd. (ms)	Desv. Padrão <i>Jitter</i> (ms)	Desv. Padrão Atraso (ms)
Reitoria	9,51	0	81,44	2,46	59,87	0,07	9,96
Campus II	2,51	0	109,69	2,78	84,52	0,96	21,86
Escola de Informática	8,62	0	71,67	2,42	61,26	0,05	9,54
Prédio dos DAs	2,39	0	77,3	2,39	55,41	0,02	13,30
Central de Atendimento	2,74	0	75,06	2,53	62,22	0,09	8,5
Contabilidade	2,74	0	120,55	2,58	73,44	0,08	18,52
<i>Backbone</i>	14,44	0	132,55	2,69	79,9	0,12	22,25

5.6.1 Análise dos Resultados

No estudo de caso, os resultados das medições demonstraram a viabilidade do uso de serviços VoIP, mesmo sem a utilização de mecanismos de Qos ou CoS. Segundo (WALLINGFORD, 2005), os valores de referência para um boa qualidade de chamadas VoIP são:

- Atraso: o atraso máximo deve ficar abaixo dos 150ms em um sentido da ligação. Na prática, o valor máximo aceitável, pode ficar em torno de 400ms.
- Perdas: abaixo de 1%, com o limite de 3% de perdas de pacotes. O efeito do nível

Tabela 5.9: Resultados das medições para intensidade de tráfego e número de ramais com 50% de acréscimo.

Segmento	<i>Jitter</i> máx. (ms)	Perdas máx. (%)	Atraso máx. (ms)	<i>Jitter</i> méd. (ms)	Atraso méd. (ms)	Desv. Padrão <i>Jitter</i> (ms)	Desv. Padrão Atraso (ms)
Reitoria	10,81	0	76,23	2,49	62,68	0,06	10,13
Campus II	3,67	0	110,37	2,98	84,32	0,32	14,13
Escola de Informática	7,52	0	80,77	2,42	60,43	0,04	11,53
Prédio dos DAs	2,46	0	85,44	2,46	69,96	0,07	11,38
Central de Atendimento	2,69	0	112,35	2,53	68,43	0,08	14,07
Contabilidade	2,81	0	113,21	2,61	76,27	0,09	16,23
<i>Backbone</i>	19,16	0	216,68	2,8	92,84	0,18	35,33

de perdas vai depender do *codec* utilizado e do número de amostras por pacote. No estudo de caso, foi utilizado o *codec* G.711 com três amostras por pacote. Com duas amostras por pacote, o *codec* G.711 mantém a qualidade da chamadas com perdas de até 3% dos pacotes. Com três amostras, o nível de perdas de pacotes deve ser menor.

- *Jitter*: o *jitter* máximo recomendado é cerca de 75ms. O valor máximo pode variar de acordo com o tipo de algoritmo utilizado no equalizador (*buffer*) de *jitter* do receptor.

Nas medições realizadas, das 600 ligações, 99% das ligações ficaram dentro dos limites de referência para atraso, perdas e *jitter*.

Em relação a perdas, foi perdido apenas 1 pacote de um total de 789871 pacotes gerados em todas a ligações. Por se tratar de uma LAN, são valores considerados normais.

No caso do atraso, cerca de 1% das ligações apresentaram valores acima dos 150ms, mas o atraso nestas ligações ficou abaixo dos 220ms. De acordo com os valores de referência definidos pelo ITU na recomendação G.114, o atraso seria aceitável, mas com algum prejuízo de qualidade. Além disto, as chamadas com atrasos acima de 150ms foram originadas dos testes com 50% de acréscimo de tráfego e 50% de acréscimo de ramais. De acordo com o desvio padrão calculado, o atraso não obteve grande variação nos testes realizados.

Os valores de *jitter* máximo e médio se mantiveram bem abaixo dos 75ms em todas as ligações, sendo o valor máximo de 19,22ms. A variação do *jitter* foi bastante baixo, ficando abaixo dos 1ms.

Nas figuras 5.8, 5.9 e 5.10 estão plotados os valores de atraso máximo e *jitter* máximo para todas as ligações realizadas no diversos segmentos, variando o volume de tráfego e o número de ramais.

De acordo com o resultado das medições realizadas, a rede do estudo de caso suportaria a intensidade de tráfego telefônico ramal para ramal, atual e com um aumento de 50%, no tráfego e no número de ramais.

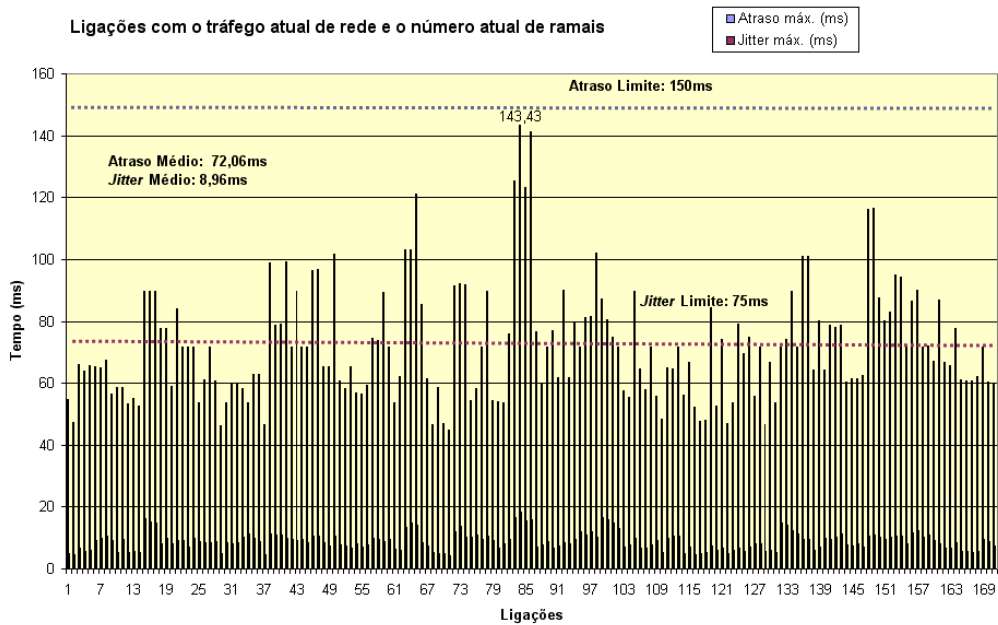


Figura 5.8: Gráfico das ligações com volume de tráfego e o número de ramais atuais.

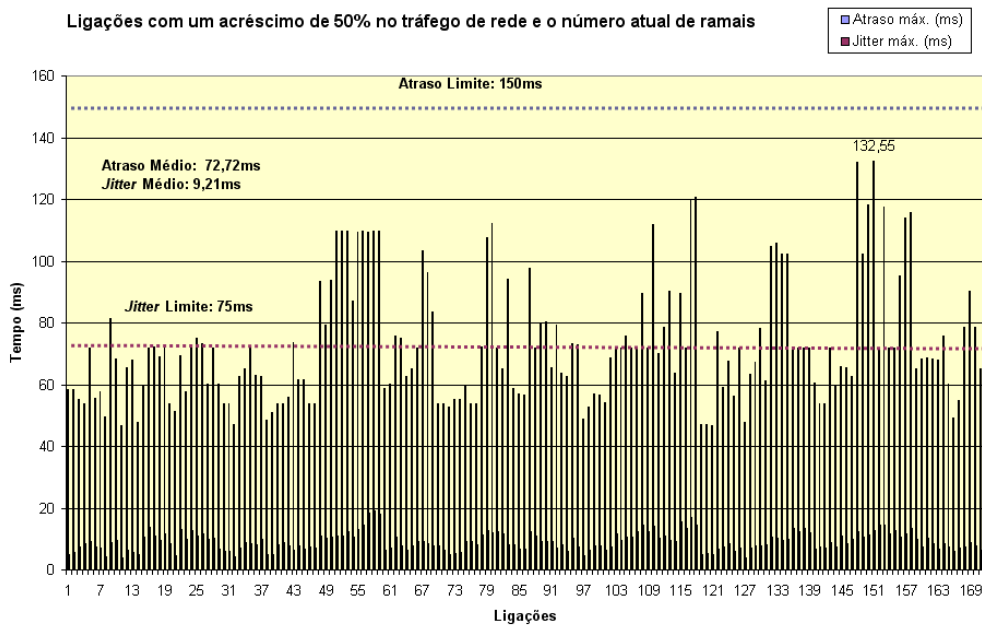


Figura 5.9: Gráfico das ligações com 50% a mais de volume de tráfego e o número de ramais atuais.

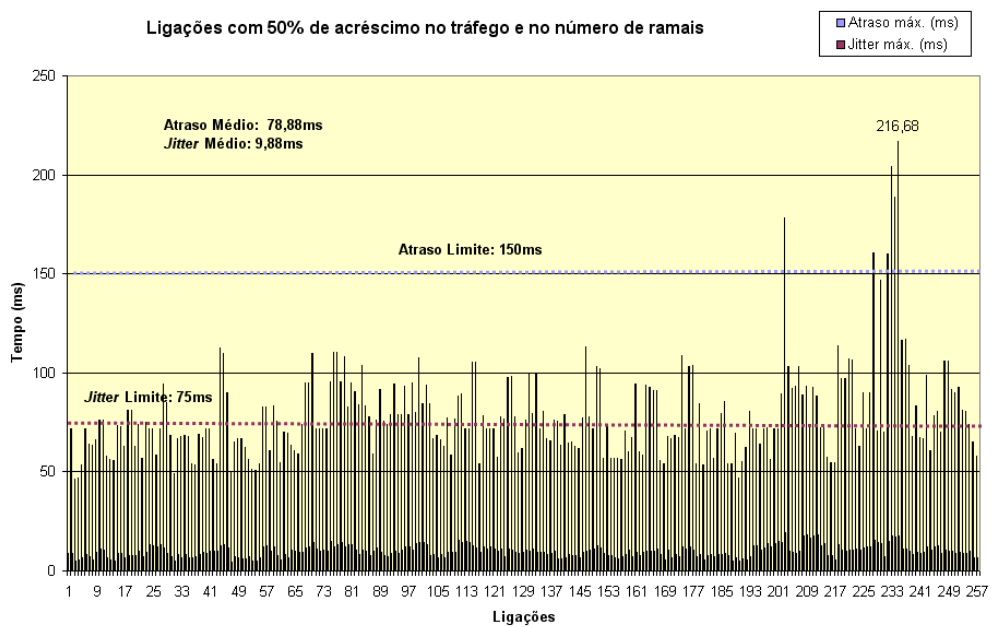


Figura 5.10: Gráfico das ligações com 50% a mais de volume de tráfego e de número de ramais.

6 CONCLUSÃO

Os serviços de VoIP estão em amplo crescimento, tanto para usuários domésticos quanto para usuários corporativos. A tendência é que haja um crescimento ainda maior nos próximos anos. Embora a popularidade dos serviços VoIP tenha se dado em serviços sem garantias de qualidade, a substituição do sistema telefônico convencional por serviços VoIP só será possível com as garantias da mesma qualidade de voz. Para isto, o planejamento de capacidade de rede desempenha papel fundamental.

O trabalho teve como objetivo principal mostrar uma metodologia para o planejamento de capacidade de rede para implantação de serviços VoIP. Foram realizados estudos sobre o planejamento de capacidade de redes telefônicas convencionais e a possibilidade de adaptação destes métodos para os serviços VoIP.

Como objetivos secundários, os fundamentos de VoIP foram abordados, principalmente o protocolo SIP, resultando em uma referência para aproveitamento em estudos futuros. Os quesitos de QoS utilizados para a implementação de serviços VoIP em redes corporativas foram apresentados e discutidos.

O estudo de caso possibilitou a aplicação da metodologia em um ambiente real. Foram utilizadas ferramentas gratuitas e de código-fonte aberto para a geração de chamadas e de tráfego de rede e análise dos fluxos de áudio. Os resultados demonstram a possibilidade de utilizar serviços de VoIP, com boa qualidade, mesmo em redes corporativas sem mecanismos de QoS ou CoS. Entretanto, para garantir a qualidade oferecida pela telefonia convencional, é indispensável a utilização de mecanismos de QoS. Em redes corporativas, podem ser utilizados mecanismos de CoS, tais como *DiffServ* e IEEE 802.1p, em *switches* e roteadores para a priorização de tráfego. Em casos de largura de banda excedente, como em LANs comutadas, os mecanismos de CoS podem garantir a qualidade das chamadas. Em enlaces ponto a ponto, podem ser utilizados mecanismos de reserva de banda, tal como RVSP, ou de redução de largura de banda, tal como a utilização de compressão de cabeçalhos.

A aplicação, no estudo de caso, da metodologia de planejamento de capacidade de rede desenvolvida demonstra a sua utilidade como uma ferramenta na implantação de serviços VoIP em redes corporativas.

Como trabalhos futuros, sugere-se:

- Aplicação da metodologia elaborada neste trabalho em outros ambientes de rede;
- Aplicação e medição dos mecanismos de garantia de QoS e CoS na rede com serviços VoIP;
- Elaboração de uma metodologia para planejamento de capacidade em redes mistas, como o uso de serviços VoIP e telefonia convencional através de *gateways*;

- Adaptação das ferramentas utilizadas para o cálculo do fator de qualidade R do Modelo E;
- Aplicação da metodologia desenvolvida neste trabalho e a implantação do serviços VoIP em um ambiente com usuários.

REFERÊNCIAS

3COM. Disponível em: <<http://www.3com.com/>>. Acesso em: set. 2005.

ANATEL. **Resolução nº 30, de 29 de junho de 1998.** Disponível em: <http://www.anatel.gov.br/index.asp?link=/biblioteca/resolucao/1998/res_030_1998.PDF>. Acesso em: ago. 2005.

ANATEL. **Anexo à Resolução nº 272, de 9 de agosto de 2001 - Regulamento do Serviço de Comunicação Multimídia.** Disponível em: <http://www.anatel.gov.br/index.asp?link=/biblioteca/resolucao/2001/res_272_2001.pdf?Cod=1944>. Acesso em: ago. 2005.

ANATEL. **Plano Geral de Metas de Qualidade (PGMQ) - janeiro a dezembro de 2005.** Disponível em: <<http://sistemas.anatel.gov.br/saci/Relatorios/PgmqConsolidado/Tela.asp?acao=Confirmar&codTipoConsolidado=2&AnoInicial=2005>>. Acesso em: ago. 2005.

ANGUS, I. An Introduction to Erlang B and Erlang C. **Telemanagement Journal**, [S.l.], n.187, June 2002. Disponível em :<<http://www.angustel.ca/>>. Acesso em: ago. 2005.

AT&T. **AT&T - History of Network Switching.** Disponível em: <<http://www.att.com/history/nethistory/switching.html>>. Acesso em: set. 2005.

AUDACITY: free audio editor and recorder. Disponível em: <<http://audacity.sourceforge.net/>>. Acesso em: set. 2005.

AVAYA. **Avaya - IP Voice Quality Network Requirements - Version 2.0.** Disponível em: <<http://www1.avaya.com/enterprise/whitepapers/lb1894.pdf>>. Acesso em: ago. 2005.

BARAN, P. On Distributed Communications Networks. In: CONGRESS OF THE INFORMATION SYSTEMS SCIENCES, 1., 1962. **Proceedings...** [S.l.: s.n.], 1962.

BASET, S. A.; SCHULZRINNE, H. **An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol.** New York - USA: Department of Computer Science, Columbia University, 2004.

BERNERS-LEE, T.; FIELDING, R.; MASINTER, L. **Uniform Resource Identifiers (URI): generic syntax: RFC 2396.** [S.l.]: Internet Engineering Task Force, Network Working Group, 1998.

BIYANI, P. et al. Early Estimation of Voice over IP quality. In: NORDUNET NETWORK CONFERENCE, 21., 2003. **Proceedings...** [S.l.: s.n.], 2003.

BLAKE, S.; BLACK, D.; CARLSON, M.; DAVIES, E.; WANG, Z.; WEISS, W. **An Architecture for Differentiated Services: RFC 2475.** [S.l.]: Internet Engineering Task Force, Network Working Group, 1998.

BRANDL, M. et al. **IP Telephony Cookbook - TERENA Report.** Amsterdã - Holanda: Trans-European Research and Education Networking Association (TERENA), 2004.

BROCKMEYER, E.; HALSTRØM, H.; JENSEN, A. The Life and Works of A.K. Erlang. **Transactions of the Danish Academy of Technical Sciences**, [S.l.], n.2, 1948. Disponível em :<<http://oldwww.com.dtu.dk/teletraffic/erlangbook/pp001-278.pdf>>. Acesso em: nov. 2005.

CASNER, S.; JACOBSON, V. **Compressing IP/UDP/RTP Headers for Low-Speed Serial Links: RFC 2508.** [S.l.]: Internet Engineering Task Force, Network Working Group, 1999.

CASTRO, M. C. F. de. **Planejamento de Redes Comutadas:** notas de aula. Disponível em: <http://www.ee.pucrs.br/decastro/pdf/Redes_Comutadas_Cap4.pdf>. Acesso em: maio 2005.

CHONG, H. M.; MATTHEWS, H. S. Comparative Analysis of Traditional Telephone and Voice-over-Internet Protocol (VoIP) Systems. **IEEE International Symposium on Electronics and the Environment**, [S.l.], p.106–111, May 2004.

CISCO. **Traffic Analysis for Voice over IP.** Disponível em: <http://www.cisco.com/univercd/cc/td/doc/cisintwk/intsolns/voipsol/ta_isd.pdf>. Acesso em: maio 2005.

CISCO. **Cisco - Understanding Delay in Packet Voice Networks.** Disponível em: <<http://www.cisco.com/warp/public/788/voip/delay-details.pdf>>. Acesso em: ago. 2005.

COFFMAN, J. E. Not Your Father's PBX. **ACM Queue**, [S.l.], Sept. 2004.

COHEN, D. **Specifications for the Network Voice Protocol (NVP): RFC 471.** [S.l.]: Internet Engineering Task Force, Network Working Group, 1977.

COLE, R. G.; ROSENBLUTH, J. H. Voice over IP performance monitoring. **SIGCOMM Comput. Commun. Rev.**, New York, NY, USA, v.31, n.2, p.9–24, 2001.

COLLINS, D. **Carrier Grade Voice over IP.** 2nd.ed. [S.l.]: McGraw-Hill, 2003.

CONWAY, A. E. A Performance Monitoring System for VoIP Gateways. In: WORKSHOP ON SOFTWARE AND PERFORMANCE, 2., 2000. **Proceedings...** [S.l.: s.n.], 2000. p.38–43.

COUNTERPATH. **Counterpath.** Disponível em: <<http://www.xten.net/>>. Acesso em: out. 2005.

CYGWIN. Disponível em: <<http://www.cygwin.com/>>. Acesso em: out. 2005.

DAVIDSON, J.; JAMES, P. **Voice over IP Fundamentals.** [S.l.]: Cisco Press, 2000.

DIGIUM. **Asterisk - the Open Source PBX**. Disponível em: <<http://www.asterisk.org>>. Acesso em: ago. 2005.

DRAPER, J. T. **Cap'n Crunch in Cyberspace**. Disponível em: <<http://www.webcrunchers.com/crunch/>>. Acesso em: ago. 2005.

ELAOU, M.; FAMOLARI, D.; GHOSH, A. Experimental VoIP Capacity Measurements for 802.11b WLANs. In: IEEE CONSUMER COMMUNICATIONS & NETWORKING CONFERENCE, 2005. **Proceedings...** [S.l.: s.n.], 2005.

ENUM.ORG. **ENUM - FAQ**. Disponível em: <<http://www.enum.org/information/faq.cfm>>. Acesso em: out. 2005.

ERLANG.COM. **Erlang Calculator**. Disponível em: <<http://www.erlang.com>>. Acesso em: out. 2005.

ERLANG.CO.UK. **Erlang for Excel**. Disponível em: <<http://www.erlang.co.uk/excel.htm>>. Acesso em: maio 2005.

ETHERREAL: a network protocol analyzer. Disponível em: <<http://www.ethereal.com/>>. Acesso em: out. 2005.

ETHERREAL: wiki. Disponível em: <<http://wiki.ethereal.com/Ethernet/>>. Acesso em: set. 2005.

FALTSTROM, P.; MEALLING, M. **The E.164 to Uniform Resource Identifiers (URI) Dynamic Delegation Discovery System (DDDS) Application (ENUM)**: RFC 3761. [S.l.]: Internet Engineering Task Force, Network Working Group, 2004.

FARLEY, T. **Tom Farley's Telephone History Series**. Disponível em: <<http://www.privateline.com/TelephoneHistory/History1.htm>>. Acesso em: ago. 2005.

FWD: free world dialup. Disponível em: <<http://www.pulver.com/fwd/>>. Acesso em: maio 2005.

GIZMO Project. Disponível em: <<http://www.gizmoproject.com>>. Acesso em: maio 2005.

GONCALVES, A. R. **Método para Planejamento de Capacidade de Redes ATM baseado em Simulação**. 2000. Dissertação (Mestrado em Ciência da Computação) — Programa de Pós-Graduação em Computação, UFRGS, Porto Alegre, Brasil.

GOODE, B. Voice Over Internet Protocol (VoIP). **Proceedings of the IEEE**, [S.l.], v.90, n.9, Sept. 2002.

HANDLEY, M.; JACOBSON, V. **SDP: session description protocol**: RFC 2327. [S.l.]: Internet Engineering Task Force, Network Working Group, 1998.

HARDY, W. C. **VoIP Service Quality - Measuring and Evaluating Packet-Switched Voice**. [S.l.]: McGraw-Hill, 2003.

HP. **VoIP Solutions Cookbook Configuring a ProCurve Voice-over-IP Solution**. Disponível em: <http://www.hp.com/rnd/pdfs/VoIP_Cookbook.pdf>. Acesso em: jan. 2005.

HUSTON, G. ENUM - Mapping the E.164 Number Space into the DNS. **The Internet Protocol Journal**, [S.l.], v.5, n.2, June 2002. Disponível em :<<http://www.cisco.com/application/pdf/en/us/guest/about/about/c644/ccmigration09186a008020eba4.pdf>>. Acesso em: ago. 2005.

IANA. **Enumservice Registrations**. Disponível em: <<ftp://ftp.iana.org/assignments/enum-services>>. Acesso em: out. 2005.

ICT. **ICT (Information and Communication Theory Group) - The Power of the Erlang Formula**. Disponível em: <http://ict.ewi.tudelft.nl/index.php?option=com_sections&id=164&Itemid=286>. Acesso em: maio 2005.

IEEE. **IEEE Std 802.3af - Part 3: carrier sense multiple access with collision detection (csma/cd) access method and physical layer specifications**. [S.l.: s.n.], 2003.

ILBCFREEWARE.ORG. Disponível em: <<http://www.ilbcfreeware.org/>>. Acesso em: out. 2005.

INTEL. **Differentiated Services - Moving towards Quality of Service on the Ethernet**. Disponível em: <http://www.intel.com/network/connectivity/resources/doc_library/white_papers/solutions/diff_serv/diffserv.pdf>. Acesso em: jan. 2005.

INTERNET2. **Projeto Internet2**. Disponível em: <<http://www.internet2.edu/>>. Acesso em: jul. 2005.

ITU-T. **ITU-T Recommendation H.323: packetd-based multimedia communications systems**. [S.l.: s.n.], 1997.

ITU-T. **ITU-T Recommendation G.107. The E-Model, a computational model for use in transmission planning**. [S.l.: s.n.], 2003.

ITU-T. **ITU-T Recommendation E.164: the international public telecommunication numbering plan**. [S.l.: s.n.], 2005.

IVERSEN, V. B. et al. **Teletraffic Engineering Handbook**. Genebra - Suíça: ITU (International Telecommunication Union), ITC (International Teletraffic Congress), 2005.

JI, W.; SCHULZRINNE, H. Assessment of VoIP Service Availability in the Current Internet. In: **PASSIVE AND ACTIVE MEASUREMENT WORKSHOP, 2003. Proceedings...** [S.l.: s.n.], 2003. Disponível em: <<http://moat.nlanr.net/PAM2003/PAM2003papers/3897.pdf>>. Acesso em: maio 2005.

JIANG, W.; KOGUCHI, K.; SCHULZRINNE, H. QoS Evaluation of VoIP End-points. In: **IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS, 2003. Proceedings...** [S.l.: s.n.], 2003. p.1917–1921.

JIANG, W.; SCHULZRINNE, H. Analysis of on-off patterns in VoIP and their effect on voice traffic aggregation. In: **IEEE INT. CONF. COMPUTER COMMUNICATION NETWORKS, 2000. Proceedings...** [S.l.: s.n.], 2000. p.82–87.

KAZAA. **Kazaa File Sharing**. Disponível em: <<http://www.kazaa.com>>. Acesso em: ago. 2005.

LUSTOSA, L. C. G. **Arquitetura de Monitoração de Qualidade de Chamadas Telefônicas IP**. 2005. Dissertação (Mestrado em Ciência da Computação) — UFRJ, Instituto de Matemática, Núcleo de Computação Eletrônica, Rio de Janeiro, Brasil.

LUSTOSA, L. C. G. et al. Utilização do Modelo E para avaliação da qualidade da fala em sistemas de comunicação baseados em voz sobre IP. In: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES, SBRC, 22., 2004, Gramado: II/UFRGS. **Anais...** [S.l.: s.n.], 2004.

MARKOPOULOU, A. P.; TOBAGI, F. A.; KARAM, M. J. Assessing the Quality of Voice Communications Over Internet Backbones. **IEEE/ACM Transactions on Networking**, [S.l.], v.11, n.5, p.747–759, Oct 2003.

MESSENGER. **Página do serviço Microsoft Messenger**. Disponível em: <<http://messenger.msn.com>>. Acesso em: set. 2005.

MILLENNIUM MATHEMATICS PROJECT, B. from. **Agner Krarup Erlang (1878 - 1929)**. Disponível em: <<http://pass.maths.org.uk/issue2/erlang/>>. Acesso em: out. 2005.

MINOLI, D.; MINOLI, E. **Delivering Voice over IP Networks**. [S.l.]: John Wiley Sons, 1998.

MORTON, D. The Electrical Century - A Snapshot of Telephony at the Turn of the Century. **Proceedings of the IEEE**, [S.l.], v.87, n.4, Apr. 1999.

MRTG: the multi router traffic grapher. Disponível em: <<http://people.ee.ethz.ch/~oetiker/webtools/mrtg/>>. Acesso em: out. 2005.

NET2PHONE. Disponível em: <<http://www.net2phone.com/>>. Acesso em: jul. 2005.

NETFILTER. Disponível em: <<http://www.netfilter.org/>>. Acesso em: out. 2005.

NICHOLS, K. et al. **Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers: RFC 2474**. [S.l.]: Internet Engineering Task Force, Network Working Group, 1998.

NLANR/DAST : iperf 1.7.0 - the tcp/udp bandwidth measurement tool. Disponível em: <<http://dast.nlanr.net/Projects/Iperf/>>. Acesso em: set. 2005.

ODLYZKO, A. **The history of communications and its implications for the Internet**. [S.l.]: AT&T Labs - Research, 2000. Disponível em: <<http://www.dtc.umn.edu/~odlyzko/doc/history.communications0.pdf>>. Acesso em: maio 2005.

OPENH323. Disponível em: <<http://www.openh323.org>>. Acesso em: jul. 2005.

PROJETO Fone@RNP. Disponível em: <<http://www.voip.nce.ufrj.br/foneatrn/>>. Acesso em: maio 2005.

RACHID, E. M. **Cancelamento de Eco em Telefonia IP**. 2004. Dissertação (Mestrado em Ciência da Computação) — DECOM/Pós-Graduação em Engenharia Elétrica, UNICAMP, Campinas, Brasil.

- ROSENBERG, J. et al. **SIP**: session initiation protocol: RFC 3261. [S.l.]: Internet Engineering Task Force, Network Working Group, 2002.
- ROSENBERG, J.; SCHULZRINNE, B. **SIP**: session initiation protocol v.2.0: RFC 3261. [S.l.]: Internet Engineering Task Force, Network Working Group, 2002.
- ROSENBERG, J.; SCHULZRINNE, H. **An Offer/Answer Model with the Session Description Protocol**: RFC 3264. [S.l.]: Internet Engineering Task Force, Network Working Group, 2002.
- SCHULZRINNE, H. et al. **RTP**: a transport protocol for real-time applications: RFC 1889. [S.l.]: Internet Engineering Task Force, Network Working Group, 1996.
- SIP.EDU: voice over ip workgroup. Disponível em: <<http://voip.internet2.edu/SIP.edu/>>. Acesso em: jul. 2005.
- SIPPHONE. **University Peering Program**. Disponível em: <<http://www.sipphone.com/university/>>. Acesso em: out. 2005.
- SIPPHONE. Disponível em: <<http://www.sipphone.com>>. Acesso em: maio 2005.
- SKYPE. Disponível em: <<http://www.skype.com>>. Acesso em: mai. 2005.
- SOARES, L. F. G.; LEMOS, G.; COLCHER, S. **Redes de Computadores - Das LANs, MANs e WANs às Redes ATM**. [S.l.]: Campus, 1995.
- SPEEX.ORG. **Speex a free codec for free speech**. Disponível em: <<http://www.speex.org/>>. Acesso em: out. 2005.
- SPURGEON, C. E. **Ethernet - O Guia Definitivo**. [S.l.]: Campus, 2000.
- STALLINGS, W. The Session Initiation Protocol. **The Internet Protocol Journal**, [S.l.], v.6, n.1, Mar. 2003. Disponível em :<http://www.cisco.com/application/pdf/en/us/guest/about/about/c644/ccmigration_09186a008044af45.pdf>. Acesso em: ago. 2005.
- STI. Disponível em: <<http://www.informatec-sp.com/>>. Acesso em: out. 2005.
- TAKAHASHI, A.; YOSHINO, H.; KITAWAKI, N. Perceptual QoS Assessment Technologies for VoIP. **IEEE Communications Magazine**, [S.l.], July 2004.
- TANENBAUM, A. S. **Redes de Computadores**. 4.ed. [S.l.]: Campus, 2003.
- TCPDUMP. **TCPDUMP public repository**. Disponível em: <<http://www.tcpdump.org/>>. Acesso em: set. 2005.
- VARSHNEY, U. et al. Voice over IP. **ACM Communications**, [S.l.], v.45, n.1, Jan. 2002.
- WALLINGFORD, T. **Switching to VoIP**. [S.l.]: O'Reilly, 2005.

APÊNDICE A GRÁFICOS POR SEGMENTO

Gráficos por segmento, dos resultados das medições com 50% de acréscimo no volume de tráfego e 50% de acréscimo no número de ramais.

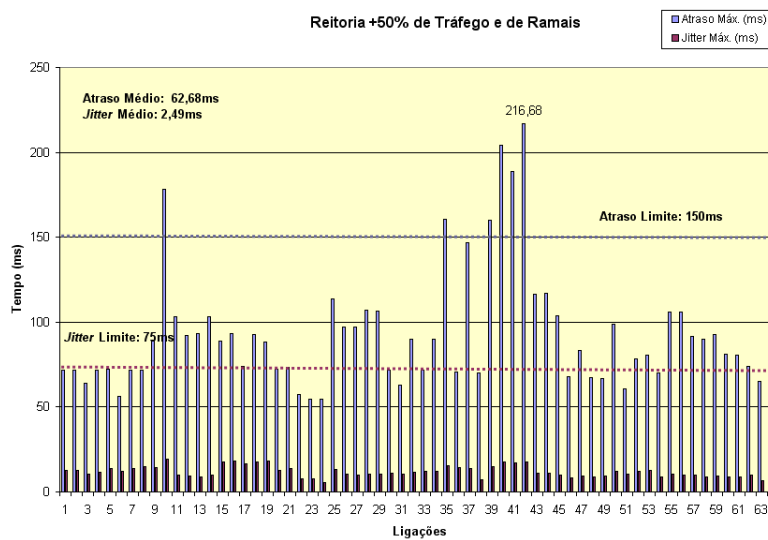


Figura A.1: Gráfico das ligações do segmento da Reitoria, com 50% a mais de tráfego e de ramais.

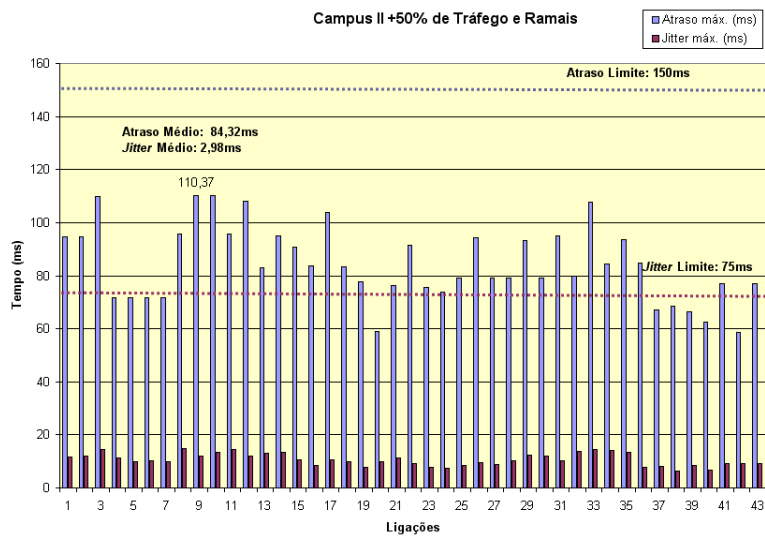


Figura A.2: Gráfico das ligações do segmento do Campus II, com 50% a mais de tráfego e de ramais.

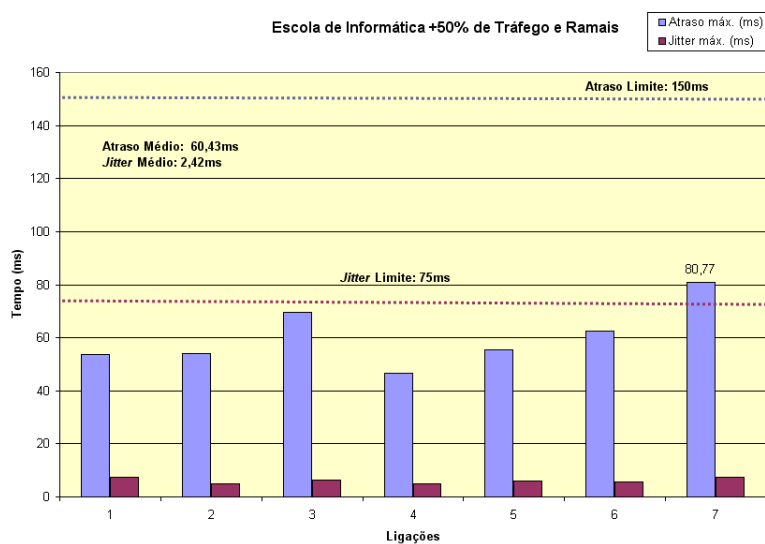


Figura A.3: Gráfico das ligações do segmento da Escola de Informática, com 50% a mais de tráfego e de ramais.

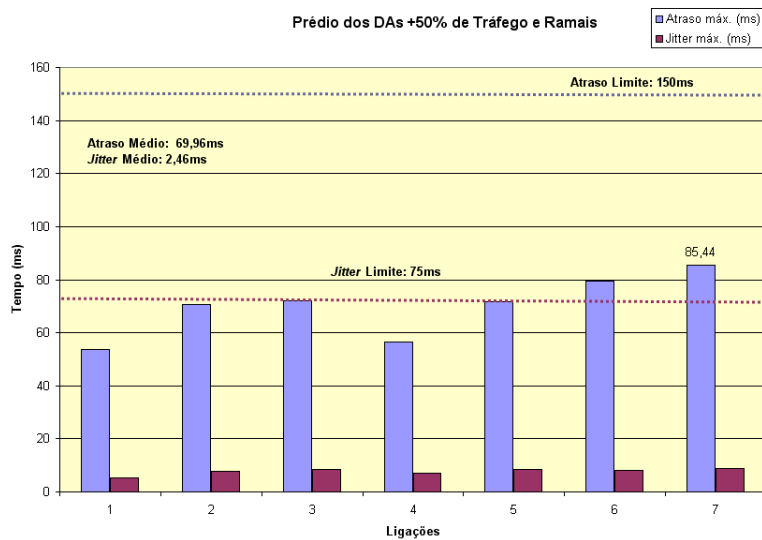


Figura A.4: Gráfico das ligações do segmento do Prédio dos DAs, com 50% a mais de tráfego e de ramais.

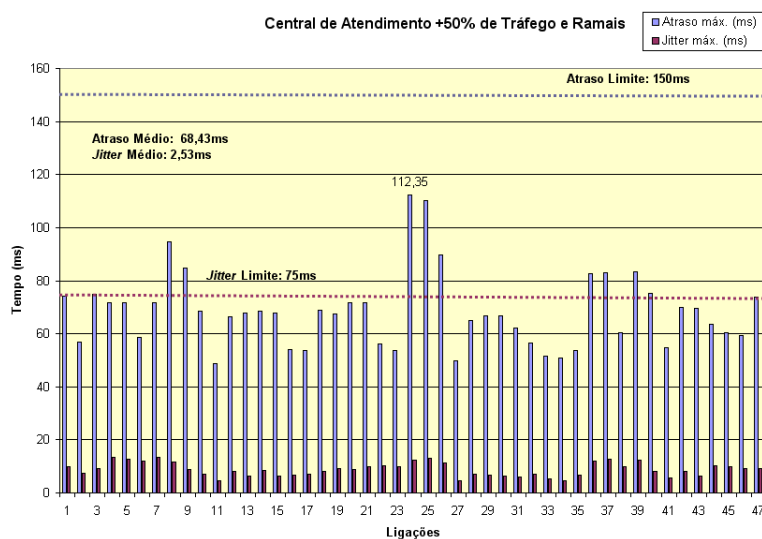


Figura A.5: Gráfico das ligações do segmento da Central de Atendimento, com 50% a mais de tráfego e de ramais.

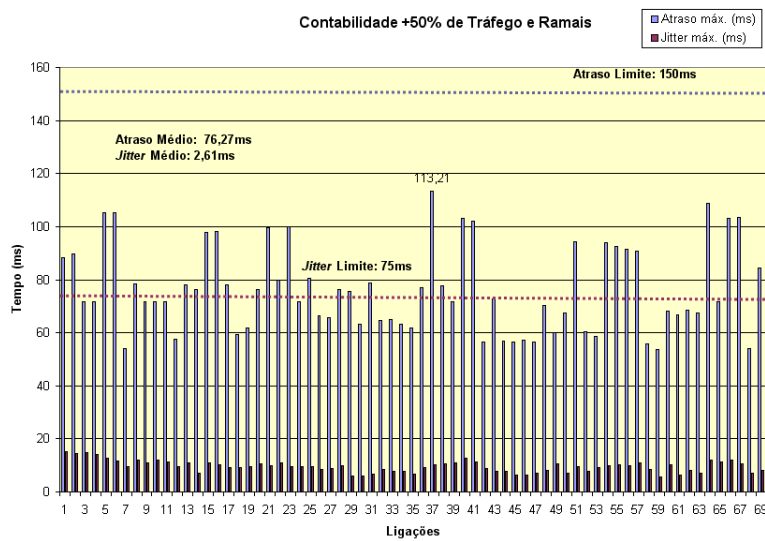


Figura A.6: Gráfico das ligações do segmento da Contabilidade, com 50% a mais de tráfego e de ramais.

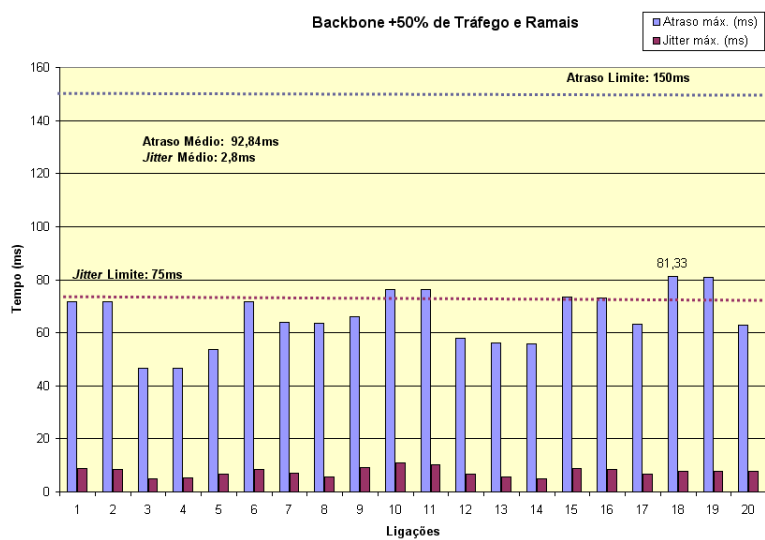


Figura A.7: Gráfico das ligações do segmento do Backbone, com 50% a mais de tráfego e de ramais.