

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

SOLUÇÕES DE VIDEOCONFERÊNCIA H.323 E SIP

por

CLEBER MACHADO ORTIZ
TI – CPGCC - UFRGS

Trabalho Individual I

Profa. Dra. Liane Margarida Rockenbach Tarouco
Orientadora

Porto Alegre, Maio de 2002.

Sumário

LISTA DE ABREVIATURAS	03
LISTA DE FIGURAS	04
RESUMO	05
ABSTRACT	06
1. INTRODUÇÃO	07
2. VIDEOCONFERÊNCIA	
2.1 SISTEMAS DE VIDEOCONFERÊNCIA	09
2.1.1 Ponto-a-ponto	09
2.1.2 Modelo centralizado.....	09
2.1.3 Modelo Descentralizado	11
2.1.4 Modelo Híbrido.....	12
2.2 SERVIÇOS AUXILIARES	13
2.3 CODIFICAÇÃO/COMPRESSÃO DE ÁUDIO/VÍDEO	14
2.3.1 Áudio	14
2.3.2 Vídeo	16
2.4 PROTOCOLO RTP E RTCP	19
2.5 ALGUNS PROBLEMAS ENFRENTADOS NA VIDEOCONFERÊNCIA	21
3 SOLUÇÕES PARA VIDEOCONFERÊNCIA	24
3.1 H.323	24
3.1.1 Componentes da arquitetura	24
3.1.2 Sinalização no H.323	29
3.2 SIP (<i>SESSION INITIATION PROTOCOL</i>).....	30
3.2.1 Componentes da arquitetura	31
3.2.2 Sinalização no SIP.....	34
3.3 RELAÇÃO ENTRE AS SOLUÇÕES H.323/SIP	39
3.3.1 Estabelecimento da chamada	40
3.3.2 Conversão de endereços	41
3.4 COMPARAÇÃO ENTRE AS PROPOSTAS	41
3.4.1 Funcionalidades que o SIP faz e o H.323 não realiza.....	41
3.4.2 Funcionalidades que o H.323 faz e o SIP não realiza.....	42
CONCLUSÃO	44
BIBLIOGRAFIA	45

Lista de Abreviaturas

A/D	Analógico/Digital
ACF	Admission Confirm
ARJ	Admission Reject
ARQ	Admission Request
B-ISDN	Broadband ISDN
CIF	Common Intermediate Format
Codec	Coder/Decoder
Cseq	Campo de cabeçalho
D/A	Digital/Analógico
DCT	Discrete Cosine Transform
DNS	Domain Name System
IETF	Internet Engineering Task Force
IETF	Internet Engineering Task Force
ITU-T	International Telecommunication Union - Telecommunication
Kbps	Kilobits per second
LCF	Location Confirm
LRJ	Location Reject
LRQ	Location Request
MBONE	Multicast Backbone
MC	Controlador Multiponto
MCU	Multipoint Control Units
MP	Processador Multiponto
NTSC	National Television Standards Committee
OSI	Open System Interconnection
PAL	Phase Alteration Line
PSTN	Public Switched Telephone Network
QCIF	Quarter Common Intermediate Format
QOS	Quality of Service
RAS	Requisição, Admissão e Status
REMAV	Redes Metropolitanas de Alta Velocidade
RFC	Request For Comment
RTCP	RTP Control Protocol
RTP	Real Time Transport Protocol
SDP	Session Description Protocol
SIP	Session Initiation Protocol
TCP	Transmission Control Protocol
UAC	User Agent Client
UAS	User Agent Server
UDP	User Datagram Protocol
UDP	User Datagram Protocol
WWW	World Wide Web

Lista de Figuras

FIGURA 2.1 – Exemplo de videoconferência entre dois participantes (não é necessária presença de um MCU)	09
FIGURA 2.2 – Videoconferência baseada em modelo centralizado entre 3 ou mais participantes.....	10
FIGURA 2.3 - Exemplo de videoconferência baseada em modelo descentralizado.....	11
FIGURA 2.4 - Exemplo de videoconferência baseada em modelo híbrido.....	12
FIGURA 2.5 – Transmissão de voz na Internet utilizando <i>codecs</i>	15
FIGURA 2.6 – Transmissão de vídeo em redes digitais.....	17
FIGURA 2.7 - RTP combinado com um formato de <i>payload</i> para formar um protocolo completo.....	19
FIGURA 2.8 – <i>Translator</i> processando pacotes RTP	20
FIGURA 2.9 – Operação de um <i>mixer</i>	20
FIGURA 2.10 – Pacote RTP e RTCP enviados pela rede	21
FIGURA 2.11 – Espaçamento nas mensagens que chegam no destino.....	22
FIGURA 2.12 – Relação entre latência e <i>jitter</i>	23
FIGURA 2.13 – O problema do <i>skew</i> entre áudio e vídeo.....	23
FIGURA 3.1 – Componentes da arquitetura H.323	25
FIGURA 3.2 – Terminal H.323.....	26
FIGURA 3.3 – Gateway H.323/H.320.....	27
FIGURA 3.4 – Zona H.323	28
FIGURA 3.5 - Sinalização no H.323	30
FIGURA 3.6 - Componentes da arquitetura SIP	32
FIGURA 3.7 - Servidor de <i>proxy</i>	33
FIGURA 3.8 - Chamada para um endereço IP conhecido com SIP	36
FIGURA 3.9 - Negociando o estabelecimento da chamada.....	38
FIGURA 3.10 - Pilha de protocolos	39

Resumo

Não resta dúvidas que a Internet se tornou o meio de comunicação universal, trazendo inúmeros benefícios para os usuários e tornando a comunicação sobre IP a forma predominante para a transmissão de dados e informação multimídia. É desta forma, que pode-se destacar também a videoconferência como sendo um dos fatores que impulsionam a evolução tecnológica.

Videoconferência consiste em uma reunião entre duas pessoas ou grupos, simultaneamente em locais distintos, mediante o uso de câmeras apropriadas para captura e processamento da imagem e monitores de vídeo para a visão do interlocutor. A grande vantagem desta solução é o fato de pessoas, independente de suas localizações geográficas, poderem por exemplo, compartilhar recursos de áudio e vídeo simultaneamente além de materiais de trabalho.

No âmbito dos serviços de videoconferências, existem algumas soluções propostas como o padrão H.323 do ITU-T e o *Session Initiation Protocol* (SIP) da IETF. O objetivo deste trabalho é dar uma visão geral sobre essas duas soluções, comentando serviços que são fornecidos pelas duas sobre forma de uma comparação, de modo que tendo um maior conhecimento sobre os dois padrões, pode-se buscar que cada vez mais essas duas soluções se interoperem.

Palavras chave: videoconferência, H.323, SIP, colaboração visual.

Abstract

Without that the Internet became the middle of universal communication, bringing countless benefits for the users and turning the communication on IP the predominant form for the transmission of data and information multimedia. It is this way, that can also stand out the videoconference as being one of the factors that impel the technological evolution.

Videoconference consists of a meeting between two people or groups, simultaneously in local different, by means of the use of cameras adapted for capture and processing of the image and video monitors for the speaker's vision. The great advantage of this solution is the people's fact, independent of its geographical locations, they cannot for example, to share audio resources and video simultaneously besides work materials.

In the ambit of the videoconferences services, some exist solutions proposals like the pattern H.323 of ITU-T and Session Initiation Protocol (SIP) of IETF. The objective of this work is to give a general vision about those two solutions, commenting services that are supplied by two about on form of a comparison, so that i tend a larger knowledge on the two patterns, it can be looked for that more and more those two solutions to relate.

Key words: videoconference, H.323, SIP, visual collaboration

1 Introdução

Hoje em dia, a existência de novas alternativas para as redes de computadores, tais como ReMAVs (Redes Metropolitanas de Alta Velocidade), Internet2 e RNP2 estão viabilizando o desenvolvimento de aplicações avançadas como videoconferência, vídeo interativo, bibliotecas digitais e laboratórios virtuais. Este fato em colaborado em grande parte para o favorecido da comunidade acadêmica e instituições de pesquisa, além do setor comercial; e tem provocado um crescimento no uso do computador como uma ferramenta para mediar a comunicação em tempo real entre indivíduos e grupos, aumentando o número de ferramentas disponíveis e também o número de usuários destes serviços [LEO 02A].

Sobre videoconferência, podemos considerar como sendo uma forma de comunicação interativa que permite que a duas ou mais pessoas que estejam em locais diferentes, à comunicação com áudio e visualização de imagem em tempo real. Reuniões, cursos, conferências, debates, palestras são conduzidas como se todos os participantes estivessem juntos no mesmo local. Com os recursos da videoconferência, pode-se conversar com os participantes e, ao mesmo tempo visualizá-los na tela de um monitor (telão ou televisão, dependendo dos recursos utilizados), trocando informações como se fosse pessoalmente.

Segundo [ZAN 99], o compartilhamento de recursos com comunidades distantes, a realização de experiências virtuais quando as reais não são possíveis, a possibilidade de trazer aos alunos as opiniões de importantes especialistas através de palestras, à aplicação de atividades conjuntas como debates e exercícios em grupo, a perspectiva de trazer uma variedade de formas novas de aprendizagem com diversas mídias, são apenas algumas vantagens trazidas pelo uso da videoconferência no ensino à distância.

Atualmente uma grande variedade de soluções para sistemas de videoconferência está disponível e cada aplicação, de acordo com o seu propósito, pode ter necessidades diferentes com relação a equipamentos, à infra-estrutura de comunicação e à qualidade de serviço. Considerando estes fatores, um sistema de videoconferência deve se adequar da melhor forma possível aos recursos que a infra-estrutura de rede oferece [LEO 02B].

Uma grande vantagem das soluções e recursos da videoconferência é o fato de podermos conversar com pessoas geograficamente distantes, através de um computador pessoal ou mesmo em salas de reuniões, eliminando com isso despesas com viagens por exemplo. Baseado neste e em outros exemplos, muitas empresas começam a buscar soluções que atinjam um número maior de produtos e infraestruturas, investindo mais em soluções “padronizadas” e deixando de lado um pouco as “proprietárias”.

O objetivo deste trabalho é apresentar alguns conceitos de videoconferência como: tipos, serviços, codificação, protocolos de comunicação e problemas enfrentados através de um estudo bibliográfico obtido através de recomendações, artigos e livros de pessoas que participam do desenvolvimento da videoconferência. Também comentar as soluções existentes e como é o mecanismo de controle no H.323.

O texto está estruturado da seguinte forma:

Na sessão 2, é apresentado os tipos de videoconferências e seus serviços auxiliares. Também é comentada algumas das técnicas de codificação e compressão de áudio/vídeo, além dos protocolos utilizados e os maiores problemas encontrados na videoconferência.

A sessão 3, mostra as soluções para videoconferência, abordando a relação existente entre as duas soluções, com isso, busca-se uma maior compatibilidade entre as novas tecnologias que estão surgindo para interoperarem melhor.

Por fim, o último capítulo busca apresentar as conclusões obtidas.

2. Videoconferência

Os sistemas de videoconferência permitem que se trabalhe de forma cooperativa e se compartilhe informações e materiais de trabalho sem a necessidade de locomoção geográfica. Através da videoconferência, são criadas novas possibilidades para reuniões de pessoas distantes umas das outras, provendo diversos meios bem mais interessantes pela comunicação visual e a interação entre os demais participantes, superando desta forma o conhecido e-mail, os diversos canais de bate-papo e até mesmo o telefone.

2.1 Sistemas de videoconferência

Segundo [LEO 02A], existem diversos sistemas disponíveis para videoconferência, classificados segundo a forma de comunicação que utilizam:

2.1.1 Ponto-a-ponto: quando existe a comunicação entre duas pessoas ou mais, conforme figura 2.1:

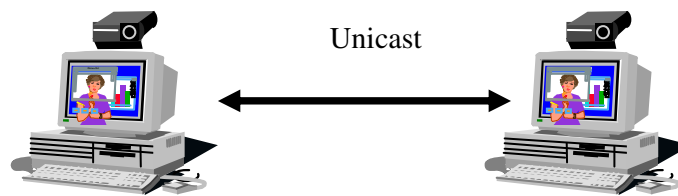


Figura 2.1 – Exemplo de videoconferência entre dois participantes (não é necessária presença de um MCU)

2.1.2 Modelo centralizado:

Este modelo é baseado no modo de comunicação ponto-a-ponto ou *unicast*. No caso de existirem três ou mais pontos para conectarem entre si, a comunicação é possível desde que haja um MCU ou Unidade de Controle Multiponto. Nas sessões de videoconferência baseadas neste modelo cada participante estabelece uma conexão com o MCU central e a distribuição do fluxo de áudio, vídeo e dados para cada participante são feitos pelo MCU que mescla os vários fluxos de áudio, seleciona o fluxo de vídeo correspondente e retransmite o resultado para todos os outros participantes, conforme figura 2.2:

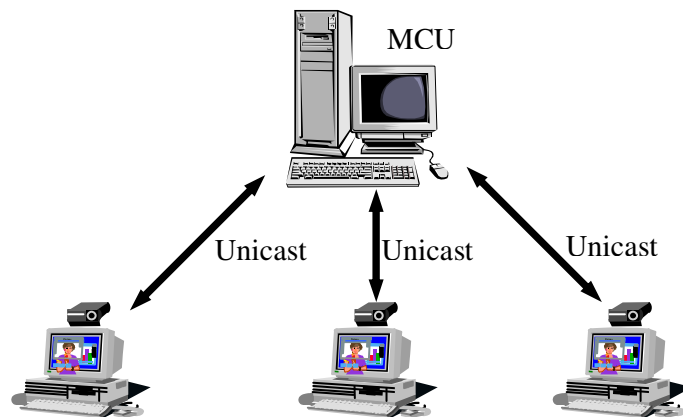


Figura 2.2 – Videoconferência baseada em modelo centralizado entre 3 ou mais participantes

Algumas vantagens deste modelo são:

- Os equipamentos dos participantes não precisam possuir recursos poderosos, pois localmente apenas codificam seu fluxo de mídia produzido e decodificam o fluxo enviado pelo MCU;
- Não exige nenhuma capacidade especial da infra-estrutura de rede, pois todas as conexões são ponto-a-ponto;
- Permite a integração mais fácil de aplicações de um único usuário em cenários orientados a grupos. O controle do material disponível é manuseado pelo MCU com redirecionamento de conteúdo para o equipamento dos participantes.

A algumas desvantagens são:

- O MCU geralmente é um recurso caro;
- Limitação em escalabilidade. Os sistemas de videoconferência são projetados para cenários específicos e não são facilmente adaptáveis para uso em outros cenários. Cada MCU suporta um número específico de participantes;
- Gera mais tráfego na rede. O fato de utilizar conexões *unicast* faz com que todo o tráfego seja roteado pelo MCU podendo provocar um congestionamento na rede.

Existem, no mercado, soluções de videoconferência multiponto centralizada que consistem somente de software, e outras que envolvem software e hardware.

2.1.3 Modelo descentralizado

O modelo descentralizado compartilha características de controle comum com o modelo centralizado, mas o fluxo de mídia é manuseado diferentemente. Uma das entidades participantes deve ser um MC que, independente do modelo de comunicação, provê o controle de três ou mais participantes durante uma sessão multiponto. O MC tipicamente é colocado com um dos participantes. Enquanto no modelo centralizado o MCU faz o processamento de mídia, no modelo descentralizado, os fluxos de mídia são enviados e recebidos por todos os participantes sobre uma base fim a fim.

A Figura a seguir ilustra um exemplo deste modelo. Não há MCU para processar os múltiplos fluxos; cada participante é responsável por sua própria mesclagem de áudio e seleção de vídeo. A mídia pode ser enviada entre todos os participantes utilizando *multicast*, ou múltiplas *unicast* se a rede não suportar *multicast* (Figura 2.3).

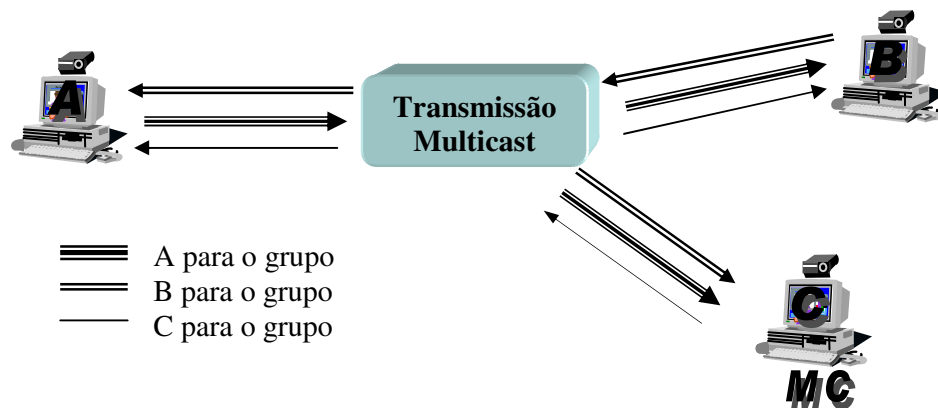


Figura 2.3 - Exemplo de videoconferência baseada em modelo descentralizado

Algumas vantagens deste modelo são:

- Não requer a presença de um MCU. Geralmente, um recurso limitado e caro;
- Permite um processamento individualizado em cada participante o que possibilita que cada um execute em seu próprio nível;
- Ao contrário do modelo centralizado, é escalável quanto ao número de participantes, mais flexível e sensível a variações nas condições da rede;
- Pode economizar largura de banda se a rede suportar *multicast*.

A algumas desvantagens são:

- Uma sessão de videoconferência é terminada quando o participante que contém o MC deixa a sessão. Para que a videoconferência não seja terminada o MC deve permanecer ativo;
- Um participante que possui um equipamento com baixo desempenho pode ser prejudicado já que a mesclagem do fluxo de áudio e a seleção do vídeo são realizadas localmente em cada participante e não só a codificação e decodificação como no modelo centralizado;

- Cada participante tem sua própria cópia dos arquivos que estão sendo compartilhados, o que torna mais difícil manter a consistência entre as cópias distribuídas no grupo.

Como exemplo deste modelo, pode-se citar as ferramentas utilizadas no MBone (*Multicast Backbone*), o qual suporta distribuição de dados *multicast*.

2.1.4 Modelo híbrido

O modelo híbrido tenta mesclar o melhor dos dois modelos anteriores, mantendo a consistência dos dados através de um armazenamento centralizado e suportando visões individualizadas através do uso de *front ends* (gráficos do modelo descentralizado no qual cada participante pode ter controle sobre sua aplicação para prover suas necessidades pessoais).

Uma implementação híbrida pode operar sobre uma rede *multicast* com cada usuário utilizando sua própria versão das ferramentas, distribuir o fluxo de mídia de acordo com o modelo descentralizado, mas conter algum mecanismo - tal como um servidor na conferência - para controlar documentos compartilhados, ou arquivar a mídia de sessões que ocorreram. Portanto, o modelo híbrido tem a vantagem de prover armazenamento centralizado para sessões de videoconferência sem ter que controlar cada instante da aplicação, para cada participante da sessão.

Uma videoconferência baseada no modelo híbrido é mostrada na figura 2.4. Neste exemplo, um MCU separado é usado para manusear o áudio, dados e controle de funções, e o vídeo é distribuído por *multicast* conservando a largura de banda.

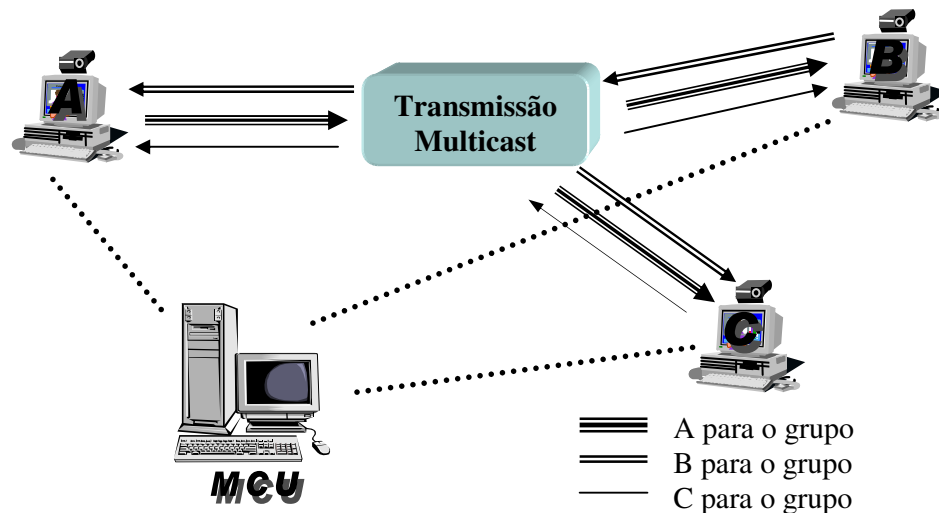


Figura 2.4 - Exemplo de videoconferência baseada em modelo híbrido

2.2 Serviços auxiliares

Atualmente o sistema de videoconferência vai muito além da transmissão de áudio e vídeo, pelo fato de existirem diversas características ou serviços auxiliares que podem ser adicionados. Esses serviços visam aumentar os benefícios que podem possuir uma videoconferência.

Conforme o [ITU 92] qualquer sistema de videoconferência deve, pelo menos, prover a transmissão e o recebimento de áudio e vídeo no mínimo entre dois participantes. Existem alguns serviços auxiliares que podem ser adicionados a uma videoconferência:

- Transmissão de imagens de alta definição e estáticas, como por exemplo as no formato MPEG. No caso de diagnóstico de uma patologia em um sistema de telemedicina.
- Na criptografia dos dados, utilizando algoritmos para garantir a privacidade de uma videoconferência, impossibilitando por exemplo, a espionagem industrial.
- Pode ser adicionado o serviço de transferência de arquivos, a exemplo do FTP.
- O sistema poderia possuir as opções de gravação das videoconferências, isso seria interessante no caso do ensino à distância para que alunos possam revisar suas aulas.
- A participação de um coordenador do sistema, este poderia coordenar a participação no grupo (entrada e saída de participantes), disponibilizar diversos recursos entre outros.
- Serviços de controle de colaboração de um documento, como exemplo: o fato de um documento estar sendo modificado por um dos participantes do grupo, os demais não poderiam modificá-lo enquanto não for finalizado o procedimento.
- O estabelecimento de prioridades no serviço, quando alguns participantes de uma palestra transmitida por videoconferência, possuem capacidades de ouvintes e interlocutores, ao passo que um outro grupo desta mesma videoconferência, apenas capacidades como ouvintes, não podendo interagir com o palestrante.
- Quando a transmissão de áudio apresentar problemas ou não for possível, pode-se disponibilizar canais para bate-papo (*chat*) como alternativa.
- A possibilidade do usuário escolher para seu equipamento, diferentes tipos de codificação ou compressão de mídias. Isto é importante, pelo fato de alguns usuários não possuírem os melhores recursos de hardware ou meios de transmissão por exemplo.

- O sistema deve oferecer um mecanismo onde o usuário possa “anotar” possíveis endereços de *sites*, para uma futura videoconferência quando desejado.

2.3 Codificação/compressão de áudio/vídeo

Para que áudio e vídeo possam ser manipulados pelo computador, estes devem ser capturados na sua forma analógica e armazenados como informações digitais. Para que exista transferência de voz e vídeo entre participantes de uma videoconferência através de uma rede de computadores, estes sinais devem sofrer um processo de codificação, ou seja, devem ser transformados em sinais digitais [ZAN 01]. Por fim, para que estes sinais possam ser apresentados para os participantes, eles devem ser novamente transformados em sinais analógicos.

Os dados de áudio e vídeo que vão ser transmitidos por uma rede de computadores, necessitam uma compressão de dados para não requererem uma grande quantidade de largura de banda. Baseado nisto, torna-se fundamental a aplicação de técnicas de compressão de dados, para que exista uma redução na quantidade de espaço que será destinada para armazenar informações como a de videoconferência.

Compressão é uma técnica de diminuição do tamanho dos dados sem mudar significativamente o seu conteúdo, a exemplo do que ocorre nos arquivos de extensão *zip*. Com isso, dados compactados podem ser transmitidos em uma rede de computadores com mais rapidez que os dados normais.

2.3.1 Áudio

Os áudios (voz humana, por exemplo) que trafegam pela Internet são sinais digitalizados, ou seja, sinais analógicos que foram convertidos para ser transportados pela rede. Segundo [KOR 01], a existência de um conversor Analógico-Digital (A/D), que tem a tarefa de transformar os sinais analógicos produzidos pelo microfone em uma corrente de dados digitais que possam ser transmitidos em pacotes através da rede. Em outro extremo da comunicação (ou destino), existe um conversor Digital-Analógico (D/A), que possui a capacidade de converter os sinais digitais para analógicos e reproduzi-los no auto-falante do computador.

Após os dados serem convertidos em analógico-digitais ou vice versa; estes devem passar por um processo de compressão ou descompressão. Os chamados *codecs*, são compostos por codificadores (*encoders*) e decodificadores (*decoders*). O codificador possui a finalidade de comprimir as amostras de dados digitais para serem transmitidos pela rede. Já os decodificadores possuem a função inversa, descomprimir os dados e enviá-los ao conversor D/A (Digital-Analógico) (Figura 2.5).

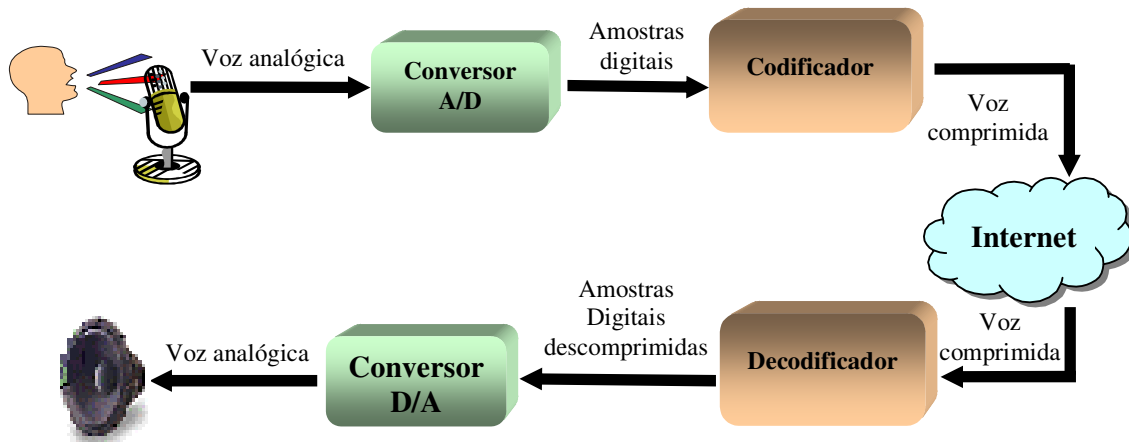


Figura 2.5 – Transmissão de voz na Internet utilizando codecs

Na compressão (encoders), existem alguns atributos relacionados para transmissão como [KOR 01]:

- **Taxa de bit:** Quanto maior a taxa de bits, maior a qualidade em sinais de vídeo. Como isso, pode-se concluir que quanto menor a taxa de bit utilizada pelo áudio melhor será aproveitada a largura de banda disponível. A diminuição se dá pelo aumento da compactação, onde ocorre algumas perdas de quadros da amostragem.
- **Qualidade:** Para aumentar a qualidade da amostragem de áudio, pode-se incluir detectores de atividade de voz. Ou seja, quando houver uma inatividade desta, o encoder é desligado e este espaço pode ser utilizado para prover maior qualidade. Porém pode existir uma desvatagem, que é o fato desses detectores influenciarem no desempenho do codec e na sua complexidade.
- **Complexidade:** A complexidade está relacionada com o número de instruções para comprimir/descomprimir os sinais que os codificadores/decodificadores utilizam.
- **Atraso:** Relaciona-se principalmente com o tamanho dos frames produzidos para transmitir na rede. Porém, pode-se também relacionar-se com o tempo de codificação de áudio.

Padrões do ITU-T para compressão de áudio digital:

- G.711

G.711 é um padrão internacional para codificação de áudio para os sistemas telefônicos em um canal de 64kbps. Trata-se de um esquema PCM operando a uma taxa de amostragem de 8 kHz com 8 bits por

amostra, segundo a recomendação do ITU-T. G.711 é capaz de codificar frequências entre 0 e 4 kHz.

- G.721

Especifica que o áudio analógico de 3 kHz deve ser codificado tomando como entrada os sinais PCM gerados pelo G.711, produzindo uma taxa de transferência de 32 Kbps.

- G.722

O padrão G.722 foi desenvolvido com a intenção de possibilitar a transmissão de voz com alta qualidade sobre canais de 65, 56 ou 48 kbps. O codec poder ser integrado em um *chip* e o seu *delay* médio é em torno de 3ms, pequeno bastante para não causar problemas de eco em redes de telecomunicações. Este padrão faz amostragens de 16 kHz e então as digitaliza em 14 bits por amostra, gerando uma largura de banda de 224 kbps. O fluxo digital é então filtrado em duas bandas, uma de 50 Hz a 4 kHz e outra para a faixa de 4 a 7 kHz.

- G.728

Define um método de codificação e compressão para voz humana. O resultado é um fluxo de dados que ocupa apenas 16 kbps.

2.3.2 Vídeo

Vídeo é uma seqüência de imagens paradas, que, apresentadas a uma taxa suficientemente rápida, causam a impressão de movimento contínuo. Por exemplo, no cinema, os quadros se sucedem 24 vezes por segundo, enquanto que na televisão essa taxa sobe para 25 (sistema PAL, adotado no Brasil) ou mesmo 30 (sistema NTSC americano). Com isso, pode se concluir que quanto maior a taxa de quadros ou frames, maior será a qualidade do vídeo e conseqüentemente, maior a largura de banda que este poderá ocupar.

Existem inúmeras soluções para minimizar o problema da largura de banda, como a compressão. Uma maneira de comprimir satisfatoriamente uma figura em movimento consiste em dividir a imagem em blocos e elaborar um chamado *vetor de movimento*. A compensação de movimento utiliza o fato de que um dado quadro é muito provavelmente similar ao que o precedeu. Portanto, essa técnica só é válida para cenas que não apresentam muita ação.

No caso de vídeo transmitido em tempo real sobre uma rede de computadores, existe a necessidade de capturar, codificar e transformar sinais de vídeo. Com essa finalidade, o ITU-T definiu dois padrões para a codificação/decodificação e também o transporte dos sinais de vídeo nos sistemas de videoconferência: H.261 e H.263 [KOR 01].

Segundo [KOR 01], quando dados são procedentes de uma fonte analógica, existe uma necessidade de conversão ou digitalização destes, para poderem trafegar em uma rede de computador. O sistema de captura deve deixar os frames ou quadros em formatos que possam ser entendidos pelos padrões H.261 ou H.263 (formatos CIF, QCIF) (Figura 2.6).

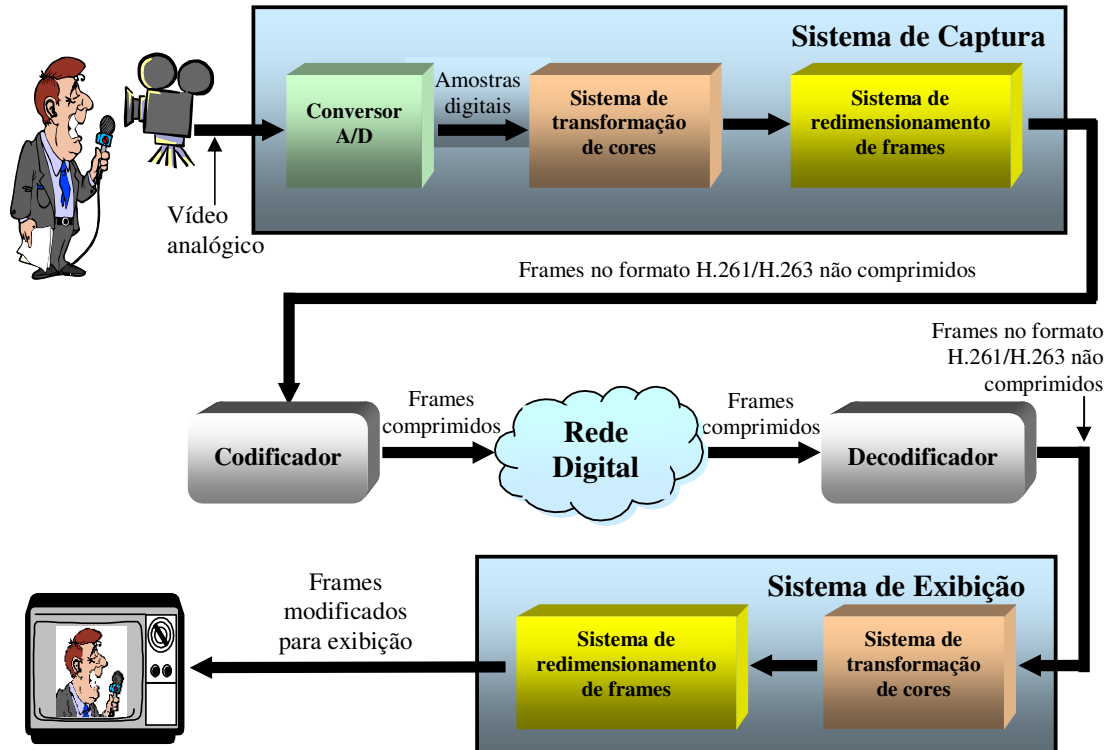


Figura 2.6 – Transmissão de vídeo em redes digitais

Segundo o ITU-T os codecs são:

- H.261:

É um padrão internacional amplamente utilizado para videoconferência na faixa de 128Kbps a 384Kbps. Ele utiliza a técnica de compressão espacial denominada codificação DCT (Discrete Cosine Transform). Trata-se de um método de codificação e decodificação de vídeo a taxas de $p \times 64$ kbits/s, onde p está na faixa de 1 a 30. Ele descreve a codificação da geração, multiplexação e transmissão de vídeo.

Este padrão foi elaborado para transportar vídeo em RDSI – particularmente para videofone e videoconferência. Videofone é menos exigente em relação à qualidade do vídeo e pode ser conseguido com $p=1$ ou 2. Para aplicações de videoconferência (em que há mais de uma pessoa no campo de visão) a qualidade da imagem deve ser melhor e p será, no mínimo, igual a 6.

H.261 define dois formatos de quadros: o CIF (common Intermediate Format) tem 288 linhas de 352 pixels/linha de informação de luminância e 144 x 180 de crominância; e QCIF (quarter common Intermediate Information) que possui 144 linhas de 180 pixels de luminância e 72 x 90 de crominância. A escolha de CIF ou QCIF depende da capacidade disponível do canal – por exemplo, QCIF é normalmente usado se $p < 3$.

O algoritmo inclui um mecanismo que otimiza a utilização da largura de banda ao contrabalançar movimento e qualidade de quadro, de modo que uma sucessão de quadros variando rapidamente redundará em uma qualidade de imagem pior do que uma imagem relativamente estática. Usado dessa maneira, o padrão H.261 é uma codificação a taxa de bit constante (*constant-bit-rate*) ao contrário do esquema de taxa de bits variável, em que a qualidade de vídeo é que é constante. Há ainda outros esquemas de taxa de bit constante, que, em vez de diminuir a qualidade da imagem à medida que a largura de banda disponível diminui, opta por não transmitir toda a imagem.

- H.263:

Trata-se de outro padrão internacional e é baseado na codificação DCT e compensação de movimento, sendo esta o principal aperfeiçoamento em relação ao antecessor H.261. Foi concebido para larguras de banda estreitas.

Segundo [ZAN 01], o H.263 é uma nova versão do padrão do H.261, onde foram adicionados alguns elementos que aumentam a eficiência da compressão. Na mesma qualidade da imagem percebida, é possível reduzir o tamanho do fluxo de 30% a 50%. Porém, a mais importante das novas características é a introdução de opções negociáveis. Estas opções podem ser negociadas entre o codificador e o decodificador no início do processo de transporte dos fluxos. O codificador deve ser capaz de utilizar todas as opções durante o estágio de compressão para tirar a máxima vantagem da efetividade do H.263. O decodificador pode implementar somente um subconjunto de opções negociáveis, pedindo fluxos de dados somente com estas opções e ser capaz de decodificar.

- H.263 versão 2:

Esta nova versão do H.263 inclui uma maior eficiência de compressão em relação a anterior. Possui uma escalabilidade de taxas de transferência para redes sem fio e de pacotes. Esta padronização também pode incluir tamanhos de frames customizáveis [KOR 01].

2.4 Protocolo RTP e RTCP

O protocolo RTP foi projetado e unificado pelo IETF (Internet Engineering Task Force) especialmente para oferecer suporte ao tráfego de aplicações que transmitem dados em tempo real (videoconferência, por exemplo); normalmente implementado como parte da aplicação e não como parte do kernel do sistema operacional [LIN 01].

Segundo [MEL 97], em algumas redes de pacotes ou mesmo na Internet, existem perdas e desordenações de pacotes, também atrasos na entrega e intervalos variáveis. Não é função do protocolo RTP fornecer mecanismos que garantam a entrega e também que efetue ordenação de pacotes. O cabeçalho do protocolo possui informações de temporização e número de seqüência, fica a aplicação de destino encarregada de montar os datagramas, passando a ter posse das informações de temporização e ordenação de pacotes da origem. Com os números de seqüência, é possível determinar a localização de um pacote sem necessariamente decodificar os pacotes em seqüência, como em decodificações de vídeo.

Cabe salientar que o protocolo possui duas divisões: o próprio RTP que está relacionado com os dados e a outra parte chamada de RTCP. O RTP é responsável pelo transporte das mídias contínuas (áudio e vídeo) e o RTCP se baseia na transmissão periódica de pacotes de controle de todos os participantes de uma sessão, utilizando o mesmo mecanismo de distribuição dos pacotes de dados. Os protocolos inferiores devem fornecer multiplexação dos dados e controle dos pacotes, utilizando por exemplo número de portas separadas para o UDP.

Segundo [LIN 01], RTP utiliza uma estrutura chamada camada de aplicação do frame. Existem padrões para codificar dados de vídeo, como MPEG, JPEG e H.261. O RTP oferece um suporte para essas estruturas ou codificações (Figura 2.7). Faz parte da estrutura RTP uma informação de *timestamping* que permite ao receptor de um vídeo sincronizar voz e movimento de figuras. Cada fonte de dado RTP marca cada mensagem com um *timestamping* indicando quando um evento acontece, dessa forma o receptor representa o dado recebido no mesmo tempo relativo. Isso garante por exemplo, que quando o computador receber o áudio, ele seja reconstruído no tempo original. Redes baseadas em TCP/IP não mantêm essa relação do tempo entre as mensagens, elas podem chegar ao destino em intervalos irregulares, fora de ordem, ou ainda podem não chegar iguais.

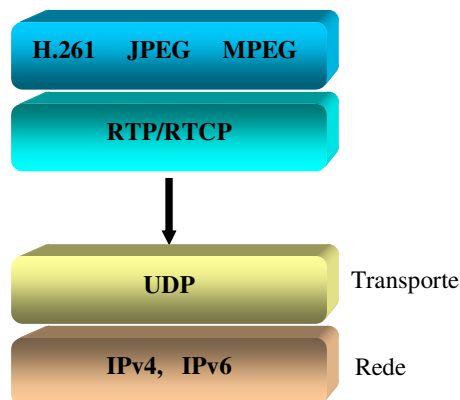


Figura 2.7 - RTP combinado com um formato de *payload* para formar um protocolo completo

O protocolo RTP também trata questões sobre os *mixers* e *translators*. *Mixers* e *translators* são usados sempre que alguns usuários tenham necessidades diferentes dos demais. Um exemplo disso é um usuário em um segmento de baixa largura de banda conectado em uma videoconferência de alta largura de banda. Ao invés de reduzir a largura de banda de toda a videoconferência, é possível colocar um *mixer* no caminho do usuário que dispõe de baixa largura de banda. Este *mixer* transforma os fluxos reduzindo a sua largura de banda.

Os *translators* podem transformar um formato *payload* em outro. Cabe relatar que este novo formato possui uma baixa qualidade em relação ao anterior, porém aliá-se a isto, o fato de possuir uma largura de banda menor (Figura 2.8).

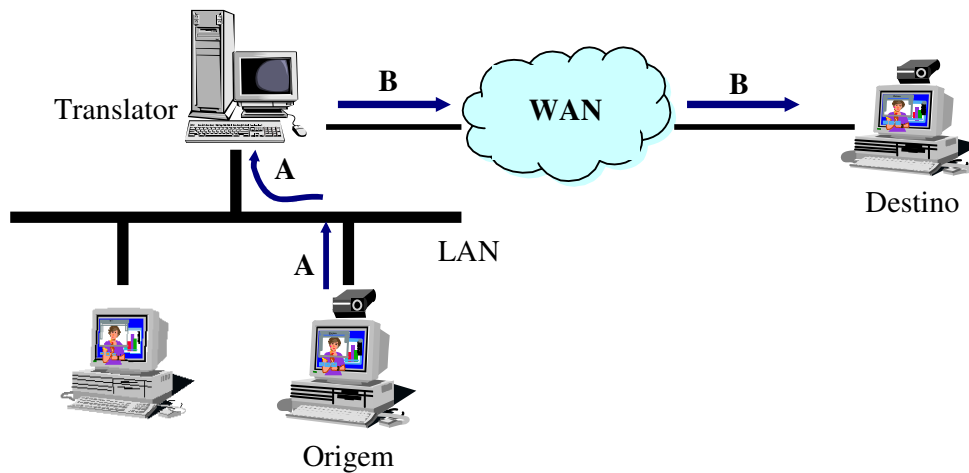


Figura 2.8 – Tradutor processando pacotes RTP

No caso dos *mixers*, não ocorre uma alteração no formato do *payload* e sim uma combinação dos múltiplos fluxos em somente um, o que acaba preservando o formato original. Esse conceito é interessante em transmissões de áudio, múltiplas fontes de vídeo por exemplo, não podem ser combinadas em somente uma. A abordagem dos *mixers* possui uma boa eficiência em conferências utilizando áudio com muitos participantes (Figura 2.9).

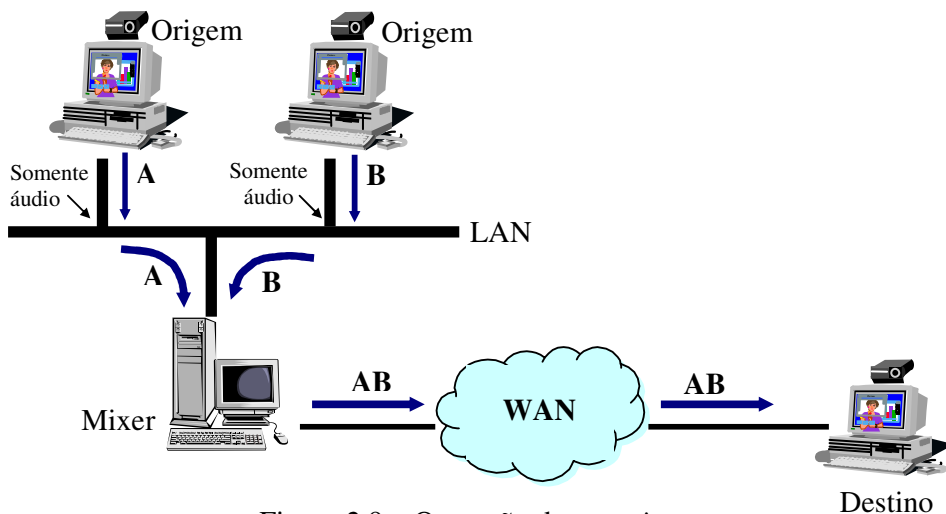


Figura 2.9 – Operação de um mixer

O RTCP prevê uma transmissão periódica de pacotes de controle entre todos os participantes de uma sessão. A transmissão dos pacotes de controle e de dados é feita em canais separados, permitindo mais flexibilidade para a avaliação. Os pacotes RTCP não encapsulam os dados de áudio e vídeo, mas ao invés disso, eles contêm relatórios do transmissor e/ou receptor informando algumas estatísticas que podem ser úteis para aplicação. Estas estatísticas incluem número de pacotes enviados, número de pacotes perdidos e *jitter*. O RTCP não indica o que a aplicação deve fazer, tarefa que cabe ao desenvolvedor do software. Os receptores podem usar as informações de *feedback* para, por exemplo, modificar suas taxas de transmissão. Estas informações podem também ser usadas pelos receptores para que eles possam diagnosticar se os problemas existentes são locais, regionais ou globais.

Os pacotes RTP e RTCP são distribuídos para todos os participantes através de IP *multicast*. Isso pode ser notado na figura 2.10.

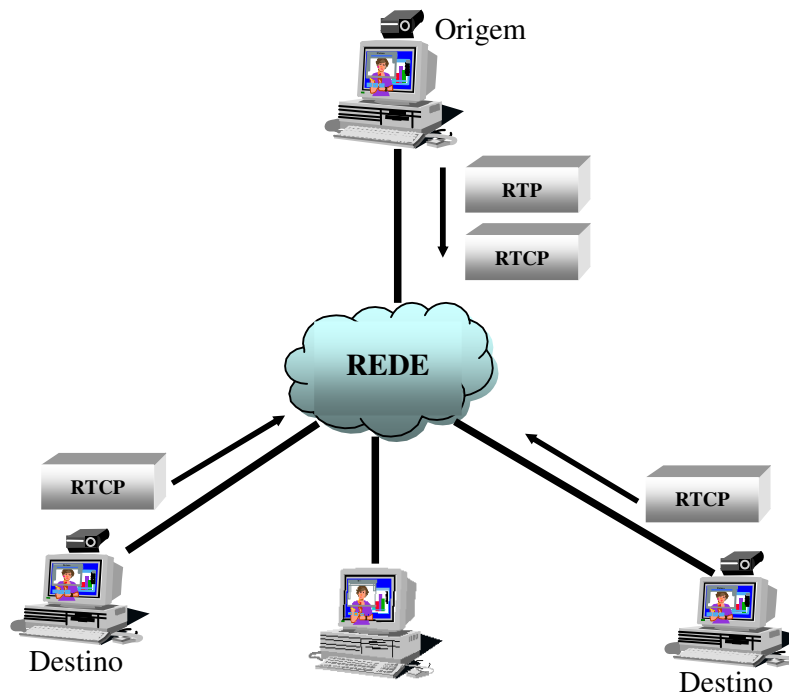


Figura 2.10 – Pacote RTP e RTCP enviados pela rede

2.5 Alguns problemas enfrentados na videoconferência

Um dos problemas mais conhecidos que dificultam as transmissões de videoconferência em redes de computadores (como a Internet, por exemplo), trata-se da largura de banda. Porém, existem outros fatores que também devem ser considerados:

- Atraso (*Delay*)

Delay é o tempo que decorre da transmissão de um pacote de informação da origem até o destino. Quanto maior o atraso, mais problemas tem o protocolo de transporte para trabalhar eficientemente. Aplicações multimídia como a videoconferência, estão diretamente

relacionadas com o problema do atraso. A solução para isso, é buscar que os projetistas destas aplicações busquem mecanismos capazes de preservar a boa qualidade dos dados mesmo na sua presença.

- Latência

A questão da latência é representada pelo atraso que é imposto pelo *link* de comunicação. Esse retardo que é introduzido torna-se constante, o que faz da latência um atraso médio esperado em transmissões fim-a-fim. Como exemplo, pode apresentar as chamadas via satélite (principalmente as internacionais), onde o atraso pode muitas vezes confundir o usuário (Figura 2.11).

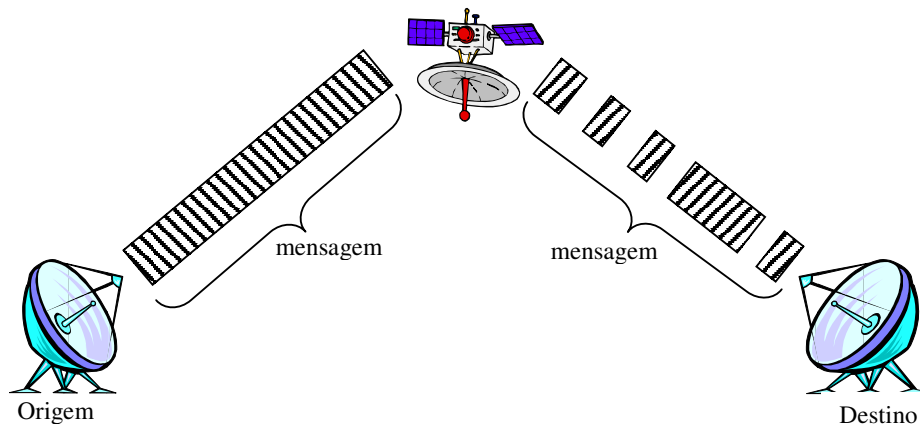


Figura 2.11 – Espaçamento nas mensagens que chegam no destino

- Jitter

Segundo [ZAN 01], o *jitter* é um dos componentes do atraso fim-a-fim, causa algumas “flutuações” na latência ou variações estáticas no atraso. Ou seja, o *jitter* pode ser causado pelo enfileiramento randômico dos pacotes em roteadores.

Na transmissão de voz sobre IP os datagramas podem tomar caminhos diferentes na rede, resultando em diferentes tempos de propagação, ou podemos ter congestionamentos momentâneos que obriguem a maiores retardos. Para contornar este problema, são usados *buffers* nas entradas dos equipamentos decodificadores, a fim de guardar alguns pacotes como “reserva” em uma fila, que é servida de forma mais constante possível. Mesmo que alguns pacotes sofram uma demora maior que a normal para chegada (guardando certos limites), os pacotes no *buffer* são enviados para o decodificador com uma cadência homogênea. A quantidade ótima de pacotes armazenados nestes *buffers* depende do tamanho dos pacotes de voz, taxa de transmissão, atraso médio da rede e como não poderia deixar de ser, exigência da qualidade de voz requerida.

O problema do *jitter* pode ser eliminado pelo uso de *sequence numbering* (enumeração da seqüência dos pacotes), *timestamps* (temporizadores) e *playout delay* (bufferização). A figura 2.12, apresenta uma relação entre a latência e o *jitter*.

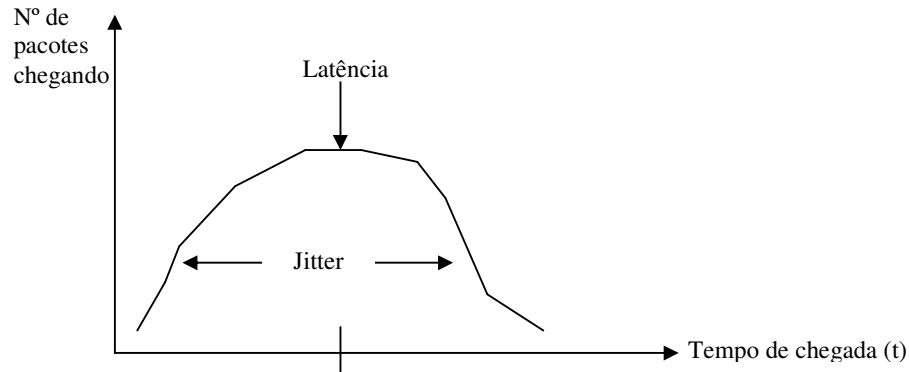


Figura 2.12 – Relação entre latência e *jitter*

- Skew

O problema do *skew* está relacionado com os tempos de apresentação de dois objetos relacionados (áudio e vídeo em uma videoconferência, por exemplo) (figura 2.13). A questão do movimento dos lábios de um locutor está relacionada com o áudio resultante. Portanto, áudio e vídeo devem estar muito bem sincronizados.

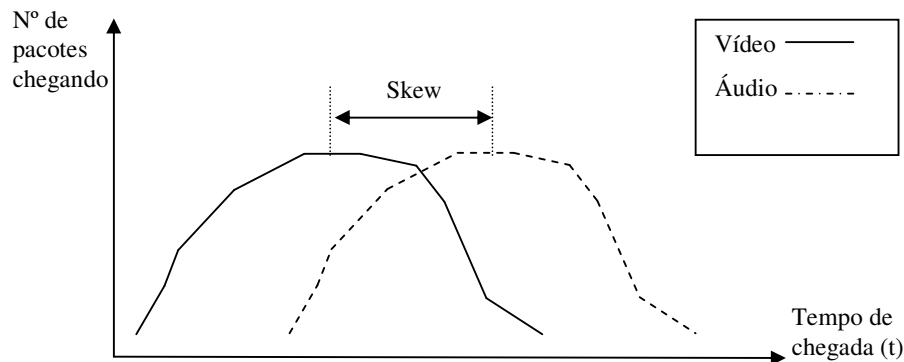


Figura 2.13 – O problema do *skew* entre áudio e vídeo

- Confiabilidade

Confiabilidade pode ser considerada como alguns dos problemas enfrentados na videoconferência. Alguns dos fatores que contribui para a falta de confiabilidade são: a alteração da ordem da transmissão inicial e perda de alguns pacotes. Esses pacotes que são perdidos, por exemplo, podem contribuir bastante para o aumento do tráfego em uma rede, devida a retransmissão (protocolo TCP). A maioria dos sistemas de videoconferência usa UDP, que trabalha com transmissão de pacotes com o mínimo de controle, a propagação de erros pode ocasionar uma distorção do sinal analógico original no receptor.

3 Soluções para videoconferência

Soluções para videoconferência permitem que pessoas que estejam separadas pela distância possam trabalhar em colaboração juntas, adicionando alguma característica à comunicação, com isso, criando um certo valor para a organização que está desfrutando da videoconferência. Dentro dos conceitos da videoconferência, serão apresentadas neste capítulo, algumas soluções.

3.1 H.323

O ITU-T (*Telecommunication Standardization Sector of International Telecommunication Union*) propôs a Recomendação H.323, que descreve uma arquitetura contendo terminais, equipamentos e serviços para comunicação multimídia sobre LANs (*Local Area Networks*), sem garantia de qualidade de serviço. Na verdade a H.323 é uma recomendação “guarda-chuva”, referenciando outras recomendações ITU-T e até o protocolo RTP (*Real-Time Transport Protocol*) do IETF. A primeira versão desta recomendação foi publicada em 1996 e a segunda em 1998. Esta última contém melhorias no tocante à segurança, diminuição do tempo necessário para liberação do canal após a chamada ter sido atendida pelo destino, incorporação de serviços adicionais como transferência, redirecionamento, etc. e maior integração com T.120 (protocolo de transmissão de dados multimídia).

Basicamente, destina-se a transmissão em ambiente com uma ou mais redes locais, sem garantia de qualidade de serviço. Como por exemplo: Ethernet (IEEE 802.3), Fast Ethernet (IEEE 802.10), FDDI (no modo sem garantia de QoS) e Token Ring (IEEE 802.5).

3.1.1 Componentes da arquitetura

O ITU-T define na recomendação H.323 a descrição dos seus componentes, visando a prover serviços de comunicação multimídia sobre redes de pacotes, tais como LAN's, MAN's e WAN's incluindo a Internet (não garantem qualidade no serviço). Os componentes descritos compreendem terminais, *gateways*, *gatekeepers*, MCU (*Multipoint Control Units*) (Figura 3.1).

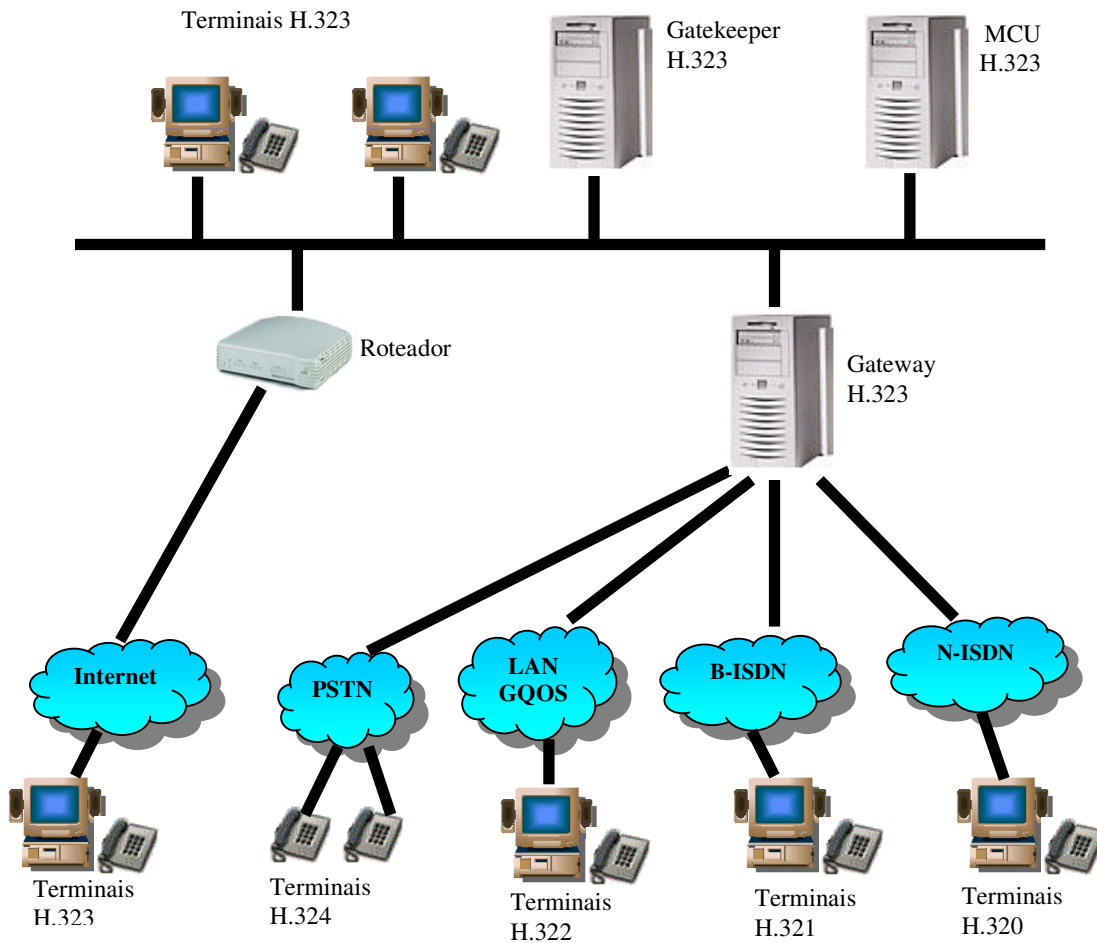


Figura 3.1 – Componentes da arquitetura H.323

- Terminais

Por definição é um ponto de rede que provê comunicação bidirecional, em tempo real, com outro equipamento terminal, *gateway* ou MCU. Pode ser um equipamento telefônico com a funcionalidade de comunicação através do protocolo IP, um microcomputador com *software* específico e *hardware* multimídia para comunicação de voz ou qualquer outro tipo de equipamento que possa estar conectado em rede e permita a transferência apenas voz; voz e dados; voz e vídeo; ou voz, dados e vídeo, entre os usuários (Figura 3.2).

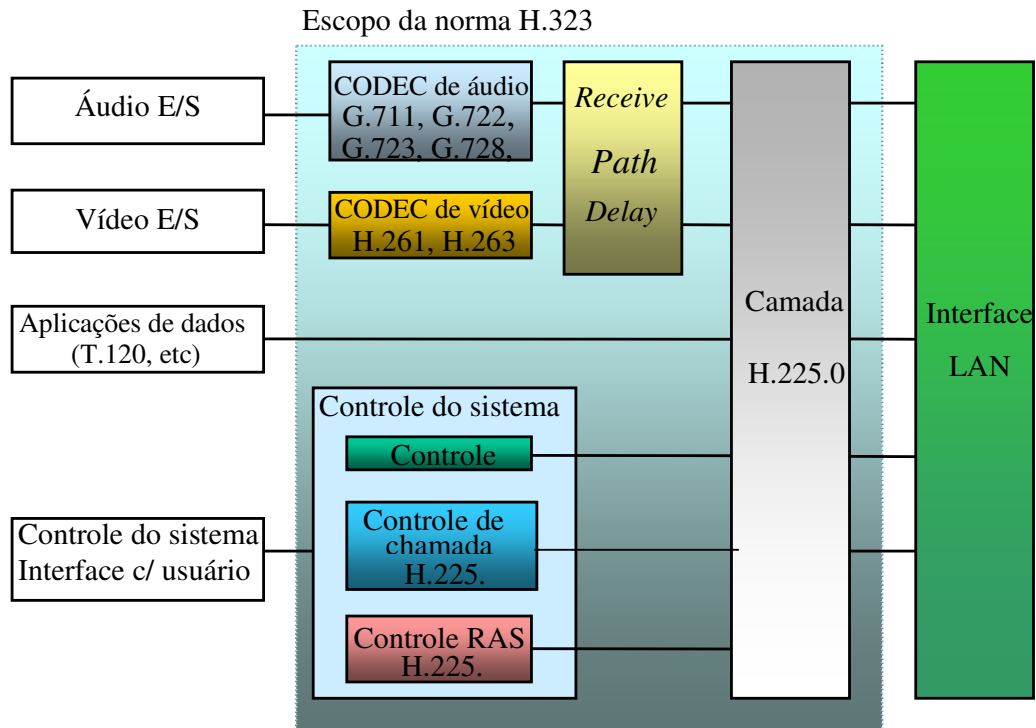


Figura 3.2 – Terminal H.323

O são ilustrados os elementos internos de um terminal H.323, sendo eles:

- Codec (codificador/decodificador) de vídeo – suportando obrigatoriamente a Recomendação H.261 e opcionalmente a H.263, referentes à codificação de sinal de vídeo. A implementação deste elemento é opcional;
- Codec de áudio – suportando obrigatoriamente a Recomendação G.711 e opcionalmente as Recomendações G.722 (sinal de 7KHz codificado em 64Kbps), G.728, G.729, áudio MPEG1 e G.723.1. A implementação deste elemento é obrigatória;
- Atraso de recepção – responsável pelo controle do *jitter* nas recepções de áudio e/ou vídeo. A implementação deste elemento é obrigatória;
- Canal de dados – apesar de não estar explicitamente indicado na figura, suporta um ou mais canais de dados, uni ou bidirecionais. O T.120 é a base padrão para interoperabilidade entre terminais H.323 ou outro tipo suportado pelos ambientes externos ligados por *gateway*. A implementação deste elemento é opcional;
- Controle H.245 – responsável pelo manuseio de mensagens de controle fim-a-fim, que governam a operação da entidade H.323, no tocante à troca do modo de operação de receptor para

transmissor e vice-versa, abertura e fechamento de canais lógicos para troca de informações de um transmissor para um ou mais receptores, requisições de preferência de modos de operação (*unicast* ou *multicast*), mensagens de controle de fluxo e comandos gerais e indicações;

- Controle RAS – controle das mensagens de sinalização para registro, admissão, troca de banda, pedido de informação sobre a situação atual e desvinculação, com o *gatekeeper*;
- Controle de Chamadas – usado para o estabelecimento de conexão entre dois terminais H.323;
- H.225.0 – estabelece o formato das mensagens trocadas pelos canais lógicos de áudio, vídeo dados ou controle.

- Gateway

Converte, apropriadamente, diferentes formatos de mensagens (exemplo: H.225.0 / H.221 - estrutura de quadros em canais de 64 a 1920 Kbps em tele-serviços audiovisuais) e procedimentos de comunicação (exemplo: H.245 / H.242 - estabelecimento de comunicação entre terminais audiovisuais em canais digitais até 2 Mbps). A conversão entre diferentes formatos de áudio, vídeo ou dados também pode ser feita pelo *gateway*.

Os terminais H.323 podem conversar entre si em uma mesma rede, sem o envolvimento de um *gateway*. Entretanto quando um terminal H.323 desejar conversar com uma pessoa que possui apenas um telefone comum, conectado na rede pública de telefonia (PSTN) ou terminal H.320 e em linhas ISDN, isto deverá executar todas as traduções necessárias para que a comunicação aconteça (Figura 3.3). A sua função é refletir as características de um terminal na rede IP para um terminal na SCN e vice-versa.

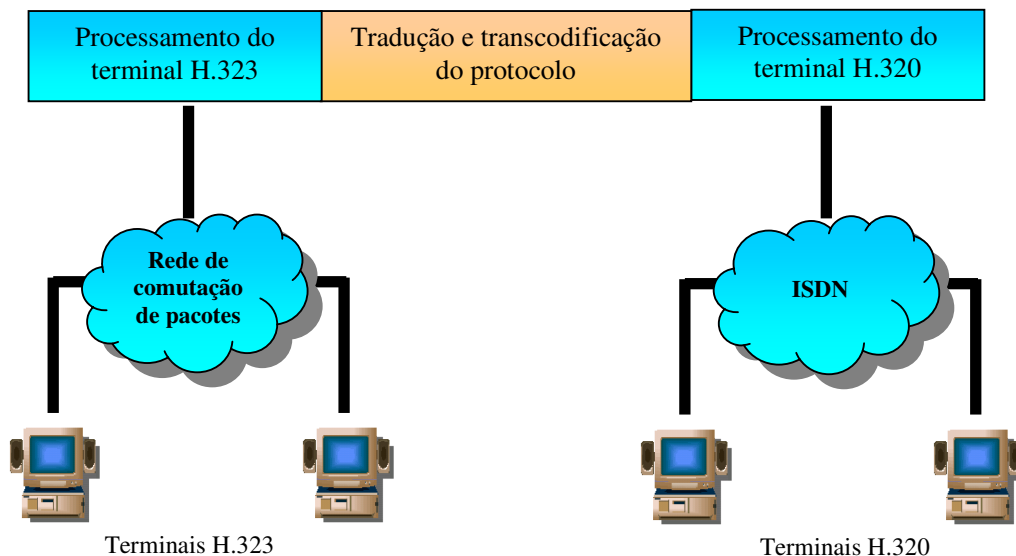


Figura 3.3 – Gateway H.323/H.320

- Gatekeeper

É um elemento opcional no sistema H.323, provendo serviços de controle de chamadas para os terminais. Possui como funções obrigatórias a tradução de endereços (permite o uso de apelidos (*alias*) em lugar dos Endereços de Transporte), controle de admissão (autoriza o acesso de recursos usando por base critérios como: permissão de acesso em período predeterminado, banda disponível, etc.), controle de banda (rege os pedidos de troca de banda em uso) e a gerência de zona (prove as funções acima para terminais, MCUs (*Multipoint Control Units*) e *gateways* nele registrados).

Opcionalmente, o *gatekeeper* pode apresentar as funções de controle de sinalização de chamadas (pode controlar todo processo de sinalização de chamadas entre terminais, ou deixar que os próprios terminais tratem da sinalização), autorização de chamadas (através da sinalização prevista na H.225.0, o *gatekeeper* pode rejeitar chamadas de um terminal sem autorização adequada), gerência de banda (controla o número de acessos simultâneos de terminais H.323 à LAN, podendo rejeitar uma chamada por falta de banda disponível), gerência de chamadas (controla que terminais possuam chamadas estabelecidas, para o caso de uma nova chamada a um terminal ocupado ser informada aos módulos de direito).

Um conceito importante, introduzido pela segunda versão do H.323 é o de Zona. Por definição, Zona abrange a coleção de todos os terminais, *gateways*, e MCUs, gerenciados por um único *gatekeeper*. Cada Zona possui apenas um *gatekeeper* e não precisa estar confinada geograficamente a mesma LAN (Figura 3.4).

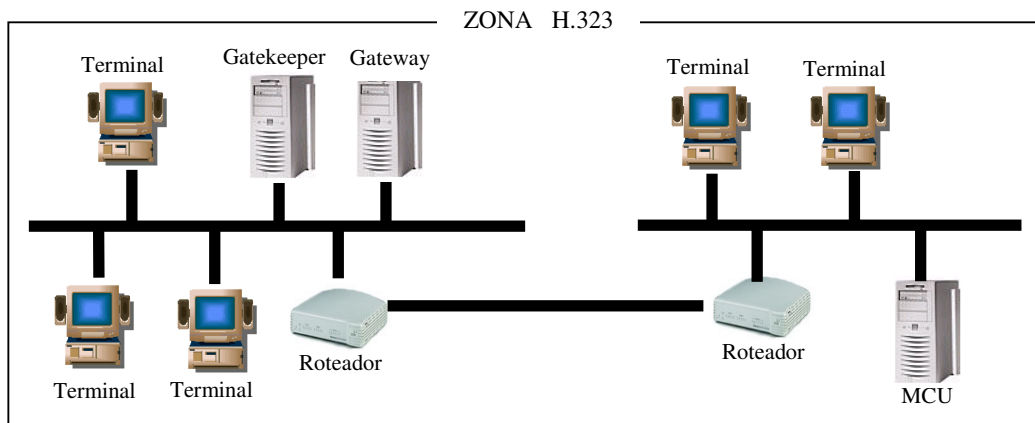


Figura 3.4 – Zona H.323

- MCU

O MCU é um componente do H.323 que provê a capacidade para suportar conferências multiponto. Pode estar dividido em um MC (obrigatório) e MPs (não obrigatório).

MCU suporta conferências multipontos centralizadas através de um MC e da agregação de sinais de áudio, vídeo e dados. Conferências descentralizadas constituem de um MC e suporte para agregação de dados através de um MP que implementa as funções conforme a recomendação T.120 usando capacidades de áudio e vídeo descentralizadas [PIN 02].

Um *gateway* pode implementar um MCU no lado interno da rede. Um *gatekeeper* poderá também conter um MCU em qualquer dos casos, as funções devem estar separadas e não podem interferir uma na operação da outra. Isto significa que embora as funções possam coexistir em um mesmo equipamento, elas devem ser implementadas separadamente e a ativação de um serviço, não implica na ativação do outro e vice-versa.

3.1.2 Sinalização no H.323

Segundo [INT 02], a sinalização H.323 é complexa, devido principalmente a sua extensa pilha de protocolos, e a conformidade com padrões antigos da ITU-T. As mensagens ARQ (*Admission Request*, ou pedido de abertura de sessão), ARJ (*Admission Reject*, ou a rejeição do pedido) e ACF (*Admission Confirm*, ou a confirmação do pedido) são exclusivas dos terminais H.323.

Estas mensagens, em conjunto com LRQ (*Location Request*), LCF (*Location Confirm*), LRJ (*Location Reject*), usadas pelos *gatekeepers*, formam o que denomina-se conjunto de mensagens RAS (Requisição, Admissão e Status). Um terminal registrado em um *gatekeeper* sempre pedirá autorização ao *gatekeeper* para iniciar e/ou aceitar chamadas. As mensagens Q.931 são SETUP (estabelecimento de chamada ISDN), Call Proceeding (processamento da chamada ou solicitação) e CONNECT (confirmação do estabelecimento de chamada).

Na fase de inicialização de mídia pelo H.245, uma porta TCP é aberta para negociação dos subconjuntos de mídias suportados e a ordem de preferência das mídias. O canal H.245 é mantido aberto caso alguém abra uma nova sessão de mídia, ou modifique uma existente. As mensagens mais básicas do H.245 são: *Capability Exchange* (troca de conjuntos de capacidades de mídia entre os terminais), *Open Logical Channel* (abertura de canal de controle do fluxo de mídia) e *Open Logical Channel Acknowledge* (confirmação do mesmo). O transporte do fluxo de mídia, após a fase de negociação, acontece no nível de rede pelo uso do protocolo de transporte RTP (*Real time Transport Protocol*) (Figura 3.5).

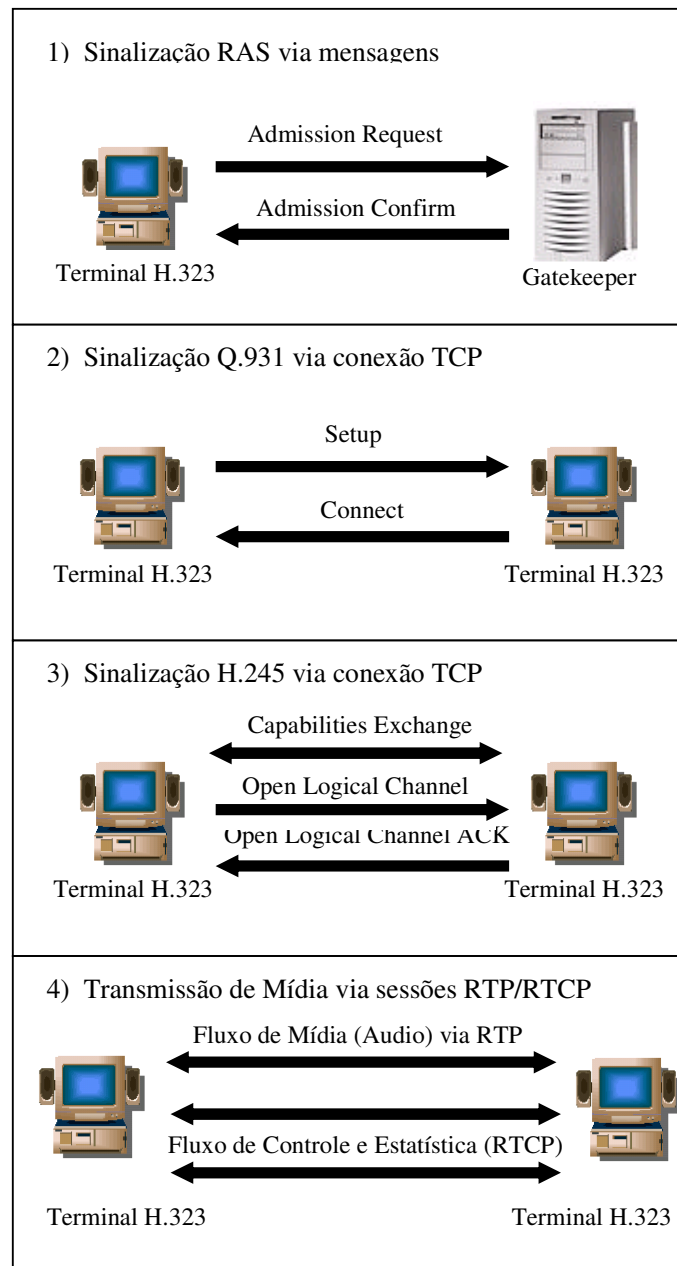


Figura 3.5 – Sinalização no H.323

3.2 SIP (*Session Initiation Protocol*)

Pode-se dizer que SIP trata-se de um protocolo de controle referente à camada de aplicações do Modelo de Referência OSI (*Open System Interconnection*), que é usado para iniciar, modificar ou terminar sessões ou chamadas multimídia entre usuários. Dentre suas funcionalidades tem-se a localização de usuários, o estabelecimento de chamadas, o suporte a *unicast* ou *multicast*, administração na participação de chamadas (transferências, conferência, entre outros) e possibilidade de participação de um usuário em terminal H.323, via *gateway*. É um protocolo cliente-

servidor similar ao HTML no tocante à sintaxe e semântica das estruturas empregadas, com campos explicitamente descritos [DAL 02] [WHA 02].

Esta solução de videoconferência, estabelece, modifica e termina sessões multimídia e/ou ligações. Estas sessões podem ser conferências multimídia, aulas pela Internet, telefonia sobre Internet, entre outras. O protocolo SIP é baseado no HTTP e, assim como este, suporta o transporte de qualquer tipo de carga em seus pacotes, pelo uso de *Mime-Types (Multipurpose Internet Mail Extensions)*. Por utilizar uma arquitetura cliente/servidor, suas operações envolvem apenas métodos de requisição e respostas, como observado também no HTTP e no RTSP [RFC2543].

Como mencionado anteriormente, o fato do SIP tratar-se de um protocolo cliente-servidor (o cliente realiza chamadas que são atendidas pelo servidor), em alguns casos, uma chamada pode envolver diversos servidores e clientes [INT 02]. Um protocolo como o SIP, deve oferecer funções básicas como:

- Conversão de nomes e localização de usuários:

Envolve o mapeamento entre nomes de diferentes tipos de abstração, tais como nomes de um domínio e o nome de um usuário em servidor *Internet*, isto é necessário para que um determinado usuário, que possui um nome qualquer, possa ser convertido em um endereço IP, de modo que possa ser localizado a qualquer momento da ligação;

- Negociação de configuração:

Permite que um grupo de usuários defina que tipo de informação será trocada e seus respectivos parâmetros. O conjunto e o tipo dos dados que estão sendo enviados não precisam ser uniformes dentro de uma chamada, como diferentes conexões ponto a ponto pode envolver diferentes tipos e parâmetros de dados. Muitos codecs são capazes de receber diferentes tipos de codificações, sendo restritos ao enviar apenas um tipo de dado em cada fluxo;

- Alteração de configuração:

Torna possível a alteração, de maneira dinâmica, ou seja, durante a utilização de uma conexão por seus usuários, dos parâmetros definidos no momento do estabelecimento da conexão.

3.2.1 Componentes da arquitetura

SIP foi concebido na Universidade de Columbia e depois submetido para aprovação do IETF (*Internet Engineering Task Force*). Foi aprovado como RFC (*Request For Comment*), sendo publicado como RFC 2543, em março de 1999 [PRO 02]. O IETF define um conjunto de componentes na sua arquitetura operando numa rede IP, este conjunto é definido como “rede” SIP. Estes componentes são apresentados na figura 3.6:

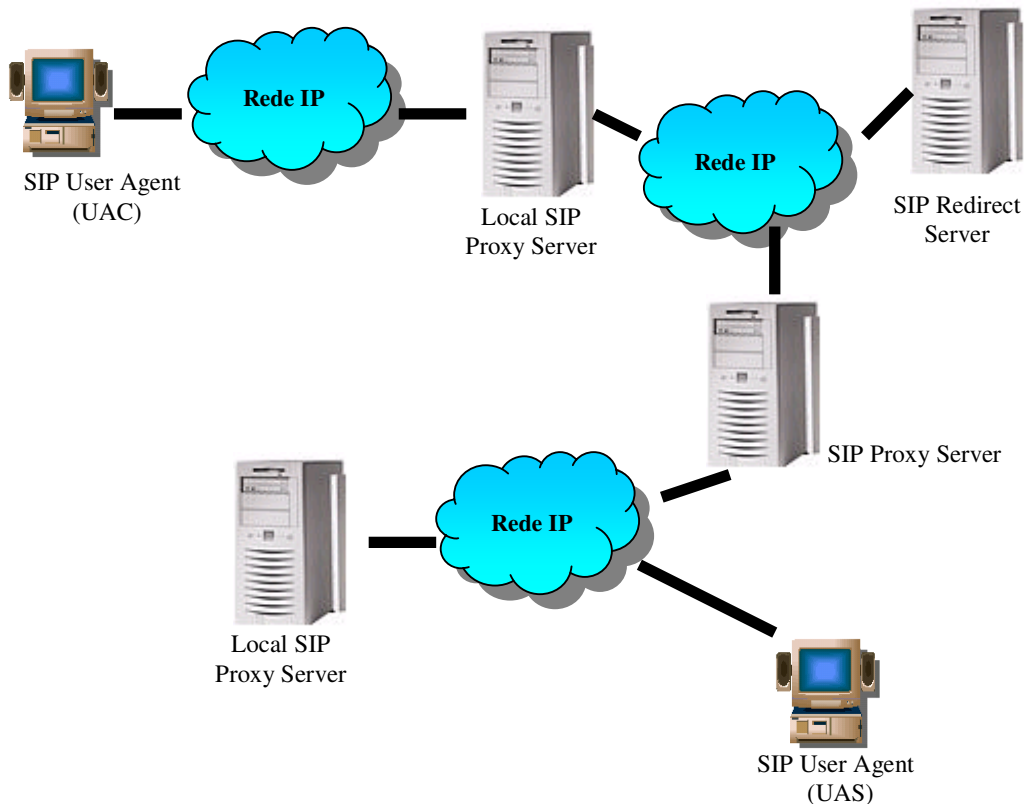


Figura 3.6 – Componentes da arquitetura SIP

- SIP User Agent

Cliente da arquitetura, ou o ponto final da comunicação multimídia que interage com o usuário. Um *user agent* tem dois componentes, um *user agent client* (UAC) e um *user agent server* (UAS). O UAC é responsável por iniciar as chamadas enviando requisições, e o UAS é responsável por responder às chamadas, enviando respostas. Uma aplicação de telefonia Internet por exemplo, contém ambos UAC e UAS. Pode-se destacar a diferença com um *Web Browser* que contém apenas um cliente.

- SIP Proxy Server

Servidor de redirecionamento de requisições e respostas SIP. Passa a realizar a sinalização como se fosse o originador da chamada, e quando a resposta lhe é enviada, ela é redirecionada para o originador real. O servidor de *proxy* também é conhecido como *next-hop*. O servidor *next-hop* pode ser outro servidor de *proxy*, um UAS ou um servidor de redirecionamento.

Um exemplo da utilização do *proxy* pode ser verificado na Figura 3.7:

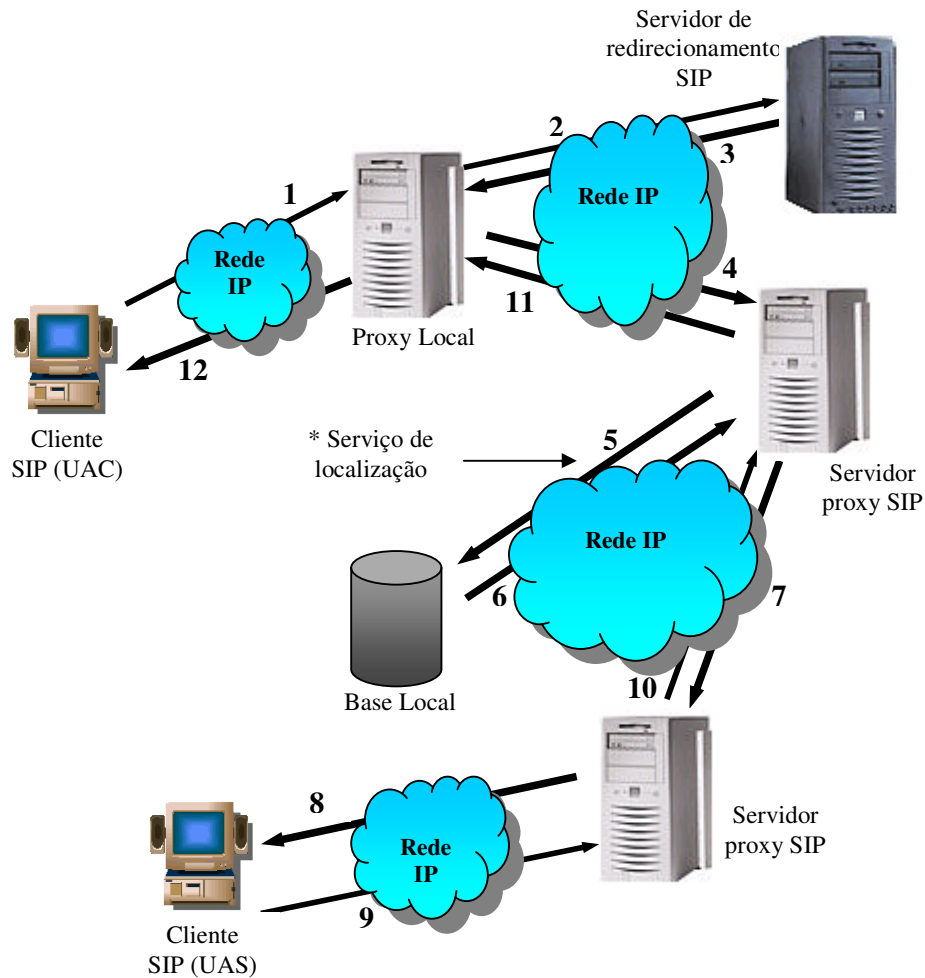


Figura 3.7 – Servidor de *proxy*

O primeiro passo que acontece na figura anterior, é quando o usuário encaminha um *invite request* para um *proxy* local, tendo como destino paulo@empresa.com (1). O *proxy* procura empresa.com no DNS (*Domain Name System*) e obtém o endereço IP do servidor que trata requisições SIP deste domínio. O *proxy* então encaminha o *request* para esse servidor (2). O servidor de empresa.com conhece o usuário paulo, mas ele está “logado” no momento como n.aluno@universidade.edu. Então, o servidor redireciona o *proxy* (3) para tentar este endereço. Também com auxílio do DNS, o *proxy* obtém o endereço IP do servidor responsável por universidade.edu e envia o *request* para esse servidor (4). O servidor consulta uma base local (5) e indica que n.aluno@universidade.edu é localmente identificado como paulo.silva@inf.universidade.edu (6). O servidor principal de universidade.edu envia o *request* para o servidor da informática (7), que por sua vez, sabendo o endereço IP do usuário, envia o *request* para ele (8). O usuário aceita a chamada e responde retornando pela cadeia de *proxy* (8), (10), (11) e (12).

- SIP Redirect Server

Recebe requisições e determina um servidor *next-hop*. Todavia, ao contrário de reenviar neste caso, ele retorna o endereço do servidor *next-hop* para o cliente. A função primária dos servidores *proxy* e de redirecionamento são roteamento de chamadas, ou seja a determinação do conjunto de servidores a serem usados no caminho para completar a chamada. Um servidor de *proxy* ou um servidor de redirecionamento pode usar qualquer meio para determinar o servidor *next-hop*, incluindo executar programas e consultar banco de dados. Um servidor *proxy* pode também duplicar uma requisição, enviando cópias para múltiplos servidores *next-hop* de uma vez. Isto permite que uma requisição de início de chamada tente diversas localizações diferentes ao mesmo tempo. A primeira localização que responder é conectada com o cliente chamador.

Em suma, redireciona requisições e respostas, enviando uma mensagem para os clientes como novo endereço SIP procurado, e não fazendo o papel de continuar a chamada.

A arquitetura pode apresentar ainda os seguintes componentes:

- SIP Register Server

Servidor SIP que suporta requisições REGISTER, usadas para registrar as informações dos usuários em algum Servidor de Localização.

- Servidor de Localização

Possui apenas as funcionalidades de armazenamento e consulta de registros de usuários SIP neste servidor são descritas, ficando a critério do implementador da solução SIP a escolha da melhor tecnologia para esta finalidade.

3.2.2 Sinalização no SIP

O protocolo SIP é baseado no HTTP e, assim como este, suporta o transporte de qualquer tipo de carga em seus pacotes, pelo uso de Mime-Types (Multipurpose Internet Mail Extensions). O SIP funciona numa arquitetura cliente/servidor, e suas operações envolvem apenas métodos de requisição e respostas, como observado também no HTTP e no RTSP. Os métodos de requisição do SIP são os seguintes:

- INVITE: Indica que o usuário está sendo convidado a participar de uma sessão multimídia. O corpo da mensagem pode conter uma descrição da sessão, utilizando-se o protocolo de descrição de sessão SDP (*Session Description Protocol*).
- ACK: Mensagem recebida como resposta final a um INVITE. A requisição ACK pode conter o SDP de descrição da sessão negociada

entre ambos os clientes. Se não contiver o SDP, o usuário chamado pode assumir a descrição dada pelo primeiro INVITE, se houver.

- **OPTIONS:** Faz uma pergunta sobre quais métodos e extensões são suportados pelo servidor e pelo usuário descrito no campo de cabeçalho <To:>. O servidor pode responder a esta pergunta com o conjunto de métodos e extensões suportado pelo usuário e por ele mesmo.
- **BYE:** Usado para liberar os recursos associados a uma ligação e forçar a desconexão da mesma.
- **CANCEL:** Cancela uma requisição que ainda esteja pendente, ou seja, em andamento. Uma requisição é considerada pendente, se e somente se, ela não foi atendida com uma resposta final.
- **REGISTER:** Um cliente usa este método para registrar o "alias" (apelido) do seu endereço em algum servidor SIP, que, por aceitar registro de usuários, chamamos de serviço REGISTRAR.

Para cada requisição ou resposta, temos um grupo de cabeçalhos, divididos em: cabeçalhos gerais, com informações importantes sobre a chamada; cabeçalhos de entidade, com meta-informação sobre o corpo da mensagem; e os cabeçalhos específicos, que permitem passar informações adicionais, que não couberam na linha de status da requisição ou da resposta.

As entidades SIP se comunicam usando “transações”. Quando requisições (transações) são atendidas, as respostas enviadas são identificadas por números, que significam a classe da resposta. Pode-se enviar diversas mensagens provisórias antes de se enviar uma resposta definitiva. Existem seis classes possíveis de resposta: Classe 1XX, respostas temporárias ou informativas; Classe 2XX, resposta final de sucesso; Classe 3XX, redirecionamento da requisição; Classe 4XX, erros no cliente; Classe 5XX, erros do servidor; e Classe 6XX, erros globais na rede.

O iniciador de um pedido SIP é chamado de cliente SIP e a entidade que responde é chamada de servidor SIP. As mensagens trocadas durante uma transação compartilham um número Cseq (campo de cabeçalho) comum, com uma exceção: a mensagem ACK usa o mesmo Cseq da transação à qual ela se aplica, mas é uma transação diferente.

A primeira etapa consiste em abrir uma conexão de sinalização entre os pontos de origem e de destino da chamada. Pontos finais SIP podem usar sinalização UDP ou TCP, a sintaxe das mensagens é independente do protocolo de transporte usado.

Quando se usa o TCP, a mesma conexão pode ser usada para todos os pedidos e respostas SIP (não par os dados de mídia) ou uma nova conexão TCP pode ser usada para cada transação. Se o UDP for usado, o endereço e a porta a serem usados para as respostas a pedidos SIP estarão contidos no parâmetro de cabeçalho ‘via’ do pedido SIP. As respostas não devem ser enviadas para o endereço IP do cliente. Se nenhuma porta for especificada no endereço SIP, a conexão é feita com a porta 5060 tanto para o TCP quanto para o UDP.

Um cliente SIP chama um outro ponto final SIP enviando uma mensagem de pedido *invite* (Figura 3.8). A mensagem *invite* normalmente contém informações suficientes para permitir que o terminal de destino estabeleça imediatamente a conexão de mídia solicitada com o ponto de origem da chamada. Essas informações incluem as capacidades de mídia que o ponto de origem da chamada pode receber (e enviar: considera-se que a capacidade do codificador estão enviando ou recebendo no SIP, a menos que os parâmetros SDP *sendonly* ou *reconly* sejam usados) e o endereço de transporte onde o ponto de origem da chamada espera que o ponto de destino da chamada envie esses dados de mídia. A maioria dos pontos finais serão capazes de receber muitas codificações diferentes para cada tipo de mídia. A codificação em particular escolhida pelo remetente aparece como parte do cabeçalho RTP.

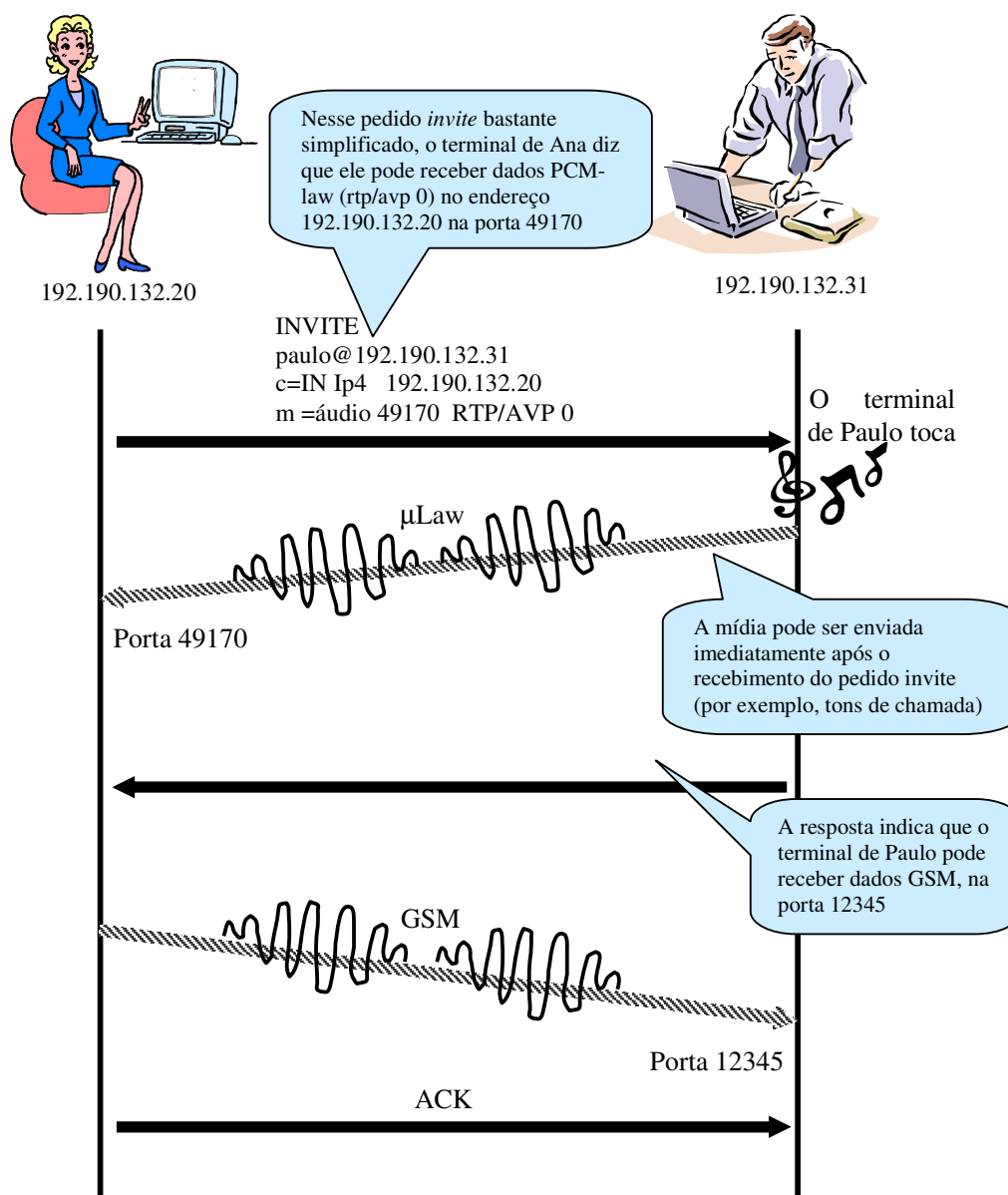


Figura 3.8 – Chamada para um endereço IP conhecido com SIP

O ponto de destino da chamada precisa indicar que ele está aceitando o pedido. Esse é o objetivo da mensagem de resposta OK. Uma vez que o pedido foi um convite, a resposta OK também contém as capacidades de mídia do ponto de destino da chamada e onde ele está esperando receber os dados de mídia. O originador da chamada precisa confirmar que recebeu de maneira adequada a resposta do ponto de destino da chamada (lembre-se que podemos estar usando UDP) com a mensagem ACK.

A partir desse exemplo, pode-se notar que o SIP é eficiente, pois o canal de mídia do receptor para o originador da chamada pode ser estabelecido com uma ida e uma volta de mensagens e o canal de mídia do originador para o receptor pode ser estabelecido com uma ida, uma volta e outra ida de mensagens. Isto é muito melhor que as várias idas-e-voltas exigidas pelo H.323 v1.

Na chamada anterior, o terminal de Ana ofereceu-se para receber um canal de áudio codificado em PCM. Isto pode não ser aceitável para o terminal de Paulo, ou porque Paulo não dispõe de largura de banda suficiente (o PCM exige 64Kbit/S, mais o *overhead* RTP/UDP/IP/PPP) ou porque o terminal de Paulo não possui um codificador PCM m-law. Nesse caso, o terminal de Paulo responderá com um 606 *Not Acceptable* e finalmente relacionará o conjunto de codificadores que ele pode usar. Com essa informação, o terminal de Ana pode enviar um novo pedido de *invite*, com o mesmo identificador de chamada, anunciando o código apropriado (mas se ele tivesse essa capacidade, poderia tê-la enviado como uma escolha no primeiro pedido *invite*) ou então reiniciar uma chamada por meio de um *proxy* de transcodificação (Figura 3.9).

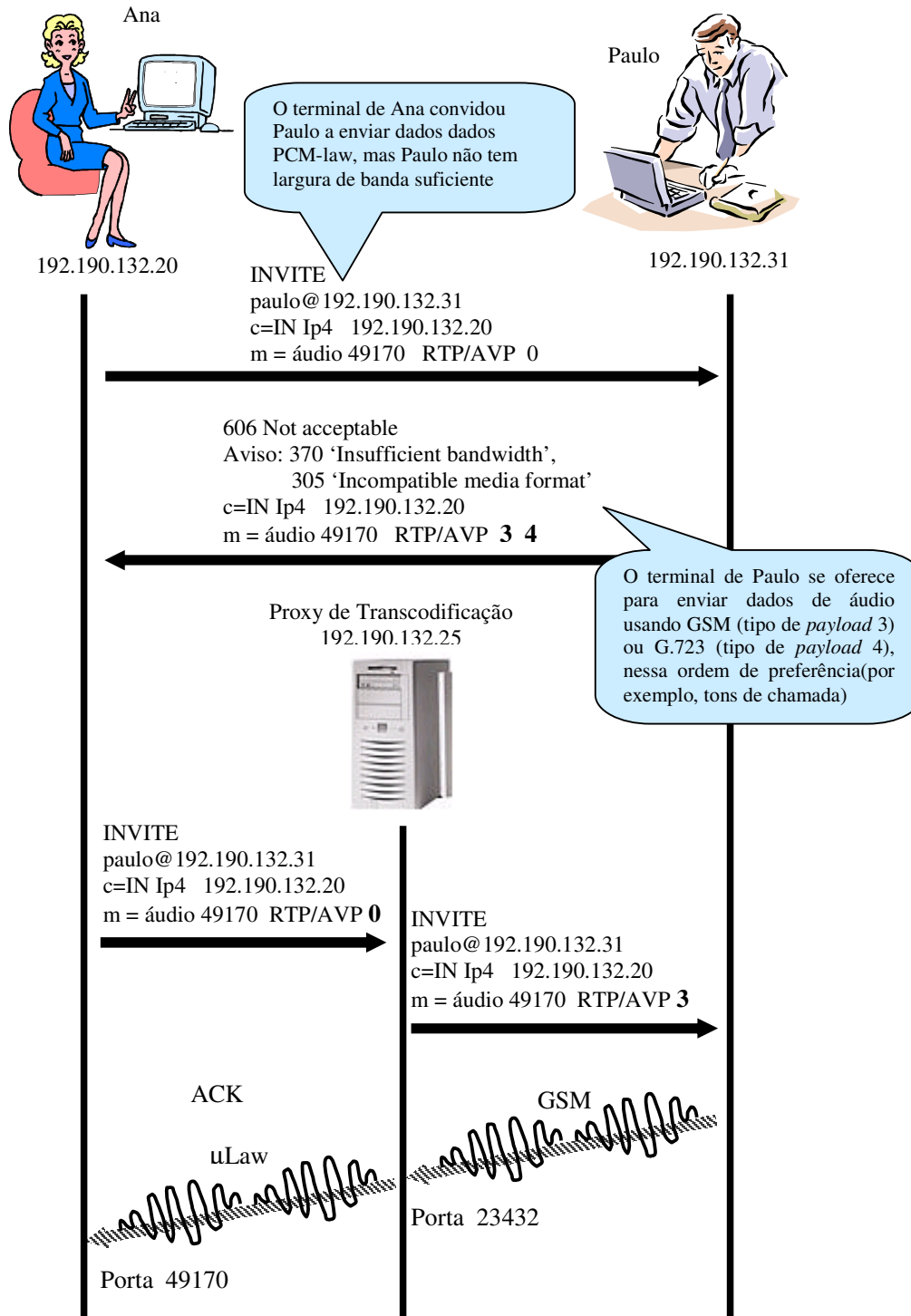


Figura 3.9 – Negociando o estabelecimento da chamada

O IETF concluiu sobre um codificador padrão a ser usado no caso de voz a baixa taxa de bits, o G.723.1, de maneira a manter mínima a probabilidade desse tipo de incompatibilidade no H.323. Nenhuma recomendação desse tipo existe para o SIP, mas a maioria dos terminais parece ser capaz de receber PCM A-law e m-law e também GSM.

O SIP não tem uma noção de canais lógicos (como definido no H.323). Quando um cliente se oferece para receber vários tipos de mídias em várias portas UDP ou TCP, ele tem de estar preparado para receber mídia imediatamente em qualquer um dessas portas. No entanto, o terminal de destino pode decidir enviar dados em apenas algumas portas (por exemplo, ele não possui capacidades de vídeo e envia apenas áudio). Nada na sinalização informa ao cliente se uma determinada porta estará ativa ou não. Em geral, isso não é realmente um problema, uma vez que a maioria dos pontos finais podem ficar “escutando” as portas não usadas sem um impacto significativo no desempenho.

Em alguns casos, no entanto, as entidades SIP precisam manter vários canais de mídia e têm de reutilizar as portas UDP da maneira mais eficiente possível, este é o caso de grandes *gateways* ou recursos de mídia centralizados. Essas entidades podem ter de multiplexar vários canais de mídia em uma única porta ou fechar portas inativas com base em heurísticas, por exemplo, após um período sem nenhuma atividade.

3.3 Relação entre as soluções H.323/SIP

Existe a idéia de que as soluções de videoconferência podem se interoperarem, essa discussão inicia pelo protocolo RTP. RTP é um protocolo ao nível de aplicação, não orientado a conexão que opera tanto sobre UDP/IP quanto TCP/IP. O RTP é utilizado principalmente pelos protocolos de sinalização H.323 e SIP (Figura 3.10).

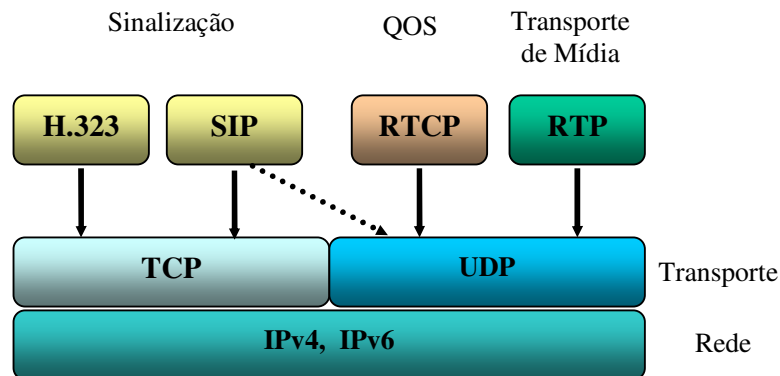


Figura 3.10 – Pilha de protocolos

A solução H.323 utiliza RTP como meio padrão para transporte de dados multimídia durante uma conferência, juntamente com RTCP, para fazer uso dos serviços de numeração de seqüências, detecção de erros, além de correção de erros apropriadas para cada tipo de mídias. O protocolo de iniciação de sessão (SIP) faz uso de RTP para transporte de dados multimídia durante uma conferência, e a parte de controle RTP também pode ser utilizada durante a transferência de dados em tempo real

numa conferência de áudio ou vídeo, já que este oferece um tipo de mensagem que identifica os participantes de uma conferência, através de e-mail, nome, telefone entre outras informações. Com isso, um participante da conferência pode convidar outro para uma conferência separado. Ambos os protocolos podem utilizar-se de RSVP para realizar reserva de recursos em roteadores e *hosts*, como por exemplo largura de banda e outros requisitos que aplicações em tempo real normalmente exigem para transmissão com qualidade de serviço [LUN 01].

As duas soluções podem utilizar-se de RTSP para armazenar parte de uma conferência num servidor de áudio ou vídeo ou recuperar *streaming* de áudio e vídeo sob demanda, no caso dos participantes desejarem assistir um vídeo durante a conferência. O último protocolo de sinalização, RTSP, também utiliza-se de RTP para transporte dos *streams* de mídias dos servidores de mídia para os clientes, e RTCP para oferecer informações sobre sincronização e a quantidade dos dados que vai ser enviada ao cliente, juntamente com informações de perda de pacotes para o servidor. Aqui também o protocolo RSVP pode ser utilizado para reservar recursos da rede.

Os protocolos de sinalização também relacionam-se entre si, SIP com RTSP e H.323 com RTSP, principalmente durante uma conferência, conforme já mencionado. SIP e H.323 também podem ser utilizados juntos; SIP localizando um usuário para uma conferência, inclusive um terminal H.323 e determinando suas capacidades (e. g., receber áudio e vídeo) e H.323 para estabelecer e terminar uma conferência. Contudo, ainda não existe uma especificação que realize a tradução entre os dois protocolos e a conversão é difícil de acontecer devido as múltiplas versões de H.323. Cabe ressaltar contudo, que SIP e H.323 oferecem serviços equivalentes como controle de chamadas, estabelecimento e liberação de conexões e são utilizados principalmente para aplicações de telefonia e videoconferência sobre a Internet. Os protocolos HTTP e SDP são também utilizados por RTSP, sendo responsáveis pelo armazenamento das descrições de apresentações das mídias de áudio e vídeo [LUN 01].

3.3.1 Estabelecimento da chamada

Para o estabelecimento de chamada nos dois protocolos, deve-se trocar, entre os terminais clientes, três informações: o endereço destino da sinalização, as capacidades de mídia sejam elas remotas ou locais, e os endereços e portas de transporte remotas ou locais de onde os terminais pretendem receber os pacotes de mídia. A dificuldade em traduzir este procedimento entre o SIP e H.323 ocorre porque, no SIP, todas estas informações estão contidas na mesma requisição SIP INVITE e respostas subsequentes, enquanto que, no H.323, estas informações estão espalhadas entre diversas mensagens de sinalização diferentes [INT 02].

A solução deste problema é formada por duas partes. Na direção SIP para H.323, deve-se dividir as informações que vieram numa única mensagem SIP INVITE, e passar estes parâmetros nas sinalizações correspondentes de H.225/H.245. Na direção H.323 para SIP, o processo é mais complicado e exige que todas as mensagens H.225 e H.245 sejam reunidas pelo lado H.323. Entretanto, o H.245 só é ativado quando a chamada já foi estabelecida pelos procedimentos do Q.931/H.225, de forma que as alternativas são: ou mandar um INVITE vazio (sem SDP e sem conter a descrição de mídia) para estabelecermos a conexão (requisito primário do H.245), ou usar uma mensagem

OPTIONS para o lado da “rede SIP”, para descobrir as capacidades do cliente SIP. Em seguida, repassar para o lado H.323 e poder negociar com o H.245 [DAL 02] [INT 02].

3.3.2 Conversão de endereços

No SIP, os endereços são formados como URIs (assim como os endereços HTTP). Por exemplo, sip: user@host. Os clientes SIP podem suportar esquemas de URL como “tel:xxx@” para números de telefone ou mesmo eventuais URLs H.323, se existirem [RFC2543].

O H.323, seus endereços são codificados em ASN.1 (no SIP, tanto a codificação das mensagens quanto a URI é baseada em texto) e podem ter várias formas, como identificadores não-estruturados (H.323-ID), E.164 para números telefônicos, URLs de diversos tipos, nome de host ou endereços IP, e endereços de e-mail (e-mail-ID). No entanto, os nomes de usuários locais e host parecem ser os mais comuns na formação do endereço H.323 [DAL 02] [INT 02].

Para realizar o mapeamento de um endereço SIP para H.323 deve-se copiar o SIP-URI para o H.323-ID. As partes para montagem do e-mail-ID são tiradas do formato user@host do SIP-URI: o *transport-ID* da parte *host* do SIP-URI, e o campo E.164 extraído da parte user do endereço SIP, se ele estiver com um número de telefone ou esquema “tel:”. Se não estiver, a tradução é mais difícil devido às várias representações do H.323 e posterior união destas num único endereço SIP [DAL 02].

3.4 Comparação entre as propostas

Baseado no estudo apresentado das soluções de videoconferência H.323 e SIP, pode-se destacar algumas comparações em ambas as soluções.

3.4.1 Funcionalidades que o SIP faz e o H.323 não realiza

- Velocidade

Pode ser notado pelo número de transações, o SIP utiliza uma transação ao passo que o H.323 versão 1 realiza em quatro ou cinco trocas de mensagens, cada uma delas especificada em documentos ITU diferentes. Além disso, o SIP pode usar UDP, ao passo que o H.323v1 e 2 têm que usar o TCP. A combinação dessas diferenças resultou em um tempo de configuração muito menor para os pontos finais SIP. Isso inspirou diretamente algumas modificações introduzidas no H.323v2. Outras melhorias vão surgindo com o passar do tempo, principalmente a questão de usar sinalização UDP.

- Multicast

O SIP foi projetado para funcionar em *backbones* com capacidade para *multicast*, não apenas para os fluxos de mídia, como o H.323, mas também para as mensagens de sinalização [RFC2543]. Por exemplo, uma mensagem *invite* pode ser enviada para um grupo *multicast*. Isso é útil para *call centers* ou quando

se quer encontrar uma pessoa em uma empresa. As versões 1 e 2 do H.323 precisam usar *multicast* para o mesmo fim.

- Uso de URLs

Primeiramente, pode-se notar não haver uma grande diferença entre um alias de e-mail H.323 (joao@nome.com) e uma URL SIP (sip:joao@nome.com). Porém, existe uma diferença: um alias de e-mail H.323 considera que o protocolo usado seja o H.323, ao passo que o SIP especifica ele mesmo o protocolo na URL. Deste modo, um servidor SIP pode redirecionar uma chamada para servidores não SIP. Um terminal SIP, quando recebe uma chamada de um outro terminal SIP, pode redirecionar a chamado para uma página Web ou para uma URL. Isso facilita a integração de aplicações de áudio e vídeo com outras aplicações multimídia.

- Priorização de chamadas

O campo de cabeçalho *priority*, trata-se de um campo adicional que foi desconsiderado no H.323 [RFC2543].

- Codificação de texto

A codificação possui uma vantagem: pode ser depurada usando *sniffers* de rede, o que faz com que problemas de interoperabilidade sejam detectáveis “visualmente” [RFC2543].

O problema pode ser desempenho e tamanho, como alguns sustentam que protocolos binários são mais fáceis de codificar e decodificar. Com certeza o tamanho das mensagens de texto são maiores que o tamanho das mensagens codificadas em binário.

3.4.2 Funcionalidades que o H.323 faz e o SIP não realiza

- Canais lógicos

O H.323 faz uma distinção clara entre os tipos de mídias que podem ser enviados ou recebidos e as combinações que podem ser válidas por um lado (capacidades) e os tipos de mídia que estão ativos e, de fato, enviados par a rede (canais lógicos) por outro lado. Segundo [RFC2543], o SIP não tem uma distinção como esta, uma vez que os pontos finais SIP divulgam apenas os codificadores que eles podem receber e não há nenhum procedimento par abrir uma conexão de mídia em separado do ato de efetivamente enviar a mídia. Isso simplifica a sinalização e pode parecer uma vantagem à primeira vista.

No entanto, em alguns casos, especialmente quando se implementam servidores que fazem uso intensivo de recursos, isso pode ser um problema, porque toda vez que o servidor divulga uma capacidade, ele precisa criar um soquete de escuta. Ao contrário, um cliente H.323 precisa abrir um soquete apenas quando ele recebe uma mensagem [DAL 02]. Esse comportamento do

SIP pode resultar em vários soquetes inativos e como a maioria dos codificadores de voz implementam detecção ativa de voz e podem estar inativos mesmo se os terminais estiverem querendo enviar dados de mídia. Porém, a tarefa de fechar os soquetes não se trata de uma tarefa simples.

- Codificação binária

As mensagens H.323 são codificadas de acordo com o Q.931 para o subconjunto de mensagens H.225 provenientes do Q.931. Todas as mensagens, bem como as extensões H.323 para mensagens Q.931, são codificadas usando-se regras de codificação de pacotes (*Packet Encoding Rules – PER*) ASN.1 (*Abstract Syntax Notation 1*). Segundo [DAL 02], o fato de misturar dois métodos de codificação com regras totalmente diferentes não é a melhor das idéias. Isso implica também grandes esforços de programação por parte das empresas que ainda não têm uma implementação Q.931 e um compilador ASN.1.

Finalmente, depurar os problemas de interoperabilidade entre os terminais H.323 exige monitores de rede com capacidades de decodificação tanto Q.931 quanto PER ASN.1 [DAL 02].

Existem algumas vantagens de se utilizar codificação binária:

- tamanho do PDU é otimizado;
- (talvez) o desempenho é melhor;

O tamanho dos PDUs de sinalização não é, em geral, um problema em redes modernas. No entanto, se o H.323 fosse usado em redes sem fio, isso poderia ser uma vantagem [INT 02].

- Descoberta dos gatekeepers

Atualmente a questão da descoberta de *gatekeepers multicast* parece melhor definida no H.323, permitindo que um terminal saiba qual *gatekeeper* está respondendo.

Conclusão

Nos dias atuais a tecnologia de videoconferência é uma realidade. Pode-se constatar que a cada dia, mais usuários têm demonstrado interesses em fazer parte destes recursos. Partindo então para escolha de qual solução pode satisfazer suas expectativas. Baseado nisso algumas soluções de videoconferências foram definidas atualmente como o H.323 do ITU-T e ainda em fase de definição o *Session Initiation Protocol* (SIP) da IETF. Buscando assim, que a cada vez mais novos adeptos possam

O H.323 é um padrão importante para comunicação de dados, áudio e vídeo. O protocolo descreve especificamente como a comunicação multimídia acontece entre terminais, equipamentos de rede, e serviços em redes locais (LANs), as quais não oferecem uma qualidade de serviço garantida (por exemplo redes IP). Atualmente a recomendação H.323 está sendo amplamente aceita devido ao suporte oferecido pela união global de fabricantes de computadores, sistemas de comunicação e vendedores de sistemas operacionais, sendo utilizado para videoconferência, computação colaborativa, aprendizado a distância, suporte e ajuda a aplicações de leitura, shoppings interativos, chamadas para conferências de negócios, entre outros. A solução é utilizada para estabelecer chamadas e iniciar conferências em grupos de tamanho moderado

A outra solução de videoconferência é o SIP, que é baseado no HTTP e, assim como este, suporta o transporte de qualquer tipo de carga em seus pacotes, pelo uso de *Mime-Types* (*Multipurpose Internet Mail Extensions*). O SIP funciona numa arquitetura cliente/servidor, e suas operações envolvem apenas métodos de requisição e respostas, como observado também no HTTP e no RTSP.

Este trabalho propôs um estudo das soluções de comunicação e colaboração visual, onde a finalidade é contribuir sobre forma de uma tutorial para um maior conhecimento sobre as duas soluções de videoconferência, podendo assim buscar que os padrões se interoperem futuramente e ocorra com isso, um aproveitamento maior de recursos e aplicações já desenvolvidas tanto para H.323 como SIP.

Bibliografia

- [DAL 02] DALGIC, I; FANG, H. Compararison of H.323 and SIP for IP Telephony Signaling. Disponível via WWW em: http://www.cs.columbia.edu/~hgs/papers/others/dalg9909_comparisn.pdf, Março de 2002.
- [INT 02] Internet Engineering Task Force. Interworking Between SIP/SDP and H.323. Disponível via WWW em: <http://alternic.net/drafts/drafts-s-t/draft-singh-sip-h323-01.html>, Março de 2002.
- [ITU 92] INTERNATIONAL TELECOMMUNICATION UNION. ITU-T Recommendation G.711 – Pulse Code Modulation (PCM) of Voice Frequencies. Fascicle III.4. ITU, 1993, 9p.
- [KOR 01] KORB, Alexei, Colaboração visual: Implementando o trabalho cooperativo com H.323 (uma visão geral do H.323). Trabalho Individual II, CGPC-UFRGS, Porto Alegre, 2001.
- [LEO 02A] LEOPOLDINO, G.; Modelos de Comunicação para Videoconferência. Disponível via WWW em: <http://www.rnp.br/newsgen/0105/video.shtml>, Março de 2002.
- [LEO 02B] LEOPOLDINO, G.; MOREIRA, E.; Avaliação de Sistemas de Videoconferência. Disponível via WWW em: <http://www.rnp.br/arquivos/AvaliacaoVideo.pdf>, Março de 2002.
- [LUN 01] LUNARDI, Sediani.; Protocolos de suporte a integração de serviços. Disponível via WWW em: <http://www.inf.pucrs.br/~sediane/Trabalho%20Individual%20HTML.htm>, 2001.
- [MEL 97] MELCHORS, Cristina. Sistemas Interpessoais de Videoconferência (MBONE). Trabalho Individual I, CGPC-UFRGS, Porto Alegre, 1997.
- [PIN 02] PINHEIRO, Cássio, Tutorial – Especificação ITU H.323. Disponível via WWW em <http://penta2.ufrgs.br/h323/indroducao.html>, Março de 2002.
- [PRO 02] Protocolos IETF para a Telefonía IP / SIP - Session Initiation Protocol. Disponível via WWW em: <http://www.lecom.dcc.ufmg.br/~sergiool/telefonía/protocol.htm>, Março de 2002.
- [RFC2543] HANDLEY, M.; SCHULZRINNE, H.; SCHOLEER, E. and ROSENBERG, J. SIP: Session Initiation Protocol, Internet Engineering Task Force, RFC 2543, march 1999

- [WHA 02] What is SIP?. Disponível via WWW em: http://www.radvision.com/papers/c1_what_is_sip.html, Março de 2002.
- [ZAN 01] ZANIN, FÁBIO A, Um modelo para videoconferência sobre redes IP. Erechin: Edifapes, 2001.
- [ZAN 99] ZANIN, FÁBIO A, O ambiente de videoconferência sobre Redes ATM, Trabalho Individual I. Porto Alegre, 1999