

420

现代信息网络实用技术讲座

视频会议系统及其应用

主 编：陶智勇 廖云霞
副主编：郭德霈 刘清堂 王大为
编 著：张江山 鲁 平

北京邮电大学出版社
• 北京 •

内 容 简 介

多媒体技术的研究和应用是目前信息技术领域最热门的话题,视频会议系统便是多媒体技术和网络技术的一项重要应用成果,是近年来发展起来的最先进的将图像、声音等资料进行异地传输的会议系统设备。

本书阐述了视频会议系统的最新发展,并全面介绍了视频会议系统的原理及应用,注重理论,联系实际,突出实用性。

全书内容包括:概述;视频会议的网络环境及交换技术;数字视频和音频;数据压缩技术;视频会议系统结构;视频会议系统国际标准;典型视频会议系统。

图书在版编目(CIP)数据

视频会议系统及其应用/张江山,鲁平编著. —北京:北京邮电大学出版社,2001

(现代信息网络实用技术讲座)

ISBN 7-5635-0570-9

I. 视… II. ①张… ②鲁… III. 视频会议系统 IV. TN948.63

中国版本图书馆 CIP 数据核字(2001)第 086990 号

视频会议系统及其应用 (现代信息网络实用技术讲座)

编 著 张江山 鲁 平

责任编辑 蒋 亮

*

北京邮电大学出版社出版发行

新华书店北京发行所发行 各地新华书店经售

北京源海印刷厂印刷

*

850 mm×1 168 mm 1/32 印张 4.75 字数 145 千字

2002 年 1 月第 1 版 2002 年 1 月第 1 次印刷

印数:1—3 000 册

ISBN 7-5635-0570-9/TN·262 全套定价:40.00 元(本册:10.00 元)

前 言

自工业革命后,科技的发达使得通信技术有了突破性的发展,电话的发明,使异地的人们可以传送声音。然而,除了言语的应用外,人类更重视的是表情及肢体的表达,声音的传送已经无法满足现代人沟通的需求,立即且互动的影像能更真实地传送信息。视频会议系统正是在这种现代人迫切需要的推动下开发成功的新一代通信产品。

视频会议系统通过各种通信网络及相关的多媒体设备,除了可将声音、影像信息互传,达到即时且互动的沟通外,并提供如共享白板、软件共享使用等功能。视频会议系统旨在为分散于不同地区的多个用户提供一个很好的讨论环境,使人们足不出户就可与远在天涯的朋友及客户进行各种交互活动。

视频会议系统有着广阔的应用前景,除了用于商业会议之外,还可使商务交流、企业客户服务和产品开发、远程教学和技术培训、市场调查和情报收集、远程医疗和会诊、科研合作和工程设计、跨国企业沟通以及招募员工等活动受益。总之,视频会议系统无论应用于什么场合,出于何种目的,其最终也是最直接的效益是节省开支,提高效率。

全书分为7章,第1章介绍视频会议的类型、现状、关键技术及发展方向等;第2章介绍视频会议系统中有关的网络交换技术;第3章介绍数字视频和音频的基本概念;第4章介绍数据压缩技术,包括音频、视频压缩原理及相关国际标准;第5章介绍视频会议系统的结构,包括系统结构、终端结构和组网结构;第6章介绍视频会议系统的相关国际标准;第7章介绍典型的视频会议系统,其中详细阐述了微软的 NetMeeting 桌面视频会议系统的使用方法。

在本书的编写过程中,得到了很多人的鼓励、支持和帮助,在此谨向陶志勇先生、朱光喜教授表示感谢。

限于作者的学术水平,错误和不妥之处在所难免,敬请读者不吝赐教。

作 者
2001 年 8 月

目 录

第 1 章 概 述	1
1.1 引言	2
1.2 视频会议的类型	2
1.3 视频会议的发展及现状	5
1.4 视频会议的关键技术	7
1.5 视频会议的发展方向	12
第 2 章 视频会议的网络环境及交换技术	15
2.1 网络上的信息交换技术	16
2.1.1 线路交换	17
2.1.2 信息包交换	18
2.1.3 面向连接服务与无连接服务	20
2.1.4 网络分类	21
2.2 网际多目标广播	23
2.2.1 多目标广播的概念	23
2.2.2 多目标广播所需要的环境	25
2.2.3 多目标广播树的概念	25
2.2.4 IP 多目标广播的地址和主机组的管理	27
2.3 实时传输协议和实时控制协议	30
2.3.1 实时传输协议 RTP	30
2.3.2 RTP 信息包标题域	32
2.3.3 实时传输控制协议 RTCP	33
2.3.4 实时流放协议 RTSP	34
2.4 资源保留协议	35
2.4.1 资源保留协议 RSVP	35

2.4.2 不同种类的接收器	36
2.4.3 接纳测试	39
2.4.4 路径消息	39
第3章 数字音频和视频	41
3.1 数字音频	42
3.1.1 声音与听觉器官	42
3.1.2 声音信号数字化	43
3.2 数字视频	46
3.2.1 彩色电视制式	47
3.2.2 电视扫描和同步	48
3.2.3 彩色电视	50
3.2.4 彩色电视信号的类型	51
3.2.5 电视图像数字化	53
第4章 数据压缩技术	57
4.1 多媒体数据压缩依据	59
4.2 话音编码	60
4.2.1 话音编码技术	60
4.2.2 音频编码国际标准	64
4.3 视频编码	65
4.3.1 数字图像编码技术	65
4.3.2 视频编码技术	67
4.3.3 视频编码的国际标准	69
第5章 视频会议系统结构	75
5.1 系统结构	76
5.1.1 物理系统结构	76
5.1.2 逻辑系统结构	77

5.2 终端结构	78
5.3 组网结构	82
5.3.1 点对点组网结构	82
5.3.2 多点会议组网结构	83
5.3.3 MCU	85
5.3.4 多点会议控制模式	87
第 6 章 视频会议系统国际标准	91
6.1 T.120 数据会议序列标准	93
6.2 H.320 视频会议标准	96
6.3 H.323 视频会议标准	99
6.3.1 H.323 标准介绍	99
6.3.2 H.323 的主要优点	102
6.3.3 H.323 的体系结构	104
6.3.4 H.323 标准下的通信	107
6.3.5 H.323 的实现	109
6.4 H.324 可视电话标准	111
6.4.1 H.324 系统结构	112
6.4.2 H.324 特性	112
第 7 章 典型视频会议系统	121
7.1 Microsoft NetMeeting 的使用	122
7.1.1 系统安装环境	122
7.1.2 系统卸载	123
7.1.3 开始使用 NetMeeting	124
7.1.4 发出呼叫与接收呼叫	126
7.1.5 白板、闲聊与文件传送	131
7.1.6 音频通讯	133

7.2 其他视频会议系统简介.....	136
7.2.1 CLI 公司及主要产品	136
7.2.2 Intel 的 PPCV200	137
7.2.3 Picture Tel 公司的 LiveLAN	137
7.2.4 Creative 的 Share Vision PC3000	138
7.2.5 Apple 的 Quick Time	139
7.2.6 华中科技大学电子与信息工程系的桌面视频会议系统.....	140
7.2.7 清华大学计算机系的 MCONF(多媒体会议系统)	140
7.2.8 InSoft 的 Communique	141
7.2.9 Sun 的 ShowMe	141

第1章

概 述

1.1 引言

人类进入信息社会的一个重要标志是计算机网络的迅速普及。在网络应用的初始阶段,人们通常只是通过网络发送电子邮件、进行文件传输或远程登录等。随着网络技术与多媒体技术的发展,计算机网络已逐渐演变成丰富多彩的多媒体信息网。越来越多的多媒体信息可通过网络进行传输与发布,一大批方便实用的多媒体应用(WWW 浏览、音视频多媒体会议、远程教学、电子商务等)异军突起,成了网络应用的主力军。在众多多媒体应用中,实时多媒体会议是一个重要方面。它能广泛用于社会各行各业,特别是教育、行政、医疗、家庭娱乐等领域,极大地提高了人们的工作效率,方便人们的工作、学习和生活,具有广阔的市场前景和不可估量的社会效益及经济效益,因而也是当前研究和开发的热点。

1.2 视频会议的类型

视频会议系统作为目前最被看好的技术,对其进行科学的、系统的、全面的分类是相当重要的,将对其发展起到巨大的推动和指导作用。可以从视频会议系统通信网络、传输内容、终端配置以及媒体选择的不同等对其分类,下面将各种分类方法作一简要描述。

(1) 从通信网络(或传输介质)角度

这是一种最直观、最简单也是最通用的分类方法。支持视频会议系统的通信网络有很多,而且各种通信网络均有其各自独特的特性,从而导致了在不同通信网络上视频会议系统设计和部署的差异性。包括现有的和未来的,通信网络实际上有很多种,但从其结构的本质来分,只有几种,即:通用电话网(POTS: Plain Old Telephone Service)、局域网(LAN: Local Area Network)、综合业务

数字网 (ISDN: Integrated Services Digital Network)、异步传输网 (ATM: Asynchronous Transfer Mode)、因特网 (Internet)。这样就形成了 5 种视频会议系统,即基于 POTS、LAN、ISDN、ATM、Internet 的视频会议系统。

(2) 从传输内容角度

在实际的计算机会议系统中,根据不同程度的需求和目的,在网络中交互的会议内容也有极大的差别,这样形成了计算机会议系统的以下几种不同形式:

- 文件会议;
- 数据会议;
- 可视会议系统;
- 桌面视频会议系统。

文件会议 (Document Conferencing) 的特点是与会者共享屏幕上的一个或多个窗口,通过这些窗口交换信息。这样的窗口称为共享白板,用户在这个白板上进行交互式的讨论或对文件进行修改等。文件会议系统可以传输图文,但不能传递语音。

数据会议 (Data Conferencing) 是在文件会议系统的基础上,在相同的通信线路上增加同时传送声音的功能,这样就成为数据会议。

可视会议系统是在数据会议系统的基础上,再增加静态图像或准动态图像传输的功能,于是便构成了可视会议系统。

桌面视频会议系统 (Desktop Video Conferencing) 可以支持语音、视频、文本、图形等多种媒体,因此也称为多媒体会议系统 (Multimedia Conference System)。桌面视频会议系统是视频会议系统发展的方向。

从狭义角度讲,前两种会议系统并不属于视频会议系统。在此处这么处理的唯一目的是使读者对视频会议系统结构有更全面的了解和体会。

(3) 从终端配置角度

为了同时且实时地提供每个与会者的活动情况,从终端角度可将视频会议系统分为两种:

- 多窗口系统(Single Monitor Screen With Segmented Windows);

- 多监视器系统(Multiple Monitors)。

多窗口系统只需要一个监视器,每个会议场点(Conference Site)的活动情况只体现为一个窗口。这种系统的通信硬件成本和处理设备成本比较低,网络为每一次会议提供一个会议桥(Conference Bridge),该会议桥收集了从所有会议场点发送的音频和视频信息,并混合音频信号,合成视频信号,然后再将结果信号分发给每一个会议场点,由于有会议桥的存在,每个会议场点只需一个输入通道(Incoming Channel)便可接收到所有与会者的信息并在本地场点实时显示出来,不需进行复杂的视频和音频信号处理,从而降低了运行成本。

而多监视器系统则恰恰相反,不需要窗口技术,远端每一个会议场点的活动情况在本地场点都体现为一个单独的监视器,而且还需要若干输入通道来接收所有会议场点的活动信息,对视频、音频信号也要做较复杂的处理,不仅导致了通信硬件成本和处理设备成本的增加,而且还显著地增加了运行成本。

(4) 从媒体选择角度

为了优化网络连接,从媒体选择角度,可将视频会议系统分为两类:

- 媒体可选系统(Selectable Media Conference);
- 媒体固定系统(Common Media Conference)。

对于媒体可选系统,每一个与会者(或会议场点)均有权选择(或授权选择和限制)在本地所需观察的特定场点的活动情况,这样,呈现在每个场点面前的会议活动情况是不尽相同的。在这种系统中,每个会议场点都需要一个特制的输入通道来接收外来信息,而且会议桥必须有足够的处理能力(或潜力)为每个会议场点处理各不相同的视频音频信息数据(包括合成视频信号、混合音频信号等功能),从而导致了成本的上升,但却明显地增加了灵活性,有利于集中管理和授权控制,为各种存取保密、私有数据保密技术

的实现提供了很强的硬件和软件设计基础。

而在媒体固定系统中,呈现在每个与会者面前的会议活动情况都是相同的,因此,会议桥只须将合成的视频信号和音频信号多目的(Multicasting)发送给所有会议场点,比媒体可选系统需要更少的带宽便可完成会议目的。虽然成本低,但灵活性很差,使各种保密技术的实现只能在高层协议中完成,且使这种实现复杂化,难于维护 and 设计。

(5) 根据与会者参加的方式

根据与会者参加的方式,视频会议系统可分为以下四种:

- 单用户系统;
- 拨号群组系统;
- 点到点系统;
- 多点可视系统。

此外,有时也将视频会议系统分为室内型会议系统和桌上型会议系统,或分为预先安排型(PreScheduled)和即时召开型(On-Demand)等。实际产品化的视频会议系统,是以上几种分类的综合体,如拨号群组系统必须设计为“预先安排型”系统,否则无法工作。

1.3 视频会议的发展及现状

和其他许多事物的发展一样,视频会议的发展也经历了一个从无序到有序、从不成熟到基本成熟的过程。在这个发展过程中,相继出现了电视会议、桌面视频会议、多媒体会议等多种远程会议系统。电视会议又分为模拟系统和数字系统两种,模拟电视会议系统传送的是模拟音视频信号,因而需要铺设专用的模拟传输线路,不仅造价昂贵,而且模拟信号的多点切换与控制手段比较单一,不利于系统的扩容与级联;数字电视会议系统则是租用 DDN 信道,通过多点控制单元(Multipoint Control Unit)进行多点交换,控制相对比较灵活,而且便于实现主、分会场的多级连接。目前我国已经建成了以北京为主会场,

下设 9 个分会场,覆盖全国 29 个省市的全国数字电视会议网络。这两种电视会议系统的共同缺点是费用昂贵,一般只有政府高级官员和重要部门能够使用,同时系统的功能比较单一,除了进行音视频多点传输和简单的会议控制外,很难增加新的功能。桌面视频会议系统是面向广大机关企事业单位、组织和个人的较理想的远程会议工具,其优点是价格便宜、带宽利用率高、接入方式灵活(PSTN、ISDN、LAN、Internet、虚拟专网 VPN 等)、具有互操作性以及便于升级扩充等等,因而在最近几年里得到了飞速的发展。多媒体会议是在桌面视频会议的基础上增加多媒体支持特性而形成的,目前对于多媒体会议还没有一个统一的定义和标准,也有人将桌面视频会议称作多媒体会议,因为许多桌面视频会议系统支持一些简单的多媒体特性(如电子白板、文字交流、文件传输、应用程序共享等)。多媒体会议系统虽然比普通视频会议系统有一定的进步,但仍然存在较大的局限性,只能召开一般意义上的互相交流信息的会议,而不能让某一群体用户合作完成一项共同的工作,或者说,多媒体会议系统的协同性不够。

在多媒体会议系统的发展过程中,音视频编码技术作为其中的关键技术之一起到了极大的推动作用。1990 年 CCITT 第 15 研究组制定了针对活动图像的 $P \times 64$ kbps 的编解码器协议 H.261 建议之后,视频压缩编码技术开始走向标准化和实用化,一批符合 H.261 建议的专用芯片和多媒体会议产品(大多基于 ISDN)相继问世。五年之后,该研究组又提出更低比特率的视频编解码方案 H.263 建议。该建议可将视频图像最少压缩到大约 20 kbps,可在普通电话线上通过 28.8 kbps 的 V.34 Modem 传送音视频信号。音频编码建议则从早先的 G.711、G.722 建议发展到以后的 G.723.1、G.728、G.729 等。在音视频编码协议不断改进与发展的同时,多媒体会议本身的协议也实现了更新换代。从基于 ISDN 环境的 H.320 建议到基于分组交换网的 H.323 建议,再到 PSTN 的 H.324 建议,另外还有 H.321 建议(B-ISDN 环境下的视频会议)、H.322 建议(等时以太网环境下的视频会议)。表 1-1 列出了这些建议的基本概况。多媒体会议发展到现在,已经有许多成熟的系统,如 NetMeeting、

CUSeeMe、PictureTel、Proshare 等等。

表 1-1 几种不同的多媒体会议国际建议

	H.320	H.321	H.322	H.323	H.324
适用网络	N-ISDN	B-ISDN	等时以太网	普通 LAN	PSTN
比特率 (bps)	<2 M	<600 M	<6/16 M	<10/100 M	<28.8 K
音频编码 协 议	G.711,G.722, G.728	G.711,G.722, G.728	G.711,G.722, G.728	G.711,G.722, G.728,G.723.1, G.729	G.723.1
视频编码 协 议	H.261	H.261,H.262	H.261	H.261,H.263	H.263
数据协议	T.120	T.120	T.120	T.120	T.120
控制协议	H.245	H.242,H.245	H.242	H.245	H.245
复用协议	H.221	H.222.0,H.222.1	H.221	H.225.0	H.223

1.4 视频会议的关键技术

1. 高速多媒体通信网络及多媒体传输协议

现代计算机网络与通信技术是多媒体会议的基础,但是目前的广域网环境及 TCP/IP 协议并不适合传送实时多媒体数据。因此,要开发全球范围的高性能多媒体会议系统首先必须建立高速的多媒体通信网络,同时配以新型的实时多媒体传输协议。新的传输协议应当提供全套服务质量(QoS)参数,包括流通量、端对端延时、时延抖动、突发性、分组差错率(PER)、误码率(BER)等。在连接建立原语中,应当可以为每个 QoS 参数确定优先等级。事实上,早在 20 世纪 80 年代初就开始了在 Internet 上进行多媒体通信的研究,并提出了两种早期的多媒体通信协议:ST2 数据流协议版本 II 和 Tenet 协议。这时已经发现了 TCP/IP 协议传输多媒体数据的一些问题,并开始了对实时传输、资源预留、QoS 控制、多址广播等关键技术的研究,并在它们的试验协议中部分实现了这些技术。之后又提出了一些新的多媒体通信协议,如 RTP、XTP、

RSVP、IPv6 等,其中 IPv6 将从本质上提高了 IPv4 的性能,支持资源预订(与 RSVP 协议结合)和 QoS 控制,同时支持完整的 Multicast(内置 Internet 组管理协议 IGMP),因而将成为下一代 IP 技术的核心。

2. 多媒体数据压缩编码技术

多媒体应用走向实用化的一个前提是必须具备高效率的压缩编码方法。迄今为止,已经提出了大量编码技术,如在图像压缩方面的预测编码、变换编码、子带编码、小波编码、分形编码、模型基编码、矢量量化、运动估计等,在音频压缩方面的自适应差分脉码调制(ADPCM)、线性预测编码(LPC)、子带编码、熵编码、矢量量化等。评价编码技术优劣的准则主要有三条:压缩比、重现精度和压缩速度,另外还有抗干扰能力、同步能力、可伸缩性等。其中压缩速度在多媒体应用中显得尤为重要。目前在音视频编码方面已经制定了一些国际标准,如用于视频编码的 H.261、H.263、H.263+、MPEG-1、MPEG-2 建议。这些建议的算法核心均采用了 DCT 加运动估计的编码技术,H.261 是 $P \times 64$ kbps 的视频编解码建议,H.263 则是低比特率视频编解码建议,H.263+ 又在 H.263 基础上进行了改进,提出了一些新的帧内和帧间编码技术,如在帧内编码时对 8×8 方块的第一行(或第一列)DCT 系数进行预测(H.263 建议只对直流分量进行预测),对帧内和帧间编码采用不同的 VLC 码表,帧间编码时可在一组缓存图像中选择一帧作为参考帧进行运动估计,另外还提出了图像分层技术和改进的 PB 帧方式等。

在音频编码方面,相继出现了 G.711、G.722、G.723.1、G.728、G.729、G.729A 等建议。除了早期的 G.711、G.722 建议,近期提出的低比特率音频编码建议其核心都包括线性预测分析/合成技术(LPAS)。码激励线性预测(CELP)是最常用的 LPAS 方法,它可分成代数码激励线性预测(ACELP)、低时延码激励线性预测(LD-CELP)、共轭结构码激励线性预测(CS-CELP)等等。表 1-2 是几种建议的性能指标对照表。其中帧大小和先导时延之和即为编码器的缓存时延,复杂度决定了编码器的计算时延,这两项

指标决定了编码器的压缩速度。

从表中可以看出,G.728 的优点是低时延,它是 ISDN 视频会议系统的推荐语音编码标准。G.723.1 的优点是低码率,因而是 Internet 视频会议系统的推荐语音编码标准,但其缺点是时延太大。G.729 是目前很有前途的一种低比特率多媒体会议语音编码建议,它的时延和码率都较低,很适合在多媒体会议系统中使用。

表 1-2 几种多媒体会议语音编解码建议的性能指标对照表

标 准	核心编码技术	比特率(kbps)	帧大小/先导 时延(ms)	算法复杂度 (MIPS)
G.711	PCM	64	0/0	0
G.722	ADPCM	48,56,64	0.125/1.5	5
G.723.1	MP-MLQ* /ACELP	5.3,6.3	30/7.5	16
G.728	LD-CELP	16	0.625/0	30
G.729	CS-ACELP	8	10/5	20
G.729A	CS-ACELP	8	10/5	11

* MP-MLQ 即为 Multipulse Maximum Likelihood Quantization

3. 视频分层编码与传输技术(媒体缩放)

多媒体会议是面向群组的应用,由于群组中各成员的终端条件和接入速率不完全相同,它们对多媒体信号(尤其是视频信号)的分辨率要求也不尽相同。另外,在网络多媒体会议应用中,由于网络带宽的动态变化,要求多媒体会议系统采取相应的措施以适应这种变化,于是出现了视频分层编码与传输技术。在 H.263+ 和 MPEG-2 建议中都提供了视频分层技术,通常有三种分层方法:

(1) 时间分层:即时间域抽样,通过调整视频流的时间分辨率(帧率)来适应信道的变化。

(2) 空间分层:即空间域抽样,将各单帧图像分成不同空间分辨率的层次信息进行传输,通过调整图像空间分辨率来适应信道的变化。

(3) 信噪比分层(SNR Scalability):将各单帧图像分成不同信噪比的层次信息进行传输,通过调整图像信噪比来适应信道的变化。

信噪比分层通常是通过频域变换、引入各类量化噪声和截断误差来实现的,分层后的信息具有相同的时间分辨率和空间分辨率。

4. 群组通信

群组通信(Group Communication)也称多站点传输或点对多、多对多通信,即将数据同时传送到群组中的所有成员。在不同类型的网络环境中,群组通信的实现方式是不同的。在分组交换网中,群组通信主要是通过众多的路由器独立选路与转发来完成的。为了支持群组通信,需要对参与群组通信的主机和路由器进行功能扩展,以便有效地建立从发送方到接收群组各成员的多目标分组传送路径。图 1-1 是群组通信的体系结构示意图。其中主机扩展的目的是使主机具有收发多目标分组的功能,内容包括群组地址管理、群组成员管理以及多目标分组的发送和接收。路由器扩展的目的是使路由器具备转发多目标分组的能力,主要内容是多目标分发树的建立和维护。资源预定子层与可靠传输子层主要是给群组通信提供一定的服务质量,甚至提供可靠的群组通信服务。在主机与路由器扩展中,路由器的扩展是一个难点,因为建立多目标路由算法将给路由器带来相当大的负担。在 ISDN 等电路交换网中,群组通信是通过多点控制单元 MCU 的多点交换功能来实现的,因而群组通信问题就转化为 MCU 内部的交换网络结构和控制算法问题。

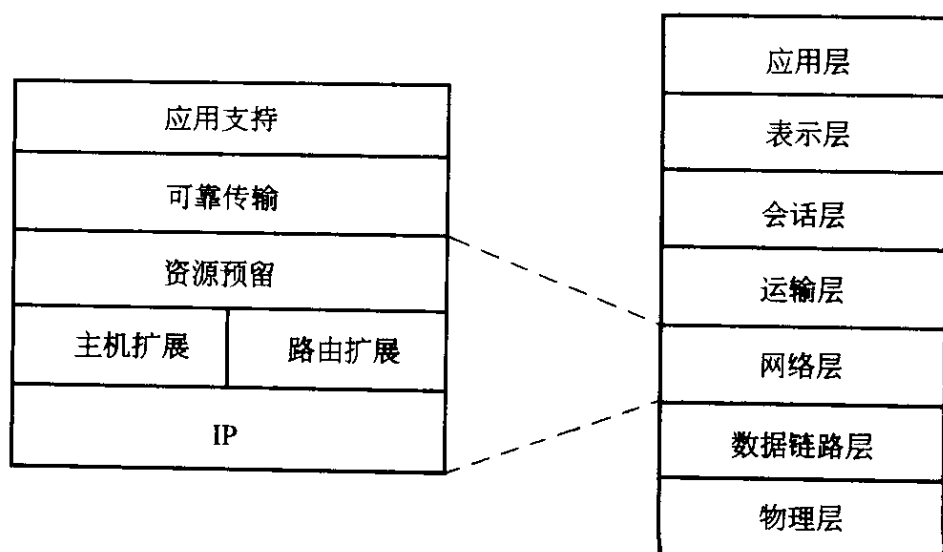


图 1-1 群组通信体系结构

5. 同步机制

在多媒体会议中,多媒体信息通常是由多个媒体流(音频、视频、文本等)组成的,各媒体流之间往往具有一定的时间关系,因此在接收端应实现同步回放,我们称这种同步为媒体同步(也称唇同步)。另一方面,由于多媒体会议是面向多个用户的,为了公平起见,一个用户发送的多媒体信息应当在所有接收的用户中同时进行回放,使每个用户有平等的反应机会,这种同步称为群同步。实现同步的方法很多,如时间戳法、多路复用技术、同步信号通道技术、对象复合 Petri 网(OCPN)方案等等。

6. 差错控制技术和流量控制技术

由于多媒体通信具有高速大容量的特性,而目前的计算机通信网还不能完全满足多媒体通信的要求,因而需要在传输协议中加入流量控制算法,同时传统协议中采用的为每一个丢失或出错分组进行重发的差错控制方法也已不适用。流量控制技术主要可分为预防式或反馈式两大类,预防式包括漏桶(LB)、广义漏桶、峰值计数器、虚时钟等,信源根据监控参数调整发送速率;反馈式大多以窗口控制机制为基础,信源根据反馈的网络情况调整速率。反馈式会引入较大的时延,因而预防式更适合于多媒体通信,但是监控参数的确定也是一个十分复杂的问题。多媒体通信的差错控制技术有选择式重发或前向纠错编码 FEC 两种,前者根据丢失(或出错)分组的重要性进行有选择的重发;后者则在发送(重要)分组时加入一定的冗余信息,以便在出错时能得到一定程度的恢复。

7. 多媒体会议系统的性能评价

目前对于多媒体会议的研究与应用已经有了相当大的发展,但是对多媒体会议系统的性能评价却非常肤浅。通常用视频尺寸、帧率、码率等指标来大致描述系统的性能,但没有一个能客观反映递交给最终用户的媒体质量(尤其是图像质量)的指标,显然这种描述是不全面的。

8. 其他

在开发多媒体会议时,还有些关键技术,如多媒体数据库、人机接口等也是决定系统成功与否的重要因素,这里不再赘述。

1.5 视频会议的发展方向

1. 向高性能、多功能方向发展

目前由于网络的限制,商用视频会议系统的功能和性能还不能令人满意,但随着计算机网络的不断发展,具有 QoS 保证的高速多媒体通信网一定会得到普及。届时,开发具有高性能(如具有 MPEG-2 音视频质量、极低时延)的多媒体会议系统将不再是梦想。另外,多媒体会议的功能也会得到进一步地扩充,比如可以同时参加多个会议;并能在多个会议间快速切换,录制会议内容;另外还可以增加一些娱乐性的工具。

2. 向协同性发展

目前的多媒体会议中,每个用户还是相对独立的个体。作为一个会议群体,相互之间的协同性比较差,这就抑制了多媒体会议的进一步发展。计算机支持的协同工作(CSCW)是近十多年来发展起来的多学科交叉的新研究领域,它研究如何利用计算机及计算机网络支持地理位置分散的人们协同工作,完成一项共同的任务。我们有理由相信,计算机网络的发展必然走向计算机协同工作,而多媒体会议也必然走向协同与高效。

3. 向智能化发展

随着 MPEG-4 技术的成熟,下一代多媒体会议将首选基于内容的编码技术。这不仅将大大改善编码效率,而且能在会议系统中融入智能和用户交互性,用户进入会场时不再需要登记和输入口令,系统会根据用户的头像及声音自动进行身份认证及登记;用户可以随意选择或调节感兴趣的画面内容等等。

4. 向集成化发展

多媒体会议产品最终将走入家庭,而体积小、价格低、操作简单多媒体会议产品才会受到消费者的欢迎。现在市场上已经有符合 H.320 标准和 H.324 标准的可视电话出售,它将音视频 I/O、压缩编码、显示、通信协议、互操作协议全部用硬件实现,使用时只要拿起电话拨一个电话号码就行,非常方便。

总之,今后的多媒体会议系统一定会变得更加方便、更加亲切、更加真实。

第 2 章

视频会议的网络 环境及交换技术

2.1 网络上的信息交换技术

远程通信网络可分成两类网络:线路交换网络(circuit-switched networks)和信息包交换网络(packet-switched networks)。信息包交换网络也称数据网络(data networks)。在线路交换网络中,双方交换消息期间通道上的全部资源(如链路带宽等)都被租用,而在信息包交换网络中双方交换消息期间不租用通道上的资源,也就是不保留资源,而是根据需要来使用资源,交换的信息包往往需要等待可用资源。打个比方,例如有两种类型的饭馆:一种需要预定的饭馆,另一种不需要或者不接收预定的饭馆。对于前者,在去饭馆之前必须要和第一个打电话或发 E-mail 的人竞争,但到了饭馆后通常马上就可与服务人员联系并定饭菜;对于后者,客户没有预定座位的问题,但到了饭馆之后也许必须等座位,然后才能定饭菜。

电话网络(telephone networks)是线路交换网络的一个例子。如果要在电话线路上与他人通话或者发送传真,你必须首先与对方建立连接方可进行,这就叫做线路连接,或者叫做线路交换。当线路连接建立之后就租用了这个线路的带宽,你就可以用常数速率传输数据。

因特网是信息包交换网络的例子。例如,如果你想在因特网上给人发送一个消息,因特网并不给你保留任何带宽。在你的消息到达接收端之前,必须要经过一系列通信链接(link)才能传递到接收端。由于许多消息都在因特网上传输,如果其中有一个链接处发生拥挤,你的消息就不得不在缓冲存储器中等待,这样就产生了延迟。虽然因特网使用了很多很好的技术来克服这些问题,但依然不能像线路交换网络那样来保证以恒定的速率来传送消息。

不是所有的远程通信网络都属于线路交换网络或者属于信息包交换网络。例如,ATM 网络技术既有保留带宽的连接又有不保

留带宽的连接。

2.1.1 线路交换

线路交换(circuit switching)的特点是,在开始通信之前通信双方由线路交换中心建立物理连接,维持连接的时间长短取决于消息交换的需要。图 2-1 是线路交换网络的示意图。在这个网络中,使用 2 条传输链路(link)把 3 个线路交换机连接在一起,每条传输链路都有“ n 条线路”,“ n 条线路”是指通过时分多路复用技术 TDM(time-division multiplexing)或者频分多路复用技术 FDM(frequency division multiplexing)来获得的,因此每条链路可同时支持 n 个连接(connection)。每个接收端(如 PC 机或者工作站)直接连接到其中的一台交换机,或者通过调制解调器直接连接到交换机。我们使用的普通电话也连接到交换机上,但没有在图上表示出来。当两台主机希望通信时,两台主机之间就要建立一个专门的端-端线路连接,每个端-端线路得到 $1/n$ 的链路带宽。

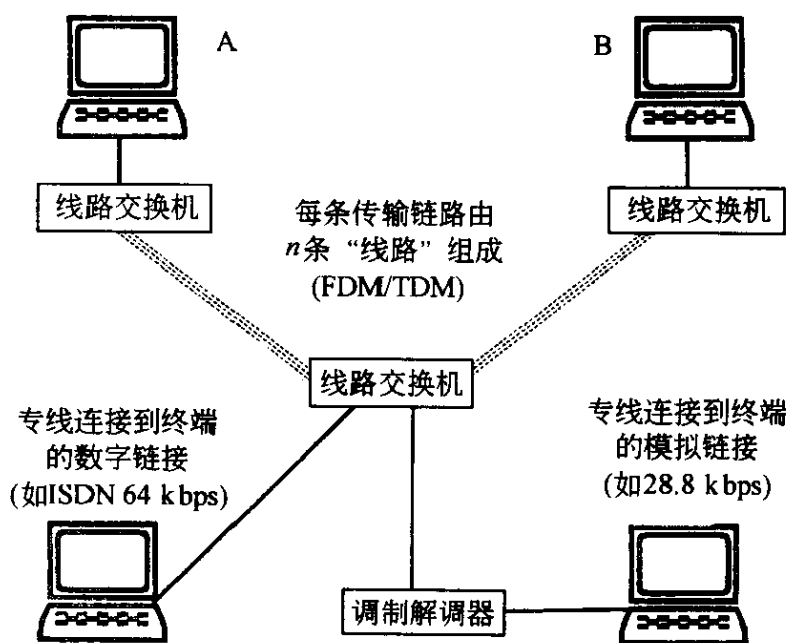
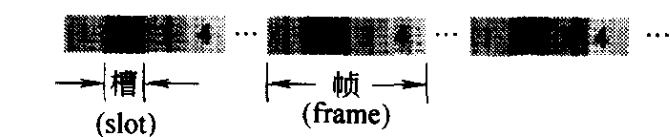


图 2-1 线路交换的概念

TDM 技术把时间分割成许许多多固定的帧(frame),而每一

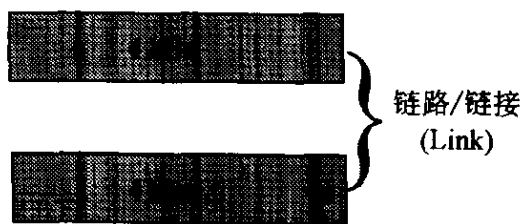
帧又划分成固定数目的时间间隔,称为时间槽(time slot)。当用户 A 与用户 B 进行通信时,系统为这对用户指定一个时间槽,仅在这个时间槽里建立连接,每对用户连续占用的资源仅是其中的一个时间槽的带宽,这样就可以使许多用户共用一条传输链路。

FDM 技术把传输线的频带分割成若干部分,每个频带均可作为一个独立的传输信道使用,因此一对传输线上也可以同时有若干对用户进行通话,每对用户连续占用资源的仅是其中的一个频段。采用频分多路复用技术的通信又称为载波通信。在电话网络中使用 FDM,每个链接使用 4 kHz 的带宽,调频广播使用 FDM 共享微波频谱。



注：标有相同时间槽号的时间槽分配给某一对收发者

(a) 时分多路复用



(b) 频分多路复用

图 2-2 FDM 和 TDM 的概念

2.1.2 信息包交换

通常人们把任何一个应用数据块称为消息(message),例如, ASCII 文件, Postscript 文件, Web 页面和声音文件等等。在现代的信息包交换(packet switching)网络中,发送端把整个消息分割成许多小的数据块,经过包装并“贴上”标签之后再发送到网络上。发送到网络上的这种数据包裹称为信息包(packet),比较准确地说,信息包是一个经过包装且具有固定大小的传输单元,这种数据

包裹既包含用户的数据又包含按照协议规定加入的标题,而标题中包含有识别号码、发送地址、接收地址等信息。在接收端把接收到的信息包拆开后重新拼接成原来的完整消息。不像点对点的电路交换需要在收发双方建立物理连接,在信息包交换网络上的每个信息包都包含有目的地址,因此一个消息分装成的许多信息包不必都沿着同一条线路到达目的地,也不必同时到达目的地,到达目的地的次序也不必按照发送的次序,哪条信道有空就从哪里传送。

发送端把一个很长的消息分割成比较小的信息包之后,在发送端和接收端之间,每个信息包通过通信链路和信息包交换机(packet switches),有时也称路由器(routers)传送到接收端。信息包在每个通信链路上以传输速率同链路传输速率相等的速度传输。大多数信息包交换机在输入端使用存储转发方法(store and forward)把信息包转发到输出链路上,这意味着交换机必须把整个信息包接收完之后才能把信息包的第 1 位(bit)转发到输出链路上,这样就产生了存储转发时延,时延的长短与信息包的大小(长度)成正比。如果信息包的长度为 L ,传输链路的数据率为 R ,时延就为 L/R 。

在每台交换机中有多个缓冲存储器,每个信息包在输出到链路之前必须要在链路缓冲存储器中排队,如果在信息包到达时缓冲存储器是空的或者没有其他信息包到达的情况下就不需要排队。因此除了前面介绍的存储转发时延(store-and-forward delays)之外,还要附加信息包的排队时延(queueing delay),它是一个不确定的时延,取决于网络上的拥挤情况。

图 2-3 表示一个简单的信息包交换网络。假设主机 A 和 B 正在向主机 E 发送信息包,信息包 A 和信息包 B 首先通过 28.8 kbps 链路传送到第一个信息包交换机,然后把它们传送到 1.544 Mbps 的链路上。如果在这条链路上出现拥挤现象,这些信息包就在链路缓存存储器中排队等待输出到链路上。输出到链路上的次序不按照任何预先定义的规则,是随机的或者是统计方式的,因此

这种信息包交换技术也称为统计多路复用技术,它与每个主机获得相同时间槽的 TDM 技术截然不同。

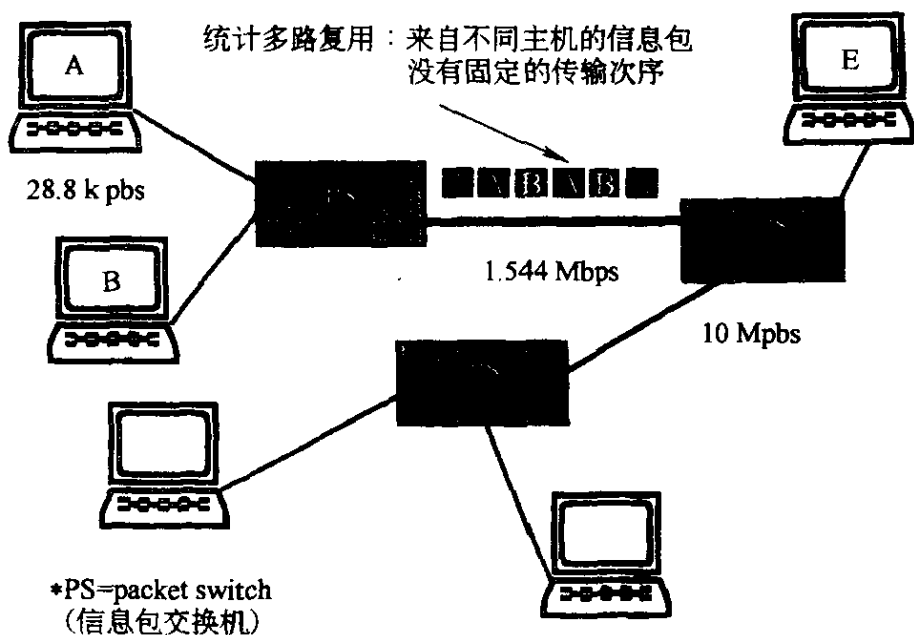


图 2-3 信息包交换的概念

2.1.3 面向连接服务与无连接服务

因特网通常叫做 TCP/IP 网络,提供两种类型的服务:无连接服务(connectionless service)和面向连接服务(connection-oriented service)。任何一种因特网服务都必须使用这两种服务之一。对这两种服务,每个信息包都带有目的地址,而交换机就是按照信息包的地址来发送信息包的。

(1) 面向连接服务(TCP)

面向连接服务是,在开始发送信息包之前发送端和接收端要进行沟通,建立直接连接,并提醒对方准备接收信息包,然后才开始进入信息包的传送过程。

面向连接服务提供下列 3 种功能来确保信息包从发送端传送到接收端:

① 确认(acknowledgements):当接收端接收到信息包时,接收端向发送端发送一个确认信号,当发送端接收到确认信号后,说明

相应的信息包已被对方接收到。

② 流程控制(flow control):接收端也许因为忙于其他任务或者限于它的处理能力,须要限制发送端发送信息包的速率,这可在开始沟通时告诉发送端所须要限制的速率。

③ 拥挤控制(congestion control):当交换机拥挤时,交换机中的缓冲存储器可能发生溢出,导致信息包丢失,面向连接服务提供了一种机制(方法)来限制传输速率。

由于采用了上述措施,因特网的面向连接服务被认为是可靠的服务,而因特网的无连接服务被认为是不可靠的服务,理由是它不能确保把数据传送到目的地。

(2) 无连接服务(UDP)

通过使用用户数据包协议 UDP(User Datagram Protocol)来调用无连接服务。无连接服务的含义是发送端简单地把信息包送到网络上,在传送信息包之前发送端和接收端没有沟通的过程,也没有对方来的确认,因而也不知道目的地是否接收到。无连接服务既没有拥挤控制功能,也没有流程控制功能。

2.1.4 网络分类

远程通信网络的分类大致如图 2-4 所示。远程通信网络(telecommunications network)使用线路交换(circuit switching)和信息包交换(packet switching)。在线路交换中,使用频分多路复用 FDM(frequency-division multiplexing)技术和时分多路复用 TDM(time-division multiplexing)技术来提高传输容量。

信息包交换网络分成两类,一类叫做数据包网络(datagram network),另一类叫做虚拟线路网络(virtual circuit network)。如果信息包按照目的地址发送,这种网络就称为数据包网络,因特网的 IP 协议就是按照目的地址发送信息包;如果信息包按照虚拟线路号发送,这种网络称为虚拟线路网络。

数据包网络(datagram networks)在许多方面与邮政服务类

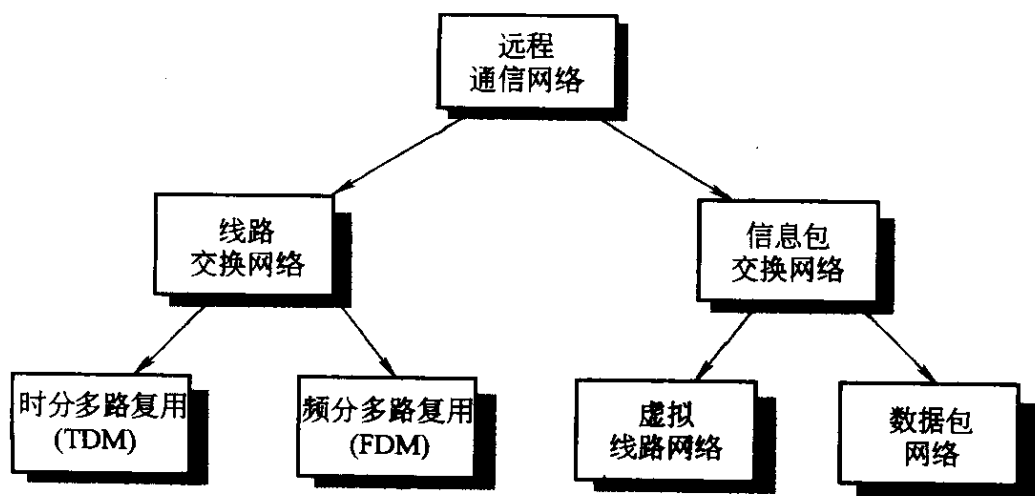


图 2-4 网络分类

似。当寄件人发送一封信到目的地时，寄件人把信装入信封，并且在信封上写上带层次结构的收件人地址，如国家—省—市—街道—门牌号—收件人。邮政服务系统根据信封上的地址把信件送达目的地。在数据包网络中，网络上传送的每个信息包都包含有目的地址，它很类似于邮件上的分等级的地址。由于每台信息包交换机都有路由表，当信息包到达网络中的交换机时，交换机就抽出信息包中的目的地址部分，并用该地址在路由表中进行检索，以找出合适的外向链路，然后把信息包送到链路上。

虚拟线路就是临时创建的通信通路，由 3 个部分组成：

① 发送端和接收端之间的一系列链路和交换设备，称为路径 (path)。

② 沿着这条路径上的虚拟线路号 (virtual circuit number)，一个号码对应路径上的一条链路。

③ 沿途交换设备里的虚拟线路转换表中的表项。发送端和接收端之间的虚拟线路一旦建立，信息包就可选择合适的虚拟线路号进行发送。由于虚拟线路在每个链路上有不同的虚拟线路号，中间的信息包交换机必须用新的虚拟线路号替代每个行进中的信息包虚拟线路号。

数据包网络对某些消息可提供无连接服务，对另一些消息可提供面向连接服务。例如，因特网是数据包网络，它既可提供无连

接服务也可提供面向连接服务。使用虚拟线路的网络一般都是面向连接服务,例如 X.25、帧中继(frame relay)和异步传输模式 ATM(Asynchronous Transfer Mode)网络。

2.2 网际多目标广播

进入 20 世纪 90 年代之后,组合声音、电视和数据流的多媒体网络应用的开发和研究迅速增加,近年来不断有产品投入市场,用户也渴望得到服务质量好、服务费用低的产品,像电视会议、协同工作、远程教学、可视电话等多媒体网络应用都是非常受欢迎的应用。这些应用都是实时的交互应用,而且即使采用了很好的压缩技术,传输多媒体数据所需要的带宽也是巨大的。在实现这些令人兴奋的应用中,正在开发的网际多目标广播(IP multicast)技术将会并且正在起着越来越重要的作用。使用这种多目标广播技术不仅可以节省宝贵的网络资源(主要是带宽)和服务器资源,而且可以保证服务质量。

如同因特网上的其他应用技术一样,多目标广播的核心技术体现在各种各样的协议上。协议是技术的精华,是人类智慧的结晶。如果不带着问题去阅读协议是很枯燥的,但如果要了解多目标广播技术就要了解协议的思想。如果要开发产品就要深入研究协议的细节,如果没有协议可遵循就自己制定和提交新的协议,让大家使用你提出的协议。

2.2.1 多目标广播的概念

假设世界上有许多用户想在因特网上接收现场电视节目(例如现场体育比赛),如何把它传送到世界各地?最直观的方法就是使用传统的 IP 寻址方法,每个信息包都使用一个唯一的 IP 地址,一次只给一个节点(即主机)传送,如图 2-5 所示,这种方法称为单目标传送(unicast)。如果使用单目标广播服务把相同内容的信息传输给 n

个目标站点,就须要传输 n 个拷贝,即要传输 n 次。

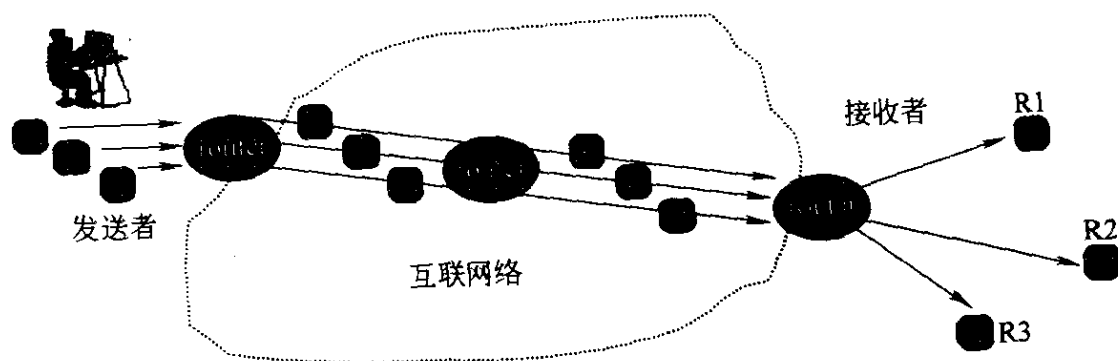


图 2-5 单目标广播服务

使用单目标广播服务有下列缺点:

- ① 浪费链路带宽,因为在链路上要传送多个相同的拷贝;
- ② 大大地加重了服务机的负担。

一个比较好的方法是把消息一次性地同时传输给 N 个目标站点,如图 2-6 所示,这就叫做多目标广播(multicasting)。在因特网上广播就称为 IP 多目标广播或者叫做网际多目标广播。

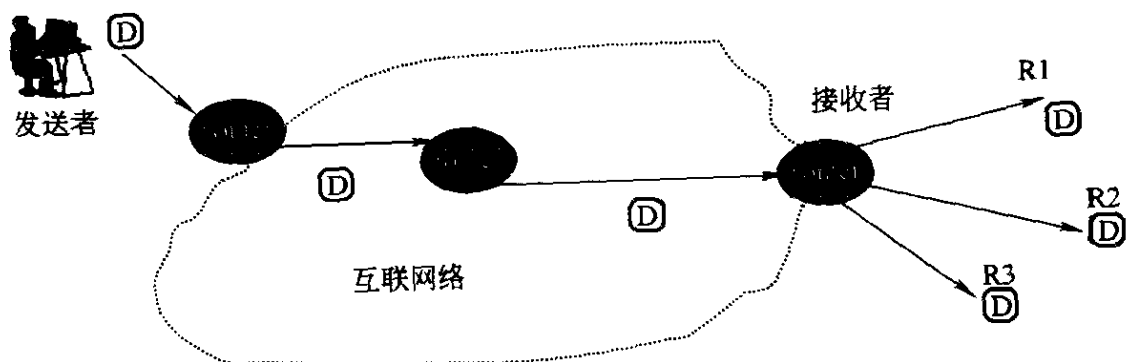


图 2-6 多目标广播服务

多目标广播是真正的分布式信息传输服务,使用 IP 多目标广播可以大大减轻网络上出现的拥挤和服务器的负担,因此可用于声音和影视的实时广播,例如,因特网电话会议、因特网电视会议。

2.2.2 多目标广播所需要的环境

主机节点所需要的环境是：

(1) TCP/IP 协议堆中可支持 IP 多目标广播。

(2) 软件支持网际主机组管理协议 IGMP (Internet Group Management Protocol), 这样就可以申请参加多目标广播组 (multicast group) 和接收多目标广播。

(3) 要有 IP 多目标广播应用软件, 例如电视会议软件。

在 WAN 网络上运行或者评估多目标广播还需要以下条件：

(1) 在接收两端之间的所有路由器都具备多目标广播的功能。许多新的路由器都有这个功能, 老的路由器也许需要升级和更新。

(2) 需要能够识别防火墙以便使多目标广播畅通。

图 2-7 描述了运行多目标广播必需要有的部件。图中的交通方向是多目标广播数据包的方向, 需要与主机组成员进行通信的消息和传输路径消息没有在图上表示。

2.2.3 多目标广播树的概念

多目标广播树 (multicast tree) 或者叫做多目标广播跨越树 (multicast spanning tree) 的概念, 用来描述服务机 (播送机) 和接收机之间构成的播放与接收关系。服务机只向外传送一个信息包流, 每当信息包到达多目标广播树中有多个分支的路由器时, 路由器就为每个分支复制一个信息包。这样就减轻了服务机的负担, 有效地利用了网络资源。

网络将使用路由协议建立从源端到所有希望参加会话的接收端之间的多目标树, 图 2-8 说明这种树的结构。网络建立多目标广播树的过程如下：

(1) 广播源把数据或者广播通知发送给所有路由器。

(2) 不想参加多目标广播的终端逆向发送一个删除消息。

(3) 删除没有成员的分支和不在最短路径树上的分支。

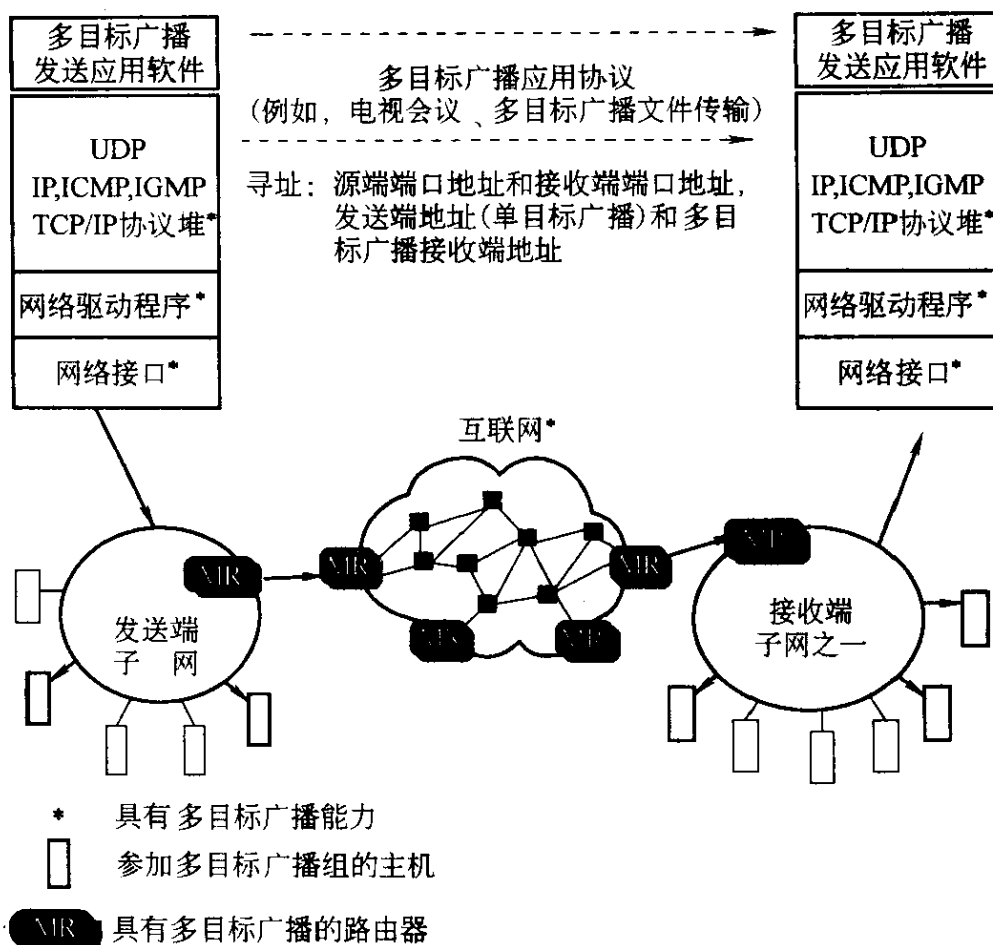


图 2-7 网际多目标广播环境

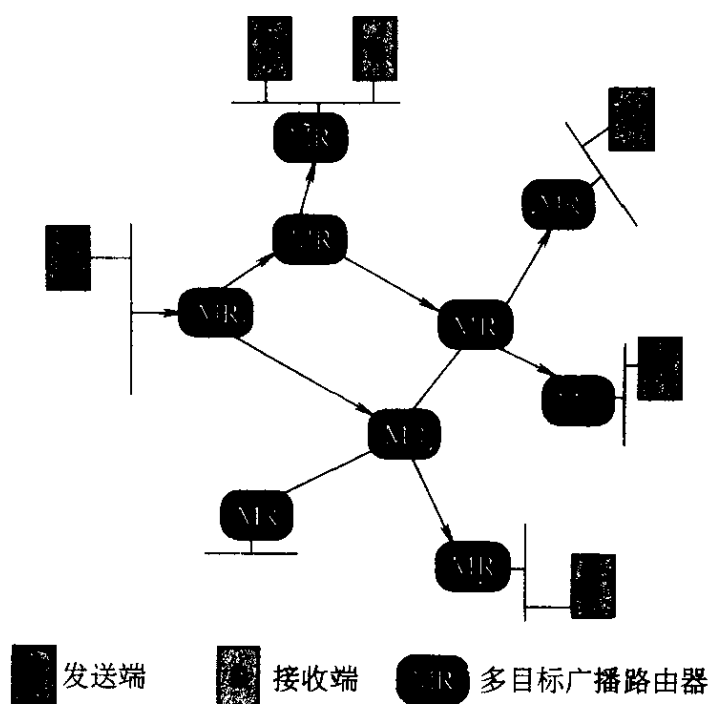


图 2-8 多目标广播树结构

- (4) 在广播源生成最短路径树。
- (5) 使用联结和删除功能改变成员之间的关系。

现在至少有三种多目标广播路由协议:

- (1) 距离矢量多目标路由协议 DVMRP (Distance Vector Multicast Routing Protocol)。
- (2) 多目标广播开放最短路径优先协议 MOSPF (Multicast Open Shortest Path First)。
- (3) 协议独立多目标广播 PIM (Protocol Independent Multicasting)。

2.2.4 IP 多目标广播的地址和主机组的管理

1. IP 多目标广播的地址

IP 多目标广播使用 D 类 IP 地址,它的最高 4 位为 1110 就是用来指定多目标广播主机组群的地址。在网际协议地址(IP 地址)的“点十进制”记法中,主机组的地址范围从 224.0.0.0 到 239.255.255.255。

主机组的地址划分为两种类型,一种称为永久性地址,另一种称为暂时性地址。

例如,永久地址 224.0.0.1 称为“所有主机组”地址,它是用来与直接连接到网络上的所有 IP 多目标广播主机组进行通信的地址,而 244.0.0.2 是与 LAN 上的所有路由器进行通信的地址。244.0.0.0 到 244.0.0.255 是为路由协议和其他用途保留的地址。其他的地址和地址范围用于应用软件,例如,244.0.13.0 到 244.0.13.255 用于网络新闻(Net News)。

发送给多目标广播主机组所有成员的数据包称为 IP 数据包(IP Multicast Datagram),这种数据包与单目标广播 IP 数据包相同。要发送 IP 多目标广播数据包(IP Multicast datagrams)时,发送端指定一个代表主机组的目标地址,使用与发送单目标数据包一样的操作“Send IP”来发送多目标广播数据包。

与发送 IP 多目标广播数据包比较,接收 IP 多目标广播数据包要复杂得多,尤其是在 WAN 上。为了接收数据包,用户主机上的应用软件需要在多目标广播主机组中申请成员资格。申请成员资格要与 LAN 上的路由器通信,如果需要的话还要与发送端和接收端之间的中间路由器进行通信。有了成员资格之后,接收端主机上的网络接口卡就可开始筛选具体的 LAN 网络硬件(数据链路层)的地址,这个地址是与新的多目标广播组地址相关联的。WAN 路由器把请求的多目标广播数据包递送到 LAN 路由器上,LAN 路由器把接收端的主机地址转换成相关的硬件地址,然后使用这个地址创建消息(例如,以太网帧)。正在注意这些地址的接收主机的网络接口卡和网络驱动程序把多目标广播消息传输给 TCP/IP 协议堆,由协议堆把广播消息变成用户应用程序的输入,例如电视播放器。

2. 网际主机组管理协议 IGMP

如果 LAN 上多目标广播组的成员要接收来自远距离广播源的广播,多目标广播信息包就必需需要通过路由器转发到 LAN 网上。多目标广播路由器通过使用网际主机组管理协议 IGMP(Internet Group Management Protocol)来查明直接附加到子网的主机组成员是否存在,它发送一个 IGMP 查询消息以获得主机组成员的情况。

网际主机组管理协议 IGMP 在 IP 上的执行过程与网际控制消息协议 ICMP(Internet Control Messages Protocol)的执行过程类似。IGMP 消息封装在 IP 数据包(IP datagram)中,而数据包有两种类型:主机成员查询(Host Membership Query)数据包和主机成员报告(Host Membership Report)数据包,它们具有相同的固定格式。

为了确定在局域子网上的任何一台主机是否属于多目标广播组,每个子网上的路由器要周期性地向 LAN 上的所有终端主机发送一个主机成员查询信息,要求它们汇报主机组群成员的情况。这个查询被送到所有主机组(网络地址=224.0.0.1),而每台主机

都回送一个主机成员报告(Host Membership Report)信息,如图 2-9 所示。这样就可确定子网上的主机是否加入多目标广播组,从而确定是否要把多目标数据包送到这个子网。

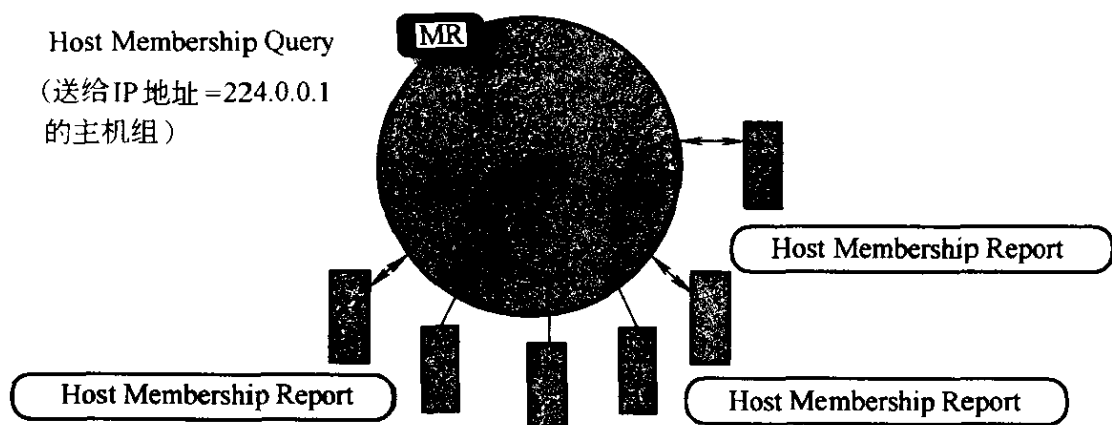


图 2-9 在 LAN 内的网际主机组成员协议(IGMP)消息

3. 接收端如何参与多目标广播

多目标广播会话消息经常通过因特网广告通知接收端。广告使用应用层的广播会话说明协议 SDP(Session Description Protocol)来发布。这个协议发送广告的地址是 224.2.2.2,UDP 端口号是 4000,这是一个特殊的多目标广播会话。发送的广告内容包括广播的名称、广播时间(active times)、媒体的类型(声音、电视、白板等等)和广播地址。任何有兴趣了解多目标广播的主机都可以参加这个特殊的多目标广播会话和接收这种广告。可用一台主机运行用户代理来收集广告并把广告摘要用图形方式向用户显示。第一个这样的用户代理叫做会话目录 SD(Session Directory),是由 LBL(Lawrence Berkeley Laboratories)的 Van Jacobson 开发的,它除了显示公告之外,也启动多目标应用程序和为任何一个新的多目标广播会话选择一个地址。

如何给多目标广播分配地址?当源端开始一个新的多目标广播时,SD 就从多目标地址空间中随机地选择一个多目标广播地址。参加多目标广播的主机需要做两件事:

①.立即给最近的路由器发送一个消息,告诉它参加多目标广播,发送这个消息是使用应用层的一个叫做网际主机组管理协议

IGMP(Internet Group Management Protocol,RFC 1812)。

② 设置 IP 进程去接收 IP 数据包,因为它在 IP 目标地址域中含有会话的多目标广播地址,当接收端退出多目标广播会话时,类似的进程也要设置。

2.3 实时传输协议和实时控制协议

实时传输协议 RTP(Real-time Transport Protocol)、实时控制协议 RTCP(Real-time Control Protocol)、资源保留协议 RSVP(Resource Reservation Protocol)和实时流放协议 RTSP(Real-time Streaming Protocol)就是为实时多媒体在网络上的应用而开发的协议。

2.3.1 实时传输协议 RTP

RTP 是一种提供端对端传输服务的实时传输协议,用来支持在单目标广播和多目标广播网络服务中传输实时数据,而实时数据的传输则由 RTCP 协议监视和控制。

使用 RTP 协议的应用程序运行在 RTP 之上,而执行 RTP 的程序运行在 UDP 的上层,目的是为了使用 UDP 的端口号和检查机制。如图 2-10 所示,RTP 可以看成是传输层的子层。由多媒体应用程序生成的声音和电视数据块被封装在 RTP 信息包中,每个 RTP 信息包被封装在 UDP 消息段中,然后再封装在 IP 数据包中。

从应用开发人员的角度来看,可把 RTP 执行程序看成是应用程序的一部分,因为开发人员必须把 RTP 集成到应用程序中。在发送端,开发人员必须把执行 RTP 协议的程序写入到创建 RTP 信息包的应用程序中,然后应用程序把 RTP 信息包发送到 UDP 的套接接口(socket interface),如图 2-11 所示。同样,在接收端,RTP 信息包通过 UDP 套接接口输入到应用程序中,因此开发人

员必须把执行 RTP 协议的程序写入到从 RTP 信息包中抽出媒体数据的应用程序中。

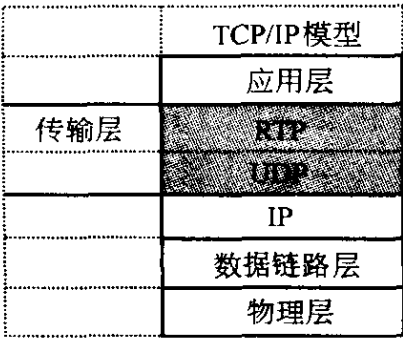


图 2-10 RTP 是传输层上的协议

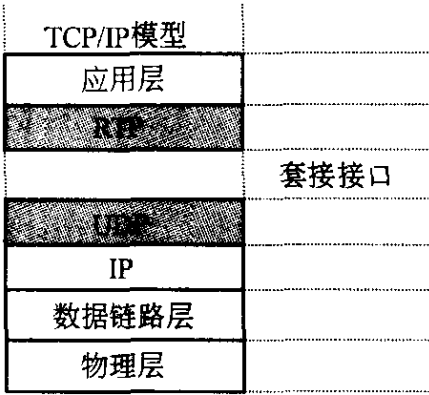


图 2-11 RTP 和 UDP 之间的接口

现以 RTP 传输声音为例来说明它的工作过程。

假设音源的声音是 64 kbps 的 PCM 编码声音,并假设应用程序取 20ms 的编码数据为一个数据块(chunk),即在一个数据块中有 160 个字节的语音数据。应用程序需要为这块声音数据添加 RTP 标题生成 RTP 信息包,这个标题包括声音数据的类型、顺序号和时间戳。然后 RTP 信息包被送到 UDP 套接接口,在那里再被封装在 UDP 信息包中。在接收端,应用程序从套接接口处接收 RTP 信息包,并从 RTP 信息包中抽出声音数据块,然后使用 RTP 信息包的标题域中的信息正确地译码和播放声音。

如果应用程序不使用专有的方案来提供有效载荷类型(payload type)、顺序号或者时间戳,而是使用标准的 RTP 协议,应用程序就更容易与其他的网络应用程序配合运行,这是大家都希望的事情。例如,如果有两个不同的公司都在开发因特网电话软件,他们都把 RTP 合并到他们的产品中,这样使用不同公司电话软件的用户之间就有希望能够进行通信。

这里需要强调的是,RTP 本身不提供任何机制来确保把数据及时递送到接收端或者确保其他的服务质量,它也不担保在递送过程中不丢失信息包或者防止信息包的次序不被打乱。的确,

RTP 的封装只是在系统端才能看到,中间的路由器并不区分哪个 IP 数据包是运载 RTP 信息包的。

RTP 允许给每个媒体源分配一个单独的 RTP 信息包流,例如,摄像机或者麦克风。例如,有两个团体参与的电视会议,这就可能打开 4 个信息包流:两台摄像机传送电视流和两个麦克风传送声音流。然而,许多流行的编码技术,包括 MPEG-1 和 MPEG-2 在编码过程中都把声音和电视图像捆绑在一起以形成单一的数据流,一个方向就生成一个 RTP 信息包流。

RTP 信息包没有被限制只可应用于单目标广播,它们也可以在一对多(one-to-many)的多目标广播树或者在多对多(many-to-many)的多目标广播树上传送。例如,多对多的多目标广播,在这种应用场合下,所有发送端通常都把他们的 RTP 信息包流发送到具有相同多目标广播地址的多目标广播树上。

2.3.2 RTP 信息包标题域

RTP 标题由 4 个信息包标题域和其他域组成:

1. 有效载荷类型

RTP 信息包中的有效载荷(Payload Type Field)域的长度为 7 位(bit),因此 RTP 可支持 128 种不同的有效载荷类型。

对于声音流,这个域用来指示声音使用的编码类型,例如 PCM、自适应增量调制或线性预测编码等等。如果发送端在会话或者广播的中途决定改变编码方法,发送端可通过这个域来通知接收端。

对于视频流,有效载荷类型可以用来指示视频编码的类型,例如 motion JPEG、MPEG1、MPEG2 或者 H.231 等等。发送端也可以在会话期间随时改变视频的编码方法。

2. 顺序号

顺序号(Sequence Number Field)域的长度为 16 位(bit)。每发送一个 RTP 信息包顺序号就加 1,接收端可以用它来检查信息包是否有丢失以及按顺序号处理信息包。例如,接收端的应用程序接收到一个 RTP 信息包流,这个 RTP 信息包在顺序号 86 和 89 之间有一个间隔,接收端就知道信息包 87 和 88 已经丢失,并且采取措施来处理丢失的数据。

3. 时间戳

时间戳(Timestamp)域的长度为 32 字节。它反映 RTP 数据信息包中第一个字节的采样时刻(时间)。接收端可以利用这个时间戳来去除由网络引起的信息包的抖动,并且在接收端为播放提供同步功能。

4. 同步源标识符

同步源标识符 SSRC(Synchronization Source Identifier)域的长度为 32 位(bit)。它用来标识 RTP 信息包流的起源,在 RTP 会话期间的每个信息包流都有一个清楚的 SSRC。SSRC 不是发送端的 IP 地址,而是在新的信息包流开始时源端随机分配的一个号码。

2.3.3 实时传输控制协议 RTCP

实时传输控制协议 RTCP(Real-time Control Protocol),当从一个或者多个发送端向多个接收端广播声音或者电视时,也就是在 RTP 会话期间,每个参与者周期性地向所有其他参与者发送 RTCP 控制信息包,如图 2-12 所示。RTCP 用来监视服务质量和传送有关与会者的信息。对于 RTP 会话或者广播,通常使用单个多目标广播地址,属于这个会话的所有 RTP 和 RTCP 信息包都使用这个多目标广播地址,通过使用不同的端口号可把 RTP 信息包和 RTCP 信息包区分开来。

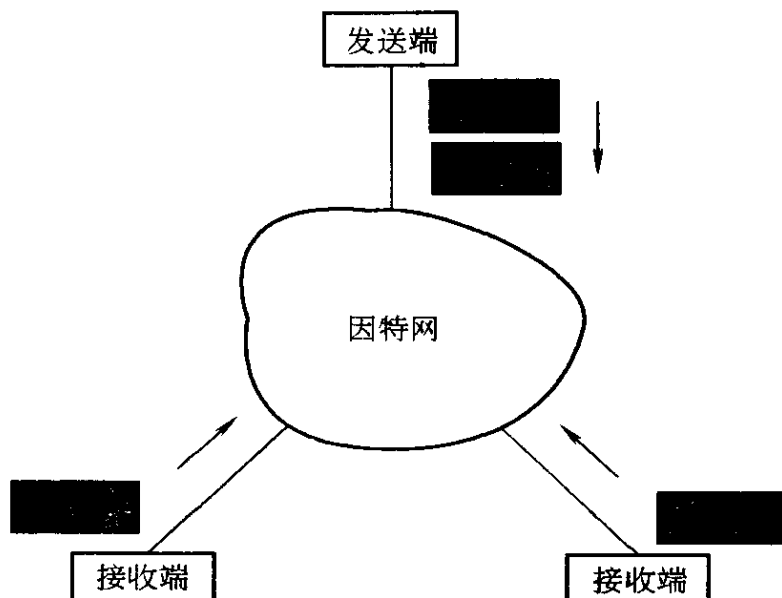


图 2-12 每个参与者周期性地发送 RTCP 控制信息包

RTCP 的主要功能是为应用程序提供会话质量或者广播性能质量的信息。每个 RTCP 信息包不封装声音数据或者电视数据，而是封装发送端和/或者接收端的统计报表。这些信息包括发送的信息包数目、丢失的信息包数目和信息包的抖动等情况，这些反馈信息对发送端、接收端以及网络管理员都是很有用的。RTCP 规格没有指定应用程序应该使用这个反馈信息做什么，这完全取决于应用程序开发人员。例如，发送端可以根据反馈信息来修改传输速率，接收端可以根据反馈信息判断问题是本地的、区域性的还是全球性的，网络管理员也可以使用 RTCP 信息包中的信息来评估网络用于多目标广播的性能。

2.3.4 实时流放协议 RTSP

实时流放协议 RTSP(Real-Time Streaming Protocol)是一个刚开始开发的协议，它是应用级的实时流放协议，其主要目标是为单目标广播和多目标广播上的流式多媒体应用提供牢靠的播放性能，以及支持不同厂家提供的客户机和服务机之间的协同工作能力。

播放的数据流被分成许多信息包，信息包的大小很适合于客户机和服务器之间的带宽。当客户机已经接收到足够多的信息包

之后,用户软件就可开始播放一个信息包,同时对另一个信息包解压缩和接收第三个信息包。这样用户不需要把整个媒体文件从服务器上下载之后就可立即播放。广播源可以是现场的数据流也可以是存储的数据流。

RTSP 协议要想提供控制多种应用数据传送的功能,则要提供一种选择传送通道的方法,例如 UDP、TCP、IP 多目标广播通道,以及提供一种基于 RTP 协议的递送方法。正在设计的 RTSP 将工作在 RTP 的上层,用来控制和传送实时的内容。

RTSP 能够与资源保留协议一起使用,用来设置和管理保留带宽的流式会话或者广播。

2.4 资源保留协议

2.4.1 资源保留协议 RSVP

资源保留协议 RSVP(Resource Reservation Protocol)是为保证服务质量而开发的一种协议,它允许应用程序为它们的数据流保留带宽。

RSVP 的两个主要特性是:

- ① 保留多目标广播树上的带宽,单目标广播是一个特殊情况。
- ② 接收端导向,也就是接收端启动和维护资源的保留。

这两个特性可用图 2-13 说明。该图表示的是一个多目标广播树(multicast tree),它的数据流向是从树的顶部到 6 个主机。虽然数据源来自发送端,但保留消息(reservation message)则发自接收端。当路由器向上给发送端转发保留消息时,路由器可以合并来自下面的保留消息。

2.4.2 不同种类的接收器

接入因特网的用户是多种多样的,有的使用 28.8 kbps 速率接收数据,有的使用 128 kbps 速率接收数据,而有的使用 10 Mbps

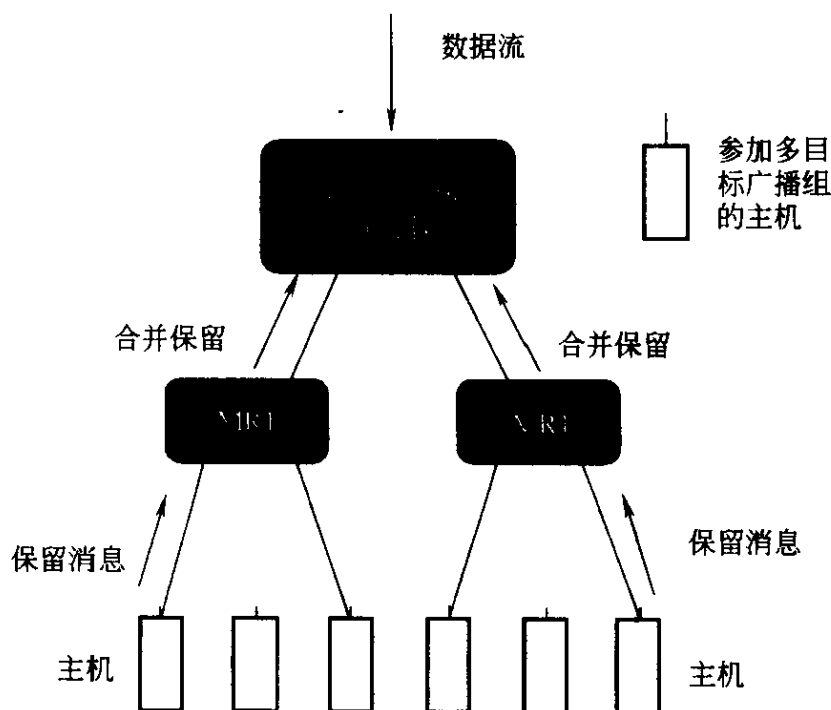


图 2-13 RSVP 的特性

甚至更高的速率接收数据。解决这个问题的一种方案是对声音或电视进行分层编码,每层的声音或电视的数据速率各不相同,以此来满足各种不同用户的要求。在广播时,发送端并不一定要知道接收端的数据接收速率,而只需要知道这些用户中使用的最大数据接收速率即可。发送端对声音或电视进行分层编码,并且把它们发送到广播树上,用户就可以根据自己的实际速率接收不同质量的广播。

例 1 多目标实况广播

假设有一场体育比赛要在因特网上进行实况转播,多目标广播地址已经在因特网上事先发布,并且假设从发送端到接收端的多目标广播树已经确立,如图 2-14 所示。此外,每个接收端的数据接收速率已经在图上标出,而且电视数据是按照接收端数据接

收速率进行分层编码的。

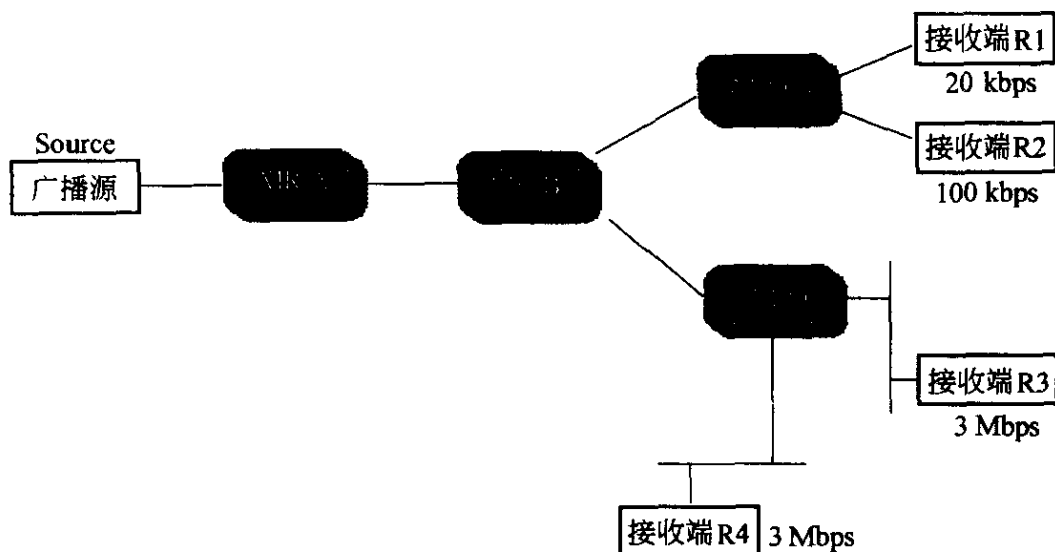


图 2-14 多目标实况广播举例

RSVP 的工作过程大致如下：

每个接收端逆向发送一个资源保留消息到多目标广播树，这个消息说明了接收端接收广播源的数据速率。

当保留消息到达一个路由器时，路由器就调整它的信息包调度程序来保留带宽，然后把这个消息送到上游路由器。

路由器逆向保留的带宽数量是根据下流的保留带宽的数量来确定的。

在本例中，接收端 R1、R2、R3 和 R4 分别保留 20 kbps、100 kbps、3 Mbps 和 3 Mbps，因此路由器 D 下面的接收端请求的最大速率为 3 Mbps。

路由器 D 把保留消息发送给路由器 B，请求路由器 B 在这两个路由器之间的链路上保留 3 Mbps 的带宽，因为 R3 和 R4 观看同一个广播，因此不需要保留 6 Mbps 的带宽。

依此类推，路由器 C 请求路由器 B 在路由器 B 和路由器 C 之间的链路上保留 100 kbps 的带宽。

分层编码的数据流保证接收端 R1 的 20 kbps 的数据流包含在 100 kbps 数据流中。一旦路由器 B 接收到来自下面的路由器

的保留消息之后就传送给它的调度程序,然后把新的保留消息递送给上一层的路由器 A。这个消息要求在从路由器 A 到路由器 B 的链路上保留 3 Mbps 的带宽,这是接收端要求保留的最大带宽。

从上面的分析可以看到,RSVP 是接收端导向(receiver-oriented)的协议,也就是由接收数据流的终端提出资源保留请求。我们也注意到,每个路由器接收到的消息依次是从多目标广播树上的下流链路上发送来的,而且只有一个带宽保留消息。

例 2 4 人电视会议

假设每人自己的计算机屏幕上开有 3 个窗口观看其他 3 人的情况,按照路由协议在 4 台计算机之间建立多目标广播树,他们都使用 3 Mbps 的速率来接收电视,如图 2-15 所示。在这个多目标广播树上的每条链路上,RSVP 就要在一个方向保留 9 Mbps 的带宽,而在其他方向上保留 3 Mbps 的带宽。在这个例子中,RSVP 不合并保留带宽,因为每个人都想接收来自其他 3 台计算机的数据流。

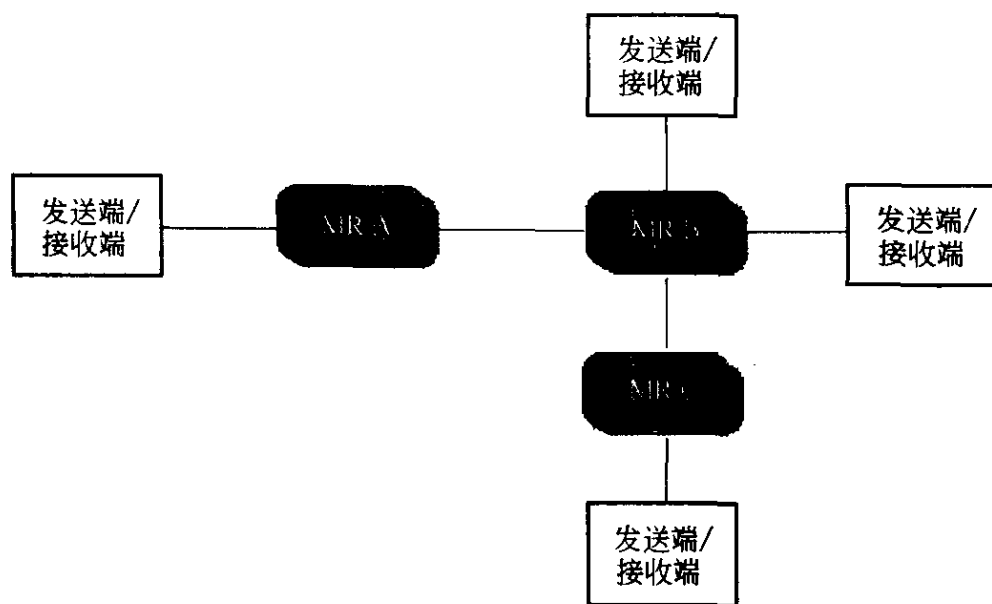


图 2-15 4 人电视会议

2.4.3 接纳测试

由于路由器在链路上保留的带宽的数量不能超过链路本身的能力,因此每当路由器接收一个新的保留消息时,它必须首先判断多目标广播树上的上流链路是否可以容纳。

如果接纳测试(Admission Test)失败,路由器就拒绝保留带宽,并且给请求保留带宽的接收端发送一个错误消息。RSVP 不定义和执行接纳测试,而是由路由器来承担这种测试。

2.4.4 路径消息

路径消息(path messages),是从发送端发出并向接收端发送的消息。

发送路径消息的主要目的是要让路由器知道在哪些链路上应该转发保留消息,尤其是在多目标广播树中从路由器 A 到路由器 B 发出的路径消息,这个消息中包含有路由器 A 的单目标广播 IP 地址。路由器 B 把这个地址放在路径状态表(path-state table)中,当它从下流节点接收到保留消息时,它就访问这张路径状态表,从而得知它应该向路由器 A 发送一个带宽保留消息。

路径消息也包含发送端的交通特性,它定义发送端将要生成的数据流的特性,这个消息可以用来预防带宽保留过多。

第 3 章

数字音频和视频

3.1 数字音频

声音是携带信息的极其重要的媒体,是多媒体技术研究中的一个重要内容。声音的种类繁多,如人的话音、乐器声、动物发出的声音、机器产生的声音以及自然界的雷声、风声、雨声等等。这些声音有许多共同的特性,也有它们各自的特性。在用计算机处理这些声音时,既要考虑它们的共性,又要利用它们各自的特性。本节将介绍声音的基础知识,重点掌握声音数字化的两个最基本的概念。此外,还介绍在上网浏览或者脱机工作时你会经常遇到的声音文件存储格式和声音工具。

3.1.1 声音与听觉器官

声音是通过空气传播的一种连续的波,叫声波。声音的强弱体现在声波压力的大小上,音调的高低体现在声音的频率上。声音用电表示时,声音信号在时间和幅度上都是连续的模拟信号。声波具有普通波所具有的特性,例如反射(reflection)、折射(refraction)和衍射(diffraction)等等。

对声音信号的分析表明,声音信号由许多频率不同的信号组成,这类信号称为复合信号,而单一频率的信号称为分量信号。声音信号的一个重要参数就是带宽,它用来描述组成复合信号的频率范围。例如,高保真音响信号(high-fidelity audio)的频率范围为10~20000 Hz,它的带宽约为20 kHz,而视频信号的带宽是6 MHz。

声音信号的两个基本参数是频率和幅度。信号的频率是指信号每秒钟变化的次数,用Hz表示。例如,大气压的变化周期很长,以小时或天数计算,一般人不容易感到这种气压信号的变化,更听不到这种变化。对于频率为几Hz到20 Hz的空气压力信号,人们也听不到,如果它的强度足够大,也许可以感觉到。人们把频

率小于 20 Hz 的信号称为亚音信号,或称为次音信号(subsonic);频率范围为 20 Hz~20 kHz 的信号称为音频信号(audio)。虽然人的发音器官发出的声音频率大约是 80~3 400 Hz,但人说话的信号频率通常为 300~3 000 Hz,人们把在这种频率范围的信号称为话音信号(speech);高于 20 kHz 的信号称为超音频信号,或称超声波信号(ultrasonic)。超音频信号具有很强的方向性,而且可以形成波束,在工业上得到广泛的应用,如超声波探测仪,超声波焊接设备等就是利用这种信号。在多媒体技术中处理的信号主要是音频信号,它包括音乐、话音、风声、雨声、鸟叫声、机器声等。

人们是否都能听到音频信号,这主要取决于每个人的年龄和耳朵的特性。一般来说,人的听觉器官能感知的声音频率大约在 20~20 000 Hz 之间,在这种频率范围里感知的声音幅度大约在 0~120 dB 之间。除此之外,人的听觉器官对声音的感知还有一些重要特性,这些特性将在后续章节中介绍,它们在声音数据压缩中已经得到广泛的应用。

3.1.2 声音信号数字化

1. 从模拟过渡到数字

回顾历史,大多数电信号的处理一直是用模拟元部件(如晶体管、变压器、电阻、电容等)对模拟信号进行处理。但是,开发一个具有相当精度,且几乎不受环境变化影响的模拟信号处理元部件是相当困难的,而且成本也很高。

如果把模拟信号转变成数字信号,用数字来表示模拟量,对数字信号做计算,那么难点就发生了转移,这就把开发模拟运算部件的问题转变成开发数字运算部件的问题,于是出现了数字信号处理器 DSP(Digital Signal Processor)。DSP 与通用微处理器相比,除了它们的结构不同外,其基本差别是,DSP 有能力响应和处理采样模拟信号得到的数据流,如做乘法和累加求和运算。

在数字域而不在模拟域中做信号处理有许多优点。首先,数

字信号计算是一种精确的运算方法,它不受时间和环境变化的影响;其次,表示部件功能的数学运算不是物理上实现的功能部件,而是仅用数学运算去模拟,其中的数学运算也相对容易实现;此外,可以对数字运算部件进行编程,如欲改变算法或改变某些功能,还可对数字部件进行再编程。

2. 模拟信号与数字信号

话音信号是典型的连续信号,不仅在时间上是连续的,而且在幅度上也是连续的。在时间上“连续”是指在一个指定的时间范围里声音信号的幅值有无穷多个,在幅度上“连续”是指幅度的数值有无穷多个。我们把在时间和幅度上都是连续的信号称为模拟信号。

在某些特定的时刻对这种模拟信号进行测量叫做采样(sampling),由这些特定时刻采样得到的信号称为离散时间信号。采样得到的幅值是无穷多个实数值中的一个,因此幅度还是连续的。如果把信号幅度取值的数目加以限定,这种由有限个数值组成的信号就称为离散幅度信号。例如,假设输入电压的范围是 $0.0 \sim 0.7 \text{ V}$,并假设它的取值只限定在 $0, 0.1, 0.2, \dots, 0.7$ 共8个值。如果采样得到的幅度值是 0.123 V ,它的取值就应算作 0.1 V ;如果采样得到的幅度值是 0.26 V ,它的取值就算作 0.3 V , $\dots$,这种数值就称为离散数值。我们把时间和幅度都用离散的数字表示的信号称为数字信号。

3. 声音信号数字化

声音进入计算机的第一步就是数字化,数字化实际上是采样和量化。如前所述,连续时间的离散化通过采样来实现,就是每隔相等的一小段时间采样一次,这种采样称为均匀采样(uniform sampling)。连续幅度的离散化通过量化(quantization)来实现,就是把信号的强度划分成一小段一小段,如果幅度的划分是等间隔的,就称为线性量化,否则就称为非线性量化。图3-1表示了声音数字化的概念。

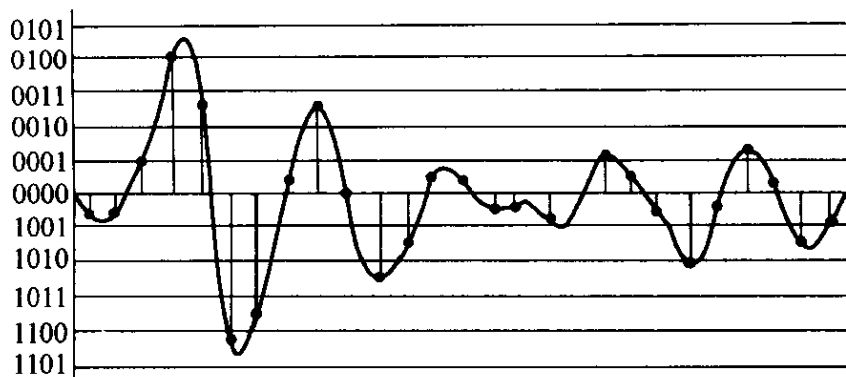


图 3-1 声音的采样和量化

声音数字化需要回答两个问题：

- ① 每秒钟需要采集多少个声音样本,也就是采样频率(f_s)是多少;
- ② 每个声音样本的比特数 bps(bit per sample)应该是多少,也就是量化精度。

4. 采样频率

采样频率的高低是根据奈奎斯特理论和声音信号本身的最高频率决定的。奈奎斯特理论指出,采样频率不应低于声音信号最高频率的两倍,这样就能把以数字表达的声音还原成原来的声音,这叫做无损数字化。

可以这样来理解奈奎斯特理论:声音信号可以看成由许许多多正弦波组成的,一个振幅为 A 、频率为 f 的正弦波至少需要两个采样样本表示,因此,如果一个信号中的最高频率为 $f_{\max}$,采样频率最低要选择 $2f_{\max}$ 。例如,电话语音的信号频率约为 3.4 kHz,采样频率就选为 8 kHz。

5. 采样精度

样本大小是用每个声音样本的比特数 bps 即 bps 表示的,它反映度量声音波形幅度的精度。例如,每个声音样本用 16 比特(2 字节)表示,测得的声音样本值是在 0~65 536 的范围里,它的精度就是输入信号的 $1/65\,536$ 。样本比特数的大小影响到声音的质

量,比特数越多,声音的质量越高,而需要的存储空间也越多;比特数越少,声音的质量越低,需要的存储空间就越少。

6. 声音质量与数据率

根据声音的频带,通常把声音的质量分成 5 个等级,由低到高分别是电话、调幅 AM 广播、调频 FM、光盘 CD 和数字录音带 DAT 的声音。在这 5 个等级中,使用的采样频率、样本精度、通道数和数据率列于表 3-1。

表 3-1 声音质量和数据率

质量	采样频率 (kHz)	样本精度 (bps)	单道声/ 立体声	数据率(未 压缩)(kB/S)	频 率 范 围 (Hz)
电 话	8	8	单道声	8	200~3 400
AM	11.025	8	单道声	11.0	20~15 000
FM	22.050	16	立体声	88.2	50~7 000
CD	44.1	16	立体声	176.4	20~20 000
DAT	48	16	立体声	192.0	20~20 000

3.2 数字视频

电视是 20 世纪 20 年代的伟大发明,在 50 年代开发电视技术时,用任何一种数字技术来传输和再现真实世界的图像和声音都是极其困难的,因此电视技术一直沿着模拟信号处理技术的方向发展,直到 70 年代才开始开发数字电视。由于数字技术具有许多优越性,而且数字技术已经发展到足以使模拟电视向数字电视过渡的水平,电视和计算机才开始融合在一起。

彩色数字电视是从模拟彩色电视发展而来的,因此本节的前半部分将介绍模拟电视的一些基本常识,而后半部分将介绍彩色数字电视的基本常识。这些基本常识对理解 MPEG 电视是极其有用的。

3.2.1 彩色电视制式

目前世界上现行的彩色电视制式有三种:NTSC 制式、PAL 制式和 SECAM 制式。这里不包括高清晰度彩色电视 HDTV (High-Definition television)。数字彩色电视是从模拟彩色电视基础上发展而来的,因此在多媒体技术中经常会碰到这些术语。

NTSC(National Television Systems Committee)彩色电视制式是 1952 年美国国家电视标准委员会定义的彩色电视广播标准,称为正交平衡调幅制式。美国、加拿大等大部分西半球国家,以及中国的台湾、日本、韩国、菲律宾等都采用这种制式。

由于 NTSC 制式存在相位敏感造成彩色失真的缺点,因此德国(当时的西德)于 1962 年制定了 PAL(Phase-Alternative Line)制式彩色电视广播标准,称为逐行倒相正交平衡调幅制式。德国、英国等一些西欧国家,以及中国、朝鲜等国家采用这种制式。

法国制定了 SECAM (法文:Sequential Couleur Avec Memoire)彩色电视广播标准,称为顺序传送彩色与存储制式。法国、苏联及东欧国家采用这种制式。世界上约有 65 个地区和国家试验这种制式。

NTSC 制式、PAL 制式和 SECAM 制式都是兼容制式。这里说的“兼容”有两层意思:一是指黑白电视机能接收彩色电视广播,显示的是黑白图像;另一层意思是彩色电视机能接收黑白电视广播,显示的也是黑白图像,这叫逆兼容性。为了既能实现兼容性而又要彩色特性,因此彩色电视系统应满足下列几方面的要求:

(1) 必须采用与黑白电视相同的一些基本参数,如扫描方式、扫描行频、场频、帧频、同步信号、图像载频、伴音载频等等。

(2) 需要将摄像机输出的三基色信号转换成一个亮度信号,以及代表色度的两个色差信号,并将它们组合成一个彩色全电视信号进行传送。在接收端,彩色电视机将彩色全电视信号重新转换成三个基色信号,在显像管上重现发送端的彩色图像。

3.2.2 电视扫描和同步

扫描有隔行扫描(interlaced scanning)和非隔行扫描之分。非隔行扫描也称逐行扫描,图 3-2 表示了这两种扫描方式的差别。黑白电视和彩色电视都用隔行扫描,而计算机显示图像时一般都采用非隔行扫描。

在非隔行扫描中,电子束从显示屏的左上角一行接一行地扫到右下角,在显示屏上扫一遍就显示一幅完整的图像,如图 3-2(a)所示。

在隔行扫描中,电子束扫完第 1 行后回到第 3 行开始的位置接着扫,如图 3-2(b)所示,然后在第 5,7,⋯,行上扫,直到最后一行。奇数行扫完后接着扫偶数行,这样就完成了一帧(frame)的扫描。由此可以看到,隔行扫描的一帧图像由两部分组成:一部分是由奇数行组成,称奇数场;另一部分是由偶数行组成,称为偶数场,两场合起来组成一帧。因此在隔行扫描中,无论是摄像机还是显示器,获取或显示一幅图像都要扫描两遍才能得到一幅完整的图像。

在隔行扫描中,扫描的行数必须是奇数。如前所述,一帧画面分两场,第一场扫描总行数的一半,第二场扫描总行数的另一半。隔行扫描要求第一场结束于最后一行的一半,不管电子束如何折回,它必须回到显示屏顶部的中央,这样就可以保证相邻的第二场扫描恰好嵌在第一场各扫描线的中间。正是这个原因,才要求总的行数必须是奇数。

每秒钟扫描多少行称为行频 f_H ;每秒钟扫描多少场称为场频 f_i ;每秒扫描多少帧称帧频 f_F 。 f_i 和 f_F 是两个不同的概念。

1. PAL 制式电视的扫描特性

PAL 电视制式的主要扫描特性是:

- ① 625 行(扫描线)/帧,25 帧/秒(40 ms/帧);
- ② 高宽比(aspect ratio)是 4:3;
- ③ 隔行扫描,2 场/帧,312.5 行/场;
- ④ 颜色模型是 YUV。

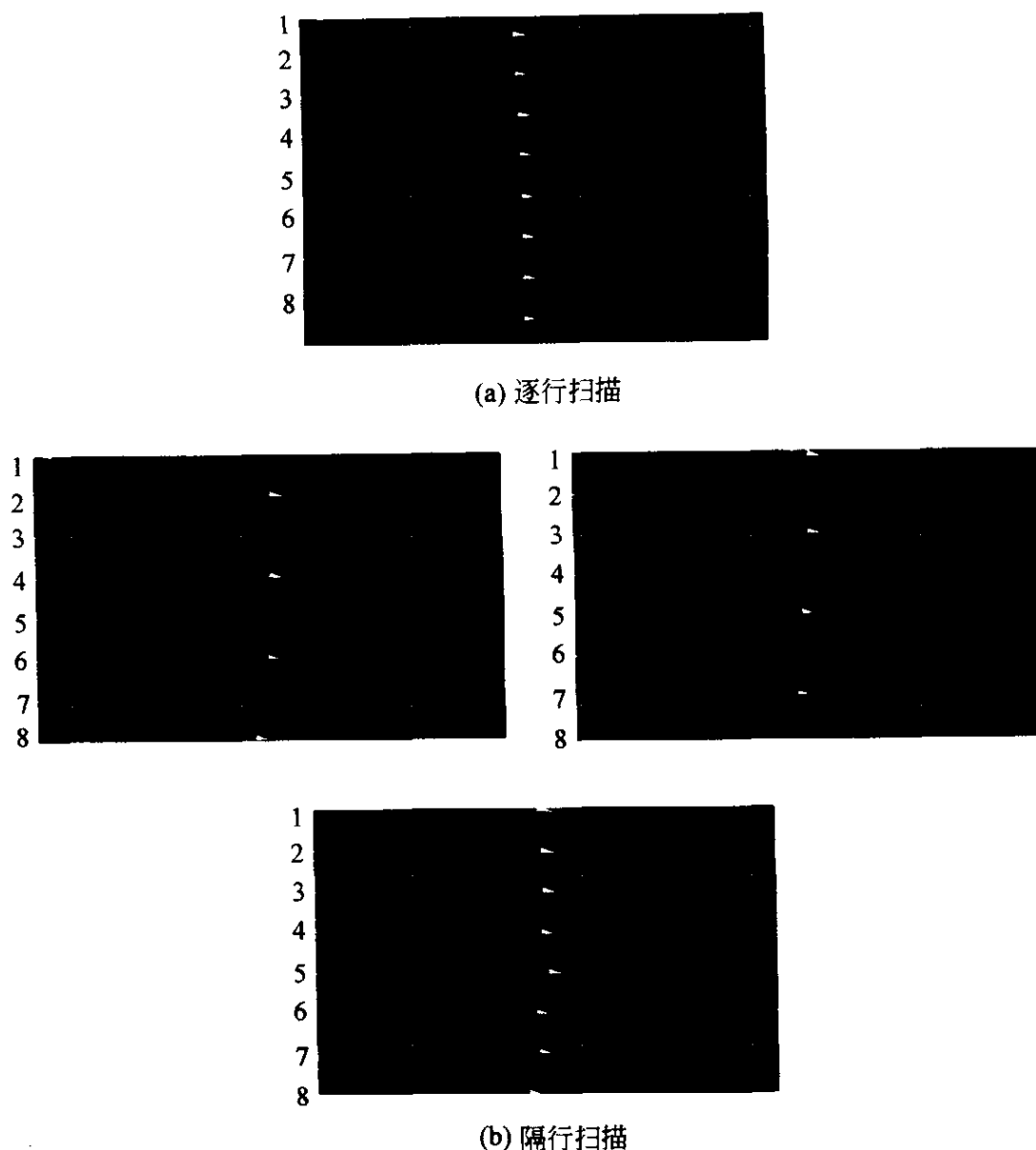


图 3-2 图像的光栅扫描

一帧图像的总行数为 625, 分两场扫描。行扫描频率是 15 625 Hz, 周期为 $64\ \mu\text{s}$; 场扫描频率是 50 Hz, 周期为 20 ms; 帧频是 25 Hz, 是场频的一半, 周期为 40 ms。在发送电视信号时, 每一行中传送图像的时间是 $52.2\ \mu\text{s}$, 其余的 $11.8\ \mu\text{s}$ 不传送图像, 是行扫描的逆程时间, 同时用作行同步及消隐用。每一场的扫描行数为 $625/2=312.5$ 行, 其中 25 行作场回扫, 不传送图像, 传送图像的行数每场只有 287.5 行, 因此每帧只有 575 行有图像显示。

2. NTSC 制式的扫描特性

NTSC 彩色电视制式的主要特性是:

- ① 525 行/帧, 30 帧/秒(29.97 fps, 33.37 ms/frame);
- ② 高宽比为电视画面的长宽比(电视为 4:3; 电影为 3:2; 高清晰度电视为 16:9);
- ③ 隔行扫描, 一帧分成 2 场(field), 262.5 线/场;
- ④ 在每场的开始部分保留 20 扫描线作为控制信息, 因此只有 485 条线的可视数据。Laserdisc 约 420 线, S-VHS 约 320 线;
- ⑤ 每行 $63.5\mu\text{s}$, 水平回扫时间 $10\mu\text{s}$ (包含 $5\mu\text{s}$ 的水平同步脉冲), 所以显示时间是 $53.5\mu\text{s}$ 。
- ⑥ 颜色模型 YIQ: 一帧图像的总行数为 525 行, 分两场扫描。行扫描频率为 15 750 Hz, 周期为 $63.5\mu\text{s}$; 场扫描频率为 60 Hz, 周期为 16.67 ms; 帧频为 30 Hz, 周期为 33.33 ms。每一场的扫描行数为 $525/2 = 262.5$ 行。除了两场的场回扫外, 实际传送图像的行数为 480 行。

3.2.3 彩色电视

彩色电视是在黑白电视基础上发展起来的。彩色电视的许多特性, 如扫描、同步等都与黑白电视相同, 不同的是显示图像的颜色不同。

根据三基色的基本原理, 任何一种颜色都可以用 R, G, B 三个彩色分量按一定的比例混合得到, 但要精确地重显自然景物中的色彩的确是相当困难的。值得庆幸的是, 科学家们对人的彩色视觉特性经过长期研究后发现, 在重显自然景物彩色过程中, 并不一定要恢复原景物辐射的所有光波成分, 而重要的是获得与原景物相同的色彩感觉。

按照色度学的基本原理, 用 R, G, B 三基色的各种线性组合可以构造出各种不同的色彩空间来表示景物的颜色。各种不同的色彩空间在不同的应用中也许会比原始的 RGB 色彩空间具有更有

用的特性,更有效且更经济。因此在彩色电视中,用 Y, C_1, C_2 色彩表示法分别表示亮度信号和两个色差信号, C_1, C_2 的含义与具体的应用有关。在 NTSC 彩色电视制式中, C_1, C_2 分别表示 I, Q 两个色差信号;在 PAL 彩色电视制式中, C_1, C_2 分别表示 U, V 两个色差信号;在 CCIR 601 数字电视标准中, C_1, C_2 分别表示 Cr, Cb 两个色差信号。所谓色差是指基色信号中的三个分量信号(即 R, G, B)与亮度信号之差。

在彩色电视中,使用 Y, C_1, C_2 有两个重要优点:

① Y 和 C_1, C_2 是独立的,因此彩色电视和黑白电视可以同时使用, Y 分量可由黑白电视接收机直接使用而不需做任何进一步的处理;

② 可以利用人的视觉特性来节省信号的带宽和功率,通过选择合适的颜色模型,可以使 C_1, C_2 的带宽明显低于 Y 的带宽,而又不明显影响重显彩色图像的观看。

因此,为了满足兼容性的要求,彩色电视系统选择了一个亮度信号和两个色差信号,而不直接选择三个基色信号进行发送和接收。

3.2.4 彩色电视信号的类型

1. 复合电视信号

包含亮度信号、色差信号和所有定时信号的单一信号叫做复合电视信号(composite video signal),或者称为全电视信号。

2. 分量电视信号

分量电视信号(component video signal)是指每个基色分量作为独立的电视信号。每个基色既可以用 RGB 表示,也可以用亮度-色差表示,如 YIQ, YUV 。使用分量电视信号是表示颜色的最好方法,但需要比较宽的带宽和同步信号。

3. S-Video 信号

分离电视信号 S-Video(Separated video-VHS)是亮度和色差

分离的一种电视信号,是分量模拟电视信号和复合模拟电视信号的一种折中方案。使用 S-Video 有两个优点:

- ① 减少亮度信号和色差信号之间的交叉干扰。
- ② 不须要使用梳状滤波器来分离亮度信号和色差信号,这样可提高亮度信号的带宽。

复合电视信号是把亮度信号和色差信号复合在一起,使用一条信号电缆线传输。而 S-Video 信号则使用单独的两条信号电缆线,一条用于亮度信号,另一条用于色差信号,这两个信号称为 Y/C 信号。S-Video 使用 4 针连接器,如图 3-3 所示。具体的规格如表 3-2 所示。

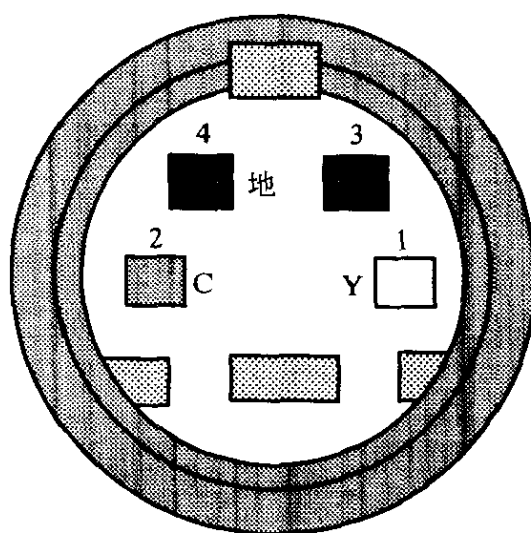


图 3-3 S-Video 连接器

表 3-2 S-Video 工业标准 4 针连接器规格

插座号	信号	信号电平	阻抗
1	地(亮度)	—	—
2	地(色度)	—	—
3	亮度(包含同步信号)	1 V	75 ohms
4	色度	0.3 V	75 ohms

需要注意的是,不要把 S-Video 和 S-VHS (Super Video Home System) 相混淆, S-VHS 是高档家用录像系统。S-Video 是定义信号电缆连接插座的硬件标准, S-VHS 或者写成 SVHS 是加强型

VHS 电视录像带的信号标准,提供的分辨率比 VHS 提供的分辨率要高一些,噪声信号要低一些。S-VHS 支持分离的亮度和色差信号输入/输出,取消了亮度和色差的复合-分离过程。

3.2.5 电视图像数字化

数字电视图像有很多优点。例如,可直接进行随机存储使电视图像的检索变得很方便,复制数字电视图像和在网络上传输数字电视图像都不会造成质量下降,很容易进行非线性电视编辑。

在大多数情况下,数字电视系统都希望用彩色分量来表示图像数据,如用 YCbCr、YUV、YIQ 或 RGB 彩色分量。因此,电视图像数字化常用“分量初始化(component digitization)”这个术语,它表示对彩色空间的每一个分量进行初始化。电视图像数字化常用的方法有两种:

(1) 先从复合彩色电视图像中分离出彩色分量,然后数字化。我们现在接触到的大多数电视信号源都是彩色全电视信号,如来自录像带、激光视盘、摄像机等的电视信号。对这类信号的数字化,通常的做法是首先把模拟的全彩色电视信号分离成 YCbCr、YUV、YIQ 或 RGB 彩色空间中的分量信号,然后用三个 A/D 转换器分别对它们数字化。

(2) 首先用一个高速 A/D 转换器对彩色全电视信号进行数字化,然后在数字域中进行分离,以获得所希望的 YCbCr, YUV, YIQ 或 RGB 分量数据。用这种方法对电视图像数字化。

早在 20 世纪 80 年代初,国际无线电咨询委员会 CCIR(International Radio Consultative Committee)就制定了彩色电视图像数字化标准,称为 CCIR 601 标准,现改为 ITU-R BT.601 标准。该标准规定了彩色电视图像转换成数字图像时使用的采样频率,RGB 和 YCbCr(或者写成 YCBCR)两个彩色空间之间的转换关系等。

CCIR 为 NTSC 制式、PAL 制式和 SECAM 制式规定了共同的电视图像采样频率。这个采样频率也适用于远程图像通信网络

中的电视图像信号采样,如 ISDN、电视会议、CCITT-H.261 及光纤通信等。

ITU-R BT.601 用于对隔行扫描电视图像进行数字化,对 NTSC 和 PAL 制式彩色电视的采样频率和有效显示分辨率都作了规定。表 3-3 给出了 ITU-R BT.601 推荐的采样格式、编码参数和采样频率。

表 3-3 彩色电视数字化参数摘要

采样格式	信号形式	采样频率 (MHz)	样本数/扫描行		数字信号取值 范围(A/D)
			NTSC	PAL	
4:2:2	Y	13.5	858(720)	864(720)	220 级(16~235)
	Cr	6.75	429(360)	432(360)	225 级(16~240)
	Cb	6.75	429(360)	432(360)	(128 □ 112)
4:4:4	Y	13.5	858(720)	864(720)	220 级(16~235)
	Cr	13.5	858(720)	864(720)	225 级(16~240)
	Cb	13.5	858(720)	864(720)	(128 □ 112)

ITU-R BT.601 推荐使用 4:2:2 的彩色电视图像采样格式。使用这种采样格式时,Y 用 13.5 MHz 的采样频率,Cr,Cb 用 6.75 MHz 的采样频率。采样时,采样频率信号要与场同步和行同步信号同步。

为了既可用 625 行的电视图像又可用 525 行的电视图像,CCITT 规定了称为公用中分辨率格式 CIF(Common Intermediate Format),1/4 公用中分辨率格式 QCIF(Quarter-CIF)和 SQCIF(Sub-Quarter Common Intermediate Format)格式,具体规格如表 3-4 所示。

CIF 格式具有如下特性:

表 3-4 CIF 和 QCIF 图象格式参数

	CIF		QCIF		SQCIF	
	行数/帧	像素/行	行数/帧	像素/行	行数/帧	像素/行
亮度(Y)	288	360(352)	144	180(176)	96	128
色度(Cb)	144	180(176)	72	90(88)	48	64
色度(Cr)	144	180(176)	72	90(88)	48	64

① 电视图像的空间分辨率为家用录像系统 VHS (Video Home System) 的分辨率, 即 352×288 。

② 使用非隔行扫描 (non-interlaced scan)。

③ 使用 NTSC 帧速率, 电视图像的最大帧速率为 $30\,000 / 1\,001 \approx 29.97$ 幅/秒。

④ 使用 $1/2$ 的 PAL 水平分辨率, 即 288 线。

⑤ 对亮度和两个色差信号 (Y、Cb 和 Cr) 分量分别进行编码, 它们的取值范围同 ITU-R BT.601。即黑色 = 16, 白色 = 235, 色差的极大值等于 240, 最小值等于 16。

第4章

数据压缩技术

随着数字化通信、计算机网络以及多媒体技术的飞速发展,符合人类信息交换方式的多媒体通信已经成为一种不可避免的趋势。据统计,人类感觉器官接收的各类信息中,语音和图像占了80%,因而在现有的技术条件下,多媒体通信主要还是以图像和语音通信为主,并辅以少量的文字信息。这些不同媒体的信息经数字化后,如果不进行压缩,其数据量十分巨大,例如:

1. 音频信息

对于电话语音质量的单声道的音频信号,如果采用8 kHz采样、8 bit 量化,其比特流达到64 kbps,这样15分钟音频信号的数据量将达到7.2 M。而采用48 kHz、16 bit 量化的立体声数字音响源的数据流将高达1 536 kbps,15分钟的数据量就有172.8 M。

2. 图像信息

对于静止图像来说,类似于FAX传真质量的二值图像,如果采用1bit 量化,空间分辨率为3.85或7.7点/毫米,一张 $210 \times 297\text{mm}^2$ 的A4纸大小的图像数据量为116 kbytes。而一个3 072像素 $\times$ 1 024像素,每个彩色分量用8比特量化的数字彩色图像的数据量为9.4M。

对于活动图像来说,大小为 160×120 、灰度级用8 bits 量化、帧率为每秒5帧的黑白图像序列的数据流的大小为768 kbps,每15分钟的数据量为173 M。而高质量的彩色活动图像的数据流可高达312 Mb/s,60分钟的数据量可高达211 G字节。

显然以现有 10~100 M 的 LAN 的带宽以及现有的存储介质和手段,传输和存储这样浩大的数据是几乎不可能的,因而多媒体数据的压缩成为了解决多媒体通信和存储问题的一个关键环节。

4.1 多媒体数据压缩依据

由于各种多媒体数据之间存在着各种形式的相关性,因而存在着很大的数据冗余。这些冗余数据对于信息的表达是意义不大的,因为在绝大多数情况下,只要能够把意义完整的信息传输或保存下来就足够了。因此,去除数据冗余的工作成为多媒体数据压缩的根据。

各种数据压缩技术与方法的提出都是与多媒体数据中数据冗余类型相关的,一般多媒体数据中存在着以下几种数据冗余:

1. 空间冗余

这是图像数据中经常存在的一种冗余。在同一幅图像中,规则物体和规则背景的表面物理特性具有相关性,这些相关性的光成像结果在数字化图像中就表现为数据冗余。

2. 时间冗余

这是序列图像和语音数据中经常包含的冗余。图像序列中前一幅图像和后一幅图像之间存在有较大的相关,这反应为时间冗余。同理,在语音中,由于人在说话时其发音的音频是一个连续和渐变过程,而不是一个在时间上完全独立的过程,因而存在时间冗余。

3. 信息熵冗余(编码冗余)

4. 结构冗余

有些图像从大区域上看有着非常强的纹理结构,例如草席图像,我们称之为结构上存在有冗余。

5. 知识冗余

许多图像的理解与某些基础知识有相当大的相关性,这类规律性的结构可由先验知识和背景知识得到,我们称之为知识冗余。

6. 视觉冗余

人类的视觉系统对于图像场变化的敏感程度是非均匀和非线性的,它并不能感知和察觉所有的图象变化。我们在对图像进行编码和解码处理时,尽管由于压缩和量化截断引入了噪声,使图像发生了一定变化,但这些变化如果不能被视觉系统所察觉的话,我们仍认为图像是完好的或着是足够好的。像这样的冗余,我们称之为视觉冗余。

7. 其他冗余

例如由图象的空间非定常特性所带来的冗余等。

4.2 话音编码

随着数字电话和数据通信日益增长,要想不降低传送话音信号的质量,除了提高通信带宽之外,对话音信号进行压缩是提高通信容量的重要措施。另一个可以说明话音数据压缩的重要性的例子是,用户无法使用 28.8 kbps 的调制解调器来接收因特网上的 64 kbps 话音数据流,这是一种单声道、8 比特/样本、采样频率为 8 kHz 的话音数据流。ITU-TSS 为此制定了并且继续制定一系列话音(speech)数据编译码标准。其中,G.711 使用 μ 率和 A 率压缩算法,信号带宽为 3.4 kHz,压缩后的数据率为 64 kbps;G.721 使用 ADPCM 压缩算法,信号带宽为 3.4 kHz,压缩后的数据率为 32 kbps;G.722 使用 ADPCM 压缩算法,信号带宽为 7 kHz,压缩后的数据率为 64 kbps。以这些标准为基础还制定了许多话音数据压缩标准,例如 G.723,G.723.1,G.728,G.729 和 G.729.A 等等。本章将重点介绍话音编码的基本思想,而详细的计算则留给那些开发和具体设计编译码器软硬件的读者去研究。

4.2.1 话音编码技术

为简单起见,通常把已有的话音编译码器分成三种类型:波形

编译码器(waveform codecs)、音源编译码器(source codecs)和混合编译码器(hybrid codecs)。一般来说,波形编译码器的话音质量高,但数据率也很高;音源编译码器的数据率很低,产生的合成话音的音质有待提高;混合编译码器使用音源编译码技术和波形编译码技术,数据率和音质介于它们之间。

1. 波形编译码器

波形编译码的想法是,不利用生成话音信号的任何知识而企图产生一种重构信号,它的波形与原始话音波形尽可能地一致。一般来说,这种编译码器的复杂程度比较低,数据速率在 16 kbps 以上,质量相当高。低于这个数据速率时,音质急剧下降。

最简单的波形编码是脉冲编码调制 PCM(Pulse Code Modulation),这种方法仅仅是对输入信号进行采样和量化。典型的窄带话音带宽限制在 4 kHz,采样频率是 8 kHz。如果要获得高一点的音质,样本精度要用 12 比特,它的数据率就等于 96 kbps。这个数据率可以使用非线性量化来降低,例如可以使用近似于对数的对数量化器(logarithmic quantizer),使用它产生的样本精度为 8 位,它的数据率为 64 kbps 时,重构的话音信号几乎与原始的话音信号没有什么差别。这种量化器在 20 世纪 80 年代就已经标准化,而且直到今天还在广泛使用。在美洲的压扩(companding)标准是 μ 律(μ -law),在欧洲的压扩标准是 A 律(A-law)。它们的优点是编译码器简单,延迟时间短,音质高。它们的不足之处是数据速率比较高,对传输通道的错误比较敏感。

在话音编码中,一种普遍使用的技术叫做预测技术,这种技术是企图以过去的样本来预测下一个样本的值。这样做的根据是认为在话音样本之间存在相关性。如果样本的预测值与样本的实际值比较接近,它们之间的差值幅度的变化就比原始话音样本幅度值的变化小,因此量化这种差值信号时就可以用比较少的比特数来表示差值。这就是差分脉冲编码调制 DPCM(Differential Pulse Code Modulation)的基础——对预测的样本值与原始的样本值之差进行编码。

这种编译码器对幅度急剧变化的输入信号会产生比较大的噪声,改进的方法之一就是使用自适应的预测器和量化器,这就产生了一种叫做自适应差分脉冲编码调制 ADPCM(Adaptive Differential PCM)。在 20 世纪 80 年代,国际电话与电报顾问委员会 CCITT (International Telephone and Telegraph Consultative Committee), 现改为 ITU-TSS (International Telecommunications Union-Telecommunications Standards Section), 即国际电信联盟-远程通信标准部制定了数据率为 32 kbps 的 ADPCM 标准,它的音质非常接近 64 kbps 的 PCM 编译码器,随后又制定了数据率为 16、24 和 40 kbps 的 ADPCM 标准。

上述的所有波形编译码器完全是在时间域里开发的,在时域里的编译码方法称为时域法(time domain approach)。在开发波形编译码器的过程中,人们还使用了另一种方法,叫做频域法(frequency domain approach)。例如,在子带编码 SBC(Sub-Band Coding)中,输入的话音信号被分成好几个频带(即子带),变换到每个子带中的话音信号都进行独立编码,如使用 ADPCM 编码器编码,在接收端,每个子带中的信号单独解码之后重新组合,然后产生重构话音信号。这样做的优点是每个子带中的噪声信号仅仅与该子带使用的编码方法有关。对听觉感知比较重要的子带信号,编码器可分配比较多的位数来表示它们,于是在这些频率范围里噪声就比较低。对于其他的子带,由于对听觉感知的重要性比较低,允许比较高的噪声,于是编码器就可以分配比较少的位数来表示这些信号。自适应位分配的方案也可以考虑用来进一步提高音质。子带编码需要用滤波器把信号分成若干个子带,这比使用简单的 ADPCM 编译码器复杂,而且还增加了更多的编码时延。即使如此,与大多数混合编译码器相比,子带编译码的复杂性和时延相对来说还是比较低的。

另一种频域波形编码技术叫做自适应变换编码 ATC(Adaptive Transform Coding)。这种方法使用快速变换(例如离散余弦变换)把话音信号分成许许多多的频带,用来表示每个变换系数的

位数取决于话音谱的性质,获得的数据率可低到16 kbps。

2. 音源编译码器

音源编译码的想法是企图从话音波形信号中提取生成话音的参数,使用这些参数通过话音生成模型重构出话音。针对话音的音源编译码器叫做声码器(vocoder)。在话音生成模型中,声道被等效成一个随时间变化的滤波器,叫做时变滤波器(time-varying filter),它由白噪声(无声话音段)、或者由脉冲串(有声话音段)激励。因此需要传送给解码器的信息就是滤波器的规格、发声或者不发声的标志和有声话音的音节周期,并且每隔10~20 ms更新一次。声码器的模型参数既可以使用时域的方法也可以使用频域的方法确定,这项任务由编码器完成。

这种声码器的数据率在2.4 kbps左右,产生的语音虽然可以听懂,但其质量远远低于自然语音。增加数据率对提高合成语音的质量无济于事,这是因为受到话音生成模型的限制。尽管它的音质比较低,但它的保密性能好,因此这种编译码器一直用在军事上。

3. 混合编译码

混合编译码的想法是企图填补波形编译码和音源编译码之间的间隔。波形编译码器虽然可提供高质量的话音,但在数据率低于16 kbps的情况下,技术上还没有解决音质的问题;声码器的数据率虽然可降到2.4 kbps甚至更低,但它的音质根本不能与自然语音相提并论。为了得到音质高而数据率低的编译码器,历史上出现过很多形式的混合编译码器,其中最成功并且普遍使用的编译码器是时域合成-分析 AbS(Analysis-by-Synthesis)编译码器。这种编译码器使用的声道线性预测滤波器模型与线性预测编码 LPC(Linear Predictive Coding)使用的模型相同,不使用两个状态(有声/无声)的模型来寻找滤波器的输入激励信号,而是企图寻找这样一种激励信号,使用这种信号激励产生的波形尽可能接近于原始语音的波形。AbS编译码器由 Atal 和 Remde 在1982年首次提出,并命名为多脉冲激励 MPE(Multi-Pulse Excited)编译码器,

在此基础上随后出现的是等间隔脉冲激励 RPE (Regular-Pulse Excited) 编译码器、码激励线性预测 CELP (Code Excited Linear Predictive) 编译码器和混合激励线性预测 MELP (Mixed Excitation Linear Prediction) 等编译码器。

4.2.2 音频编码国际标准

1. G.711

该标准针对 50 Hz 到 3.4 kHz 的一般语音信号进行 PCM 编码, 取样频率为 8 kHz, 比特率为 64 kbps。

2. G.721

该标准通过自适应差分脉冲编码调制 ADPCM 方法将 64 kbps 的数字语音信号压缩到 32 kbps。

3. G.722

这是一种高质量语音信号(带宽为 50 Hz ~ 7 kHz)的压缩标准。采样频率为 16 kHz, 每个样值量化到 14b。它采用子带编码和 ADPCM (SB-ADPCM) 进行压缩编码, 压缩后的比特率为 48, 56 或 64 kbps。

4. G.728

该标准利用低延时码本激励线性预测 LD-CELP 方法将一般语音信号压缩至 16 kbps。CELP 是一种综合了波形编码和参数编码的混合编码方法。

5. G.729

它是一种低码率的语音压缩标准, 采用共轭结构代数码本激励线性预测 CS-ACELP 算法, 将一般语音信号压缩至 8 kbps。

6. G.723.1

该标准将一般语音信号压缩到 5.3 kbps 和 6.3 kbps, 二者均使用 ACELP 算法, 其主要区别在于激励码本的不同。在 6.3 kbps

编码器内采用了多脉冲激励,重建的音质稍高一些,可以与 G.729 在 8 kb/s 上的音质相比拟,但编码延时较大。G.723.1 在低码率的多媒体通信中占有重要的地位,特别是在公用电话网、移动网和因特网上的语音通信中都得到了广泛的应用。

4.3 视频编码

4.3.1 数字图像编码技术

数字图像编码方法是视频编码的基础,它主要是通过去除图像中非时间相关的其他相关冗余信息来达到压缩编码的目的。现有的数字图像编码可分为无失真编码和有限失真编码两大类。

无失真编码是利用图像的统计特性来进行数据压缩的。其主要思想是对图像数据中出现概率大的字符编以短字节的码,而对图像数据中出现概率小的字符编以长字节的码。众所周知的 Huffman 码就是一种典型的无失真编码。无失真编码的主要缺陷是受信源熵值极限的影响,压缩比很低,一般只有几倍。

对于一些结构复杂、细节丰富、灰度层次较多的数字图像(如人像、自然景物等)来说,其数据量往往十分巨大,仅仅进行无失真编码是达不到应用的要求的。但是由于视觉特性的存在,只要把失真度限制在某一个容许的限度内,就可以使人的眼睛看不出或可以容忍这种失真。我们把这种通过将失真限制在某一允许限度以内,使图像编码达到更高压缩比的编码方法称为率失真编码。

最常用的率失真编码是预测编码和变换编码。

预测编码是利用空间(或时间)相关性,用前面出现的 M 个像素之值来预测当前值,再将当前值与预测值之差进行量化编码的方法,差分脉冲调制法(DPCM)是预测编码的典型代表。

变换编码的主要思想是基于图像数据的统计模型。由于图像像素之间存在高度的相关性,因而需要进行某种变换来消除这种相关性。多数图像的统计特性表明,经过变换后一般数值较大的

方差总是集中在少数系数中,而且大幅值的系数往往集中在低频区域内,这样就可以给那些小幅值系数分配很少的比特数,从而达到压缩的目的。正交变换编码和子带编码都是典型的变换编码。正交变换编码不是直接对空域图像信号进行编码,而是先将空域图像信号映射变换到另一个正交矢量空间(如离散 Fourier 变换、DCT 变换、KL 变换和 Walsh 变换等),产生一组变换系数,然后对这些系数量化、编码和传输。子带编码的基本思想是利用一组滤波器,通过重复卷积的方法,经取样将输入信号分解为高频分量和低频分量,然后分别对高频和低频分量进行量化和编码。解码时,高频和低频分量经过插值和共轭滤波合成原信号。

以上所述的编码方法都属于经典的第一代编码方法,而一些现代的图像压缩方法则突破了经典的信源模型与理论框架。如 Kunt 等人提出的利用人的视觉特性的第二代图像压缩编码方法,它充分考虑了人类视觉生理、心理特性,将原始图像在频域内作多层分解,提取边缘、纹理或方向信息等对人眼观察有意义的内容后,再对这些信息灵活地、有选择地编码。80 年代末至 90 年代初出现的小波压缩编码方法就属于这一范畴。

小波编码方法是利用离散小波变换将图像分解为低频分量,水平高频分量,垂直高频分量及对角高频分量,对这四个分量根据人类的视觉生理及心理特点分别作不同的量化与编码处理。如对于低频分量可以采用保真度较好的 DCT 变换结合 Huffman 编码进行压缩,对于其余三个分量,可以采用去掉高频成分,阈值量化和均匀量化结合 Huffman 编码的策略。

另外,常见的现代编码方法还有分形编码。分形编码主要是建立在拼贴定理和迭代函数系统理论的基础上,它利用图像局部与局部、局部与整体间的自相似性冗余进行压缩,不仅消除了图像中的近程相关性,而且有效地消除了自然图像中的远程相关性,因而能够获得相当高的压缩比,特别是在高压缩比的情况下,分形编码的恢复图像质量要优于其他编码方法。目前常见的自动分形图像压缩编码方法都是以 Jacquin 的方法为基础,这类方法用仿射

变换来消除图像块中间的确定自相似性冗余。在编码时将图像分为互不重叠的子块,通常称这些子块为值域块,对每一个值域块编码时都要用搜索的方法在图像中寻找与其相似的定义域块,并按照某种误差准则确定一个相应的收缩仿射变换,使得编码误差满足给定要求。

4.3.2 视频编码技术

视频编码是多媒体通信、数字电视等一系列新兴的多媒体应用中一个非常关键的技术,它使得大数据量和高实时性的视觉信息交换和传播成为可能。无论是在桌面视频会议以及计算机支持的协同工作(CSCW: Computer Supported Cooperative Work)中,还是在提供视频点播、游戏和远程购物服务的交互式娱乐网络中,都包含有大量需要存储和传输的视频信息,这些视频信息因为存储容量和网络带宽的限制,必需进行高效、可靠和灵活地压缩。正是这种需要促使了活动图像数据压缩技术的长足发展。

1990年12月国际标准化组织ITU-T通过了用于可视电话和视频会议的H.261标准,该标准又简称 $P \times 64$ 。当 $P=1$ 或2时,码率最大为128 kbps,只支持QCIF(Quarter Common Intermediate Format)格式,适用于低帧率的可视电话。当 $P>6$ 时,码率大于384 kbps,适用于CIF(Common Intermediate Format)图像格式的视频会议环境。1992年由国际标准化组织ISO和国际电工委员会IEC共同制定了MPEG-1标准。该标准分为三个部分:第一部分为系统描述,介绍了语音和图像数据信号的同步和复用标准;第二部分为视频编码标准;第三部分为音频编码标准,其中视频编码部分采用的压缩方法与H.261的方法基本类似。MPEG-1已广泛应用到活动图像及其数字伴音的介质存储压缩上(如VCD,CD-I)。MPEG-1的码流大小为1.5 Mbps,图像采用CIF格式,分辨率为 352×288 ,帧率为30帧/s,两路立体声伴音的质量接近于CD声音质量。1994年,MPEG组织又为数字广播电视制定了活动图像及其数字伴音压缩

编码标准 MPEG-2, 编号为 IS13818, 其中 IS13818-1 是对 MPEG-2 系统的阐述, IS13818-2 是对视频压缩编码的阐述, IS13818-3 是对音频压缩编码的阐述, 而且视频压缩编码与 ITU-T 的 H.262 完全一致。MPEG-2 的码流为 2~60 Mbps, 支持四个等级的图像分辨率(352×288, 720×576, 1 440×1 152, 1 920×1 152)。

在 1995 和 1996 年, ITU-T 又制定了用于低比特流通信的 H.263 视频编码标准, 最初的 H.263 草案主要是为了支持 64 kbps 以下的视频数据流编码, 但不久这个限制又被取消了, 使得 H.263 可以支持大范围变化的视频比特流编码, 从而可以完全取代 H.261。从视频压缩的效率上说, H.263 编码方案是现有的活动图像编码方法中最好的一种。

现在 MPEG 专家组又在致力于发展用于交互式多媒体应用的 MPEG-4 标准。在 MPEG-4 的视频编码器中采用了基于内容的编码技术, 即图像场景由几个视频对象 (Video Objects) 组成, 如背景、实物、文本等, 对图像场景的编解码建立在这些视频对象上, 这样用户就可以对这些对象分别进行解码和重建, 使得使用者对原始场景能够以一种灵活的方式来操作。MPEG-4 视频编码算法不仅支持所有 MPEG-1 和 MPEG-2 提供的功能, 还能提供基于内容的功能。

扩展的 MPEG-4 基于内容的算法和工具与标准的 MPEG-1 和 MPEG-2 对长方形图像序列进行编码不同, MPEG-4 的视频对象通常被认为是任意形状的。MPEG-4 基于内容的方法可以看作是 MPEG-4 VLBV 和 HBV 核的编码方法向任意形状的视频图像作逻辑扩展的结果, 而传统的长方形图像序列只是 MPEG-4 视频编码对象中的一类特殊对象。

此外, 在超低比特流的视频编码方法中, 基于模型的编码技术越来越受到人们的重视, 有文献报道采用基于模型的编码技术, 可以在不降低图像分辨率和重建质量的情况下实现码流低于 1 000 bps 的高倍压缩。在基于模型的图像编码系统中, 编码器用三维模型对场景进行解析, 解码端在解码器用同样的模型将编码器传来的

参数合成为场景。基于模型的编码的一个重要应用就是可视电话,对于可视电话中的头肩序列,编码器将引起面部表情变化的全局和局部的肌肉形变和刚体运动提取出来,然后将三维运动参数进行编码传输。在解码端已知头肩部模型的情况下,只要传递很少的运动参数就能够重建整幅图像。考虑到模型的局部失真,可以传递一些形状和纹理参数给解码端。

用的各种特性和采用的方法归纳在表 4-1 中。从表中可以看到,电视图像本身在时间上和空间上都含有许多冗余信息,图像自身的构造也有冗余性。此外,正如前面所介绍的,利用人的视觉特性也可对图像进行压缩,这叫做视角冗余。

表 4-1 视频编码利用的各种冗余信息

种 类		内 容	目前用的主要方法
统 计	空间冗余	像素间的相关性	变换编码,预测编码
特 性	时间冗余	时间方向上的相关性	帧间预测,移动补偿
图像构造冗余		图像本身的构造	轮廓编码,区域分割
知识冗余		收发两端对人物的共有认识	基于知识的编码
视觉冗余		人的视觉特性	非线性量化,位分配
其 他		不确定性因素	

视频编码技术的基本思想和方法可以归纳成两个要点:

① 在空间方向上,图像数据压缩采用 JPEG (Joint Photographic Experts Group) 压缩算法来去掉冗余信息。

② 在时间方向上,图像数据压缩采用移动补偿(motion compensation)算法来去掉冗余信息。

4.3.3 视频编码的国际标准

1. H.261 和 H.263 标准简介

H.261 和 H.263 是可视电话和视频会议系统中最为常用的两个视频压缩编码协议,采用的都是基于 DCT 编码和带有运动预测

的 DPCM 预测编码的混合编码方案,能够对运动图像进行高倍的压缩。其中 H.261 标准制定于 1990 年 12 月,常被简称为 $P \times 64$ 。H.263 协议的编码器大致同 H.261 相同,只是在某些方面作了改进,能够获得更佳的图像质量,得到更低的视频码率(低于 64 kbps)。

H.261 和 H.263 主要对图像的亮度和两个色差信号(Y, Cb, Cr)分别进行编码,其中 Cb 和 Cr 信号矩阵是 Y 信号矩阵的 1/4。它们支持的输入图像源的格式如表 4-2 所示。H.261 与 H.263 编码器结构相似。

表 4-2 H.261 和 H.263 支持的图像源格式

图 像 格 式	亮度信号 每行像 素 数	亮度信号 每帧的 行 数	H.261 支持与否	H.263 支持与否	未压缩信号比特流(Mbps)			
					10 帧/秒		30 帧/秒	
					灰 度	彩 色	灰 度	彩 色
SQCIF	128	96		是	1.0	1.5	3.0	4.4
QCIF	176	144	是	是	2.0	3.0	6.1	9.1
CIF	352	288	可选	可选	8.1	12.2	24.3	36.5
4CIF	704	576		可选	32.4	48.7	97.3	146.0
16CIF	1 408	1 152		可选	129.8	194.6	389.3	583.9

H.263 在 H.261 基础上,除了在输入图像格式选项上有所增加外,还对视频编码器作了如下改进:

- (1) 运动补偿采用了半像素精度并使用了环形滤波器。
- (2) 数据流的一些层次结构内容是可选的,使编码器的参数配置达到最优,以取得更低的码流,并能进行有效的差错掩盖和恢复。
- (3) 采用了四种可选择项用以提高执行效率:无范围限制的
运动矢量、基于语义的算术编码、前向预测及能进行前向和后向预测的 PB 帧。这些选项使得 H.263 的编码器能够只用 H.261 一半的码率就能获得同 H.261 相同的解码图像质量。

(4) H.263 支持五种分辨率的图像格式:QCIF, CIF, SQCIF, 4CIF, 16CIF。其中前两种是 H.261 所支持的格式, SQCIF 格式大约只能达到 QCIF 一半的分辨率,而 4CIF 和 16CIF 图像格式的分辨率分别为 CIF 格式的 4 和 16 倍。对 4CIF 和 16CIF 格式的支

持意味着 H.263 编码器在高比特流视频编码中也能够同 MPEG 标准编码器一争高下。近两年来,为了保证能在各种信道、不同的 QoS 需求下进行可靠通信,H.263 编码器又做了许多改进,如提出了用于无线信道的与标准兼容的扩充 H.263 算法等。

这些改进和增加的选项使得 H.263 的编码器能够只用 H.261 一半的码率就能获得同 H.261 相同的解码图像质量。

2. MPEG-1 标准

MPEG-1 处理的是标准图像交换格式 SIF (Standard Interchange format) 或者称为 SIF (Source Input Format) 的电视,即 NTSC 制式为 352 像素 $\times$ 240 行/帧 $\times$ 30 帧/秒,PAL 制式为 352 像素 $\times$ 288 行/帧 $\times$ 25 帧/秒,压缩的输出速率定义在 1.5 Mbps 以下。这个标准主要是针对当时具有这种数据传输率的 CD-ROM 和网络而开发的,用于在 CD-ROM 上存储数字影视和在网络上传输数字影视。

3. MPEG-2 标准

MPEG-2 标准从 1990 年开始研究,1994 发布。它是一个直接与数字电视广播有关的高质量图像和声音编码标准。MPEG-2 可以说是 MPEG-1 的扩充,因为它们的基本编码算法都相同。但 MPEG-2 增加了许多 MPEG-1 所没有的功能,例如增加了隔行扫描电视的编码,提供了位速率的可变性能 (scalability) 功能。MPEG-2 要达到的最基本目标是:位速率为 4~9 Mbps,最高达 15 Mbps。

4. MPEG-4 标准

MPEG-4 从 1994 年开始工作,它是为视听 (audio-visual) 数据的编码和交互播放而开发的算法和工具,是一个数据速率很低的多媒体通信标准。MPEG-4 的目标是要在异构网络环境下能够高度可靠地工作,并且具有很强的交互功能。

为了达到这个目标,MPEG-4 引入了对象基表达的概念,用来表达视听对象。MPEG-4 扩充了编码的数据类型,由自然数据对象扩展到计算机生成的合成数据对象,采用合成对象/自然对象混

合编码 SNHC 算法;在实现交互功能和重用对象中引入了组合、合成和编排等重要概念。MPEG-4 系统构造如图 4-1 所示,接收端的构造部件如图 4-2 所示。

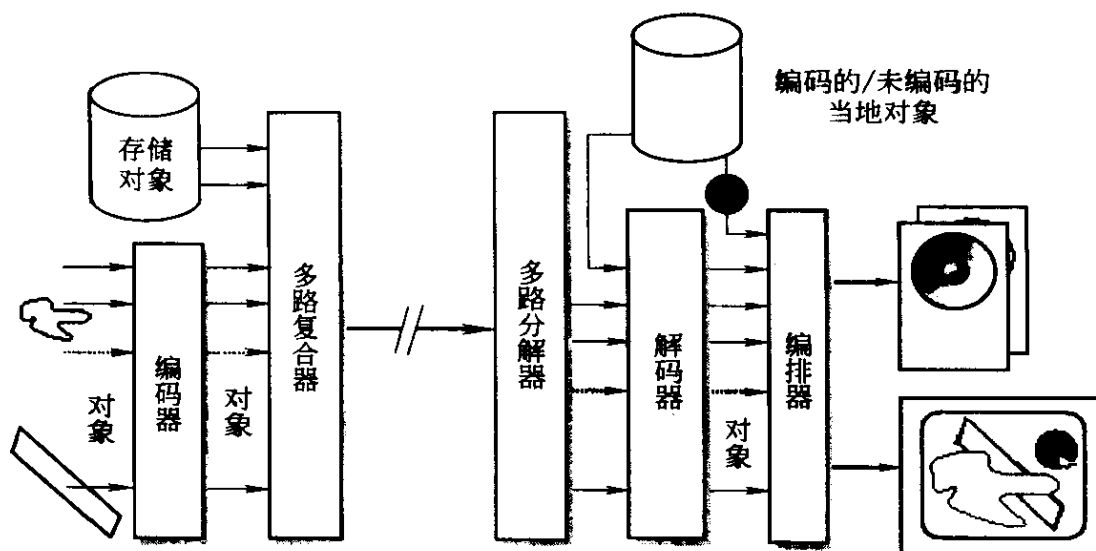


图 4-1 MPEG-4 系统示意图

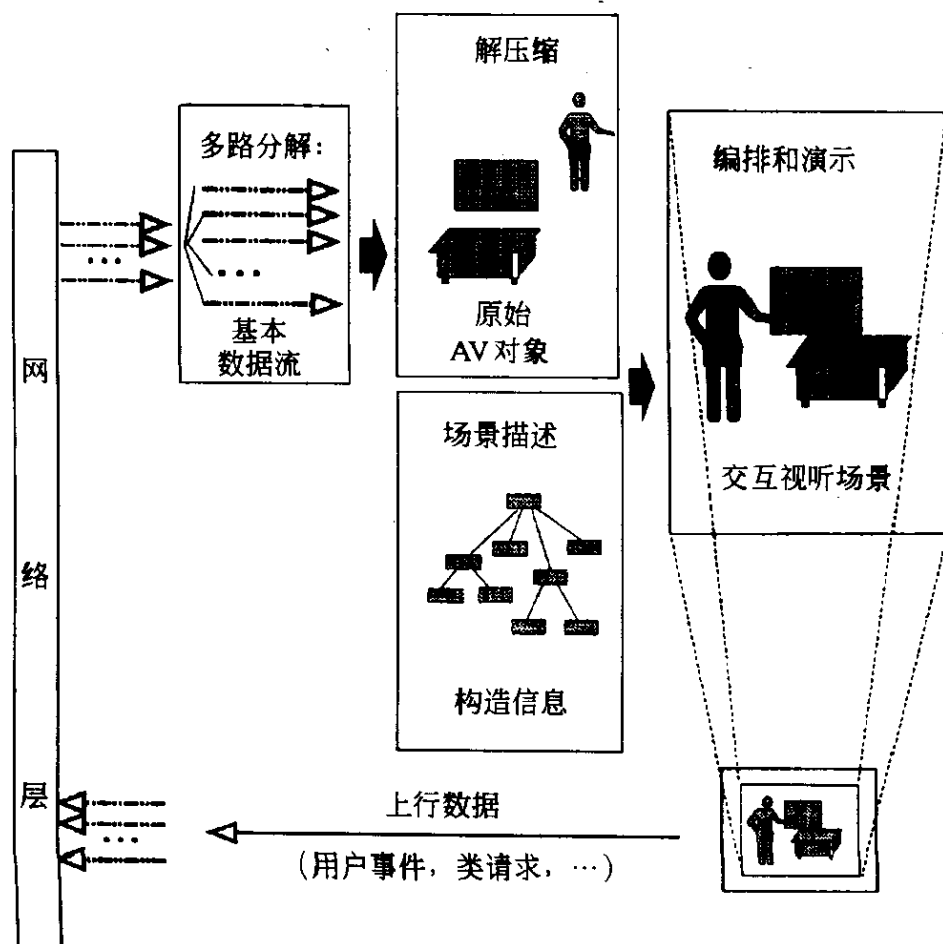


图 4-2 MPEG-4 接收端的主要部件

MPEG-4 中制定了一个称为传输多媒体集成框架 DMIF 的会话协议,它用来管理多媒体数据流。该协议在原则上与文件传输协议 FTP 类似,其差别是:FTP 返回的是数据,而 DMIF 返回的是指向到何处获取数据流的指针。DMIF 覆盖了三种主要技术:广播技术、交互网络技术和光盘技术,如图 4-3 所示。

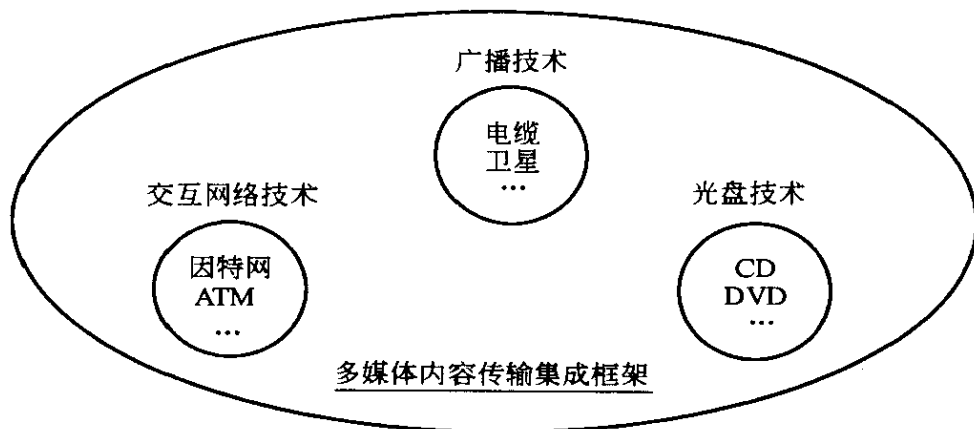


图 4-3 DMIF 覆盖的三种主要技术

MPEG-4 将应用在移动通信和公用电话交换网 PSTN 上,并支持可视电话、电视邮件、电子报纸和其他低数据传输速率场合下的应用。

第5章

视频会议系统结构

视频会议系统是集通信技术、计算机技术、微电子技术于一体的远程异地通信方式,它将计算机的交互性、通信的分布性、以及电视的真实性有机地结合在一起,具有明显的优越性,因而成为当今计算机领域的一个研究热点。本章首先从整体上介绍视频会议的系统结构,使读者对其有个清晰的结构概念;然后再介绍内涵于系统结构中的终端系统结构和组网结构;最后介绍有关视频会议系统的国际标准分布结构。

5.1 系统结构

本节从两个角度来介绍视频会议系统的整体系统结构,即物理系统结构和逻辑系统结构。事实上,这两种结构密不可分,前者侧重于硬件结构,后者侧重于软件结构,在视频会议系统中表现为一个有机的统一整体。

5.1.1 物理系统结构

视频会议系统的物理系统结构如图 5-1 所示。

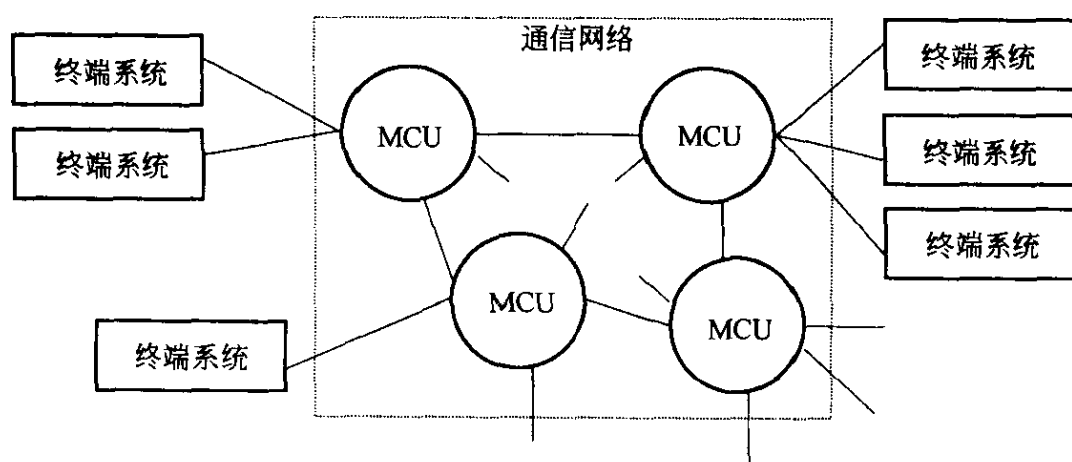


图 5-1 视频会议系统的物理结构

其中,有以下几点值得注意:

(1) 通信网络可以是 PSTN, LAN, ISDN, Internet, FDDI/ATM 等,但由于不同的通信网络原理及结构差异很大,导致了视频会议系统微观部署结构(包括终端系统的连接结构,MCU 的配置方案结构等)的差异性。

(2) MCU(Multipoint Control Unit, 多点控制设备)并不是必需的。只有在多个会议场点(Conference Site)进行多点视频会议时,才设置一台或多台 MCU,以进行图像与语音的分配和切换,这是由于视频与音频是连续传递的数据流,多个信道之间不能直

接并联连接,否则来自不同会议场点的视频和音频信号将重叠在一起。MCU 通常设置在网络节点处。

(3) 终端系统的配置在同一视频会议系统中并不要求完全一致;对终端系统(包括软件和硬件)的要求比较统一,只需符合一定的国际标准即可。

5.1.2 逻辑系统结构

从功能上看,完整的视频会议系统应具有会议管理、协作处理、视频/音频处理、多点控制、通信服务等功能模块,其逻辑系统结构如图 5-2 所示。

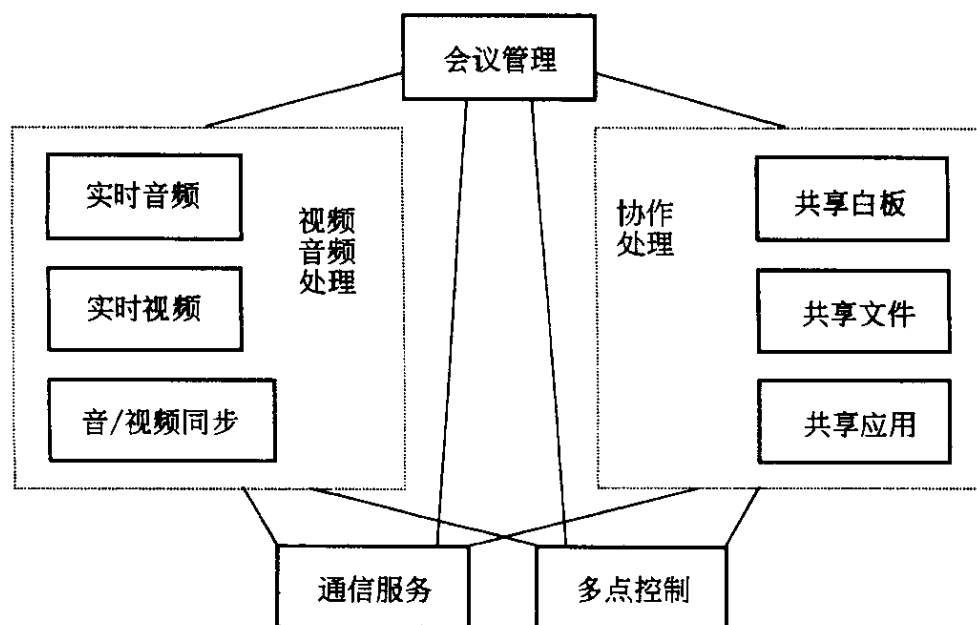


图 5-2 视频会议逻辑结构

(1) 会议管理

完成会议通知,召集任务;初始化会议环境;在会议进行中,协调系统各部分,管理与会者的身份与权力;进行系统各项性能参数的设置和调整。

(2) 协作处理

主要提供共享白板、共享文件、共享应用等形式的协作方式。

其中,共享白板的作用是实现在会人员的公共显示和修改窗口,实时传送修改信息;内含文件等数据的传送功能,完成文件传阅任务;具有 OLE 功能。

(3) 视频/音频处理程序

完成视频/音频信息的采集、转化,实时压缩本地媒体产生的数据,实时解压缩和播放远地媒体产生的并经过网络传送过来的数据。

(4) 通信服务和多点控制

具有网络管理的功能,能集中处理各种媒体产生的信息流的调度、传输等一系列问题,实现点对点、组广播、广播方式等通信方式;完成相应进程的数据的连接;保证网络传输的效率,以维护一定的系统性能。

视频会议系统应具有的特点是:其谈话和活动图像必须是自然的,用户之间必须共享信息,必须要具有友好的用户接口。

5.2 终端结构

终端系统实际上代表视频会议中的本地会议场点,其结构如图 5-3 所示,主要包括以下几部分。

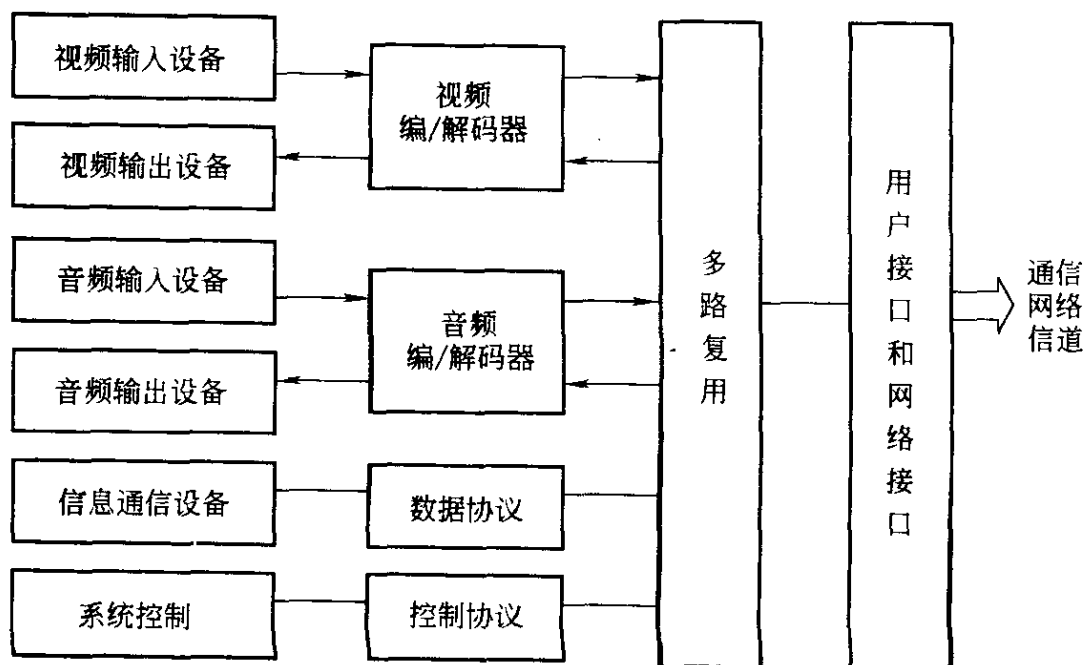


图 5-3 视频会议系统终端结构

(1) 视频输入/输出设备

① 视频输入设备

视频输入设备采集的各种视频信号都须经过视频输入口,将视频信号(模拟或数字)送入视频编码器内进行处理(变换、压缩等)。

视频输入设备包括摄像机及录像机。摄像机主要有主摄像机、辅助摄像机和图文摄像机。

其中,主摄像机主要用来摄取发言人的特写镜头,但主摄像机可被控制转动,这种控制可以是手动的,也可以是自动的。手动方式主要由参加会议人员通过控制器,控制摄像机上、下、左、右转动以及焦距的调节,也可以控制对方会场的主摄像机的转动。自动方式涉及到摄像头方向性的技术问题,可以加入声学扫描技术,用数字信号处理方法将发言者的声音叠加起来,形成一个很窄的声波,再自动调整摄像头对准本会议场点发言者,如果没有太大干扰,反应速度是比较满意的。

辅助摄像机主要用来摄取会场全景图像,或不同角度的部分场面镜头,或摄取白板上的内容。辅助摄像机主要由人工操作控制,属于手动方式。

图文摄像机一般固定在某一特定位置,用来摄取文件、图表等,其焦距已事先调整好。

录像机可播放事先录制好的活动和静止图像。

小型视频会议系统一般只采用主摄像机。

主摄像机镜头一般采用 25 万像素或 30 万像素的分辨率,体积较小,易于安装。辅助摄像机可以选择广角镜头、特殊监控镜头等。

② 视频输出设备

视频输出设备包括监视器、投影机、电视墙、多画面、视频处理器等。监视器用于显示接收的图像;会议人数较多时,可采用投影机或电视墙。为了在监视器上既可以显示其他会议场点的图像,同时又可以显示本会议场点的画面,一般采用多窗口系统,每个会

议场点的情况在屏幕上只表现为一个窗口,并且可以允许这种窗口随意放大缩小,而且不失真。

(2) 音频输入/输出设备

音频输入/输出设备主要包括麦克风(话筒)、扬声器、调音设备以及提供语言激励、多麦克风混合、回声抑制器等附加的语音设备。具体会议对音频设备的配备情况主要由会议对音频质量的要求决定。

话筒和扬声器主要用于与会者的发言和收听其他会议场点的发言。

调音设备主要用于调节本会议场点话筒的音色和音量。

(3) 视频编码解码器

视频编码解码器(Video Codec)是视频会议系统的核心,主要功能有以下几点:

① 将来自本地会议场点视频输入设备的模拟视频信号数字化后进行压缩编码处理,以适应窄带数字信道的传送。目前,已有许多数字式摄像头产品面世,这种摄像头输出的信号已经被数字化,视频编码解码器对此数字信号流直接进行压缩编码处理。

② 将来自远地会议场点的已压缩视频信号解压缩后,送给相应的视频输出设备。

③ 可对不同电视制式的视频信号进行处理,以使不同电视制式的视频会议系统直接无缝互通,如 PAL 与 NTSC 间的互通。

④ 在多点视频会议通信的环境下,视频编码解码器应支持 MCU 进行多点切换控制。

(4) 音频编码解码器

在视频会议系统中,音频编码解码器(Audio Codec)与视频编码解码器具有同等的核心地位,但由于音频数据量与视频数据量相比要小得多,因此,音频编码解码器在视频会议系统设计中并不会成为瓶颈问题。

音频编码解码器功能主要包括两个方面:

① 对来自本地会议场点音频输入设备的模拟信号数字化,以 PCM、ADPCM 或 LDCELP 方式进行编码。这类模拟信号频率通常为 50 Hz~3.4 kHz 或 50 Hz~7 kHz。编码后的数字音频信号的速率可为 16, 48, 56, 64 bps 四种。

② 对来自远地会议场点已压缩的音频信号解压缩后,送到相应的音频输出设备。

(5) 时延

由于视频编码解码器会引入一定的时延,造成发言人的语言与唇部的动作不协调,其口形动作与语音相比有一个延迟,因此在音频编码器中必须对编码的音频信号增加适当的时延,以便使解码器中的视频信号和音频信号同步,即所谓的唇同步问题。

(6) 信息通信设备

信息通信设备是视觉的辅助设备,可增强视频通信能力。

信息通信设备包括白板、书写电话、传真机等。白板供本会议场点与会人员与对方会场人员进行讨论问题时写字画图用,通过辅助摄像机的摄取而输入编码器,传送到对端,在对方会场的监视器上显示。书写电话为书本大小的电子写字板,供与会人员将要说的话写在此板上,变换成电信号后输入到视频编码解码器,再传送到对方会场,并显示在监视器上。

(7) 数据协议

数据协议是所有会议场点之间进行各种数据通信的基础,它必须支持电子白板、静止图像传输、文件交换及数据库存取等应用类型。

(8) 控制协议和系统控制

控制协议提供各终端系统正确运行端到端信令,在系统之间进行能力交换、发送命令和指示信号,以及提供打开和描述逻辑信道的信息。

系统控制是利用控制协议的控制信令对系统进行控制。视频会议系统各终端系统之间的互通一般是依据一定的步骤和规程通过系统的控制来实现的。每进行一项步骤都由相关的信令信号完成。

(9) 多路复用和解复用设备

该设备可将视频、音频、数据、信令等各种多媒体数字信号组合为 64~1 920 kbps 的数字码流,成为与用户/网络接口兼容的信号格式。同时,也可把接收到来自远地会议场点的比特流分解为各种多媒体信号。此外,其中包含的复用协议还具有能对图像序列进行编号、进行误差检测以及采用重传输的方式实现误差校正等功能。

(10) 用户/网络接口

用户/网络接口是用户端的终端系统与通信网络信道的连接点,该连接点称为接口。该接口主要完成通信网络与多路复用和解复用模块的匹配问题。

其中有几点值得注意:

① 终端系统结构中各模块并不是独立存在的,在实际设计时可能会将若干模块集成或镶嵌在一起协调工作,如时延电路模块就内嵌在音频编码解码器电路中。

② 各模块并不是都由硬件实现。事实上,目前上市的产品中,除了视频、音频输入输出设备和用户/网络接口模块外,其余模块均可由软件实现。

5.3 组网结构

视频会议系统的组网结构随与会者参加方式的不同有所不同,从整体上看,有两种组网结构:点对点组网结构和多点会议组网结构。

5.3.1 点对点组网结构

点对点视频会议系统只涉及到两个会议终端系统,其组网结构非常简单,不需要 MCU,也不需要增加额外的网络设备,只须

在终端系统的系统控制模块中增加会议管理功能即可实现。其组网结构如图 5-4 所示,图中控制协议虚线实际上并不存在,其内容也是通过接口相互传递的。

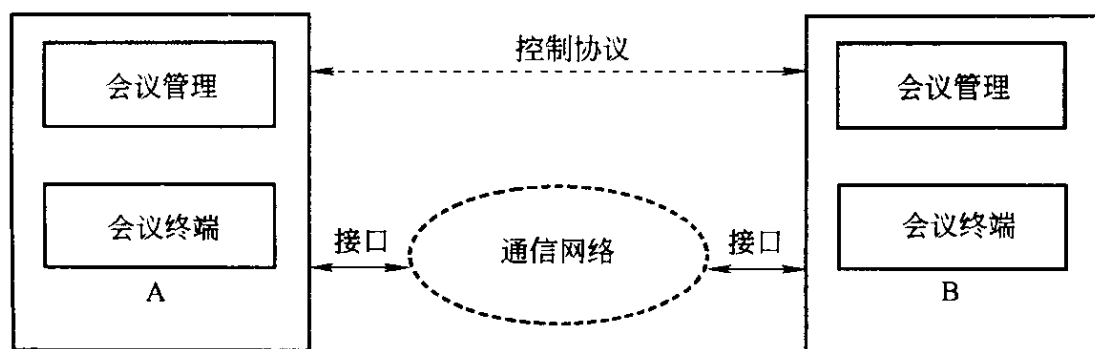


图 5-4 点对点组网结构

两个会议场点(终端系统)只须相互拨号呼叫对方并得到对方确认后便可召开视频会议。目前比较流行的可视电话的通信网络是 PSTN,实际上这是点对点结构的一种特例。

5.3.2 多点会议组网结构

在多个会议场点进行多点会议时,必须设置一台或多台 MCU(多点控制设备)。MCU 是一个数字处理单元,通常设置在网络节点处,可供多个会议场点同时进行相互间的通信。MCU 应在数字域中实现音频、视频、数据信令等数字信号的混合和切换(分配),但不得影响音频、视频等信号的质量。

多点会议组网结构比较复杂,根据 MCU 数目可分为两类:单 MCU 方式和多 MCU 方式。而多 MCU 方式一般又可分为两种:星型组网结构和层级组网结构。

(1) 单 MCU 方式

在会议场点数目不多且地域分布比较集中时,可采用单 MCU 方式,其组网结构如图 5-5 所示。图中 $T_A, T_B, \dots, T_F$ 均为视频会议终端系统设备。

各会议场点依次加入会议时,必须经过 MCU 确认并通知先

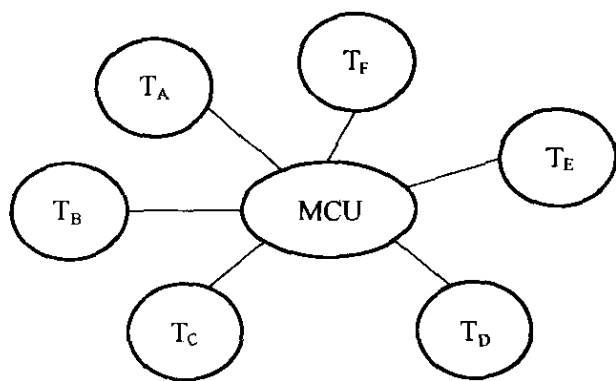


图 5-5 单 MCU 组网结构

于它加入会议的会议场点。

(2) 星型组网结构

多 MCU 连接的星型组网结构如图 5-6 所示,其中 VCT 是视频会议终端 Video Conference Terminal 的缩写。

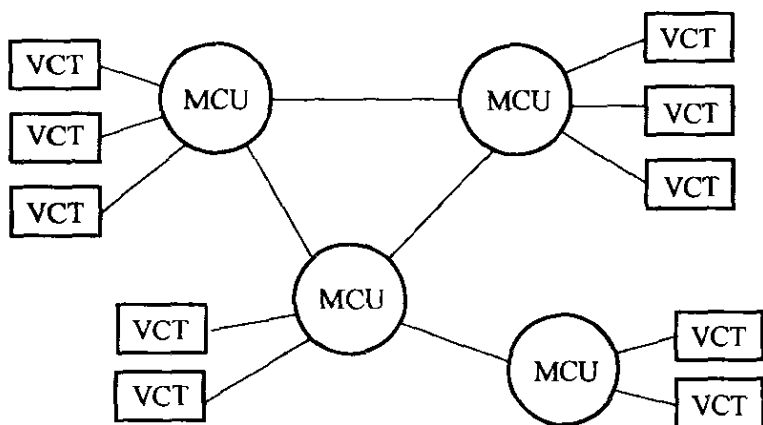


图 5-6 多 MCU 连接的星型组网结构

这种星型结构对会议终端系统要求较低,增加新会议场点时易扩展。MCU 功能类似于交换机,各 MCU 在这种组网结构中地位平等。由于该组网方式的会议场点数目较多,其会议控制模式宜采用主席控制模式。

(3) 层级组网结构

多 MCU 连接的层级组网结构最适宜于布署在各会议场点地域上很分散的情况,可利用 ISDN, B-ISDN 或 DDN(长途数字传输网)等通信网络。其组网结构如图 5-7 所示。

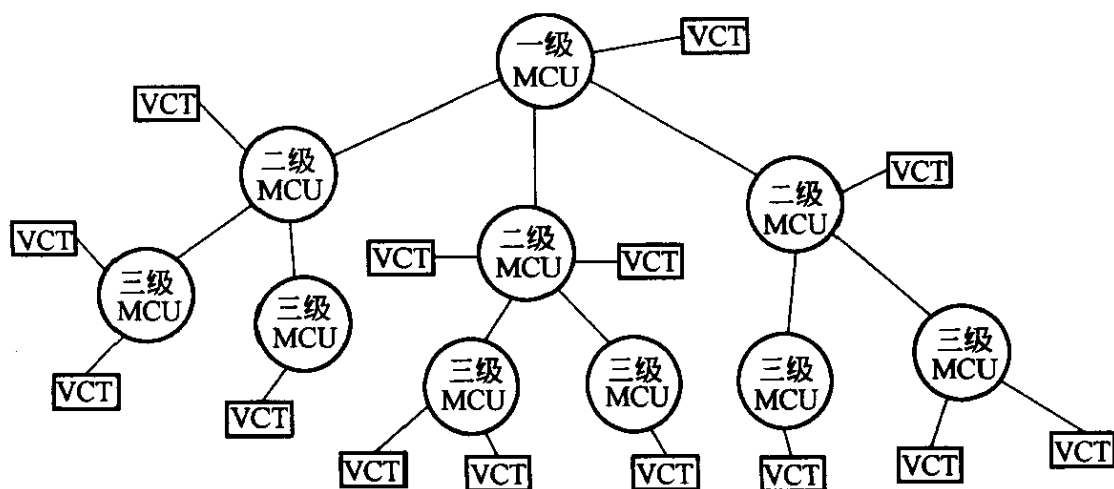


图 5-7 多 MCU 连接的层级组网结构

这种层级结构覆盖的地域很广,也可以进行国际间视频会议,不仅易于扩充,而且更易于管理。多个 MCU 在组网结构中地位是不相同的,有等级之分,下层的 MCU 受上层的 MCU 控制和制约。

这种结构将是未来发展中最理想的组网结构,其最理想的通信网络结构是现存的 Internet,因为在分布式结构与分布式管理机制方面,二者有惊人的相似之处。但令人遗憾的是将 Internet 引入视频会议系统目前仍有许多无法克服的技术问题。

最适于这种结构的会议控制模式是声控模式和主席控制模式。

5.3.3 MCU

MCU(多点控制设备)是多点视频会议系统的关键设备,它的作用相当于一个交换机,但它又与一般电话网中的交换机不一样。前者是对数字信号进行切换,而后者是对模拟信号进行切换。MCU 将来自各会议场点的信息流,经过同步分离后,抽取出音频、视频、数据等信息和信令,再将各会议场点的信息和信令,送入同一种处理模块,完成相应的音频混合或切换、视频混合或切换、数据广播和路由选择、定时和会议控制等过程,最后将各会议场点

所需要的各种信息重新组合起来,送往各相应的终端系统设备。

MCU 还有自动统一传输速率的功能,同一次会议的所有终端系统应该工作在同一速率上,如果与它连接的终端系统速率不一致,它会自动选择所有终端系统的最低速率为工作速率。

MCU 主要处理以下三类数据:

1. 视频信号

主要由视频处理器完成。

MCU 对视频信号一般采用直接分配的方式,若某会议场点有人发言,它的图像信号便会传送到 MCU,MCU 将其切换到与它连接的所有其他会议场点。

如果每个会议场点需要同时观看多个会议场点的图像(多窗口系统或多监视器系统)时,MCU 的视频处理器才对多路视频信号进行混合处理。

2. 音频信号

音频信号主要由音频处理器完成。

如果只有一个会议场点发言,MCU 将其音频信号切换到其他会议场点;若同时有几个会议场点发言,MCU 根据会议控制模式选出一个音频信号,将其切换到其他会议场点。

音频处理器由语音代码转换器和语音混合模块组成。前者从各个端口输入的数据流的帧结构中分离出各种语音信号,并进行译码,然后送入语音混合器进行线性叠加,最后送入编码器,形成合适的编码形式,插入到输出的数据流中。

3. 数据信号

数据信号主要由数据处理器完成。

MCU 采用广播方式将某一会议场点的数据切换到其他会议场点,这主要用于主席控制模式。

此外,MCU 结构中网络接口模块和控制处理器也是必不可少的。

其中,网络接口模块分输入、输出两方向,该模块校正输入数

据流和输出数据流,并按本地系统的时钟定位输入的数据流。在接口模块的输出方向插入所需的各种信令和信息,形成信道帧,以便输出到通信网络数字信道。

控制处理器主要负责决定正确的路由选择,混合或切换音频、视频、数据信号,并对会议进行控制。

5.3.4 多点会议控制模式

在多点视频会议中,与会者既能看到其他会议场点的与会者,又能听到他们的讲话,但这个过程可能并不是同时发生的。那么,在某一时刻到底能看到谁呢?它由多点视频会议的控制模式来决定。

目前,业界已根据视频会议用户的各种不同需要开发出一系列控制模式,其中五种比较重要的控制模式是声控模式、发言人控制模式、主席控制模式、广播/自动扫描模式以及连续模式。

1. 声控模式

声控模式的使用极为普遍,是全自动工作模式,按照“谁发言显示谁”的原则,由声音信号控制图像的自动切换。

多点会议进行过程中,一般只有一方发言,其他会议场点显示发言者的会场图像。当同时有多个会议场点要求发言时,MCU从这些会议场点终端系统送来的数据流中抽取出音频信号,在语言处理器中进行电平比较,选出电平最高的音频信号,即与会者讲话声音最大的那个会议场点,将其图像与声音信号广播到其他的会议场点。

同时为防止由于咳嗽、噪声之类的短促干扰造成误切换,双方同时发言造成图像信息的重叠输出等问题,设置声音判决延迟电路,声音持续 1~3s 后,方能显示发言者的图像。无发言者时,输出主会场全景或其他图像。此外,在有人发言时应将该系统锁定,这样,由背景噪声等引起的干扰就不会将画面切换到其他场所,从而保证视频会议画面的稳定性。

声控模式对项目组会议是十分理想的,与会代表可以自由发言。

该控制模式仅适于参加会议的会场数目不多的情况,一般控制在十几个会议场点之内。因为如果要比较的声音信号数目愈多,则背景噪声愈大,MCU 的语言处理器将很难选出最高电平的语言信号。

2. 发言人控制模式

该控制模式一般与声控模式混合使用,与声控模式一样是全自动工作模式,也仅适用于参加会议的会场数目不多的情况。

当召开一次多点会议时,要发言的人通过编码译码器向 MCU 请求发言。此时如按桌上的按钮,或触摸控制盘上相应键按钮,编码译码器便给 MCU 一个请求信号,如 MCU 认可便将它的图像、语音信号播放到所有与 MCU 相连接的会议终端,同时 MCU 给发言人会场终端一个已“播放”的指示,使发言者知道它的图像、语音信号已被其他会议场点收到。当发言者讲话完毕时,MCU 自动切换恢复到声控模式。

3. 主席控制模式

该控制模式将所有会议场点分为主会场(只有一个)和分会场两类,由主会场组织者(或称主席)行使会议的控制权,它根据会议进行情况和各分会场发言情况,决定在某个时刻人们会看到哪个会场,而不必考虑此刻是谁在发言。

主席可点名某分会场发言,并与之对话,其他分会场收听它们的发言,收看发言人图像。分会场发言需向主席申请,但须经主席认可后发言有效,此时申请发言的会场图像才被传送到其他各分会场。

这种控制模式具有很大的主动性,控制效果比较好,避免了声控模式中频繁切换图像造成的混乱现象。

当然,主会场和分会场的地位在同一次会议中也可以动态变化。

4. 广播/自动扫描模式

该控制模式实际上是主席控制模式的一个变种。这种模式可以将画面设置为某个会场(这个会场被称为广播机构),而这个会场中的代表则可定时、轮流地看到其他各个分会场。这种模式按照事先设定好的扫描间隔自动地切换广播机构的画面,而不论此刻是谁在发言。

5. 连续模式

连续模式是一种最新发展的控制模式。这种模式通过将屏幕分割成若干窗口,而使与会者可以同时看到多个分会场的情况。

值得注意的是,控制模式是由应用程序所驱动的,当视频会议中出现新的应用需求时,就会产生新的控制模式来支持它们。

第 6 章

视频会议系统国际标准

总部位于瑞士日内瓦的国际电信联盟 (ITU: International Telecommunications Union), 主要帮助制定通信标准, 以方便世界范围内通信的相互操作。ITU 的成员代表了不同的国家和公司。在 20 世纪 80 年代, ITU 专门成立了一个小组研究视频会议, 从那时起建立了一系列的体系和标准。ITU 是联合国的一个机构, 遵从 ITU 标准的产品使得任何呼叫方都能同其他呼叫方建立联系, 并且, 不管产品的厂商是谁, 均可具有透明的互用能力。

同样位于瑞士日内瓦的国际电报电话咨询委员会 CCITT, 长时间以来致力于制定国际通信标准, 它的成员包括国家、公司及科学组织。两年前, 它改成了国际电信联盟-通信标准部 (ITU-TSS: Telecommunication Standardization Sector 或 ITU-T)。

由 ITU-T 制定的标准正是为视频会议系统而制定的, 实质同该小组为数据调制解调器和传真机所制定的“V.dot”和“Group-III”标准一样, 这些标准能够通过使遵从标准的设备可以加入到世界上所有的基础设备 (包括通信网络) 之中以保证其兼容性。

ITU 制定的有关视频会议系统标准即 H.32X 标准如表 6-1 所示。

表 6-1 ITU 视频会议系统标准 H.32X 一览表

视频会议系统 标准名称	标准功能和应用范围
H.320	主要应用于 ISDN 及交换线路上(包括 15 个子标准)
H.321/ H.310	定义了应用于 B-ISDN 通道(ATM)上的宽带视频会议终端系统
H.322	定义了 H.320 终端系统如何适用于 B-ISDN 通道
H.323	定义了 LAN 上的视频会议终端系统
H.324	定义了 PSTN 上的视频会议终端系统

6.1 T.120 数据会议序列标准

T.120 是由国际电信联盟通信标准部(ITU-T)开发的系列国际标准,此标准是为支持多点和多媒体会议系统中发送数据而制定的。

T.120 标准也为连接白板和非视频会议应用及文件传输提供应用规程。T.120 系列标准又称为多层协议(MLP: Multi Layer Protocol),既可包含在 H.32X 框架之中,对现有的视频会议进行补充和增强,也可独立地支持声像(Audio Graphic)会议(传送语音、静止图像、白板、加注等信息的实时电信会议)。

T.120 模型规定了一个通信基础结构和在此基础结构上的应用规程。该模型遵循 ISO 开放系统互连(OSI)七层模型,每层向上一层提供业务并向上一层发送由低层提供业务使用的数据。T.120 通信基础结构设计灵活,它能同时处理多个应用,允许与电路交换的通信网和基于分组的局域网以及数据网任意组合进行连接。

T.120 系列主要标准如表 6-2 所示。

表 6-2 T.120 系列标准一览表

标准名称	主 要 内 容	通过时间
T.120	声像会议系列标准概貌,是有关多媒体数据传输草案的文件	1996 年
T.121	常规应用模板,它是声像会议系列标准中所有应用规程和细节方面所涉及到的通用程序要素的说明	1996 年
T.122	用于声像会议和视听会议的多点通信业务(MCS),它确定声像会议和视听会议业务中多点通信的数据传送、令牌管理的机制及原理	1993 年
T.123	用于声像会议和视听会议应用的网路特定传输规程,它确定终端使用各种网路所对应着各种特殊的规程,包括 OSI 模型中多至七层的一系列规程	1993 年

续表

标准名称	主要内容	通过时间
T.124	用于声像和视听终端、多点控制设备的通用会议控制(GCC),GCC功能包括会议的建立、保持与退出、管理会议登记、管理应用登记、应用登记服务、会议指挥等	1996年
T.125	用于声像会议和视听会议的多点通信服务规程的详述,是T.122的补充	1994年
T.126 (T.ST)	多点静止图像和注释规程利用由MCS,GCC提供的原语实现静止图像的交换,并通过扩展能力实现注释交换规程。其中包括:数据单元的结构和它们的交换协议、一组控制资源管理的规则,能力协商等等	1995年
T.127 (T.MBFT)	多点二值文件传输规程,它包括有关二值文件在多点环境下进行广播、选择性分发与确认、对远端索引簿的访问、压缩档案的转移等应用规程	1995年
T.128 (T.AVC)	音频和视频控制,它确定多点多媒体系统的音频、视频控制模式、辅助设备的控制方式	1996年
T.BWC	带宽协调与控制	
T.RESV	视听和声像会议的预约业务	
T.CFAX	用于视听和声像会议的多点传真	

T.120系列标准的关系结构如图6-1所示。

1. 网路特定传输规程

网路特定传输规程框架T.123规定了不同网络之间的连接,它包括:

- 在物理层使用V系列标准调制解调器的公共交换网;
- 使用X.25规程的分组交换数据网;
- ISDN和电路交换数字网,网上的音频、视频和T.120数据根据H.221与H.320标准内的成帧规程复用。

LAN规程不在ITU-T权限内,但对TCP/IP和NOVELL公司的IPX的参考框架已开发出来,它包括一个增加的传输层能力,因此起到了面向连接的网络作用。

2. 多点通信业务

多点通信业务MCS(Multipoint Communication Service)包括

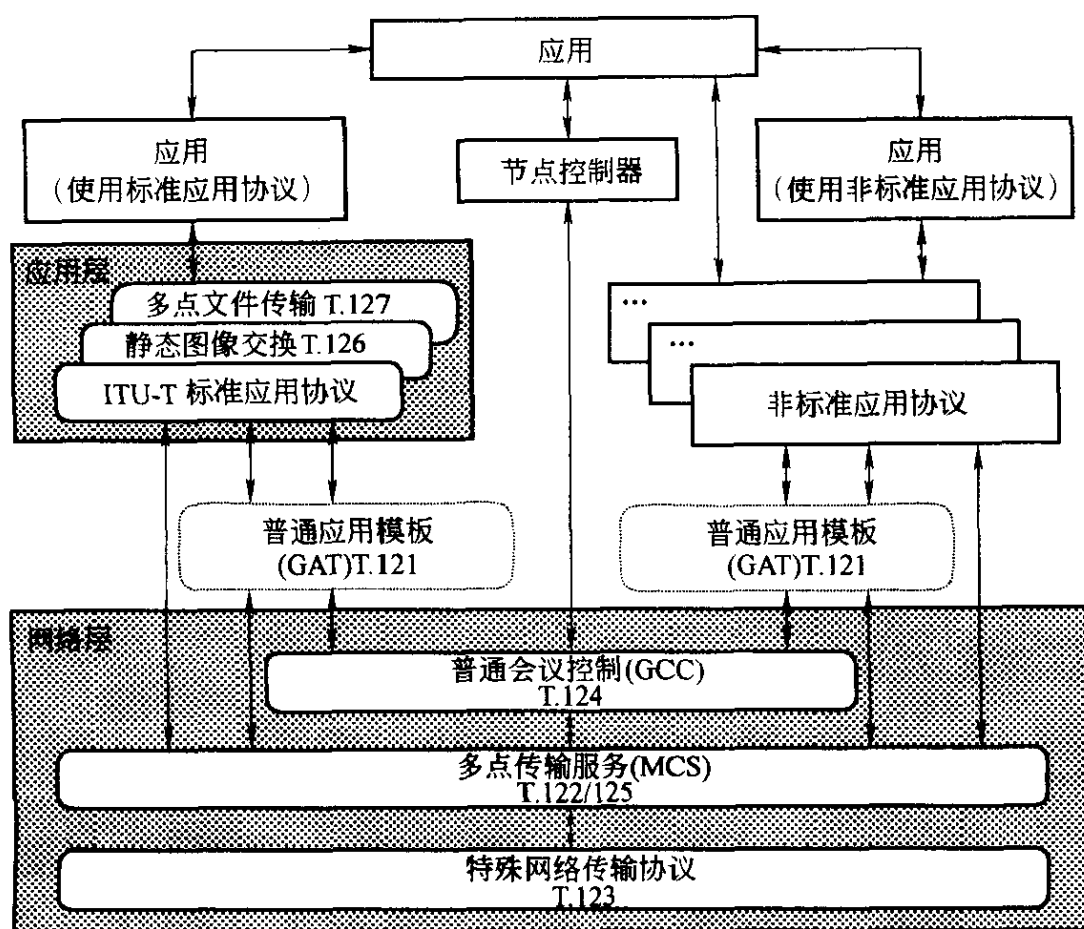


图 6-1 T.120 建议的体系结构图

T.122 业务定义和 T.125 规程规范。MCS 提供面向连接的业务，它与 T.123 传输堆栈操作无关。MCS 在 OSI 模型的会话层，将多个点对点传输连接分到多点通信领域。在 MCS 域内的节点按层次组织，向上连到 MCU，MCU 可以是台式或独立设备中的一个独立节点，如会议桥。

数据要沿最有效的路径送到节点，MCS 能确保从多个源节点来的数据在所有的节点按相同顺序接收，这可以通过在 MCS 内规定适当的优先级和为不同类型的数据确定选路的级别来实现。

MCS 支持的会议种类包括一对一、一对多和多对一连接，它也可以在多种类型网路和连接上操作。

3. 通用会议控制 GCC 和应用规程

通用会议控制 T.124 (GCC: General Conference Control) 可以

通过指定某种业务和管理多点会议来提供会议能力。这种能力包括建立和中止会议、多种应用中的协商能力和提供通用会议管理能力,如参加人员的清单、主持人控制功能等。GCC 也装备了安全机制,如口令编码保护,能限制未被邀请的人参加会议。

T.120 系列标准为应用开发提供规范,使不同厂家的产品和不同的业务可以互通。ITU-T 在 T.120 中规定了两个高层规范:一个是 T.127,它为用户提供同时初始化多点文件传输能力;另一个是 T.126,它允许用户在多点文件会议中查阅图像或对它作注释,共享应用和交换传真图像,共享可以发生在运行不同操作系统中的应用。

6.2 H.320 视频会议标准

H.320 标准是视频会议中最重要的标准,这个标准包括视频、音频的压缩和解压缩,静止图像,多点会议,加密及一些更新的特性。

H.320 标准中包括 H.200 系列标准和 T.120 系列标准。H.200 系列指的是视听业务(Audio Visual Service),具体来说是以传送活动图像为主的通信业务。T.120 系列主要针对声像业务(Audio Graphic Service),即传送静止图像的通信业务。

H.320 系列标准可以分成六个部分,其互操作示意图见图 6-2。

- 通用体系;
- 音频;
- 多点会议;
- 加密;
- 数据/远端摄像机控制 FECC;
- T.120 数据会议。

H.320 各部分标准主要内容由表 6-3 给出。

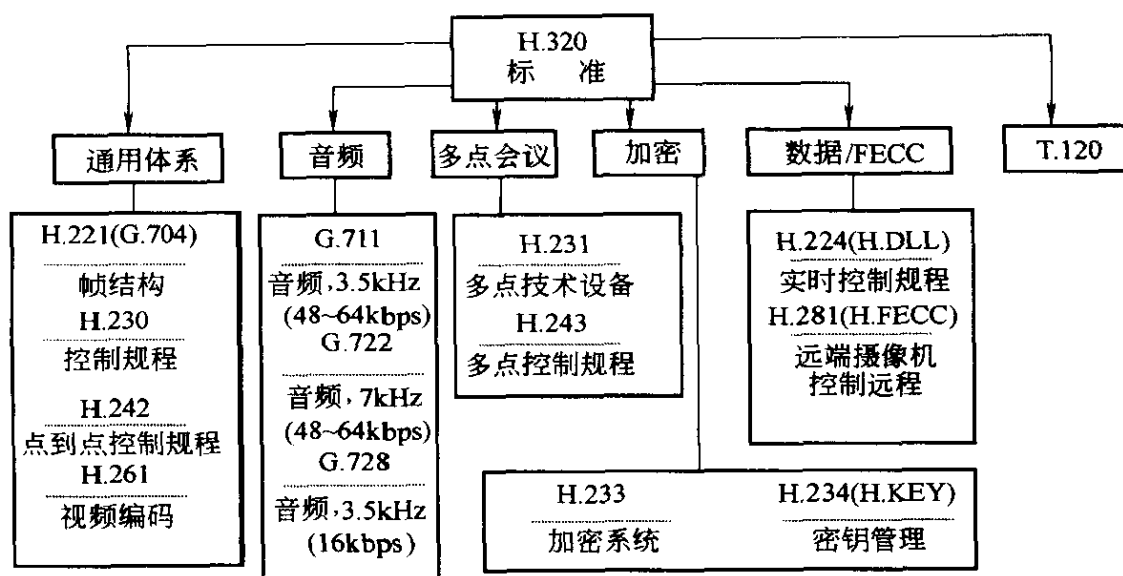


图 6-2 H.320 互操作性示意图

表 6-3 H.320 标准框架一览表

组成部分	标准名称	主要内容
通用体系	H.221	视频会议中 64~1 920 kbps 信道的帧结构
	G.704	有关 PCM30 路帧结构内容
	H.230	视频会议的帧同步控制和指示信号 C&I, 负责处理基于 H.320 的 CODEC 之间传送的控制信息
	H.242	关于建立使用 2 Mbps 以下数字信道的视听终端间建立通信和设置呼叫的通信系统, 定义了基于 H.320 设备之间传送压缩视频及音频信号的协议, 实际上是端到端之间的互通规程
	H.261	关于 $P \times 64$ kbps 视听业务的视频编码解码器
音频	G.711	一种脉冲编码调制的音频编码方式, 数据速率为 48~64 kbps
	G.722	一种自适应差分脉冲编码的音频编码方式, 数据速率为 48~64 kbps
	G.728	一种低时延码本激励线性预测音频编码方式, 数据速率为 16 kbps

续表

组成部分	标准名称	主要内容
多点会议	H.231	用于 2 Mbps 以下数字信道的视频会议系统中的 MCU, 定义了什么是 MCU, 以及它如何连接 3 个或多个基于 H.320 的 CODEC 以组成一个多点会议
	H.243	利用 2 Mbps 信道在 2~3 个以上的视听终端间建立通信的方法, 实际上为多个终端与 MCU 之间的通信规程, 是 H.231 的一个补充
加密	H.233	提供了基于 H.320 设备线性加密的 3 种不同方法的选择, 数据加密标准 (DES) 主要用于美国, FEAL 主要用于日本, BCRYPT 主要用于英国
	H.234 (H.KEY)	即 H.KEY, 是密钥管理标准, 确定了如何在不同点之间传送密钥及一些其他的管理问题
数据 /FECC	H.224 (H.DLL)	全称为“使用 H.221LSD/HSD/MLP 信道的单工应用的实时控制协议”, 提供了诸如使用 H.221 和 H.243 中定义的 LSD, HSD, MLP 信道的单工, 低时延应用, 它为远端摄像机控制提供了一个简单而灵活的协议
数据 /FECC	H.281 (H.FECC)	为采用数据链路协议的视频会议远端摄像控制规程, 它涉及到支持远端摄像机控制协议标准的过程元素和字段的格式, FECC 协议在 H.224 建议规定的的数据链路协议之上, 利用 H.224 才能实现
T.120	—	用于声像和视听会议的一系列标准
	G.703	脉冲编码调制通信系统网路中的数字监控参数
其他	G.735	工作在 2 Mbps 并提供同步的 384 kbps 数字接入和/或同步的 64 kbps 数字接入集群复用设备的特性

为了使读者对 H.320 框架协议有个更清晰的层次结构, 图 6-3 给出, H.320 层次结构框架图。

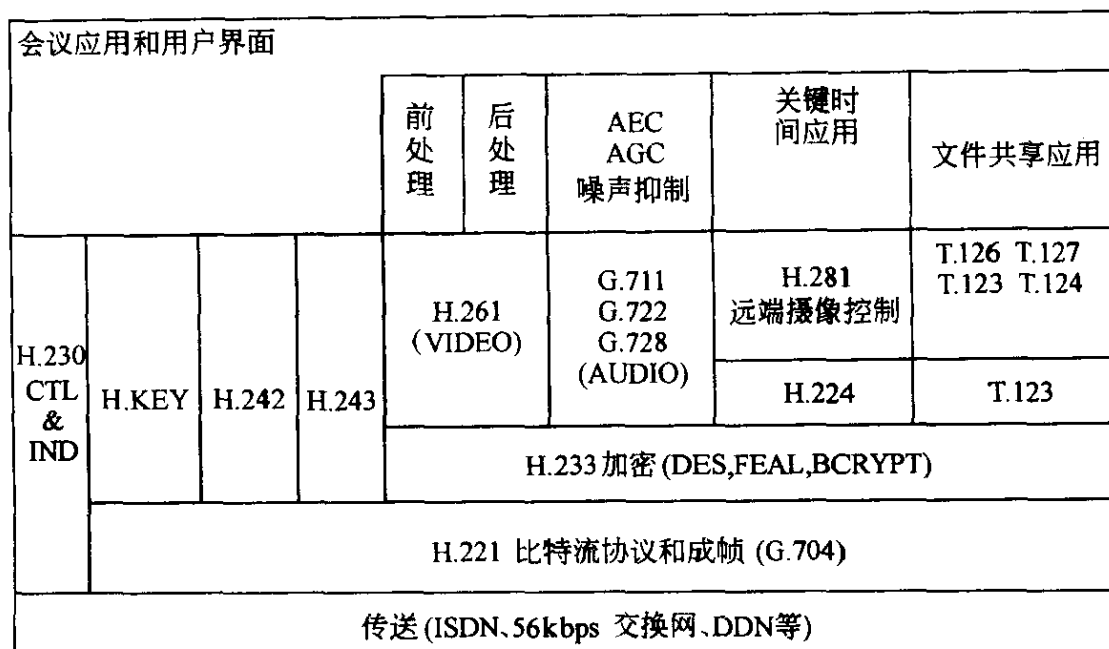


图 6-3 H.320 层次结构框架图

6.3 H.323 视频会议标准

6.3.1 H.323 标准介绍

H.323 标准涵盖了音频、视频及数据在以 IP 包为基础的网络——LAN, Intranet, Extranet 和 Internet 上的通信, 建立 H.323 标准是为了允许不同厂商的多媒体产品和应用能够互操作。

1996 年, ITU 批准了 H.323 规范。该标准范围广泛, 涵盖了各种独立设备、个人计算机技术以及点对点和多点视频会议技术。

该标准解决了点对点及多点视频会议中诸如呼叫与会话控制、多媒体与带宽管理等许多问题。

H.323 只是 ITU 制定的使视频会议能够在各种网络上进行的一系列通信标准的一部分, 这一标准系列称之为 H.32X, 包括 H.320 和 H.324, 分别解决在 ISDN 和 POTS(普通电话网)上的通信问题。

H.323 建议是由 ITU 制定的用于分组交换网的多媒体会议协议体系。与 H.320, H.324 等用于固定带宽信道的建议不同, H.323 建议

建立在 RTP 基础之上,它所定义的多媒体会议系统由终端、守门人(gatekeeper)、网关(gateway)、多点控制器(MC)、多点处理器(MP)和多点控制单元(MCU)等组成。下面分别加以介绍。

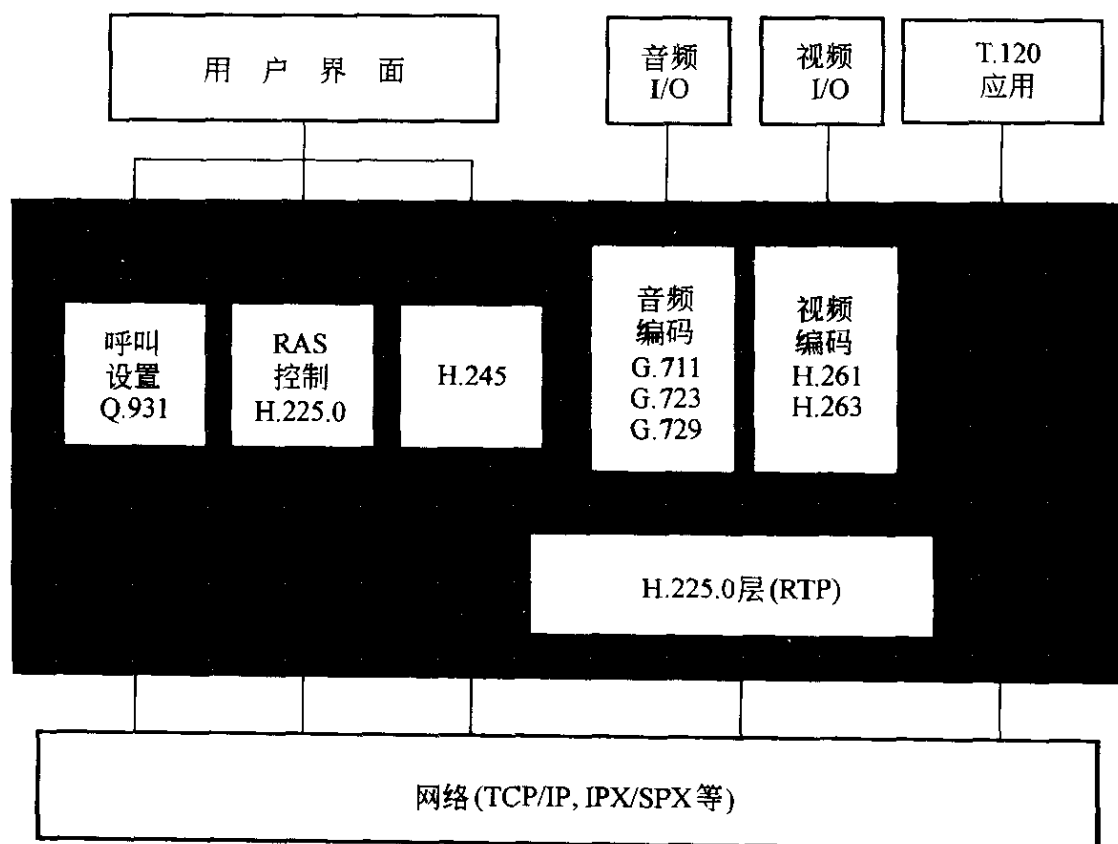


图 6-4 H.323 终端的协议层次图

图 6-4 为 H.323 终端的协议层次图,其中涉及到四类数据,即音频、视频、数据和控制信号。对于音频和视频流,H.323 建议规定每个终端至少必须支持 G.711 音频编解码和 H.261QCIF 格式的视频编解码,推荐使用具有更高效率的编解码建议(如 G.723.1,G.729,H.263 等),但同时必须支持 G.711 和 H.261 格式。数据信号是指符合 T.120 系列建议的数据流信号(包括文件传输、白板、应用程序共享等)。T.120 建议包括 T.121(通用应用模板),T.122/T.125(多点通信服务),T.123(网络传输协议),T.124(通用会议控制),T.126(静态图像交换),T.127(多点文件传输),T.128(应用程序共享)等子建议,T.120 建议也可以独立应用以召开单纯的数据会议。控制信号则主要是指 H.245 控制

信令、RAS 信号、呼叫信号和 RTCP 信号。其中 RAS 信号是指终端与守门人之间进行登记(Registration)、接收(Admission)、状态(Status)改变等操作而交换的控制信号。这四类数据最终采用 H.225.0 建议格式进行打包,成为统一的 RTP 数据包。

在 H.323 会议系统中,守门人是一个可选的角色,它的主要职责有:

- (1) 地址翻译:将用户别名翻译成网络地址;
- (2) 用户进入会场许可的控制与管理;
- (3) 用户带宽控制与管理;
- (4) 代理呼叫或呼叫转移。

守门人是一个独立的逻辑设备,但可以与终端、MC、MP、MCU、网关等在同一个物理设备上。

多点控制单元(MCU)是由多点控制器(MC)和多点处理器(MP)组成的。其中多点控制器提供支持多点会议的控制能力,协商各会议终端的音视频编解码能力(Capability),确定会议的公共能力;多点处理器则负责对来自各终端的多种媒体流进行处理,主要有对视频画面的切换和复合,对音频信号的混合与抑噪,以及对多媒体数据的多点传输。与 H.320 系统不同,在 H.323 系统中,MC 和 MP 既可以合在一起作为一个单独设备 MCU,也可以分散在其他设备(如终端、守门人、网关等)中。

网关用来把位于不同物理网络或不同子网上的会议终端连接起来组成一个会议。网关的主要任务有:

- (1) 实现不同物理网络(ISDN, PSTN, LAN 等)的多媒体数据流格式(如 H.220.0, H.221)之间的互译;
- (2) 实现音视频及数据编码格式之间的互译;
- (3) 实现不同物理网络的控制信令(如 H.245, H.242)之间的互译。通过网关, H.323 会议终端既可以与不同子网上的 H.323 会议终端进行互连,也可以与不同物理网络上的非 H.323 会议终端一起召开范围广泛的多媒体会议,如图 6-5 所示。

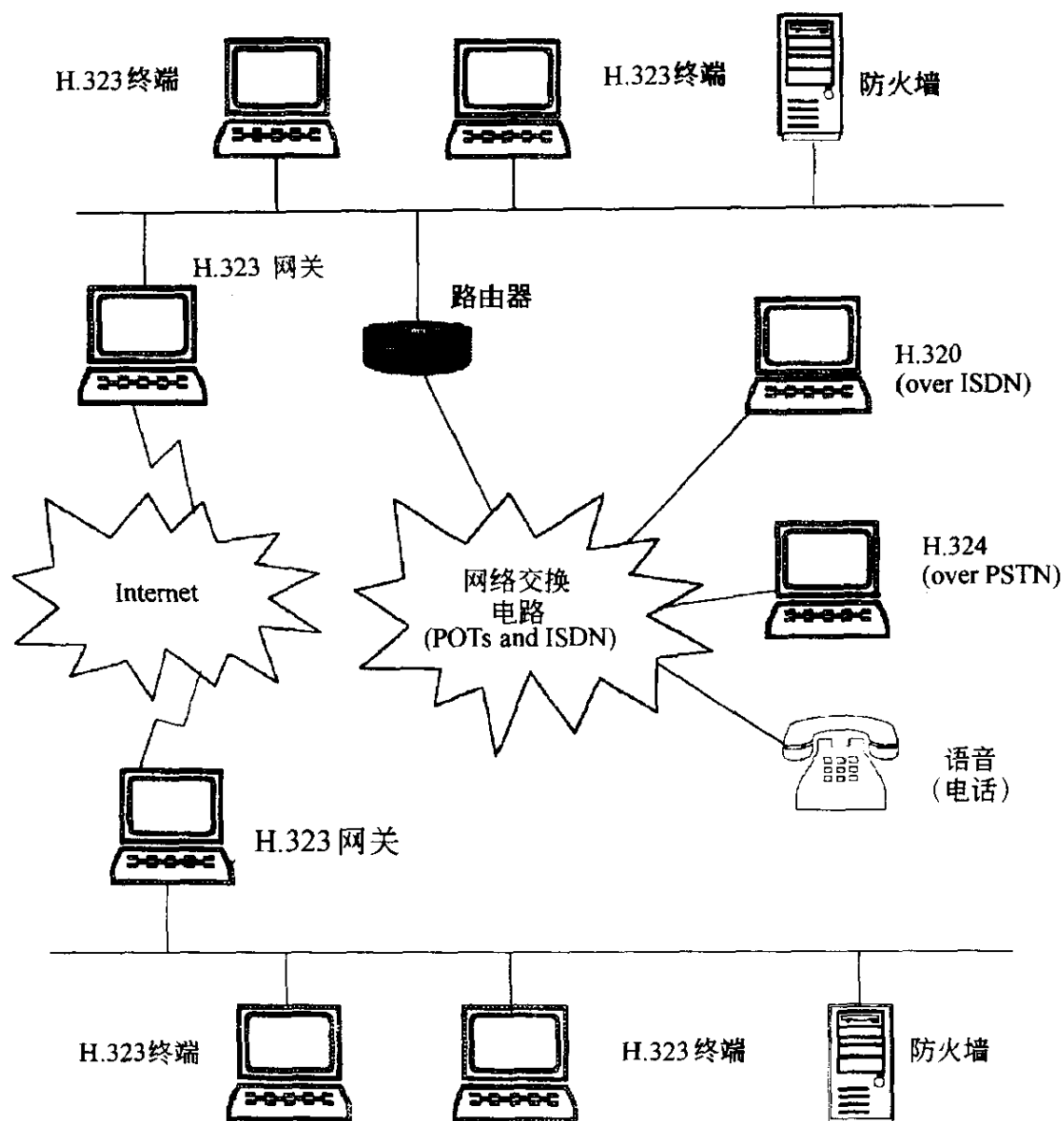


图 6-5 H.323 网络结构图

6.3.2 H.323 的主要优点

1. 网络独立性

H.323 被设计成运行在通用的网络体系结构之上。随着网络技术的发展以及带宽管理技术的进步,基于 H.323 的解决方案将能获得由于这些进步所带来的好处。

2. 带宽管理

视频和音频传输对带宽要求很高,若不加以监控,会阻塞企业的网络。H.323 通过提供带宽管理来解决这一问题,能够对网络中并发的 H.323 连接数以及 H.323 应用可获得的带宽总量进行限制。一方面,这些限制保证了网络上的关键应用不会被中断;另一方面,又为网络上的其他应用保留了足够的资源。每个 H.323 终端都能针对每一会话提供带宽管理,该机制(称之为自动带宽管理)能根据网络的状况(如:延迟、抖动和丢包率)提高或降低视频传输率。

3. 网间会议

许多用户希望与远地进行视频会议。在 H.323 环境中,通过直接的 PPP 连接或企业 Intranet, Extranet 甚至 Internet 进行交流。H.323 也提供一种连接基于包交换的视频会议系统与基于电路交换(H.320, H.324)或基于 ISDN 的视频会议系统的手段。H.323 使用不同视频会议标准中共同的编解码技术来减少或消除协议转换,从而提供优化的性能。

4. 平台及应用的独立性

H.323 不依赖于任何特定的硬件和操作系统, H.323 兼容的平台可以有多种规格和类型,包括支持视频的个人计算机、专用平台以及功能齐全的产品包。

5. 多点支持

H.323 能支持三点或更多点的多点会议,可以有也可以没有集中式的 MCU(多点控制器)的支持。多点控制功能可以是分布式的,可作为 H.323 终端功能的一部分来实现。

6. Multicast(多点广播)支持

当网络支持“工作组管理协议”(如 IGMP)时, H.323 能在多点会议中支持 Multicast 功能。Multicast 将单一信息包送至网上多个目标,而无需重复传送。Multicast 传输能更有效地使用带宽,因为 Multicast 组中所有的工作站(且只有这些工作站)将从单一信息包中读取信息。

7. 互操作性

用户希望在通信时无需担心接收端的兼容性。除了保证接收端能解压信息,H.323 还提供了在客户端之间调整性能,为会议设置共同的性能的方法。标准也建立了公共的呼叫设置和控制协议。

8. 编解码标准

H.323 为音频、视频数据流的压缩和解压建立了标准,保证来自于不同厂家的产品有共同支持的领域。它也为基于厂商特殊硬件和软件的附加特性与增强性能提供了极大的灵活性。

9. 灵活性

一个 H.323 会议能够包括具有不同功能的终端。比如说,只有音频功能的终端也能参加具有音频、视频和/或数据功能的终端所进行的会议。而且,一个 H.323 多媒体终端能够与一个只具有 T.120 数据功能的终端分享会议的数据部分,同时与其他的 H.323 终端分享声音、图像和数据。

6.3.3 H.323 的体系结构

H.323 标准覆盖了在包交换网络上提供音频通信服务的技术需求。H.323 引入了针对数据会议的 T.120 建议,使得包括数据功能的会议成为可能。H.323 不包括网络本身以及用于连接各种网络的网络传输层的内容。

H.323 为基于网络的会议系统定义了四个主要的组成部分:终端、网关(GATEWAY)、守门人(GATEKEEPER)以及多点控制器(MCU)。

1. 终端

终端是提供单向或双向实时通信的客户端。所有终端必须支持声音通信,视频和数据是可选的。H.323 规定了不同的音频、视频和/或数据终端共同工作所需的操作模式。它将成为下一代 In-

ternet 电话、电话会议和电视会议技术领域占统治地位的标准。

所有的 H.323 终端必须支持 H.245, 该标准用来协商通道的使用及容量。另外三个组成部分是: Q.931 的缩略版, 用于呼叫信令和呼叫建立; 注册/许可/状态(RAS), 用于 GATEKEEPER 进行通信; RTP/RTCP, 用于音频和视频包排队。

2. 多点控制器(MCU)

MCU 支持三点或更多点之间的会议, 在 H.323 情况下, MCU 由必不可少的多点控制器(MC)以及零个或多个多点处理器(MP)组成。MC 处理所有终端之间的 H.245 协议, 以确定共同的音频和视频处理能力。MC 也通过确定哪些音频和视频流将进行多点广播(Multicast), 从而对会议资源进行控制。

MC 不直接处理任何多媒体信息流, 这由 MP 来进行处理。MP 混合、交换及处理音频、视频和数据流。MC 和 MP 功能可以存在于一个专门的组件中, 或作为其他 H.323 组件的一部分。

3. 多点会议

在 H.323 情况下, 多点会议功能可以用多种方法和配置进行处理。H.323 标准使用了集中和分散会议的概念, 如图 6-6 和图 6-7 所示。

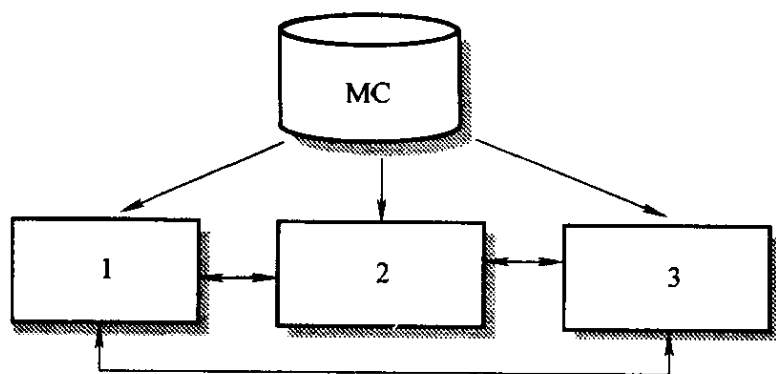


图 6-6 使用 Multicast 的分散会议

集中多点会议需要 MCU 的存在以进行多点会议。所有的终端以点对点的方式向 MCU 发送音频、视频、数据和控制流。MC

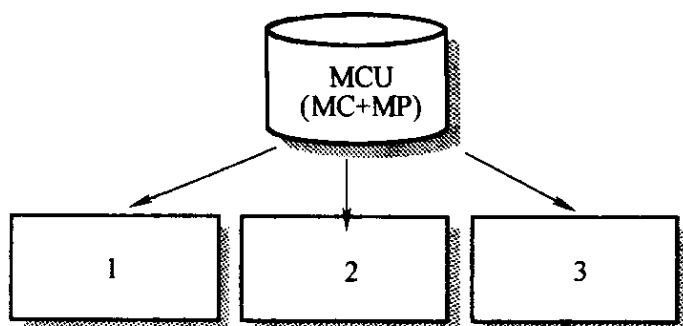


图 6-7 使用 Unicast 的集中会议

通过 H.245 控制功能集中管理会议,该功能也为每个终端指定性能。MP 进行声音混合、数据分发以及视频切换/混合,并将生成的信息流送回参加会议的终端。MP 也提供不同编解码器和数据率之间的转换,并利用 Multicast 功能来发送经过处理的音频、视频和数据流。一个典型的支持集中多点会议的 MCU 由一个 MC 和一个音频、视频和数据 MP 组成。

分散式多点会议利用了 Multicast 技术。参加会议的 H.323 终端将音频和视频信息多点广播(Multicast)给其他参加会议的终端,无需将这些信息送往 MCU。但多点控制数据仍然由 MC 集中处理,而且 H.245 控制通道信息仍以点对点的方式传送给 MC。

接收终端负责处理多个输入的音频和视频流,终端利用 H.245 控制通道信息告知 MC,它能同时处理多个音频和视频流。一个终端的并发处理能力并不限制会议中多点广播的音频和视频流的数量。在分散式多点会议的情况下,MP 也能提供音频、视频选择。

混合式多点会议综合了集中式和分散式多点会议的特点。H.245 决定音频和视频二者择一以点对点的方式送往 MC,剩下的部分(音频或视频)通过多点广播传送至参加会议的 H.323 终端。

H.323 也支持混合的多点会议。在这种情况下,一些终端在集中式会议中,另一些终端在分散式会议中,MCU 作为二者之间的桥梁。

6.3.4 H.323 标准下的通信

H.323 标准下的通信可以看成是音频、视频、数据和控制包的混合体。音频功能、Q.931 呼叫建立、RAS 控制以及 H.245 信令是必须的。所有其他功能,包括视频和数据会议,都是可选的。当编码器支持多种编码算法时,编码器使用的算法由解码器根据 H.245 协议传送过来的信息决定。H.323 终端也能不对称地工作(不同的编码和解码算法),并能发送/接收多个视频和音频通道。

1. 控制

呼叫控制功能是 H.323 终端的核心。整个系统的控制由三个不同的信道提供:H.245 控制信道、Q.931 呼叫信道和 RAS 信道,控制功能包括呼叫建立信号、性能协商、命令和指示信号以及开通并描述逻辑信道内容的信息。所有的音频、视频和控制信息经过控制层处理,输出到网络接口,对于输入信息流,过程正好相反。

H.245 控制信道是传送控制信息的可靠信道,这些控制信息支配 H.323 实体的运作,包括性能协商、开通和关闭逻辑信道、优先选择请求、流控制信息以及通常的命令和指示。性能协商是 ITU 建议中的基本功能之一,H.245 将接收和发送能力分隔开来以及提供 H.323 终端描述这些能力的方法。在任何两个终端之间只有一个 H.245 控制信道。

呼叫信道使用 Q.931 在两个终端之间建立连接。RAS 信令功能在终端与 Gatekeeper 之间执行注册、许可、改变带宽、状态及脱离程序。如果 Gatekeeper 不存在,RAS 不需使用。

2. 音频

音频信号包含经压缩的数字化语音,H.323 支持的压缩算法都是 ITU 的标准。H.323 终端必须支持 G.711 声音压缩标准。对 ITU 其他声音标准的支持是可选的。

不同的 ITU 音频数字化和压缩标准反映了在声音质量、比特率、计算机性能以及信号延迟之间的折衷。G.711 通常以 56 kbps

或 64 kbps 的速率传输声音。由于 G.723 以非常的数据效率工作,在 H.323 应用中是一种非常流行的音频编解码标准,这一标准也消除了 H.323-H.324 网关中对代码转换模块的需要,并且支持内置静音探测,以节省更多的带宽,也支持半双工的音频系统。功能更强的 H.323 终端将实现高质量的 16 kbps 的 G.728 编解码标准,它消除了 H.323-H.320 网关中相应的代码转换模块。由于需要较高的处理性能,该算法仅在基于硬件的编解码器中实现。H.323 会议室系统也支持 G.720 编解码算法,以获得高级的音频质量。

3. 视频

虽然视频功能是可选的,但任何具有视频功能的 H.323 终端必须支持 H.261 编解码算法(支持 H.263 是可选的)。视频信息以不超过性能协商时所选定的速率进行传输。H.261 提供了许多不同的 ITU 标准之间的兼容性尺度,在 $N \times 64$ kbps 的通信信道上使用。

灵活的 H.263 实现可以是任何数据率,不一定是 64 kbps 的倍数。这种类型的编解码器在许多情况下提高了视频质量。例如:有一个 128 kbps 的会话,并使用了 G.728 音频,那么有超过 100 kbps 的带宽可提供视频使用。

H.263 是对 H.261 向后兼容的修改。在低数据传输率的情况下,通过使用运动估算技术、帧预处理,以及优化的霍夫曼编码表,H.263 的画面质量得到了巨大的提高,H.263 定义了 5 种标准的图像格式。H.261 系统和 H.263 系统之间的通信是容易的,因为两者都支持 QCIF。

如果 H.323 终端既支持 H.261,也支持 H.263,建议在低速率会话时使用 H.263,在高速率会话时使用 H.261,实际使用的数据率取决于厂家特定的实现及应用需求。基于硬件的编解码器的优点在于能够支持大图像格式的高比特和帧率: CIF, 4CIF 及 16CIF。

4. ITU 视频会议图像格式

ITU 视频会议图像格式如表 6-4 所示。

表 6-4 ITU 视频会议图像信号

视频会议图像格式	图像大小(像素)	H.261	H.263
Sub-QCIF	128×96	可选的	必须的
QCIF	176×144	必须的	必须的
CIF	352×288	可选的	可选的
4CIF	702×576	N/A	可选的
16CIF	1 408×1 152	N/A	可选的

5. 数据

数据会议是可选的功能。若支持此功能,则进行协同工作,如共享的电子白版、共享应用及文件传输等。

H.323 通过 T.120 规范支持数据会议。作为 ITU 标准, T.120 解决了点对点及多点数据会议的问题。它在应用层、网络层和传输层提供了互操作性。

H.323 系统通过将 T.120 功能结合进终端和多点控制器来支持数据会议。MCU 控制和混合数据会议信息。目前, T.120 只支持树型 Unicast 结构,将会进行修改以支持 Multicast。

6.3.5 H.323 的实现

H.323 使用可靠和不可靠的传输通道。控制信号和数据需要可靠的传输,因为这些信号必须按照它们发送的顺序接收,且不能丢失。而对于音频和视频流来说,丢失一个包对终端用户没有什么影响。音频和视频使用更加高效但可以是不可靠的传输。现有的大多数 H.323 产品使用 IP 来进行信息传输。

可靠的信息传输使用面向连接的数据传输模式。在 IP 协议栈中,这种类型的传输靠 TCP 来完成,可靠传输保证顺序的、无错的、流控制的包传输,使传输延迟和吞吐率降低。H.323 使用可靠的 TCP 端到端服务,建立 H.245 控制信道、T.120 数据信道和呼

叫信道。

在 IP 协议栈,不可靠的服务由用户数据报协议(UDP)提供。非可靠传输是一种只承诺“尽好地”传送的传输模式。UDP 使用最少限度的控制信息。H.323 使用 UDP 进行音频、视频传输和建立 RAS 信道。

实时协议(RTP)是由 Internet Engineering Task Force(IETF)制定,用来处理在 Internet Multicast 主干上进行音频、视频多媒体信息传输。ITU 对它进行了少量的修改以适应 H.323 的需求。

RTP 信息头包含时间标记和顺序号,附加在音频和视频包外面。通过在接收端适当地缓存,时间和顺序信息使应用能消除重复的数据包,重新排序,同步声音、图像和数据及实现连续的回放。

实时控制协议(RTCP)用于控制 RTP 和监控服务质量,传送有关会话参加者的信息,定期向所有会话参加者分发包含质量和标识信息的控制包。

在多点视频会议中,H.323 终端可以使用 Internet Group Management Protocol(IGMP)协议进行音频、视频多点广播。主要的交换机和路由器都支持 IGMP 协议,以便在 Intranet、Extranet 和 Internet 上提供更高的传输效率。

在大型包交换网络(如 Internet 或企业 Intranet)中,为多媒体应用提供并保证足够的带宽是关键。另一个 IEEE 协议,资源保留协议(RSVP),允许终端为一个特定的数据流请求一个指定的带宽、优先权、最大延迟和处理器时间,并接受一个指明请求是否被批准的答复。

随着 H.323 标准开始在市场上扎根,设备制造商和软件开发商正面临着实现丰富的 H.323 标准的挑战。

几百位来自于 VCON, 8×8 , RadVision 和 Video Server 的工程师已加入了具有互操作性的高质量的 H.323 产品的开发,从 H.323 终端、网关、Gatekeeper 到 MCU。

从这些公司将可获得对这些技术的不断维护和修改。它们是标准化组织的积极参加者,并不断跟踪 H.323 规范的变化。

6.4 H.324 可视电话标准

V.34Modem 和音频、视频压缩新技术的出现,同时伴随着计算机硬件设备功能的日益增强,价格却日益低廉而导致其应用领域愈来愈广,使得在低比特率电路交换网上多媒体视频会议成为现实。ITU-T H.324 正是在这种迫切需要的推动下通过的。基于 PSTN 网的低比特率多媒体可视电话国际标准,主要用于 PSTN/GSTN/POTS 网上。该标准能够确保不同的 H.324 终端的互操作,包括基于 PC 的多媒体视频会议系统、廉价的声音/数据 modem、加密电话、WWW 实时视频浏览器、远端安全摄像头和单独可视电话等。

H.324 是一个“工具箱”标准,可灵活选择产品所需要的媒体类型(音频、数据、视频等);并且对支持相同特性的系统,通过规定共同的操作基线模式(Baseline Mode)而确保互操作性。除基线模式外,H.324 还允许其他可选模式(标准的或非标准的)。更重要的是,H.324 在低比特率网上提供了最好的性能(视频和音频质量、延迟等)。

H.324 标准建立在 ITU-T H.320 工业标准的实践之上(1990 年通过的 ISDN 视频会议国际标准)。H.324 可共享 H.320 的基本结构,包括一个复用器(将几种媒体类型复用成一条比特流,H.223)、音频和视频压缩算法(G.723.1 与 H.263)、控制协议(自动执行容量协商和逻辑信道控制,H.245)。H.324 其他部分,如 H.261 视频压缩、H.233/H.234 加密协议、H.224/281 远端摄像机控制,直接来源于 H.320。H.324 标准可通过网关同以下标准设备进行互操作,即 H.320 ISDN 系统、LANs 与 ATM 系统、H.323 和 H.310、ITU-T T.120 系列数据会议标准。

作为第二代标准,H.324 充分改进了音频和视频压缩技术,避免了 H.320 中出现的问题和局限性。为此,H.324 增加了 H.320 所没有的许多新功能,如接收器控制模式优先策略、单个媒体类型

支持多通道的能力、不同通道带宽的动态分配等。

尽管 H.324 作了以上改进,但它的视频质量远不如 H.320 视频会议系统,因为 H.320 ISDN 呼叫比特率为 128 kbps 或更高,而 H.324 呼叫比特率最大为 33.6 kbps。

6.4.1 H.324 系统结构

图 6-8 是 H.324 系统原理的基本结构图。V.34 modem (GSTN/PSTN)、H.223 复用器、H.245 控制协议是基本组成部分;数据、视频和音频流是可选的,在呼叫中动态建立。H.324 终端同独立的远端终端协商一套通用的容量集用于每个传输方向(终端间可允许容量集不同);通过使用 MCU(又称为桥,bridge),H.324 可进行多点操作。

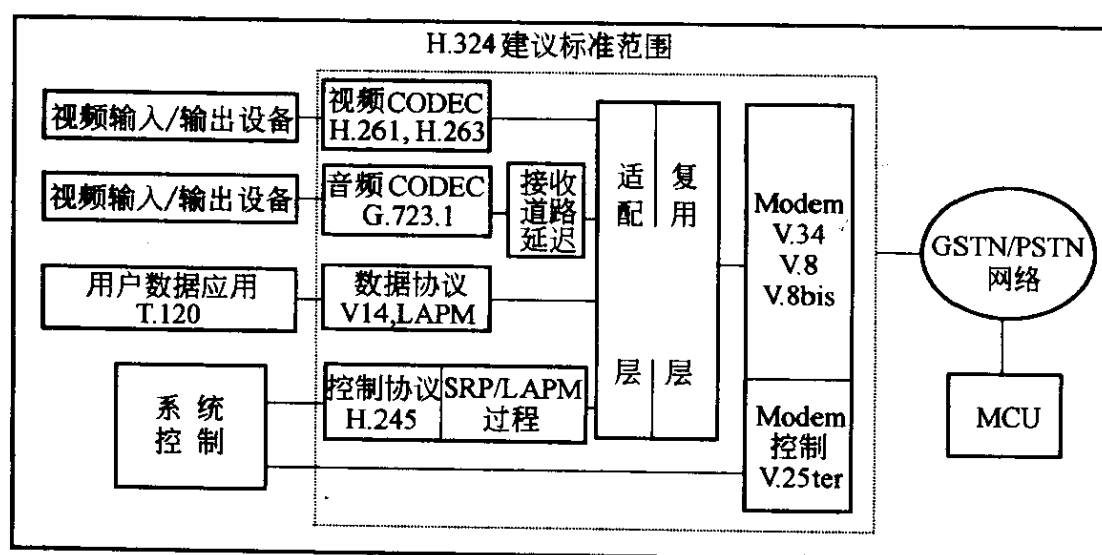


图 6-8 H.324 系统原理框图

6.4.2 H.324 特性

1. 调制解调器

当运行在 PSTN 线路上时,H.324 使用全双工 V.34 modem(如果媒体采用近期推出的规范扩展时,速率可达 28 800 bps 或 33 600 bps);受线路条件所限,modem 也可工作在低速率,如

2 400 bps。H.324 工作的最低速率为 9 600 bps。在呼叫启动时运用必需的 V.8 协议或可选的 V.8 bis 协议可分辨出 modem 类型和操作模式。在大多数情况下,V.8 bis 协议允许一个正常普通电话呼叫在任何时间切换成一个多媒体呼叫。由 ITU-T V.25 ter(“AT”命令集)控制本地 modem 功能(如拨号)。V.34 调制和解调过程比较复杂,一般端到端延迟将增加 30~40 ms。

H.324 系统设计中 V.34 一个重要的特征是错误猝发行为。由于 V.34 格栅编码器和其他因素的影响,比特错误很少单独发生,而是以猝发方式发生。这些错误猝发的速率与 modem 的比特率息息相关,2 400 bps 的一个“单步”(single stop)操作改变比特率不止一个数量级(more than an order of magnitude)。这会影响用于 H.324 modem 的正确调整。既然重传比速率减缓造成的损失消耗更少的带宽,所以用于简单 V.42 数据传输的 modem 通常在“更高速率”时更有效。但是,由于在许多事例中实时需求造成重传出错,需要更低的出错率,H.324 系统应该运行在更低速率上。

H.324 将 V.34 modem 看作是一个同步数据“泵”,用来发送和接收由 H.223 复用器产生的比特流。数据压缩(如 V.42 bis)和重传协议(如 LAPM V.42; Link Access Procedure for Modems 与 MNP: Microcom Network Protocol)不用在 modem 级,尽管这些相同的协议对于某些类型的数据流能用在更高层。

同大多数 modem 支持的 V.42 纠错协议相同,H.324 需要一个与 V.34 数据泵同步接口。既然大多数 PC 只有异步 RS-232 modem 接口,因此基于 PC 的 H.324 终端需要同步接口硬件或用一些其他方法将同步比特流传送给运行在 PC 异步接口的 modem。ITU-T V.80 提供了一种 PC 和 modem 间标准的异步一同步“通道”协议。

当运行在除 PSTN 外的低比特率线路交换网上时,V.34 modem 被另一种适当的网络接口取代。H.324 也可以看作是 ISDN 网上 H.320 的第二代替品。

2. 呼叫建立

一个 H.324 呼叫建立和终止过程分为 7 个阶段(1~7)。

(1) 用正常的过程(拨号、响铃、响应)建立一个普通电话呼叫。这可以完全手动或自动完成,使用 V.25 ter 协议(“AT”命令集)控制 modem 拨号和响应功能。

(2) 一旦呼叫连接,H.324 终端能立即开始多媒体呼叫,或者如果两终端都支持可选的 V.8 bis 协议,于是两终端用户可以首先进行普通的语言交谈,直到其中一个用户决定将呼叫切换到 H.324 多媒体模式时,这种谈话才会结束。V.8 bis 允许这种“晚启动(late start)”模式,也允许一个普通电话呼叫在任何时候都可切换到多媒体模式。

(3) 每个终端发送 V.8 或 V.8 bis 消息,开始进行 modem 协商,使用 V.23 300 bpsFSK 调制(频移键控,不需要任何“训练”时间)。modem 交换容量选择 V.34 调制方式和 H.324 协议。PC 或其他外部 DTE(Data Terminal Equipment,数据终端设备)能够控制这个 V.8 或 V.8 bis 协商过程。如果使用 V.8 bis 协议,支持 H.324 视频、音频、数据或加密的终端大容量也互相通信以迅速决定预期的模式是否可用。然后,V.34 启动过程需要 10 s,建立起运行在电话连接质量允许的全 V.34 数据速率上的端到端数字通信,最高速率可达 33 600 bps。

(4) H.324 终端首先在 H.245 控制信道上互相通信;然后终端被初始化,用 H.245 交流详细终端容量。由于这类控制信息传递可用 V.34 比特率,所以上述信息的交换发生很快,不超过 1 s。最后为各类不同的媒体类型打开各自的逻辑信道,同时 H.323 复用器下载表入口(table entries)来支持这些逻辑信道。

(5) 此时,终端进入正常多媒体通信模式,在传呼期间可以改变逻辑信道的数目和类型,每个终端也可动态改变自己的容量。

(6) 当某个用户想终止呼叫进行下一阶段时,假如逻辑信通被关闭,一个 H.245 消息通知远端终端终止呼叫的 H.324 多媒体

模式部分,并设定使用的新模式(断开,回到声音模式或另一种数字模式)。

(7) 在第 7 阶段,终端断开,回到声音电话模式,或进入所设定的新模式(传真,V.34 数据通信)。

3. 复用

H.324 使用一个新的复用器标准 H.223,多个数据流(视频、音频、数据)和控制信道混合成单个比特流,在 V.34modem 上进行传输。为了降低复用器延迟并高效率地合并处理来自可变数目逻辑通道的突发数据通信量,H.324 应用程序必须开发一种新的复用方法。

TDM(时分多路转换,例如 H.221,被认为不合适,因为它们不能很好地适应动态变化的 modem 数据速率和负载数据速率,而且由于复杂的帧同步和面向比特的信道分配问题,TDM 在软件层很难实现。

分组复用器(例如:V.42 (LAPM), Q.922(LAPF: Link Access Procedure Forframes))可以避免这些问题,但却要忍受“块延迟”,即紧急数据(如音频)传输必须等到已开始传输的大分组完成之后才能进行。当主要信道比特率低得使单个分组传输时间变得过分明显时,会导致“块延迟”问题。在传统的分组复用器中,这个延迟可通过限制分组最大尺寸或禁止大分组传输的方法降低,但这两种方法同时会降低效率。

H.223 复用器结合了 TDM 和分组复用的长处,并增加了新的特征。它比 TDM 和分组复用延迟短,且开销低,易扩展到每种数据类型的多个信通。H.223 面向字节,易于运行,可填充标志字节来适应不同的数据速率,且使用一个在有效数据中永远不会出现同步字符。在 H.223 中,每个 HDLC (High-Level data Link Control)-MUX-PDC(Multiplex-Protocol Data Unit,复用协议数据单元)可按不同比例携带多个不同的数据流,并允许不同信道的动态带宽分配。

H.223 由一个低级复用层(混合不同媒体流)和一组适应层

(执行逻辑成帧、序列编号、检错和重传纠错)组成,适用于每一种媒体类型。

4. 控制

H.324 使用新的 H.245 多媒体系统控制协议(得益于 H.320 系统控制实践经验)。H.245 也被 H.310(ATM 网)和 H.323(用于分组交换 LANs 和公司 Intranet)系统。H.245 控制模型基于逻辑信道,该逻辑信道是容易固定的独立的单向比特流,由发送者随意选择的唯一编号所识别,最多可达 65 535 个逻辑信道。

H.245 控制信道携带端到端控制信令,这些信令控制 H.324 系统的操作,包括容量交换、逻辑信道的打开和关闭、模式优先请求、复用表入口传输、流控制消息、通用命令和指示。H.245 结构允许将来扩充额外容量,同时也允许厂商进行自定义非标准扩展。

H.245 信令用 ASN.1 定义,并根据 ITU-T X.691 的 PER 进行编码;(它提供规范的灵活性和定义的简明性)。在 H.324 中, H.245 控制信道运行在一个同所有媒体类型都分开的通道上,即 LC0。当 H.324 呼叫启动时,LC0 早已经打开。H.245 将 V.42 或基于 V.42 XID(exchange identification)步骤的简单重传协议看作是一个可靠的链路层。H.245 本身假设该链路层能保证信令正确有序传递。

5. 音频信道

H.324 最基本的音频模式是 G.723.1 声音编码解码器,它带宽为 5.3 或 6.4 kbps。G.723.1 提供市内电话质量的声音,帧长度为 30 ms,在通用的 DSP(数字信号处理器)上运行 G.723.1,估计需要 18~20 定长 MIPS。发送者可使用任何一种速率,由于编码器速率作为每帧语法的一部分发送,所以每一传输帧都可改变速率。利用“无声抑制”(silence suppression)方法进一步降低平均音频比特率,“无声抑制”是指不传输无声帧或用更小的帧只传输背景噪音信息。在典型谈话中,双方很少同时讲话,可以节省一部分带宽给视频和数据信道使用。

接收端可以利用 H.245 信令发出信号,优先使用低速率或高速率音频。音频信道使用 H.223 AL2,它包括一个 8 比特的 CRC (每音频帧或每组帧有一个 CRC)。

G.723.1-CODEC 有大约 97.5 ms 端到端音频延迟,如果加上 modem,抖动缓存,传输时间,复用器和其他延迟,最终将导致 150 ms 总的端到端音频延迟,不包括传播延迟。

声音延迟一般都小于视频 codec 延迟,所以如需要“唇同步”,在音频传送通路上需增加额外延迟,只需在接收端增加声音延迟即可。用 H.245 发送一个包含传输视频和音频信号间时间偏差的信令,接收端知道每个信号流的解码延迟后,时间偏差信令 (time skew message) 允许接收端插入正确的声音延迟或忽略唇同步使音频延迟达最小。当传输多个音频信道时,H.324 并不立即提供用于立体声音频正确取样的信道同步。

许多 H.324 应用不需要唇同步,或根本不需要视频。对于这样的应用,让 H.324 音频 codec 可选是非常重要的,如选用 G.729 (8 kb/s 声音 codec) 可将总端到端延迟降到大约 85 ms。非标准音频格式可以同标准 codec 以相同的方式使用。例如,建议的 V.34Q 声音/数据 modem 调制可将端到端声音延迟减少到大约 60 ms,可以作为一种非标准模式使用。其他可选的音频模式也可以同 H.324 一起使用,如即将推出的 16 kbps ITU-T 宽带 (7 kHz) codec,提供的音频质量比传统电话线高得多。

6. 视频信道

在 modem 可用带宽的任意子带宽上,H.324 都可发送彩色运动视频。H.324 支持 H.261 和 H.263 视频编码,优先选用 H.263,H.261 用于同 ISDN H.320 视频会议系统互操作,而不需转换视频格式,否则会增加难以忍受的延迟。

H.263 同 H.261 一样基于相同的视频压缩技术,但都增强了功能,包括改进的运动补偿,使 H.324 只需 50%~100% 的比特率就可达到与 H.261 相同的视频质量。当 H.324 用于低比特率情况时,这种改进尤为明显;但 H.263 用于高比特率时,与 H.261 差

别很小。H.263 还包括范围更广的图像格式。

受比特率、图像格式、H.263 可选项及场景复杂性和活动量大小的影响, H.324 视频可达 5~15 帧/秒。H.245 控制信令: VideoTemporalSpatialTradeOff 允许接收端在帧速和图像分辨率之间进行一定的折衷。

视频信道使用 H.223 AL3, 它包括一个 16bitCRC 和序列编号, 以及用于接收端错误视频数据重传的相关信息。

既然 H.324 系统支持视频多信道(尽管许多应用中由于带宽所限而无法实现), 通过多个视频逻辑信道可以很容易实现连续显示的多点操作, 此时, 接收端可以用“Hollywood Squares”方式显示每个会议场点的不同图像, 由接收者安排用于本地显示的合适信道集。

7. 数据信道

除视频和音频信道外, H.324 多媒体系统也携带数据信道。用与普通 modem 一样的方式, 数据信道能用于任何数据协议或应用。标准协议包括 T.120 数据(用于会议应用, 如电子白板、应用共享)、用户数据 V.14 或 V.42(带重传功能), 用于远端摄像头摇动、倾斜和推拉功能控制的 H.224/H.281, 以及 ISO/IEC TR9577 网络层协议, 如 IP 和 IETF PPP(可用于提供实时视频内容的 Web 浏览)。正如其他媒体类型, H.324 也提供非标准协议的扩展能力, 这些非标准协议自动协商, 其使用与标准协议完全相同。

为视频和音频信道定义的相同容量交换和逻辑信道信令交流过程也适用于数据信道, 因此数据容量的协商也可自动完成。这意味着目前 GSTN modem 从仅有数据的应用向前推进了一大步, 而在仅有数据的应用中, 数据协议和应用必须在呼叫前由人工安排好。

与所有其他媒体类型一样, 数据信道可作为不同的逻辑信道在 H.223 复用器上传输(携带), 而 H.223 复用器通过动态分配不同信道间的带宽来满足突发数据通信量。例如, 当数据信道需要额外带宽时(如传输文件或静止图片), 视频信道速率会降低或

停止。

8. 多点操作和 H.320 ISDN 互操作性

H.324 终端可直接加入多点会议,通过 MCU(中间桥作用)三到四个用户可加入同一会议中。既然每个多点连接的 modem 可以运行在不同的速率上,MCU 可发送 H.245 信令:FlowControlCommand 将一个或多个逻辑信道,或者全部应用的所有比特率限制到一种大家均支持的通用模式。

当 H.324 终端同 ISDN H.320 终端互操作时,类似的情况也会发生。同现存的 H.320 终端相互工作时,H.324/H.320 网关可以转译 H.324 与 H.320 协议间的 H.223 和 H.221 复用、控制内容以及声频、数据逻辑信道内容,使用 H.261 QCIF 视频格式来消除视频转译延迟。在上例中,网关类似于多点会议中的 MCU,可以发送 H.245 信令:FlowControlCommand 强制被传输的 H.324 视频比特率同 H.221 复用器使用的 H.320 视频比特率匹配。

将来,在 ISDN 上只需一个“虚拟 modem”,ISDN 上的双模式(dual-mode: H.320 ISDN 和 H.324 GSTN)终端便可直接同 H.324 终端互工作,虚拟 modem 可产生代表 V.34 模拟 modem 信号的 G.711 音频比特流。

第7章

典型视频会议系统

7.1 Microsoft NetMeeting 的使用

Microsoft®NetMeeting®允许在 Internet 上进行实时的音频、视频和数据通信。

7.1.1 系统安装环境

- Microsoft Windows 95 或 Windows 98;
- 至少带 8 MB 内存的 486/66 处理器(推荐使用带 12MB 内存的奔腾机);

-或-

- Microsoft Windows NT 4.0;
- 至少带 16 MB 内存的 486/66 处理器;
- 必须有 Microsoft Windows NT 4.0 Service Pack 3 才能在 Windows NT4.0 上共享应用程序。



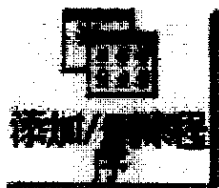
重要须知

- NetMeeting 无法在 Windows 3.1 或 Windows 98 Beta 3 之前的 Beta 版本上运行;
- 特别推荐在安装 NetMeeting 2.1 版本之前,卸载以前的 Microsoft NetMeeting Beta 版本;
- 安装 NetMeeting 过程中在键入安装路径时,如果使用了特殊的字符(例如“\”或双字节字符),NetMeeting 可能无法正确安装;
- NetMeeting 不再支持通过 IPX 或 PSTN 发出呼叫;
- 与 Internet 建立快速连接时,Microsoft NetMeeting 性能最佳(28.8 kbps 的调制解调器或更快的调制解调器或局域网);

- 要使用 Netmeeting 的音频特性,必须配备声卡、扬声器及麦克风;
- 要使用 NetMeeting 的视频特性,必须配备视频捕获卡和摄影机,或者连接到计算机并口(打印机)的视频摄影机。在处理器速度低于奔腾的某些计算机上无法发送视频图像,带双向(ECP 或 EPP)并口的计算机运行并行视频摄影机的效果明显好于带单向并口的计算机。连接到视频捕获卡的摄影机比连接到计算机并口的摄影机占用处理器的时间要少,建议在计算机处理器低于奔腾 133 时,不要使用彩色并口摄影机。
- 使用通过慢速连接的音频和视频时,应用程序的共享特性可能会出现问题。另外,使用视频时可能会听到间隔跳跃的声音。通过 28.8 kbps 的调制解调器连接时的默认视频设置质量一般。要更改设置,可单击“工具”菜单上的“选项”命令,单击“视频”选项卡,然后更改“视频质量”选项。
- 如果 Windows NT 的用户安装了 Service Pack 3 for Windows NT 4.0,则可以共享应用程序。

7.1.2 系统卸载

(1) 双击控制面板中的



(添加/删除程序)图标。

(2) 在“安装/卸载”选项卡上,单击“Microsoft NetMeeting”,然后单击“添加/删除程序”。

(3) 如果显示对话框询问是否要重新启动计算机,则单击“是”。

或者按以下步骤操作:

(1) 双击“控制面板”中的“添加/删除程序”图标。

(2) 在“安装 Windows”选项卡上,单击“通讯”,然后单击“详

细资料”，如图 7-1 所示。

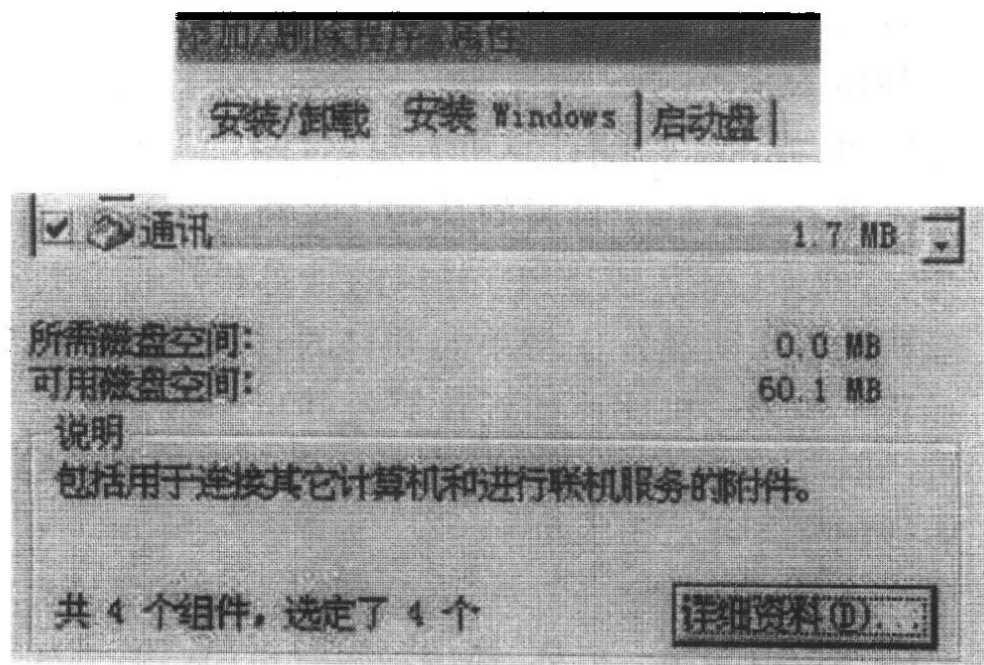


图 7-1 系统安装与卸载

(3) 清除“Microsoft NetMeeting”复选框。

7.1.3 开始使用 NetMeeting

Microsoft NetMeeting 包括了对 H.323 音频和视频会议标准的支持。NetMeeting 可用于向 H.323 兼容产品(包括 Intel Internet Video Phone)中发出呼叫和从这些产品接收呼叫。通过目前由第三方开发的适当的设备和服务，NetMeeting 可以使用 H.323 网关向电话发出呼叫。

Microsoft 在 Internet 上保留了下列用户目录服务器：

- ils.microsoft.com
- ils1.microsoft.com
- ils2.microsoft.com
- ils3.microsoft.com
- ils4.microsoft.com
- ils5.microsoft.com

无论服务器是否出现在向导的服务器列表中，都可以使用上

面所列的任何一个目录服务器呼叫。

另外,也可以使用不由 Microsoft 维护的下列服务器:

ils.four11.com

ils.business.four11.com

ils.family.four11.com



重要须知

- 如果通过不支持远程 WinSock 的代理服务器与 Internet 连接,则无法使用这些服务器查找 Internet 上的用户。
- 只有通过 TCP/IP 连接并且在同一时刻只与另一个用户连接时,音频和视频才能正常工作。
- 如果使用此版本的 NetMeeting 呼叫正在使用 NetMeeting 2.0 的用户,呼叫切换功能可能无法正常工作。
- 此版本的 NetMeeting 允许多达 32 位用户共享同一应用程序。但是,如果有一台运行 NetMeeting 2.0 的计算机也在会议中,并且有三个以上用户在共享应用程序,那么,这台计算机将无法运行该应用程序。另外,在同一时刻通过音频和视频只能与其中某一用户连接。
- 可成功参加会议的总人数取决于可用的网络带宽和与会者的计算机速度。
- 某些活动会导致参加会议的计算机之间发送大量数据(例如,在传输大文件时交谈而且共享着几个应用程序)。在极端情况下,这可能会引起会议中的计算机变慢。为解决这一问题,请停止会议中的一项或几项活动。
- 如果想在局域网上进行 TCP/IP 呼叫,或者想通过代理服务器和 MSN,或通过拨号网络显示“连接”对话框,请单击控制面板上的 Internet 图标,然后确认未选中“自动拨号”复选框,这样可以禁止显示该对话框。
- 共享 DirectX,OpenGL,MS-DOS,增强图形游戏及 AVI 都

不能良好运行,而且不被支持。

- 如果在开会期间各计算机屏幕的分辨率不同,共享窗口大小可能与最高分辨率桌面所允许的窗口大小相同。使用低分辨率桌面的用户必须滚动窗口,才能查看到超出屏幕的窗口部分。

- 必须有 32 位的 TCP/IP 堆栈和 Windows 套接字接口,才能通过 TCP/IP 使用 NetMeeting。NetMeeting 已通过内置于 Windows 95 和 Windows NT 4.0 中的 32 位 TCP/IP 和 Windows 套接字测试。

- 应用程序共享特性、白板及交谈程序,在使用不同语种设置和键盘布局的计算机之间无法正常运行。

- Microsoft NetMeeting 支持 Windows 95 和 Windows NT 4.0 系统策略的使用,以便在企业环境中设置默认配置。

- 如果 NetMeeting 在启动时就已挂断,请检查在“字体”文

件夹中是否存在已损坏的字体



- 如果在脱机方式下使用 Internet Explorer 4.0,并试图启动 NetMeeting,NetMeeting 将无法自动连接到 Internet。为解决这一问题,对于 Windows NT 的计算机请使用 RAS 拨号,而对于 Windows 95 的计算机则使用拨号网络,或者在 Internet Explorer 中,单击“文件”菜单上的“脱机工作”命令清除复选标记。

- 如果建立的连接为 SLIRP 或其他模拟的 SLIP/PPP 类型,NetMeeting 将无法正常运行。

7.1.4 发出呼叫与接收呼叫

您可以向多个用户发出 NetMeeting 呼叫。参加会议的任何人都可以通过 NetMeeting 向其他人发出呼叫,如图 7-2 所示。除此之外,其他人可以通过呼叫其中一位与会者,从而加入到正在进行的 NetMeeting 会议中。

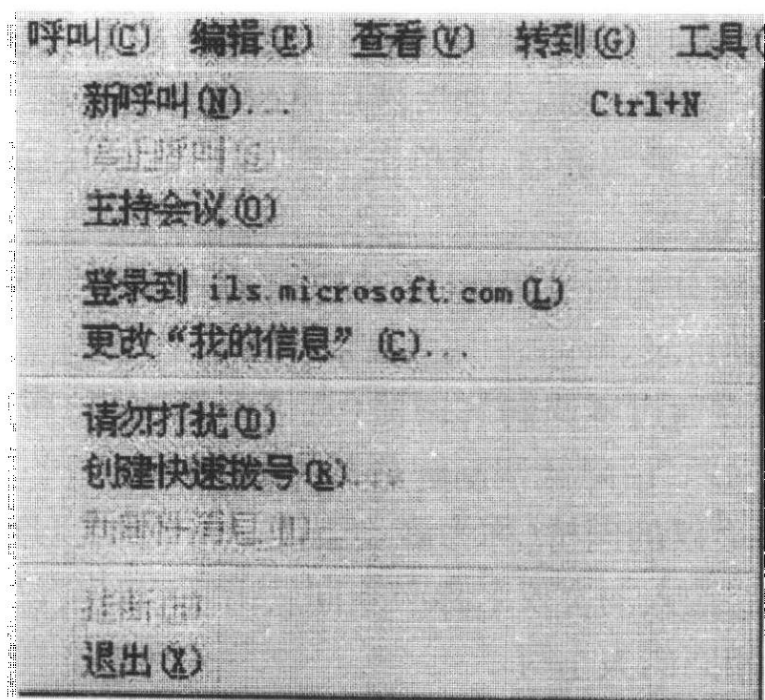


图 7-2 向用户发出呼叫

- 如果无法通过用户的计算机名与某位用户连接,请试用他们的 IP 地址。
- 如果单击 Web 目录中某位用户的名称,屏幕上将显示一则消息,询问是否打开或保存文件,这时请选择打开文件。
- 如果单击 Web 目录中某位用户的名称,可能会遇到一则错误信息,指出不识别这种类型的文件,这时需要配置浏览器,将 ILS 目录文件和 NetMeeting 关联。

关于如何关联文件的步骤,请参阅浏览器的相关文档。并且需要输入以下信息:

MIME 文件类型:“text/iuls”

命令行:“rundll32.exe msconf.dll,NewMediaPhone %l”

MIME 文件类型:“text/h323”

命令行:“rundll32.exe msconf.dll,NewMediaPhone %l”

- 某些 ISDN 设备配置为与网络自动连接。这可能会导致

ISDN 设备在运行 NetMeeting 时与网络试连接。为防止这种情况出现,请关闭 ISDN 设备的自动拨号功能。

- 如果在目录中未看见自己的姓名,可能是因为您未登录,或者登录到其他目录服务器上。若要更改登录的目录服务器上,请单击“工具”菜单上的“选项”,然后更改服务器名称。

要登录到目录服务器,请单击“呼叫”菜单,然后单击“登录到[您的目录服务]”。

- 如果您分别通过两块网卡建立了两个有效的网络连接,可能无法连接到目录。

- 如果在通过零调制解调器(直接电缆连接)所建立的连接拨号时遇到问题,请降低调制解调器的波特率。

- 某些会议的网桥(称为多点会议单元或 MCU)是区分大小写的。因此您输入的会议名称必须同发出呼叫时会议网桥所显示的名称具有相同的大小写。

- 若想通过零调制解调器(直接连接)电缆接收呼叫,请依次单击“工具”菜单、“选项”命令、“协议”选项卡。选中“零调制解调器”协议,然后单击“属性”。选择连接零调制解调器电缆的 COM 端口。

- NetMeeting 无法检测通过与零调制解调器连接的 COM 端口上拨入的呼叫。若想通过零调制解调器的电缆接收呼叫,必须通过 COM 端口发出呼叫。

- 在 NetMeeting 中对 COM 端口属性的更改仅对 NetMeeting 有效。建议在发出或参与 NetMeeting 呼叫时,不要更改 COM 端口的设置。

- NetMeeting 使用在“新呼叫”对话框内键入的 COM 端口发出呼叫,而不是用“通讯端口协议的属性”对话框中指定的 COM 端口。

- NetMeeting 可以只用 COM1 和 COM2 发出和接收零调制解调器呼叫。

- 零调制解调器呼叫可能无法以高波特率工作。

- 如果通过代理服务器连接到其他用户,可能需要在加入会议后手工连接音频和视频。



接收呼叫

如果正在运行 Microsoft NetMeeting,但未选择“呼叫”菜单上的“请勿打扰”,则可以接收拨入的呼叫。



警告和已知问题

- 如果无法接收呼叫,可能是由于 IP 地址未能正确注册到目录服务。为避免这种情况的发生,可以在连接到 Internet 之前释放局域网的 IP 地址。可执行下列操作释放 IP 地址。

① 在 Windows 95 桌面上单击“开始”菜单,单击“运行”,然后键入 Winipcfg.exe,如图 7-3 所示;

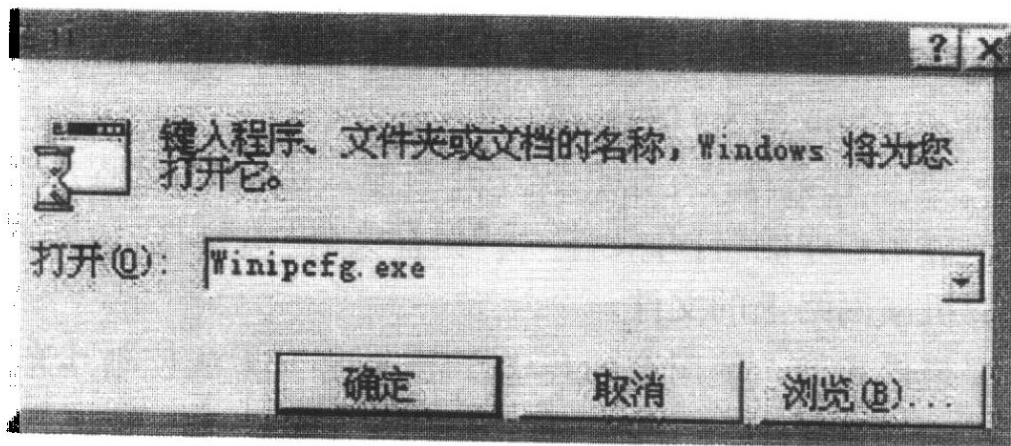


图 7-3 在“运行”程序对话框中键入 Winipcfg.exe

② 在 Windows NT 桌面上单击“开始”菜单,单击“运行”,然后键入 Ipconfig /release。

共享应用程序参加会议的任何人都可以与其他与会者共享应用程序。共享应用程序时,其他人可以看到应用程序。

采用“协作”方式共享时,其他人都可以看到并使用该应用程序。



警告和已知问题

- Windows NT 的用户只要安装了 Service Pack 3 for Windows NT 4.0 就可以共享应用程序。

- 如果 NetMeeting 安装在包含扩展字符的目录下,应用程序的共享功能可能无法执行。

- 共享“Windows 资源管理器”窗口时,该窗口中所有的资源都可以共享。如果在共享“Windows 资源管理器”时运行某个程序,该程序也将被共享。

“Windows 资源管理器”窗口的例子包括控制面板、公文包、“接收文件”文件夹或者任何 Windows 资源管理器文件夹。

- Internet Explorer 4.0 的用户如果共享“Windows 资源管理器”窗口,而且在允许协作的情况下,正参与协作的某一位用户关闭了窗口,那么,以后您打开的所有应用程序和窗口都将被共享。

在这种情况下,如果想取消共享,请再次打开“Windows 资源管理器”窗口,取消窗口的共享。

- 当您在共享某个应用程序并与其他人协作时,远程用户可以使用该应用程序的“打开文件”或“保存文件”对话框,访问或删除您计算机或网络上的文件。

要想停止他人使用您的共享程序,而您又无法控制其光标时,请单击 ESC。

当您可以控制其光标时,要想停止他人使用共享程序,请单击“停止协作”按钮。

建议在共享应用程序和协作时,不要将计算机置于一旁不管。

- 从正在共享的应用程序(例如:在 Microsoft Word 中共享 Microsoft Excel 电子表格)中启动另一应用程序时,后启动的应用程序可能无法正常共享。

- 无法将对象拖到共享应用程序或将对象从共享应用程序拖到桌面上。

- 与输入法编辑器(IME)共享应用程序时应该显示 IME 状态栏,以便其他人可以用鼠标激活 IME。

如果 IME 不支持显示状态栏,或者其他人也难以激活 IME,则应该为他们激活和关闭 IME。

- 如果在会议期间 IME 窗口无法刷新,可以单击桌面上的任何位置以刷新 IME 窗口。

- 对于计算机上安装的包含应用程序共享或远程控制功能的产品,您可能无法与其共享应用程序。

7.1.5 白板、闲聊与文件传送

1. 白板使您可以与其他用户同时绘图。会议中的每个人都可以在白板上看到所画的内容。一旦有人运行白板,在每个人的屏幕上都将显示白板,如图 7-4 所示。

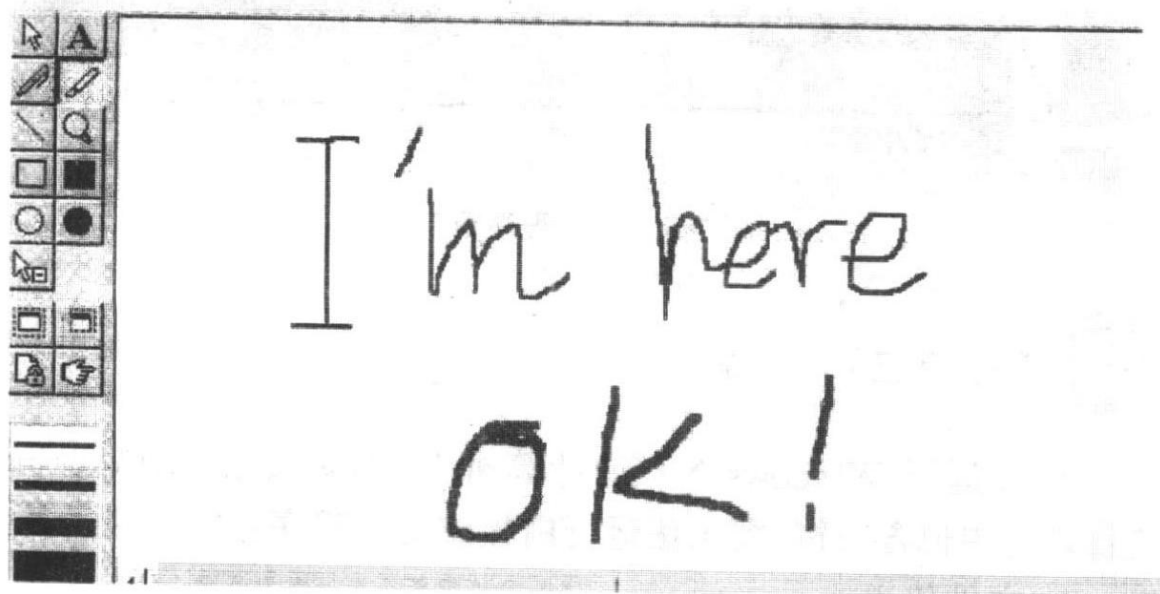


图 7-4 使用白板



警告和已知问题

- 在运行 Windows NT 的计算机上,如果安装 NetMeeting 的目录名中包含空格,则无法运行白板或交谈程序。

- 在 $1\ 152 \times 864$ 或更高的分辨率下,白板程序的窗口不能最大化。

2. 闲聊

闲聊程序允许您输入信息与其他用户对话。一旦有人运行闲聊程序,在每人的屏幕上都将显示该程序的窗口,如图 7-5 所示。

闲聊文件以 .txt 扩展名存储。如果已经在您的计算机上安装了 Microsoft 记事本或其他可以阅读 .txt 格式的程序,可以用它来打开闲聊文件。

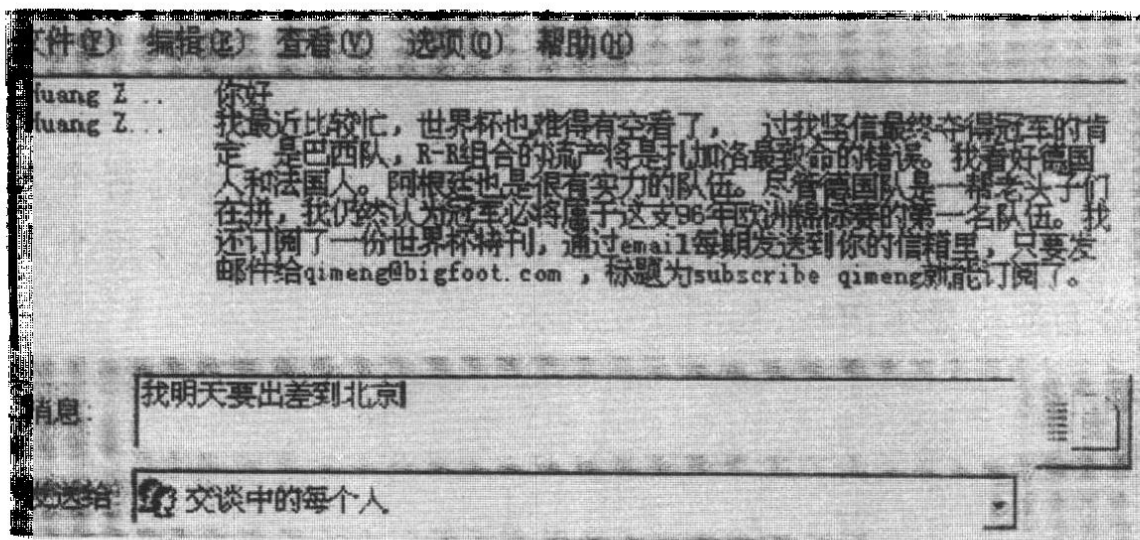


图 7-5 闲聊程序窗口



警告和已知的问题

- 在运行 Windows NT 的计算机上,如果安装 NetMeeting 的目录名中包含空格,则无法运行白板或交谈程序。

3. 文件传送

- 如果文件被拖动到“当前会议”选项卡上的用户清单中,该文件将传送到所有参加会议的人。

- 要向特定的人传送文件,请在与会者列表中,用右键单击这个人的名称,然后单击“发送文件”。

- 要向所有与会者传送文件,请单击“工具”菜单,指向“文件传送”,然后单击“发送文件”。



警告和已知问题

- 传送文件的快捷方式实际上只传送文件而不传送快捷方式。
- 用 Intel ProShare 2.0 交换文件有时会出问题。

7.1.6 音频通讯

- 只能通过 TCP/IP 进行音频通讯,而且人数不得超过两人。
- 音质的差别比较大,这取决于声卡和麦克风的质量。
- 使用全双工音频对带宽和 CPU 的要求比半双工要高,但可能会导致音质较差。

• 无论以何种方式修改声卡设备驱动程序,例如将其升级到全双工驱动程序,都必须重新运行音频调整向导才能使 NetMeeting 正常工作。



警告和已知问题

• 如果使用 Videum (Winnov) 的视频捕获卡并且拥有新的 2.0 驱动程序,您可以通过打开 Videum 的配置文件(在 WinnovVideum 程序组下)和选择 NetMeeting 使用的音频输入设备来更改捕获卡的音频输入设备。默认情况下,对于 MXC 视频连接,可使用摄像机麦克风。而对于其他视频连接,则使用 Line 输入。这样,在启动 NetMeeting 时,将选择正确的音频输入。

• 使用视频时,WinNov 摄像机用户的音频输入将自动切换。如果视频是通过 MXC 连接,那么摄像机输入将被用于音频。如果视频是用 Composite 或 S-Video 连接的,Line 输入将用于音频。您可以通过升级到新的 2.0 驱动程序并使用 Videum 配置应用程序解决此问题。

• Microsoft NetMeeting 和 Intel Internet Video Phone 的用户只有都使用奔腾处理器的计算机时才能相互交谈。

- 如果允许使用全双工的声霸卡,而且在使用 NetMeeting 时与计算机之间的配合出现问题,可能需要更新您的声霸卡驱动程序。

- 您可能会在音频调节向导中遇到不支持这种声卡的信息。当声卡不支持 NetMeeting 所需的特性时会发生上述情况。即使您得到该信息,NetMeeting 的音频特性也能工作,但音质可能会很差。

如果您的声卡不被支持,可以同制造商联系以获得新的声卡驱动程序。

- 如果全双工方式音质不好,请执行以下步骤将 NetMeeting 改成半双工音质:

- ① 确认目前未进行 NetMeeting 呼叫;
- ② 单击“工具”菜单上的“选项”命令;
- ③ 在“音频”选项卡上,清除“启用全双工音频”复选框,如图 7-6 所示。



图 7-6 音频选项卡

- 建议您在参加音频会议时不要在全双工和半双工音频之间进行切换。

- 使用 NetMeeting 时所遇到的音频质量或声卡方面的问题可能与硬件配置或驱动程序安装有关。若想了解最新的支持信息,请单击“帮助”菜单,然后单击“联机支持”。

- 如果您的计算机有多个音频设备,必须确保在音频调节向导中选择的音频设备与控制面板“多媒体”属性中的选择一致。

- 如果计算机的处理器升级,则应该重新运行音频调节向导。

- 全双工音频可能不支持 Media Vision Pro Audio Spectrum

16 声卡。

- 486 计算机不能通过拨号网络(调制解调器)利用基于 H.323 音频会议标准的其他软件使用音频,但可以通过拨号网络或局域网使用其他基于 H.323 音频会议标准的产品与其他 NetMeeting 用户使用音频进行通信。

- 使用 NetMeeting 参加音频会议时,具有录音功能的应用程序(如 Power Point)将按指定的过程录音。然而,由于 NetMeeting 已使用声卡,所以实际上其他应用程序将无法录音。

- 如果有 Tecra 740,则需要在“控制面板”的“多媒体属性”



多媒体

中手工设置麦克风音量。

- 如果音质较差,请试着清除“麦克风”复选框,然后再次选中。
- 如果您使用 Gravis Ultrasound 声卡并听到间隔跳跃的声音,请按如下步骤操作:

- ① 确认没有进行 NetMeeting 呼叫;
- ② 单击“工具”菜单上的“选项”命令;
- ③ 在“常规”选项卡上的“网络带宽”区域,选择低于当前选定设置的带宽。



挂断

挂断会使所有您呼叫或呼叫您的人被断开。如果被断开的用户已邀请或接受其他用户的邀请参加会议,他们也断开连接。用户之间的音频连接可能不会断开。



警告和已知问题

- 如果与两个或多个用户通过 56 kbps 的调制解调器使用 NetMeeting 相互呼叫,挂断功能可能工作不正常。在呼叫期间切

换音频和视频可能会导致部分参与者断开连接。

7.2 其他视频会议系统简介

7.2.1 CLI 公司及主要产品

在美国几家生产视频会议设备的公司中,CLI 公司是市场占有率较大的公司,全球市场占有率在 35% 以上,中国市场占有率在 60% 以上,且行销 64 个国家,其全球销售量在 8 000 台以上。它拥有一套 CLI 专有的 CTX 及 GTXPLUS 的编码算法。这种算法可提供高于 H.261 编码算法的图像质量。该公司生产的 Rembrand II VP 编译码器,采用了 H.261 规定的 CIF, QCIF 图像格式,同时又具有高图像质量的 CTX, CTS+ 的图像格式,因此保证了与具有 H.261 标准的编译码器的互通,并已实验证实 CLI 公司与英国 GTP、BT 公司,日本 NEC 公司的编译码器进行了互通。

CLI 公司于 1994 年推出了超集成电视画质的视频会议系统 Radiane(含编译码器)。由于该设备采用图像压缩可编程芯片,其处理速度比采用 PC 芯片的同类设备的处理速度快 1 000 倍。同时它使用 CLI 的自导技术(Self-guide),因此可应用基于触模板的图形象标来指示每项功能,包括 H.243 建议的主席控制模式的操作。从标准及功能来考虑,它增加了有关 H.243 的内容以及语音压缩编码 G.728、远端摄像控制标准 H.281 的软硬件单元,使得设备功能更加完善。

CLI 公司生产的“ECLIPSE 系统”是工作在低速率(56/64 kbps, 112/128 kbps)具有拨号功能的视频会议终端设备,支持 H.261 建议,以及 CTX 等编码算法,其低速图像质量达到良好状态。

CLI 公司生产的多点控制设备 MCU II 具有 12 个 2 Mbps (E1)端口,这样便可减少网中的 MCU 的个数而达到相同的功能,

从而降低了网路的复杂性。MCU II 除符合 H.200 等建议规范外,尤其符合 H.243,H.224,H.281 等建议,以实现全网的主席控制模式以及远端摄像机控制的功能。

由于 CLI 公司领先的视频压缩编码技术和严格的质量管理,从而于 1994 年 4 月获得 ISO9001 最高质量证书。

7.2.2 Intel 的 PPCV200

PPCV200 是 ProShare Personal Conferencing Video System 200 的简称。

它是运行于 ISDN 环境的一个典型产品。它运行于 486 及其以上的 PC 机上,需要为摄像机服务的视频数字卡即 CODEC,负责视频信号的采集、编码、压缩,以及一块 ISDN 网接口卡,同时进行声音信号的采集、编码、压缩和传送。PPCV200 的 CODEC 采用 Intel 自己的 Indeo 专有算法,从本地送出的图像通过 CODEC 压缩,而远端送进的图像则通过软件进行解压缩。因此,它的视频质量依赖于 PC 平台,在 486 上测试,每秒可提供 15 帧画面,动作的光滑度较好,但图像的彩色准确度和饱和度有待提高。

PPCV200 提供两个窗口,可同时看到自己和对方的模样。它的软件运行在 Windows 环境下,提供一个支持 OLE 的白板软件作为双方屏幕的共享部分,凡是调入白板的内容,双方均能同时看到并进行协同修改。在白板的下方是用户的私人区,供用户作笔记或进行注释,而对方看不到。

7.2.3 Picture Tel 公司的 LiveLAN

运行在 LAN 上的视频会议系统必须要与其他信号共享带宽,一旦发生冲突其视频数据必然受到一定程度的堵塞,从而影响画面的质量,因此要求软件提供动态监视网络的能力,即软件的智能化。LiveLAN 便具有这一功能,它根据网络上的负载自动调整

视频信号,同时还能监视用户 PC 的 CPU 工作状态,既保证视频会议的顺利进行,又不使网络超载。

LiveLAN 当今运行在 LAN 上 PC 平台的视频会议系统中具有典型意义,它提供三个模块:Live-A,LiveLAN-V 和 Live-Manager。Live-A 可把 PC 变为一个免提电话,使用户能够通过 LAN 与他人通话,也可以进行数据共享。在此基础上,如果加上后两个模块即可升级为一个视频会议系统,并可通过路由器跨越局域网。由此,LiveLAN 为用户提供了不同的使用层次。此思路与清华大学的 TH-Multimedia Conference System 类似。

7.2.4 Creative 的 Share Vision PC3000

该产品通过高速 modem,在普通电话线上实现了声、像、文的动态传送,还可共享应用程序和数据。它要求一台 486 或以上 PC 机、一块音频压缩卡、一块视频采集压缩卡 Video Blaster RT300、一台彩色摄像机和外置 V.Fast 28.8kbpsFax/modem。连接 RT300 的是高性能的视频捕获卡,它采用 Intel i750 视频处理器,使用 Indeo 压缩算法,可以在 320×200 或者 160×120 的窗口中进行视频捕获和回放。

Share Vision PC3000 在 PC 屏幕上提供两个窗口,对话双方可同时看到自己和同伴的动态图像。它的软件置于 Windows 下与标准 Windows 应用软件一同工作。用户可以很方便而快速地发送和接收文件,而且提供实时的软件共享,两个人可以同时用一个应用程序来编制文件,而这个应用程序只需要一端用户具有即可。它的白板软件提供了一系列的工具,两个人可迅速而方便的用图文来表述意见,同时可捕捉动态视频窗口或显示器上的图像或扫描仪进行的图像,供两人一起讨论。由于 Share Vision 系统在普通电话线上即可运行,因此视频的稳定性取决于各时刻用户所进行的作业,传输文件时会降低视频的速率。

从技术上讲,Share Vision 最优先考虑的是语音,然后是数据,

最后才是视频信息。视频是以每秒 10 帧的全彩色在 160×120 像素的屏幕窗口上显示的,若在更小的窗口上显示的话,速率可达每秒 15 帧。

7.2.5 Apple 的 Quick Time

1995 年 2 月,Apple 公司发布了一种跨平台的会议系统合作多媒体技术,它允许个人电脑用户共享全世界任何地方的实时信息、图像和语音,这个技术被称为 Quick Time 视频会议系统技术。它有如下几种功能:

(1) 举行视频会议并进行合作。分布在世界各地的与会人员能够实时地共享和评注文字、图像、屏幕捕捉和虚拟场景。

(2) 在 Macintosh、PC、UNIX 系统和室内会议系统间实现跨平台的视频会议连接。该连接采用 H.320 全球远程会议标准。

(3) 在局域网或广域网上播放及查看多媒体资料,包括数字语音、音乐和视频图像。

通过与合作厂商的合作,大批产品将会在 Quick Time 基础上开发出来,例如:

(1) 苹果多媒体会议成套系统包括 Quick Time 视频会议系统功能扩展、苹果多媒体会议应用程序和一个高质量的彩色摄像机。

(2) 苹果多媒体会议专业成套系统包括上述设备和一个 H.320 编码译码器/ISDN 适配卡。

(3) 完全多媒体视频会议系统包括苹果多媒体会议成套系统,一台 Power Macintosh 7100AV、一个 17 英寸彩色监视器、外接扬声器和键盘。

由于 Quick Time 是一个基于软件的系统,它很容易被集成到新的或现有的第三方产品中。设计的目的是为用户开发提供视频通信和合作机会的解决方案。Quick Time 是第一个采用软件视频图像压缩技术的产品,它尽管显示的图像很小且图像模糊,但毕

竟给用户一种耳目一新的感觉。

7.2.6 华中科技大学电子与信息工程系的桌面视频会议系统

采用 JPEG 压缩/解压缩卡对视频、音频进行同步压缩与解压缩,压缩比可以调整。同时,采用中文 Windows 98 作为标准人机界面。

该系统提供多点双向实时声音与视频的同步传输服务以及提供广播(broadcasting)、组广播(multicasting)和共享式白板(shared white board)功能。每个 PC 节点的屏幕可同时显示多个视频窗口、一个共享白板和—个控制窗口。其中—个视频窗口播放本地 PC 机附带的摄像机采集的视频画面,其他视频窗口显示远程 PC 机附带的摄像机采集的经过远程 JPEG 卡压缩后通过以太网传送过来的由本地 JPEG 卡解压缩的视频信息,共享白板可以使与会人员共同浏览、修改和编辑文件,共享鼠标,共同绘制各种图表,控制窗口除控制会议召集、会议响应、视频窗口的调整 and 共享白板的控制外,还包括视频的颜色、亮度、对比度的调节、监视功能等。

7.2.7 清华大学计算机系的 MCONF (多媒体会议系统)

系统的网络环境是细缆以太网,网络操作系统为 WFW(Microsoft Windows for Workgroup)V3.1。通信功能以 NetBIOS 实现。多媒体设备采用标准的 Windows MCI 设备。语音方面采用自己的声霸卡或兼容卡,视频则采用自己开发的 TH-VIDEO 视频采集播放卡。

MCONF = 会议平台 + 客户程序 + 会议工具

会议工具可完成会议的准备、邀请等工作。主要利用 WFW

的已有应用程序:CHAT(适时会话(字符方式)程序),SCHEDULE+(会议安排与调度),MAIL(电子邮件程序)。

会议举行期间的程序包括多媒体会议平台和客户程序两部分。APP.EXE 在系统中充当会议平台的作用,提供会议连接、管理功能,简单协作功能和通信管理、调度功能。其他各个程序是 APP 的会议客户程序,完成各种会议支持功能,APP 与客户程序之间通过标准的 Windows DDW(Dynamic Data Exchange)接口进行通信。目前,系统所提供的客户程序包括:

ACLI.EXE 完成语音的采集与播放,EDITC.EXE 完成基于文本的编辑功能的共享显示功能,PAINT.EXE 完成基于图形编辑功能的共享显示功能,VIDEO.EXE 完成视频采/播功能。

7.2.8 InSoft 的 Communique

该产品是一个基于 UNIX 的多点视频会议系统,它支持 HP-UX,OSF/1,AIX 及 Windows NT,不久将推出 Windows 版。

该系统区别于当前许多视频会议系统产品的地方是,其多点体系结构支持多个(两个以上)用户的使用。它运行在基于 TCP/IP 的以太网 LAN 和 WAN 上,InSoft 公司最近与新桥公司合作,希望 Communique 能支持 ATM。

它的另一个特点是:其 API 允许用户集成会议工具,如白板、音频控制和数据会议,并将应用嵌入视频会议分组。

InSoft 公司承诺,将允许 Windows 客户机上的用户与 UNIX 工作站上的用户进行视频会议。

7.2.9 SUN 的 ShowMe

SUN 公司推出的一种视频会议系统软件 ShowMe,基于 SUN 工作站上的 Openwin 窗口环境。它也实现了多点会议功能。

该系统运行时,在屏幕上有一个控制窗口,完成会议召集功

能,同时具有初始化共享白板、视频、音频窗口的功能。共享白板窗口可进行与会人员的共同编辑和文件等数据信息的适时传输,类似于 Windows 中的画笔软件(Paintbrush)的功能,但较画笔程序功能简单。它可由多人同时编辑,视频窗口可开零个到多个,每个对应一个与会人员,可实时显示视频信息。其性能参数的设置集成于相应窗口的菜单中,如设置视频信息的采集速率等。同时它具有监测网络性能的能力,能实时显示当前网络数据的传输量,以便用户参考来调节系统参数,提高系统性能。音频控制窗口具有调节音量,以及设置讲话方式的能力。

该系统是一个平等方式的多点视频会议系统,但缺乏对会议进程的控制功能,会议的协调要通过与会人员的预先约定。但由于该系统是建立在 SUN 工作站这种高性能的平台上的,因此具有较高的整体性能,不失为一个好的视频会议产品。

《现代信息网络实用技术讲座》

总 序

在知识经济初见端倪和网络浪潮潮涨潮落的世纪之初,世界经济体系更加依赖于人类的知识和智慧。人们更加需要信息,并且需要随时随地地获取信息。以因特网技术为核心的现代信息网络从根本上影响着人们获取信息的方式,因而,也进一步的影响了人们的工作、学习、消费、休闲、交往等生活方式。

人们迫切需要高速实时的现代信息网络。现代信息网络就是城市、国家或国际的信息基础设施(俗称“信息高速公路”)。因为只有具备了这个平台才有可能加速传播和共享知识,大大降低知识传递的成本,开创新的工作方式、管理方式、商贸方式、金融方式、人际交往方式、文化教育方式、医疗保健方式、消费与生活方式以及休闲方式。使知识经济真正成为数字经济和网络经济。

近年来人们一直在探讨现代信息网络的未来模式与发展趋势。这一切仁者见仁,智者见智,尚有很多的争议。但有一点可以肯定:现代信息网络要对全社会一般家庭成员的生活方式产生一般性的影响或者使之发生根本性转变,仅仅靠局部的个别技术和少数人的参与是不够的。网络购物在中国为何举步难艰,原因很清楚,我们缺少配套的其他信息技术以及百姓的认同意识。比如中国的信用卡持有率很低、通用性很差,邮政服务不尽人意,商用数据库信息匮乏等等。

现代信息网络的技术必须向百姓普及,必须为百姓所接受,其技术及应用已经不再只是信息技术行业从业人员关注的话题,它已深入到当今社会越来越广泛的人群之中,并被他们所应用,与我们的生活、工作越来越息息相关。现代信息网络的技术及其应用

作为实现社会信息化的一项重要支持技术和操作平台,其发展、应用和普及尤其令人瞩目,也受到世界各国政府的广泛重视。人们对现代信息网络实用技术及应用了解的需求也日益突出。

目前,有关现代信息网络方面的图书数量尽管很大,但大多内容庞杂、面面俱到、理论深奥、价格不菲,难于为广大普通百姓所接受,更谈不上向这类人群普及。为了帮助大家及时了解信息技术的发展、掌握新技术的应用方法,为顺应网络通信技术的发展大潮和网络通信关注者的迫切需求,北京邮电大学出版社周明老师与我们共同策划编写了这套《现代信息网络实用技术讲座》丛书,供大家学习使用,希望为广大关注现代信息技术领域的普通百姓提供专业的技术咨询和实用的信息服务。这套丛书紧密结合实际,重点介绍近年来迅速出现并发展起来的新技术、新应用。主要内容有:IP电话的基本原理和应用技术、电子商务、视频会议系统、智能大厦等。丛书的特点是结合发展,全面介绍新技术、新概念、新应用,突出实用性、系统性,内容详实、丰富,便于读者理解和掌握。相信不管是普通百姓还是专业人士都会开卷有益。

这一套丛书的宗旨在于:引导大众迎接信息网络时代的生活,以通俗的方式普及信息技术。纵览这套丛书的作者,他们中有华中科技大学、华中师大的博士和中青年教师、东风汽车工程研究院的高级经济师和工程师,他们把自己对现代信息网络的深刻理解和实际经验,凝聚在这套现代信息网络实用技术讲座中,奉献给大家。这些中青年才俊,他们为了整个文稿的简捷、通俗易懂而不厌其烦,几易其稿,这令我们既感动又宽慰,北京邮电大学出版社为这套丛书的出版倾注了大量的精力,我们谨此致以诚挚的谢意。殷切希望广大读者和各有关方面提出宝贵意见和建议,以便这套丛书日臻完善。

《现代信息网络实用技术讲座》丛书编委会

Images have been losslessly embedded. Information about the original file can be found in PDF attachments. Some stats (more in the PDF attachments):

```
{
  "filename": "MTA5MzYxODEuemlw",
  "filename_decoded": "10936181.zip",
  "filesize": 11395217,
  "md5": "3804307b0b653c71fbafe515f8114422",
  "header_md5": "d03a96fad2c72cddc494c65c407bbfa1",
  "sha1": "94f1c6d5e9409502345bb60e1d5c2ef08ead2336",
  "sha256": "1c582304057627364e70324a90c06f64f296ec38f59ba24a088d8ae02bb47a7f",
  "crc32": 1131015920,
  "zip_password": "",
  "uncompressed_size": 11745931,
  "pdg_dir_name": "\u2569\u2559\u255e\u2561\u2557\u00df\u2565\u0398\u2567\u2561\u2550\u2502\u255d\u2591\u255e\u03a3\u2559\u00aa\u2559\u251c_10936181",
  "pdg_main_pages_found": 142,
  "pdg_main_pages_max": 142,
  "total_pages": 151,
  "total_pixels": 602731200,
  "pdf_generation_missing_pages": false
}
```