

863

黄永峰 编著

李 星 审

人 民 邮 电 出 版 社

CIP

因特网语音通信技术及其应用 / 黄永峰 编著. —北京：人民邮电出版社，2002.1

（“863”通信高技术丛书）

ISBN 7-115-09739-9

I. 因... II. 黄... III. 因特网—语音信号处理 IV. TN912.3

中国版本图书馆 CIP 数据核字（2001）第 088029 号

内 容 提 要

本书全面介绍了因特网语音通信技术的原理和应用，全书共分九章。第 1 章介绍了因特网的发展过程、特点和因特网语音通信的关键技术。第 2 章介绍了 IP 网络的特点和 VoIP 基本原理及过程。第 3 章介绍了语音编码技术和回声消除技术。第 4 章介绍了流媒体服务质量动态监测和控制技术。第 5 章介绍了因特网语音通信的关键设备，如网关、关守和多点会议单元。第 6 章介绍了 ITU-T H.323 协议的体系结构、基本内容和呼叫建立过程以及会议系统的实现方法。第 7 章介绍了因特网语音通信技术的典型应用。第 8 章介绍了企业级的数据与电话网融合方案和实现方法。第 9 章主要介绍了计算机集成技术和 Windows 2000 的电话应用编程接口 TAPI 的编程方法和函数。

全书内容翔实，结构清晰，适合广大对通信知识感兴趣的读者阅读，尤其适合作为高等院校通信和计算机专业的教材。

“863”通信高技术丛书

因特网语音通信技术及其应用

◆ 编 著 黄永峰

审 李 星

责任编辑 陈万寿

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号

邮编 100061 电子函件 315@ pptph.com.cn

网址 <http://www.pptph.com.cn>

北京汉魂图文设计有限公司制作

北京 印刷厂印刷

新华书店总店北京发行所经销

◆ 开本：787×1092 1/16

印张：16

字数：385 千字 2002 年 1 月第 1 版

印数：1-5 000 册 2002 年 1 月北京第 1 次印刷

ISBN 7-115-09739-9/TN · 1792

定价：28.00 元

本书如有印装质量问题，请与本社联系 电话：(010)67129223

早期的因特网主要是用于电子邮件、电子布告、远程登录、全球浏览、数据通信和文件传输，但是随着计算机网络技术的飞速发展和多媒体技术应用的日益深入和普及，基于因特网的多媒体通信已成为当前因特网发展的趋势之一，而发展基于因特网的语音通信也是技术发展的必然趋势。语音和数据的综合可大量节省通信的开销，如果在因特网上的语音应用技术实现得好，节省的开销最高可达 90%。随着因特网的多媒体通信技术的发展，已经或将会产生许多类型的因特网多媒体通信业务，如：多媒体会议型业务、多媒体会话型业务、多媒体分配型业务、多媒体检索型业务、多媒体消息型业务和多媒体采集型业务等。实现上述业务的核心技术是采用因特网进行声、文、图等多媒体信息的传输技术，其中最关键的技术是音频和视频信号的实时传输。

作者多年来一直从事因特网语音信号的实时传输的研究和开发工作，深深体验到因特网语音通信的巨大发展潜力和广阔的应用前景。作者参加了国家 863 项目——IP 电话与视频电话的呼叫中心与目录服务的研究（项目编号：863-317-01-66-99）。该项目已通过国家 863 项目小组的鉴定和验收，取得了多项研究成果，并由此申请了多项专利。例如书中介绍的 Cool-Audio 软件就是其中的成果之一，它是在我国率先推出的具有自主知识产权的基于因特网的语音通信软件，目前用户已达几十万人。

本书主要是围绕因特网的语音通信技术及其应用这两个方面的内容展开，其中包括了因特网发展过程，因特网的特点和因特网语音通信的关键技术；IP 网络特点和 VoIP 基本原理及过程；语音编码技术、回声消除技术以及低速率语音编码及其实现；流媒体服务质量动态监测和控制技术及其应用；因特网语音通信的关键设备，如网关、关守和多点会议单元；ITU-T H.323 协议的体系结构、基本内容和呼叫建立过程以及会议系统的实现方法；因特网语音通信技术的一些典型应用，包括因特网电话及会议系统、基于 Web 的聊天室、基于因特网的呼叫中心，基于因特网的音、视频点播等；企业级的数据与电话网融合方案和实现方法；计算机集成技术和 Windows 2000 的电话应用编程接口 TAPI 的编程方法和函数。

本书的完成得到了李星教授的大力支持，李老师认真审阅了全书，在结构、内容组织等方面作了详细的指导。作者在研究工作中还得到其他老师和同事的协助。在此一并表示由衷的感谢。

作者

第 1 章 因特网语音通信技术概况	1
1.1 因特网业务的发展趋势	1
1.2 因特网的特性	2
1.3 因特网语音通信的典型应用	5
1.3.1 计算机电话集成	5
1.3.2 视频会议系统	6
1.3.3 基于 Web 网的呼叫中心	6
1.3.4 因特网媒体流技术	6
1.3.5 IP 电话	7
1.4 因特网语音通信的特点和关键技术	9
1.4.1 因特网语音通信的特点	9
1.4.2 因特网语音通信的关键技术	11
第 2 章 IP 网络和 VoIP	16
2.1 IP 网络的基本概念	16
2.1.1 数字数据交换方式	16
2.1.2 OSI 模型与 TCP/IP 模型	17
2.1.3 IP 包的传送	19
2.1.4 数据报的分段与重组	20
2.1.5 差错报告机制	22
2.1.6 TCP 数据传送可靠性服务	23
2.1.7 用户数据报协议 UDP	26
2.1.8 下一代 IP 网	26
2.2 IP 网络的关键技术——IP QoS	29
2.2.1 IP QoS 的概念	30
2.2.2 IP QoS 的流量管理	34
2.2.3 IP QoS 的实现策略	37
2.3 VoIP 的原理及技术	41
2.3.1 VoIP 的基本传输过程	42
2.3.2 推动 VoIP 发展的动力	43
2.3.3 软件技术及协议	46

2.3.4	VoIP 管理技术	48
2.3.5	VoIP 技术的应用	49
2.4	VoIP 的服务质量	50
2.4.1	影响语音质量的因素	50
2.4.2	VoIP 服务质量的分析	53
第 3 章 语音处理技术		60
3.1	低速率语音编码	60
3.1.1	低速率语音编码的原理	60
3.1.2	编码器的设计目标	63
3.2	G.723.1 双速率语音编码器的算法	65
3.2.1	编码算法	65
3.2.2	解码算法	76
3.2.3	静音检测和舒适噪声生成算法	80
3.3	G.729.A 编码器的算法	81
3.3.1	编码算法	81
3.3.2	解码算法	83
3.4	多通道语音编码器的实现	83
3.4.1	TMS320C6201 DSP 简介	83
3.4.2	编码器的结构	85
3.4.3	多通道语音编解码算法	86
3.4.4	编码器的执行速度优化技术	87
3.5	回声消除技术	88
3.5.1	因特网语音通信中回声的特点	88
3.5.2	声学回声消除器的结构和相关算法	89
3.5.3	自适应滤波器及其自适应算法的比较	94
3.6	回声消除器的实现	95
3.6.1	声学回声消除器的实现平台	96
3.6.2	声学回声消除器的关键技术	96
3.6.3	声学回声消除器的性能分析	99
3.7	用 16 位定点 DSP 实现回声消除器的误差分析	101
3.7.1	误差的理论计算	101
3.7.2	计算机仿真	103
第 4 章 RTP/RTCP 协议及流媒体的服务质量控制技术		105
4.1	实时传输协议 RTP	105
4.1.1	RTP 数据包格式及功能	105
4.1.2	RTCP 控制包格式及功能	106

4.1.3	RTP 的有关讨论	109
4.2	服务质量的动态监测	110
4.2.1	流媒体实时传输中的典型服务质量参数	111
4.2.2	服务质量的动态监测	111
4.2.3	服务质量参数在实际应用中的意义	115
4.3	服务质量控制方法	116
4.4	服务质量动态反馈控制	117
4.4.1	服务质量动态反馈控制的步骤	117
4.4.2	服务质量动态控制判据	117
4.4.3	服务质量动态控制算法	118
4.5	服务质量动态控制策略的可扩展性	120
4.5.1	可扩展性问题产生的原因	120
4.5.2	RTCP 扩展的技术难点	120
4.5.3	RTCP 可扩展性问题的解决方案及比较	121
4.6	服务质量动态监控在流媒体技术中的应用	123
4.6.1	流媒体的概念及意义	123
4.6.2	实时流媒体协议 RTSP 研究	124
4.6.3	RTSP 状态机	125
4.6.4	RTSP 系统设计和实现	125
4.6.5	服务质量动态监控在 RTSP 实现中的应用	128
4.7	自适应策略在语音编码器中的应用	129
4.7.1	自适应变速率语音编码器的基本结构	129
4.7.2	网络带宽的监测和自适应算法	131
第 5 章	因特网语音通信的关键设备	134
5.1	IP 电话网关	134
5.1.1	ETSI 的网关模型	134
5.1.2	实现策略	135
5.1.3	IP 电话网关的主要特征	137
5.1.4	一种小型 IP 电话网关的实现	137
5.1.5	小型网关的 Petri 网模型及性能分析	140
5.1.6	IP 电话网关的缓冲区大小的计算	144
5.2	关守	147
5.2.1	关守的类型	147
5.2.2	关守的功能	147
5.3	多点控制单元(MCU)	149
5.3.1	多点会议系统的实现方式	149
5.3.2	MCU 的主要功能和技术要求	150

5.3.3	集中式的会议系统中的 MCU	150
5.3.4	分布式 MCU	152
第 6 章	H.323 协议簇	155
6.1	H.323 协议简介	155
6.1.1	H.323 的体系结构	155
6.1.2	H.323 终端的组成	156
6.1.3	H.323 的标准协议簇	157
6.2	基于 H.323 协议的呼叫建立	161
6.2.1	终端到终端的简单呼叫建立	161
6.2.2	终端到终端的快速呼叫的建立	165
6.2.3	H.245 隧道	166
6.2.4	关守模式的呼叫建立过程	168
6.3	H.323 协议支持的会议	170
6.3.1	基本概念	170
6.3.2	创建或加入一个会议	170
6.4	SIP 协议	172
6.4.1	SIP 协议概述	173
6.4.2	H.323 和 SIP 协议的比较	174
第 7 章	因特网语音通信技术的应用	175
7.1	因特网电话	175
7.1.1	Cool-Audio	175
7.1.2	MediaRing Talk 99	178
7.1.3	Net2phone	179
7.1.4	Internet Phone	181
7.1.5	Video VoxPhone Gold	182
7.2	因特网传真	183
7.2.1	IP 传真与传统传真的比较	183
7.2.2	基本原理和技术	184
7.2.3	IP 传真的典型应用	186
7.2.4	3Com 公司的 IP 传真解决方案	187
7.3	因特网会议系统	188
7.3.1	NetMeeting 简介	189
7.3.2	基本操作	190
7.3.3	NetMeeting 音频/视频操作	192
7.3.4	聊天操作	193
7.3.5	白板操作	193

7.3.6 共享程序和桌面 193

7.4 基于 Web 的语音通信 195

7.4.1 Web 电话 195

7.4.2 Web 方式的交互式语音应答 195

7.4.3 基于 Web 的呼叫中心 196

7.5 流式媒体和音视频点播 198

7.5.1 流式媒体技术 198

7.5.2 流式媒体技术应用前景 200

7.5.3 音/视频点播软件简介 200

7.5.4 微软的 Windows Media 203

第 8 章 企业级数据与电话网的融合 208

8.1 “三网融合”的技术基础 208

8.1.1 “三网融合”的物理网的选择 209

8.1.2 IP 将作为未来三网融合的公共平台 209

8.1.3 网络带宽飞速增长 210

8.2 企业数据与电话网融合的基本技术 211

8.2.1 企业 IP 网络的结构 211

8.2.2 企业 IP 电话网的结构 214

8.2.3 企业 IP 网电话呼叫过程 216

8.2.4 完整的 IP 电话系统结构 216

8.3 企业数据与电话网融合的解决方案 219

8.3.1 采用 VPN 的融合方案 220

8.3.2 直接与因特网互连方案 221

8.4 运营商级 VoIP 解决方案 222

8.4.1 公司发展战略 223

8.4.2 运营商级 VoIP 解决方案 223

第 9 章 计算机电话集成 228

9.1 电话集成基础知识 228

9.1.1 设备类和媒体模式 229

9.1.2 拥有者和监控程序 229

9.1.3 呼叫状态 230

9.1.4 号码转换 230

9.1.5 服务提供者 230

9.2 电话服务 230

9.3 TAPI 消息 232

9.4 TAPI 提供的电话功能 232

9.5 TAPI 应用程序执行步骤 233

9.6 电话记录 234

9.7 TAPI 3.0 将会成为 CTI 标准的 API 234

 9.7.1 CTI 应用程序接口 234

 9.7.2 TAPI 3.0 的特性 235

 9.7.3 TAPI 3.0 的结构 236

 9.7.4 TAPI 3.0 编程 237

缩写词 239

参考文献 244

1

本章介绍了因特网业务的发展趋势，指出信息产业将成为 21 世纪的带头产业和支柱产业。同时介绍了因特网的基本特征，分析了因特网用于语音通信的必然趋势。介绍了因特网语音通信技术的一些典型应用，如基于 Web 方式的呼叫中心、IP 电话、VOD 等等。最后介绍了因特网语音通信所涉及的关键技术问题，如：标准化问题、语音压缩和因特网组网技术问题等。

1.1 因特网业务的发展趋势

经济学家康道拉恰夫的景气波动论认为：人类进入工业社会以来，便经历着景气波动，并把这种标志着经济发展情况的景气波动称之为“康道拉恰夫浪潮”，人类至今已经历过四次这种浪潮。第一次浪潮是由产业革命引起的，第二次浪潮是由铁路和钢铁业的发展而引起的，第三次是电力和化工的发展而引起的，第四次是汽车和电子的发展而引起的。这四次浪潮经历的时间是 1789-1995 年，前后大约经历了 200 年。从现在起推动经济回升的是第五次浪潮，它的原动力主要是因特网技术和多媒体技术的发展。其对应的时间是 1995~2055 年，2055 年将达到高峰。由于通信技术与计算机技术的结合，尤其是因特网技术和多媒体技术的结合，形成了一种新兴产业，即信息产业。信息产业将成为 21 世纪的带头产业和支柱产业。

早期的因特网主要是用于电子邮件、电子布告、远程登录、全球浏览、数据通信和文件传输，但是随着计算机网络技术的飞速发展和多媒体技术应用的日益深入和普及，基于因特网的多媒体通信已成为当前因特网发展的趋势之一。随着因特网的多媒体通信技术的发展，已经或将会产生许多类型的因特网多媒体通信业务，如：多媒体会议型业务（MM Conference Service）、多媒体会话型业务（MM Conversation Service）、多媒体分配型业务（MM Distribution Service）、多媒体检索型业务（MM Retrieval Service）、多媒体消息型业务（MM Message Service）和多媒体采集型业务（MM Collection Service）。实现上述业务的核心技术是采用因特网进行声、文、图等多媒体信息的传输技术，其中最关键的是音频和视频信号的实时传输。

虽然目前受因特网带宽等因素限制，基于因特网的实时视频信号的传输的应用还不是很普遍，但是在因特网上进行音频、特别是语音信号的实时传输已是相当普遍，技术也相当成熟，其中最为典型的应用是因特网电话。

因特网电话是因特网多媒体通信的一个典型业务，成为当前计算机网络技术和通信技术研究的热点，因此它也是因特网增长最快的业务。据国际数据公司统计，因特网电话的市场业务量已从 1995 年的 350 万元上升到 1999 年的 56000 万元，2000 年营业额为 30 亿元，并且全球有 15% 的用户使用 IP 电话，预计在 2005 年，使用 IP 电话的用户数将回增至 34%。五年的增长率为 149%，而传统电话的增长率为 15%。人们之所以对因特网电话抱着如此大的

兴趣和保持如此乐观的态度，有三个主要原因：(1) 因特网电话在很大程度上减少了长途(国际、国内)电话费用。(2) 因特网电话的使用标志着一种全新的功能更强的通信方式的产生，从而可以打破电话的垄断市场。因为与传统的传输媒体相比较，因特网电话不仅具有传统媒体所具有的一切表现形式和特点，而且传播信息容量大、不受时空限制，突破了传统地缘政治、地缘经济的概念，形成以传输信息为中心的跨国界、跨文化、跨语言的全新的传媒方式。因此，因特网电话作为现有电话的竞争对手在将来的发展中会形成很大的市场。(3) 符合未来“三网融一”(电话网、有线电视网、数据网)的发展方向。许多网络专家都指出，网络的发展趋势是“IP 一统天下”，“Every thing over IP”是网络发展的目标。据报道，全球电信业巨子 AT&T 的首席执行官 Michael Armstrong 最近说“我们要在全球把电话变成最普通的 IP 设备”，这或多或少地代表了对今后网络发展的一种观点。同时，很多专家也都指出，IP 电话目前所面临的问题是如何保证因特网语音传输的质量。

总之，发展基于因特网的实时语音，无论是从技术发展趋势，还是从市场前景和投资回报来分析，都是完全正确的。另外，促进因特网语音通信发展的更多因素是：

- 不断增加的语音业务和数据业务的综合，这是因特网语音通信发展的动力。
- IP 协议目前已成为事实上的公用协议，这是因特网语音通信发展的基础。
- 先进的分组语音压缩技术所产生的良好的成本效益，是因特网语音通信发展的前提。
- 企业内部网 (Intranet) 和企业外部外延网 (Extranet) 快速增长，这是因特网语音通信发展的平台。

1.2 因特网的特性

因特网在计算机和计算机之间传输数据的机理可以归纳为两点：第一点是采用自适应的路由和存储转发机制；第二点是传输的对象采用以分组为单位的数据包。存储转发和自适应的路由机制意味着数据包根据特定时刻的网络状况(如拥塞、链路故障等)在网络上可能采取不同的传输路由。这种路由机制可能会导致目的用户所收到的数据包和发送顺序不同，而且还会导致接收端数据包的到达速率不一样；数据包到达的时延也不相同。由于传输的单位是分组，这就意味着在接收端点还需进行数据包的重组。根据因特网的传输机理可以将其特性归纳如下：

1. 因特网是一个无连接的计算机网络系统

因特网无连接的概念和自适应路由概念密切相关，它就意味着因特网上的机器之间并没有紧密的联系，因此因特网不保存用户信息传输过程中的信息，而且从源主机到目的主机之间无需建立固定路径进行交换。实际上，IP 协议的寻路是无状态的，即不需要建立路由表来保留有关连接的信息，因为它本身就是无连接的。但是在电话网中，使用的是另外一种相反的体系结构，即主叫和被叫用户之间建立面向连接的固定通路。电话网中的这种方法对支持语音实时、固定时延的要求是必需的。然而，因特网是一个数据网，大多数数据业务并不要求实时传输。

2. 因特网是一个“尽力而为 (best effort)”的传输网络

所谓“尽力而为”是指因特网会尽量去传送业务流量，但如果发生了问题(由于噪声引

起的比特差错、路由器中的拥塞等)或无法找到目的主机,数据就会被丢弃。在大多数情况下,发送端主机上的 TCP 会重新发送已丢失或损坏的数据包。

3. 传输延迟的不确定性

由于传输路径的不确定,存储和转发时间的不确定,因此造成数据包往返时间(RTT)波动,数据包往返时间(RTT)是指发送一个分组到目的节点再从该节点收到响应所需要的时间。RTT 包括在两个方向上的传输时间和目的节点的处理时间。

虽然 RTT 值可能会有大的波动,但在因特网上 RTT 的典型值一般是 70~160ms。ITU-T G.114 协议规定电话业务的 RTT 值应小于或等于 300ms,因为较长的时延会使电话网中的通话双方感觉使用的是半双工电路。可见在正常情况下,因特网上 RTT 是能支持语音通信的,但是当网络负荷很重时,RTT 值可能超过 300ms。

4. 分组丢失的突发性

因特网特性中对语音和视频应用至关重要的一个因素是分组丢失。它涉及到两个方面,即分组丢失的频度和有多少连续的分组受影响。在数据业务中,分组丢失可以利用 TCP 的重传机制来补偿。但是, TCP 的重传机制不适合实时业务。所幸的是,随机的、少量的分组丢失所产生的影响对语音通信来说不像数据那么严重。

然而,大量的或连续的分组丢失问题在语音和视频应用中还是影响很大的,因为这会影响到接收端解码的结果,而且终端用户可以听出或看出这种变化。不过,如果语音分组丢失是随机的、不相关的,当前的语音编码器在分组丢失率为 10%的情况下,仍能恢复出高质量的语音信号。采用 G.732.1 语音编码标准,利用前面的语音分组来模拟丢失的分组中的音频信号的特性,从而补偿分组丢失。因特网中的数据丢失是突发的,在少量突发中会发生大量分组丢失。这一特征使得因特网支持语音通信业务在技术上变得更复杂了。利用前向纠错(FEC)方案可以减小分组丢失的影响,而且已经设计出了补偿突发丢失的方法。

5. 数据包的顺序

因为 TCP 协议可以对 TCP 数据段重新排序并以正确次序把数据提供给应用程序,所以在由 TCP 支持的数据应用中并不关心接收端数据包的到达顺序问题。但是,在语音和视频中并未使用 TCP 协议,因此对这些应用来说数据包的到达次序问题就显得很重要了。

从上述对因特网性能分析来看,因特网传送语音业务并不是最佳选择。但为什么仍然选择这样一种技术来支持语音业务呢?主要有三个方面的原因:

(1) 语音和数据业务的融合

首先,许多应用软件不可避免地要求 Web 服务器必须具有与用户进行数据、语音和视频图像交互的能力,显然这就要求语音和数据业务的融合。只进行带有静态图像的文本交互已不能满足用户的需要。

同时,语音和数据业务的融合可使带宽合并,而带宽合并使得数据信道的利用率得到提高。传统的电话技术,语音独占以时隙划分的信道(如信道组,数据业务单元 DSU 等),这不是支持数据应用的有效技术。在电话交谈中一般有相当长的静音期(即电话用户在交谈中停下来思考的时间,及轮流讲话中间的停顿时间等),在过去的电话通信机制中静音期间也占用带宽,而如果使用新的统计时分复用(STDM)的数据通信机制则可以更加有效地使用宝贵的带宽资源。STDM 方式使得只有在用户需要时才占用带宽资源,这样当这个用户不讲话时,空闲的带宽资源可以为其他用户所用。

一般来讲,交谈中有 50%的静音期(至少大多数交谈都是这样的),由此可以看出时分复用(TDM)方式是多么地浪费带宽。基于 TDM 方式构筑的电话网必须占用带宽来承载大量的静音期业务,而数据通信网就无需如此。语音中存在的 20%左右的冗余还可以通过压缩算法消除,传统的 TDM 方式却没有利用这一优势。利用先进的模拟-数字转换技术,一条高质量的语音信道只需运行在 4.8kbit/s~8kbit/s 带宽上,而目前使用的 TDM 电话信道都运行在 64kbit/s 带宽上。

由于带宽合并和语音压缩,将使长途通信的费用大幅降低,这也是利用因特网承载语音业务的主要原因。

(2) IP 协议的普遍存在和成熟

IP 及相关协议在用户和网络设备中已经大量存在。现在已经有许多人利用 PC 机来帮助他们完成电话呼叫。不久以后,基于计算机的电话就会变得很普遍,并成为现有电话系统的自然扩展。

同时,成熟的技术使 IP 电话现在就切实可行。三项关键技术的成熟将有力地推动信息技术革命。这就是:

- 通信信道容量不断增加;
- 计算机 CPU 处理能力不断增强;
- 具有人工智能的可重用即插即用软件代码技术的出现。

特别是 IPv6 的出现,使因特网的特性有很大的改善,随着传输语音等实时业务相关的协议也不断推出,例如 RTP 等,使得基于因特网的语音通信成为可能。

(3) 语音自身有一定错误容忍能力

语音传输对误码有很强的耐受力。如果偶尔有一个包损坏了,并不会严重影响语音的保真度。然而,数据业务就不能容忍误码。一个比特误码就会导致数据意义的改变。而且,语音分组可以容忍偶尔的丢失或损坏。如果网络时延过大,语音分组经过很长时间才到达接收端,这些语音分组就会因无用而被丢弃。当然,如果丢失的包只占总发送包数的不到 10%也不会严重影响声音的保真度。但数据分组就不能容忍这种包丢失或丢弃。

通过上述的分析,可以预见基于因特网的语音通信是未来因特网的一个主要业务,也是因特网的主要发展方向。但是就目前因特网的现状,在采用因特网传输语音时,有下列几方面因素是必须考虑的,即分组时延、带宽要求、计算量以及可变比特率。

分组时延描述数据包从发送方传到接收方所需的时间,分组时延包括两方面的内容:第一是指数据包从发送方到接收方需要多长时间;第二是指数据包到达接收方时延的变化,这种时延的变化叫做时延抖动。

带宽是影响语音质量的一个主要因素,也是因特网当前最难解决的问题。目前能做到一方面是提高因特网的固有带宽,另一方面是采用高效的语音压缩算法。另外,支持语音和视频传输所需带宽的计算除了要考虑表示语音信号所需要的比特外,还要考虑传输语音信号所需头部开销(协议控制信息)所占的比特。

计算量是指支持音频应用业务所涉及到的计算开销和复杂度。简而言之就是支持这一应用所需的 MIPS(百万条指令/秒)数及内存大小,也就是语音编解码器的复杂度和开销。因为在因特网上传输语音,都需要进行压缩,一般来说,压缩比较高的算法都比较复杂,所需的计算量都比较大。如何选择压缩算法是确定计算量的一个关键因素。

描述业务流量性质的一个有效方法是使用可变比特率 (VBR) 和恒定比特率 (CBR) 这两个概念。使用 VBR 机制的业务无需恒定连续的带宽分配。这些业务也叫突发业务, 即数据的发送和接收是按异步方式的 (随时都可能出现不发送或不接收的情形)。VBR 机制是最常见的数据通信方式。

从时间角度来看, 这种应用允许排队长度可变, 在收发方之间无需固定的定时关系。因此, 如果数据流从发送方发出, 然后缓存 (排队), 缓存时间可长可短, 这对接收方不会产生影响。使用 VBR 机制的典型应用有: 终端到终端的交互式会话、查询/响应, 客户机/服务器系统及大容量数据传输。需要强调的是, 虽然 VBR 的定时机制比较松散而且收发端是异步工作的, 但是大部分 VBR 应用还是需要某种类型的定时限制的。

语音打包被归为 VBR 业务, 这是因为分组语音是在基于 VBR 机制的数据网上而不是在基于 CBR 的电话网上传输的。因此, 这就产生了一个难题, 即在接收端接收数据网的突发性语音分组, 然后将这种 VBR 数据流平滑为 CBR 数据流, 这样才能对传统的 CBR 数据流进行数-模转换。与此相反, 采用 CBR 机制的业务需要恒定而连续 (或近似连续) 的带宽分配。这种业务是非突发的, 基于 CBR 的典型应用就是语音传输。这种应用需要有带宽保证而且在收发方之间要有恒定而连续的定时关系, 同时还需要收发方之间的时延是可预测的。

1.3 因特网语音通信的典型应用

从前面的数据可以看出, 因特网语音通信正在改变着产值数以万亿美元的全球通信业, 过几年, 至少有 70% 的语音业务将以分组形式通过因特网传输, 在如此高涨的因特网语音通信大潮的冲击下, 当今的通信公司必须重组其经营之道和承载语音的方法才能成为新生的“下一代通信公司”。

随着因特网实时业务的发展, 基于因特网语音通信的应用越来越多, 除了 IP 电话以外, 还有如下的一些应用:

小型办公室和家庭办公室的连接, 利用该连接可以通过因特网业务提供商和 IP 语音技术来获取对大型计算机和电话服务的透明接入。

企业内部互连网电话, 这种电话对于一个机构分布在全国各地, 特别是全球各地的企业来说是非常必要的, 它能节省很大一笔通信费用。

基于双音多频 (DTMF) 的安全机制: 许多地域分布特别广泛的金融系统经常使用双音多频作为远程输入身份证号码或者是密码等信息的安全监视手段, 使用因特网来传输这些信息比传统的电话线路要便宜得多。

另外, 电话会议系统、语音信箱、基于 Web 网的呼叫中心、文档共享、计算机电话集成技术等都是利用因特网来进行语音实时传输的典型应用。下面就详细介绍几种重要的应用。

1.3.1 计算机电话集成

SoundWare 公司的电话操作系统 (TOS) 的产生标志着计算机电话集成产业的开始, TOS 是与电话应用程序接口 (TAPI) 兼容的 Windows 程序, 它可以在 X86 或 Pentium 处理器上运行, 代替硬件来管理基于主机的信令处理和通信算法, 使得高级的 PC 电话无须昂贵的硬

件支持就能实现。TOS 还支持大多数音频芯片，它的广泛使用加快了语音数字化的发展。另外，TOS 还用于管理呼叫处理的各个方面，为信令处理和通道算法的开发者提供一个灵活开放的电话平台。

TAPI (Telephony Application Programming Interface) 是微软提供的计算机和电话网相联系的编程接口，使程序员可以利用这个接口通过电话线完成多种计算机复杂的通信工作。TAPI 能提供的功能主要有：自动拨号；以文件、传真、电子邮件的方式传送文件；访问 Internet 或其他形式信息服务、组织会议呼叫、使用主叫识别处理入呼叫、计算机间通过电话线的互相协作等等。而且 TAPI IMAP4——Microsoft 的 Telephony API 是被当今的 CTI (计算机电话系统集成)应用广泛使用的标准。特别是微软在 Windows 2000 中发布的 TAPI3.0，它能够支持传统的 PSTN 电话和 IP 电话两种电话，为用户提供了一个良好的开发环境。随着电话和呼叫控制在个人计算机中应用的越来越普遍，需要一个通用的电话接口来支持应用程序，使其能够访问任何一台计算机上的电话，通过一个统一的标准，一个呼叫的媒介和数据对应用程序也必须是可用的。随着 Windows 2000 的广泛应用，基于 TAPI 3.0 所开发的 Call Center 将以其优异的性能价格比和几乎无限的扩展空间，得到越来越多用户的青睐。

1.3.2 视频会议系统

国际电联已经为电视会议产业指定了一系列的标准，使得不同的厂商的产品能够在兼容。这些标准系统包括用于传统电话系统的 H.324、用于 ISDN 的 H.320 和用于因特网的 H.323。由于因特网的业务量的随机性，电视会议所需的带宽无法保证，但是随着视频压缩、因特网的组网技术的发展，基于因特网的电视会议系统的普遍应用为时不远。特别是基于局域网或广域网的电视电话系统的应用已经相当普遍了，有关此类产品已是层出不穷。

1.3.3 基于 Web 网的呼叫中心

基于 Web 网的呼叫中心是因特网语音通信飞速发展的一个应用，呼叫中心和因特网的结合，能实现因特网的语音技术、交互式的网页浏览技术和现有呼叫中心的交换机技术的有机结合，极大地丰富了呼叫中心的功能，而且使得因特网成为真正的呼叫中心的扩展。

目前，基于 Web 网的呼叫中心的产品和解决方案有许多种，而且价格便宜。其中比较有影响的公司有朗讯科技、NetSpeak 等。这些产品的普遍采用的过程为：当用户需要检索信息时，可以通过网页上的按钮采用因特网语音方式或基于文本聊天工具将用户和信息服务器连接起来，然后按需要呼叫相关信息。例如，如果在网页设置一个所谓的因特网语音按钮，当用户需要访问一个企业网站时，按一下该按钮，就能呼叫该企业下订单、或者是请求服务、或者是检索信息。客户不用拨号码，仅采用这个按钮会话就能实现，而且不用花一分钱。

1.3.4 因特网媒体流技术

随着因特网应用的普及，在网络上传输的资料不仅仅限于文字和图形。声音和影视的传播给广大网民带来了新的享受。在此情况下世界各地传统影视的媒体纷纷加入到因特网领域中，使自身的传播方式得到了扩充。

面对因特网有限的带宽和拥挤的拨号网络，实现窄带网络的视、音频传输最好的解决方

案就是流式媒体的传输方式。采用这种流式媒体技术可以提供视频点播、音频点播、MTV 播放、音乐会播放、音乐欣赏、多媒体广告发布等服务。如条件成熟还可提供电视节目的网络直播等功能。

随着因特网的发展，流式媒体（Streaming Media）越来越普及，流式媒体是通过网络传输的音频、视频或多媒体文件，流式媒体在播放前并不下载整个文件，流式媒体的数据流随时传送随时播放，只是在开始时有一些延迟。当流式媒体文件传输到用户的计算机时，在播放之前该文件的部分内容已存入内存。随着因特网应用的日益深入，基于因特网的流式媒体技术有着广泛的应用前景，可以用于娱乐、培训和在线教育等方面：

1. 带图片的广播（Illustrated Audio）

可以把 PowerPoint 讲座录制到 CD 或 Web 站点，用户通过 IE，就可以看到一张张 PPT 在讲解员讲解的同时会自动翻转。工作方式是：只要提供讲解员的声音（磁带或电子格式均可）和 PPT 文件，Windows Media 提供了 Encoder（压缩）工具可以在声音中插入 Marker，然后在声音播放的过程中，这些 Marker 就会翻转 PPT 图片，使声音和 PPT 图片保持同步。

2. 流视频播出(Streaming Video)

用摄像机或投影仪获得视频信号后，就可以通过 Web 站点进行基于因特网的现场直播；或者保存为 .nsf 文件后，以按需播放。需要在一台较高配置的 PC 机上安装上普通视频采集卡和声卡，然后分别通过视频采集卡、声卡输入视频和声音信号就可以用实时 Encoder（压缩）工具来直播或录制成流媒体。

3. 远程教学（Remote seminar）

教学者事先在因特网/Intranet 上发出通知，听众在讲座开始前访问某个 URL 地址，当讲座开始时，听众可以看到演讲者的图像并能听到他的声音，还有 PPT 图片。在演讲者翻转 PPT 的时候，用户端的 PPT 图片也自动翻转了。整个讲座也可以记录下来，以后按需播放。需要教学者事先用 Multicast 把 PowerPoint 文件传给 IIS 服务器，当听众等待讲座开始的时候，PPT 图片下载到用户的浏览器 Cache 中。用摄像机录制演讲者的图像和声音，通过实时 Encoder 直播出来。直播中包含了 PPT 翻转的命令，可以使听众的浏览器同步地翻转 PPT 图片。

4. 提供收费电视（Pay by view）

当没有许可证的用户需要看内容供应商直播或按需点播的内容时，则 Windows Media Player 会带他到内容供应商的网站，来（付费）申请许可。Windows Media 4.0 中提供的 Digital Rights Management 功能可以加密内容供应商的内容，确保收费后信息才会播发。

除了上述应用外，一个最为典型的应用为音视频的点播（AoD 或 VoD），目前，由于受带宽的限制，在因特网上实现视频点播（Video on Demand）还不是很成熟，但实现音频点播（Audio on Demand）已是很普遍，而且基于 IP 协议的 VoD 在局域网传输平台上是完全可以实现的。

目前比较流行的视音频网上直播技术产品有：

- RealNetworks 公司的 Real System；
- 微软公司的 Windows Media。

1.3.5 IP 电话

因特网语音通信最典型、也是最成熟最广泛的应用莫过于 IP 电话了。早在 1995 年初, Vocaltec 公司推出了一种叫作 Internet Phone 的客户软件, 虽然当时该公司还没有提出在 IP 上传输话音的概念, 但这确实是 IP 电话第一次成功的商业化和市场化。自从 IP 电话出现至今, IP 电话的发展经历了三个阶段: 从最初的 PC 到 PC 的简单通话方式, 到基于插卡模式的网关出现; 现在由于各个电信设备商的介入, IP 电话已经走到了第三个阶段, 这就是电信级 IP 电话网络。目前, IP 电话已经有成熟的产品出现, 它和打普通电话一样方便, 而费用只有普通电话的几分之一到几十分之一。

1. IP 电话的优点

(1) 更有效地利用带宽和设备

IP Phone 并不为每个呼叫都分配高达 64kbit/s 的一条通路, 它可以在多个逻辑连接中动态地共享带宽, 这在处理突发业务时是极为理想的。同时, IP 电话还可以承载现有电话交换的业务量。

(2) 降低电话网传送的费用

传统电话网要求大量的设备来建立和分离高带宽传送电路, 这必须在 64kbit/s (DS0) 的级别上才能进行处理。IP 电话在最大限度的可用传送链路上复用话音业务量, 这意味着可以极大地节省设备和运营费用。

(3) 综合的话音和数据网络费用

Internet 业务量每 6~9 个月翻一番, 按照这个速度, 目前重叠在话音网上的数据网络将很快承载大量的业务。因此, 话音网必须尽快向分组交换结构转变, 创建具有公共交换和传送系统的单一综合通信网络。

(4) 来自新业务的收入不断增长

IP Phone 实现了新的综合业务——宽带质量的音频、统一消息、实时话音数据的综合、视频电话和可管理的数据/话音/视频业务。这些业务都将会增值, 它们使服务提供商与众不同, 同时也改变了其市场定位。

(5) 统一通信

与开放式分组电话 (OPT) 结构相结合, IP 电话可以实现统一通信。它采用 IP 基础设施, 将以前独立的通信方法统一起来, 包括传真、语音函件、电子函件、有线电话、蜂窝电话和 Web 等, 向用户提供接入消息和初始化实时通信的通用方法, 而且用户不必考虑时间、地点或设备。

2. IP 电话的工作原理

当电话用户拨号到因特网电话经营者的访问节点后, IP 电话网关便作出响应, 于是便输入对方电话号码。IP 电话网关以输入的电话号码为依据, 向“关守 (gatekeeper)”查询作为分组发送目的地的 IP 电话网关的 IP 地址。知道了 IP 地址后, IP 电话网关便完成了它同对方的 IP 电话网关之间设定 IP 连接而需要的手续。H.323 建议中的“H.225.0”便规定了上述操作过程。H.225.0 是在名为“Q.931”这一 ISDN 使用的呼叫控制协议的子集中, 加入 H.323 独自的功能。它规定了 IP 电话网关等 H.323 终端与其他 H.323 终端间交换电话号码信息和线路设定信息时的操作过程, 以及与关守交换信息时所需要的操作过程。然后, 与对方的 IP 电