摘要

本报告描述了下一代网络的体系架构，阐明了下一代网络承载与业务分离的设计思想；系统地介绍了 IPTV 系统和业务的基本概念，研究了 IPTV 业务系统与网络体系架构，提出了 IPTV 业务提供的技术需求；研究了 IPTV 业务保障问题，提出了从承载网络、内容传送网络和业务体验三个方面保障用户业务体验的架构；研究了 IPTV 业务提供的接纳控制方案，结合公司的产品现状提出了一个接纳控制机制的实施方案；重点研究了基于 P2P 技术的 IPTV 业务提供关键技术：分析了 P2P 技术在 IPTV 业务提供中的作用，研究了基于 P2P 技术的 IPTV 系统的基本架构，提出了一种新的用于 P2P 流媒体系统的内容调度算法。我们认为，用户业务体验保障对 IPTV 业务提供成功的关键，接纳控制机制是保障用户业务体验质量的基本手段；由于 P2P 技术具有良好可扩展性和经济性，P2P 可以直接应用到 IPTV 业务提供系统设计或作为其它技术的补充。

关键词：IPTV 业务提供 P2P 业务保障 接纳控制 下一代网络

Abstract

This report describes the architecture of next generation networks (NGN), emphasizing the idea of separating transport and service in NGN; Present the basic concepts of IPTV system and services in some detail, and summarize the technical requirements of IPTV service providing based on the study of IPTV system and network architecture; Investigate issues of IPTV service assurance and propose a framework of IPTV service assuance from transport networks, content delivery networks and user service experience; Study the admission control mechanism in IPTV service providing, and design a practical admission control solution with products available in Alcatel-Lucent. Another part in this report is about the P2P based IPTV service providing. The value of P2P technique for IPTV is analysised, and the architecture of P2P based IPTV system is studied and a noval content scheduling algorithm for P2P media streaming is presented. In our opinions, user service experience is critical to the success of IPTV service providing and admission control can take a key role in guaranteeing users' service experience; P2P technique can be employed for IPTV system design directly or as compensation due to its good scalability and low cost.

Key Words: IPTV; Service Providing; P2P; Service Assurance; Admission Control; Next-Generation Networks

第1章 引言

电信行业的发展正面临严峻的挑战。传统的电信业以话音业务为主，由于其垄断的市场地位带来了丰厚的利润。Internet 的出现改变了这种垄断局面，VoIP、即时通信等作为替代技术和应用以其低使用成本吸引了一部分用户，更为严重的是大大压缩话音服务的利润空间。除了 Internet 带来的影响，话音业务向移动转移的趋势对于固网运营商而言更是雪上加霜。ARPU 值不断下降，用户离网率高是全世界的电信运营商都需要解决的问题。为了摆脱当前的困境，电信运营商需要进一步降低运营成本，同时开发新业务，寻找新的业务增长点。建设下一代网络、提供视频业务并加强多业务绑定（bundling）是目前公认的有效办法之一。

下一代网络（Next Generation Networks, NGN）是基于分组技术的多业务网络，能够提供包括传统电信业务在内的多种业务，从而有效地降低过去为每种业务建立一个专用网络的建设和运维成本。现阶段，以话音业务为主固网电信运营商正在经历用户 ARPU 值降低、赢利能力下降的困扰。为了增加运营收入，需要提供新的业务通过业务绑定提高用户忠诚度。IPTV 正是被寄予厚望的 NGN 业务。

提供融合的家庭业务，即在电视机、台式计算机、电话和移动设备等各种终端上提供交互式业务，代表了电信运营商的未来发展方向。IPTV 是当前最热门技术领域之一，主要的电信服务提供商都在努力推出自己的 IPTV 业务。简单而言，IPTV 是指在通信网络基础设施上使用 IP 协议提供数字电视业务的系统。IPTV 的优势包括传统电视技术所不具备的双向能力，它使业务提供商能够提供用户所需的视频广播和点播（Video on Demand, VoD）及其它增值服务。广播业务可以节约带宽以提供更多的内容，点播业务则可以满足用户的互动需求。

IPTV 产业广阔的发展前景和巨大的市场空间正是电信运营商所期待的新业务。国外的电信运营商大多以发展 IPTV 入手，提供语音视频数据捆绑的三重播放(Triple Play)业务，促进 ARPU 的增长。鉴于国外的经验，中国的两大固网电信运营商都在积极开展 IPTV 业务的网络试验，力求摸索出理想的网络架构和商业模型。正是电信运营商对 IPTV 业务的巨大需求，才推动了 IPTV 的商业发展进程。

随着中国市场经济的发展，内容提供商同样面临产品市场化的问题。传统单一的内容发行模式，容易造成内容的滞销。一般内容产品的生产前期投入巨大，如果没有广泛的发售渠道，很难形成良性发展的产业结构。内容提供商为了自身发展一直在努力寻找更多的发行渠道，延长内容产品的生命周期。IPTV 业务的出现为内容提供商提供了一个新的展现和利用内容的平台。因此，无论是电信运营商还是内容提供商，都可以从 IPTV 业务提供中获益。

对全球的电信运营商而言，IPTV 都是个全新的业务，不仅没有相关的运营经验可以借鉴，由于需要建设一个新的业务系统，还面临着很大的风险。国内的电信运营商也不例外，甚至情况更加严重。首先，在政策层面，电信和广电分业经营的基本政策没有改变，短期内也不大可能有大的变化。另外，国内的内容监管比较严格。上述两个方面增加了内容与通信整合的困难。其次，在市场层面，由于产业链环境高度复杂，商业模式融合比较

困难。广电领域和 Internet 的低资费传统使 IPTV 的潜在用户对资费很敏感，当前较高的机顶盒价格也限制了业务发展。具有差异化的业务和运营模式需要开发和检验。缺乏吸引新兴用户的节目内容也是当前 IPTV 发展的障碍。即便是政策和商业的问题都得到了妥善解决，IPTV 业务提供还需要解决技术方面的问题。

和传统的电信业务相比，IPTV 有很大的区别，其中最重要的区别是引入了视频业务。尽管视频业务并不是一个新业务，但在一直由广电行业通过专网提供和运营。从技术角度来看，视频业务有以下几个特点：1）带宽要求高。标准清晰度视频信号需要的 2M 带宽，而在专网上提供每路话音业务需要 64K 带宽，通过 VoIP 提供话音所需带宽更少；2）具有实时性要求。视频业务不仅要提供大量的数字信号，还要求信号到达接收端时要满足严格的时间要求，否则将失去作用；3）支持交互性。交互性意味着广播方式不能满足需要，从而大大提高了网络容量要求。交互性还要求快速响应，对传输时延也有严格的要求。尽管 Internet 上已经存在一些简单的视频业务，但由于商业模式的不同，Internet 上的视频业务只提供尽力而为的质量保证。电信行业的目标是提供大规模高质量的视频业务，这为 IPTV 业务提供带来了很大的技术挑战。

IPTV 业务提供至少需要解决两个层次的问题。首先建设和改造承载网络，提供足够的网络容量。当前，宽带网络设备已经存在，但网络建设和维护的经验相对缺乏。相对于需求，容量的大小总是相对的，增加容量意味着增加建设成本；如何加强管理，合理地利用网络容量是降低成本的重要手段。其次，也是更加重要的方面，建设 IPTV 业务系统，为用户提供良好的业务体验。带宽只是 IPTV 业务提供最基本的要求，实现 IPTV 业务提供还涉及到内容的传送管理以及业务运营管理系统。用户的业务体验是决定业务成败的关键，IPTV 的建设成本很高，运营失败可能对电信运营商带来致命性的打击，因此需要充分发挥各种资源的作用保障用户的业务使用体验。

1.1 报告的内容安排

本报告针对在下一代网络中提供 IPTV 业务存在的问题，研究了 IPTV 业务提供相关的关键技术，特别是基于 P2P 技术的 IPTV 业务提供关键技术。报告的研究内容安排如下：

第 2 章描述了下一代网络的体系架构，阐明了下一代网络承载与业务分离的设计思想。第 3 章较为系统地介绍了 IPTV 系统和业务的基本概念，研究了 IPTV 业务系统与网络体系架构，在此基础上提出了 IPTV 业务提供的技术需求。第 4 章研究了 IPTV 业务保障问题，提出了从个承载网络、内容传送网络和业务体验三个方面保障用户业务体验的架构，为后续的研究规划了方向。第 5 章研究了 IPTV 业务提供的接纳控制方案。接纳控制是避免网络拥塞的重要手段，本章结合公司的产品现状提出了一个接纳控制机制的实施方案，对 IPTV 网络建设有一定的参考价值。第 6 章和第 7 章重点研究了基于 P2P 技术的 IPTV 业务提供关键技术。第 6 章比较了 IPTV 业务的提供方式，分析了 P2P 技术在 IPTV 业务提供中的价值，然后研究了基于 P2P 技术的 IPTV 系统的基本架构，最后针对运营商对 P2P 流量的顾虑提出了 P2P 流量管理与控制的基本策略。第 7 章研究了基于 P2P 的 IPTV 系统的内容调度，提出一种新的用于 P2P 流媒体系统的内容调度算法，并通过仿真实验进行了验证与分析。第 8 章对报告进行了简单的总结，指出了进一步研究的方向。

第 2 章　下一代网络体系架构概述

第一代分组网络的设计迄今已有数十年的历史。NGN 中新的应用和部署需求带来了一些在第一代分组网络设计时难以考虑到的新需求。近年来，ITU-T 一直努力致力于 NGN 的标准化工作。ITU-T 推荐标准 Y.2001[1]明确了 NGN 中必须考虑的一些特征。此外，ITU-T 推荐标准 Y.2011[2]提供了一个体系结构基础以满足这些特征的实现要求。

一般认为，传统电信服务与 NGN 之间的主要区别在于从应用特定的垂直集成网络迁移至可以承载任意和所有服务的单一网络。对于电话业务，这种迁移包括了从电路交换基础设施转换到分组交换基础设施。当前 NGN 研究的目标是确保下一代基于 IP 的网络能够满足通常部署于公用电话网的服务标准，不仅是基本的电话服务，还包括现在和将来的各种可能的多媒体应用。

鉴于 Internet 的通用特性和现存的相关基础设施，包括地址规划、地址分配以及域名解析系统等，以及诸如电子邮件、文件传输（FTP）和 WWW 应用，基于 IP 的通信系统将构成下一代网络（Next-Generation Network, NGN）的基础。然而，从商业角度来看，由于电信与 Internet 的商业模型有很大的不同，对通信网络的技术需求也有一定的差异。因此，可以认为 NGN 是经过增强的满足电信业务需要的 IP 网络 。另一方面，随着竞争环境的变化，用户对业务多样性的需求更加迫切，网络在电信业中的重要性逐渐下降。NGN 不仅关心承载网络技术，还需要考虑如何更加高效地开发和部署新的业务。

2.1 ITU-T 推荐标准 Y.2001：NGN 概述

Y.2011 的主要目的是给 NGN 下一个通用的定义，如下文：

下一代网络（NGN）：能够支持电信服务**分组网络**，该网络能够利用各种**具有 QoS 保证的宽带传送**技术，其中与服务相关的功能**独立**于底层传输相关的技术。它使用户可以根据自己的选择自由地接入或访问网络、具有竞争关系的服务提供商和服务。它支持**一般移动性**，能够对用户提供一致的无所不在的服务。

上述定义中明确地表明服务与承载的分离原则，并提出在 IP 承载网络中增加 QoS 能力。第二句包含了创新的服务提供能力，允许服务提供商自由地创建、用户自由地选择服务。最后一句包含了在移动与固网之间以及适当条件下固网之间的移动扩展。

Y.2001 还进一步定义了 NGN 的基本特征：

- 分组传送
- 控制功能从承载、呼叫/会话、应用/业务中分离
- 服务提供与网络传送分离，提供开放接口
- 利用各基本的服务组成模块，提供广泛的服务和应用(包括实时、流式、非实时及多媒体服务)
- 具有端到端 QoS（Quality of Service）的宽带传送能力
- 通过开放的接口规范与传统网络实现互通
- 具有一般移动性

- 允许用户自由地接入不同服务提供商
- 支持多样化的身份识别机制
- 用户对同一种服务具有统一的服务特征感受
- 融合固定与移动服务
- 服务相关功能独立于底层的传送技术
- 支持各种最后一公里技术
- 适应各种监管需求，比如紧急通信，安全，隐私和合法监听等

2.2 ITU-T 推荐标准 Y.2011：NGN 一般原则与参考模型

Y.2011 的主要目的是为开发 NGN 服务功能模型提供基础。首先，它指出 NGN 分层系统和由 ITU-T 推荐标准 X.200 定义的 7 层开放系统互连基本参考模型（OSI BRM）的潜在区别。例如，考虑编号和 OSI BRM 7 层系统的特征带来的困难。在 NGN 系统中，可能遇到下面列出的部分或全部情形：

- 层数可能不是 7 层；
- 个别层的功能可能与 OSI BRM 不相互对应；
- 某些在 OSI BRM 中被确定或禁止的情形可能不再适用；
- 涉及到的协议可能不是 OSI 协议（显著的例子是 IP）；
- OSI BRM 的兼容性需求可能不再适用。

Y.2011 的附录详细说明了 OSI BRM 中适用于 NGN 以及不适用于 NGN 的条款。

服务需要由功能来构建，因而服务和功能之间是相关的。可以方便地将全部 NGN 功能分为两个不同的组或平面：一个包含全部控制功能，另一个包含全部管理功能。将同类型的功能分组便于定义同一组内的功能相互关系和信息流。

在功能分组的基础上，Y.2011 进一步考虑了各功能部分的系统实现。特别的，它提出了如图 2.1 所示的高层模型，图中显示了系统开发时各种功能的相互联系。图中的功能实体可以进一步细分为子组以便于分组实现和分布式系统描述。

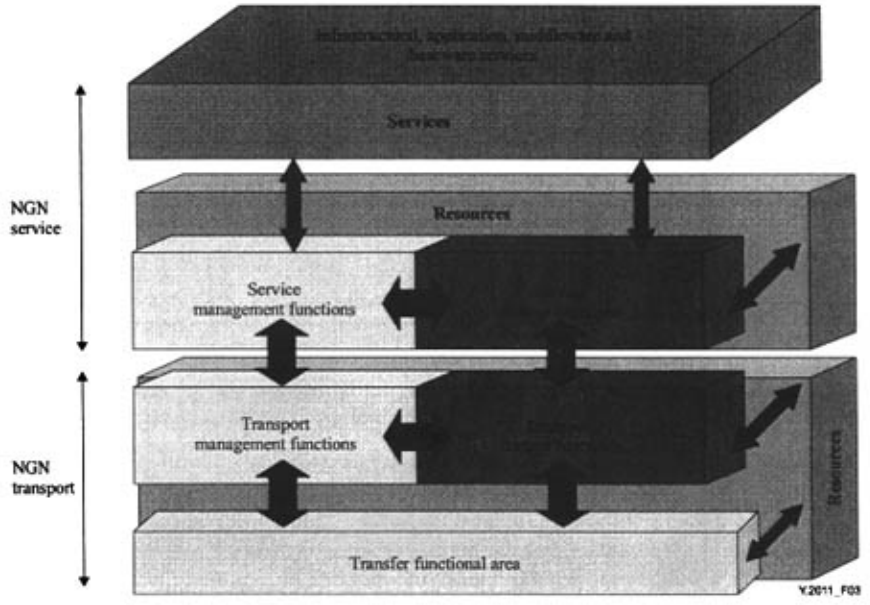

图 2.1 NGN一般功能模型 [Y.2011]

2.3 NGN 体系结构

本节描述 ITU FGNGN（Focus Group for Next-Generation Networks）讨论定义的 NGN 体系结构。由于 NGN 的标准化工作还在进行中，根据进一步的研究结果，最终的描述可能有所变化。

NGN 服务包括基于会话的服务和非会话服务；其中会话服务包括 IP 电话、视频会议、视频聊天等，非会话服务则有视频流化和广播等。此外，NGN 支持 PSTN/ISDN 替换（ITU-T 称之为 PSTN 仿真(emulation)）。

图 2.2 显示了 NGN 体系结构的概况[3]。根据 Y.2011 定义的框架，NGN 的功能被分为服务层和传送层。

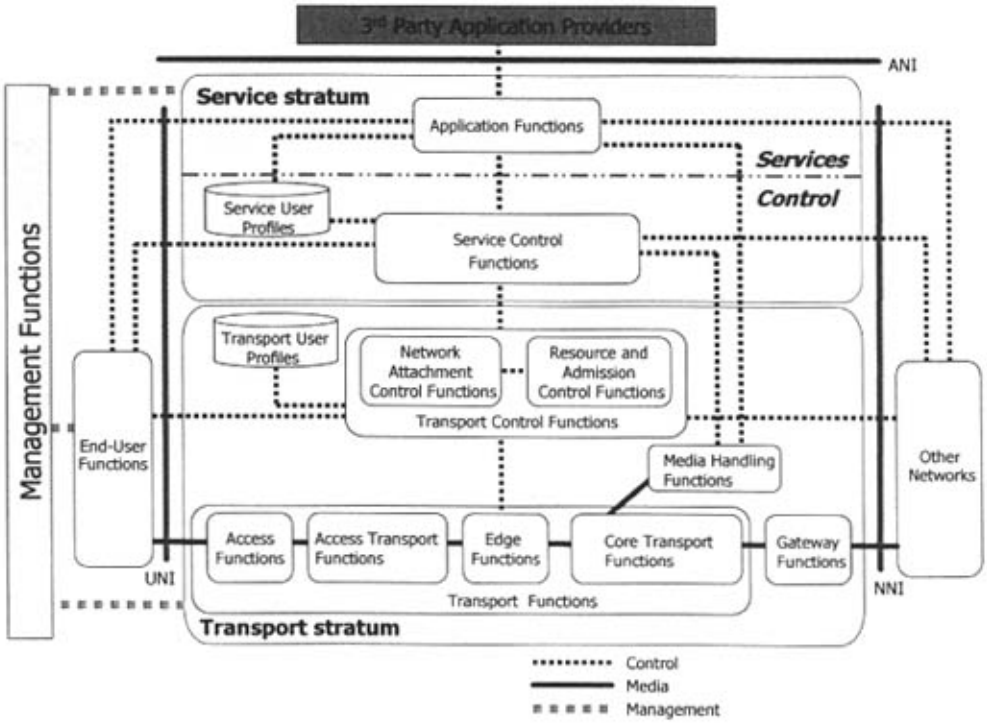

图2.2 NGN 体系结构概述

端用户的功能通过用户网络接口（UNI）连接到 NGN 网络，网络之间通过网络网络接口（NNI）实现互连。清晰地区分 UNI 和 NNI 对于使用各种现成设备的同时保持 NGN 环境中的商业边界和分割点具有重要意义。对于第三方应用提供者，应用网络接口（ANI）形成一个边界。

2.3.1 传送层功能

传送层为 NGN 各组成部分和物理上分离的组成部分提供连接。IP 被认为最有希望成为 NGN 的承载技术。这样，传送层将为位于 NGN 之外的用户终端和位于 NGN 之内的控制与支持服务器提供 IP 连接。传送层需求提供端到端的 QoS 保证能力，这是 NGN 需要的技术特征。传送层分为接入网与核心网，两者相互连接。

- **接入功能**

接入功能管理端用户访问网络。接入功能与接入技术相关，如 W-CDMA 和数字用户线（xDSL）。接入网包括和各种接入技术相关的功能，如 cable，DSL 技术，无线技术，以太网技术和光接入技术等。

- **接入传送功能**

负责在接入网中传送信息。接入传送同样需要 QoS 控制机制直接处理用户流量，包括缓存管理，队列管理，分组调度与过滤，流量分类、标记、监管及整形。

- **网络边缘功能**

用于在接入流量汇入核心网络时进行适当的流量处理。

- **核心传送功能**

负责确保信息穿过核心网络。根据与传送控制功能交互的结果，核心传送提供差异化的传输质量。核心传送也提供直接处理用户流量的 QoS 机制，包括缓存管理，队列管理，分组调度与过滤，流量分类、标记、监管及整形，连接控制和防火墙等。

- **网络附属控制功能**

负责接入层面的注册和接入 NGN 服务时的终端用户初始化。提供网络层面的身份识别与认证，管理接入网络的 IP 地址空间以及接入会话的认证。该功能向端用户通告 NGN 服务的接触点和应用功能，即网络附属控制功能辅助用户设备进行注册并启动服务。

- **资源与接纳控制功能（RACF）**

RACF 提供接纳控制与网络控制功能，包括控制网络地址与端口转换（NAPT）和区分服务域码点。接纳控制涉及到通过网络附属控制功能根据用户特征数据（User Profile）检查认证。该功能还包括结合运营相关的策略规则和资源可用性进行基于用户数据的认证。检查资源可用性意味着接纳控制功能验证在当前剩余资源的条件下，新的资源请求是否可以被满足。RACF 与传送功能交互控制一个或多个传送层的功能，包括分组过滤，流分类，标记与监管，带宽预留与分配，NAPT， IP 地址防伪造（anti-spoofing），NAPT/FW（防火墙）穿越和用户监测。

- **传送用户描述功能**

该功能模块在传送层将用户和其它控制数据编辑成单独的用户特征数据。该功能可以被规范和实现成一组协同工作的数据库，并且可能在 NGN 的任何一部分实现。

- **网关功能**

网关功能提供与其它网络互通的能力，包括各种现存的网络，如 PSTN/ISDN 以及 Internet。这些功能还支持与不同管理者的 NGN 相通。连接其它网络的 NNI 接口对控制层和传送层同样适用。控制与传送层可能直接进行交互也可能通过传送控制功能实现。

- **媒体处理功能**

媒体处理系列功能为满足服务提供需要进行媒体资源处理，如产生音调（tone）信号，编码转换及会议桥接等。

2.3.2 服务层功能

提供基于会话的非会话的服务，包括预订/通告存在信息和即时消息传递方法。服务层功能还提供所有和 PSTN/ISDN 服务相关的网络功能及与旧式用户设备相通的能力和接口。

- **服务与控制功能**

该功能包括会话控制功能，服务层的认证与授权功能，还包括专用媒体资源的控制功能。

- **服务用户描述功能**

该功能模块在服务层将用户和其它控制数据编辑成单独的用户特征数据。该功能可以被规范和实现成一组协同工作的数据库，并且可能在 NGN 的任何一部分实现。

- **应用功能**

NGN 支持开放 API，这样第三方服务提供商可以利用 NGN 的能力为 NGN 用户创建增强的服务。所有应用功能（可信的和非可信的）和第三方服务提供商通过服务层的服务器或网关获取 NGN 服务层的能力和资源。

2.3.3 管理功能

网管支持是 NGN 运营的基本要求。管理功能使 NGN 运营商管理网络并提供具有可预期质量、安全和可靠性的 NGN 服务。

这些功能在各功能实体间（FE）以分布的方式分配。该功能与网元管理、网络管理及服务管理功能实体相互作用。

管理功能包括计费和账单功能。这些功能在 NGN 中相互作用以在 NGN 中获取计帐信息，为 NGN 运营商提供资源使用数据从而对用户进行合理计费与收费。计费与帐单功能收集相关数据，支持后处理（离线计费）和与预付费服务的近似实时交互（在线计费）。

2.3.4 端用户功能

如图 2 所示，面向端用户的接口同时是物理和功能（控制）接口。对于可能接入 NGN 的各种用户接口和网络没有做任何假设。NGN 支持所有的客户设备种类，从单线的老式电话到复杂的企业网络。端用户设备既可以是固定的也可以是移动的。

2.4 小结

本章介绍了 ITU-T 推荐标准中规范的下一代网络（NGN）的基本原则和参考模型。文中还检查了 NGN 功能体系结构的最新概念，它们正在成为 ITU-T 的推荐标准。和传统的电信网络相比，NGN 最大的特点是实现了承载与业务分离，在统一的承载网络上可以支持多种业务。当前，IP 技术是承载网基本技术，基于 IP 承载网络实现具有三重播放能力的 IPTV 业务是 NGN 业务发展的重点。

参考文献

[1] ITU-T Rec. Y.2001, General Overview of NGN.
[2] ITU-T Rec. Y.2011, General Principles and General Reference Model for Next Generation Network.
[3] K. Knightson, N. Morita and T. Towle, NGN architecture: Generic Principles, Functional Architecture, and Implementation, IEEE Comm. Magazine, 2005. Vol. 43(10): 49- 56

第 3 章　IPTV 业务提供的技术需求分析

3.1 IPTV 业务的定义

根据中国通信标准化委员会的定义，IPTV 是指在 IP 网络上传送包含视频、音频和数据，提供 QoS/QoE、安全、交互性和可靠性的可管理的多媒体业务，其表现形式可以是电视、文本或图形等 [1]。

IPTV 业务具有以下特征：

- IPTV业务用于向用户提供有QoS保障和安全保障机制的多媒体内容传输业务，这些内容在最终用户的终端上播放或显示。
- IPTV业务向用户提供流式服务，主要用于传输音视频媒体流（当然也可以携带传输文本、数据、语音等），可以同时为众多用户提供业务支持。
- IPTV业务能够向用户提供检索能力和交互能力，交互表现形式为用户能够对IPTV业务所提供的音视频节目进行主观控制。
- IPTV的承载网络是IP网络，IPTV业务网络具备内容缓存、分发和存储功能。
- 向用户呈现IPTV业务的终端包括电视机＋机顶盒、PC电脑、PDA或手机等终端设备。

根据 IPTV 的业务提供情况来看，IPTV 的业务参与者基本构成如图 3.1 所示。

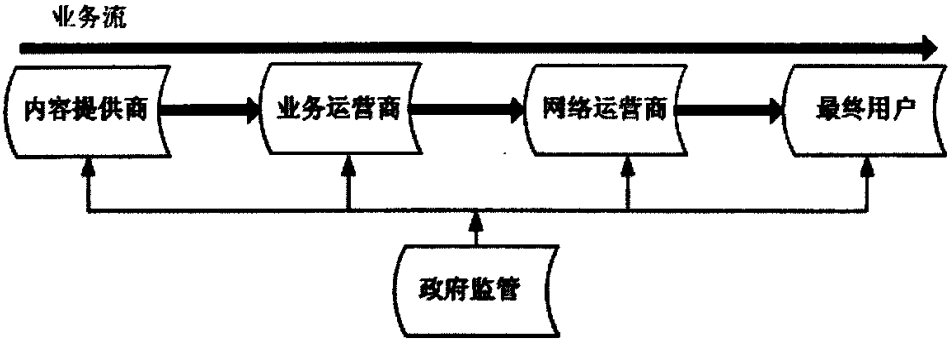

图 3.1 IPTV 的业务参与者

IPTV 业务参与者包括内容提供商、业务运营商、网络运营商和最终用户，其中业务流是从内容提供商向下游最终用户发送的。

最终用户通过 STB+TV、PC、手机或其他终端系统(设备)接入网络，获得 IPTV 服务。

网络运营商主要负责利用自己的网络将 IPTV 业务运营商所提供的内容传送到最终用户并获取收益。网络运营商包括基础网络运营商、接入网络运营商。基础网络运营商包括电信、广电、卫星等，接入网络运营商包括有线、无线等宽带网络接入服务的提供商。

业务运营商负责内容的集成(管理)和运营，并在自己的 IPTV 平台上将各种内容集成在一起，利用网络运营商的网络为用户提供内容和应用。为了促进业务推广，业务运营商通常建设自己的接入网络。

内容提供商负责内容的制作，并将所制作的内容版权提供给业务运营商。在 IPTV 业务产业链中，内容提供商是整个产业链的源头。这里的内容提供商泛指向业务运营商提供内容的各种机构或组织，这些机构或组织在某种程度上也进行内容运营，但不直接面对最终用户。

政府监管部门主要对 IPTV 业务在产生、集成、传输和消费等各个环节进行监控，控制网上不良音视频信息的传播以及防止对正常节目的恶意破坏和篡改。

3.2 IPTV 的业务类型

目前，大部分 IPTV 业务解决方案可以提供基于 IP 网络的直播电视、时移电视以及视频点播等基本业务。在基本业务之外还可以提供诸如 Internet 浏览、在线游戏、视频短信、远程教学、可视电话等增值业务。IPTV 系统通常提供以下基本业务：

- **直播电视**

直播电视类似于广播电视、卫星电视和有线电视所提供的服务，这是宽带服务提供商为与传统电视运营商进行竞争的一种基础服务。直播电视通过组播方式实现直播功能。

- **时移电视**

时移电视能够让用户体验到每天实时的电视节目，或是今天可以看到昨天的电视节目。时移电视是基于网络的个人存储技术的应用。时移电视功能将用户从传统的节目时刻表中解放出来，能够让用户在收看节目的同时，实现对节目的暂停、后退操作，并能够快进到当前直播电视正在播放的时刻。

- **视频点播**

IPTV 的视频点播是真正意义上的 VOD 服务，它能够让用户在任何时间任何地点观看系统可提供的任何内容。通过简单易用的遥控器，让用户有了充分支配自己观看时间的权利。这种新的视频服务方式让传统的节目播出时刻表失去意义，使用户在想看的时候立即得到视频服务的乐趣。

- **准视频点播**

准视频点播是服务提供商通过组播技术将相同的视频内容通过时间交错方式在一组虚拟频道上实现准视频点播功能。通过增加组播频道的数量，用户可以在电子节目单中选择同一内容的不同组播频道，从而选择当前要看的节目内容。

IPTV 系统具有三重播放的能力，且支持良好的交互性，因而能够提供满足用户需求的各种增值服务，典型应用包括：

- **网络游戏**

网络游戏是指在互联网上运行的游戏，分为服务器端和客户端，用户安装客户端软件后，需要登录运营商服务器才可以进行游戏。网络游戏是一种网络服务，游戏者需要向运营商缴纳一定的费用才能进行网络游戏。

网络游戏近几年随着互联网技术和网络带宽的提高得到蓬勃发展，引起政府和社会各界广泛关注，国家对网游产业采取积极扶植和大力支持的态度，最近由国家新闻出版总署公布了首批"中国民族网络游戏出版工程"名单。这说明我国的网游产业正向着多元化、本土化的方向发展，已成为一个健康向上的朝阳产业。网络游戏在今后发展中必将成为一种规范化、标准化、产业化的新兴行业，是未来发展不可忽视的领域。网游产业发展壮大，将彻底改变人们生活娱乐方式，使我们的生活越来越丰富多彩。

- **电视上网**

尽管目前个人电脑日益普及，但仍有相当一部分人认为电脑过于昂贵、太复杂。这个群体的人们只是喜欢偶尔上上网，收发电子邮件，而不想费神去拥有或学习使用电脑。IPTV 业务的出现使他们的愿望得以实现。他们可以利用机顶盒的无线键盘或遥控器在电视机上享受定制的互联网服务，浏览网页和收发电子邮件，享受高科技带来的丰富的信息资源。

- **远程教育**

远程教育是指通过音频、视频(直播或录像)以及包括实时和非实时在内的课程通过计算机、多媒体与远程通讯技术相结合的网络传送到校园外的教育。IPTV 所具有的点播功能完全符合远程教育的需求，是远程教育课件点播很好的应用平台。

如今国内的 IPTV 部署正处于初级阶段，还存在许多不定的因素，特别是 IPTV 业务被用户接受和认可的程度将决定着 IPTV 的成败。因此业务将是 IPTV 业务提供商需要着重考虑的问题。特别是电信运营商，在部署 IPTV 网络的过程中，准确的界定一下 IPTV 所涉及的业务是非常重要的。事实上，IPTV 不仅仅是 IP 视频，电信运营商必须从这个角度出发，考虑他们和传统的电视运营商之间的区别。虽然所有针对于 IPTV 的框架选择和基础技术都集中在如何开展不同的广播和点播业务，但是电信所擅长语音和高速数据业务使得他们可以提供其他一些捆绑业务来作为整个 IPTV 业务的一部分。

首先要部署的基础业务毫无疑问是视频广播和点播业务，但是通过集成语音、通信、电视商务和广告等业务来对基本业务进行增强的也应该考虑。事实上，业务肯定是在不断的演进的，无论是其定义还是涉猎的范围。由于世界上大多数的电信运营商仍处在 IPTV 部署的早期阶段，所以许多 IPTV 业务的定义还不清晰；随着基础设施的建设到位和用户需求的进一步了解，这些业务的含义将会更加明确。

3.3 IPTV 业务系统体系架构

IPTV 是一个全新的业务，在商业模型上有别于其它传统的电信和广电业务。IPTV 的产业链构成更加复杂，包括了内容提供商、业务提供商、网络运营商和最终用户等主要环节。IPTV 业务需要产业链各个环节的支持才能构成完整的业务系统，从技术角度，IPTV 系统可以分为三个组成部分：IPTV 业务平台、IP 承载网络和用户终端，如图 3.2 所示。每个组成部分由一些关键设备或软件系统组成，完成相应的基本功能。

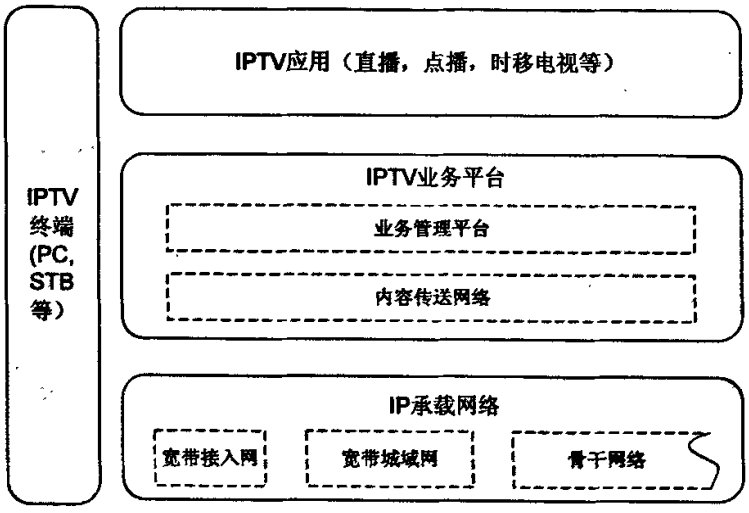

图 3.2 IPTV 系统体系架构

3.3.1 IPTV 业务平台

根据业务运营需要，IPTV 业务平台包括内容处理系统、内容分发网络和运营管理平台。内容处理系统负责内容的数字化、压缩编码以及分布式存储等。视频内容信息极其丰富，经过数字化后仍将需要大量的存储空间，压缩编码技术可以减小视频的存储容量和传送带宽需求。即便经过高度压缩，数字化后的视频文件仍然相当庞大，因此存储系统必须具备海量的特性。

由于多媒体内容的高带宽和实时性要求，底层的承载网络通常难以满足直接传送视频内容的需求，而需要在承载网络的基础上建设专用的 IPTV 内容分发网络。广义而言，内容分发网络是在内容源和用户终端之间进行内容分发的业务网络。内容源包括提供组播和点播服务的媒体服务器，它将存储系统中经过数字化和压缩处理的媒体内容以流化的形式推送到网络中。

运营管理平台包括用户管理、业务网络管理以及其它运营信息管理功能。其中用户管理包括对 IPTV 业务用户的认证、计费、授权等功能，以保证合法用户可以得到安全高质量的服务。

3.3.2 IP 承载网络

IPTV 承载网络是运行 TCP/IP 协议的网络，包括骨干/城域网络及宽带接入网络。骨干/城域网络是各种业务共享的通信平台。目前城域网络主要采用千兆/万兆以太网络，而长距离的骨干网络则较多选用 SDH 或 DWDM 作为 IP 流量的传送网络。

宽带接入网络主要实现用户到城域网络的宽带连接，目前常用的宽带接入技术包括xDSL，LAN，WLAN和双向HFC等，可以为用户提供数百kbps~100Mbps的带宽。

3.3.3 IPTV用户终端

IPTV系统的用户终端一般有三种类型，即PC（个人计算机）、机顶盒+电视和智能移动终端。PC终端包括了各种台式计算机和笔记本式计算机，此类设备的特点是自身集成了较强的处理能力，不仅可以独立完成视频解码回显任务，还可以安装其它软件完成信息交互、自动升级和远程管理等功能(如浏览器和终端管理代理等)。

电视机一般仅具备显示模拟和数字视频信号的能力，而不具备交互通信能力。机顶盒主要作为数字视频信号的接收和处理设备并通过网络与媒体服务器进行交互。因此，目前采用机顶盒+电视机的终端形式才能满足全部的IPTV业务要求。

智能移动终端主要是指智能手机及个人数字助理（PDA）。目前，市场上具有显示动态画面功能的手机由于受网络传输速率和视频解码处理能力的限制还无法提供流畅的视频信号，因此只有在无线宽带接入网络，如3G网络投入运营并结合更高效的编码方案后，移动终端才会逐步成为IPTV的终端设备。

3.4 IPTV网络参考体系架构

和传统的电信业务不同，IPTV业务的视频内容主要来源于专业的内容提供者，服务提供商通过网络运营商的网络把内容提供者与最终消费者联系起来。在提供IPTV服务的过程中，尽管通信网络仍然是业务提供的基础，内容已经成为不可缺少的组成部分。图3.3所示为ATIS定义的IPTV网络的参考体系架构[2]。

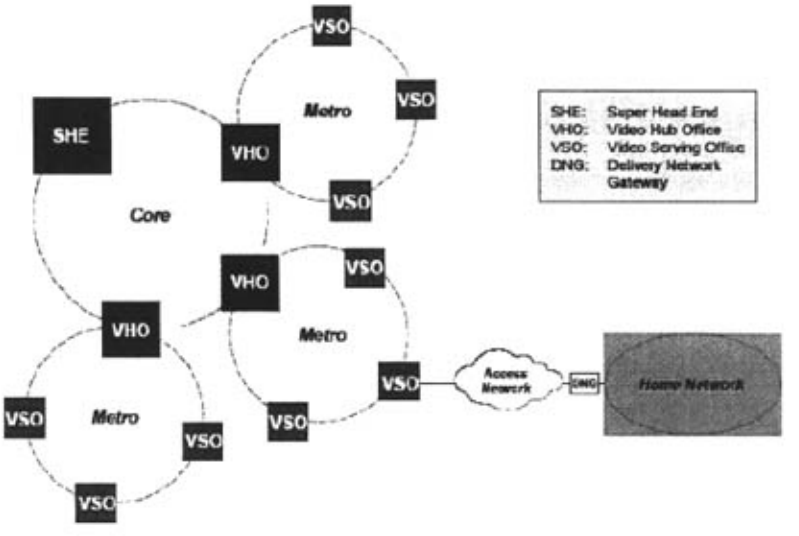

图3.3 IPTV网络参考体系架构

图3.3所示的IPTV网络参考架构包括了内容网络和传送网络两个层面。在内容网络层面分别定义了如下结点：

- 超级头端（Super Head End, SHE），是国家级的广播视频获取与汇聚点，也是插入点播视频的中心点。广播视频通常通过卫星接收并经过处理后分发到下游结点。点播内容从各种渠道获取并进行处理。SHE是整个IPTV网络的内容源，其任务包括视频内容的存储及区域无关内容的传送。

- 视频分发中心（Video Hub Office, VHO），是区域性的视频分发点，它的目的是为本地的IPTV网络提供内容。VHO是区域性的广播视频源，负责存储本地媒体内容，接收上游的VHE内容并为下游用户提供连接。本地广告的插入也在VHO中执行。IPTV业务由VHO通过汇聚/接入网络提供给最终用户，VoD服务器和其它应用服务器通常置于VHO中。

- 视频服务中心（Video Service Office, VSO），也称为中心机房（Central Office, CO），存放和管理用于将用户连接到IPTV网络中的所有接入系统。此外，汇聚设备可以使网络连接更加高效，VSO也是汇聚设备的存放点。

- 传送网关（Delivery Network Gateway, DNG），专门服务于单个用户/家庭。DNG支持连接运营商IPTV网络的接口和连接家庭网络的LAN接口。

在网络层面，端到端的视频业务传送路径上的关键组成部分包括：

- 核心网络：包括宽带业务路由器（Broadband service routers，BSR），位于视频分发中心VHO，用以连接提供音视频源的业务后端基础设施SHE和广域Internet。

- 城域/汇聚网络：包括宽带业务汇聚结点（Broadband service aggregators，BSA），用于汇聚用户流量并在城域范围内为多个业务中心机房分发流量，BSA一般位于VSO中。

- 接入网络：包括宽带业务接入结点（ Broadband service access node，BSAN），在业务中心机房（CO）提供用户接入。

- 家庭网络部分：包括家庭网关（RG）和机顶盒（STB）等，用以传送音视频信号到终端设备。

相对于话音和数据业务，视频业务对网络提出了更高的要求，为了满足用户对IPTV业务的期望需要考虑以下的问题：

- 设计一个良好的网络架构，这个网络架构内在具有支持BTV和VoD业务的可扩展性、可靠性和QoS需求，同时可以优化业务提供成本并具有足够的灵活性适应视频业务的变化和将来流量的变化。

- 规划和优化网络的容量、QoS机制和资源保护机制，使得在正常情况下端到端路径上网络拥塞最小化。

- 在特殊情况下，为了避免网络出现过载采用接纳控制机制。

- 测量和验证用户服务质量的SLA满足情况，如有必要采取相应的纠正行动。

3.5 IPTV 业务提供的技术需求

IPTV 也称为网络电视。确切地说，IPTV 是利用宽带网络作为基础设施，以家用电视机或个人电脑作为主要显示终端，利用一系列 Internet 协议承载和传输经过编码压缩的多媒体数字信号，为家庭用户提供包括电视节目在内的多种交互式数字多媒体服务以及各种增值服务的技术。人们可以通过 PC 个人电脑、机顶盒+电视机、智能移动终端等多种方式享受 IPTV 服务。从用户角度看，IPTV 业务可以提供具有个性化和实时交互特点的点播服务和类似于传统电视的非交互式广播服务。

IPTV 业务的突出特点是具有交互性和实时性，从而改变了传统电视的被动收看模式，实现了无论何时何地“按需观看”的交互式电视功能。IPTV 业务的实现过程对业务网络、IP 承载网、宽带接入、终端，以及媒体编解码、数字版权保护等一系列技术提出了新的要求。特别是由于视频的高带宽和实时性要求，视频内容存储与分发是 IPTV 业务提供的关键技术。

IPTV 是一种三重播放业务，需要同时支持视频、语音和数据业务，其系统架构必须要满足各种复杂的技术需求。用户需求是业务发展的基础，提供良好的用户业务体验是 IPTV 系统的设计目标。对于一个新业务，用户希望它在提供更加丰富的影视内容的同时，还需要具有与传统有线电视相比更加优良的特点。除了与有线电视节目相当的高质量图像效果、类似于遥控器的简单操作方式，用户还需要个性化节目定制，需要能够在任何时间和地点自由选择观看内容的全新业务体验。

为了满足用户的需求，IPTV 业务对现有网络技术、计算技术、多媒体技术等都提出了很高的要求，对传统的商业模型、网络融合、安全机制等也提出了许多新的挑战。IPTV 业务的技术需求主要有以下几点：

- 承载网络需求

  为确保良好的服务质量，需要底层的 IP 承载网络和接入网络具有足够的容量，并可以灵活地管理和控制网络资源。

- 信源编解码技术需求

  要求在提供高质量图像画面的同时尽量降低编码速率，为了减少传输差错的影响，还需要编码具有一定的容错能力。

- 业务性能需求

  IPTV 系统需要提供明晰的节目单和业务导航能力以及大容量内容存储、管理与分发技术。IPTV 是以内容为中心的实时业务，用户的使用目的多数是为了娱乐和消遣，因而需要高质量保证。

- 数字安全保护需求

出于对内容提供商内容资源的保护，IPTV 系统要有良好的安全管理和版权保护措施，一方面对内容进行管理，另一方面对用户进行认证授权。针对用户的游牧性，需要采取统一的认证机制。

综上所述，IPTV 业务提供对承载网络质量、信源编解码、业务提供能力、内容安全管理等提出了一系列的技术要求。在构建 IPTV 业务体系架构时要充分地利用现有网络和技术优势，尽可能多地支持各类技术需求；在某些业务不明朗、技术不成熟的方面应留有余地，以将 IPTV 系统建成具有融合、开放、安全特点的多媒体业务平台。

3.6 小结

本章较为系统地介绍了 IPTV 业务与系统的基础知识，研究了 IPTV 业务提供平台和 IPTV 网络的基本模型，在此基础上分析了 IPTV 业务提供的技术需求，为后续的研究工作提供了技术背景和基础知识。

参考文献

[1] IPTV 范围、定义和场景（讨论稿），中国通信标准化协会，2006 年 2 月
[2] ATIS IPTV Interoperability Forum (IIF), IPTV Architecture Requirements, ATIS Standard, May 2006

第 4 章　IPTV 业务保障及其关键技术

IPTV 提供商是希望通过提供有吸引力的、经济便捷的业务绑定从而与用户建立可赢利的长期业务关系，然而实现这一目标还存在诸多风险。在国外，有线多业务运营商(MSOs）通过提供语音、视频和数据业务绑定取得了很大的成功，已经吸引了大量的电信用户，对传统的电信运营商构成了严重的威胁。随着监管政策的放开，国内的电信运营商迟早要面临相同的问题。有线多业务运营商也在谋求开展 IPTV 业务。网络基础设施改造需要高达数十亿美元的投入，失败的风险对任何运营商都是难以承受的。因此，IPTV 业务竞争会非常激烈，价格战也将在所难免；为了赢得市场，电信运营商必须推出具有质量保证的 IPTV 服务以满足用户的需求。

4.1 IPTV 业务保障的重要性

和传统的电信业务不同，IPTV 业务的视频内容主要来源于专业的内容提供者，服务提供商通过网络运营商的网络把内容提供者与最终消费者联系起来。在提供 IPTV 服务的过程中，尽管通信网络仍然是业务提供的基础，内容已经成为不可缺少的组成部分。相对于话音和数据业务，视频业务对网络提出了更高的要求，

IPTV 是一种多媒体业务，对网络的 QoS 具有非常严格的要求。首先，由于视频成分的存在，IPTV 要求很高的传输带宽。其次，IPTV 中的音频成分对传输时延提出了很高的要求。为了保证理想的声音质量，端到端的传输时延应小于 200 毫秒，时延抖动应小于数十毫秒。由于多种媒体成分的同时存在，在播放时要保证各种媒体的时间同步。与用户的交互能力是 IPTV 业务的重要特征，具备交互能力的同时保证良好的业务体验对 IPTV 业务部署提出了更高的要求。

传统的的卫星电视和有线电视采用专网传输的方法，它们为用户体验质量（Quality of Experience, QoE）确立了很高的标准。保障良好的用户业务体验为电信运营商的业务部署带来了技术挑战和潜在风险。视频业务对电信运营商来说是全新的领域。为了开展视频业务，电信运营商将要面临许多重大的挑战，其中包括建设新的城域和接入网络满足 IPTV 大量的带宽需求，重新规划核心网络以支持视频业务传输，增加新技术实现内容管理、分发与计费等。这些挑战不仅是技术上，而且会带来很大的财务开销。由于需要处理复杂的内容部分，IPTV 业务提供非常昂贵。网络基础设施的升级就需要花费数十亿美元。例如，美国的 AT&T 公司将花费 46 亿美元实现在 2008 年底前为 40 个新兴市场的 1900 万家庭部署 IPTV 业务。

因为现有的有线和卫星电视业务都提供可靠的业务体验，在所有的技术挑战和投资之上，用户质量期望被格外重视。因为投资巨大而且用户获取成本很高，电信运营商需要以质取胜。IPTV 被认为是非常有潜力的业务，但它必须可靠、可定制、足够灵活，这样才能推动新用户和运营收入的增加。不仅它们的业务提供要有足够的吸引力从其它运营商那里吸引用户，电信运营商还希望 IPTV 能够满足用户更高的期望以降低用户流失的风险。

对普通用户而言，画面质量、频道可靠性和频道响应时间意味着一切。用户已经习惯于传统的电视业务，他们对相同的质量体验期望很高而容忍力很低。如前所述，市场竞争

非常激烈，不能提供高质量的服务将使用户转投其它运营商。电信运营商的风险不仅在于丢掉 IPTV 用户，而在于可能失去整个业务绑定。只有通过提供一致的、具有质量保证的业务，运营商才可以培养用户的忠诚度并减少用户流失。正因为此，业务质量保障意义重大，电信运营商需要在 IPTV 业务推广的早期做好准备，确定实现业务保障的技术方案。

4.2 IPTV 业务保障关键技术

为了保证运营商的正常运营，端到端的业务保障是 IPTV 运营商的强烈需求，但无论是视频质量管理还是 IP 网络资源管理对电信运营商而言都是全新的挑战。IPTV 业务质量保障的目标是提供理想的用户体验质量。通常业务质量通过业务提供商与用户之间的服务等级约定（SLA）来确定。大多数用户并不了解 IPTV 业务提供的技术，通常他们不能帮助确定问题的原因。IPTV 系统由多个独立的部分组成，问题可能出现在电视机，机顶盒，内容编码/加密，用户家庭网络，接入网络或核心网络等。实现 IPTV 业务保障有必要开发工具完成以下任务：

- 设计一个良好的网络架构，规划和优化网络的容量、QoS机制和资源保护机制，使得在正常情况下端到端路径上网络拥塞最小化。

- 测量和验证用户服务质量的SLA满足情况，如有必要采取相应的纠正行动。包括监视端到端的全部系统；如果某个部分发生故障发出指示信号，启动自动故障管理与恢复的过程；判断未解决的故障对服务效果和范围的影响。

4.3 IPTV 业务保障体系架构

电信运营商在竞争中获胜的关键是通过提供SLA对用户的服务等级负责。因此，运营商必须保证能够实时监视并报告SLA。因为用户也会根据业务使用情况对运营商提出某种要求，运营商必须有适当的机制以清楚明确的方式通知用户当前的业务状态。从用户为中心的角度，所有重要的质量指标都要能够通过网络实时监测获得，即业务必须是端到端可视的；只要需要，可以用数据确定客户的体验。电信运营商甚至可以定义新的测量准则适应新的业务以及业务等级。

IPTV 业务保障解决方案需要能够提供完整的端到端的业务可视化。网络的复杂性和视频内容创建的动态特性都可能成为对用户体验产生有害影响的潜在故障点。传统的运营支撑系统和网络管理系统不能提供所需的端到端可视化，也不能有效地监视实时流量和业务。

IPTV 业务涉及三个主要的组成部分：底层承载网络，视频内容以及总的用户体验。IPTV 业务保障方案需要对这个三个方面的质量提供可视化。通过连续不断地监视网络和业务的性能，电信运营商利用业务保障方案识别 IP 传输问题或者视频质量损伤；通过端到端的可视化检查视频质量，对高价值用户的总的业务体验进行预先管理。因此 IPTV 业务保障体系结构包括三个关键方面：承载网络、内容以及用户的业务体验，如图 4.1 所示。

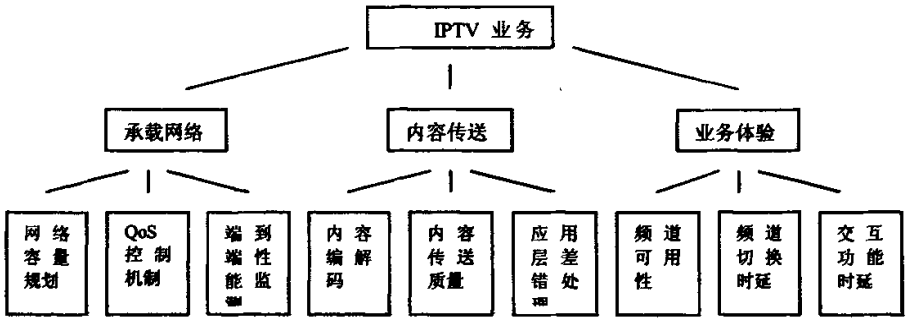

图 4.1 IPTV 业务保障体系架构

4.3.1 承载网络质量保障关键技术

承载网络是业务开展的平台，承载网络的质量保障是 IPTV 业务保障的基础。业务质量保障需要充足的网络容量。为了保证 IPTV 的业务质量，网络必须提供足够的带宽以保证传输时分组丢失率和时延限制在一定范围之内。承载网络的质量保证首先要提高接入网的带宽配备和管理核心网的带宽使用 [2]。

通常标准清晰度视频流需要 2M 的带宽；取决于压缩标准和质量要求，高清晰度视频流需要 4~8M 的带宽。考虑到三重播放的需要，每个家庭至少需要 20M 以上的带宽才能满足 IPTV 的部署需求。当前常用的接入技术包括 xDSL 和 xPON。从技术上而言，两者都可以通过适当的配置满足 IPTV 业务部署的需求。

现阶段城域核心网以 IP/MPLS 及以太网技术为主，相对而言容量较为充足。但核心网是一个共享网络，不同业务竞争其网络容量，因此有必要加强核心网容量管理以对其合理使用。对于广播应用，IP 组播技术可以有效地节约带宽，但是将不同的频道分配到多个组播组中也带来复杂的带宽管理问题。假定采用恒定比特率编码方式，则不必考虑统计复用效应。一种极端情况是每个频道分配单独的组播树，其优点是不同的组播分支只需接收其想要的频道内容，因而传输效率较高。但如果频道较多，这种方式将带来很大的组播管理负担，还会增加频道切换的时间。另一种极端的情况是将所有的频道放到一个 IP 组播树中。显然，用户不可能同时观看多个频道，这种方式会造成严重的传输资源浪费，但频道切换快捷。多个组播树可以进行灵活的流量管理，而少量组播树传输多个频道可以大大降低信令负担。因而，IPTV 业务支撑系统需要系统地评估用户和频道的实际情况，为多个频道合理地规划组播树数量。

在 IPTV 业务部署之初，视频点播的流量会小于广播流量。但由于点播采用单播技术，取决于视频服务器在网络中的位置，点播将消耗相当大的核心网带宽。VoD 服务器的负载和连接服务器链路的利用率需要配置适当并在运行过程中进行监测。为了保证点播业务的质量，可以对点播流量采用高优先级传输，但优先级的配置需要兼顾其它业务的带宽需求。此外，为防止网络过载造成传输质量下降，接纳控制机制是改善用户体验的有效方法，我们将在下一章中详细说明在 IPTV 网络中实现接纳控制的方案。

实现承载网质量保障的另一个关键因素是了解和理解网络在传输视频流量时的表现。有效的业务保障解决方案需要让运营商能够快速地判断出当前网络出现问题的位置，并且在业务质量下降之前予以修复。因此，承载网的质量保障除了需要合理的容量配置与管理，还需要监视、测量、诊断及维护等系统的方法。

4.3.2 内容传送质量保障关键技术

对用户来说，画面质量是业务体验的重要方面。定义画面质量模型是评价内容传送质量的前提。衡量视频质量有多种指标和模型。对视频质量指标的严格定义仍然没有定论，是当前主要的标准化组织和产业界论坛讨论的话题之一。由于视频质量来源于最终用户的主观感受，客观地评价较为困难。平均主观分数（Mean Opinion Score，MSO）是描述视频业务质量体验常用工具。和话音类似，视频 MSO 也分为五个等级，其中 5 级表示最高质量等级。对于高质量的视频传输服务，通常要求分组丢失率在10^{-4}到10^{-7}之间，甚至更小；可以容忍的分组时延在数百毫秒数量级，时延抖动在数十毫秒数量级。

不难理解，分组网络在分组丢失、时延及时延抖动方面的缺陷对画面质量感受有很大的影响。实际上，用户的 IPTV 业务质量体验受到多方面的因素影响：内容源质量，编码质量，及由于传输丢失、差错及时延造成的图象质量损伤等。尽管合理的规划和配置承载网络对实现端到端的用户体验具有重要作用，但承载网络建设并不能解决 IPTV 业务保障的全部问题。首先，承载网络是一个多业务平台，并不是专门针对 IPTV 业务进行优化。其次，减少或消除网络的所有问题会大大增加基础设施的建设成本。另一方面，用户并不能感知也不在意良好的业务体验来源于何处。为了在保障业务质量的同时维持合理的成本结构，业务提供商有必要采取其它方法来改善用户的视频业务体验，应用层内容传送技术是常用的技术手段。差错掩盖（Error Concealment）就是一种有效的应用层视频传送质量保证技术。

H.264 (MPEG-4 part 10) 定义了网络适应层（Network Adaptation Layer）用以克服网络传输产生的数据差错和丢失。关键比特的丢失，如图象头部信息会对解码过程产生严重的负面影响，所以关键数据必须被分成多个部分以一种特殊的方式进行处理。为了保证媒体数据的时间特性，多媒体数据通常采用轻权的传输协议，如 UDP。因此，音视频流在传输过程中很可能会出现差错。检测视频流中的严重差错并且采取措施对之进行掩盖比简单地任由其降低画面质量效果要好得多，而没有检测出来的差错可以使图象质量严重下降。常用的差错掩盖技术是在丢失少量视频帧时简单地保持图象不变或插入相邻帧。音频质量也是 QoE 的组成部分，分组丢失带来的差错也需要进行适当的掩盖。 如果运用得当并结合更底层类似的机制，差错掩盖以及重传和前向纠错能够帮助减少偶尔的丢失和错误分组。

用户已经习惯了传统电视的画面质量，因而低劣的图象质量是难以接受的。不幸的是，IPTV的传输质量难以控制，IP网络传输的视频质量可以在数秒内发生变化。为了保障内容质量，自动地、连续地视频质量主动测量与被动监视，可以确保内容经过网络传递到用户时能够保持质量一致性。为了实现内容质量保障，运营商需要自动地监测视频流并产生视频质量的评价分数。相对于音频情况，视频质量测量带来了更多的挑战。视频质量与视频生成的设备有关，视频编码算法的复杂性，不同的应用场合，视频广播和点播采用的传输技术等也会影响视频的质量体验。连续地测量视频质量以准确地反映用户的实际体验对IPTV业务运营非常关键。

4.3.3 业务体验质量保障关键技术

保障用户业务体验对IPTV业务的成功至关重要。业务保障机制需要连续测量频道切换时间及视频功能的时延，提供运营商需要用来保证用户体验的基本使用统计数据。如果没有提供用户使用情况的业务保障机制，运营商就没有事先确保用户期望能够得到满足的必要依据。完整的业务体验质量保障功能包括：

- 监测频道可用性及频道切换性能；
- 提供用户体验报告和使用的统计信息；
- 通过仿真机顶盒测试频道选择和执行VoD交互功能的时延。

因为内容已经推送到终端，传统的电视频道选择反映非常快捷，用户体验良好。由于IPTV的协议特征，频道选择时延成为一个新问题。IPTV采用IP组播技术，其中的基本协议是IP组管理协议（IGMP）。组播避免了视频头端对每个用户发送频道内容，大大节约了带宽，增加了频道容量。但采用IGMP意味着频道切换在网络中进行，而不是本地的机顶盒。IGMP的加入与离开操作时延成为影响用户质量体验的重要因素，同时发生的频道加入请求数量将影响请求用户加入组播组的性能。

为了使频道转换更加快捷，Microsoft在其IPTV系统中设计了瞬时频道切换（Instant Channel Change，ICC）技术。当用户切换频道时，ICC保证画面可以在瞬间出现。ICC的主要工作原理是在内容传送网络中引入一个分发服务器（D服务器）来响应用户的频道切换请求。当用户请求频道切换时，刚开始时的内容不是通过加入相应的组播组获得，而是由D服务器以正常频道速率的130%向请求终端发送一个单播流。ICC避免了频道切换时的组播信令操作，因而可以消除用户频道切换的等待时间。另一方面，D服务器和请求终端之间的单播流量增加了视频的带宽使用，对承载核心和接入网络有一定影响。

除了频道切换时延外，以下几个因素也会对用户业务体验产生重要影响。视频点播（VoD）和个人视频记录（Personal Video Recorder，PVR）业务都是交互式业务，控制命令响应时间是交互式业务体验质量的关键因素。用户暂停、回退及快进控制命令的响应时间IPTV质量的基本指标。机顶盒的重启时间也可能成为用户的烦恼，特别是重启可能丢掉与网络的同步。在IPTV系统需要处理多个同时认证请求时，这个问题也可能出现。在多媒体数据流中，音视频回放可能会出现时间不一致，即“唇同步”问题，当音视频流需要通过IPTV网络的不同部分处理时此问题往往会变得更加严重。对于上述问题，都需要实时地监测网络状态和相关的技术指标，在问题出现之前或初期就及时加以解决以保证用户的业务体验质量。

4.4 小结

由于技术的进步和替代网络和应用的出现，电信行业的发展正面临严峻的挑战。为了摆脱当前的困境，提供具有三重播放能力的 IPTV 业务并加强多业务绑定是目前公认的有效办法。IPTV 具有传统电视所不具备的技术优势，它使业务提供商能够提供用户所需的视频广播和点播及其它增值服务。然而作为一项全新的业务，IPTV 在技术上和财务上都

对电信运营商带来了很大的风险。为了保证 IPTV 业务的成功，提供可靠的业务质量保障将发挥关键作用。

参考文献

[1] Brix Networks Inc, Service Quality Matters for IPTV: A Lifecycle Approach，white paper, 2006
[2] K. Kerpez, D. Waring, G. Lapiotis, et al, IPTV service assurance, Communications Magazine, IEEE, Sept. 2006, Vol.44(9): 166 – 172

第 5 章　IPTV 业务提供接纳控制方案

电信运营商正在保留老客户、吸引新客户和增加 ARPU 值等方面的经历着前所未有的激烈竞争，这是促使它们提供 IPTV 业务的主要推动力。下一代宽带接入、汇聚和边缘网络将是 IPTV 业务提供的基础，但确保用户体验水平的业务保证和控制方案将最终决定这些努力的成败与否。

在 IPTV 业务部署中，影响用户获得较高体验质量（Quality of Experience, QoE）的因素是多方面的。首先，各种业务流的混合是相对不可预测的，将影响广播和单播流量在网络中的比例及网络架构的设计。点播视频（VoD）会话的并发直接影响网络中单播流的数量，因此也是影响网络设计和业务可靠提供的主要因素。同时观看的广播频道的数量将影响网络中的组播复制，而高清晰度（High-Definition, HD）内容的增加则直接增加了带宽的使用。其它因素，诸如用户使用 VoD 业务的模式，机顶盒（Set-Top Box，STB）数量的增加等，会进一步增加资源需求并影响内容在网络中的优化分布。基于网络的业务智能和 QoS 机制（如层次化 QoS）需要能够根据事先定义的策略高效地处理实时变化的动态流量。考虑上述因素及日趋增加的用户基础，从 IPTV 系统中间件获取相关信息，准确地理解资源可用性、使用情况和性能细节对提供 IPTV 业务保障非常重要。

视频接纳控制（Video Admission Control, VAC）正在成为电信网络和设备的新需求。VAC 相当于保证用户体验的安全阀门，它可以防止当网络资源不足时（如新片上映时极端的 VoD 并发高峰，设备故障时容量骤减）新的业务流进入网络。因为 VoD 在每次使用时付费同时也是最消耗带宽的业务，VAC 最重要的应用是控制 VoD 会话的接纳。尽管 VAC 是在业务质量和可用性之间的折衷，由于网络资源不足而拒绝服务的情况仍应视作为意外而非常规做法。考虑到网络拥塞总是难以避免，高效的接纳控制机制无论对于网络基础设施、策略控制系统还是 IPTV 中间件都具有重要意义。

本章分析了 IPTV 系统的体系架构和各种接纳控制机制，在此基础上提出了一种易于部署的城域网视频接纳控制机制；该方法实现简单、易于部署，为保证用户的视频业务体验质量提供了新的手段。而且，该方法可以与基于策略服务器业务管理架构集成，从而有效控制了 IPTV 业务的部署成本。文中还介绍了 Alcatel 基于策略服务器的视频接纳控制解决方案。

5.1 IPTV 系统的视频接纳控制机制

通信网络资源规划是在网络建设之前对网络资源的需求进行分析并对规划周期内的资源使用情况进行预测。考虑到成本因素和实际运营需要，网络的建设和扩容具有相对稳定的周期。然而在实际运营过程中，需求总是不断变化的，而且有些变化是在网络规划时难以预测的。因此需要一种机制，当网络资源不足时可以对后来的或不重要的需求采取拒绝措施，以保证已经在进行中的或重要的通信任务可以顺利完成。根据网络资源使用情况或特定的商业策略拒绝某些通信请求的措施就是网络的接纳控制机制。

根据接纳控制实施的位置，可以将接纳控制机制分为三类：端到端的接纳控制，端点间的接纳控制和瓶颈点接纳控制。在上述的三种方法中，分别存在一个或多个接纳控制

点，即实现接纳控制算法并实施接纳控制的位置。在接纳控制的实施点，接纳控制主要由三个基本的组成部分构成：流量描述、接纳准则和测量过程[1]。接纳控制的根本目标是拥塞避免，但根据应用对性能的要求不同，具体的接纳准可能是带宽约束也可能是分组丢失率约束，其表达形式既可以是确定的，也可以是统计的。关于实现统计 QoS 的接纳控制机制可参考[2]

5.1.1 IPTV 视频业务对资源的需求分析

IPTV 业务包括视频，音频和数据部分，其中视频是最主要的成分。视频信号传递的信息最为丰富，但同时对资源的消耗也非常大。采用先进的 MPEG-4 多媒体压缩编码算法，传输高清晰度视频需要 8Mbps 的带宽，而标准清晰度（Standard Definition, SD）也需要 2Mbps 的带宽。

业务提供商通常在网络的每个部分为 BTV 业务配置一定数量的带宽，包括 HD 和 SD 内容。VoD 业务的资源消耗与用户的使用习惯有关，特别是峰值并发率（peak concurrency rate），即在同一时间内用户使用 VoD 业务的比率。显然，峰值并发率越高，意味着同一时间内使用 VoD 业务的用户数越多，所需的资源就越多。

下面通过一个例子来分析 IPTV 系统的带宽资源需求。假定某运营商有 400 万用户，在分布式的城域网络架构中，通过 400 个中心机房提供服务，则平均每个中心机房服务 10000 个用户。若 IPTV 业务的订购率为 40%，则每个中心机房需要服务 4000 个 IPTV 用户。对于 VoD 业务，假定视频点播的峰值并发率为 20%，每个 IPTV 的订购家庭拥有两台电视机，在点播视频中标准清晰度流占 90%，高清晰度流只占 10%，则提供标准清晰度为主的视频点播，VHO 到中心机房的链路就需要约 4Gbps 带宽。具体计算为：2*(4000*20%) = 1600; 2*(1600*90%) + 8*(1600*10%) = 4160Mbps。对于 BTV 业务，假定总共提供 200 个频道，其中 HD 频道占 20%，其余为 SD 频道，则所需带宽为：2*(200*80%) + 8*(200*20%) = 640Mbps。VoD 与 BTV 业务总的带宽需求约为 5Gbps。

从上面的分析可见，视频业务，特别是 VoD，对网络资源的需求比高速上网和 VoIP 业务要多得多，因此网络设计时必须尽力优化以对其提供最高效的支持。在部署高效、可靠和可扩展的视频服务时，服务提供商可以为网络配备最大的带宽，即以最坏情况下估算出来的并发峰值所需的带宽，包括网络出现故障的情况，也可以部署接纳控制机制管理偶尔出现的峰值超出可用带宽的情况。第一种方法非常昂贵和浪费。由于视频加入到数据和语音中使拥塞难以完全避免，第二种方法平衡了资本投入和高质量的用户体验，已经成为运营商的优先选择。

BTV 被认为是任何 IPTV 业务的基础，无论如何应当保护其正常工作。用户在观看有线电视时，通常不会出现因为资源不足而被拒绝服务的情况，有线电视用户也不会担心频道切换失败。因而，在网络的设计和资源规划时，应该保证 BTV 业务的带宽请求无阻塞。相对于 VoD 业务而言，BTV 业务所需带宽较少且基本确定，保证 BTV 业务的带宽请求无阻塞代价可以接受，因此我们认为暂不需考虑对 BTV（组播流）进行接纳控制。

然而，VoD 业务的单播特性决定了它可能消耗大量的资源，随着用户对 IPTV 业务认知和认同程度的提高，VoD 业务将给运营商的网络带来严重的挑战。在点播高发时期，20%甚至更多的用户可能使用点播业务观看 VoD 视频。同时，网络链路的故障也可能使峰

值点播需求超过网络容量。VoD 业务的用户使用习惯 （并发数量）通常难以准确预测，因而也难以进行准确的资源规划，为了避免网络在某些特殊情况下出现过载，有必要实施基于资源使用状态的接纳控制机制。

5.1.2 城域网架构模型与资源瓶颈分析

图 5.1 所示为城域网的基本架构模型，为了便于分析其资源瓶颈，即潜在拥塞点，这部分网络可被分成四段 [3]：

- 第一英里 （First Mile），用户家庭和第一个接入网元之间的链路；
- 第二英里 （Second Mile）， 宽带接入网元和宽带汇聚结点之间的链路；
- 第三英里 （Third Mile），宽带汇聚结点和宽带边缘路由器之间的链路；
- 第四英里 （Fourth Mile），第四英里，宽带边缘路由器和视频源之间的链路。应当指出，在分布式VoD服务器架构中，和区域或本地内容被放置在距离用户更近的位置时，“第四英里”可能是视频源与汇聚结点间的直连链路。

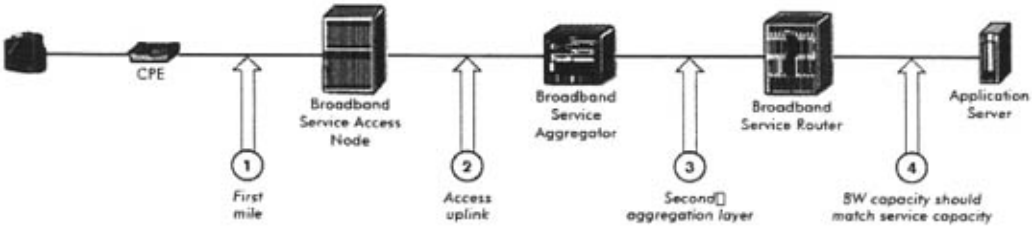

图 5.1 IP 城域网潜在的容量瓶颈

在第一英里处，用户与所需容量之间存在着简单的一一对应关系。然而接入技术的限制使得该点容量难以大规模扩大，因而很有必要实施接纳控制和执行适当的策略控制以保证业务请求不超过规范的和实际的带宽容量。

第二英里处链路容量和结点容量限制了可用带宽，需要层次化 QoS（Hierarchical QoS, H-QoS ）机制在业务与用户之间分配可用带宽及应对实时的流量模式变化。在这部分网络中，用户数量（对应多个服务实例）和网络容量之间的对应关系更加复杂一些。虽然第二英里的汇聚链路容量通常已知并且是确定的，链路的混合流量是动态变化的。

第三英里处的资源汇聚和流量统计复用增益意味着这一部分在VoD高峰时段可能成为瓶颈。相比于第二英里，这部分还需支持一些其它业务（比如商用VPN）。此外，因为它通过可路由网络连接应用和内容服务器，这部分网络路径的容量可能是动态变化的。不同于第二英里，如果有必要，扩容相对容易一些。

应用服务器本身可能成为瓶颈，但由于匹配单个或多个业务的资源需求较为容易，因而不需要在连接服务器的链路上实施接纳控制。

在通常的IPTV业务提供模型中，不必要对IP/MPLS核心网络进行资源预留和接纳控制。如前所述，VoD传送成本与距离成比例，QoS和接纳控制也需要增加到端到端路径中的每一跳。这些成本和控制问题通常采用分布的方式将VoD服务器置于地区的VHO中加以避免。

综上所述，现阶段第一英里和第二英里是主要的拥塞点，至少需要考虑在此处实施接纳控制。

5.2 视频接纳控制的实施方案

接纳控制对避免用户过度点播，防止拥塞、分组丢弃和可能的用户体验降低至关重要。随着更多的 VoD 和高清晰度广播频道需求的增加，端到端的 IPTV 业务流量按预期增长，视频传输将会遇到麻烦，视频服务的接纳控制也将更加迫切。幸运的是，用户对 VoD 业务的阻塞通常要宽容一些，因此网络规划时应该让 VoD 使用 BTV 业务剩余的带宽资源。为了避免用户过多导致网络阻塞和用户体验劣化，有必要对 VoD 业务实施接纳控制策略。

现阶段，在基于 IP 的电信网络中，由于大容量链路和结点的出现且成本降低，核心网的资源相对较为充足，而城域汇聚和接入部分是拥塞发生的潜在位置。而且，城域和接入网的改造较为困难，通过频繁扩容解决拥塞问题代价很高。幸而，如果能确定潜在的拥塞点，即可假定其它点不会发生拥塞，只需在少量潜在的拥塞点实施接纳控制。

在分布式的城域网架构中，接入链路的主要功能是用户接入，而业务管理与控制功能在汇聚结点和城域边缘实现。根据上面的分析，可以将 IPTV 视频接纳控制方案概括如下：

- 在接入接点，用户的需求比较简单，可以通过制定业务策略，比如，每个家庭只允许不超过 3 个视频会话，实施基于策略的接纳控制。
- 在汇聚结点实施基于资源的视频业务接纳控制机制。

5.2.1 资源瓶颈点的视频接纳控制算法

接纳控制的目标是在保证通信质量的同时尽量提高网络的资源利用率。相对而言，接纳控制的决策并不复杂，但资源需求与供给信息难以获取导致了问题的复杂。通常采用粗略的方法，实施比较简单，但要牺牲资源的利用率；如需提高资源利用率，则需求采用复杂的方法描述资源需求，获取资源可用信息。通常越是简单的方法越是便于实际部署，因此我们把可用带宽约束做为接纳控制的基本准则。假设：1）资源请求是没有并发的，即某一个时刻只有一个流申请新的资源。2）流之间不存在优先级，即资源是不可抢占的（non-preemptive）当可用带宽可以满足新的流的资源需求时，则接纳该流；相反，当可用带宽不能满足新流的资源请求时，则拒绝该流。

根据前面的分析，视频（VoD）接纳控制在城域汇聚结点实施。设v是预留带宽资源总和，r是表示新的用户带宽请求，u是分配该类流量的带宽总和，则该算法可以表示为：

资源瓶颈点的接纳控制算法可以简单描述为：

if ($v+r \le u$), request is allowed;

else, request is denied. //$v+r > u$

应用资源请求描述可采用平均带宽（average bandwidth），保守的做法是选择峰值带宽（peak bandwidth）来描述资源请求。显然，以峰值带宽做为资源描述可以绝对保证接纳控制的可靠性，但有可能造成很大的资源浪费（取决于请求流量的突发程度，突发程度越高则浪费越严重）。

对于接纳控制点，如不考虑统计复用增益，则可以从资源总和中直接减去正在使用的资源（资源请求的总和扣除已经释放的资源），剩余的即为当前可用资源；考虑统计复用增益，实际可用资源将大于简单扣除的结果，通常需要采用基于测量的方法获得当前资源可用信息。

5.2.2 方案分析

在分组网络中实施拥塞控制并不一件容易的事。首先，需求信息难以准确描述与传递。对于某些应用，资源的需求信息比较明确，比如采用恒定比特率编码的话音业务。而另外一些应用，如视频传输，因为流量速率是可变的，而且变化的范围很大，需要用多个参数进行描述。将端系统的资源需求传递到网络中间结点，通常需要借助于信令完成；另一种方法是将需求信息传递到资源代理点，再由资源代理与资源控制点进行协调。其次，资源使用情况难以判断。对于汇聚和核心结点，考虑到统计复用的结果，数量变化的变比特率流叠加的结果通常是难以预测的。因此，判断资源的可用情况，比如可用带宽的大小，需要采用基于测量的方法。

上述的视频接纳控制方案，根据现阶段网络资源分布和业务需求的实际情况对接纳控制的实施做了合理的简化。首先明确了接纳控制实施的对象和具体位置。根据我们的分析，VoD 业务是最主要的消耗资源者，而且用户对 VoD 的业务阻塞相对宽容一些，因而是适当的控制目标。理论上在 IP 网络中任何结点任何时间都可能出现拥塞，但在实际的网络中接入和城域汇聚是主要的潜在拥塞点。其次，根据拥塞点的资源使用情况，采用了不同的接纳控制策略。在接入接点实施基于策略的接纳控制而在汇聚结点实施基于资源的接纳控制。再次，借助网络结点提供的信息确定资源可用情况，避免了用实时测量方法获取资源信息的复杂性。

5.3 Alcatel 基于策略的 IPTV 接纳控制方案

尽管网络资源在网络建设时经过了仔细的规划，但由于资源的相对有限性和流量分布的动态变化，总会出现资源不足难以满足业务请求的情况。通常有两种类型的策略控制功能来处理这个问题：

- 业务提供时措施：例如，当网络资源不足时，禁止用户接收或激活某类服务，或限制某种业务类型（如HD和SD TV）或同时启用某类业务的数量。这些静态配置的控制确保大部分业务在大多数时间内正常工作。

- 业务请求时措施：在下个请求（比如开始观看一部VoD电影）被接受之前先验证是否有足够的资源。这种措施被称作会话接纳控制（Session Admission Control, SAC）。这种机制还必须处理资源突然失去的情形，比如在故障情形。

通常 SAC 请求需要在请求客户端（如机顶盒 STB 或 VoIP 终端）和网络服务器之间进行协商。VoD 业务流取决于 SAC 策略服务器的决策，此决策的依据是第二或第三英里带宽是否可以传输 VoD 流和其它策略决策准则。

在三重播放（Triple Play）的业务提供架构中多种业务，如 BTV，VOD 及其它 IP 多媒体应用将竞争带宽，因而需要在各种业务之间进行合理的资源分配以保证提供适当的业务。Alcatel 提供基于 Alcatel 5750 用户业务控制器（Subscriber Service Controller, SSC）的集中式宽带策略服务器为三重播放业务提供统一的业务控制 [4]。通过 Alcatel 5750 用户业务控制器将基于策略的 AAA 和 SAC 功能紧密集成起来并集中管理，从而实现了基于用户业务特征数据和网络资源可用性的业务授权完美结合。

通过集成 Alcatel 5620 业务感知管理服务器（Service Aware Manager， SAM），可以获得拓扑和链路容量信息并且利用拥塞通告消息交换带宽更新信息和验证业务性能。例如，BSAN 可以准确地知道 DSL 环路的传输速率，当传输速率低于宽带策略服务器所知的配置容量时，将发送更新通告消息。这个机制需要考虑变动的阈值以及时延以避免振荡效应。如果一个用户想要更新业务或增加新的终端设备，BSAN 可以被策略服务器触发增加容量配置。图 5.2 显示了 Alcatel 用于多种业务混合部署的网络与业务管理的端到端的体系架构。

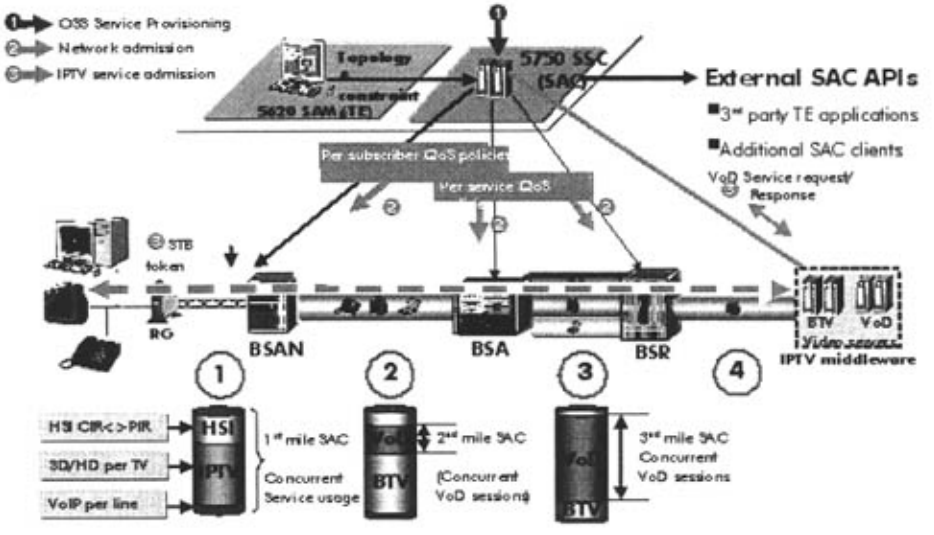

图 5.2 业务接纳控制体系架构

在图 5.2 的架构中描述了多个接纳控制场景及实施情况：

- 业务提供：通常这由运营支撑系统（Operations Support System，OSS）驱动向网络中增加/删除用户或改变用户的业务订制内容。AAA服务器接受并存储用户及其业务特征数据，在宽带策略服务器处实施统计接纳控制确定现存容量是否可以满足新用户接入网络的要求，包括接入环路容量的验证。网络中静态QoS策略也在此时配置。

- 网络接纳控制：AAA服务器根据由RADIUS和DHCP或其它授权协议传递过来的信任证书做出动态策略决策，决定接受或拒绝用户和终端设备进入网络。动态业务策略（如BSAN和BSA处的每用户/每业务策略）的配置及点播业务更新请求时会话中间策略的改变都可以通过用户自助服务门户进行改变。

- IPTV业务接纳控制：IPTV中间件平台利用令牌机制（图中底部的水平线）执行端到端的业务流分配和接纳控制。在允许HD或SD视频流的令牌请求之前，通过宽带策略服务器验证（会话接纳控制）业务请求的资源没有超出业务的订制资源及此用户或此类业务的可用带宽。

不同的三重播放业务创建过程略有不同。例如，为了建立VoD业务会话，客户端先联系VoD服务器中的控制模块，控制模块会与宽带策略服务器进行联系执行会话接纳控制对请求进行业务授权。每用户/每业务的QoS保证是通过Alcatel 7450业务交换机和Alcatel 7750业务路由器中的H-QoS机制实现的。

H-QoS 提供了一种精细化资源分配机制，通过适当的分组标识和多级分组调度可以对共享资源的进行多层次的分配。比如，在网络结点首先对端用户或下游结点进行资源分配，然后在已经分配的资源份额中按照业务类型进行再次分配。给合 H-QoS 可以使接纳控制机制对资源的分配更加准确和高效。图 5.3 所示为 Alcatel 实现的会话接纳控制与 H-QoS 相结合的系统架构。

在第一英里处，接纳控制确保用户的业务请求满足业务订购合约、用户所在位置和可用网络资源的要求。接入环路通常是静态配置供所有业务使用，用户的业务特征数据也是在业务提供时与第一英里处的资源和带宽分配静态匹配。第一英里处对具体的业务实施优先级调度，同时可以实施基于策略的接纳控制。

第二英里处应用了每用户和每业务的 H-QoS 机制。第一英里处的业务接纳控制确保用户的资源预算没有超出，第二英里处基于会话的业务接纳控制防止 VoD 业务超出其资源预算，即所有 VoD 业务的资源请求总和不超过第二英里处分配给 VoD 业务的资源。

第三英里处实施基于每结点（BSAN）和每业务的 H-QoS 机制。该结点中 VoD 流量占链路容量预算的大部分，因此接纳控制应当保证通过 BSAN 申请的 VoD 带宽不超过 BSA 的容量预算。如果第二英里的资源预算定义恰当且接纳控制有效执行，第三英里通常进行无阻塞的资源配置。

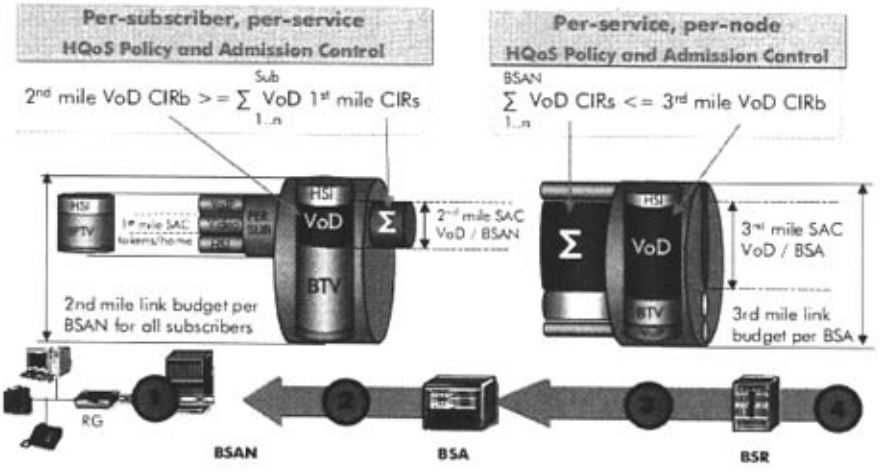

图 5.3 会话接纳控制与 H-QoS 相结合的系统架构

5.4 小结

接纳控制可以在网络资源不足时（或不满足预先定义的商务策略）拒绝新的业务请求，因而可以有效地避免拥塞的发生。一般情况下，由于难以对资源请求进行准确地描述和对资源实际情况进行准确地进行估计，实施接纳控制并非易事。然而，考虑 IP 电信网络的现状，可以判断出现阶段 IP 电信网的资源瓶颈主要在城域范围内，因而可以在城域接入接点和汇聚接点实施接纳控制，从而简化了接纳控制机制的实现和部署。本文分析了 IPTV 系统的的资源需求和网络架构，在此基础上提出了一种实现简单、易于部署的城域网视频接纳控制实现机制，为保证用户的视频业务体验质量提供了新的手段。而且，该方法可以与基于策略服务器业务管理架构集成，从而有效控制了 IPTV 业务的部署成本。

Alcatel 在全球率先倡导三重播放业务，是业界领先的 IPTV 解决方案供应商。继基于 Alcatel 7750 SR 和 Alcatel 7450 ESS 的城域网方案之后， Alcatel 又推出了基于 Alcatel 5750 策略服务器的视频接纳控制解决方案，该方案不仅可以实现基于策略的接纳控制，而且很容易集成基于资源的会话接纳控制机制，将为保证用户的 IPTV 业务体验水平、增加运营商收入提供强大的支持。

参考文献

[1] N. Tang, S. Tsui and L. Wang, A Survey of Admission Control Algorithms, UCLA, Tech. Rep. CS215, Dec. 1998. [Online]. http://www.cs.ucla.edu/~tang/papers/admission_control_paper.pdf
[2] E. Knightly and N. Shroff, Admission Control for Statistical QoS: Theory and Practice. IEEE Network, March/April, 1999
[3] Assuring Quality of Experience for IPTV, Heavy Reading white paper, July 2006
[4] Assuring Quality of Experience for IPTV-The Role of Video Admission Control, Alcatel Application Note. May 2006.

第 6 章　基于 P2P 的 IPTV 业务系统设计与分析

6.1 P2P 技术概述

P2P（Peer-to-Peer）技术是近年来国内外计算机研究界与通信技术人员关注的一个热点问题。国际上，P2P 技术引起人们的关注起源于一个非常流行的 Internet 应用 Napster。Napster 诞生于 1999 年，是当时在美国年轻人中盛行的音乐下载程序，它打破了传统的 C/S 结构，将连接在 Internet 中的端用户直接联系起来，从而使音乐下载变得更加容易，在短时间内吸引了大量的用户。尽管 Napster 后来由于在音乐版权所有者的法律诉讼中败诉而被迫停止运营，其所倡导的 P2P 技术思想得到了学术界与工业界的广泛认同与深入研究。在计算机领域，P2P 被看成是一种新的计算与应用模式，对分布式计算和 Internet 的发展产生了深远而深刻的影响。

P2P 的核心思想是“对等”，即在 P2P 系统中相互作用的双方或多方具有完全平等的关系，计算机网络的最初设计目标就是让连接在网络中的主机可以平等地自主通信。当前，客户/服务器（client/server，C/S）结构是计算机网络中主要的资源共享方式；在 C/S 结构中，服务器是资源的所有者，客户是资源的消费者，因而两者之间的关系不具有对等性。

对于计算机研究人员而言，P2P 是一种计算资源的组织形式，通过聚合网络边缘的大量空闲资源可以得到相当于大型计算机但更加廉价的计算能力；在通信领域，P2P 技术可以用于一种新信息共享与协作方式，通信系统中可以无需作为控制中心的服务器。

当前，有很多应用和服务采用了 P2P 技术。SETI@home (setiathome.ssl.berkeley.edu)有效集成了 Internet 边缘的计算能力，BitTorrent (BT)，eDonkey（电驴）等是网民中流行的下载工具，MSN，Skype，gTalk 等的即时消息与话音通信服务已经逐渐成为人与人之间一种重要的沟通方式。据统计，当前网络中的 P2P 流量已经超过网络总流量的 60%以上。P2P 应用产生的大量业务流量对电信网络造成了很大的冲击，曾经遭到电信运营商的强力反对；另一方面，P2P 技术在 Internet 领域的成功应用对电信界具有很强的吸引力，电信领域也在思考如何将 P2P 技术“为我所用”。

6.1.1 P2P 技术的发展与应用

从根本上来说，P2P 系统是为了避免存在类似于 C/S 结构中的中心结点约束而出现的一种新的计算模型。我们通过对比 P2P 计算模式与 C/S 结构可以更好地理解 P2P 的基本含义。图 6.1 所示为 P2P 与 C/S 结构的简单对比。

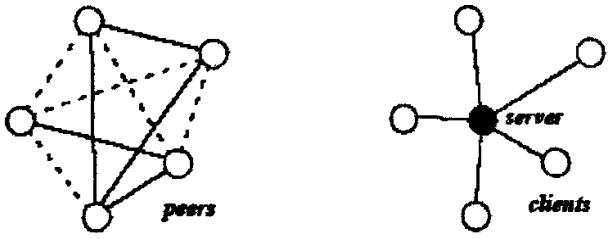

图 6.1 P2P 与 C/S 结构对照图

从图 6.1 可见，P2P 系统各结点之间具有对等的关系，它们直接交换信息和资源，相互之间的连接和所形成的拓扑具有不确定性；而在 C/S 结构中，所有的 clients 都和 server 想连才能够获取信息和资源，最终形成了一个星形结构。图 6.1 中的 P2P 系统是最基本的模型，实际应用当中 P2P 经历了一个发展过程，CacheLogic 将 P2P 系统的发展分为三代[1]。第一代 P2P 网络采用集中控制的体系结构，系统中包括维护内容和用户信息的索引服务器，典型的代表如 Napster。它采用快速搜索算法，查询响应时间短，使用简单的协议，能够提供高性能和良好的健壮性，缺点是服务容易中断。

第二代 P2P 采用完全无中心的分布式网络体系结构。它不再使用中央服务器，用户的 PC 具有多种功能，包括索引服务器、搜索本地资源以及在结点间中继查寻信息的路由器。由于每次查询都要在全网中“泛洪”，造成大量网络流量，使得其查询速度慢，响应时间长。用户的 PC 性能及其连接方式决定了网络健壮性和性能；没有中心控制服务器，也不存在单点故障失效的情况。

第三代 P2P 系统采用混合组网方式，具有层次化网络结构。混合模式综合第一代和第二代 P2P 系统的优点，用超级结点代替中央索引服务器，分布的超级结点构成一个骨干网络。超级结点负责搜集并存储端用户和可共享的内容信息，当用户登录到系统时，只与其中单个超级结点相连以获取相关信息。分层次的快速搜索改进了搜索性能，缩短了查询响应时间，并且每次查询产生的流量少于完全分布的网络。超级结点的部署提供了高性能和良好的健壮性。超级结点的失效对系统性能具有较大的影响。

P2P 技术仍然处于不断发展之中，现阶段还没有被广泛接受的严格定义。文献[2]中的定义比较全面地概括了 P2P 技术的本质特征：P2P 系统是由互连的结点以自组织方式构成的分布式系统，这些系统以共享内容、计算能力、存储能力和带宽为目的，能够适应某些结点失效并在只有一定数量结点存在时保持可以接受的连接和性能，不需要中间设施、全局中心服务器支持或它方授权。根据这个定义，传统的电话系统虽然实现了用户之间的对等通信，但由于需要大量的中间设施支持，因而不能算是 P2P 系统。MSN、Skype 等通信系统对索引服务器依赖性很强，自组织能力较差，也不是严格的 P2P 系统。

从技术发展上来说，P2P 系统可以分为三代。自从 P2P 技术出现以来，各种应用层出不穷，既有为大众所知的工具软件，如 BT、eDonkey 等网络下载工具，也有用于科学研究的实验项目，如 SETI@home，总得来说，P2P 的应用系统可以分为三类，即并行计算，内容与文件共享和协作[3]。

- **并行计算**

可并行化的 P2P 计算应用将计算量很大的大型计算任务分解成为可以在大量对等结点上并行执行的子任务。通常的应用是在不同的结点上采用不同的参数集同时执行相同的任务。另一种应用将任务分解成为粒度很小的子模块然后在大量的结点上并行执行，这种情况下每个结点执行的任务并不相同。

- **内容与文件管理**

内容与文件管理应用的目标是利用网络中的对等结点存储及检索信息。最基本的应用是内容与文件共享，用户可以直接连接到其它用户的计算机下载感兴趣的文件，如著名的 Napster、Gnutella 等。这类应用可以充分利用用户闲置的存储空间。P2P 文件系统和利用 P2P 技术进行文件过滤与搜索是另外的两种内容与文件管理应用，其目的不是共享而是在 P2P 网络中建立可搜索的索引实现协作文件过滤。

- **协作**

协作 P2P 应用允许用户无需借助中心服务器进行实时协作。即时消息应用，如 MSN masseger 已经有了大量的用户 。允许不同用户在数千里之外同时观看并编辑相同信息的共享应用也已出现，如分布式 PowerPoint [4]。游戏是协作 P2P 应用的另一个实例。P2P 游戏在对等结点上运行，更新信息可以无需任何中心服务器分发至所有结点。图 6.2 总结了 P2P 应用的分类。

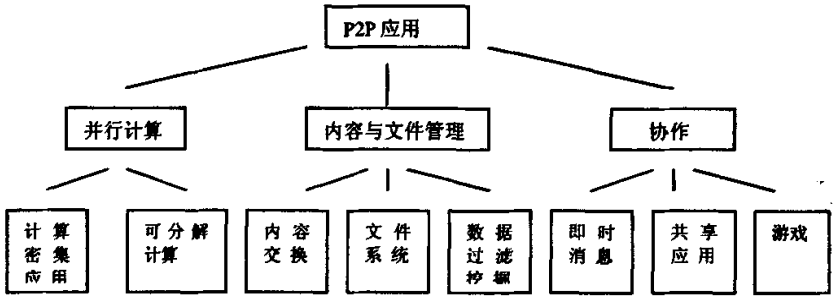

图 6.2 P2P 应用系统分类

6.1.2 P2P 内容传送关键技术

内容传送是当前 P2P 技术最重要的应用，相对于并行计算和协作应用，内容传送与人们的工作与生活关系更加紧密，因而具有更加广泛的应用基础。内容传送可以进一步分为文件共享与流媒体应用，前者用于共享或分发任何文件，后者主要是用于分发存储或实时媒体内容。使用者并不希望保存媒体内容，而是通过 P2P 网络获取内容后即时消费。为实现内容传送，P2P 网络需要具有内容路由与内容交换两个基本功能。

- **P2P 内容路由技术**

实现内容传送首先要有内容查找与搜索的机制，这种机制在 P2P 网络中又称为内容路由（Content routing）。P2P 网络是一种新型的分布式系统构建模型，系统中的对等点

（Peers）通过自组织的方式在底层网络上构成一个叠加（Overlay）网络。不同于传统的C/S 系统结构，在这种模型中，系统不能借助于中心服务器实现内容索引与查找功能。当前，P2P 的内容路由技术可以归纳为结构化与非结构化两类。

非结构化网络采用了随机图的组织方式构建叠加网络，理论上其结点度数服从"Power-law"规律，从而能够较快发现目标内容。非结构化内容路由方式在网络动态变化时体现了较好的容错能力，因此具有较好的可用性。它还可以支持复杂查询，如带有规则表达式的多关键词查询、模糊查询等，其典型代表是Gnutella采用的文件查找方式。

Gnutella 是一个 P2P 文件共享系统，它和 Napster 最大的区别在于 Gnutella 是完全分布结构的系统，没有索引服务器；在 Gnutella 分布式对等网络模型中，每一个连网计算机在功能上都是对等的，既是客户机同时又是服务器，所以被称为 Servent（英文 Server+Client 的组合）。Guntella 采用了基于完全随机图的洪泛（Flooding）发现和随机游走（Random Walker）机制。为了控制搜索消息的传输，通过 TTL (Time To Live)的减值来实现。

由于没有确定拓扑结构的支持，非结构化网络无法保证资源发现的效率。随着连网结点的不断增加，网络规模不断扩大，通过洪泛方式定位对等点和内容的方法将造成网络流量急剧增加，从而导致网络中部分低带宽节点因网络资源过载而失效。所以在初期的 Gnutella 网络中，存在比较严重的分区断链现象，即一个查询访问只能在网络的很小范围内进行，因此网络的可扩展性不好 [5]。因此内容搜索的准确性和可扩展性是非结构化网络面临的两个重要问题。目前对此类结构的研究主要集中于改进发现算法和复制策略以提高发现的准确率和性能。

非结构化 P2P 网络基于完全随机图模型，结点之间的链路构建没有遵循某些预先定义的拓扑。这类系统容错性好，支持复杂的查询，并受结点频繁加入和退出系统的影响小。但是查询的结果可能不完全，查询速度较慢，采用广播查询的系统对网络带宽的消耗非常大，并由此带来可扩展性差等问题。由于非结构化系统中的随机搜索造成的不可扩展性，构建结构化系统成为另一个选择，结构化 P2P 系统通常采用分布式散列表（DHT）的实现内容的分布式发现和路由算法。

DHT 实际上是一个由广域范围大量结点共同维护的巨大散列表。散列表被分割成不连续的块，每个结点被分配给一个属于自己的散列块，并成为这个散列块的管理者。DHT 的结点是动态的且数量巨大，因此非中心化和原子自组织成为两个设计的重要目标。通过加密散列函数，一个对象的名字或关键词被映射为 128 位或 160 位的散列值。一个采用 DHT 的系统内所有结点被映射到一个空间 $L=[0,1)$，如果散列函数映射一个 h 位的名字到一个散列值 H，则有 $H/2^h \in L$。结构化内容路由方法的一个研究重点是采用新的拓扑图构建叠加路由网络以减少路由表容量和路由延时，其基本原理是在 DHT 一维空间的基础上引入更多的拓扑结构图来反映底层网络的结构。

结构化内容路由算法都避免了类似 Napster 的中央服务器，也不是像 Gnutella 那样基于广播进行查找，而是通过分布式散列函数，将输入的关键字惟一映射到某个结点上，然后通过路由算法同该结点建立连接。由于重叠网络采用了确定性拓扑结构，DHT 可以提供精确的发现。只要目的结点存在于网络中 DHT 总能发现它，发现的准确性得到了保证。但当 P2P 结点的动态加入/退出时，DHT 需要重建以适应结点变化，带来了很大的开销。此外，DHT 的方法支持关键字查询比较困难。典型的结构化 P2P 系统包括Chord、CAN 和

Pastry等[6]。由于结构化与非结构方式路由方式都有各自的优缺点，实际应用中，可以将两种方法结合起来构成混合式内容查找方法。

- **P2P 内容交换技术**

在 P2P 系统中，当某个结点通过路由机制搜索到所需的内容，即可从内容拥有者那里获取内容，这个过程称为内容交换。早期的 P2P 系统中，如 Napster，内容交换非常简单，只需在内容的所有者 O 与请求者 R 之间建立直接连接进行文件传输，直到 R 得到了全部内容后或网络发生故障连接才会终止。由于在 P2P 系统中，某个文件通常有多个副本，采用唯一连接的方式不能发挥其它内容源的作用，多径并行下载就可以很好地解决这个问题，大大提高了内容共享的效率。

动态地并行访问重复内容最早用于具有多个内容镜像的客户/服务器系统 [7]，由于 P2P 系统存在类似的多源特性和应用场景，因而被引入到 P2P 系统中。在后来的 P2P 文件下载和流媒体应用，如 eDonkey2000/eMule，BitTorrent 及 CoolStreaming 和 PPLive 都采用了这种方式。实践表明，采用这种方式可以大大提高内容共享的效率。

为了实现并行内容共享，需要对内容数据进行适当的分割以保证并行地从不同对等结点获取部分内容。当前，通常把内容分成长度相同的数据片段，为了可以在接收端进行准确地重组，需要对数据片段按照原来的位置顺序进行编号。内容分块带来的附加的好处是，即使是某个内容源没有全部内容，也可以利用已有的内容片段为其它结点提供服务，进一步扩大了内容源的范围。但同时，由于内容片段的数量较多，查找或定位相对困难，并带来了一定的网络负载。

实现并行内容共享需要解决的另一个问题是内容接收者决定何时从哪个内容源选择某个内容分片，即分片调度机制。并行调度是经典的 NP 难解问题，通常只能通过启发式算法求解。然而，分片调度机制对于内容共享效率和系统功能实现具有重要的影响。BT 采用了本地最称缺优先（Local Rarest First，LRF）进行内容分片调度，研究表明这种方法可以高效地实现文件下载 [8]。

6.2 IPTV 的业务提供方式分析

作为一种新的电信业务，IPTV 在国内的发展面临着政策、商业和技术等方面的风险。在技术层面，网络容量和结构的可扩展性是一个重大的挑战。由于视频信息需要消耗大量的带宽资源，且流媒体内容需要实时播放，与传统的话音与数据业务相比，IPTV 资源需求要大得多。现在正进行的 IPTV 试点项目只支持数千用户，而大规模部署需要支持百万数量级用户，对网络容量带来前所未有的挑战。此外，网络服务质量的保证、视频编码标准的选择、内容的安全性保障以及用户体验质量等问题都是 IPTV 系统亟待解决的技术问题。

随着高速路由设备、城域汇聚与接入设备的发展与成熟，承载网络的容量已经有了很大的提高。在城域网内已经达到 10Gbps 的量级，核心网的超大容量路由器也有相应产品可选。相对而言，宽带接入技术有所落后，但高速 DSL 和无源光网络接入技术有望提供大容量的接入带宽。然而，根据美国 AT&T（原 SBC）公司的部署经验，为了支持高质量的三重播放业务，每个家庭需要 20M 以上的带宽。尽管承载网络的容量已经很大，还必

须进行合理的利用才能在成本和质量两方面取得更好的平衡。当前，为了提供大规模的IPTV业务，主要有三种技术可用于提高网络容量的扩展性，分别是IP组播，CDN技术和P2P技术，下面分别对上述技术进行分析。

6.2.1 IP组播

IP组播中，接受相同内容的所有用户被分成一个组播组，并共享相同的IP组播地址。借助于路由协议，IP组播可以在网络中间结点实现数据报文的多点复制，从而避免了每一个用户都要到最初的源结点获取数据，在用户数量较多的情况下，可以大大节约网络带宽的消耗。IP组播的提出最初是为了实现Internet中的多媒体会议系统，并通过MBone实验网进行了验证。当前，在IPTV业务提供方案中，IP组播技术是重要的组成部分。

然而，IP组播组中的用户可以动态地加入和退出，用户的组管理增加了路由设备的复杂性和负担，因而其本身具有可扩展性问题[9]。其次，IP组播遵循Internet设计中路由与传输分离的原则，只提供尽力而为的服务，为组播提供服务质量保证比单播要复杂得多。此外，由于实现IP组播需要对IP路由协议进行修改，部署IP组播需要替换原来的IP路由器，因而会带来很大的经济负担。事实上，IP组播技术已经有超过15年的历史，但并没有得到大规模的应用。

相对于核心网络，城域网络的带宽资源更加紧张，而且由于城域网通常需要新建，因而在城域中部署减化的IP组播技术不失为一个可行的方案。为了实现城域组播，需要相应的接入汇聚、城域汇聚以及IP路由设备都支持组播能力，因而实现组播会带来很大的成本负担。由于组播技术存在可扩展性问题，它也不适合在核心网中部署。此外，组播技术只能用于广播应用，对点播能够提供的帮助有限，因而并不能完全解决IPTV系统的容量可扩展性问题。

6.2.2 内容传送网络

IP组播是一种承载网技术，具有较好的通用性。但由于IP组播的部署成本很高，还需要其它技术做为补充。应用层组播技术是近年来出现的一种替代方案。应用层组播不需要对IP层的协议进行修改，可以根据需要有选择性的实施，因而具有良好的可部署性。内容传送网络（CDN）技术可以看作是一种运用于多媒体传输的应用层组播技术，已经在Internet多媒体应用中得到了广泛的应用。

CDN是一个叠加在骨干/城域网络之上的应用系统，其主要任务是将节目内容由中心服务器推送到尽量靠近用户的边缘服务器中，用户终端依据就近原则和负载均衡原则选择尽量接近用户的边缘服务器获取节目内容。目前的CDN大都依据"80/20"规律进行规划，即80%的用户可在边缘服务器直接得到节目，而边缘服务器只存贮占总体20%的热点节目。边缘服务器一般连接在靠近用户的城域设备上，由于终端与服务器之间的路径较短，可以大大提高响应速度；媒体流量不需经过骨干网络，从而节约了骨干网资源，提高网络容量的扩展性；媒体内容均衡分布到大量边缘服务器上，有效地减少网络和服务器发生拥塞的可能。

相对而言，CDN是一个比较成熟的技术，国内外有很多集成商可以提供相关方案。由于内容已经存储在边缘服务器中，点播应用所消耗的网络资源也很有限。尽管CDN可用

于支持媒体广播，但更多是用于 IP 组播进行补充以支持点播业务。部署 CDN 虽然没有技术上障碍，由于需要购置大量的服务器和管理软件，其成本投入从网络层转移到应用层，仍然是不容忽视的问题。

6.2.3 P2P 技术

P2P 技术是近年来新兴的内容分发技术。P2P 技术充分利用终端的资源，每个对等结点既是内容的获取者，同时又是内容的提供者，为其它结点提供服务。P2P 系统最初用于文件共享，但其内容交换特性也可用于 IPTV 的流媒体应用。PPLive 是 Internet 中比较成功的 P2P 流媒体应用[10]，图 6.3 所示为 PPLive 的基本工作原理。需要获取内容的结点（终端 1）先到频道列表服务器中得到所需的频道，然后登录到结点列表服务器中获取正在观看同一频道的其它结点列表，这样该结点就可以和列表中的其它结点进行数据交换，观看感兴趣的频道内容。

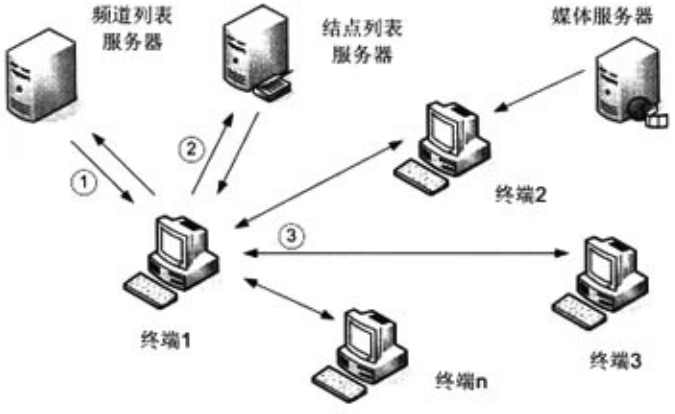

图 6.3 PPLive 系统工作原理示意图

除了使用单纯的 P2P 技术，还可以将传统的 CDN 技术与 P2P 技术结合起来。在 CDN 中，由中心服务器向边缘节点分发节目内容一般采用传统的 FTP 技术，节目内容采用文件格式或者分片文件格式。目前的一种趋势是采用 P2P 技术构建 CDN 网络，将节目内容预先进行流化处理后分块存贮在多个边缘服务器中，由调度服务器按照就近原则、负载均衡原则进行集中控制，并可基于网络状况实时选择和切换向用户提供流服务的边缘服务器。图 6.4 是一个基于 P2P 的 CDN 网络结构示意图。采用这种分布式存贮和集中控制的 CDN 结构，用户观看一个点播节目时通常是由多个边缘服务器协同完成流服务，可以在用户集中点播时将负载在整个 CDN 内部进行更加合理的分布，避免过于集中在某个边缘服务器中造成拥塞。采用基于 P2P 的内容存储与检索技术，可以有效地降低 CDN 系统的存储成本。

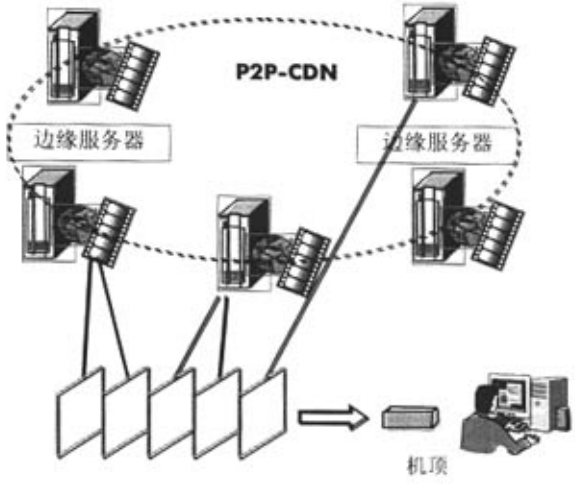

图 6.4 基于 P2P 的 CDN 网络结构示意图

6.2.4 IPTV 系统的可扩展性问题与 P2P 技术的优点

下一代网络的目标是多业务网络，在众多的业务中，IPTV 以其三重播放能力可以有效地进行业务绑定，从而提高用户的忠诚度，保证电信运营商的稳健运营。然而由于 IPTV 中视频业务与传统电信业务相比需要大量的通信和计算资源，在 IPTV 业务实际部署过程中，提高网络容量与业务系统的可扩展性成为一个具有挑战性的基本问题。当前，IP 组播、CDN 技术以及 P2P 技术都有助于提高 IPTV 系统的可扩展性。

P2P 系统的最重要的特性是具有自扩展性，即参加的结点越多，系统的容量就越大。在 P2P 系统中，对等结点不需要到源服务器就可以得到所需的内容，因而服务器的连接带宽和处理能力不再是系统的瓶颈。由于 P2P 系统通过终端软件实现，因而也是一种应用层组播技术。和 CDN 相比，区别在于 P2P 网络不需要部署任何用于内容转发的服务器，因而具有良好的经济性。但是 P2P 系统不关心底层网络连接服务质量，因而对等结点间的连接可以无法满足流媒体的实时性要求。此外，P2P 系统中结点可以自由地加入和退出，因而系统的容量不稳定，难以提供可靠的性能。

P2P 技术是网络经济中出现的新技术，可以支持一些新的网络应用和服务。P2P 技术在文件共享方面的成功应用已经表明了其独特的技术优势，P2P 流媒体应用也在 Internet 中出现了很多的案例。尽管 IP 组播结合 CDN 的方式是当前的主流方案，但 P2P 技术具有明显的成本优势，在用户活跃程度较高的地区或时段可以发挥作用，作为一种新的 IPTV 业务提供方式值得进一步研究。

6.3 基于 P2P 技术的 IPTV 系统架构分析与设计

P2P 内容共享系统架构自下而上由多个层次构成，其中结点管理层的主要功能是资源的发现以及内容的搜索与定位 [6]，是 P2P 系统的基本功能，也是 P2P 技术研究与系统设计的重点。根据所采用的内容路由技术，可以将 P2P 系统分成结构化与非结构化两类 。结构化系统中 P2P 叠加网络拓扑被严格控制，内容不是随机分布于结点中，而是具有明确的位置，从而可以提高系统资源查询的效率。结构化 P2P 系统采用分布式哈希表作为系统的底层支撑，数据对象（值）的位置信息确定存放在某个结点，该结点的标识与数据对象的唯一键值相对应。因为基于键值的路由具有良好的可扩展性，结构化 P2P 网络可以高效地定位数量稀少的数据对象。

然而，由于 P2P 网络的结点具有很强的动态性，维护结构化的 P2P 网络会带来很大的负担。考虑到 P2P 共享主要针对流行的内容，在今天的 Internet 中，去中心的非结构化 P2P 叠加网络应用更加广泛。Gnutella 是典型的非结构化 P2P 网络。Gnutella 是完全无中心的网络，其共享内容搜索采用简单的查询消息泛洪方式。这种方式实现较为简单，但由于搜索的对象位置不明确，对于网络中存在较少的数据对象查询时延长，而且大量的查询与应答消息使网络规模难以扩展。在 C/S 结构系统中，由于内容在服务器集中索引，查询效率较高。容易想到在 P2P 网络中增加目录服务可以大大提高查询的效率。

为了实现目录服务，需要在网络中引入集中点。纯 P2P 网络假定网络中的每个结点都是相同的（从能力到功能）。所谓混合式 P2P 网络是指网络中存在一些结点有别于其它结点，通常它们是一些分布的局部中心点，这些结点为与之相连的其它普通结点提供目录索引或资源管理服务。由于混合式结构 P2P 系统集合了 C/S 系统的优点，因而是一种更加合理和实际的 P2P 网络架构。

当前，比较成功的 P2P 应用系统大多采用了混合式架构，如 BitTorrent，KaZaA 等。然而，现阶段的混合式 P2P 系统都是基于私有的技术，公开的技术信息很少。本节对几种典型的混合式 P2P 应用系统架构进行了比较分析，尽管它们的主要功能是实现文件共享，但可以从中深入理解混合式架构的特点，并将 P2P 系统架构的的基本结构与设计重点运用 IPTV 系统设计中。

6.3.1 混合结构 P2P 系统概述

分布式系统架构的两个极端的情形是纯 P2P 结构和 C/S 结构。纯 P2P 架构是完全去中心的，对等点是高度自治的平等实体。纯 P2P 网络假定各对等点具有相同的能力。事实上，在 P2P 网络中，参与的结点在存储能力、处理能力、带宽及在线时间等方面具有相当大的差异。为了使系统的运行更加稳定，功能更强，有必要利用这个差异，让连接可靠、能力较强的结点完成一些重要的基本功能。为了避免系统的查询消息风暴，提高系统的可扩展性，这些结点通常提供局部的目录服务。相对于普通结点，这些结点被称为超级结点（Super-peer 或者 Ultra-peer）。在有些系统中，超级结点的角色由专用的服务器充当，这些服务器专门负责共享文件索引和查询管理。这类 P2P 系统融合了 C/S 结构的特征，被称为混合式（hybrid）P2P 系统。KaZaA 是混合式 P2P 系统的典型代表。

混合式 P2P 系统架构最明显的特点是其层次化结构。超级结点或专用服务器构成一个层次，普通结点与超级结点相连，构成另一个层次。超级结点相互连接组成一个骨干叠加网络，并且可通过应用层广播协议在此叠加网络上实现分布式查询服务。超级结点是一个

本地的索引中心，为与之相连的其它结点建立共享的文件索引，当查询请求超出本地索引的范围时，它还要代表查询结点向其它超级结点转发查询消息。

系统的层次化结构结合了集中式和纯 P2P 系统的优点。新层次的引入增加了查询消息转发的范围和效率。和普通结点相比，超级结点加入和离开的频率要低些，系统的结构更加稳定。而且，基于超级结点的内容路由机制避免了大量的广播消息，从而明显减少了网络中查询消息流量。

6.3.2 典型混合结构 P2P 系统分析

P2P 系统的基本功能包括 P2P 网络构建，如结点的加入和退出协议等，和内容/用户的搜索与定位机制。良好的可靠性与安全性是对 P2P 系统的更高要求。根据系统的具体应用，系统架构还需要考虑其它问题，比如在文件共享系统中，如何提高共享的速率以及用户的激励机制等。相应地，P2P 网络体系架构需要考虑的问题包括网络的自组织构建方法、对象的定位及路由机制，需要设计的协议可以分为普通结点间、普通结点与超级结点间以及超级结点之间三类。混合式 P2P 网络通过这些协议实现内容搜索定位与共享功能。

本小节的目标是分析混合式 P2P 系统的体系架构和基本功能，将主要围绕 P2P 网络构建以及搜索与定位机制等方面对典型的混合式 P2P 文件共享应用系统进行分析。

■ **KaZaA**

KaZaA（www.kazaa.com）是一种 P2P 文件共享应用的客户端软件。无论从活跃用户数量还是从产生的流量来看，KaZaA 都曾经是最重要的 P2P 应用之一。KaZaA 基于 FastTrack 技术，是最早利用结点间差异的 P2P 系统。图 6.5 所示为 KaZaA 系统的基本结构图。 图中的对等点被分成两类，超级结点（Supper Node，SN）和普通结点（Ordinary Node，ON）；SN 以部分网状的方式相连构成 KaZaA 网络的上层，ON 分别连接到不同的 SN 上构成底层。SN 通常在接入带宽、处理能力等方面具有优势；SN 的另一个选择依据是结点不需经过网络地址转换（NAT）连接到广域网中。在 KaZaA 中，每个 ON 都有一个父亲 SN。

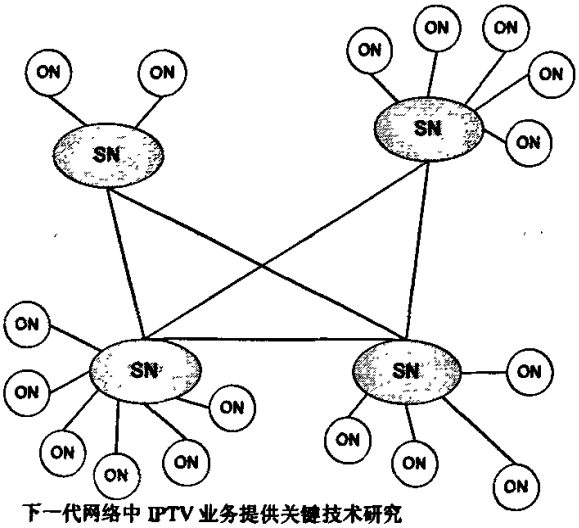

图 6.5 混合式 P2P 系统基本架构示意图

每个 SN 维护一个本地索引，其中包括所有与之相连的 ON 的文件共享信息，因此，每个 SN 就象是一小的索引服务器。但 SN 不是专用的服务器，只是属于某个个人用户的普通终端，包括 ON 的所有功能。为了扩大查询范围，文件索引在 SN 之间进行分发。当用户试图搜索某个文件时，ON 先通过 TCP 连接发送带有关键字的查询消息到父亲 SN，如果匹配成功，SN 将反馈匹配文件对应的 IP 地址、服务端口号以及元数据。每个 SN 还与其它 SN 维持长期的连接，当 SN 收到查询消息时，它会把该查询转到其它与之相连的 SN。当客户端得到相匹配的查询结果后，就可以直接与该源文件所在结点建立连接并下载文件。KaZaA 采用简化的 HTTP 协议实现文件下载；由于 KaZaA 采用的是私有的协议，没有详细的公开信息描述 KaZaA 的文件下载过程。为了便于进行用户管理，当客户端试图连接到 KaZaA 网络时，它先要到一个中心服务器去注册。注册服务器是系统中唯一的集中点。

在 KaZaA 系统中，超级结点构成网络的骨干层，普通结点不具备独立工作的能力，因而当结点加入网络时必需先与某个超级结点相连。当 ON 启动 KaZaA 客户端时，第一个任务就是选择一个超级结点并与之建立和维持一个半永久性的 TCP 连接，并将它正在共享的文件信息上传到该 SN。文献[5]通过流量监测，发现了 ON 与 SN 建立连接的步骤：

- ON 维护一个 SN 列表缓存。ON 启动时从列表中选择多个（通常为 5 个）候选 SN，通过发送 UDP 分组测试每个候选对象。如果候选对象当前处于活动状态，ON 可以收到其发回的 UDP 反馈分组；
- 对于每个发回反馈分组的 SN，ON 试图与之建立 TCP 连接。通过这个连接，SN 和 ON 交换加密密钥。然后 ON 向 SN 发送结点信息，SN 则向 ON 发送更新的 SN 列表。
- ON 从中选择一个 SN 并与其它 SN 断开连接。这个保持连接的 SN 成为 ON 的父亲 SN。
- 当需要下载文件时，ON 发送查询消息到选定的父亲 SN。

如前所述，SN 之间维持长期的 TCP 连接，这些连接除了用于转发查询消息外，还用于发送 SN 更新消息。ON 的列表信息来源于 SN，为了保证 ON 可以连接有效的 SN，需要维护 SN 的状态信息并对之及时更新。KaZaA 通过在相互连接的 SN 之间交换各自维护的 SN 列表保持整个系统的 SN 状态更新。KaZaA 对信令交换采用了加密措施，ON-SN 及 SN-SN 在开始交换信息之间先进行密钥交换，从而可对后续的信息交换进行加密。

在 KaZaA 中，当 ON 试图与 SN 建立连接时， ON 并不是随机地选择 SN，而是遵循一定的规则。根据现有测量结果，KaZaA 主要使用两条准则来进行 ON-SN 以及 SN-SN 之间的邻居选择：流量负荷和本地性 [11]。流量负荷可以用 TCP 连接数近似，本地性可以用 RTT 进行度量，即邻居选择时会优先考虑 TCP 连接数较少的结点和 RTT 较小的结点。流量负荷原则可以实现结点的负载均衡，而本地性原则可以减少流经骨干网的流量。

良好的 NAT 穿越能力是 KaZaA 的重要特色，ON 需要借助 SN 实现 NAT 穿越，本文不详细讨论其工作原理。Skype(www.skype.com)由 KaZaA 的创建者开发出来，虽然两者的应用目标不同，但系统架构有很多相似之处。

■ **eDonkey/eMule**

eDonkey(www.edonkey2000.com，因为版权问题，该网站已不再提供服务)是继 KaZaA 之后另一个流行的文件共享应用，eMule（emule.sourceforge.net）是 eDonkey 的开放源代码版本，两者差别不大。eDonkey 网络由数百个服务器和数百万的客户端组成。服务器构成网络的一个层次，作为索引服务器用于文件定位和向客户端分发其它服务器的地址；所有的客户端构成网络的另一个层次，用于共享和下载文件，文件传输无需通过任何服务器。每个用户都可以建立索引服务器。图 6.6 显示了 eDonkey 系统架构的示意图。

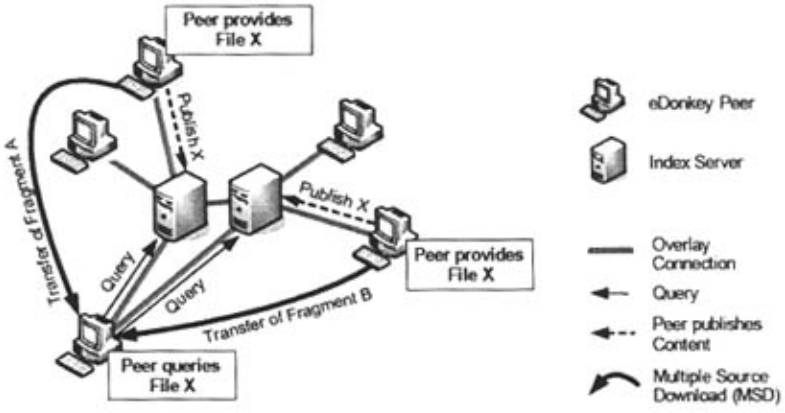

图 6.6 eDonkey 系统架构示意图

每个 eDonkey/eMule 客户端预先配置了服务器列表和共享文件列表。客户端通过一个 TCP 连接与服务器建立连接并注册到网络中，获取所希望得到的文件信息和可用的对等结点信息。

服务器不保存任何文件，作为中心索引点只保存文件的位置信息。服务器曾经有类似于 KaZaA 中辅助穿越防火墙的功能，但这个功能后来没有被保留，因为这种桥接功能大大增加了服务器的负担。服务器提供局部的中心索引服务，客户端需要连接到服务器上以获得文件共享服务。只要系统中有客户端存在，则服务器的连接保持开放。在 eDonkey 中，服务器之间有少量的通信，在 eMule 中，服务器之间不相互通信。此外，eDonkey 的客户端还相互交换服务器地址信息。

eDonkey客户端有文件愿意共享时，它将向某个索引服务器发布共享信息。每个服务器保存与其相连的客户端的所有共享文件列表。当客户端搜索一个文件时，它向主服务器发送查询消息，服务器反馈匹配的文件及所在地址列表。如果服务器没有反馈或匹配的文件数量较少，客户端可以向其它服务器再发送查询消息。

为了提高下载的效率，eDonkey 将文件分成独立的块（chunks）。一个文件块的大小接近 10M。下载客户端可以通过组装所得文件块得到完整的文件。为了保证文件块的正确性，对每个文件块计算了 MD4 校验。校验值可以在客户端之间按需传递。一个客户端收

到文件块并校验正确以后即可用于共享。通过分块机制，eDonkey/eMule 可以实现多源并行下载，并且不需要源结点有完整的文件，大大提高了下载的效率。

当eDonkey的客户端需要下载文件时，它通过查询消息获得所有可能的文件提供者列表，然后向文件提供结点申请上载时间片。当文件提供结点收到下载请求时，它将该请求置于上载队列中。当该下载请求获得可用的时间片时， 文件提供客户端发起TCP连接到文件请求客户端，协商下载的文件块并传输数据。

eDonkey的网络协议是私有的，并没有公开的技术文档。然而，很多技术细节被自由软件的开发者探索出来，如通信协议及端口等。eMule是最著名的eDonkey自由软件版本。EDonkey/EMule实现文件共享的协议包括服务器-服务器，服务器-客户端以及客户端-客户端之间的通信 [12]。

在文件下载之前，每个客户端必须连接到一个服务器。客户端与服务器之间的通信一般采用TCP，为了提高系统的鲁棒性，UDP也可以用于消息交换。客户端启动时，它首先在可用服务器的侦听端口上打开一个socket连接，并通过这个socket连接向服务器发送消息。服务器在客户端的侦听端口上打开一个临时连接以检查客户端是否可以收发文件，然后关闭这个连接。如检查成功，服务器记录客户端的存在，并通过最初的连接向客户端发送消息，通知它所知道的资源情况。类似地，客户端向服务器发送消息通知它愿意共享的文件。

文件块传输也通过在客户端间交换一系列的消息实现。这些消息通过客户端间的 TCP 连接交换，该类连接只在文件下载期间被维护。客户端首先给服务器发送一个消息以找到存储文件的结点，然后它在一个或多个对等结点的侦听端口上打开 socket 连接，并询问它们所存的文件块。确定所需的文件块存在时，它将发送文件块请求到对等客户端，并接收请求的文件块。整个文件下载通过把文件块连接起来组装完成。

在 eDonkey 中，每个服务器注册其它服务器的存在信息，这样可以用于文件搜索。eDonkey 服务器之间的通信非常有限。服务器之间以很低的频率周期性地通告自己的存在并发布其它服务器的列表信息。通过这种方式，eDonkey 服务器可以维持经常更新的工作服务器列表，从而提高搜索的效率。

- **BitTorrent**

BitTorrent（www.bittorrent.com，简称 BT）是当前国内最流行的 P2P 文件共享系统。根据 CacheLogic 的统计，2005 年，在国内 BT 所产生的流量相当于其它 P2P 文件共享应用产生的流量总和。BT 客户端是共享内容的所有者，同时也是共享的请求者。除了大量的客户端以外，网络中还存在一些跟踪服务器（Tracker）用于管理共享文件的下载，其作用类似于 eDonkey/eMule 网络中的服务器。因此 BT 网络也采用了混合式的系统架构。

不同于 KaZaA 和 eDonkey/eMule，BitTorrent 本身没有提供搜索机制，而是通过目录网站实现内容查找。共享文件的拥有者将文件的共享信息发布到公开的目录网站上，如果用户希望获取某个文件，则需要到相关的网站去查找最近发布的共享信息。当前，国内用户常访问www.btchina.net获取文件共享信息。

文件共享者发布的共享信息保存在后缀为.torrent 的数据文件中，该文件的一个重要内容是提供管理文件共享的跟踪服务器地址。用户获得了跟踪服务器的地址后，即可与跟踪服务器建立连接。因为每个文件共享者都与跟踪服务器建立通信连接并交换信息，跟踪服务器知道所有共享/下载该文件的客户端的文件共享信息。当它收到某个客户端的文件下载请求以后，将从共享结点集合中按照一定策略选择一个子集作为响应发送给请求用户，该用户即可直接与这些结点建立连接获取文件内容，同时它也作为一个新结点加入到这个集合中。在 BT 中，一个文件共享会话被称为 torrent，共享信息的最初发布者称为 torrent 的种子结点。当新加入的结点下载了部分内容以后也可能成为内容源，为会话中的其它结点服务。

在 BT 系统中，每个跟踪服务器可以维持多个 torrent。虽然跟踪服务器可能有多个，但它们之间并不相互通信。一个跟踪服务器可以同时跟踪多个文件的下载，不同的.torrent 文件对应的相同的数据文件可以连接不同的跟踪服务器。通过上面的分析，我们可以看出，BT 系统包括客户端软件，.torrent 文件制作工具（可以集成到客户端中），发布.torrent 文件的网站及跟踪服务器。图 6.7 描述了 BT 系统的工作流程。

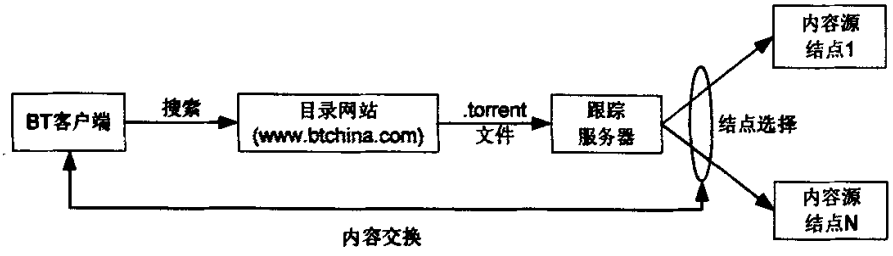

图 6.7 BitTorrent 系统的文件共享流程

尽管 BT 没有内在的搜索功能，但是由于其采用了适当的激励机制和并行下载机制，内容共享的效率很高。在 BT 中，共享的文件被分成“块（piece）”，每个文件块的大小通常为 256K 字节。种子结点为每个文件块计算 SHA1 散列值并将其存入.torrent 文件中；当接收结点收到文件块时，可以根据散列值检验数据的正确性。每个文件块被进一步分成“片（block）”，每片的大小通常为 16K 字节。接收结点可以同时从多个结点接收同一块文件的多个片。但只有收到完整的文件块并检验正确以后，接收者才能共享文件块中的各个片。与 eDonkey/eMule 相比，BT 的文件块要小得多，因而其共享也更加灵活。

当BT的客户端与潜在的源结点建立连接时，基于它所了解的文件块分布情况，它需要确定从哪些源结点获取哪个文件块。BT采用了最稀缺优先的策略，即根据邻居结点的文件块选择最稀缺的文件块下载[13]。最稀缺内容优先分发的策略使网络中保存尽量多的不同文件块供其它结点下载。

与eDonkey/eMule不同，BT对结点间的通信采取了一定的控制措施。在BT系统中，文件共享者通过下载和上传文件块来和其它对等端进行交换，这种方式可以有效地避免不劳而获者的存在，是一种有效的激励机制。每个对等点通过连接适当的对等点以使其下载速率最大化，那些具有高上传速率的结点更有可能获得高的下载速率。这个机制被称作Tit for Tat，BT通过阻塞（Choke）算法来实现这个机制[7]。BT采取了一个公平的策略防止搭便车者（free rider，只下载不上传的结点）。特别地，在任何时间时刻，BT选择四个贡献

最大的结点解除阻塞，解除阻塞的评估每10秒进行一次。因为种子结点不下载任何数据，它无法根据对等点的上载能力做出解除阻塞判断，而是根据其下载能力。

除了基于过去上传性能选择四个结点解除阻塞，BT还结合了“乐观解除阻塞机制（Optimistic Unchoking mechanism），即允许BT客户端不根据对等点以前的上载性能对其解除阻塞。解除阻塞的对象以循环的方式进行选择，其间隔为30秒。乐观解除阻塞允许一个新的BT结点有机会接收文件块并参与贡献。此外，这个机制也给那些已经建立稳定对等关系的结点一个机会找到新的连接，这些连接可以提供更好的下载带宽。

6.3.3 比较分析

我们从系统架构和功能两个角度对上述三种 P2P 系统进行比较分析。

三种系统都采用了混合式的层次化架构，从而具有可扩展性好、搜索效率高的特点。KaZaA 利用了对等点之间的异构性，选择能力较强的结点构成网络系统的骨干，而 eDonkey 和 BT 都采用主动部署服务器作为超级结点的方式。KaZaA 的方式降低了系统的设备成本，但牺牲了系统可靠性，同时也增加了软件复杂度。KaZaA 具有良好的 NAT 和防火墙穿越能力是其它两个系统所不具备的。

为了实现文件搜索，普通结点都需要借助于超级结点或服务器。由于没有内在搜索机制，BT 的客户端与跟踪服务器之间通信协议要简单一些。由于 eDonkey 和 BT 都采用了内容分片机制，因而内容源的管理更加复杂，也增加了普通结点与服务器之间的通信协议的复杂性。在 KaZaA 中，超级结点间的通信是完成搜索功能的必要组成部分，因而需要相应的协议。eDonkey/eMule 中，服务器之间的通信仅仅是为了交换可用服务器信息，BT 的跟踪服务器则是完全独立工作，无需任何通信。

文件下载都是在对等结点之间进行，无需通过服务器，对等点之间的通信协议是必需的。由于 BT 采取了独特的激励机制，因而除了内容交换协议以外，而需要统计内容交换的状态信息并以此实现激励机制。

相对而言，KaZaA 是一种更加完善的分布式系统架构。由于采用了加密手段，因而其协议最不透明。eDonkey 有开源版本，BT 本身即是开放源码的，因而这两种系统更加容易理解。

这三个系统都提供文件共享的服务，其基本功能包括内容搜索、结点选择以及内容交换。共享内容发布是搜索的前提，在三种系统中，内容源结点都需要向超级结点或服务器发布其共享信息，而 BT 采用了一种间接的方式，通过发布于目录网站的.torrent 文件指向作为汇聚点的跟踪服务器。

内容源，即对等结点的选择是内容路由的基本功能，在 KaZaA 中，对等点的选择考虑了结点的负载能力和本地性。BT 则设计了复杂的激励机制，相比于 KaZaA 和 eDonkey，BT 的结点选择机制最为复杂。相对而言，eDonkey/eMule 没有太多的限制。由于分片机制的引入，eDonkey 和 BT 都是以文件块为单位进行数据交换，多源下载是两者共同的特征，但 BT 的内容分片更小，因而也更加灵活，但增加了内容管理的难度。在

KaZaA 中没有明显的证据表明它实现了内容分片和多源下载的方法。表 6.1 总结了上述的比较和分析。

表 6.1 三种 P2P 文件共享系统比较表

	系统架构	内容发布	内容搜索	结点选择	内容共享
KaZaA/FastTrack	超级结点与普通结点	共享结点向超级结点注册	全局索引机制，借助于超级结点实现全局搜索	考虑了结点的负载均衡和本地性	对等点之间直接通信
eDonkey/eMule	客户端与索引服务器	共享结点向索引服务器注册	搜索索引服务器	无额外约束	内容分片，多源下载
BitTorrent	客户端与跟踪服务器	将.torrent 文件发布于目录网站	指定跟踪服务器以会话为单位进行共享管理	结合激励机制，优先分发系统中稀缺内容片	内容分片更小，多源下载

KaZaA，eDonkey/eMule 和 BitTorrent 都是非常成功的文件共享应用，对应其出现和盛行的时间，可以认为它们代表了文件共享应用的三个不同阶段。总的发展趋势表现为：网络结构越来越松散，内容下载的效率越来越高。这表明了 P2P 系统的发展从重视网络构建向重视应用性能的转变。

高性能P2P系统的基本需求是可扩展、可靠和高效。从上面的分析可以得出P2P系统设计的两个基本基本原则：混合式层次化结构是提高系统可扩展性和可靠性的重要手段，同时也有助于提高系统的搜索效率。结合内容分片机制的多源下载可以充分利用网络的空闲处理能力和带宽资源，从而可以实现高效的内容共享。

P2P 系统和应用的出现为通信与信息系统的设计提供了新的思路，系统架构的选择对系统的功能实现具有重要的影响。混合式层次化结构是 P2P 系统发展的最新阶段，具有很多明显的优势和良好的应用前景。本节比较了 KaZaA、eDonkey/eMule 和 BitTorrent 三种典型的 P2P 文件共享系统架构，分析了其各自的特点，总结了高性能 P2P 系统的基本设计原则。

流媒体是 P2P 应用的新领域，尽管已经存在一些基于 P2P 的流媒体应用系统，如 CoolStreaming [14]，但这些系统并没有被广泛应用和认同。混合式结构 P2P 系统在文件共享应用中取得了很大的成功，可以相信，混合式 P2P 系统结构同样适用于流媒体应用。改造混合式 P2P 系统架构以适应 IPTV 系统是另一个值得研究的问题。我们认为，基于 P2P 技术的 IPTV 系统可以继承 P2P 文件共享应用的内容路由技术，而需对内容交换技术加以改造。

P2P 应用的广为流行产生了大量的流量，对运营商的基础通信设施带来了很大的冲击。由于受到计费方式的限制，运营商在增加扩容成本的同时并不能从中得到收益，因而通常对 P2P 的发展和应用采取敌视的态度。除了在 KaZaA 中考虑了 P2P 网络的拓扑信息，eDonkey/eMule 和 BT 并没有关心底层网络的使用。在混合式架构 P2P 系统中实现流量本地化机制是实现双方共赢的一个关键因素。

6.4 P2P 应用流量的管理与控制

长期以来，电信运营商在发展 IP 业务时，通常仅仅提供一种接入手段，用户从运营商那里获得的只是没有差异化的端口速率，即便是带宽一再提高，却难以为运营商带来如话音时代的丰厚利润。相反，接入宽带化使得即时消息（Instant Messaging, IM）、P2P 文件共享等一批替代性极强的 Internet 服务大量涌现。对运营商而言，投资建设宽带网便成了“双刃剑”，为用户提供了新服务的同时，却让部分传统业务量悄然流失。全球电信运营商现在正普遍面临一个难以破解的问题，作为利润主要来源的话音业务增长缓慢，高速增长的 IP 业务却“增量不增收”。电信运营商面临的现实情况越来越严峻：带宽增长快于业务量增长，业务量增长快于成本增长，成本增长又快于收入增长 [15]。

P2P 应用的发展非常典型地反映了当前电信业的困境。现在 P2P 应用的发展可谓突飞猛进，已经逐渐渗透到各个主流应用中，从音乐下载到在线游戏，从 VoIP 到分布式计算与协作工具，P2P 的身影随处可见。据欧洲的统计数据，迄今为止，P2P 成为网络资源的最大消耗者，其产生的流量占本地网络流量的 80%以上，占骨干网络流量的 60%以上 [16]。由于 P2P 的技术特性，它使得宽带通信产业价值链缩短，并且绕开了宽带运营商。除了增加宽带运营商的成本负担，P2P 不仅不能为宽带运营商带来额外的收益，而且还占用网络资源，从而影响宽带网络正常的运营，给宽带运营商的正常业务带来了巨大的冲击。随着 P2P 业务量的增加，网络拥塞和安全隐患等问题层出不穷，严重影响了其它关键应用的普及，也使得客户对宽带接入的抱怨率大幅增加。

现阶段，对网络运营商而言，P2P 应用在刺激宽带接入市场的同时带来了更多的负面影响，特别是对网络带宽的大量消耗大大增加了网络建设与维护的成本。尽管当前还没有找到统一和合理的应对策略，对 P2P 应用和流量进行管理与控制是当前网络运营商和服务提供商普遍关心的问题。如果采用 P2P 技术建设 IPTV 系统，对流量的控制和管理是决定其成败的重要因素。

6.4.1 P2P 应用流量对电信运营商的影响

内容共享服务是当前最常见的 P2P 应用形式，主要用于音乐与视频文件下载，当前比较流行的下载工具包括 BitTorrent 和 eDonkey（电驴）等。这类应用使得文件下载更加容易，很多用户长时间地在线下载产生了大量的流量。对 Internet 服务提供商（ISP）而言，P2P 应用流量消耗了大部分的网络带宽，经常造成网络拥塞，严重影响其它应用的性能。P2P 技术用于通信服务是一类正在兴起的应用形式，主要指基于 P2P 技术的 VoIP 和流媒体（streaming media）技术等。

应当明确，P2P 内容共享应用产生的大量流量来自于用户的实际使用，P2P 技术本身并没有产生额外的流量负担，基于 P2P 的流媒体技术甚至可以优化网络带宽的使用效率。但 P2P 技术引入改变了用户的使用习惯和网络中流量的分布特征，因而对电信运营商的网络运维有较大的影响。另一方面，P2P 内容共享涉及到内容，这是当前电信运营商难以控制的环节。尽管 P2P 内容共享是目前宽带网络典型的“杀手应用”，但它给 ISP 带来了技术、安全及法律方面的诸多问题[16]。

- **技术问题**

在 P2P 系统中，当用户通过系统底层找到了感兴趣的数据文件或通信对象时，将在对等结点直接建立连接。因而，P2P 文件共享应用可能在一个宽带链路上产生大量的连续 TCP 流。TCP 连接通过窗口调节机制试图最大化其吞吐量，因而 P2P 连接能够占用 ISP 网络的全部可用带宽。更严重的问题在于许多文件共享应用一直连接在 P2P 网络上连续地下载或上传数据文件。一般来说，ISP 的网络容量配置并不能满足其卖给用户的全部容量要求。因此，不断增加的 P2P 应用导致了网络拥塞和 Internet 连接的总体质量下降。

最糟糕的事情是时延和分组丢失的增加可能使那些使用交互式应用的用户无法忍受，P2P 内容共享用户可能破坏了大多数 Internet 用户的使用体验。在有些网络中，P2P 应用产生的大量连接甚至会阻碍其它新的 TCP 连接产生。显然，对网络进行扩容可以适当缓解这个问题。有些网络瓶颈，比如容量不足的转接链路，很容易升级。但接入网络涉及到的点多面广，扩容问题则比较难以解决。

- **安全问题**

大量的音乐和电影文件下载消耗了 ISP 网络的带宽并导致网络服务质量下降。除此之外还有一些不良的东西，比如蠕虫和木马病毒等也通过 P2P 文件共享应用进行传播。病毒程序可能是应用的一部分（特别是 P2P 的客户端）或伪装成无害的可下载内容。一旦被打开，它们就会为发送者提供对被侵害的计算机、网络和信息资源不同程度的访问和控制权。通过 P2P 网络传播的病毒比传统的邮件蠕虫病毒更加难以处理，因为集中式的网络过滤措施难以在分布式的 P2P 网络中实施。

- **法律问题**

数字版权机制还没有被广泛采用。音乐与电影文件的下载主要还是一种“地下”活动，使用者有可能侵犯了内容所有者的版权，因而 P2P 文件共享可能带来一些法律问题。事实上，Napster 和 KaZaa 已经因为在和音乐版权所有者的诉讼中败诉而被迫关闭。ISP 间接地介入侵权内容的传播，即便不承担法律责任，也有可能卷入到相关的法律纠纷中。ISP 应该考虑到这个问题，必要时采取一定的补救措施以避免陷入法律问题。

6.4.2 P2P 应用与电信运营

P2P技术最初来源于Internet研究界，其应用设计思想也遵循Internet的开放自由的基本思想，并没有考虑电信运营的需求。但是由于大多数电信运营商提供Internet接入服务或直接提供某些Internet服务，因此P2P应用与电信运营之间存在着密切的关系。具体地，P2P应用中的两类，即内容和文件共享及协作应用与电信运营关系密切。

当前P2P内容与文件共享服务占用了大量的网络带宽，各结点之间直接进行通信使运营商彻底沦为数字管道。尽管运营商承担了巨大的流量负担，甚至由于网络拥塞影响了其它业务，但由于包月方式的平面计费方法而无法获取相应的收益。另一方面，P2P协作应用中的即时消息正在年轻人当中广为流行，在一定程度上侵蚀了电信市场；更直接地，基于P2P的IP电话技术，如Skype已经开始和电信运营商争夺市场，随着电信服务市场的进一步开放，这种低成本的IP电话技术必将严重影响传统电信运营商的话音通信业务收入。但是面对当前的这种形势，除了简单地“封杀”以外，电信运营商并没有找到有效的应对策略。

我们认为，出现这种情况的原因在于当前的P2P应用对于电信网络而言更象是一个入侵者，它们来自于电信网的外部，具有自己的设计与工作理念，但是通过购买接入服务的方式获得了合法使用电信网资源的权利。P2P应用与电信运营之间的矛盾客观上损害了电信运营商的利益。作为一种新生事物，P2P顽强的生命力已经是不争之实。这种情况下，电信运营商有两种策略可以选择：一、改变自身。改变网络的服务模型，限制P2P应用对带宽的无限占用。比如对P2P流量加以识别并与其它应用的流量进行隔离。这种方法目前在技术还有一定的困难。首先是P2P流量可能采用动态端口而难以识别，另外P2P也可以借用Web端口隐身于其中。

二、改造P2P应用，将P2P技术与应用纳入到电信运营的轨道中。Internet商业化以来，由于其拥有的计算能力带来强大的创新能力产生了很多新型的服务，有些已经对传统的电信业务产生了直接的冲击。在这种形势下，电信运营商需要改变思路，积极推进战略转型，由传统的以语音为主的通信运营商向综合信息服务提供商转变，从而在与新兴网络的竞争当中保持领先优势。对于P2P应用，运营商要重新定位的自己在信息服务中的角色，接受并利用P2P技术，创建新的价值链。比如与内容所有者合作，加强版权管理，使得现阶段几乎失控的内容共享服务可持续发展；加强用户管理，使得即时消息应用成为真正的安全、可信、健康的个人通信平台；适当引入P2P技术，为用户提供更加廉价的视频服务等。由于涉及到内容传送，IPTV就是一个很好的备选对象。

6.4.3 P2P 应用流量的控制策略

当前，以 BitTorrent 为代表的 P2P 内容共享流量占用了宽带接入的大量带宽，如果采用 P2P 技术实现 IPTV 系统，将产生更大的流量，这对于以太网接入等共享带宽的宽带接入方式提出了很大的挑战。大量的流量使接入层交换机的端口长期工作在线速状态，严重影响了用户使用正常的 Web、E-mail 以及视频点播等业务。因此，运营商和企业都有对这类流量进行控制的要求。考虑到 ISP 内部网络容量一般较为充足，流量成本主要在出口链路处产生，允许用户使用 P2P 的同时通过一定的策略将用户流量限制在内部网络不失为较好的方法。

流量本地化的实现可以在 P2P 应用的内部或外部实施。根据具体的 P2P 应用，可以采用内容缓存、连接重定向和内部网络中主动部署超级结点几种方法来实现这个目标。缓存类似于 WWW 应用中的网页缓存技术，在内部网络中部署一些缓存服务器将 P2P 用户感兴趣的内容存入其中，这样用户不必到外部网络中获取这些内容。在一段时间内，用户感兴趣的内容相对比较集中且变化不大（比如近期热播的影视剧），缓存的方法是可行的。但这种方法面临版权问题，除非 ISP 事先购买版权内容的使用权。

P2P 重定向要求网络内部的 P2P 客户端产生的信令流先经过重定向服务器进行处理，重定向服务器检查信令的内容并决定用户所需文件从内部还是外部网络的 P2P 用户处获取。P2P 重定向服务器需要理解所有的 P2P 应用信令且需设在网络的关键位置，其中心结点特性有可能成为性能瓶颈并带来扩展性问题。

对某些 P2P 文件共享系统，如 KaZaa，其中存在超级结点以实现文件的快速搜索，因而超级结点是实施流量引导的有利位置。在超级结点实施本地流量感知的对等结点选择算法可以将实现负载均衡，将流量尽量限制在本地网络中。对于已经存在的 P2P 应用，ISP

可以主动地在网络中部署一些超级结点并广播其存在。这种方法是重定向服务器的退化形式，可以将流量尽可能地限制在内部网络中；而且由于主动部署的超级结点位置已知，可以增加流量分布的可预测性。对于电信运营商主导开发的应用，可以将具有流量控制能力的结点选择算法内嵌到系统中。

6.5 小结

本章介绍了 P2P 技术的基本概念及 P2P 内容传送的关键技术，然后分析了三种 IPTV 业务提供技术。我们认为 P2P 技术尽管还不成熟，技术本身也存在一定的缺陷，但由于具有可扩展性良好，系统构建性价比高，因此是一个值得研究的方向。接下来研究了基于 P2P 的 IPTV 系统的体系架构，结果表明混合式体系架构具有明显的优势。本章还讨论了 P2P 流量对网络和基于 P2P 的业务部署的影响，我们认为只要管理与控制得当，P2P 业务流量不会对网络产生严重影响，P2P 技术可以应用到电信业务开发中。

参考文献

[1] CacheLogic, Understanding peer-to-peer , http://www.cachelogic.com/p2p/p2poverview.php

[2] S. Androutsellis-Theotokis and D. Spinellis, A Survey of Peer-to-Peer Content Distribution Technologies, ACM Computing Surveys, Vol. 36, No. 4, December 2004, pp. 335–371.

[3] Dejan S. Milojicic, Vana Kalogeraki, Rajan Lukose, et al, Peer-to-Peer Computing, HP Laboratories Palo Alto, HPL-2002-57 (R.1), July 3rd , 2003

[4] J. T. von Hoffman, Guide to Distributed PowerPoint, http://www.accessgrid.org/agdp/guide/dppt.html

[5] Y. Chawathe, S. Ratnasamy, L. Breslau, et al. Making Gnutella-like P2P systems scalable.In Proc. of the ACM SIGCOMM 2003.New York:ACM Press,2003.407-418

[6] E. K. Lua, J. Crowcroft, M. Pias, R. Sharma, S. Lim, A survey and comparison of peer-to-peer overlay network schemes Communications Surveys & Tutorials, IEEE (2005), pp. 72-93.

[7] Pablo Rodriguez and Ernst W. Biersack, Dynamic Parallel Access to Replicated Content in the Internet, IEEE/ACM Trans. On Networking, Aug. 2002. VOL. 10 (4): 455-465,

[8] Arnaud Legout, G. UrvoyKeller and P. Michiardi, Rarest First and Choke Algorithms Are Enough, in Proc. of IMC'06, Oct. 06

[9] C. Diot, B. Levine, B. Lyles, H. Kassem, D. Balensiefen, Deployment issues for the IP multicast service and architecture [J], IEEE Network 14 (1) (2000). p78~88

[10] X. Hei, C.Liang, J. Liang, et al., Insights into PPLive: A Measurement Study of a Large-Scale P2P IPTV System, In Proc. of IPTV Workshop, International World Wide Web Conference, 2006

[11] Jian Liang, Rakesh Kumar and Keith W. Ross. The FastTrack overlay: A measurement study. Computer Networks，2006. 50(6): 842 ~ 858

[12] Ian G. Gosling, eDonkey/ed2k: Study of A Young File Sharing Protocol, SANS Institute Technical. Report online: www.giac.org/practical/GCIH/Ian gosling GCIH.pdf. May 2003

[13] Bram cohen, Incentive Build Robustness in BitTorrent, In 1st Workshop on the Economics of Peer-2-Peer Systems, Berkley, CA, June 5-6 2003.

[14] X. Zhang, J. Liu, B. Li, and T.P. Yum. Coolstreaming/donet: A data-driven overlay network for peer-to-peer live media streaming. In Proc. of INFOCOM'05, Miami, March 2005.

[15] http://www.cisco.com/global/CN/about/news_info/press_release/leadship/2005_06_1.shtml

[16] Klaus Nieminen. The P2P Problem and Solutions - An ISP Perspective. In Report of the Licentiate Course on. Networking Technology, HUT Spring 2004. http://www.netlab.hut.fi/opetus/s38030/K04/report_4.04.pdf

第 7 章　一种基于 P2P 技术的 IPTV 系统内容调度算法

根据第 6 章的分析，内容路由与内容交换是 P2P 内容共享系统两个最基本的功能；类似于 BT 的混合式网络架构可以高效地完成内容路由功能，包含跟踪服务器的架构也有助于实现电信业务的管理，但是为了实现 IPTV 系统中流媒体应用中，还需要设计新的内容交换机制。

7.1 问题描述

BitTorrent [1]是一种 P2P 应用，其设计目标是通过端系统合作实现快速高效地共享大型文件。和其它 P2P 应用一样，BT 的某个结点在下载文件的同时也要贡献其上传宽带。BT 的基本思想是将一个文件分成等长的块（通常是 32~256KB 大小），这些文件块可以被多个对等结点同时下载。在 BT 中，为了避免请求-响应的时延，文件块被进一步分成了子块，从而保证请求和响应以一种流水线的方式（pipelining）工作 [2]。

在 BT 中，每个结点都在寻找机会和邻居结点进行数据交换。通常，结点可能有多个数据块可以下载，这时需要从中选择一个。BT 采用了本地最稀缺优先（local rarest first，LRF）的策略选择数据块，即选择下载在邻居结点中重复数量最少的数据块。

P2P 应用的一个重要的问题是存在“搭便车者”：某些用户只愿意下载，而不愿意为其它结点提供上传服务。“搭便车者”只消耗而不贡献资源，不利于系统的扩展。为了防止过多的“搭便车者”存在，需要采用一定的激励机制鼓励结点贡献自己的资源。BT 采用了“投桃报李”（tit-for-tat，TFT）机制防止搭便车用户，即结点优先给下载速率最高的结点上传数据。每个结点限制上载的并发连接数量，一般为 5 个。种子结点不下载数据，但也遵循相似的策略，即只给 5 个下载速率最高的结点上传数据。

IPTV 系统中，流媒体应用是最重要的组成部分。BT 的设计目标是文件共享，并不能直接用于流媒体应用。与文件共享不同，在流媒体应用中，数据需要被即时播放，因而数据到达必须满足媒体播放的时间约束。当某个结点收到了其伙伴结点发来的数据块保有状态信息时，需要决定从哪个结点获取所希望得到的数据块，这个过程即是内容调度。对于同构和静态的网络，由于数据分布比较均匀，简单的轮循调度就可以满足要求；但是在异构网络中，需要设计智能化的调度算法才能更好地满足媒体内容的实时播放要求。特别地，调度算法需要满足两个约束条件：每个数据块的回放时限和伙伴结点的带宽约束。如果第一个条件不能满足，应该让错过回放时限的数据块数量尽可能地少。

在文件共享系统中，为了保证文件的完整性，必须获取所有文件块，但是对获取的时间没有要求，数据可以在任何可能的时候被下载。而且用户习惯于接受很长的下载时间，即便整个下载过程持续数十小时，用户通常也可以忍耐。由于采用类似于 BT 的文件分块和搜索机制，P2P 流媒体应用好象 BT 的文件下载方式非常相似，但关键的区别在媒体文件的实时性特点要求文件块的下载必须保持相同的时间约束。因此，P2P 流媒体应用中，内容调度算法是非常关键的组成部分，它将从不同的伙伴结点中选择必须下载的数据块来满足回放时限要求。当然，流媒体应用对数据的完整性要求有所减弱，在回放时间到达时错失的数据块可以通过差错掩盖机制进行补偿或干脆降低画面的质量。

研究表明，BT 在文件下载方面非常高效。BT 的系统架构，特别是文件的搜索与分块机制可以作为 P2P 流媒体应用的基础，但需要修改内容交换机制以满足流媒体应用的要求。

7.2 内容分块机制

在 P2P 文件共享应用中，为了让文件的各个部分可以并行下载，需要将文件分割成多个小块。文件块通过顺序的序号来表示，编号通常从 0 开始，直至最后一块。类似地，在 P2P 流媒体应用中，对媒体内容进行分块也是实现并行下载的前提。在 BT 中，每个文件块被进一步分成了子块，通常大小为 16K 字节。当结点需要下载某个文件块时，发送端会收到多个子块的请求，由于其中的一些请求处于等待状态，一个子块发送完成时可以立刻发送下一个子块，从而以流水线方式工作，省略了等待下一个请求到达的时间。流水线工作过程中可以控制等待的请求数量以保证充分利用大部分连接的容量。

在CoolStreaming [3] 和PPLive [4]中，媒体流仅被分成长度均匀的段，没有公开的信息表明各个段又被进一步分成了子块。在本文的基于P2P的IPTV系统中，媒体内容只被分成等长的块[1]，BT的流水线机制被视作一种数据传输的优化而被设定为可选项。

- **内容状态表示**

缓存是减少消息或数据报文到达时延抖动的常用方法。在 P2P 流媒体系统中，采用缓存来保存已获取但还没有用于回放的内容分块。根据内容的实时性约束，内容块以回放顺序存放在缓存中。在 P2P 系统中，结点动态加入系统，也可能在没有任何通告的情况下退出系统；结点间连接的质量也是不可靠的。因为内容源和结点间通信连接都包含不确定因素，一般来说 P2P 流媒体系统不能严格按照播放顺序获取内容块。若系统中暂时没有所需的内容块，缓存中会有空缺部分存在。

类似于 CoolStreaming 系统，P2P 流媒体系统中采用缓存图（Buffer Map, BM）来表示当前缓存中内容块的存储状态。在缓存图中，内容的存储状态用二元状态来记录，1 表示存在，0 表示不存在。图 7.1 显示了一个 BM 的例子。缓存图表示了结点当前缓存中内容的可用性情况，它是内容调度中内容块选择的参考依据。

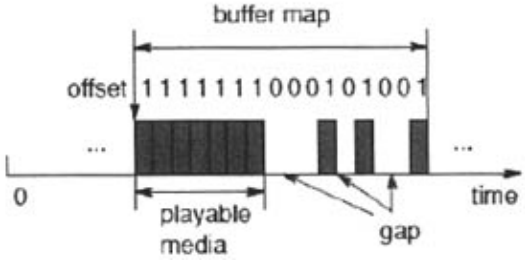

图 7.1 缓存图示例

[1] 这种等长划分方式或许并不是最优的，因为媒体内容编码的单元可能是变长的。

- **内容状态传播**

当结点收到另一个新结点的请求同一媒体内容的数据时，它会把这个结点的信息加入到伙伴结点列表中。每个结点伙伴结点列表长度是有限的，如果列表已满，则新结点将代替伙伴结点列表中“最老的”（保持不变时间最长）的结点。伙伴结点列表会在伙伴结点间周期性的交换以将结点列表信息扩散开来。

每个结点还在其保存的伙伴结点之间周期性地交换当前的缓存图，这样结点可以知道当前系统中内容块的可用状态，从而决定从哪个伙伴结点中获取哪个内容块。状态传播的关键问题是信息交换的周期，如果周期过短，则将会产生大量的控制消息；如果周期过长，则不能及时把有用的信息传播给伙伴结点。

7.3 P2P 流媒体应用的分块选择算法

在文件共享应用中，文件的所有数据块都可以选择下载。在 P2P 流媒体应用中，由于数据块在缓存中保存，只有那些落在缓存范围内的数据块才可以被选择。缓存的大小有限，由于实时播放的要求，随着时间推移，缓存中需要保存的数据块序号也随之推移，数据块的选择范围也随之改变，类似于滑动窗口协议的工作方式。在回放时限到达时还没有到达的数据块以后也不再有用，也不需再去获取它。流媒体内容的数据块的选择有多种算法，包括本文提出的算法。

- **随机选择 (Randomly Selection, RND)**

随机选择缓存范围内不存在的数据块。显然，RND 算法没有直接考虑媒体内容回放的时限要求。

- **本地最稀缺优先（Local rarest first, LRF)**

CoolStreaming系统采用了这种算法。就象BT一样，结点的内容调度器优先选择本地所知的网络中存在但重复数量最少的数据块。该算法首先计算每个所需数据块潜在提供者（保有该数据块的伙伴结点）的数量。因为内容源少的结点在满足播放时限约束方面更加困难，该算法优先调度内容源最少的数据块，然后再调度内容源更多的数据块。在多个内容源中，选择带宽最宽服务时间充足的伙伴结点。

这个算法可以有效地确保结点拥有其它伙伴结点感兴趣的数据块，可以在收到请求时及时上传。它也可以保证更加普遍的数据块留在系统中供以后下载，由于系统中数据块更加分散，也可以减少当前提供上传服务的结点在上传结束以后发现系统中没有感兴趣的数据的可能性。由于缓存较大，LRF 可能会下载稀缺但不紧急的数据块。

- **最早时限优先（Early deadline first, EDF)**

优先选择播放时限最近的数据块，如果系统中没有则选择时限次近的数据块，以此类推。EDF 直接考虑了播放时限要求，但可能造成系统中数据块分布比较集中，不利于负载分担，且对于异步系统会增加除了种子结点外找不到其它内容源的概率。

- **最早时限优先与本地最稀缺优先结合（Mixed)**

通过上面的分析，我们可以看到内容调度算法需要在媒体的实时性约束和系统中可用内容数量之间进行折衷。我们结合 LRF 和 EDF 的优点设计了一种将两者结合的内容块选择算法，即在缓存空间的前半部分采用 EDF 策略，而在缓存空间的后半部分采用 LRF 策略。这种算法既考虑了实时性要求，但连接容量较大的情况下，还可以获取稀缺内容，不仅为本结点的后续内容获取提供保证，也可以更好地为其它结点提供数据。分割两种算法的边界值决定了实时性约束和系统内容可用性的折衷倾向程度。

7.4 结点选择算法

每个结点都在本地维护一个伙伴结点列表。当结点收到来自于其它对等点的内容请求时，表明这些请求结点也在相同的流媒体会话（session）中，它将用请求结点来更新（增加或替代）伙伴结点列表。

每个结点周期性地向伙伴列表中的所有结点请求获得新的结点信息。当这些结点收到请求消息时，它会将自己的伙伴结点列表发送给请求者。系统中的所有结点通过这种方式保持其结点列表信息始终是最新的。伙伴结点列表确定了结点选择的范围。

根据接入链路的带宽，系统中的每个结点都有下载和上传连接数量限制。当结点试图下载或上传一个内容数据块时，它首先要检查这些连接限制。当某个结点收到来自于其它结点的下载请求而它又没有上传容量时，它将给请求结点发送 NACK 消息。当请求结点收到 NACK 消息时，它将在伙伴结点列表中将请求结点中的状态设为 BUSY。

当结点选择伙伴结点时，它会先选择那些非 BUSY 状态的结点，如果找不到这样的结点，它将尝试性地向 BUSY 结点发送数据请求。

最好的结点选择方法应该是根据其当前的负载情况，负载情况可以通过结点的连接（下载或上传）数量来表示。

7.5 仿真结果

7.5.1 仿真模型

OMNeT++是基于 C++的离散事件仿真器，它可以用来仿真通信网络、多处理器系统和其它分布式和并行系统。OMNeT++是开放源代码软件，它可以按照 GNU 的通用公开允可协议或软件本身的允可协议用于非赢利场合。

OMNeT++模型包括了通过消息传递进行通信的模块构成。构成模型的基本模块被称之为简单模块（simple modules）；简单模块可以被组成复合模块，复合模块可以进一步组合；组合的层次没有限制。图 7.2 显示了一个 OMNeT++系统模型，它包括了简单模块和复合模块。

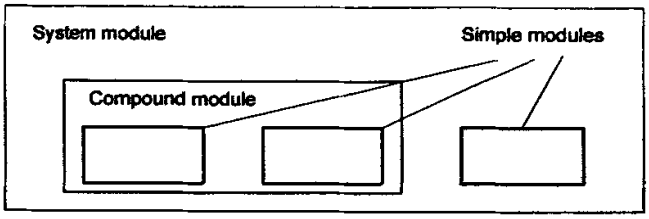

Fig. 7.2 OMNeT++ 系统模型

多路并行下载是 P2P 内容交换应用的核心部分，P2P 流媒体仿真模型基于基本的 OMNeT++仿真平台，我们在 OMNeT++平台上实现了类似于 BT 和 eDonkey/eMule 内容路由的基本功能，以及各种内容调度算法用于仿真 P2P 流媒体应用的多路并行下载机制。

- **结点动态模型**

启动模型：在 P2P 系统中，结点动态地加入和退出系统。我们用随机时延仿真结点的启动时间，系统中的每个结点在经历一段随机时延后再加入到系统中，因此系统中的结点数量不固定。当结点加入系统以后它就根据内容路由得到内容信息和结点的初始配置开始选择选择伙伴结点和所需的媒体数据块进行内容交换。

退出模型：在仿真系统中，当一个结点因为故障或对媒体内容失去兴趣而主动退出时，将会给流媒体系统发送通告消息以终此本次流媒体会话。尽管这与实际情况并不相符但可以简化仿真过程。此外，我们配置仿真系统使得即使结点的流媒体播放结束也不会退出系统，而是继续为其它结点提供数据。

- **连接模型**

表 7.1 所示为 P2P 流媒体系统中的接入类型。

表 7.1: P2P 流媒体系统的终端接入类型表

接入类型	ADSL_256	ADSL_512	ADSL_1500	SDSL_512	DS1	ETH_10	ETH_100
下载容量(KB)	32	64	196	64	256	1250	12500
上传容量(KB)	8	16	32	64	256	1250	12500

不同的接入类型在系统中占不同的比例，表 7.2 给出一个接入类型比例分配的例子。

表 7.2: P2P 流媒体系统中终端接入类型比例配置

接入类型	ADSL_256	ADSL_512	ADSL_1500	SDSL_512	DS1	ETH_10	ETH_100
百分比 (%)	0	50	20	10	10	5	5

种子结点的接入类型一般具有较高的带宽。不同的接入类型对应不同的下载和上传连接限制。

- 其它系统参数

数据块大小（BLOCK_SIZE）： 64KB = 0.5Mb

媒体内容长度：数据块大小确定时，媒体内容长度由数据块的数量（NUM_BLOCKS）决定，160 块大小为 64KB 的数据量为 10MB(160 x 64 = 10240 KB)

缓存大小（BUF_SIZE）是一个非常重要参数，对系统的性能有比较重要的影响，图 7.3 所示为 P2P 流媒体系统的缓存结构示意图。流媒体系统中，缓存是一个由起点数据块序号（B0）和终点数据块序号（B1）决定的滑动窗口，在混合内容调度算法中还包括分一个割 EDF 和 LRF 策略的边界值（BT）。

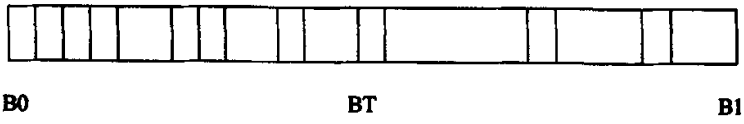

图7.3 P2P流媒体系统缓存结构图

与缓存有关的其它两个参数：

- 缓存移动时延（DELAY_BUF_MOVING）： 缓存窗口滑动时延，用于仿真流媒体的媒体播放过程；
- 缓存移动步长（STEP_BUF_MOVING）：缓存滑动窗口每次滑过的数据块数量。

DELAY_BUF_MOVING 和 STEP_BUF_MOVING 决定了缓存窗口滑动的速度，也即流媒体播放的速率，例如，假定每个数据块为 64KB, 则每 4 秒钟缓存移动 1 块数据块长度，相当于媒体回放速率为 128Kbps，每 4 秒钟滑动 2 块数据块长度则相当于媒体回放速率为 256Kbps；尽管每 2 秒钟滑动 1 块与每 4 秒钟滑动 2 块所仿真媒体回放速率相同，但缓存窗口变化的速率不同，对数据块的选择有一定的影响。

7.5.2 关键性能指标

关键性能指标（Key Performance Indicators，KPIs)是衡量 P2P 流媒体性能的重要参数，我们选择了两个重要指标做为仿真实验的测量对象：

数据块错失率（Chunk Missing Rate，CMR）：一段时间内错过回放时限的数据块数量与这段时间内应到的全部数据块数量的比值。在我们的实验中，数据块错失率在流媒体播放过程结束时计算，它仅仅是一个相对的近似，不能准确反映整个流媒体播放过程的数据块到达情况，但可用于同等条件下不同算法的性能对比。

源结点带宽消耗：仿真实验的第二个目标是检查 P2P 流媒体系统中对源结点上传带宽的依赖程度，在仿真结束时分别记录源结点的数据上传与控制消息的产生负载。这个性能参数与系统的可扩展性紧密相关。在相同结点数量的条件，实现系统功能的同时对源结

点的依赖越小，则意味着系统的可扩展性越高。下面的实验中将主要使用上述参数来描述系统的性能。

7.5.3 实验与结果

- **仿真参数配置**

结点动态模型：结点启动时间分成四类，10, 20, 30 和 40，在此基础上分别加上一定的随机扰动。连接模型如 Table 7.2 所示。

其它系统参数配置如下：
BLOCK_SIZE = 64KB = 512Kb；NUM_BLOCKS = 160；
BUF_SIZE = 20 (blocks)，缓存初始位置：B0 = 0; BT = BM_SIZE/2; B1 = BM_SIZE；
DELAY_BUF_MOVING = 16，STEP_BUF_MOVING = 4；对应的缓存滑动速率为 128K bps 。

- **实验 1: P2P 流媒体系统中结点数量对关键性能指标的影响**

本实验检查系统中结点数量对性能的影响。数据块错失率（CMR）与媒体播放的连续性密切相关，CMR 越高则媒体播放质量的下降越严重。图 7.4 所示为数据块丢失率随着系统中结点数量增加的变化情况。对于四种内容调度算法，随着结点数量的增加，CMR 也随之增加。相对而言，采用 EDF 和 RND 时增加较为平缓，采用 EDF 时 CMR 最小。这是因为 EDF 算法考虑了数据的时间约束。从系统的扩展性角度，曲线上升平缓表示结点数量对 CMR 影响较小，从而可扩展性更好。当结点数量较少时，Mixed 算法表现较好，但随着结点数量增加，其性能优势不再明显。因为数据块选择时考虑的其本地稀缺性而非实时性，采用 LRF 算法时数据错失率较高。

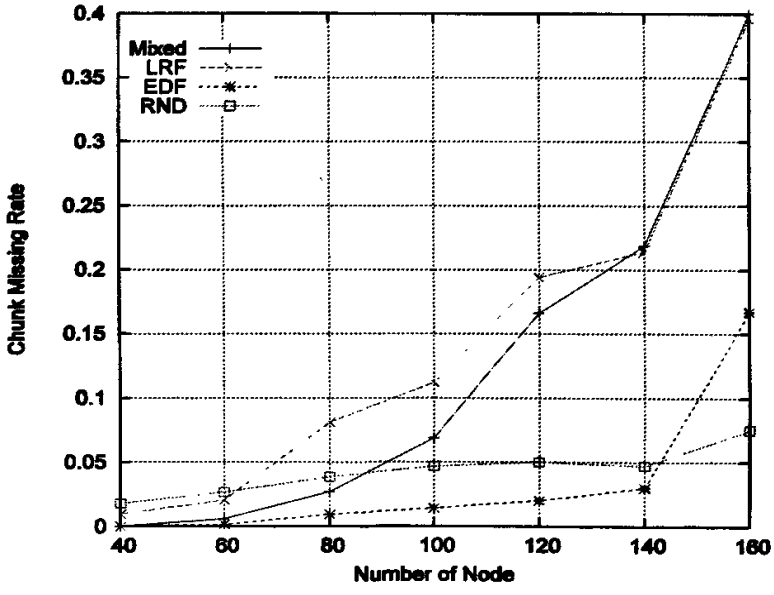

图7.4 数据块错失率随结点数量变化图

本实验中假定：1) 在 P2P 流媒体系统中只有一个种子结点，但非种子结点下载完数据块以后会保存起来为其它结点提供下载服务，哪怕是已经数据块内容已经回放；2) 当结点完成媒体流回放时并不退出，而是象最初的种子结点一样为其它结点提供服务。显然，当总的下载数据量基本固定时，原始种子结点上传的数据越少，则意味着其它结点对系统的贡献越多，系统的可扩展性就越好。图 7.5 显示了种子结点上传数据量与结点数量的变化关系。从图中可见，采用 RND 和 LRF 算法可以使数据块在系统中的存在更加分散，结点更容易从非种子结点得到数据，因而种子结点上传数据量要少于采用 EDF 和 Mixed 算法的情形。

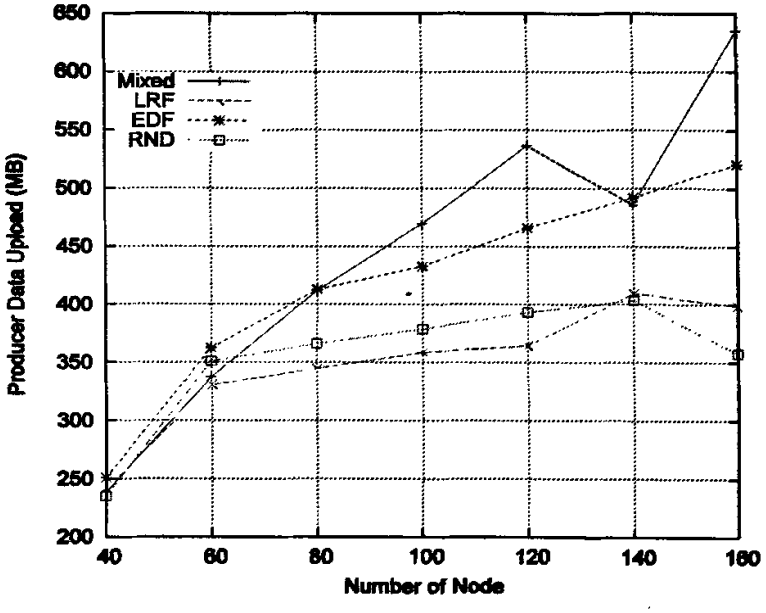

图7.5 种子结点上传数据量与系统中结点数量关系图

结点的控制消息负载包括伙伴结点间交换的伙伴结点列表及结点缓存图消息、数据块的请求与应答消息以及其它控制消息。因为种子结点是系统中每个结点的缺省源结点，其贡献的数据量较大，其控制消息负载一般比其它普通结点要高一些。图 7.5 显示了种子结点的控制消息负载随结点数量增长时的变化情况。从图中可见，相对于上传的数据量，控制消息的负载非常有限，相差两个数量级。尽管 EDF 算法可以带来最小的数据块错失率，但也产生了最大的控制消息负载。我们注意到采用 Mixed 算法时控制消息负载相对较低，但没有找到原因所在。

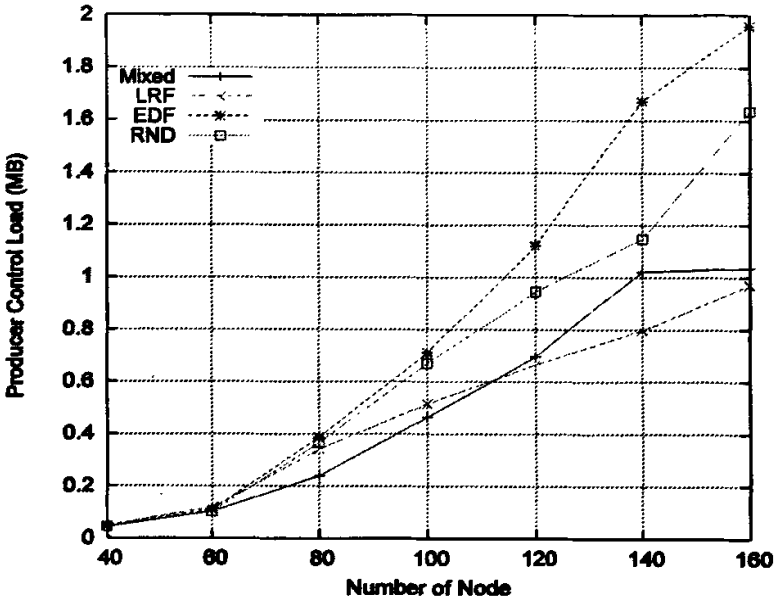

图7.6 种子结点控制消息负载与系统中结点数量关系图

- **实验 2: P2P 流媒体系统中伙伴结点数量对关键性能指标的影响**

本实验检查 P2P 流媒体系统中伙伴结点的数量对关键性能指标的影响。一个结点可能维护数十个伙伴结点以交换数据与控制消息。缓存图和数据交换只在伙伴结点之间进行。一般地，更多的伙伴结点意味着更多的内容源，将会原始种子结点的数据上传量，但系统的控制消息负载将会增加。

图 7.7 所示为数据块错失率随伙伴结点数量增加时的变化情况。对于所有的四种算法，表现出与图 7.4 类似的相对性能关系。对于具体算法，我们注意到，CMR 并不是一直随着伙伴结点数量增加而增加。当伙伴结点数量为 16 或 20 时，对于所有算法，系统表现出最小的数据块错失率；而当伙伴结点数量增加至 24 以上，采用 LRF 和 Mixed 算法时，CMR 增加幅度较大，而采用 EDF 和 RND 算法时，CMR 增加相对平缓，这种变化趋势也和图 7.4 所示相同，应该由相同的原因引起。

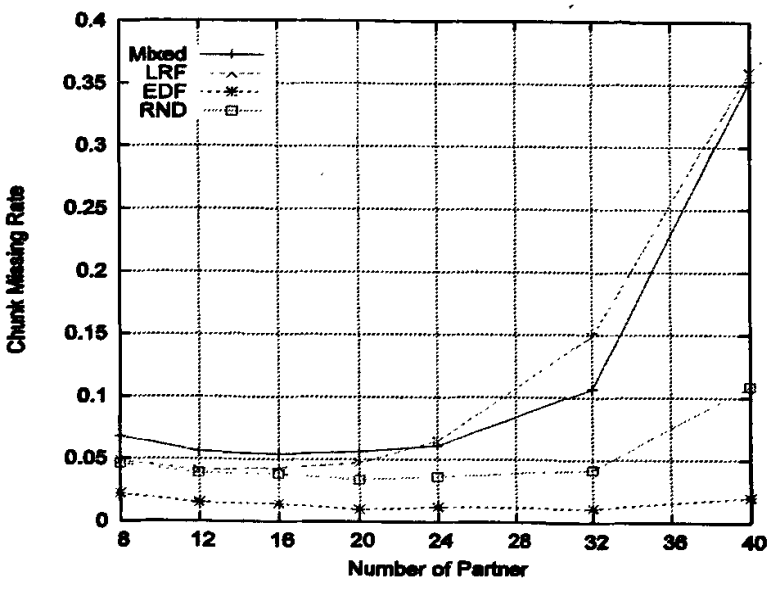

图7.7 数据块错失率随伙伴结点数量变化图

图 7.8 所示为随着伙伴结点数量增加，原始种子结点上传数据量的变化情况。正如期望的结果，伙伴结点越多，内容贡献者越多，对原始种子结点的依赖越小。

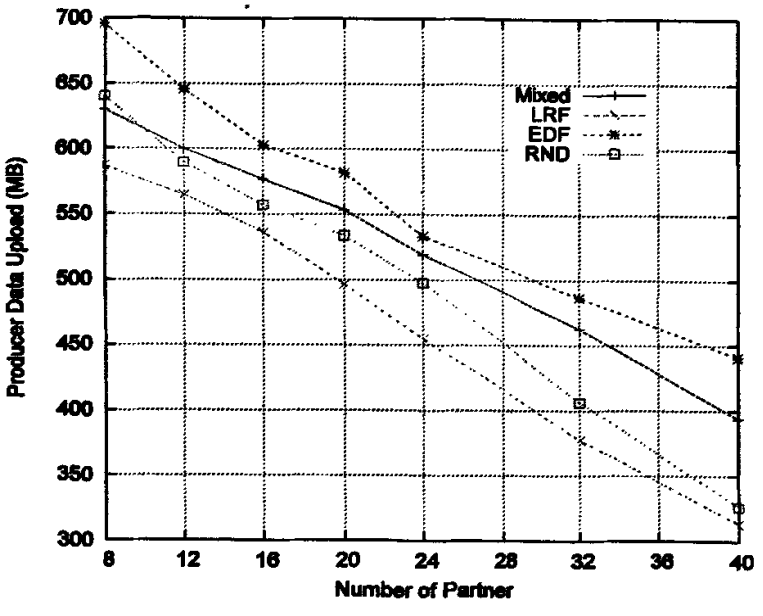

图7.8 种子结点上传数据量与伙伴结点数量关系图

种子结点的控制消息负载随伙伴结点数量增加而变化的情形与数据块错失率类似。除了采用 LRF 算法，采用其它三种算法都在伙伴结点数量为 20 时控制消息负载最轻。我们猜测这是由于当伙伴结点数量较少时，控制消息主要来源于数据块的请求与应答消息，而当伙伙伴结点数量较大时，伙伴结点列表信息与缓存图交换占了控制消息负载的主要部分。伙伴结点列表信息与缓存图交换的消息负载将随着伙伴结点数量增加而增加。

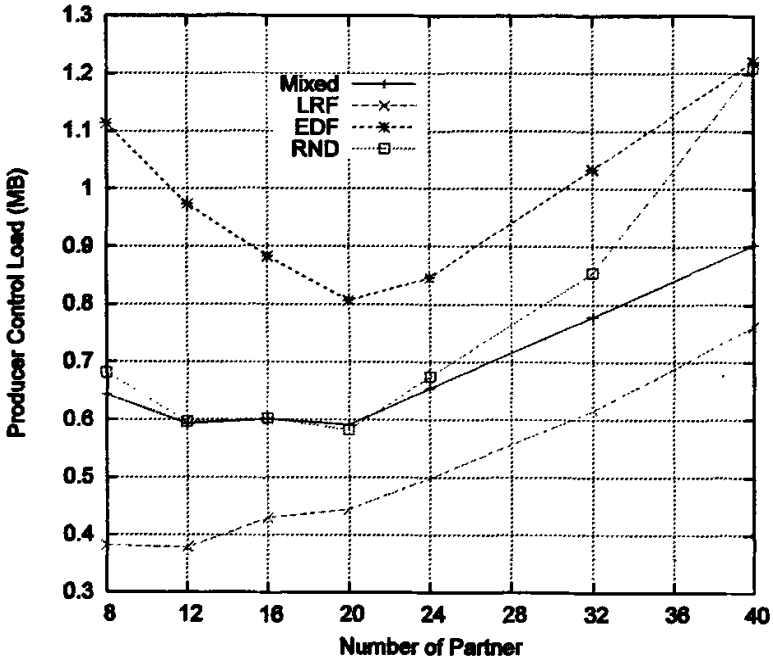

图7.9 种子结点控制消息负载与伙伴结点数量关系图

7.6 小结

P2P 文件共享应用在 Internet 中应用非常广泛，BT、eDonkey 等都具有高效快捷的文件下载能力。由于流媒体系统中内容获取具有严格的时间约束，P2P 文件共享应用并不能直接用于流媒体系统。为了满足流媒体的实时性要求，必须设计新的内容交换算法来弥补 P2P 文件共享应用的不足。本章重点研究了 P2P 流媒体系统的数据块调度算法，对四种内容调度算法的性能进行了比较。仿真实验结果表明，用于 BT 的最稀缺优先调度（LRF）策略并不适合流媒体系统。最早时限优先（EDF）的调度算法可以使媒体回放的连续性更好，但可扩展性相对较差。我们设计了一种将 LRF 和 EDF 相结合的内容调度算法（Mixed）在结点数量较少时表现良好，但在系统中结点数量较多时没有表现出期望的性能优势。

在后续的工作中，我们将进一步研究实验结果的产生原因，提高 Mixed 算法的性能。此外，如何确定 P2P 流媒体系统的缓存大小也需要更深入的探讨。

参考文献

[1] BitTorrent, www.bittorrent.com
[2] Bram Cohen, Incentives Build Robustness in BitTorrent, in Proc. of P2PEcon2003, http://bittorrent.com/bittorrentecon.pdf.
[3] X. Zhang, J. Liu, B. Li and T.-S. P. Yum, DONet/CoolStreaming: A data-driven overlay network for live media streaming, in Proc. INFOCOM'05, Miami, FL, USA, March 2005.
[4] X. Hei, C. Liang, J. Liang, Y. Liu, and K.W. Ross, Insights into PPLive: A measurement study of a large-scale P2P IPTV system, in Proc. Workshop on Internet Protocol TV (IPTV) services over World Wide Web in conjunction with WWW2006, Edinburgh, Scotland, May 2006.

第 8 章　结束语

世界范围内，电信行业的发展正处在转型阶段。Internet 的出现改变了通信服务的垄断局面，VoIP、即时通信等作为替代技术和应用成为用户的新宠，大大压缩通信服务的利润空间。除了 Interent 带来的影响，话音业务向移动转移的趋势使固网运营商的业务和收入都呈现下降趋势。ARPU 值不断下降、用户离网率高是困扰全世界的电信运营商的问题。整个电信界都在思考如何应对当前的困难，当前的结论是一方面需要降低运营成本，另一方面需要开发新的业务，创造新的利润增长点。建设下一代网络、提供业务绑定是目前公认的有效办法之一。

下一代网络基于分组技术，能够提供包括传统电信业务在内的多种业务。多业务网络不需要为每种业务建立一个专用网络，从而大大降低了网络建设和运维成本。促进业务融合，在各种终端上提供三重播放业务，代表了电信运营商的未来发展方向。IPTV 业务正是具有三重播放能力的新业务，提供 IPTV 业务是电信运营商，特别是固网运营商的转型成功的关键。

然而，对电信运营商而言，IPTV 都是个全新的业务， IPTV 业务提供面临着政策、商业模型以及技术等方面的风险。特别是由于 IPTV 中的视频业务具有高带宽、实时播放和交互性支持要求，为 IPTV 业务提供带来了很大的技术挑战。本报告针对在下一代网络中提供 IPTV 业务存在的问题，研究了 IPTV 业务提供的关键技术，特别是基于 P2P 技术的 IPTV 业务提供关键技术。

本报告描述了下一代网络的体系架构，阐明了下一代网络承载与业务分离的设计思想；系统地介绍了 IPTV 系统和业务的基本概念，研究了 IPTV 业务系统与网络体系架构，提出了 IPTV 业务提供的技术需求；研究了 IPTV 业务保障问题，提出了从承载网络、内容传送网络和业务体验三个方面保障用户业务体验的架构；研究了 IPTV 业务提供的接纳控制方案，结合公司的产品现状提出了一个接纳控制机制的实施方案；重点研究了基于 P2P 技术的 IPTV 业务提供关键技术：分析了 P2P 技术在 IPTV 业务提供中的作用，研究了基于 P2P 技术的 IPTV 系统的基本架构，提出了一种新的用于 P2P 流媒体系统的内容调度算法。我们认为，用户业务体验保障对 IPTV 业务提供成功的关键，接纳控制机制是保障用户业务体验质量的基本手段；由于 P2P 技术具有良好可扩展性和经济性，P2P 可以直接应用到 IPTV 业务提供系统设计或作为其它技术的补充。

本报告主要研究了 IPTV 业务提供的承载网络建设与内容传送技术方面的问题，其中基于 P2P 技术的流媒体应用主要以直播方式为目标，视频点播是 IPTV 系统的另一个基本业务，支持点播比直播更加困难，需要更加深入的研究。此外，IPTV 业务提供还需要考虑内容编解码以及内容安全等方面的问题，如何将这些问题与 P2P 技术结合起来也是下一步研究的方向。

发表文章

- 一种链路负载自适应的主动队列管理算法，软件学报，2006 年 5 月
- IPTV 系统的视频接纳控制机制，邮电设计技术，2006 年第 12 期
- VoIP 网络接纳控制机制必要性探讨，电信网技术，2007 年第 2 期
- 移动 WiMAX 与 3G 技术关系统的探讨，移动通信，2007 年 第 7 期
- IPTV 业务保障及关键技术研究，现代电信科技，2007 年第 8 期

致谢

人生如旅游，旅行者不辞辛劳，长途跋涉至人迹罕至处，或许景色迷人，或许不过尔尔，但他们乐此不疲，因为他们看到别人不曾看到的风景；不管结果如何，都是美好的体验。

无论如何，博士算是社会上稀有动物。有人戏称女博士为金庸小说中武功绝顶的老尼，心中已将其定位于最高境界。成为博士又敢更进一步要成为“后”的，无论男女，恐怕不是无奈便是无知了。如是而言，求学不再是旅游，而是颇有冒天下之大不韪的历险行为了。

一晃又是两年，零零总总，山高水长，都已成故事。聊以自慰者，光阴不曾虚度，风景这边尚好。旅行是自己的事，但总离不开结伴者和路人的帮助，没有他们也许旅途难以继续，至少要少了很多的快乐。感谢是一种心意，但是敞开的心灵一定能够接收到这样的讯息。感谢万永根先生的领导，这里有非常好的工作氛围；感谢陈端先生的合作，大部分的工作都是我们共同完成的；感谢孙军教授，尽管并不常常见面，但在我需要帮助的时候总能及时出现；感谢公司博士后工作站的同事们和上海交通大学博士后流动站的老师们，谢谢你们的支持！

纪其进

2007 年 8 月