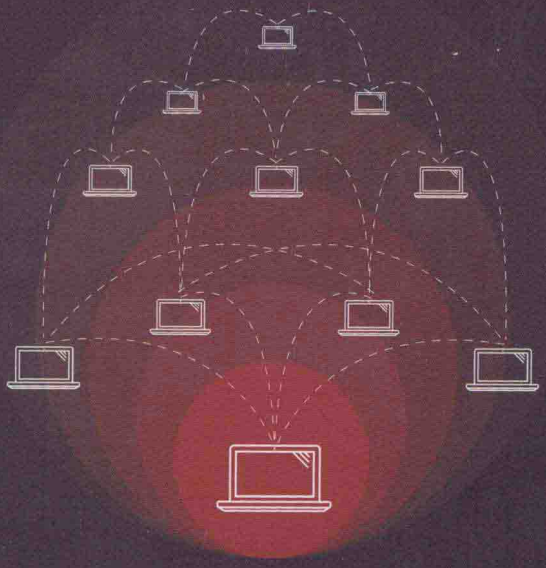
P2P 流媒体

系统关键技术

廖丹 孙罡 曾帅 著



国防工业出版社
National Defense Industry Press

P2P 流媒体系统关键技术

廖丹 孙罡 曾帅 著

国防工业出版社

·北京·

内 容 简 介

本书基于作者多年的研究成果，详细介绍了 P2P 流媒体系统的关键技术。本书内容围绕数据源服务器负载、用户上传带宽利用率、多媒体数据类型支持特性、用户节点接收流率以及转码策略下的覆盖网结构等核心问题展开。本书从理论分析到算法设计、从数学建模到仿真试验对 P2P 流媒体系统做了多方面论述。

本书可作为 P2P 流媒体相关技术的参考指南，可供 P2P 流媒体领域的科研人员、IT 企业系统和网络管理人员以及想要了解、使用 P2P 流媒体技术的读者使用。

图书在版编目（CIP）数据

P2P 流媒体系统关键技术/廖丹，孙罡，曾帅著.
—北京：国防工业出版社，2014.8
ISBN 978-7-118-09619-4

Ⅰ. ①P… Ⅱ. ①廖… ②孙… ③曾… Ⅲ. ①计算机网络－多媒体技术－研究 Ⅳ. ①TP37

中国版本图书馆 CIP 数据核字（2014）第 196968 号

※

国防工业出版社出版发行
（北京市海淀区紫竹院南路 23 号 邮政编码 100048）
北京京华虎彩印刷有限公司印刷
新华书店经售

*

开本 880×1230 1/32 印张 6⅛ 字数 176 千字
2014 年 8 月第 1 版第 1 次印刷 印数 1—2000 册 定价 68.00 元

（本书如有印装错误，我社负责调换）

国防书店：（010）88540777 发行邮购：（010）88540776
发行传真：（010）88540755 发行业务：（010）88540717

前　言

随着近年互联网规模的飞速发展，传统数据下载系统经受重大考验。这主要是由于孤立的服务器的上传带宽、处理能力等性能相对有限，而用户规模却在成几何级数的增长。在这样的背景下，点对点系统（Peer-to-Peer System，P2P）得到了广泛的重视和研究。在P2P系统中，传统设计下的信源和信宿角色不再固定不变。P2P系统通过对网络中各个子系统乃至用户终端本身的分布式利用和管理，达到子系统或者用户终端之间的协作与互助。P2P系统减少了对数据源服务器（或称为源节点）的依赖，数据的存储、传输、运算等任务被相对平均地分摊到系统的各个部分。因此，P2P系统在提高下载速度和减轻数据源服务器负载方面表现出了巨大优势。

作者从2009年开始从事P2P流媒体系统的相关研究工作，先后承担了国家自然基金项目“基于部分合作的分布式重叠网络资源管理机制研究”和“可控可管的P2P业务平台模型及关键技术研究”以及中央高校科研业务费项目“多域网络虚拟化环境的资源管理机制研究”等与P2P流媒体系统相关的项目。目前国内系统介绍P2P流媒体系统的书籍较为匮乏，为了给广大读者一个关于P2P流媒体系统的较为完整和系统的介绍，本书汇集了权威期刊和会议关于P2P流媒体系统的介绍以及作者近年来在P2P流媒体系统方面的一些重要研究成果，系统地介绍了P2P流媒体系统的主要概念和各类P2P流媒体系统的关键技术。希望本书能成为P2P研究人员的有价值的读物，起到“抛砖引玉”的作用。

研究生黄筱聪、杨晓玲、闫书刚、唐汨、卜思桐、袁亚欣、赵东成等参与了全书的资料整理、图表绘制和文字校对工作。李乐民

院士对本课题组多年的研究工作给予了很多指导和帮助，在本书的编写过程中也给予了很多宝贵的意见和建议，在此一并表示衷心的感谢。

由于作者知识和水平有限，书中疏漏和不当之处在所难免，恳请广大读者不吝批评指正。

作　者

2014.5

目　录

第1章 绪 论

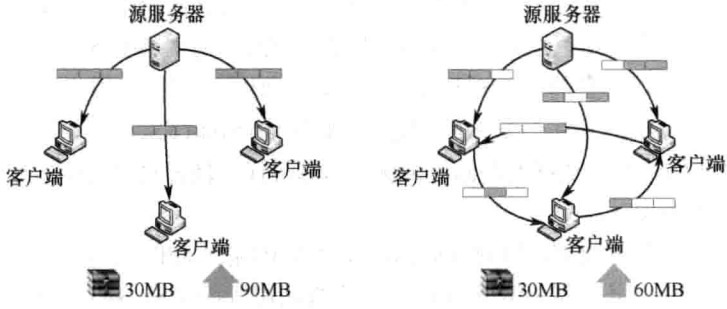
1.1 P2P 文件分享系统

随着近年国际互联网（Internet）规模的飞速发展，传统数据下载系统将经受更多考验。这主要是由于孤立的服务器上传带宽、处理能力等性能相对有限，而用户规模却在呈几何级数的增长。在这样的背景下，P2P 系统得到了广泛的重视和研究[1]。在 P2P 系统中，传统设计下的信源和信宿角色不再固定不变。P2P 系统通过对网络中各个子系统乃至用户终端本身的分布式利用和管理，达到子系统或者用户终端之间的协作与互助。由于减少了对数据源服务器（Source Server）（或称为源节点）的依赖，数据的存储、传输、运算等任务被相对平均地分摊到系统的各个部分。

基于这种原理，针对互联网以及个人电脑的 P2P 应用程序被陆续开发出来，并大量投入实际运行。这些应用程序最初主要是用于文件下载和分享，即 P2P 文件分享系统。例如，著名的 Napster 免费音乐获取软件，以及随后出现的 BitTorrent、eMule、eDonkey 等一大批应用程序[2-4]，都是 P2P 文件分享系统的典型实例。虽然 P2P 文件分享系统的相关应用和研究备受数字版权争议，但是其与传统客户机/服务器结构（Client/Server，C/S）模式相比，在提高下载速度和减轻数据源服务器负载方面还是表现出了巨大优势。例如，如图 1-1 所示，一个 30MB 的文件从数据源服务器发送到 3 个客户端：在传统 C/S 模式下（图 1-1（a）），每个客户端都要独立从数据源服务器下载 30MB 数据，所以数据源服务器总共需要上传 90MB 数据；而在 P2P 模式下（图 1-1（b）），30MB 的文件被分割为 3×10MB 的区块，数据源服务器按图示的方式向每个客户端各发送 2×10MB 的区块，每个客户

端得到的区块不完全相同，3 个客户端可以相互索取自己缺少的部分，最终获得 30MB 的完整文件。按照这样的分享方式，数据源服务器只需承担60MB的上传任务，额外的30MB上传任务由客户端完成。

图 1-1（b）的 P2P 文件分享算法并不一定是最优化的。实际上，在对这类应用程序的性能分析中[5-10]可以看到，一个典型 P2P 文件分享系统分享文件时，客户端（在 P2P 技术相关研究中也称为用户节点）通常可以提供 60%～80%的总上传带宽需求；同时，P2P 系统在鲁棒性以及网络适应性方面也较传统 C/S 模式优秀。但是 P2P 文件分享系统一个比较主要的缺点在于用户节点上线和离线时间不稳定，这使得各个终端设备很难在相对集中的时间内达到协同工作，上传能力不能得到充分的发挥，导致数据源服务器服务时间延长和资源可用性下降等问题。这个弱点在 P2P 流媒体系统的用户行为下显得不那么突出。



（a）C/S 模式　　（b）P2P 模式

图 1-1　文件分享系统：C/S 模式与 P2P 模式对比

1.2　流媒体系统及其发展演变

多媒体技术的发展和互联网规模的急速扩张，为多媒体数据传递找到了新的载体。传统上，像电视、广播这样的传播手段需要建立专用数据传送网络，而新的解决方案依靠互联网实现多媒体数据的传输和分享。

具体来说，这类新的解决方案首先将视频、音频等节目内容通过一定的编码算法编码为易于互联网传输的多媒体数据，然后在互联网

上建立连接数据源服务器和用户节点的覆盖网（Overlay Network），最后将包含视频、音频等节目的多媒体数据封装为网络数据包发送给客户端。

值得注意的是，在这类系统下，用户通常不必等到所有多媒体数据全部下载完毕才能播放视频、音频等内容，而是接收多少数据就可以播放多少。所以，从编码到播放，从数据源服务器到各个用户节点，覆盖网上传输的数据包非常类似于水流的运动过程。于是，这类新兴媒体传输系统被形象地称为流媒体系统。

由于省去了独立建网费用，基于互联网技术的流媒体系统迅速发展壮大[11]。流媒体系统从形式上看，有直播和点播两个大类，即流媒体直播系统（Live Streaming System）和流媒体点播系统（Vod Streaming System）。本书以流媒体直播系统作为研究对象。在不特别说明的情况下，本书所称的流媒体系统就是指流媒体直播系统。

目前，包括互联网电视、远程会议、实时监控等在内的一大批流媒体系统已经活跃在互联网上，成为数据业务中非常重要的组成部分[11]；同时，又有相当数量的流媒体系统，特别是各种新兴P2P流媒体系统，正处于研发阶段，蓄势待发。纵观整个流媒体系统的发展历程，大体上是以覆盖网的结构演变为线索的。这个推进过程主要经历了以下几个阶段：

1. 单播（Unicast）

单播作为最基本最简单的通信方式，在早期流媒体系统中得到了广泛的应用。例如，微软公司开发的Windows Media应用程序[12, 13]就可以非常方便地建立一个简单的单播流媒体系统。如图1-2所示，Windows Media服务端负责采集编码视频、音频数据，等待客户端连接，而Windows Media客户端通过TCP/IP网络直接连接服务端获取多媒体数据，并最终播放。

这类单播覆盖网组织起来的流媒体系统优点在于部署的简易性，对于用户规模不大的情况非常实用；缺点在于整个系统只依靠服务端来承担数据上传任务，一旦用户规模扩大，服务端有限的负载能力很难独立满足需求。

2. 组播和广播（Multicast and Broadcast）

通过组播方式建立的流媒体系统以网络层组播（Network-layer Multicast）[14-16]或者应用层组播（Application-level Multicast）[17-21]作为覆盖网的系统最为普遍。其中，网络层组播以 IP 组播（IP Multicast）技术为代表。而应用层组播相关设计相当丰富，其基本思想是：以数据源服务器为根节点，将用户节点连接到一个树形结构的覆盖网上，实现数据包在应用层的树形转发。如图 1-3 所示，系统首先在数据源服务器和各个客户端之间建立应用层组播网，然后将包含视频、音频等多媒体数据的数据包通过这个网络从数据源服务器层层推送至各个客户端。

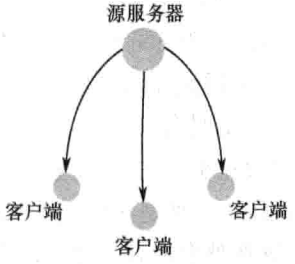


图 1-2　单播流媒体覆盖网示例

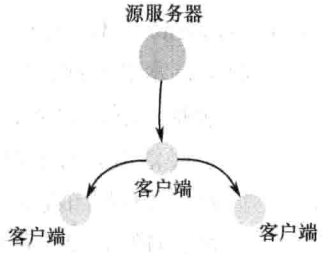


图 1-3　组播流媒体覆盖网示例

和组播方式类似，基于广播网络建立的流媒体系统是在数据源服务器和客户端之间建立广播网络，然后通过该网络分享多媒体数据。以广播模式建立的流媒体系统覆盖网并不常见，只在早期个别基于 CSMA/CD 以太网的流媒体系统中有所应用。

以组播网络或者广播网络组建的流媒体系统优点在于数据源服务器只需要付出较小的上传带宽即可满足多个用户需求。但缺点在于：①网络层组播和广播网络的建立通常需要专门设备，对硬件要求较高。在设备越来越多元化的互联网应用中，缺乏普遍适用性；②虽然应用层组播相对来说对设备依赖性较低，但是居于组播末端的客户端只是被动接收来自组播网络的数据，自身的上传能力没有得到有效利用。

3. 多树（Multiple-Tree）

为了克服单一组播网络不能利用末端客户端上传能力的问题，多树网络的概念被提出并应用于流媒体系统的覆盖网建立[22-26]。如图 1-4 所示，在基于多树网络的流媒体系统中存在多个组播网络（图示的实线和虚线），这些组播网络的起始端（根节点）不但可以是数据源服务器，同时也可以是客户端。作为根节点的客户端，一方面接收来自数据源服务器或者其他客户端的数据包，另一方面又将这些数据包进一步转发给更多客户端。与单独的应用层组播相比，多树网络提高了客户端的上传带宽利用率，降低了源节点负载。

不过，基于多树网络的流媒体系统也有不足之处：①多树网络的组织算法比较复杂，而且处于树形结构中部的节点动态性对于网络整体性能影响较大；②虽然多树和单一组播相比，客户端上传能力得到更充分利用，但是客户端上传效率仍然有可提升和改进的空间[27]。

4. 网状网（Mesh）

以网状网作为覆盖网的流媒体系统把每个客户端都视作潜在的数据转发设备。数据源服务器与客户端、客户端与客户端都可以进行一对一或者一对多的直接通信。如图 1-5 所示，在多媒体直播的初段，数据源服务器首先将数据包发送给部分客户端，然后这些获得数据的客户端进一步将数据包转发到其他客户端，以此类推，最终让所有客户端都满足需求。其中，数据包的具体大小、发送路径、调度组装等问题，则根据不同的算法设计而不同。一般来说，

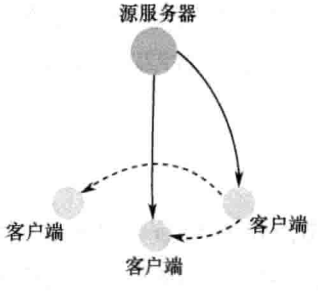


图 1-4　多树流媒体覆盖网示例

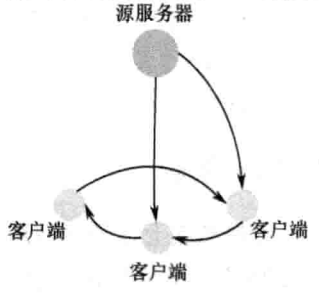


图 1-5　网状网流媒体覆盖网示例

预先设定好参数的网状网通常称为结构化的覆盖网；加入更多随机成分和分布式特性的网状网通常称为非结构化的覆盖网。基于非结构化的网状网流媒体系统通常较易于部署，在网络动态条件下稳定性也较高；而基于结构化的网状网流媒体系统通常算法相对复杂，分布式特性较弱，但传输效率较高。总的来说，和多树相比，网状网覆盖网对客户端上传带宽有更高效的利用，且具有更强的鲁棒性和灵活性[27, 28]。正因为这些优势，目前大量理论研究和实际应用都是以网状网作为系统的覆盖网。

事实上，在应用这些各式覆盖网结构的流媒体系统中，应用层组播、多树和网状网都符合P2P系统的定义，即利用了除数据源服务器以外的系统资源来传递多媒体数据。对这类系统的讨论通常被归入P2P流媒体系统研究的范畴。

1.3 P2P流媒体系统

P2P流媒体系统是一套通过P2P技术来实现流媒体数据传输的系统。它是P2P文件分享系统与流媒体系统结合的产物。与P2P文件分享系统一样，P2P流媒体系统利用了用户终端软硬件资源，特别是存储和上传带宽，来帮助整个系统分享多媒体数据。不过，和P2P文件分享系统对比，P2P流媒体系统也有一些自身特点。

本节首先介绍P2P流媒体系统研究涉及的主要问题，讨论P2P流媒体系统与P2P文件分享系统的异同，给出一些P2P流媒体系统设计实例。然后在此基础上，对比和归纳了以覆盖网结构为索引的各式P2P流媒体系统的特点及其相关研究文献。最后介绍P2P流媒体系统研究的前沿和热点。

1.3.1 核心问题

1. 用户节点上传带宽利用和源节点负载（Server Load）优化

与P2P文件分享系统提出时主要考虑的问题类似，对用户节点上传带宽的利用和对源节点负载（通常以源节点上传带宽消耗为表征）

的优化是P2P流媒体系统研究最根本和最核心的问题[29-33]。这也是以挖掘用户节点资源为目标的P2P技术的出发点。研究人员提出的各种P2P流媒体系统，无论从商业还是技术角度，总是希望能够尽可能利用用户节点包括传输、存储、运算在内的各种潜在能力，提高系统性能，减轻源节点负载，降低系统运营成本。

在本书研究中，提高用户节点的上传带宽利用率以及降低源节点负载是相关设计和算法的主要优化目标。

2. 覆盖网设计

P2P系统中的覆盖网是建立在物理网络上的一种虚拟网络[34]。覆盖网设计存在两个层面的问题：①如何在实际物理网络上构建覆盖网，即提供覆盖网与物理网络的接口问题[35-41]；②在已经建立网络接口的情况下，每个具体系统如何安排各节点间的数据分享规则。前面提到的各种覆盖网结构，只是从节点间数据流的传递形态来考察的，并不涉及实际物理网络如何实现数据传输。后面提到的覆盖网设计或者覆盖网结构也都是研究第二个层面的问题，并不涉及具体的物理网络架构。

一个P2P流媒体系统必然包含一套关于上下游关系、数据流大小、资源获取方法等问题的规则或者算法，来实现多媒体数据在各节点间的分享。这套规则或者算法也就决定了具体系统的覆盖网结构。覆盖网结构对本小节提到的几乎所有系统性能都有非常关键的影响。例如，对于上传带宽利用率和源节点负载性能，不同规则或者算法下形成的数据流动会造成不同的上传带宽利用率以及不同的源节点负载；再如，对于数据调度或者系统鲁棒性也是如此，相关性能也和覆盖网结构密切相关。事实上，从P2P流媒体系统的整个发展历程上看，在名目繁多的P2P流媒体系统设计中，通常都是以覆盖网结构作为优先考虑的问题，从这个问题的解决出发去推进其他问题的解决。

在本书研究中，会结合具体场景，通过系统建模，提出一系列转码策略下的路径选择与带宽分配算法实现各系统的覆盖网设计。

3. 多媒体编码算法（Multimedia Coding Algorithms）

多媒体编码算法主要是指针对视频、音频、图像等多媒体数据的压缩算法。这类算法研究作为一个专门的学科，家族庞大[42, 43]。

其中，编码格式和编码率（Encoding Rate）是非常重要的两个参数。一方面，不同的压缩算法产生不同的多媒体编码格式；另一方面，编码率表征多媒体播放时单位时间读取的数据量大小。一般来说，在编码格式相同的情况下，选用的编码率越高，效果越好，但是生成的数据文件越大；反之亦然。在P2P流媒体系统中，媒体编码的目的在于根据互联网网络环境变化，特别是根据用户节点软硬件解码算法和带宽条件，将多媒体节目压缩为不同格式以及不同编码率的数据方便传输。本书章节标题中的“转码”，就是指多媒体数据在不同编码格式、不同编码率间的转换。目前常用的多媒体编码算法包括MPEG1、MPEG2、MPEG4、H.263、H.264等[44-49]。

本书主要讨论P2P流媒体系统的数据传输部分相关问题。系统只是应用转码技术，不限定具体的多媒体编码算法，不针对编码算法本身进行设计。

4. 数据调度（Schedule）

P2P流媒体系统的媒体推送过程通常是一个将多媒体数据分割、传输，并最终组合的过程。数据调度策略规定了源节点和各个用户节点具体如何存储、分割、获取、组合数据。

在较早提出的P2P流媒体系统中，拥有多媒体数据的源节点或者用户节点首先将已有数据分割为多个区块，然后需要这些数据的目的节点通过一定的算法查找并索取这些区块，接着各个区块通过通常被称为子流（Substream）的各条路径汇集到目的节点，最后目的节点将这些数据组合、存储、播放。同理，目的节点还可以重复以上过程，使多媒体数据进一步扩散，达到分享给更多用户节点的目的。这类算法的主要问题在于区块的多路传输过程中，冗余数据较多。各个节点、各条子流的数据差异很难控制，分布式特性较弱。于是，随后提出了以网络编码（Network Coding）[50, 51]、多描述编码（Multiple Description Coding，MDC）[23, 26, 54, 55]等算法为基础的数据调度方法，这些算法都是以提高数据差异性、减少冗余传输和适应不同上传带宽为目的。需要注意的是，除了多描述编码与多媒体编码可以有相互渗透的地方外，通常数据调度中使用的编码算法和多媒体编码算法并不相互排斥，属于信道编码与信源编码的关系。