

P2P视频点播内容分发策略*

郑常熠, 王新⁺, 赵进, 薛向阳

(复旦大学 计算机科学与工程系, 上海 200433)

P2P Video-on-Demand Content Distribution Schemes

ZHENG Chang-Yi, WANG Xin⁺, ZHAO Jin, XUE Xiang-Yang

(Department of Computer Science and Engineering, Fudan University, Shanghai 200433, China)

+ Corresponding author: Phn: +86-21-55664617, E-mail: xinw@fudan.edu.cn, http://it.fudan.edu.cn/sonic

Zheng CY, Wang X, Zhao J, Xue XY. P2P video-on-demand content distribution schemes. *Journal of Software*, 2007,18(11):2942-2954. <http://www.jos.org.cn/1000-9825/18/2942.htm>

Abstract: Video-on-Demand, which is an important application in peer-to-peer (P2P) networks, has attracted a lot of research interests recently. As P2P network is easy to deploy and can provide a large-scale substrate networks, many P2P VoD content distribution schemes have emerged to offer the basic function of data transportation. This paper presents a survey on the state-of-the-art P2P VoD content distribution schemes. It first summarizes the key issues in designing P2P VoD schemes and then categorizes these schemes into four groups according to data transportation method. Finally, it discusses the application-level performance and outlines possible future work.

Key words: peer-to-peer (P2P); video-on-demand (VoD); content distribution; transport path

摘要: 视频点播目前已成为对等(peer-to-peer,简称 P2P)网络中一项重要的应用,引起了人们的不少研究兴趣.由于P2P网络能够为VoD(video-on-demand)应用的大规模实现提供底层网络的支持,许多正在出现的P2P VoD分发策略都能够提供在P2P网络中最基本的数据传输方式.对以往主要的P2P VoD内容分发策略进行了总结和概括.首先介绍了设计P2P VoD策略的相关重要问题,并把策略根据内容分发方式的不同分成4种类型.最后讨论了它们的应用层性能,并提出未来可以延续的工作.

关键词: 点对点;视频点播(video-on-demand,简称VoD);内容分发;传输路径

中图法分类号: TP393 **文献标识码:** A

随着 Internet 的迅速发展,交互式多媒体业务日益成为网络业务的发展方向,而 VoD(video-on-demand)是最重要的网络多媒体业务之一.

传统的电视节目只能把特定的内容“推”给用户,用户无法任意选择想看的影片.而视频点播系统的设计目标就是在任何时候为用户提供所需要的节目,并支持与用户的互动.VoD 应用根据内容源的不同可以分为两大类:1) 实时流节目直播;2) 预保存节目点播.实时流直播是指实时信号(电视、卫星信号等)通过网络实时传送到用户端的业务;预先保存节目点播是指用户根据自己的兴趣点播预先保存的节目源,这种节目并不会随着时间

* Supported by the National Natural Science Foundation of China under Grant No.60702054 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2006AA01Z203 (国家高技术研究发展计划(863))

Received 2006-12-23; Accepted 2007-05-10

的变化而改变.这两种应用在内容分发策略方面存在着区别.本文所提到的 VoD 服务主要针对第 2 种业务.

VoD 系统最常用的是客户端/服务器(C/S)模式的 VoD 系统,C/S 模式下基本上所有的功能都集中在服务器端,客户端只需要向服务器请求并接收内容.随着用户数的增加,服务器端的性能成为整个系统性能的瓶颈.由于用户点播的行为具有一定的汇聚效果,这可以利用到 IP 组播^[1]的优势.IP 组播能够把多个单播流汇聚成一个组播流,降低服务器端的压力.但是由于安全与管理上的原因,绝大多数的 Internet 服务提供商都没有在路由器上开放 IP 组播功能^[2],因此无法大规模地实现.而且用户的点播行为具有异步性^[3],IP 组播的情况下,用户需要一定的等待时间才能得到服务,无法实现 T-VoD^[4],这有可能是用户无法忍受的.在文件共享系统^[5]和数据库应用^[6]中使用得十分成熟的对等网络(peer-to-peer,简称 P2P)技术能够充分利用闲置的网络资源,实现有效的负载均衡,十分适合于应用在大规模的 VoD 系统中.

无论是哪种方式的 VoD 服务,都需要从一台或数台服务器或者端系统上把视频流数据传输到另一台端系统上,我们称这个过程为“内容分发”,与之相关的其他功能模块有资源定位、传输路由的选择等.内容分发对于系统的服务性能是至关重要的,因此,它也是本文主要关心的方面.

本文主要介绍 P2P VoD 分发策略方面的技术.第 1 节介绍 VoD 的概况.P2P VoD 分发策略的相关问题将在第 2 节中介绍.第 3 节对所提到的策略进行比较.最后总结全文并指出进一步的研究方向.

1 VoD 流服务策略概述

P2P 的 VoD 策略是在传统的 VoD 系统基础上发展而来,当传统的 VoD 策略无法满足现代流媒体的需求时,产生了 P2P 方式的 VoD 系统.

1.1 VoD 流服务特点

典型的 VoD 系统将视频节目存储于一台或多台服务器,接受用户请求,并通过网络将节目内容传递给各用户,用户可以随时随意点播,以观看高清晰的视频节目.VoD 业务可在不同的网络平台提供,如驻地网、ATM 网、CATV 网或者因特网等.与 WWW,FTP,E-MAIL 等一般数据业务相比,VoD 业务需要为每一个用户提供一段较长时间的稳定视频流.该视频流占用的网络带宽高,且要求网络的抖动较小,因此,VoD 业务对承载网络提出了很高的要求.

VoD 的业务要求包括开始延迟(starting delay)、播放质量(playback quality)、VCR 交互操作(VCR-like interaction)、可扩展性(scalability)等.

1.2 传统的流服务策略

传统的 VoD 采用 C/S 单播模式,服务器的负载能力基本上就是整个系统的服务能力,因此不具有较高的扩展能力.在简单 C/S 单播模式的基础上,为了保证用户的服务质量,还可以采用流媒体内容分发网络(content distribution network,简称 CDN)^[7]技术,把视频源推向网络边缘.但要实现大规模的应用,需要部署大量的边缘服务器,代价较为昂贵.

研究表明,由于人类行为的固有特征,绝大多数的用户请求都集中在少数的热门影片,而且由于用户数量巨大,同一时间点播相同视频节目的可能性就高,故此,可以利用一些基于 IP 组播技术的策略来设计 VoD 系统,以节省网络带宽,提高系统容量.采用 IP 组播技术的 VoD 系统一般使用静态或动态服务器通道调度方案,大致分为以下几类:周期广播(period broadcasting)方式^[8]、批处理(batching)方案^[9]、补丁(patching)方案^[10]、Piggyback 方案^[11]和流合并方案^[12]等.

对于流媒体点播业务,组播可以把多个业务流汇聚成一个流来服务.但 IP 组播的难以控制性,使得现在 Internet 的主干网上都没有开放这项功能,只能在局域网的范围内使用.用户行为的异步性也决定了 IP 组播不适用于预保存节目的 VoD 业务.

C/S 构架的服务器瓶颈效应和 IP 组播的难以实现,使得更多人开始研究 P2P 在 VoD 中的应用.在 P2P 网络中,每个节点(peer)要承担双重任务,首先要从其他节点获得服务,其次还要网络中的其他节点提供服务,一般还

要具有一定的路由和查询功能.P2P最大的优势在于不必改变传统的网络结构,在现有 Internet 层次模型上,只需改动应用层的分发和路由策略,就可以相对 C/S 模式大幅度提高扩展性,且易于实现,代价是整个网络的负载增加.这些增加的负载大部分是平时闲置的节点上行带宽.此外,P2P VoD 还存在播放质量无法保证、VCR 交互操作实现困难等问题.

表 1 列出了 3 种流服务策略对 VoD 业务性能的影响.

Table 1 The effects of different streaming strategies on VoD service performance

表 1 不同流服务策略对 VoD 业务性能的影响

	Starting delay	Playback quality	VCR-Like interaction	Scalability
Client/server (including CDN)	Low	QoS assurance	Full function	Small
IP multicast	High	QoS assurance	Hard to implement	Large
P2P model	Depending on schemes	Without QoS assurance	Hard to implement	Large

1.3 P2P预保存节目点播与实时流节目直播的区别

最早的P2P与流媒体技术的结合最先产生的是P2P的实时流节目直播系统,从传统的树型分发,如ZIGZAG^[13],到现在基于Gossip的纯Mesh分发,如Coolstreaming^[14]和Anysee^[15].实时流节目系统已经先于P2P预保存节目实现大规模的应用.

实时流节目的直播与预先保存节目的点播在实现策略上存在着区别,这是由这两种业务的不同之处而形成的:

- 点播的资源数目:P2P 预保存节目点播有较多数量的影片供点播,P2P 直播的频道数较少,用户有可能长时间观看.
- 内容源:P2P 预保存节目点播已知要播放的全部内容,可以安排调度播放的内容,而 P2P 直播一般内容未知.
- 用户观看要求:P2P 预保存节目点播用户可以从一部影片的任何部分开始观看,P2P 直播随到随看,各个用户间的观看内容基本保持一定程度的同步.
- 对端到端延时的敏感程度:P2P 预保存节目点播只要满足播放要求,并不要求实效性,P2P 直播对与内容源的时间同步要求较高,如观看现在直播节目.
- 交互操作:P2P 预保存节目点播存在各种交互式操作,例如快进、快退等,在 QoS 没有保障的情况下,用户有可能停止一段时间等缓冲结束后再继续观看,P2P 直播一般只提供播放和停止操作.

其中与内容分发策略联系较为紧密的是用户的观看要求,大量用户在相同的时间内观看相同的内容会极大地简化系统的设计,直播流节目点播满足这种情况.预保存节目点播的用户请求存在异步性,因此在设计 P2P 系统时应考虑,必要时采用其他模式和策略,如处理冷门影片时,另外还应加入 VCR 操作.

2 P2P VoD 系统的实现

2.1 P2P VoD系统功能组成

P2P 系统很重要的一个特征就是节点智能化,节点开始承担系统中的更多功能.在不同的 P2P 系统中,它的功能分布是不同的,但一个完整的 P2P VoD 系统中一般具有以下几个部分的功能(如图 1 所示):

- (1) 安全认证机制:P2P 网络的一个优势在于其匿名性,但这也增加了不安全因素.通过身份验证可以确定用户的使用权限,控制管理在其权限范围内获得资源,并保护其利益不受其他非法用户的影响.
- (2) 查找定位机制:新加入的节点需要知道到哪些节点上获得特定的资源或者知道某些节点上有哪些资源,系统中必须有节点来提供这种服务,不论是集中的还是分布的.
- (3) 内容分发策略:如何调度数据包的分发以及分发的路径关系到网络的负载平衡和节点的 QoS 保证.这也是本文主要关注的内容,其中重要的研究技术有覆盖网络的构建和分发策略等.内容分发策略也是本文关注的重点.

- (4) 数字版权管理(digital right management,简称 DRM):理论上,P2P 网络中的每个节点都可以成为内容的传播者,因此,需要保证具有版权的内容不被未授权用户所获得.在 VoD 的业务中需要涉及到的是数字版权管理.
- (5) 用户行为分析:一个完善的服务系统需要考虑到它的服务对象的行为特征,针对这些特征设计不同的策略模型.
- (6) 激励机制:P2P 强调的是节点的参与和共享,各种不同的激励策略可以有效地激励用户更多地参与到共享中来.

安全认证是所有功能模块的基础,查找定位资源是内容分发策略的前提,其他的一些功能模块需要通过底层的这些机制来实现.与 C/S 不同,P2P 网络的一些特征使这些功能模块的实现存在一定的难度.

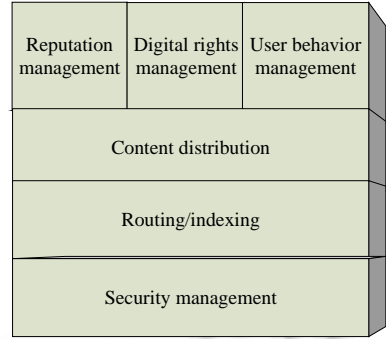


Fig.1 P2P VoD function model

图 1 P2P VoD 功能模块

2.2 P2P VoD系统的分类

现有的研究对 P2P 流媒体一般采取以下几种分类方式:

- (1) 按照P2P分发数据的overlay结构分为结构化和非结构化^[16];非结构化进一步又可分为Mesh First和 Tree First^[17];
- (2) 根据服务器的体系结构,可以分为单/多索引节点多服务器(SIMS/MIMS)、多独立索引服务器(MIIS)和完全分布式多服务器(FAMS)^[18].

另外,根据发送节点的情况可以分为单源和多源策略;根据适用的网络不同可以分为应用于小规模网络和应用于 Internet 的 P2P VoD 系统等.随着研究的进展,特别是内容分发策略的不断改进和编码技术的发展,需要更进一步地对现有的 P2P 流媒体特别是 VoD 根据内容分发策略重新分类.我们在第 3 节中的分类结合了以往的分类方法,对于编码方式的应用作了特别的说明.

2.3 P2P VoD实现的难点

不论哪种 P2P VoD 系统,P2P 网络本身存在的问题让这项技术的应用在许多方面都存在着巨大的挑战.

- 非中心化管理:P2P 的最大特征就是不存在制约整个网络性能的中心节点,但作为服务提供商会面临着许多难以控制的后果,如非法用户的恶意攻击、各个节点想尽办法逃避义务而同时还可以获得其他节点的服务等等.
- 节点的不可靠性:在 P2P 系统中,每个节点都是独立的个体.VoD 在播放时实时性要求较高,不论节点的加入或退出都会对其他节点产生影响.因此,如何把节点的不合作行为对其他节点造成的影响降到最低,对于保证服务的质量是十分重要的.
- 网络异构性:现在的 Internet 是一张很大的混合网络,ADSL,ISDN,T1 等不同的接入技术使得不同的节点间具有不一致的特性,如不同的上下行带宽、不同的拓扑结构、不同的通信协议和不同的服务质量等等,如何克服这些不一致是系统设计者需要解决的.

P2P VoD 的各种内容分发策略在设计过程中都需要考虑到以上的问题,对这些问题解决的程度是评价一个 P2P VoD 系统的标准之一.

3 P2P VoD 内容分发策略

3.1 P2P VoD内容分发策略模型描述

在 P2P 的网络中,所有的节点都是对等的,除非某些网络中设置一些超级节点负责管理局部网络的事务.另外,每个节点都对网络中的某些内容具有兴趣,或者其所拥有的某些内容其他节点会感兴趣.

内容分发算法的目标是建立起从源到目标接收节点的最小延时和满足播放质量的分发网络.由于网络中

资源的存放方式不同,分发策略可以分为单源的和多源的策略.另外,根据分发过程中数据的传输途径的不同,又可以分为单路径和多路径的策略.还有的策略对不同的节点进行了分代或分组,根据内容,分发策略可以分为以下4类:单源单路径、单源多路径、单源分代和多源多路径.

3.2 单源单路径策略

大部分单源单路径策略不仅内容源是集中式的,控制信息一般也是集中管理的.这种方式对网络和服务器要求较高.

3.2.1 链式结构(chaining)^[19]

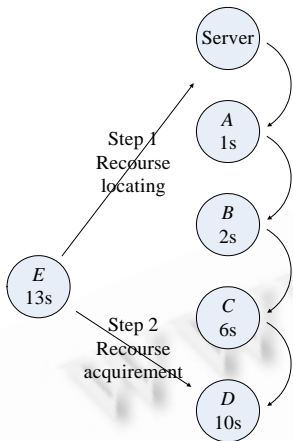


Fig.2 Chain structure

图2 链式结构

Chaining 是最早把 P2P 思想引入视频点播的服务模型,所有请求同一部影片的节点按照到达的顺序串联成一条链状结构.在节点链中,前面的节点在缓冲区保存一部分已经看过的视频片断,为后面的节点提供服务.后到的请求是否可以得到前一个节点的服务,取决于两者到达时间的间隔,如果这个时间长于前面节点的缓冲区大小,就无法形成一条连续的链,还需要直接到服务器获取资源.

为了能够形成一条较长的链,可以有几种改进措施,例如,GloVE^[20]采用 Patching^[21]+Chaining 的策略,链上的节点可能丢失的影片开头部分可以到服务器获得;ACVoD^[22]利用空闲节点,将原本属于不同链的节点或者不能够连接到链的节点最终加入到同一条链,并增强链的稳定性.

在所有的策略相同的节点数的情况下,链的层次最深,这就意味着节点中途离开造成的影响最大.而且,现在的 Internet 不一定支持一对一大流量、稳定的视频传输.因此,链式结构的传输只能支持局域网内的小规模应用.

在最基本的链式节点的策略中,节点的加入过程如图 2 所示,设缓冲区窗口为 5s,新加入节点 E 要先到服务器进行资源定位,由于它与链尾节点的到达时间相差在缓冲区窗口内,可以直接接到链尾,否则,服务器要产生另一条视频流.

3.2.2 树状(tree)结构

Chaining 实际上是树状结构的极端情况.在树状组播树中,为了减小上层节点失败对下层节点的影响,可以通过减少层次来实现.在带宽和计算能力允许的情况下,增加每个节点所带的子节点,而且尽量保证上层节点的稳定性.典型的策略有 DirectStream^[23],P2Cast^[24]等.

以 P2Cast 为例,P2Cast 是在 DirectStream 的基础上改进的,节点按照到达的时间加入一个应用层组播树(如图 3 所示).新节点加入组播树,选取与自己到达时间较为接近的节点为父节点,这个节点有可能已经把最开始的一部分内容抛弃,即它们到达时间的间隔已经超过缓冲区的大小,如果是 DirectStream 策略,该节点就无法成为新节点的父节点.但在 P2Cast 策略中,如果这个时间没有大于一个所谓 patching 窗口的大小,则新节点仍然可以从父节点获取视频流,同时还需要从其他拥有开头部分节目的节点处获得该部分内容,即需要寻找一个 patching 流的父节点.这样能使更多的节点加入到组播树中,减少了服务器新开组播树的数量,并且能够均衡负载.

在所有单源单路径的实现中,对系统性能影响最大的策略就是父节点的选择策略.

一个理想的父节点需要包括以下一些特性:1) 稳定性:理想状态是在子节点的播放期间都不退出;2) 高带宽:父节点与子节点间保持良好而稳定的满足播放的带宽;3) 低延迟:父节点与子节点间的路由延迟最低,许多情况下跳数越少,延迟越低.在所有可以为新加节点提供服务的节点中如何选择父节点,一般有以下几种算法:

- 1) 随机选择.
- 2) 最大带宽选择策略:选择测量带宽最宽的节点为自己的父节点.
- 3) 最小延迟选择策略:选择到达新加节点延迟最小的节点为父节点.
- 4) 最小深度选择策略:使最后形成的多播树的平均层次最小.

- 5) 基于用户历史行为的选择策略:根据用户的历史行为,估计其将来的行为特征,把行为较为稳定的节点尽量靠近多播树的根部.
- 6) 基于经济学原理^[25-27]:假设每个用户的决策是自私的,通过引入市场竞争机制,并通过优胜劣汰来决定父节点.

一些算法还兼有两种或两种以上的优点,如DirectStream策略计算距离和带宽的比值,即 n_i/x_i (n_i 为端到端的跳数, x_i 为可用带宽),选择比值最小的为父节点.

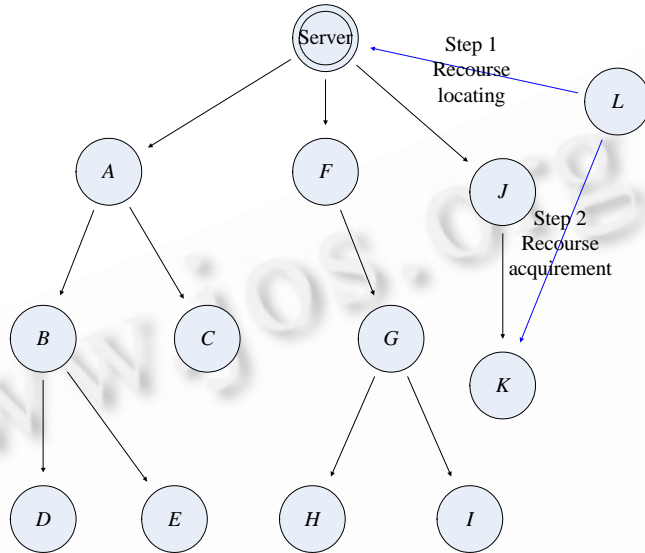


Fig.3 Tree structure
图3 树状结构

单源单路径的策略是P2P VoD内容分发策略中最简单的策略,存在着许多问题,例如:1) 所有的叶子节点都没有承担转发负载,这就会使非叶子节点的负担过重,有可能成为网络中的新瓶颈;2) 对网络性能的要求很高,不适用于大规模的应用;3) 在节点中途退出情况下很难实现无缝播放;4) 实现VCR操作有难度等.

3.3 单源多路径策略

为解决单源单路径的负载不均衡和其他问题,可以采用多路径分发策略,即源节点把需要分发的内容分成多个小段数据,通过不同的路径发送,形成多棵组播树.一个节点在某棵树上里是叶子节点而在某棵树上又是内部节点,从而较好地达到负载平衡.节点在接受时从多个路径接收,再通过解码恢复成视频流.另外,可以通过编码技术增加冗余,增强容错性,即使有少数链路失效,也可以观看.

3.3.1 普通多路径方案

SplitStream^[28]把视频流分割成不同的Stripes,分别发送给不同的节点,由这些节点形成新的不同的组播树.这样同,完整视频需要到多个节点处获得.一个节点往往是在一个组播树中充当转发的角色,在另一个组播树中就是叶子节点.这样,当一个节点发生故障或退出时,只会影响到小部分的节点.但是在SplitStream策略中,空闲的节点可能需要为其他节点转发自己不感兴趣的视频流.

类似的策略有CoopNet^[29],Scribe^[30]等,不同的分块数目会对系统性能造成影响.文献[31]对此进行了讨论.

3.3.2 分层编码、多重描述编码与P2P VoD分发策略

Internet的异构性不能实现点对点传输一条稳定视频流的QoS保证,因此需要减小调度分发数据的颗粒度,以达到一种全局与部分的优化.要把内容源分成多个小段的数据,需要用到视频的编解码技术.

分层编码(layered coding,简称LC)^[32]和多重描述编码(multiple description coding,简称MDC)^[33]是两种最基本的在不可靠信道多路径传输的编码方式,其基本思想都是把数据编码到多个字码流上,在播放时根据需要或

网络状况的限制,只需接收到全部或部分字码流就可以播放.LC把视频流编码成一个基本层和若干个增强层.基本层包含最基本的服务质量,可以单独编码,增强层的内容用于提高服务质量.而且各个增强层之间也具有层次关系,即需要接收到所有底下的增强层才能解码上面的增强层.考虑到在Internet中,不同网络中的节点具有异质性(能够接受的最大流量不同),因此需要采用不同的传输速率.对于流媒体而言,LC是一种不错的应用于多码率传输系统的编码方式.传输带宽大的节点可以接收较多层的流量,带宽小的节点只能接收较少层次的流量.在不同层次的传输过程中,优先照顾底层的传输要求,如若带宽不足,则优先抛弃较高的层.这样能够保证大部分节点的正常观看,具有一定的公平性.

MDC 的不同流之间则没有层次关系,每个都能独立地保证一定的视频质量,接受到的流越多,视频质量越高,适合高误码率的信道传输,但 MDC 编、解码的代价较高.

不同的P2P分发策略根据需要,采用不同的编码方式,如CoopNet采用的是MDC编码方式,而文献[34]和PALS^[35]采用的是LC编码方式.其他编码技术还有最近新兴的网络编码.

3.3.3 网络编码与 P2P VoD 分发策略

网络编码(network coding,简称NC)^[36]技术在P2P流媒体上的主要应用方式是源节点在发送时把视频流分成小段数据后,多个块进行编码后发出.接收节点只有在收到足够的线性无关数据包并解码之后才能获得原始数据.网络编码技术结合P2P流媒体传输能够提高网络的吞吐能力,同时又因为经过编码之后的每个块中都包含了多个原始块的信息,在一个视频段内不存在需要调度的先后问题,这样就很好地解决了分段传输的调度问题.但同时,网络编码也会引入编、解码的开销以及需要收集到该段所有块才能解码而造成的延时问题.

网络编码的理论基础是最大流最小割定理,多源的最大流最小割定理目前在理论上还没有得到证明,因此,目前网络编码在 P2P VoD 中的应用只限于单源多路径.主要有以下几种:

3.3.3.1 按时间先后分段方案

文献[37]中提出的方案是把视频流按照时间先后分成不同的 GOP,每个 GOP 作为网络编码的单元再分成 N 组,这样需要收到一个 GOP 中所有 N 组的信息才能复原出 GOP.编码方式采用随机编码,在编码之后的包中还附着 GOP 的组标识及其所含的向量信息,这样不需要查看完整的包就可以通过这些附带的信息辨别这个包是否是自己所需要的.

仿真实验的结果表明,在请求强度不大的情况下采用网络编码的方式可以提高服务的质量和系统性能,但是在请求强度达到一定程度时,就不存在这种优势,反而会因为编、解开销影响了性能.因此,应用于 P2P 流媒体中的网络编码策略只有在一定的情况下才会体现出它的优势.

由于流媒体传输是时延敏感性的网络应用,网络编码需要接收节点在收到所有的相关数据包时才能解码得到原来的数据,这会带来不小的延时.文献[38]提出了一种对播放时刻敏感的网络编码方案.发送方要根据接收方的播放时刻要求,确定编码的窗口大小,窗口的大小等于缓存中保存的块的最大时刻与下一需要播放的时刻的差值,这样就能保证在播放到某个块时已经把这个块的信息接受并解码了.对于需要及时传输的数据服务器可以不予编码,直接发送给接收方,所有节点只在可以接受编码延时范围内进行网络编码.

3.3.3.2 多重描述编码(MDC)与网络编码结合方案

CodedStream^[39]是一种混合了MDC与网络编码的传输策略.视频流首先按照MDC编码方式产生 k 个互相补充的描述子,这 k 个描述子再相继利用网络编码方式传送到一个建立好的多播图中,图中每个接收节点需要有 k 个邻居节点,这样才能保证从 k 个不同的路径收到不同的编码包.它混合了MDC和NC的优势,MDC增强了网络的容错性,即使没有接受到全部的描述子,也能顺利地播放视频,而NC增加了网络传输的吞吐量,在同样的时间内,传输的信息量最大.

CodedStream 的缺点在于事先需要建立网络传输的多播图,而且这个多播图中需要利用到一些核心节点作为传输的主干,但这些核心节点又不是服务的需求者,这增加了网络部署的难度.

3.3.3.3 分层编码(LC)与网络编码结合方案

文献[40]主要结合了 LC 和网络编码的策略.这种策略首先根据节点所能接收的最大码率把节点分成不同

的子集,并根据其大小排序.码率大的节点加入码率小的节点集中,根据这些节点集把图分割成不同的子图.在每个子图中求出每个节点的最大流的路径,即在每个子图中满足不同码率要求的节点的最大流.

LION^[41]在源节点采用的也是LC,而且网络编码是在层内进行的.在原节点分层编码之后,每个层次的数据再分成小块,分发到实现建立好的覆盖网络中(覆盖网络的建立与Narada^[2]的策略相同),通过不同的路径传输.接收节点采用尽量多的符合质量要求的路径(max-disjoint path),如果这些路径的容量允许,则节点将尽量加入更多的层次.这种方法能增加中间节点转发包时采用网络编码的概率.

这两种方案也存在一些缺点,如没有考虑到节点的动态加入与退出:如果有节点临时加入,网络已经基本饱和,如何保证其也能够获得基本的层以保证观看?如果有节点中途退出,如何再分配这些多余的带宽,是重新计算最大流还是保持原状?另外,LION 策略中采用了 Flood 算法,也会带来额外的开销.

3.4 单源分代

对于在某些到达时间相近的节点,可以根据其固有的缓冲区特性把它们归为一个层次或者一个“代”.这样,不同的“代”之间具有了前后相继的特点,在逻辑上可以形成一条链式或树状结构,但在每一代的内部可以是其他拓扑结构,例如中心式、完全对等或者树状结构等.

分代结构采用的是一种混合式的拓扑结构.它综合了几种结构的各自优势,更好地发挥了 P2P 构架的优势.但要充分发挥分代结构的优势,必须首先满足有足够密集的点播请求,才能够保证在每个层次中具有一定数量的节点,否则有可能退化成简单的类似单源单路径结构,甚至完全链式结构.

采用分代结构的主要策略有P2VoD^[42],LEMP^[43]等.

P2VoD 最早引入了“代”的概念,到达时间在一定范围内的节点为一“代”,上一代为下一代提供服务.代不仅仅是时间相近的节点的集合,还要满足缓冲区的起始内容是一致的,这就造成不同到达时间的节点的缓冲区大小不一样.当同代中有节点失效时,能够在同一代中快速找到替代的节点.

轻量高效广播协议(LEMP)是让在一个窗口时间中的所有节点都加入一个层次,在每个层次中服务器会根据一定的标准来选择出一个节点作为LR(local representative).在正常情况下,一个层次中只有LR与服务器的联系,因此,一般选择距离服务器最近的为LR.在一个层次中还要选出另一个节点作为备用的LR(BLR),在LR退出时替代它.

在同一层次中,LR 作为中心节点,与其他节点形成星型结构,为其他节点提供服务.除了从服务器到第 1 层以外,其他层次之间通过 LR 传递视频流.网络逻辑拓扑结构如图 4 所示.

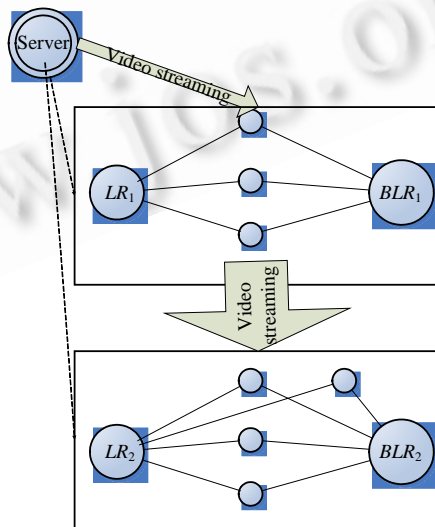


Fig.4 LEMP layered structure
图 4 LEMP 的分层结构

3.5 多源多路径策略

在某些特定的应用环境中,并不是只有一个节点拥有视频源,因此可以充分利用多个视频源来接收资源.在多源多路径结构中,没有相对的中心节点,内容源可能已经实现分布在不同的节点上(经过一定时间的观看或下载),所有节点都可以成为发布者和接受者.

如图 5 所示,节点 A,B,C 分别拥有一部视频的第①,②,③段,为了正常播放,A 需要到 B 和 C 获取第②,③部分;同理,B 要到 A,C 获得第①,③部分,C 要到 A,B 获得第①,②部分,D 要到 A,B,C 获得第①,②,③部分.

采用多源多路径结构的主要策略有 GnuStream^[44]和 PROMISE^[45]等.

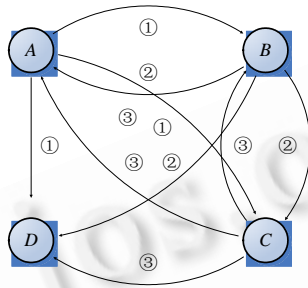


Fig.5 Mesh structure

图 5 网状结构

3.5.1 GnuStream

GnuStream 是一种完全对等的网络结构,不存在前面几种策略中所需要的中心控制索引节点.它的资源查找利用了了在文件共享系统中比较著名的 Gnutella 协议,其命名也因此而来.数据分发采用接收者驱动模式,即接收者主动搜索感兴趣的内容,向拥有这些内容的节点请求下载.这种方式类似于传统的 P2P 文件下载方式.

可以把这种多对一的 P2P VoD 传输的内容分发问题归结为两种算法来解决:(1) 如何挑选最为合适的节点集为一个节点服务;(2) 挑选出的节点如何分配传输的内容和带宽.文献[46]提出的数据指派算法 OST_{P2P}和接纳控制协议 DAC_{P2P}就是针对这两个问题的一种方案.它把不同的源节点根据不同的带宽分成不同的集合,数据分配根据的就是其所在的集合及其所拥有的数据.首先,要保证下载的带宽大于或等于其所需带宽;其次,播放之前所需要缓冲的最迟播放块由带宽最大的源节点传送,以此类推.另外,源节点在处理接收节点的请求时,根据节点的等级以不同概率传送,一般来说,具有较大下载带宽和提供较大上传的节点具有较高的等级.低等级的节点在几次拒绝后会不断提高它们的概率,防止“饿死”.

但是在这个方案中,没有解决节点的动态性问题.另外, DAC_{P2P}算法也存在缺乏公平性问题.

3.5.2 PROMISE

在 GnuStream 的策略上, PROMISE 提出了新的应用层组播策略,称为 CollectCast.

PROMISE 的底层网络可以是任何已有的 P2P 网络,作者选择了 Pastry 作为其底层的结构.UDP 传送数据包, TCP 传送信令.在选择最优的发送端时采用了拓扑感知技术,把网络的状况作为选择的依据.

通过底层的网络得到资源节点的候选集合,集合中的每个节点有两个属性:带宽速率和有效性.在这个集合中找到速率期望值最高而且能够满足需要的集合,确定活动集合(active set).在播放过程中,还需要不断监测网络节点的状况,包括丢包率和传送速率,淘汰掉不适合的节点,更新活动集.

这种方法的缺点也是很明显的,即需要了解太多的底层参数,算法复杂度高.

3.5.3 Mesh-Cast^[47]

CollectCast 算法模型是基于多对一传输的优化算法,而实际中 P2P 流媒体的传输方式是多对多,因为有可能存在同时相同的请求.因此,CollectCast 算法在处理这些后来的相同请求时,可能产生更多的网络阻塞.文献[47]指出,这实际上是多对多的传输.Mesh-Cast 策略是在 CollectCast 基础上,当后来的请求发生网络拥塞时,选取一个外围联接节点(peripheral articulation node,简称 PAN)加入,成为数据的中继站,从而改变数据的传输路

线,避免网络拥塞.PAN 的特点是,当引入该节点后会增加原来数据流图中新的路由线路.

4 P2P VoD 内容分发策略比较

4.1 各内容分发策略算法比较

表 2 对比了前面提到的部分内容分发策略的算法,具体算法详见第 2 节.

Table 2 Comparison of content distribution strategies

表 2 内容分发策略的比较

	Transmitting path structure	Construction of overlay	Dealing with partners	Coding strategy	Type of management	Idle peers employment	Scalability
Standard chaining	Chain	Base on arriving time	Fixed	None	Centralized	No	Small
ACVoD	Chain	Base on arriving time	Fixed	None	Centralized	Yes	Small
DirectStream/P2Cast	Tree	Best-Fit arithmetic ^[24]	Fixed	None	Centralized	No	Small
SplitStream	Multiple Tree	Pastry/scribe/ etc.	Fixed	None	Centralized	No	Medium
Network coding based on GOP	Mesh	Gossip	Periodic diversification	NC	Distributed	No	Large
Deadline-Aware network coding	Mesh	Gossip	Instantaneous	NC	Distributed	No	Large
CodedStream	Mesh	Narada mesh ^[2]	Periodic diversification	MDC and NC	Distributed	Yes	Small
Multirate media streaming	Mesh	LIF arithmetic ^[48]	Instantaneous	LC and NC	Distributed	Yes	Small
LION	Mesh	Narada mesh/ etc.	Instantaneous	LC and NC	Distributed	Yes	Small
P2VoD	Hybrid	Generation arithmetic ^[42]	Fixed	None	Centralized	No	Medium
LEMP	Hybrid	LEMP ^[43]	Fixed	None	Group control	No	Medium
GnuStream	Mesh	Gnutella	Instantaneous	None	Distributed	No	Small
PROMISE	Mesh	Pastry/ chord/etc.	Instantaneous and periodic diversification	None	Distributed	No	Small

4.2 各种内容分发策略性能比较(见表3)

Table 3 Comparison of content distribution performances

表 3 各种内容分发性能的比较

	Client/ Server	Single source and single path	Single source and multiple path	Single source and generation partition	Multiple source and multiple path
Load balance	Poor	Medium	Good	Medium	Good
Management overhead	Light	Medium	Heavy	Medium	Heavy
Fairness	Medium	Poor	Good	Medium	Good
Fault tolerance	Poor	Poor	Good	Medium	Good
Adjustability to network heterogeneity	Medium	Poor	Good	Poor	Good
Scalability	Small	Medium	Large	Medium	Large

4.2.1 负载均衡

P2P 与传统的 C/S 模式相比,其最大的进步就在于负载均衡的改善,参与的节点承载了很大一部分原来服务器的负担,基本上解决了单点失效的问题.但是在单路径传输策略中,大量的叶子节点没有参与承担上传负载,处于上层的节点的负载仍较大,一旦失效,也能造成相当大的影响.在多路径传输策略中,每个参与的节点都要承担一部分负载,因此,网络中存在较少的瓶颈节点.在分代策略中,每个代中的管理节点虽然分担了很大一部分源服务器的负担,但仍有可能成为性能瓶颈.

4.2.2 管理开销

管理开销与所管理内容的颗粒和编码方式紧密相关.多路径策略需要把视频分割成比较小的块(或段),不仅编、解码需要较大开销,而且分配多个节点数据的分发需要全局的调度策略,节点间需要相互通过大量及时的通信才能协调工作,也会带来很大的开销,这也是多路径策略设计中的困难所在.相比而言,C/S 结构和单路径策略调度的对象是单个流,所需管理的信息也较少,开销较小.单路径策略还要考虑到资源的定位和负载均衡,因此开销比 C/S 传输要大.

4.2.3 公平性

公平性包括每个节点的负载是否均衡、获得的服务质量是否公平等.负载均衡前面已经提到.服务质量在不同的策略中也无法保证公平性.在单路径策略中,位于下层的节点面临着服务失效的更大风险,因为上层任意一个节点离开都会影响到自己;但在多路径传输中尽管也存在这个问题,但其完全失效的概率大为降低,因为多个路径中大部分父节点同时失效的概率是很小的.

4.2.4 容错性

C/S 模式和单路径策略都要满足端到端的 QoS 保证,才能保证节点的正常播放,一旦传输中某个环节发生错误就影响到正常播放.而多路径策略中一般引入了冗余算法,即使有部分内容缺失也可以继续播放,只影响播放的质量,而不会中断播放.

4.2.5 对异质网络的适应性

现在的 Internet 是由许多不同性质的网络族组成,不同网络间存在着差异.这些差异对 P2P VoD 业务的影响可以归结为带宽、稳定性等.不同网络间的通信状况一般不十分稳定,往往满足不了一对一的传输任务.因此,单路径策略只能在小规模的局域网中实现,多路径传输必定是目前 Internet 的选择.

4.2.6 扩展性

C/S 架构是几种策略中扩展性最差的,服务器是整个系统的瓶颈.当服务器达到服务能力的上限时就不可能再为其他请求服务了.P2P 的其他策略很好地分散了服务器的负载,使得在扩展性方面具有很大优势,一个新加入的节点如果已经存在所需内容的组播树,则只需加入到一个或几个组播树中就可以获得服务,而无须占用服务器的带宽,而且理论上只要每个节点具有为一个以上节点提供服务的能力,这种扩展性是无限制的.但是在实际设计中,这种扩展性还受到节点之间通信代价的限制.

4.2.7 分发策略比较总结

对于 P2P VoD 分发策略,多路径策略体现出了在负载均衡以及扩展性方面的诸多优势,比较适合于现在的异构环境的 Internet.单路径策略尽管在这些方面不具备优势,但是如果未来网络性能较优的环境下,可以体现出它在管理开销和系统设计难易程度上的优势.

5 总结与展望

VoD 作为目前网络上最主要的应用之一,传统的 C/S 模式的 VoD 系统已经不能够满足现在的需求,无法实现大规模的应用.P2P 作为一种新兴的网络模型,为 VoD 的大规模应用提供了解决方案.在本文中,我们讨论了 P2P 网络在 VoD 中的应用,提出了 P2P VoD 的功能架构,重点介绍了单源与多源、单路径与多路径的 P2P VoD 分发策略,并对这些策略进行了比较.但是,目前的这些工作与实际的大规模应用还有一定差距.

大规模 VoD 的应用是未来网络中十分重要的应用,P2P 是目前其实现最为经济、有效的方式.广电网的双向改造为 P2P 流媒体的实现提供了双向、高带宽、低延时的网络;同时,广电网大规模的用户群也为 P2P 的实现提供了足够的用户规模.层次化的网络架构也使其比较适合流媒体的分发,邻近的节点之间可以获得较小的开始延时,而且局部的负载对主干网络的流量影响不大.在改造后的广电网实现 P2P 的流媒体分发是一种可行的方案,可以有效解决服务器流量瓶颈的问题.

但是,P2P 固有的一些性质会阻碍其在商业中的应用.因此我们认为,P2P 还需要与传统的流服务策略进行相应的结合.例如,P2P 与 CDN 网络融合,互补优劣,才能实现真正大规模的应用.另外,为了保证流媒体的可靠传

输,结合相应的编码手段,如网络编码技术,也是十分必要的。

References:

- [1] Quinn B, Almeroth K. IP multicast applications: Challenges and solutions. IETF RFC 3170, 2001.
- [2] Chu YH, Rao SG, Zhang H. A case for end system multicast. In: Kurose J, Nain P, eds. Proc. of the ACM SIGMETRICS 2000. Santa Clara: ACM Press, 2000. 1–12.
- [3] Ma HD, Shin KG. Multicast video-on-demand services. ACM Computer Communication Review, 2002,32(1):31–43.
- [4] Little T, Venkatesh D. Prospects for interactive video-on-demand. IEEE Multimedia Magazine, 1994,25(3):14–24.
- [5] Stephanos AT, Diomidis S. A survey of peer-to-peer content distribution technologies. ACM Computing Surveys, 2004,36(4): 335–371.
- [6] Yu M, Li ZH, Zhang LB. P2P data management. Journal of Software, 2006,17(8):1717–1730 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/17/1717.htm>
- [7] Day M, Gilletti D. Content distribution network peering scenarios. draft-day-cdn-scenarios-02.txt, 2000. <http://www.contentalliance.org/docs/draft-day-cdn-scenarios-02.html>
- [8] Almeroth KC, Ammar MMH. On the use of multicast delivery to provide a scalable and interactive video-on-demand service. IEEE Journal of Selected Areas in Communications, 1996,14(6):1110–1122.
- [9] Dan A, Sitaram D, Shahabuddin P. Scheduling policies for an on-demand video server with batching. In: Proc. of the ACM Multimedia. New York: ACM Press, 1994. 15–23.
- [10] Cai Y, Hua KA, Vu K. Optimizing patching performance. In: Dilip D, ed. Proc. of the MMCN'99. Washington: SPIE Press, 1999. 204–216.
- [11] Aggarwal CC, Wolf JL, Yu PS. On optimal piggyback merging policies for video-on-demand systems. In: Proc. of the Conf. on Measurement and Modeling of Computer Systems. New York: ACM Press, 1996. 200–209.
- [12] Eager D, Vernon M, Zahorjan J. Minimizing bandwidth requirements for on-demand data delivery. IEEE Trans. on Knowledge and Data Engineering, 2001,13(5):742–757.
- [13] Tran DA, Hua KA, Do T. Zigzag: An efficient peer-to-peer scheme for media streaming. 2006. <http://www.ist.psu.edu/tran03zigzag.html>
- [14] Zhang X, Liu JC, Li B, Yum P. Coolstreaming/donet: A data-driven overlay network for efficient live media streaming. In: Proc. of the IEEE INFOCOM. 2005. <http://citeseer.ist.psu.edu/zhang05coolstreamingdonet.html>
- [15] Liao X, Jin H, Liu Y, Ni L, Deng D. Anysee: Peer-to-Peer live streaming. In: Proc. of the INFOCOM. 2006. 1–10.
- [16] Lua EK, Crowcroft J, Pias M, Sharma R, Lim S. A survey and comparison of peer-to-peer overlay network schemes. Journal of IEEE Communications Survey and Tutorial, 2005,7(2):72–93.
- [17] Abad C, Yurcik W, Campbell RH. A survey and comparison of end-system overlay multicast solutions suitable for network centric warfare. In: Proc. of the SPIE, Battlespace Digitization and Network-Centric Systems IV. 2004. 215–226.
- [18] Kalogeraki V, Delis A, Gunopulos D. Peer-to-Peer architectures for scalable efficient and reliable media services. In: Proc. of the 17th Int'l Symp. on Parallel and Distributed Processing. Washington: IEEE Computer Society, 2003. 29–32.
- [19] Sheu S, Hua KA, Tavanapong W. Chaining: A generalized batching technique for video-on-demand. In: Proc. of the Int'l Conf. on Multimedia Computing and System. Washington: IEEE Computer Society, 1997. 110–117.
- [20] de Pinho LB, de Amorim CL, Ishikawa E. GloVE: A distributed environment for low cost scalable VoD systems. In: Proc. of the Computer Architecture and High Performance Computing. Washington: IEEE Computer Society, 2002. 117–124.
- [21] Hua KA, Cai Y, Sheu S. Patching: A multicast technique for true video-on-demand services. In: Proc. of the 6th ACM Int'l Conf. on Multimedia. New York: ACM Press, 1998. 191–200.
- [22] Lin F, Zheng C, Wang X, Xue X. ACVoD: A peer-to-peer based video-on-demand scheme in broadband residential access networks. Int'l Journal of Ad Hoc and Ubiquitous Computing, 2007,2(4):225–231.
- [23] Guo Y, Suh K, Kurose J, Towsley D. A peer-to-peer on-demand streaming service and its performance evaluation. In: Proc. of the IEEE ICME 2003. IEEE Computer Society, 2003. 649–652.
- [24] Guo Y, Suh K, Kurose J, Towsley D. P2cast: Peer-to-Peer patching scheme for VoD service. In: Proc. of the 12th Int'l Conf. on World Wide Web. New York: ACM Press, 2003. 301–309.
- [25] Shrivastava V, Banerjee S. Natural selection in peer-to-peer streaming: From the cathedral to the bazaar. In: Proc. of the Int'l Workshop on Network and Operating Systems Support for Digital Audio and Video. New York: ACM Press, 2005. 93–98.
- [26] Vishnumurthy V, Chandrakumar S, Sizer EG. Karma: A secure economic framework for P2P resource sharing. In: Proc. of the 2005 Conf. of the Centre for Advanced Studies on Collaborative research. IBM Press, 2005. 185–199.
- [27] Chu Y, Chuang J, Zhang H. A case for taxation in peer-to-peer streaming broadcast. In: Proc. of the ACM SIGCOMM Workshop on Practice and Theory of Incentives in Networked Systems. New York: ACM Press, 2002. 205–212.
- [28] Castro M, Druschel P, KermarrecA, Nandi A, Rowstron A, Singh A. SplitStream: High-Bandwidth content distribution in cooperative environments. 2006. <http://citeseer.ist.psu.edu/castro03splitstream.html>
- [29] Padmanabhan VN, Wang HJ, Chou PA, Sripanidkulchai K. Distributing streaming media content using cooperative networking. In:

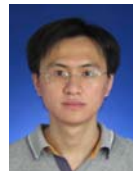
- Kevin A, ed. Proc. of the ACM/IEEE NOSSDAV 2002. New York: ACM Press, 2002. 177–186.
- [30] Castro M, Druschel P, Kermarrec AM, Rowstron A. SCRIBE: A large-scale and decentralized application-level multicast infrastructure. IEEE Journal on Selected Areas in Communications, 2002,20(8):1489–1499.
- [31] Xu X, Wang Y, Panwar SP, Ross KW. A peer-to-peer video-on-demand system using multiple description coding and server diversity. In: Proc. of the Int'l Conf. on Image Processing. Washington: IEEE Computer Society, 2004. 1759–1762.
- [32] McCanne S, Jacobson V, Vetterli M. Receiver-Driven layered multicast. In: Proc. of the ACM SIGCOMM. 1996. <http://citeseer.ist.psu.edu/steven96receiverdriven.html>
- [33] Goyal VK. Multiple description coding: Compression meets the network. IEEE Signal Processing Magazine, 2001,18(5):74–93.
- [34] Cui Y, Nahrstedt K. Layered peer-to-peer streaming. In: Proc. of the 13th ACM NOSSDAV. New York: ACM Press, 2003. 162–171.
- [35] Rejaie R, Ortega A. PALS: Peer-to-peer adaptive layered streaming. In: Christos P, Kevin CA, eds. Proc. of the ACM NOSSDAV 2003. New York: ACM Press, 2003. 153–161.
- [36] Ahlswede R, Cai N, Li SYR. Network information flow. IEEE Trans. on Information Theory, 2000,46(4):1204–1216.
- [37] Liu Y, Dou W. The P2P streaming media based on network coding. Computer Engineering and Science, 2006,28(9):33–38 (in Chinese with English abstract).
- [38] Chi H, Zhang Q. Deadline-Aware network coding for video on demand service over P2P networks. Journal of Zhejiang University Science, 2006,7(5):755–763.
- [39] Guo J, Zhu Y, Li B. CodedStream: Live media streaming with overlay coded multicast. In: Proc. of the SPIE/ACM Conf. on Multimedia Computing and Networking. 2004. <http://citeseer.ist.psu.edu/guo03codedstream.html>
- [40] Sundaram N, Ramanathan P, Banerjee S. Multirate media streaming using network coding. 2006. http://www.ece.wisc.edu/~wander/papers/rate_sp_05.pdf
- [41] Zhao J, Yang F, Zhang Q, Zhang ZS, Zhang FY. LION: Layered overlay multicast with network coding. IEEE Trans. on Multimedia, 2006,8(5):1021–1032.
- [42] Do T, Hua K, Tantaoui M. P2VoD: Providing fault tolerant video-on-demand streaming in peer-to-peer environment. In: Proc. of the IEEE ICC 2004. Paris: IEEE Communications Society, 2004. 1467–1472.
- [43] Fouliras P, Xanthos S, Tsantalis N, Manitsaris A. LEMP: Lightweight efficient multicast protocol. In: Proc. of the 2004 ACM Symp. on Applied Computing. New York: ACM Press, 2004. 1226–1231.
- [44] Jiang X, Dong Y, Xu D, Bhargava B. GnuStream: A P2P media streaming system prototype. In: Proc. of the IEEE Int'l Conf. on Multimedia and Expo, Vol.2. Los Alamitos: IEEE Computer Society Press, 2003. 325–328.
- [45] Hefeeda M, Habib A, Botev B, Xu D, Bhargava B. PROMISE: Peer-to-Peer media streaming using CollectCast. In: Proc. of the 11th ACM Int'l Conf. on Multimedia. New York: ACM Press, 2003. 45–54.
- [46] Xu D, Hefeeda M, Hambruch S, Bhargava B. On peer-to-peer media streaming. In: Proc of ICDCS2002, Vol.1. 2002. 363–371.
- [47] Ma Y, Aygün RS. The methodology of mesh-casts streaming in P2P networks. In: Proc. of the 7th IEEE Int'l Symp. on Multimedia. Washington: IEEE Computer Society, 2005. 611–617.
- [48] Sanders P, Egner S, Tolhuizen L. Polynomial time algorithms for network information flow. In: Proc. of the 5th Annual ACM Symp. on Parallel Algorithms and Architectures. San Diego: ACM Press, 2003. 286–294.

附中文参考文献:

- [6] 余敏,李战怀,张龙波.P2P 数据管理.软件学报,2006,17(8):1717–1730. <http://www.jos.org.cn/1000-9825/17/1717.htm>
- [37] 刘亚杰,窦文华.基于网络编码的 P2P 流媒体.计算机科学与工程,2006,28(9):33–38.



郑常熠(1983—),男,福建宁德人,主要研究领域为 P2P 网络流媒体,网络编码。



赵进(1979—),男,博士,讲师,CCF 会员,主要研究领域为对等网,网络编码。



王新(1973—),男,博士,副教授,CCF 高级会员,主要研究领域为自组织网络,多媒体传输。



薛向阳(1968—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为多媒体信息处理与检索,网络流媒体。