

# 大规模流媒体应用中关键技术的研究

尹 浩<sup>1)</sup> 林 闯<sup>1)</sup> 文 浩<sup>1)</sup> 陈治佳<sup>1)</sup> 吴大鹏<sup>2)</sup>

<sup>1)</sup>(清华大学计算机科学与技术系 北京 100084)

<sup>2)</sup>(佛罗里达大学(美)电子与计算机工程系 佛罗里达 32611-6130 美国)

**摘 要** 支持大规模用户在线使用的流媒体应用是 Internet 中极富潜力的一项“重磅级应用”,但由于 Internet 缺乏服务质量(QoS)与相应的安全保障,并且网络和终端系统又存在着较大的异构性,这使得在 Internet 上构建支持大规模用户的在线流媒体应用面临很多的挑战. 该文从支持该应用的流媒体编码技术和网络技术两个角度出发,针对其面临的挑战,深入、全面地综述了编码技术与网络技术的发展与现状. 提出了一个新的流媒体应用体系结构,以同时解决大规模流媒体应用中的性能瓶颈、异构性、安全传输以及服务质量等问题,并指出了大规模流媒体应用中关键技术的研究方向.

**关键词** 流媒体;服务质量;视频编码;对等网络

**中图法分类号** TP393

## Research on Key Technologies of Large-Scale Streaming Media

YIN Hao<sup>1)</sup> LIN Chuang<sup>1)</sup> WEN Hao<sup>1)</sup> CHEN Zhi-Jia<sup>1)</sup> WU Da-Peng<sup>2)</sup>

<sup>1)</sup>(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

<sup>2)</sup>(Department of Electrical & Computer Engineering, University of Florida, Florida 32611-6130 USA)

**Abstract** Large-scale online streaming video application has gradually become the most important and potential technology of multimedia applications. Since Internet is heterogeneous and there is no Quality of Service (QoS) and security guarantee, large scale online streaming video application over the Internet faces many challenges. This paper, following the development of video sourcing coding and network transmission technology, comprehensively and deeply presents the scalable encoding and error-resilient encoding, and the development of multicast and P2P technology. Specially a new streaming architecture is proposed to satisfy the requirement of QoS, heterogeneity, and security guarantee in practical application. Some detailed analysis and comparisons are conducted in this paper. At last, future research direction is proposed.

**Keywords** streaming video; QoS; video sourcing coding; peer to peer

## 1 引 言

最近十年来,随着流媒体与网络技术的飞速发

展,人们对视频会议、视频点播、远程教学、在线游戏和娱乐等应用的需求越来越广泛. 大规模流媒体技术是使这些应用得以迅速发展的关键,从而也成为学术界和业界关心的研究热点.

收稿日期:2006-08-28;最终修改稿收到日期:2008-03-13. 本课题得到国家自然科学基金(60673184)、国家“八六三”高技术研究发展计划项目基金(2007AA01Z419)和国家“九七三”重点基础研究发展规划项目前期研究专项基金(2008CB317101)资助. 尹 浩,男,1974年生,副研究员,主要研究方向为流媒体技术、网络安全与性能评价等. E-mail: h-yin@mail. tsinghua. edu. cn. 林 闯,男,1948年生,教授,博士生导师,主要研究领域为计算机网络和系统性能模型及评价. 文 浩,男,1983年生,博士研究生,主要研究方向为网络性能评价、无线网络. 陈治佳,男,1981年生,博士研究生,主要研究方向为对等计算、流媒体等. 吴大鹏,男,博士,佛罗里达大学助理教授,主要研究方向为流媒体技术、无线通信等.

现有 Internet 上构建商业的大规模流媒体应用系统始终面临着几大挑战:网络服务质量(QoS)问题<sup>[1]</sup>;异构性问题;安全问题;可扩展性问题.由于流媒体应用对网络的带宽、丢包、延迟和抖动等网络服务质量 QoS 都有严格要求,而目前的 Internet 难以提供 QoS 保证,这使得流媒体的大规模商业应用难以满足用户对服务质量的需求.另外网络与客户端通常都具有较大的异构性,如何满足具有不同接入速度的用户享受不同服务质量的流媒体业务,也是成功的流媒体系统需要解决的问题.与此同时,伴随着人们对多媒体业务需求的日益增长,流媒体应用中的安全问题逐渐成为制约该应用进一步发展的关键,其安全问题主要表现在两个方面:一个是数字版权保护问题,另一个是传输安全问题.可扩展性问题主要表现在,很多的流媒体应用系统在用户人数不多的情况下质量不错,但是当人数达到一定的数量

时,其系统的服务质量就很难满足用户的需要,如何解决大规模流媒体应用中的可扩展性问题,一直是学术界关注的重点.

为了有效地解决上述的挑战,大量的新技术、新体系被提出,其中编码技术和网络传输技术则是构建大规模流媒体系统的两个关键技术.一方面,新的视频编码技术的出现(如可扩展编码 FGS<sup>[2]</sup>、PFGS<sup>[3]</sup>和多描述编码 MDC<sup>[4]</sup>等)推动了流媒体技术的发展,使该领域不断出现新的突破.同时,从 CDN 的内容分发技术、网络 IP 组播技术到 P2P 技术,这些网络传输技术的不断发展也给流媒体应用带来了极大的促进.

从网络传输技术的特点来分,可以将现有的大规模流媒体应用方案分成图 1 所示的两类:IP 组播技术支持的流媒体方案与 P2P 技术支持的流媒体方案.

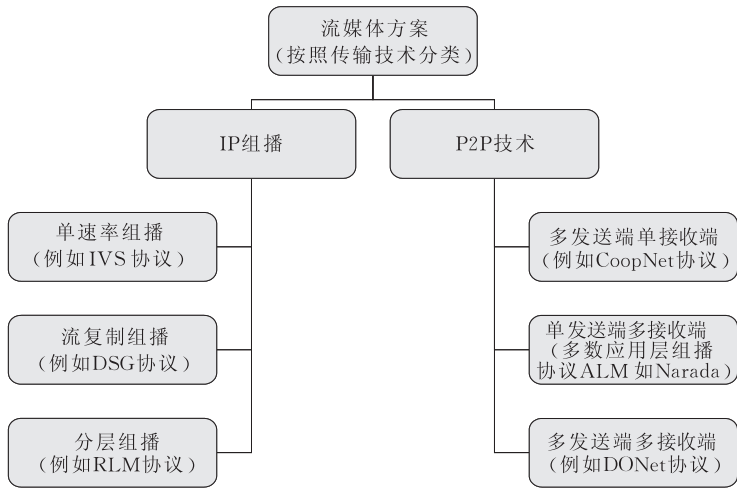


图 1 按照传输技术分类

按照网络传输技术的发展历程,大规模流媒体系统经历了如下 3 个阶段:

第 1 个阶段是利用 IP 组播协议<sup>[5]</sup>来承载流媒体传输,由于 IP 层组播在拥塞控制、可扩展性、可用性等方面存在一系列的问题,所以基于 IP 组播的视频通信应用一直未在 Internet 上得到广泛的使用.随之,出现了基于 CDN 网络(Content Distribution Network)的解决方案<sup>[6]</sup>,通过部署高性能的中心服务器和靠近用户的边缘代理服务器为用户提供高质量的流媒体服务.然而,CDN 昂贵的造价限制了其大范围的使用.

作为对等网络 P2P 传输模式的一种特殊实现方式,应用层组播(Application Layer Multicast, ALM)协议的提出,特别是基于 Narada 应用层组播

协议的端系统组播系统(End System Multicast)<sup>[7-8]</sup>在 2003 年 SIGCOMM 会议的使用,标志着流媒体播放方案进入了发展的第 2 阶段.在多数应用层组播协议构成的数据覆盖拓扑(也就是组播树)中,节点既可以从上级父亲节点接收数据,同时又能够向多个下一级节点发送数据,因此本文将多数 ALM 协议归于对等网络 P2P 方式中的单发送端多接收端一类,但其中一部分基于 ALM 协议的系统(如 CoopNet)在具体实现方式上属于多发送端单接收端的 P2P 传输方式.

在 2004 年,香港科技大学张欣研等开发的 CoolStreaming 系统<sup>[9]</sup>,又揭开了利用多点对多点的 P2P 流媒体技术进行实时视频流传输的第 3 个阶段.在多发送端多接收端的传输方式中,每个节点

既可以从多个节点接收数据也可以向多个节点发送数据,节点之间的数据拓扑构成了网状结构,极大提高了系统的扩展性.但在 P2P 技术给解决当前大规模流媒体应用中的网络与系统瓶颈问题带来新的机遇的同时,也给大规模流媒体系统带来新的挑战<sup>[10]</sup>.

全文第 2 节简要介绍近年来对流媒体系统产生过重要影响的编码技术;接下来,按照网络传输技术的分类和发展历程,在第 3 节介绍并比较基于 IP 组播的各种流媒体技术;第 4 节围绕 P2P 技术的分类,分别介绍并分析单发送端多接收端、多发送端单接收端和多发送端多接收端 3 类典型协议及其应用,同时提出基于 P2P 流媒体应用面临的挑战;最后提出一种新的体系结构,以期同时完整地解决上述的问题,并指出下一步的研究方向.

## 2 视频编码技术

支持大规模流媒体应用的视频编码技术大致分为两类:单码率与可伸缩性编码.

在单码率编码中,服务器始终以单一速率向所有接收端发送流媒体数据,并根据各个接收端的反馈信息调整数据发送速率.该编码方式的控制粒度较粗,不能同时公平有效地对待多个接收端.在接收端的网络状况差异很大时,接入带宽高于发送速率的接收端的接收能力没有得到充分利用,带宽低于该速率的接收端处会发生拥塞.

而在可伸缩性编码方式(Scalable Encoding)中,为了适应网络异构的特性,将视频内容编码分为若干个互不相交的视频层,接收端只需要接收一定数量的视频层,就可以解码还原得到视频画面,质量取决于接收到视频层的数量.由于各个层的内容之间并不重叠,接收到多个层的信息可以大大提高视频的接收质量,从而更有效地节省了带宽,因此分层组播在研究领域和实际应用中得到了广泛的重视.因此本节主要介绍并分析可伸缩编码方式.

可伸缩性编码最常见的两种编码方式就是分层编码和容错编码:前者属于累积分层,编码产生的视频层有主次之分;而后者属于非累积分层,编码产生的视频层优先级相同.

### 2.1 分层编码(Layered Encoding)

作为可伸缩编码的一种,分层编码将视频内容编码分为占用少量带宽的基本层和若干可以提高视频质量的增加层,接收端可以根据自己的网络状况

选择要接收的层,解码产生相应质量的视频图像.

在分层编码方式中,通常可在时间、空间或信噪比 3 个方面实现扩展<sup>[1-2]</sup>.基本层(base layer)作为最重要的层,包含流媒体的最重要特征的数据;其它层称为增强层(enhancement layer),包含进一步提高视频质量的数据.在流媒体传播方案中,通过仅向接收端发送它能够处理的那些层,可以很好地适应网络与终端用户的异构性.

传统的分层方案中增强层的层次粒度比较粗,精细可伸缩编码 FGS(Fine Granularity Scalability)<sup>[2]</sup>采用位平面(bitplane)的编码方式来产生增强视频层,由于增强层数据流可以任意截断来获得不同的比特率,所以能够获得不同输出速率的流媒体.虽然 FGS 具有细粒度的伸缩性和错误恢复能力,但是由于其分层数量太少,实际应用效果不好.累进的精细编码 PFGS(Progressive Fine Granularity Scalability)在继承 FGS 编码优点同时,利用了多个视频层进行预测,可伸缩性和容错性都增强了.此外,把细粒度的可伸缩性和空间可分级、时间可分级结合起来,又产生了精细空域可伸缩编码 FGSS(Fine-Granularity Spatially Scalable)和精细时域可伸缩编码 FGST(Fine-Granularity Scalable Temporal).FGSS 基本层采用传统空间编码方式,FGST 基本层采用时域编码方式,两者的增强层均采用位平面编码.这两种编码一方面保持了 FGS 的细粒度可分特性,又具有了在空域或者时域上的可分性.尽管如此,上述的技术大部分都依赖于 DCT 变换,这样虽然取得了编码的可伸缩性,但是对视频编码的效率和质量都产生了较大的影响,因而基于小波变换<sup>[11]</sup>的编码方式近年来成为了研究的热点.

小波变换从本质上是可以提供多分辨率或多尺度的信号分析,在提供可伸缩性的同时对视频压缩质量的影响很小,能够很好地保持原图像在各种分辨率下的精细结构,对图片和视频的压缩率较大,视频重构输出质量高.正是由于这些原因,小波编码在较高压缩比的图像编码领域被非常看好,而且可以同时有效地实现空间、时间与视频质量上的可伸缩性.基于三维小波变换(3D Wavelets)的可伸缩编码方法目前已成为 MPEG21 研究的重点.

### 2.2 容错编码(Error-Resilient Encoding)

容错编码是以多描述视频编码(Multiple Description Coding, MDC)<sup>[4]</sup>为典型代表的.多描述编码最初是作为一种容错编码(Error-Resilient Encoding)技术提出的,其目标是通过在编码中加入适当

冗余信息使得发生数据丢失后能够最大限度地减少对视频质量的影响。

MDC 编码方式中,编码器会生成原始信号的多个描述(相当于层),各个描述相互独立并具有相同的优先级.接收端可以对其中的任意多个描述进行解码;参与解码的描述越多,解码后的视频质量越高.但相比单描述编码,由于需要在原视频信号多描述之间加入冗余信息,会降低压缩的效率。

2.3 编码方式的比较

在流媒体传输中,分层编码如 FGS 是通过前向纠错机制 FEC 和重传进行差错复原的,但是在某些情况下(例如在误码率高的无线环境中),仍然难以保证数据的无损传输,此时基本层中的损失会使接收到的视频质量严重下降。

与此相对,多描述视频编码的主要优点是各层具有相同的优先级,没有需要特殊保护的层,在目前 Internet 不提供区分服务的现状下具有较好的稳定性。

而基于三维小波变换的可伸缩编码方法利用了小波变换本身的特点,使得视频压缩率高,只要选取的小波函数和相关滤波器合适,就能使视频能量集中在低频分量上.所以即使在编码过程中取较大的压缩比,还原图像的质量仍然较好,而且可以提供较大的可伸缩性,同时便于在终端进行不同层次之间的同步工作,因此具有很好的应用前景,其缺点是计算复杂度高,但尽管如此仍具有相对其它编码的巨

大优势。

从传统的单码率编码到适合流媒体传输的可伸缩编码方式,编码技术的发展在一定程度上解决了大规模流媒体系统中的异构性问题,能够满足不同接入带宽的用户不同的服务需求.但如何解决大规模流媒体系统服务质量 QoS 和可扩展性问题,始终离不开网络传输技术的不断发展创新。

3 基于 IP 组播的流媒体方案

3.1 方案综述

IP 组播通信方式由于在进行一对多通信时能有效地降低网络的冗余数据包的数量,所以在流媒体传输领域得到了深入的研究.下面将按照技术发展的脉络介绍这些方案.为了便于对比,将传统的单播数据传输列为最基本的解决方案。

图 2(a)是传统的点到点数据传输,发送端和每个接收端之间均需要单独的数据通道.骨干链路上需要传输大量重复分组,浪费了带宽资源,因此这种方案在支持大规模流媒体传输方面存在天然缺陷。

图 2(b)是单一速率自适应组播,发送端使用单一速率向所有接收端发送视频流,并根据各个接收端的反馈信息调整数据发送速率.与单播相比,该方案通过牺牲控制粒度来换取高的带宽利用效率.该方案的典型代表是 IVS(INRIA Video-Conference System)<sup>[12-13]</sup>。

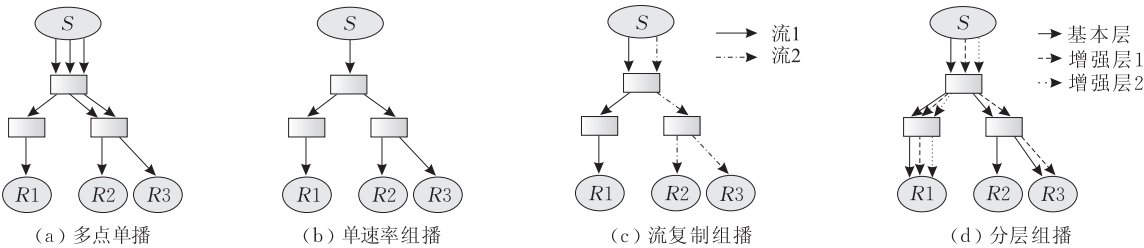


图 2 现有的视频组播方式

图 2(c)是流复制(stream replication)自适应组播,也称为 simulcast,指发送端同时维护同一视频内容的若干个(比较典型的是 3 个,图 2(c)中以两个为例)不同速率的视频流,每个视频流通过一个组播组发送出去,接收端可以根据自己的网络状况从中选择一个组播组加入,从而接收想要接收的流.可以认为该方案是前两种方案的一种折衷,既在一定程度上节省了带宽,同时接收端也有了一些选择的余地.该方案的典型代表是 DSG(Destination Set Grouping)协议<sup>[14]</sup>。

图 2(d)则表示分层组播(layered multicast)方案,利用分层视频编码技术,将视频内容分为若干个互不相交的层,接收端根据自己的网络状况选择要接收的层.例如在图 2(d)中,原始视频内容被编码为基本层、增强层 1 和增强层 2,接收端 R1 接收全部 3 个层,R2 仅接收基本层,R3 接收基本层和增强层 1。

根据具体的视频分层编码的不同,分层组播可分为累积分层组播(利用分层编码)和非累积分层(利用容错编码)组播。

在累积分层组播方案中,采用了如 FGS 等可扩



展编码,将原始视频压缩为优先级不同的若干个层(包括基本层和若干增强层),每一层通过一个单独的组播组被组播出去.通过仅向接收端发送它能够处理的那些层,可以很好地适应网络与终端用户的异构性.累积分层组播作为一种很有前途的视频组播解决方案,在最近几年里一直是研究领域中的热点问题.

McCanne、Jacobson 和 Vetterli<sup>[15]</sup> 综合利用分层压缩和分层传输技术,提出了累积分层视频组播方面第一个实用的自适应协议 RLM(Receiver-driven Layered Multicast).但 RLM 作为接收端驱动的协议,主要缺点是视频流的层数和发送速率是固定的,这样自适应的控制粒度比较粗.针对 RLM 的缺点,Vickers、Albuquerque 和 Suda<sup>[16]</sup> 提出了源端自适应的分层组播算法 SAMM(Source Adaptive Multi-Layered Multicast).SAMM 算法的网络体系结构由自适应分层视频发送端、分层视频接收端、支持组播的路由器和具有反馈聚集能力的节点 4 部分组成.同时 Liu、Li 和 Zhang<sup>[17]</sup> 在他们设计的端到端混合自适应分层组播协议 HALM(Hybrid Adaptation Layered Multicast)中提出了一个称为应用公平系数(Application-aware Fairness Index)的速率分配优化目标,用于衡量接收端的满意程度,并将分层速率分配的优化目标确定为使所有接收端的平均公平系数达到最大值,还推导出了该优化问题的多项式时间的解法.

非累积分层组播则利用了多描述视频编码(Multiple Description Coding, MDC)技术对原始视频编码,生成原始信号的多个描述(相当于层).接收端可以对其中的任意个描述进行解码;参与解码的描述越多,解码后的视频质量越高.

目前非累积分层组播的研究仍处于起步阶段,不过累积分层组播中的自适应机制(例如接收端驱动的自适应、发送端动态调整发送速率等算法)对于非累积分层组播都是适用的.

表 1 对一些代表性算法的主要特点进行了总结.

表 1 主要视频组播算法总结

算法	类别	传输速率	自适应方式	接收端是否发送反馈	控制粒度
IVS	单速率组播	单速率	发送端自适应	是	粗
DSG	流复制组播	多速率	混合自适应	是	较粗
RLM	累积分层组播	多速率	接收端自适应	否	较粗
SAMM/ HALM	累积分层组播	多速率	混合自适应	是	细

3.2 性能分析和对比

本节从以下几个方面对现有 IP 组播方案的性能进行分析和对比:接收端之间的公平性、带宽利用效率、反馈风暴的抑制方法、编码复杂度和效率、拥塞控制机制、错误控制机制和接收端的加入退出算法.

3.2.1 非接收端之间的公平性和带宽利用效率

一般而言,接收端之间的公平性和带宽利用效率是相互矛盾的.如图 3 所示,传统的多点单播需要传输大量重复分组,带宽利用效率最低,但它可以 根据每个接收端的情况调整到该接收端的数据速率,因此接收端之间的公平性最高.与之相对的另一个极端是单速率组播,发送端仅维护一个视频流,带宽利用效率最高.但单一的视频速率使它无法满足多个异构接收端的不同需求,因而导致公平性较差.流复制组播、分层组播等多速率组播介于上述两个极端之间,是一类折衷的方案.

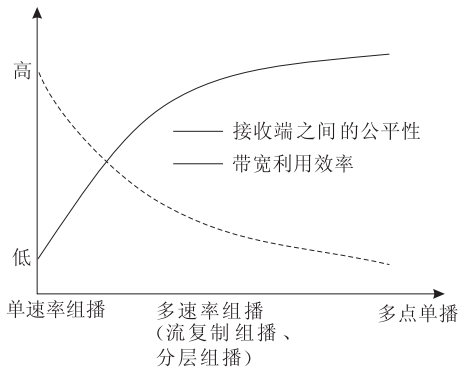


图 3 公平性和带宽利用效率的对比

3.2.2 反馈风暴(feedback implosion)问题

与单播相同,组播也需要进行流量控制、拥塞控制和错误控制.错误控制确保接收端正确地接收到发送端发送的分组;流量控制确保发送端的数据发送速率不超过接收端的处理能力;拥塞控制确保发送端的数据发送速率不超过网络的负载能力.目前这些控制均需要通过接收端向发送端发送反馈信息来完成.在组播中,这很容易引起反馈风暴,即所有接收端同时向发送端发送状态信息,从而使发送端在处理反馈信息方面消耗大量处理能力和带宽.为避免发生反馈风暴,目前的解决方法主要有以下 3 类:

第 1 类方法是在组播树中设置多层反馈聚集节点,将反馈控制的任务分担到整个组播树中(SAMM 中使用的就是这种方法).

第 2 类方法是减少发送反馈信息的接收端的数

目. 较早采用这种方法的是 Bolot、Turletti 和 Wakeman. 在他们设计的算法中, 每个接收端均以某种概率发送反馈信息, 该概率是接收端总个数的函数.

第 3 类方法是尽量减少需要发送的反馈信息,

即对拥塞、出错等问题采取预防性措施而不是纠正性措施. 例如, 在错误控制方面, 使用前向错误纠正 (FEC) 编码<sup>[18-19]</sup>代替简单的错误检测编码. 在流量控制和拥塞控制方面, 使用资源预留<sup>[20]</sup>方法使接收端和网络中间节点能够支持发送端的数据速率.

表 2 抑制反馈风暴的方法

代表性协议	协议类型	抑制反馈风暴的方法
SAMM	混合自适应累积分层组播	设置反馈聚集节点
IVS	单速率组播	减少发送反馈的接收端的数目
SARLM	混合自适应累积分层组播	在接收端设置定时器, 以减少发送反馈的接收端的数目
HALM	混合自适应累积分层组播	延长发送反馈的周期, 以减少需要发送的反馈信息

3. 2. 3 编码复杂度和效率

单速率组播的编码复杂度最低, 编码效率最高. 流复制组播如果使用转换编码, 将带来约 5% 的额外带宽开销, 计算复杂度较低; 由于 DCT 系数和运动向量都可以重用, 因此转换编码所需的时间也比较少. 相比之下, 累积分层编码需要对所有层进行迭代运动估计和 DCT 变换, 所以通常具有较高的计算复杂度. 在进行传输时, 层之间的同步信息等数据

也需要占用一些额外带宽. MDC 编码的每一层都需要提供足够多的原始视频信息, 因此压缩效率比累积分层编码更低. 表 3 列出了分别用 MPEG-4 单速率编码、FGS 编码和 MDC 编码技术对两个标准测试序列 Foreman 和 Akiyo 进行编码的比特率. 从表中可以看到, 累积分层编码和非累积分层编码的额外带宽开销均高达 20% 以上. 因此在某些网络拓扑中, 流复制组播的整体性能要优于分层组播.

表 3 各种组播方案的编码效率对比

	Foreman, QCIF (PSNR=33.3 dB)		Akiyo, QCIF (PSNR=35.4 dB)	
	比特率/(kbps)	额外带宽开销/%	比特率/(kbps)	额外带宽开销/%
单速率	79.3	—	19.3	—
FGS	102.3	29	23.8	23
MDC	114.6	44	25.7	33

累积分层视频流和 MDC 视频流要求接收端具有较高的计算能力, 以便完成多层的联合解码. 而流复制组播中的各个流可以采用比较简单的标准解码算法, 甚至各个流可以具有不同的格式, 因而更为灵活实用.

3. 2. 4 拥塞控制机制

拥塞控制主要由速率控制、速率自适应编码技术和速率整形技术实现<sup>[21]</sup>. 速率控制可以由发送端或接收端单独实现, 也可由二者联合进行控制. 图 4

是一种基于发送端的拥塞控制体系结构的示意图. 其中, 发送端的压缩层使用一种速率自适应的编码技术压缩视频, 压缩后的视频比特流经速率整形后通过 RTP/UDP/IP 层协议传输至接收端. 接收端的 QoS 监视器根据接收到的分组推测出网络的拥塞状况, 并通过反馈控制协议将该信息反馈给发送端. 在基于接收端的拥塞控制方案和混合拥塞控制方案中, 接收端也可以利用该信息选择要接收的视频流或视频层, 从而达到控制速率的目的.

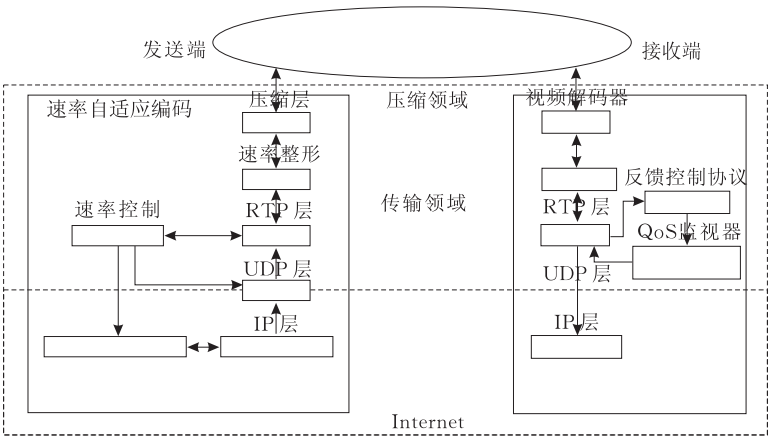


图 4 基于发送端的拥塞控制体系结构

组播拥塞控制的研究可分为单速率(Single-Rate Multicast Congestion Control, SR-MCC)和多速率(MR-MCC)两个方向. 最早的累积分层组播协议 RLM 是在接收端通过定期试探性地加入新层进行拥塞控制的. 这种方法有两个主要缺点. 一是某个接收端的加入试验可能会导致其它接收端经历分组丢失,从而具有潜在的不公平性;二是在现有的 IP 组播中,退出一个组播组所需的时间代价很大(秒数量级),因而这类算法的效率较低,收敛速度较慢. 另外,RLM 不具有 TCP 友好性,同时对组播会话内的接收端也不能做到公平的对待.

针对它的局限性,Vicisano 等人提出了 RLM 的 TCP 友好版本 RLC (Receiver-Driven Layered Congestion Control). RLC 按照指数规律设定各层的速率,即每层的速率是其下一层速率的 2 倍. 虽然这种方案对 TCP 流量比较公平,但它会使视频流的流量出现像 TCP 流那样的锯齿状变化,导致接收到的视频质量不稳定. FLID-DL(Fair Layer Increase Decrease with Dynamic Layering)是对 RLC 的进一步改进,它将 RLC 的 TCP 友好的控制机制推广到各视频层速率为任意值的情况,并通过动态调整各

层速率来减少退出组播组的次数,从而加快了收敛速度.

PLM(Packet-pair Layered Multicast)是第一个不使用试探性加入方法的拥塞控制机制,它利用包对(Packet Pair)技术直接探测从发送端到接收端的可用带宽,接收端根据探测到的可用带宽决定加入或退出相应分层视频组的操作. 该算法的局限性是它要求网络中的所有路由器均支持公平排队算法(Fair Queueing). 最近几年来,单速率组播和多速率组播都开始采用基于 TCP 友好吞吐量公式<sup>[22]</sup>的拥塞控制机制. 例如,HALM<sup>[17]</sup>、MLDA<sup>[23]</sup>和 SMCC<sup>[24]</sup>等利用该公式来估计相应路径上的 TCP 友好的可用网络带宽,然后参考该带宽值进行接收端(加入/退出组)或发送端(调整层数和各层速率)的自适应拥塞控制,以达到长程的 TCP 友好性. 在这类控制机制中,TCP 流量模型的准确性和 RTT 等参数的准确测量还有待进一步研究. 表 4 对这些算法的特点进行了总结.

表 5 是对现有的代表性组播协议及其各方面特点的综合总结与对比.

表 4 代表性拥塞控制算法的特点

拥塞控制算法	类别	检测拥塞的方法	探测空闲带宽的方法	TCP 友好性	缺点
RLM	基于接收端	分组丢失率	试探性加入	较差	收敛速度慢
RLC	基于接收端	分组丢失率	试探性加入	一般	视频质量不稳定,收敛速度慢
FLID-DL	混合	分组丢失率	试探性加入	一般	视频质量不稳定
PLM	混合	直接利用包对技术估计可用带宽	较好	需要路由器的支持	
SMCC/HALM	混合	直接用 TCP 吞吐量公式估计可用带宽	较好	难以保证参数的准确测量	

表 5 现有基于 IP 组播流媒体的特点

协议	类别	组内公平性	带宽利用效率	TCP 友好性	抑制反馈风暴	编码复杂度	速率控制方式	接收端加入/退出算法
IVS	单速率组播	差	高	一般	减少发送反馈的接收端的数目	低	基于探测	—
DSG	流复制组播	好	一般	一般	设置反馈聚集节点	低	基于探测	—
RLM	接收端驱动累积分层组播	差	一般	差	无需反馈	高	基于探测	共享学习
RLC	接收端驱动累积分层组播	差	一般	好	无需反馈	高	基于探测	同步
SAMM	混和累积分层组播	差	一般	一般	设置反馈聚集节点	高	基于探测	—
HALM	混和累积分层组播	好	一般	好	减少反馈信息	高	基于模型	同步

现有的分层组播(layered multicast)流媒体方案,由于综合利用了分层视频编码技术和 IP 组播技术,在一定程度上解决了异构问题、网络服务质量问题和可扩展问题. 但针对流媒体中的安全问题,一个成熟的大规模流媒体方案需要考虑数字版权保护和传输安全两个子问题,因此还需要依赖如下 4 个关键技术的支持:(1)高可扩展的密钥管理机制;(2)视频加密算法;(3)用户的身份认证机制;(4)流

媒体的版权验证机制. 在文献[25]中,我们提出了一种基于媒体内容的安全组播协议,其包含了可扩展并且轻量级的密钥管理机制,基于媒体内容的密钥嵌入算法、两层的选择加密算法,兼容各种底层编码技术,在保证流媒体系统的扩展性和鲁棒性的同时,在安全性和流媒体运行效率之间做出一个很好的权衡.

## 4 基于 P2P 技术的流媒体方案

尽管从 IETF(Internet Engineering Task Force)提出关于网络层组播的 RFC(Request for Comments)至今已有 20 余年了,但是网络层组播由于可扩展性差、缺乏拥塞控制、难以管理、部署难度大等各种技术或者非技术的问题,一直得不到大规模的应用。

在此背景下,P2P 技术的快速发展又为大规模流媒体应用提供了新的方案.P2P(Peer-to-Peer)意为对等互联或点对点技术,通过建立对等互联的体系结构达到去中心化(decentralize)的效果,克服了传统的客户机/服务器(Client/Server)结构容易出现的服务器瓶颈问题,可扩展性和鲁棒性更高,网络异构的问题也可以得到很好的解决。

P2P 流媒体技术综合利用 P2P 技术和流媒体技术作为大规模流媒体应用的解决方案,通过 P2P 的网络结构,可以实现不同网络情况下的可适应的流媒体传输.其中每个节点既是接收数据的客户端,同时又是发送数据的服务器,因此没有严格意义的数据源,每个客户端都可能充当数据源,这样的方案给每个节点带来了一定的运算负担,但在整体的传输上更为灵活和高效.虽然 P2P 流媒体方案具有不少优势,但是一方面除了异构问题和扩展性问题,基于 P2P 技术的流媒体方案同样需要考虑流媒体应用中如延时、抖动程度等 QoS 性能要求.另一方面,由于 P2P 技术是通过网络终端对数据包的复制来实现组播的,所以存在延迟比较大、传输效率不如 IP 组播等劣势,在这种情况下设计相关的密钥管理机制面临着很大的挑战,安全问题带来的挑战始终存在。

对 P2P 技术的划分方法很多,在本文中我们按照数据的接收和传输方式,将 P2P 流媒体应用分成单发送端多接收端、多发送端单接收端和多发送端多接收端 3 类。

单发送端多接收端方式,指一个节点的数据是来自单一的节点,而将其发送到其它多个节点,多数应用层组播协议(比如 Narada)基本属于这类.此类应用适合于当前大多数应用环境,有利于增进组播网络的鲁棒性.但由于存在组播树深度和节点负载之间的平衡问题,在控制节点的连接和分离上往往需要比较复杂的算法支持。

多发送端单接收端方式,指一个发出请求的节

点可能需要多个节点给它发送数据,这类发出请求的节点往往是异构网络中性能比较低或者对数据可靠性要求很高的客户端节点,典型例子是微软研究院开发的 CoopNet 系统。

多发送端多接收端方式,指任何一个节点既可以接收多个节点的数据,也可以向多个节点发送数据,由于综合了前两种的特点,往往被称为纯粹的 P2P(Pure P2P).以 DONet 协议和 Bit-Torrent 协议为典型代表的这类技术,在实际中得到了最广泛的应用,正在因特网中引起一阵 P2P 应用的风暴。

在这 3 种基于 P2P 技术的流媒体方案中,非常关键的评价标准在于方案的扩展性、维护状态开销的大小、传输的有效性、合理的延迟以及是否具有自适应性和高健壮性的运行机制.比如关于运行机制的研究就涉及到有效的拓扑生成、节点的加入删除处理、节点的出错处理和快速恢复等等.这几点评价标准的差异,往往决定了设计的流媒体方案适用于何种网络环境,能够提供为用户提供怎样的服务。

### 4.1 单发送多接收端方式(应用层组播协议)

与 IP 组播通过网络服务器实现不同,应用层组播是由参与的终端节点构成一个逻辑覆盖网络(overlay network),然后在覆盖网络上建立组播树结构,实现一对多的组播功能.在应用层的组播实现过程中,对目标主机的查询、对数据的打包、数据包的接收和转发工作都在终端节点的应用层上完成,屏蔽了物理底层的细节。

应用层组播继承了组播模式的通信效率,克服了 IP 层组播难以在 Internet 中应用的缺点.基于应用层组播的大规模流媒体传输体系近年来成为了研究热点和多媒体应用的发展方向.当前,基于应用层组播的流媒体系统研究工作已有不少的成果,如 Narada<sup>[8]</sup>, NICE<sup>[26]</sup>, HMTP<sup>[27]</sup>, OMNI<sup>[28]</sup>, Spread It<sup>[29]</sup>, CoopNet<sup>[30]</sup> 和 SplitStream<sup>[31]</sup> 等.其中一些系统已在 Internet 中应用,如 2003 年 SIGCOMM 会议曾利用“基于 Narada 协议的端系统组播(End System Multicast)”系统进行现场直播,位于全球不同位置的用户运行该系统通过 Internet 可观看此次会议实况。

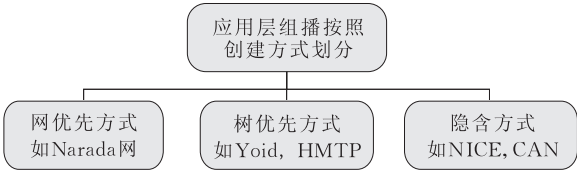
由于应用层组播通过组成员组织和维护有效的覆盖网络(overlay network)实现数据的传输,如何构造一个稳定、高效的覆盖网络就是应用层组播设计的目的.衡量一个应用层组播协议是否优秀需要考虑的性能指标主要有:压力度(stress)、伸张度(stretch)、控制开销(control overhead)等<sup>[8]</sup>.压力



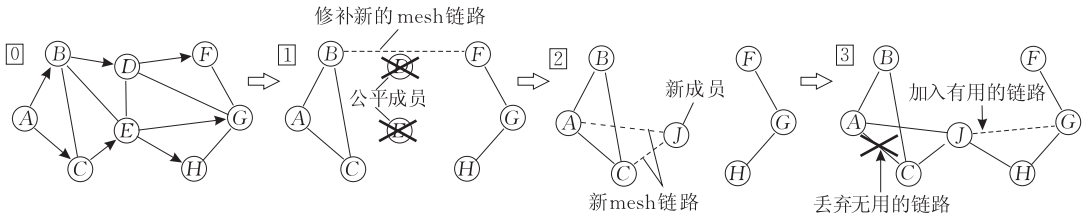
度是计算拓扑网络中经过一条链路或者路由由节点相同数据包的次数。伸张度计算的是从数据源节点到一个成员节点在覆盖网络的路径长度或者延时和在直接单播的路径长度或者延时的比例。压力度和伸张度都用来衡量网络的数据传输性能。而控制开销则用来度量覆盖网络中不同成员节点之间交互信息数量(包括加入、退出组播消息和状态信息等)。

应用层组播协议将组成员组织成两种网络拓扑结构:控制拓扑和数据拓扑。控制拓扑又被称为网(mesh), mesh 成员会不停地交换状态信息来提高拓扑网络的稳定性,而数据拓扑又称为树(tree),构成了组播协议中的数据传输路径,它通常都是 mesh 的子集。

如图 5,依据创建覆盖网络的方式可以将应用层组播分为 3 类:网优先(mesh-first)、树优先(tree-first)、隐含式(implicit)。



在网优先组播方式中,组成员先按照一定策略和方式产生覆盖控制拓扑,然后针对每个组成员计算到其它节点成员的路径,按照一定的衡量标准产生相应的覆盖数据拓扑。



Narada 协议采用的自组织、自改进的方式能够适应组成员的动态改变和端节点的异常行为,表现了强大的健壮性。但由于组成员之间需要不停地交互状态信息,而且每个成员节点需要保存其它所有节点的状态信息,这会不可避免地造成大量的控制开销和占用存储空间,使其扩展性受到制约,因此基于 Narada 协议的端系统组播(End System Multicast)方案只适用于视频会议等小规模流媒体应用,难以直接应用于大规模流媒体方案。

(2) 树优先组播协议:HMTP<sup>[27]</sup> 和 Yoid<sup>①</sup>。

树优先组播协议的核心工作就是建立并维护数

在树优先组播方式中,组成员先构成一个覆盖数据拓扑,然后每个组成员通过获得它所有相邻成员节点和一些非相邻成员的状态信息形成新的覆盖控制拓扑。

在隐含式组播方式中,在按照一定的标准产生控制拓扑同时,根据包转发的规则也就隐式地产生了数据拓扑。换句话说,两者是同时产生的,不存在产生顺序上的区别。

按照所述的分类方式,我们将主要的应用层组播协议分为三类。

(1) 网优先组播协议:Narada<sup>[8]</sup>。

Narada 协议作为最早的应用层组播协议之一,采用了自组织(self-organizing)和分布式(fully distributed)的设计思想来构建覆盖网络。在 Narada 建立的覆盖网络中,为了获得更高的鲁棒性,组成员之间通过周期性地控制信息更新交互,将使每个成员保存一份组内其它节点的信息列表,这样整个覆盖网络的信息就分散到每一个节点,而不依靠单独的核心节点。在控制拓扑 mesh 基础上,可以利用不同数据源为根构造组播树,并对每个源进行优化。此外,Narada 协议中提出了启发式渐进改良 mesh 性能的算法。组成员通过随机互相地探测,通过对链路利用率(utility)的计算来决定是否加入新的链路或者丢弃无用的链路。mesh 的动态改进机制使得 Narada 协议区别于很多早期协议。图 6 显示说明了 mesh 的修补和改进过程<sup>[32]</sup>。

据组播树,它的设计思路是先按照一定的规则(比如节点的邻接节点数目)构建数据组播树,然后在覆盖数据拓扑基础上构建覆盖控制拓扑。由于组播树的节点只需要保存邻居节点和部分非邻居节点的状态信息,维护开销不大,和网优先组播协议相比,扩展性得到了提高。

图 7<sup>[32]</sup> 说明了 HMTP 协议的新成员加入过程:一个新加入成员 N 通过引导程序确定组播树的

① Yoid F P. Extending the multicast Internet architecture, 1999. White paper. <http://www.aciri.org/yoid/>

根节点 A, 由于 A 的子节点已经达到自身上限值 3, 然后 N 在 A 的 3 个子节点 (B、C、D) 中寻找最优的节点 D. 但是由于 D 的子节点数目也达到上限值 2,

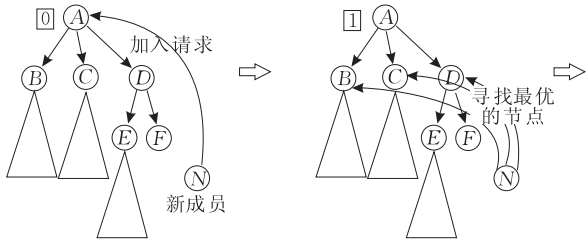


图 7 新成员加入过程

在组播树形成的过程中, 需要保证不产生回路并且不违背子节点数目限制, 以保证组播树结构的稳定. Yoid 协议中采用的是回路检测和避免 (loop detection and avoidance) 机制, 而 HMTTP 协议采用的是环路检测和消除 (loop detection and resolution) 策略, 这是两者的不同之处. 此外, Yoid 协议中的覆盖控制拓扑的建立是通过组播共享树中的每个节点随机选择一些非邻居节点交互状态信息实现的, 而在 HMTTP 协议中, 成员节点周期性地获得组播树其它部分成员节点的状态信息来提高组播树的鲁棒性, 不会特别产生覆盖控制拓扑.

在树优先组播协议中, 一方面在节点的加入和退出过程中都需要考虑环路检测的问题, 需要增加大量的控制信息, 相应增加了协议的复杂性. 另一方面, 由于网络中始终无法避免由于节点成员的意外离开或者失效造成拓扑结构的突变和传输的暂停, 而树优先组播协议往往不能提供足够的状态信息和健壮的恢复机制来快速修复组播树, 因此可靠性成为了制约其大规模应用的瓶颈.

(3) 隐含式组播协议: NICE<sup>[26]</sup>, Can-multicast<sup>[33]</sup>, Scribe<sup>[34]</sup> 和 Bayeux<sup>[35]</sup>.

针对网优先和树优先组播协议的优缺点, 基于可靠性、扩展性、控制开销之间性能和效率的平衡问题, 学术界又提出了一系列隐含式组播协议.

4.1.1 NICE 协议

NICE 协议核心思想是在组播树的基础上“分层”(Hierarchical)和“分簇”(Cluster). 如图 8, 不同于树优先或者网优先协议, 组成员被组织为层次控制拓扑. 从整体上看层次结构是树状的, 但从局部(每层内节点构成的簇)又是以网状结构组织的. 由于成员只和少量固定数目的节点联系, 控制开销不大, 同时考虑网络异构的特点, 由选择的领导节点负责簇内的管理和数据分发, 从而很好地融合了网优

于是 N 又递归搜索到 D 的下一层子节点 F 最优并且 F 子节点数目没超过上限, 最终 N 便加入成为 F 的子节点.

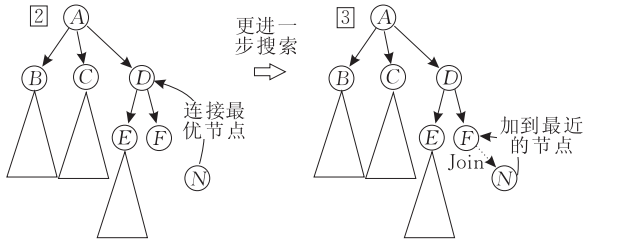
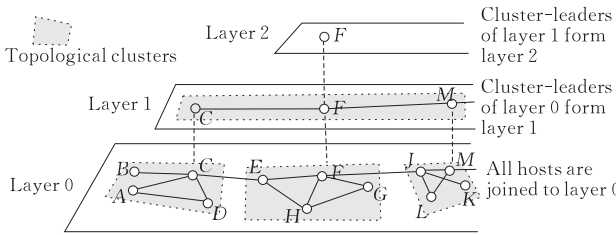


图 8 NICE 层级结构

先和树优先两种思想的优点, 一定程度上兼顾了网络异构、可扩展性和鲁棒性.



但我们看到在 NICE 协议中, 由于本质上仍采用组播树的拓扑结构, 处在高层 (比如第  $i$  层) 的成员, 也是所有相对低层 (从  $0 \sim i-1$ ) 的领导节点, 这样高层的领导节点往往容易形成系统瓶颈.

4.1.2 CAN-Multicast 协议

P2P 技术最新的研究集中在基于分布式散列表 DHT (Distribute Hash Table) 的分布式发现和路由算法方面. 此类算法通过分布式散列 HASH 函数, 将输入的关键字映射到某个逻辑节点上, 然后通过路由算法查找该节点, 避免了借助中央服务器或者利用广播进行洪泛查找, 可以提高路由的效率, 减少路由表容量和路由延时, 克服了非结构化拓扑中的路由查找存在极大的不可扩展性的问题.

基于 DHT 的结构化拓扑结构能够自适应节点的动态加入/退出, 有着良好的可扩展性、鲁棒性、结点 ID 分配的均匀性和自组织能力精确的发现机制, 正好为应用层组播的实现提供了良好的底层平台. 采用 DHT 结构的应用层组播协议, 一方面能够实现节点的动态加入和退出, 保证节点之间的均匀性和自组织能力. 另一方面也能够实现对目标节点的快速路由发现和路由查找, 减少状态维护开销和转发开销, 比较典型的例子是 CAN, Pastry 和 Tapes-try, 它们之间的差别在于具体的路由策略和发现方

式不同。

在内容可寻址网络 CAN(Content-Addressable Network)协议中,所有的成员节点被组织成一个虚拟的  $d$  维笛卡尔坐标空间(可以看作一个新的逻辑网络),同时在这个新的逻辑网络上使用散列函数对成员节点进行重新编址(比如对关键值  $key$  进行散列计算)。由于这个逻辑网络是通过笛卡尔坐标空间的方式构造的,根据坐标值就可以根据最近相邻原则进行路由查找。所以,通过选择合适的散列函数,我们可以对坐标空间的进行合理的分配和管理,实现插入、查询和删除等功能。在文献[33],Ratna-samg 等提出了基于 CAN 架构的应用层组播模式。

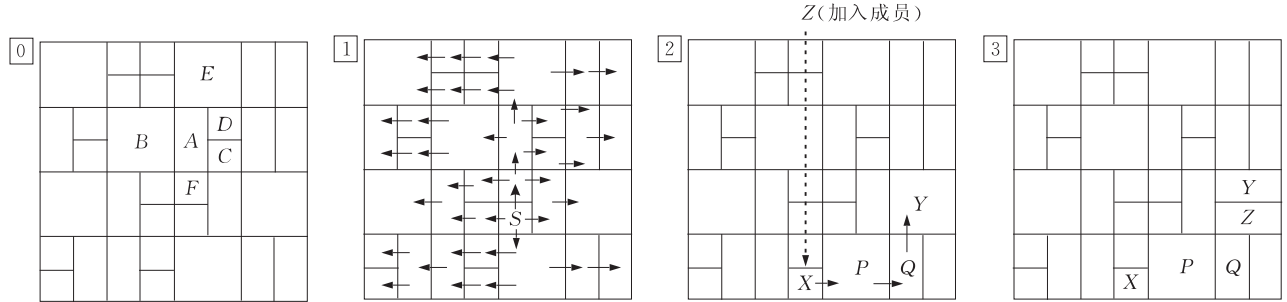


图 9 CAN 坐标空间示意图

在  $d$  维笛卡尔坐标空间中,任何一个节点的邻居节点只有  $2d$  个邻居节点,因此它只需要保留最多  $2d$  个邻近路由信息,维护的路由表信息和网络规模无关,维护开销很小。而数据拓扑结构隐含在控制信息的洪泛过程中。此外,针对笛卡儿坐标空间的路径长度能控制在  $O(d \times N^{1/d})$  的特点,通过增加维数  $d$ ,可以进一步降低路径长度,降低网络延迟,同时随着维数的增加,邻居节点也相应增加,鲁棒性也得到了提高。

由于加入过程中坐标区域的选择是比较随机的,忽略了对成员节点间相对距离的影响,因此成员节点的分布也没有规律,造成覆盖拓扑网络伸张度(stretch)过大。为此作者又提出了“分布式储存”(distributed binning)的改进思路,使得路径距离较近的节点分配临近的区域空间,以此降低覆盖网络的伸张度。

4.1.3 Scribe 协议

Pastry<sup>[36]</sup>作为可扩展的分布式对象定位和路由协议,是另一种典型的 DHT 路由协议,由位于英国剑桥的微软研究院和莱斯(Rice)大学提出,可用于构建大规模的 P2P 系统,而 Scribe 系统正是建立在底层 Pastry 网络上基于主题(topic-based)的大规模

一个新节点要加入 CAN,关键是采用散列函数对  $(key, value)$  中的  $key$  进行散列运算,找到其坐标空间中对应的区域,并将  $(key, value)$  存储在拥有该点所在区域的节点内。如果所对应的区域已经被占用,则已存在的成员节点分割其所在子块的区域空间,把其中的一半区域分配给新加入成员节点。图 9 所示是一个由 34 个成员构成的二维坐标空间,被相应分成了 34 个区域(zone)。当节点  $Z$  欲加入 CAN 时,首先通过引导程序找到一个已存在的节点  $X$ ,通过  $X$  路由随机找到属于节点  $Y$  的坐标区域空间,然后分割  $Y$  节点的一半区域给  $Z$ ,最后通知邻接区域  $Z$  的加入。

发布-预订(publish-subscribe)事件通告系统。

加入 Pastry 网络中的每个节点都会被赋予一个唯一的节点标识(node identifier),节点标识一般通过计算成员公钥或者 IP 地址的 Hash 值得到,标识值每位数字取值范围  $0 \sim 2^b - 1$  ( $b$  为很小的常数)。在 Pastry 形成的覆盖网络中,只要能够知道节点标识,就能够通过路由机制找到路径。

对于 Pastry 中的每个成员节点都有一个路由表,一个邻居节点集合和一个叶节点集合。如图 10 所显示的是标识为 2313 的路由表, $b$  取 2,每位能取值为 0,1,2,3。路由表中每个矩阵框内的值对应其它节点成员的标识值,2313 作为自身的标识值是隐藏存在的。路由表第  $i$  排的节点标识值和本节点的标识(2313)具有  $i-1$  个相同的前缀。图中第 3 排的节点标识 2301、2330 就和 2313 的前两位相同。因此我们可以知道所有节点的路由表一共具有  $\log_2^b N$ , Pastry 路由的复杂度为  $O(\log_2^b N)$ ,其中  $N$  表示 Pastry 中成员节点的数目。在具体的路由查询中,如果指定一个目标节点的标识,通过标识前缀最大匹配,节点将会把消息路由到在标识值和目标标识最接近的那个节点。

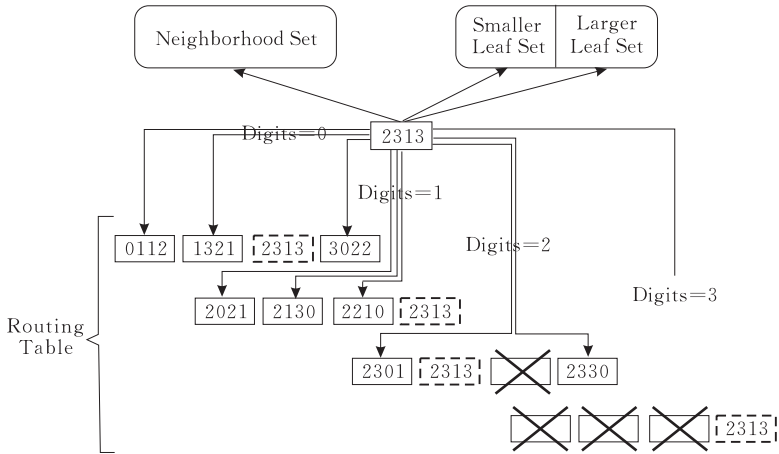


图 10 Pastry 节点路由表

Scribe 应用层组播协议采用 Pastry 网络提供底层路由支持. 因此 Scribe 的控制拓扑和 Pastry 网络的控制拓扑是相同的, 包括了成员节点的路由表信息、邻居节点集合和叶节点集合.

对于数据拓扑结构, 则是 Scribe 协议通过匹配最大前缀的原则, 建立源节点到其它节点的数据传输路径, 即使不能直接查询到目标节点的路径, 总能保证找到一条路径更接近目标节点.

需要指出的是, Scribe 协议中可以支持多个组播组, 因此需要给每个组播组分配一个唯一的标识地址, 也被称为主题标识 (topic identifier). 自身标识和主题标识最近的成员节点成为当前组播组的引导节点, 负责引导其它节点加入组播覆盖网络.

相比 CAN 协议, Pastry<sup>[36]</sup> 引入了叶节点集合和邻居节点集合的概念, 能够快速准确地获取路由信息, 大大加快了路由查找的速度.

4. 1. 4 其它应用层组播协议

Bayeux<sup>[35]</sup> 应用层组播协议采用的是隐含的方式构造覆盖网络, 底层依靠对等式目标定位系统 Tapestry<sup>[37]</sup>. Tapestry 覆盖体系和 Scribe 的底层结构 Pastry 相似, Bayeux 和 Scribe 不同在于组播数据拓扑的产生方式.

Scattercast<sup>[38]</sup> 和 Overcast<sup>[39]</sup> 协议采用应用层网关方式, 通过部署一些代理节点 (proxy) 组成应用层网络的分布式组播树, 具有比较高的稳定性, 但灵活性比较差.

ALMI<sup>[40]</sup> 没有采用分布式设计思想, 而是采用集中式的树优先构造方法. 集中式协议假设服务器知道成员之间的拓扑结构, 然后由服务器按照性能要求构建组播树, 然后转发拓扑信息给相应节点. 由于存在服务器节点性能瓶颈, 鲁棒性很低, 扩展性

不好.

4. 1. 5 应用层组播协议的比较

除了按照创建覆盖网络的方式划分应用层组播协议, 根据构建转发树的策略, 还可以将它们分成集中式和分布式两类: 集中式协议以 ALMI、HBM 为代表; 而分布式协议则有 NICE, Overcast, Yoid, HMTTP, Narada 等等. 由于分布式协议健壮性更高, 可扩展性更好, 因此更多地被采用.

按照覆盖网络成员节点的性质, 可以分为架构式、对等式和混合式协议 3 类: 架构式以 Overcast 为代表, 存在代理节点; 对等式协议以 Narada 为代表, 成员节点的地位相同; 混合式综合上述两者特点, 以 CoopNet 为代表.

一般说来, 除了集中式组播协议因为扩展性上的缺陷外, 其它分布式应用层组播协议都有各自不同的优缺点和适用环境.

流媒体应用有较高的实时需求, 对延迟敏感. 这与协议的最大路径长度有关, 长度越小相应的延迟也会越小. 多媒体应用对网络带宽也有较大的需求, 这与协议的最大子树度数有关, 度越小, 能获得的网络带宽也越大. 另外平均控制开销也希望越小越好.

表 6 给出几类典型的应用层组播协议的性能对比 ( $N$  是组播组成员数,  $d$  是 CAN 组播协议中组成员形成的笛卡尔坐标空间的维数), 主要比较最大路径长度、最大子树度、平均控制开销等. 从中可以看出比较适合多媒体应用的应用层组播协议是隐含式方法, 而且在隐含式应用层协议中, 可以看到 CAN 这类分布式结构化协议通过采用多维的标识符空间来实现分布式散列 (DHT) 算法, 具有更加良好的可扩展性, 而且节点维护的路由表信息和网络规模无关, 路径长度也能控制在  $O(d \times N^{1/d})$  规模上.



表 6 各类应用层组播数据比较

	基本类型	组播树类型	拓扑结构类型	最大路径长度	最大子树度数	平均控制开销
Narada	网优先	特定源端	中心化拓扑	无上界	无上界	$O(N)$
Yoid/HMTP	树优先	共享树	中心化拓扑	无上界	$O(\text{最大节点度数})$	$O(\text{最大节点度数})$
Scribe/Bayeux	隐含式	特定源端	分布式结构化	$O(\log N)$	$O(\log N)$	$O(\log N)$
CAN	隐含式	特定源端	分布式结构化	$O(d \times N^{1/d})$	常数	常数
NICE	隐含式	特定源端	中心化拓扑	$O(\log N)$	$O(\log N)$	常数

4.2 多发送端单接收端方式

在多发送端单接收端传输方式中,考虑到异构网络中的多数 peer 节点性能不稳定,一个发出请求的节点通过接收多个节点发送的数据,以此提高传输的效率和质量.按照具体实现方式不同,又可以分为混合方案(以 CoopNet 为例)和标准方案(以 PROMISE 为例).

4.2.1 混合方案

CoopNet 系统基于中心服务器实现,结合了以 Overcast 为代表的架构式系统和以 Narada 为代表的对等式协议的特点,实现了 Client/Server 和 P2P 两种模式的融合.为了保证整个流媒体系统的健壮性和适应性,一方面采用了多描述视频编码 MDC 技术,另一方面通过在组播成员之间维护多个组播树来实现.

在 CoopNet 系统中,一个或者一组高性能中心服务器负责组织节点建立和管理多个数据分发树,这些数据分发树就构成整个系统的 P2P 对等传输网络.

在一般的流媒体点播服务时,数据分发树所构成的 P2P 对等网络只是为了完善传统的服务器/客户端模式.当服务器没有超过负载阈值时,用户的请求通过服务器直接发送数据来响应.若服务器满载,

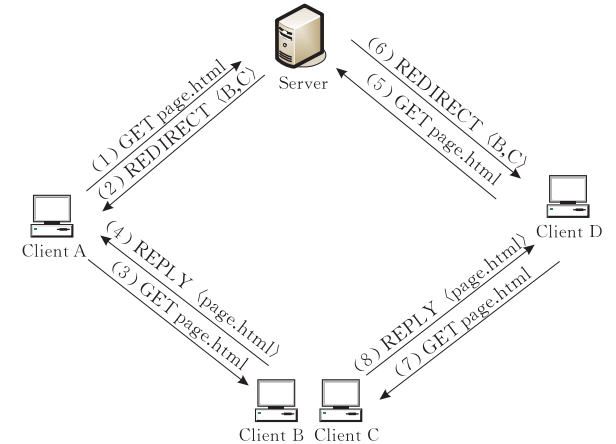


图 11 CoopNet 基本结构

仍有用户发送请求,则服务器通过查询记录找到历史用户列表,选出一定的待选节点转发应答(redirect response)至当前请求的用户,这样发出请求的节点就可以有选择性地与待选节点以 P2P 的方式传输所请求的流媒体内容.

在 CoopNet 系统应用于对实时性要求较高的流媒体直播时,所有参与直播的用户节点构成以多个数据分发树为核心的 P2P 数据拓扑,不同的 MDC 数据流沿着不同分发树路径传输至不同用户节点.通过在网络路径和数据编码两方面引入冗余机制来提高 CoopNet 系统的健壮性.

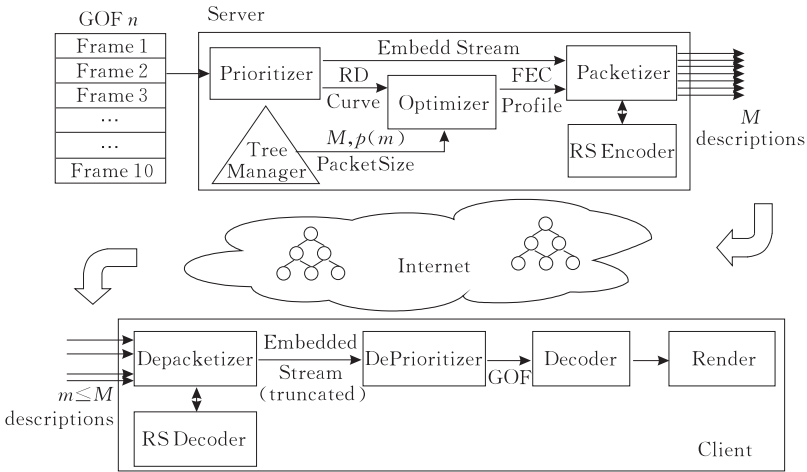


图 12 CoopNet 编码系统

在 CoopNet 系统中,由于需要同时维护多个组播树,控制开销大.同时由于 MDC 采用多个分层编

码,因此还必须考虑解决多路传送时的数据同步问题.



#### 4.2.2 标准方案

相比于 CoopNet 兼有两种传输模式, Promise 系统<sup>[41]</sup>属于名副其实的多发送端单接收端的 P2P 流媒体系统. 它的实现独立于底层网络, 可以构建在多种 P2P 底层网络之上(如 Pastry, CAN), 因此灵活性强, 扩展性好.

在 Promise 系统中, 节点之间的连接控制, 对成员节点的管理和对目标节点的查询都由底层网络实现. 当一个节点发出数据请求后, 通过底层网络查询返回一系列满足要求的候选节点, 然后按照基于拓扑(topology-aware)的原则选择传输性能相对较高的节点组成活跃发送集合(active sender set), 其它节点作为备用发送集合(standby sender set). 最后, 由原接收端节点向活跃发送集合的所有节点发起连接, 并行从多个节点接收数据, 如图 13 所示. 在连接建立后, 由接收端来控制每个发送节点的发送速率和数据分配, 发送端只需要按照接收到的控制信息来执行.

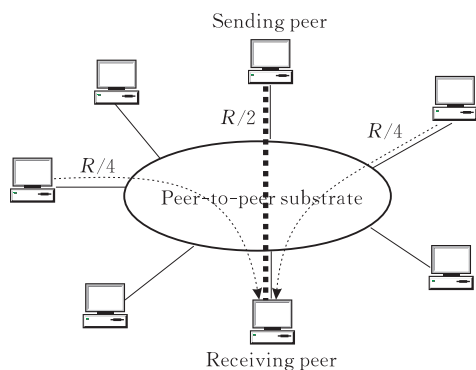


图 13 Promise 结构图

在 Promise 系统中, 当网络或者发送节点出现意外故障时, 接收端通过检测传输的速率判断网络状况和发送节点是否发生故障. 如果是网络的波动造成大范围的传输速率降低, 接收端节点就相应地动态调整总传输率和每个发送端节点的发送速率; 如果认为是发送节点出现故障, 就依据路径原则从备用发送集合选出新的节点进行替换.

同时, Promise 也通过采用前向纠错编码 FEC (Forward Error Correction) 增加视频编码的冗余性, 以此提高健壮性. 具体来说, 就是把视频流分成等长的数据段, 对每段数据进行编码, 接收端对接收到的数据进行纠错处理, 可以一定程度容忍数据包的丢失, 提高信道的传输性能.

#### 4.3 多发送端多接收端方式

作为大规模流媒体应用的解决方案, 单发送多

接收端和多发送单接收端两种模式都具备一定的优点. 但在前两种模式中, 我们可以发现在数据拓扑结构中, 不同的节点(根据发送或接收来区分)的负荷和地位并不完全相同. 比如应用层组播树中的父节点除了接收数据还需要转发数据, 显然比子节点的负荷更大. 所以当父节点成员离开或者失效时, 往往容易造成拓扑结构的突变和传输的暂停. 虽然已经有一些组播树修复算法被提出<sup>[42]</sup>, 但在动态变化的网络环境中, 组播树的破裂和修补在所难免.

多发送多接收端的 P2P 方式融合了前两种方式的特点, 一定程度克服了前两者的不足. 任何节点既可以接收多个节点的数据, 也可以向多个节点发送数据, 通过建立真正对等互联的体系结构达到了去中心化(decentralize)的效果, 因此被称为纯粹的 P2P 模式(Pure P2P).

在此类协议中, 一个媒体文件会按照一定长度被分为许多块, 块的长度主要由网络环境和客户端性能来决定. 节点通过发送数据请求给中心节点或者通过洪泛至其它节点来获得网络上数据的分布信息, 按照某种规则(比如延时)来选择每个数据块的来源. 而接收到数据请求的节点则负责响应或者转发数据请求. 由于任何一个节点都涉及到同时担任服务器端和客户端, 因此实现机制的设计直接关系到运行的效率.

根据获取数据块信息的获取方式, 我们可以将多发送多接收端协议分成: 中心化获取方式(Centralized Request Way)和分布式获取方法(Decentralized Request Way). 在中心化方式中, 节点无法通过 P2P 网络本身进行目标寻找, 而只能通过目录服务器来查找目标节点. 而在分布式获取方式中, 节点往往通过支持分布式操作的通信协议(如 Gossip 协议)获得其它邻居节点的状态信息, 从而寻找到目标节点.

本节分别介绍多发送端多接收端协议两种方式的特点.

##### 4.3.1 中心化获取方式

作为中心化获取方式的典型代表, BitTorrent 协议是当前最为流行的提供文件和其它内容共享的 P2P 网络协议, 具备了高扩展性、差错容忍性和独立性, 易于部署应用, 得到了大范围的使用.

在典型的 BitTorrent 协议中, 节点通过目录服务器来查找目标节点(这一点与最早的 Napster 系统类似). 当一个节点把文件共享为种子(seed)时, BitTorrent 协议把共享的文件按 256 KB 大小分成

数据块,同时把共享的文件信息发布在目录服务器上.其它对该内容感兴趣的用户节点,只需要点击种子信息,即可通过源节点和所有用户节点构成的 P2P 网络传输数据.其中种子信息包含了所有参与节点已经下载的数据和尚未下载数据的情况.

在用户节点接收数据同时,其已经下载的数据块作为新的种子,同时又可以被其它多个节点下载.这样源文件就通过多个种子的方式分布于整个 P2P 网络之中,即使拥有完整文件的节点离开,只要所有存在节点的数据块能构成一个完整的文件,就能保证每个节点都能获得完整的文件.因此系统的鲁棒性和自适应能力很强,少数节点的故障或者退出都不会对整个结构造成重大影响.对于简单的文件传输,这种策略是容易实现的,且只需要占用较少的网络带宽,因此得到了大范围的推广应用.

但在 BitTorrent 协议中,服务质量(如延迟和抖动)始终很难保证,因此 BitTorrent 协议在对延迟和抖动要求不高的文件共享中显得游刃有余,但不能直接适用于流媒体传输.传统的 CDN 网络通过部署高性能的中心服务器和靠近用户的边缘代理服务器,能够为用户提供高质量的流媒体服务,但限于成本问题,又很难大范围应用.

在文献[43]中,Skevik 等提出了一种混合 BitTorrent 和 CDN 技术的流媒体方案,一方面利用了 P2P 对等网络易于部署,具备高扩展性和鲁棒性,另一方面利用 CDN 的内容分发技术和流量负载均衡技术提供一定的安全保障和满足用户需求的服务质量,能够有效降低主干网络的流量负担.同时提出基于代理的结构(proxy based structure)以解决防火墙对 P2P 网络下载流量的影响.

如图 14,最左边的用户端应用程序(Client app),如视频播放软件,负责播放本地主机(LHC)所接收的视频文件.本地主机和代理服务器,即本地内容缓存 SCC(Site Content Cache)通信,同时从代理服务器和本地网其它主机获得数据,并发送已获得的数据至其它主机.本地的代理服务器又从主内容服务器(main content server)和其它代理服务器获得数据.

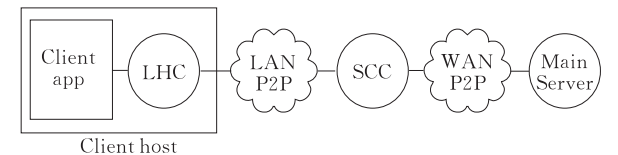


图 14 融合式流媒体方案结构图

在整个结构中,本地的主机在代理服务器的组织下构成一个底层 P2P 网络,所有的代理服务器在主内容服务器的组织下构成一个高层的 P2P 网络.在功能上,代理服务器除了存储容量更大,功能更强,而且还需要根据本地主机的工作情况,进行流量均衡处理,比如缓存注入、缓存替换等.

#### 4.3.2 分布式获取方法

在文献[9]中,作者提出了一种典型属于分布式获取方法的 DONet(Data-driven Overlay Network)协议,通过构建纯粹的 P2P(Peer to Peer)覆盖网络实现数据的传输,无需构建复杂的控制结构.基于 DONet 协议的实时流媒体播放系统 Cool-Streaming,其出色的播放效果、较低的延迟已经在实际运行中得到了证实和肯定.

DONet 的核心思想非常简单:每个节点通过 SCAM(Scalable Gossip Membership protocol)协议周期性和协作节点交互有关数据有效性的信息,从若干伙伴节点那里获得自身没有的数据,同时发送自身拥有的数据给其它需要数据的节点.除了提供节目的数据源节点(origin node),其它节点既可以作为数据接收方也可以作为数据提供方,完全取决于数据的有效信息.由于不需要构造复杂的全局拓扑结构,所以具备很强的可扩展性、高效性和鲁棒性.

从 DONet 的结构图中可以看到其具有 3 个核心模块:(1) 成员管理模块,负责帮助节点获得其它部分覆盖节点的信息.(2) 协作管理模块,负责建立和保存与其它节点的合作信息同时负责周期性地从成员列表中选择一些更优的节点(带宽更高或者有效数据更多)建立协作关系,这样可以提高系统的性能和鲁棒性.(3) 调度模块,采用启发式算法安排视频数据的传输.这 3 个核心模块的设计方法直接关系到整个实时流媒体系统的运行效率.

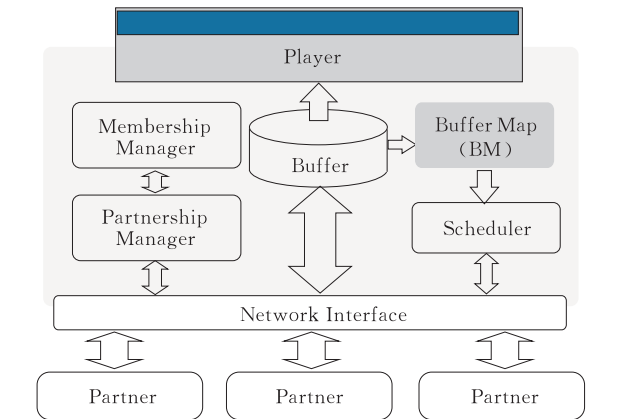


图 15 DONet 系统结构图

在 DONet 协议中,视频都被分成了固定长度的片段.节点缓存中每个片段的有效性通过缓冲图 BM(Buffer Map)来表示.比如采用 120 位来记录 BM,其中 1 表示数据有效,0 表示数据无效,则 120 位的 BM 可以表示 120 个片段的有效信息.节点通过不停地和协作节点交互 BM 获得数据的有效信息,然后确定可以从哪些协作节点获取自身没有的视频片段.

考虑实况流媒体直播对实时性的要求,节点都是半同步的.因此在确定片段长短后,一方面需要节点具有一定的数据缓冲区来缓存一定长度的视频,保证播放的流畅,更重要的就是设计一个高效的调度算法,来满足视频回放的时间要求和节点间的带宽传输限制. DONet 协议中采用的启发式调度算法,通过视频片段的提供者数量来确定优先顺序,然后从少到多来进行处理.针对同一片段的提供者,又按照节点的带宽和延时来确定优先顺序.

根据前文对基于 P2P 技术的流媒体方案的分析可以看到,虽然异构性、网络服务质量和可扩展性的问题得到了一定程度的解决,但安全性问题却始终是研究中的一大难点.首先,由于大规模流媒体服务的成员是频繁变化的,可能每时每刻都有很多的人员加入或退出组播组,这种动态变化下的安全问题是复杂的,同时 P2P 技术的中间节点不可靠性增加了特殊的安全性挑战.其次是高开销的问题,当规模从几个节点到上万个节点甚至更多,保存密钥所占用的节点存储空间、密钥生成所需要的计算量、密钥发送所占用的网络带宽、密钥更新的时间延迟和密钥更新的频率都会相应增加.而且利用基于 P2P 技术的大规模流媒体应用,其对带宽与计算资源的消耗要明显大于基于网络层组播的方案.另外,由于视频流传输的数据量非常大,同时对实时性的要求很高.所以对视频加密时必须考虑到其中的时延敏感性. P2P 技术传输效率低于 IP 层组播,其实现机制本质上是多次单播,时延敏感性问题尤为突出.对于大规模流媒体应用,如何减少密钥传输时延是一个重要问题.此外,现有的多媒体通信通常需要利用转码、应用层码率的自适应控制和速率整形等机制来提高视频通信中的服务质量,加密后的视频由于语法结构的变化可能无法有效地实施上述的操作,所以这就要求加密与数据嵌入算法要保持对多媒体通信中 QoS 控制机制的透明性.

#### 4.4 P2P 和 CDN 的结合

对等网(P2P)技术的出现给解决当前大规模流

媒体应用中的网络与系统瓶颈问题带来新的机遇,现有的 P2P 技术面临的主要问题包括:缺乏中心化的管理和健壮性,其动态性使其缺乏对服务质量的保障;纯粹的 P2P 技术占用了大量骨干带宽资源,耗费大量跨 ISP(电信运营商)的带宽;传统的服务器辅助的 P2P 系统中,松散组织的超级节点的过重负载容易引起 SN 的链式崩溃反应.文献[44]基于大规模 P2P 流媒体系统 CoolStreaming,研究了视频直播系统的工作负载,系统动态性和性能测试,并发现了随机邻居节点选择和多子流能有效解决 P2P 系统动态性问题.在详尽数据结果和理论分析的基础上,作者证明:(1)在视频直播中一个关键问题是在过多用户同时加入(flash crowd)导致的过长初始接入时间和过高的接入拒绝率;(2)系统动态性是影响整体系统性能的最关键参数;(3)不同节点在系统中的上传带宽严重不均衡,极大影响了系统资源的分配;(4)在不同系统参数下,系统需要考虑关键设计的折中.一方面过长的初始接入时间和过高的失效率是基于 P2P 技术的流媒体系统本身带来的问题,这种情况在 NAT 和防火墙后有大量不可利用节点时尤其严重.同时当系统中节点较少时,寻找合适的伙伴节点需要耗用较长的时间.因此在直播中部署适当的服务器节点变得必要<sup>[10,45]</sup>.但我们不能回避的是纯粹基于服务器模式的方式不具备很好的扩展性,其部署和维护代价昂贵.因此,一个大型的网络视频直播系统需要是一种融合服务器与 P2P 的混和结构.

在产业界作为网络加速技术的 CDN(内容分发网络)得到广泛应用. CDN 采用分布式缓存、负载均衡、流量工程等技术在已有的 Internet 上构筑一个分布式的覆盖网络,通过将内容从信源推送到网络边缘设备,一方面,使得用户得以在“就近”的位置快速访问到所需的内容,降低了端到端时延,提升了用户服务质量;另一方面,突破了中心服务器的性能瓶颈,减轻了骨干网络流量,有效缓解了高吞吐率内容传输对骨干网络的压力,在一定程度上也增加了系统容量.近年来 CDN 得到越来越多的重视并在国内外得到广泛的部署,有代表性的 CDN 服务提供商有 Akamai<sup>①</sup>等.

由于 P2P 与 CDN 具有较强的技术互补性,设计新的架构将两者优势相结合,是克服当前大规模

① Akamai. Akamai Technologies, Inc., [www.akamai.com](http://www.akamai.com), 2008

流媒体解决面临挑战的有效途径. P2P 技术的优点在于低成本、高可扩展性,这是传统 CDN 所欠缺的;而 CDN 的可靠性和可管理性将能解决 P2P 技术许多顽疾. 融合 CDN 与 P2P 的混合式流媒体系统成为大规模应用的体系发展趋势,但将两者结合的研究工作尚处步阶段<sup>[46-47]①</sup>. 现有研究要么是 1+1 式的 P2P 与 CDN 模式简单相加<sup>[47]</sup>,要么是服务器支撑 P2P 的模式,它们只能解决大规模流媒体应用所面临的部分问题,不能同时满足服务质量、可扩展性、安全性和异构性需求.

## 5 安全可扩展的流媒体系统 TrustStream

从 IP 组播、应用层组播到 P2P 技术的发展以及各种技术的综合使用,虽然一定程度上解决了可扩展性的问题,但是这些方案还不能令人满意地解决 QoS、网络异构和安全等问题.

在文献[48-49]中,我们在应用层流媒体组播领域首次将可扩展分层视频编码(PFGS)的可扩展性、可控性与 P2P 的传输思想结合了起来,设计并实现了一套新的安全流媒体系统(TrustStream)和相关的安全应用层组播协议(Secure-ALM)与调度算法.

如图 16,PFGS 编码服务器首先对视频采集的流媒体文件进行编码,产生流媒体基本层和增强层.

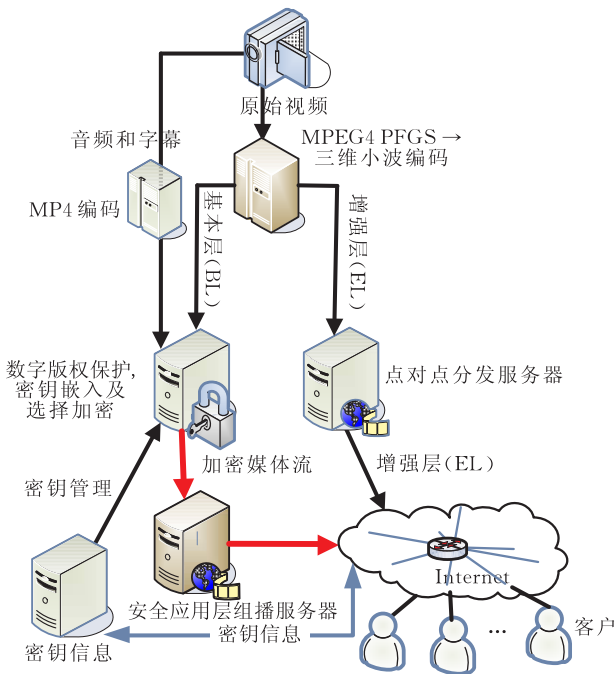


图 16 TrustStream 的系统架构

由于基本层包含了最重要特征的数据,因此需要版权保护及加密服务器对基本层进行加密和版权保护,防止非法用户的接入和非法拷贝. 加密后的流媒体基本层通过单发送多接收端方式的应用层组播传输方式进行传播发布. 对于用作进一步提高视频质量的增强层,直接利用多发送多接收端的 P2P 传输方式进行传播发布.

一方面采用 PFGS 编码,由于利用了多个视频层进行预测,增强了可伸缩性和容错性,能够获得不同输出速率的流媒体,可以很好地适应网络与终端用户的异构性.

另一方面为了实现系统的可扩展性、可控性和安全性,本系统架构创新地提出了针对基本层进行版权保护及选择加密,降低了服务器的负载. 由于 PFGS 的解码是以成功接收基本层为前提,即使未经授权的用户获取到增强层,也无法正常进行解码播放. 同时对增强层采用多发送多接收端的传输方式,进一步增强了系统的可扩展性.

TrustStream 系统不仅解决了大规模流媒体应用中服务器容易成为性能瓶颈和实际应用中的网络与用户异构性问题,更创新地解决了流媒体的加密、密钥分发等一系列安全性问题,对大规模流媒体应用中安全性这一研究的难点问题做出了一定的突破.

## 6 总结和展望

由于流媒体组播技术可以大幅度提高网络传输的效率,具有良好的应用前景,非常可能成为 Internet 上最受欢迎的应用之一. 但是 Internet 规模庞大,异构性强,视频传输对网络带宽和延迟等 QoS 性能有特殊的要求;同时,流媒体应用中的安全保护和版权保护始终都是非常棘手的问题.

一个完善的流媒体解决方案必须综合编码技术和网络传输技术来解决上述问题的. 一方面,可以结合新的视频编码技术(如可扩展编码 FGS 和多描述编码 MDC 等)和视频加密技术提高视频流对网络异构和安全的支持. 同时,可以综合 CDN 的内容分发技术、网络组播技术到 P2P 技术等网络传输技术进一步提高大规模流媒体系统的可扩展性. 根据网络应用环境和用户需求,我们可以设计不同的流媒体方案. 比如 CoopNet 系统即是为了满足对数据可

① Cachelogic. <http://www.cachelogic.com>, 2007



靠性要求很高的用户需求,结合 MDC 编码和多组播树的 P2P 技术来实现的典型例子.其它可伸缩编码和最新的 P2P 技术(比如 DHT)的有效结合,成为了流媒体应用的未来方向.

除了 FGS 编码、MDC 编码外,3D 小波编码作为一种可伸缩性的视频编码技术,目前已经成为 MPEG-21 可伸缩视频编码组的重点研究方案.但由于小波编码在处理时需要把整个一帧或一帧中的一大块图像作为一个单元来处理,要占用较大的系统资源,因此需要通过新的技术来改进编码和解码的处理速度使其实用化.

从网络传输技术角度来看,虽然基于分布式散列(DHT)的结构化 P2P 技术得到了迅猛的发展,但要应用于大规模的流媒体应用,我们还面临不小的挑战.目前分布式散列算法一般采用的是 consistency Hash(Consistent Hash),比如 SHA-1 Hash 函数.这类一致性 Hash 函数虽然能够兼顾负载均衡和一定安全保障,但却存在明显的缺陷.比如在构建逻辑结构的时候,并没有太好的方法解决物理地址和逻辑地址不一致的情况,因此在一定程度上降低了在大规模流媒体方案中的实际效率.同时也不能保证相同类型的流媒体资源能够在物理上邻近存放,很可能出现源两个内容相关度很高的多媒体资源由于 Hash 生成了完全不同的散列值,被存放到了完全随机的两个节点.因此针对应用于流媒体应用和其它 P2P 应用的需要,我们需要进一步研究更适合分布式散列的 Hash 函数,使其能够实现负载均衡和一定安全性的前提下,解决逻辑网络和物理网络的不匹配问题,在一定程度上提高内容和语义的耦合度,这对整个 P2P 技术的发展都是具有深远意义的.

## 参 考 文 献

- [1] Wu D-P, Hou Y-T, Zhu W, Zhang Y-Q, Peha J. Streaming video over the internet: Approaches and directions. *IEEE Transactions on Circuits System Video Technology*, 2001, 11(3): 282-300
- [2] Li Wei-Ping. Overview of fine grannlarity scalability in MPEG-4 video standard. *IEEE Transaction on Circuits and System for Video Technology*, 2001, 11(3): 301-317
- [3] Wu F, Li S, Zhang Y-Q. A framework for efficient progressive fine granularity scalable video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 2001, 11(3): 282-300
- [4] Wang Y, Lin S. Error-resilient video coding using multiple description motion compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 2002, 12(6): 438-453
- [5] Deering S, Cheriton D. Multicast routing in datagram inter-networks and extended LANs. *ACM Transactions on Computer Systems*, 1990, 8(2): 85-110
- [6] Day M, Gilletti D. Distribution network peering scenarios. *IETF Internet-Draft*, 2001-05
- [7] Chu Yang-Hua, Ra Sanjay, Seshan Srinivasan, Zhang Hui. Enabling conferencing applications on the Internet using an overlay multicast architecture//*Proceedings of the ACM SIGCOMM*. San Diego, 2001: 55-67
- [8] Chu Yang-Hua, Rao Sanjay, Zhang Hui. A case for end system multicast//*Proceedings of the ACM Sigmetrics*. Santa Clara, CA, 2000: 1-12
- [9] Zhang X, Liu J, Li B, Yum T-SP. CoolStreaming/DoNet: A data-driven overlay network for peer-to-peer live media streaming//*Proceedings of the IEEE INFOCOM'05*. Miami, FL, 2005: 2102-2111
- [10] Li B, Yin H. The peer-to-peer live video streaming in the Internet: Issues, existing approaches and challenges. *IEEE Communications Magazine*, 2007, 45(6): 94-99
- [11] Heising G, Marpe D, Cycon H L, Petukhov. A Proposal for ITU-T H.26L: A wavelet-based video coding scheme using OBMC and image warping prediction. *ITU-T*, Mosterey, 1999
- [12] Bolot J-C, Turletti T. A rate control mechanism for packet video//*Proceedings of the IEEE INFOCOM'94*. Toronto, Canada, 1994: 1216-1223
- [13] Bolot J-C, Turletti T, Wakeman I. Scalable feedback control for multicast video distribution in the Internet//*Proceedings of the ACM SIGCOMM'94*. London, UK, 1994: 58-67
- [14] Li X, Ammar M H. Bandwidth control for replicated-stream multicast video distribution//*Proceedings of the HPDC'96*. Syracuse, USA, 1996: 6-9
- [15] McCanne S, Jacobson V, Vetterli M. Receiver-driven layered multicast//*Proceedings of the ACM SIGCOMM'96*. California, USA, 1996: 117-130
- [16] Vickers B, Albuquerque C, Suda T. Source adaptive multi-layered multicast algorithms for real-time video distribution. *IEEE/ACM Transactions on Networking*, 2000, 8(6): 720-733
- [17] Liu J-C, Li B, Zhang Y-Q. An end-to-end adaptation protocol for layered video multicast using optimal rate allocation. *IEEE Transactions on Multimedia*, 2004, 6(1): 87-102
- [18] Shacham N, McKenney P. Packet recovery in high-speed networks using coding and buffer management//*Proceedings of the IEEE INFOCOM'90*. San Francisco, USA, 1990: 124-131
- [19] Biersack E W. Performance evaluation of forward error correction in ATM networks. *Computer Communications Review*, 1992, 22(4): 248-257



- [20] Zhang L, Deering S, Estrin D, Shenker S, Zappala D. RSVP: A new resource reservation protocol. *IEEE Network*, 1993, 7(5): 8-18
- [21] Wu D, Hou Y T, Zhang Y Q. Transporting real-time video over the Internet: Challenges and approaches. *Proceedings of IEEE*, 2000, 88(12): 1855-1875
- [22] Padhye J, Firoiu V, Towsley D, Kurose J. Modeling TCP throughput: A simple model and its empirical validation//*Proceedings of the ACM SIGCOMM 98*. Vancouver, Canada, 1998: 303-314
- [23] Sisalem D, Wolisz A. MLDA: A TCP-friendly congestion control framework for heterogeneous multicast environments//*Proceedings of the IWQoS 2000*. Pittsburgh, USA, 2000: 65-74
- [24] Kwon Gu-In, Byers J. Smooth multirate multicast congestion control//*Proceedings of the IEEE INFOCOM'03*. San Francisco, USA, 2003: 1022-1032
- [25] Yin Hao, Lin Chuang, Qiu Feng, Liu Jiang-Chuan, Min Ge-Yong, Li Bo. CASM: A content-aware protocol for secure video multicast. *IEEE Transactions on Multimedia*, 2006, 8(2): 270 - 277
- [26] Banerjee S, Bhattacharjee B, Kommareddy C. Scalable application layer multicast//*Proceedings of the ACM SIGCOMM*. Pittsburgh, USA, 2002: 43-51
- [27] Zhang B, Jamin S, Zhang L. Host multicast: A framework for delivering multicast to end users//*Proceedings of the IEEE Infocom*. New York, USA, 2002: 1366-1375
- [28] Banerjee S, Kommareddy C, Kar K et al. Construction of an efficient overlay multicast infrastructure for real-time applications//*Proceedings of the IEEE INFOCOM*. San Francisco, 2003: 1521-1531
- [29] Deshpande H, Bawa M, Garcia Molina H. Streaming live media over a peer-to-peer network. Stanford University, Stanford, CA, USA: Technical Report, 2001
- [30] Padmanabhan V, Wang H, Chou P et al. Distributing streaming media content using cooperative networking//*Proceedings of the ACM NOSSDAV*. Miami, 2002: 177-186
- [31] Castro M, Druschel P, Kermarrec A M et al. SplitStream: High-bandwidth content distribution in a cooperative environment//*Proceedings of the IPTPS'03*. Berkeley, 2003: 292-303
- [32] Banerjee S, Bhattacharjee B. A comparative study of application layer multicast protocols. University of Maryland, College Park, MD, USA: Technical Report, 2002
- [33] Ratnasamy S, Handley M, Karp R. Application-level multicast using content-addressable networks//*Proceedings of the 3rd International Workshop on Networked Group Communication*. London, UK, 2001: 14-29
- [34] Castro M, Druschel P, Kermarrec A-M, Rowstron A. SCRIBE: A large-scale and decentralized application-level multicast infrastructure. *IEEE Journal on Selected Areas in Communications (JSAC)*, 2002, 20(8): 1489-1499
- [35] Zhuang S Q, Zhao B Y, Joseph A D, Katz R H, Kubiatiowicz J D. Bayeux: An architecture for scalable and fault-tolerant wide-area data dissemination//*Proceedings of the 11th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV 2001)*. New York, USA, 2001: 11-20
- [36] Rowstron A, Druschel P. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems//*Proceedings of the IFIP/ACM International Conference on Distributed Systems Platforms (Middleware)*. Heidelberg, Germany, 2001: 329-350
- [37] Zhao B Y, Kubiatiowicz J, Joseph A. Tapestry: An Infrastructure for fault-tolerant wide-area location and routing. University of California, Berkeley, CA, USA: Technical Report UCB/CSD-01-1141, 2001
- [38] Chawathe Y. Scattercast: An architecture for Internet broadcast distribution as an infrastructure service [Ph.D. dissertation]. University of California, Berkeley, CA, USA, 2000
- [39] Jannotti J, Gifford D K, Johnson K L, Kaashoek M F, James W O'Toole, Jr. Overcast: Reliable multicast with an overlay network//*Proceedings of the OSDI*. San Diego, 2000: 197-212
- [40] Pendarakis D, Shi S, Verma D et al. ALMI: An application level multicast Infrastructure//*Proceedings of the 3rd USENIX Symposium on Internet Technologies and Systems*. California, USA, 2001: 49-60
- [41] Hefeeda M, Habib A, Botev B, Xu D, Bhargave B. PROMISE: Peer-to-peer media streaming using CollectCast//*Proceedings of the ACM Multimedia 2003*. Berkeley, CA, 2003: 45-54
- [42] Yang M, Fei Z. A proactive approach to reconstructing overlay multicast trees//*Proceedings of the INFOCOM'04*. Hong Kong, China, 2004: 2743-2753
- [43] Skevik K-A, Goebel V, Plagemann T. Analysis of BitTorrent and its use for the design of a P2P based streaming protocol for a hybrid CDN. Department of Informatics, University of Oslo, Norway: Technical Report, 2004
- [44] Li B, Xie S-S, Keung G Y, Liu J-C, Stoica I, Zhang H. An empirical study of the CoolStreaming+System. *IEEE Journal on Selected Areas in Communications*, Special Issue on Advances in Peer-to-Peer Streaming System, 2007, 25(9): 1-13
- [45] Li B, Xie S-S, Qu Y, Keung G, Liu J-C, Lin C, Zhang X-Y. Inside the new coolstreaming: Principles, measurements and performance implications//*Proceedings of the IEEE Infocom'2008*. Phoenix, AZ, 2008
- [46] Tu Y, Sun J, Hefeeda M, Prabhakar S. An analytical study of peer-to-peer media streaming systems. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2005, 1(4): 354-376
- [47] Xu D, Kulkarni S, Rosenberg C, Chai H. A CDN-P2P hybrid architecture for cost-effective streaming media distribution. *Computer Networks*, 2004, 44(3): 353-382

[48] Yin H, Lin C, Qiu F, Liu X, Wu D. TrustStream: A novel secure and scalable media streaming architecture//Proceedings of the 13th ACM International Conference on Multimedia. Singapore, 2005: 295-298

[49] Yin Hao, Lin Chuang, Zhang Qian, Chen Zhi-Jia, Wu Da-Peng. TrustStream: A secure and scalable architecture for large-scale Internet media streaming. IEEE Transactions on Circuits and Systems for Video Technology, 2008



**YIN Hao**, born in 1974, Ph. D. , associate professor. His research interests include performance evaluation for Internet and wireless network, image/video coding, multimedia over wireless network, and security.

**LIN Chuang**, born in 1948, Ph. D. , professor. His current research interests include computer networks, performance evaluation, logic reasoning, and Petri net theory and its applications.

**WEN Hao**, born in 1983, Ph. D. candidate. His research interests include modeling and performance analysis for wireless network.

**CHEN Zhi-Jia**, born in 1981, Ph. D. candidate. His research area is in computer network and media streaming, especially for architecture design and analytical modeling for P2P media streaming system.

**WU Da-Peng**, Ph. D. , assistant professor. His research interests are in the areas of networking, communications, multimedia, signal processing, and information and network security.