

# 面向 P2P 流媒体直播的数据分发机制研究进展

彭雪娜<sup>1,2</sup> 李 佳<sup>1</sup> 闻英友<sup>1,2</sup> 赵 宏<sup>1,2</sup>

(东北大学复杂网络系统安全保障技术教育部工程研究中心 沈阳 110004)<sup>1</sup>

(东软集团股份有限公司东软研究院 沈阳 110004)<sup>2</sup>

**摘 要** 数据分发机制是 P2P 流媒体直播系统中的核心技术。根据流媒体数据分发过程中数据分发驱动因素的不同,将应用于 P2P 流媒体直播业务的数据分发机制分为 3 类,分别是路径驱动的、数据驱动的和混合驱动的。通过对以上 3 类数据分发机制的典型技术进行深入的研究与分析,概括并总结了现有技术的优缺点。最后,基于研究现状分析,指出了该领域未来的发展方向。

**关键词** P2P 网络,流媒体直播,数据分发机制,应用层组播

**中图法分类号** TP393 **文献标识码** A

## Recent Advances of Data Distribution Mechanisms for Live P2P Streaming

PENG Xue-na<sup>1,2</sup> LI Jia<sup>1</sup> WEN Ying-you<sup>1,2</sup> ZHAO Hong<sup>1,2</sup>

(Education Ministry Engineering Research Center of Safety and Security of Complex Network System,  
Northeastern University, Shenyang 110004, China)<sup>1</sup>

(Neusoft Research, Neusoft Co. Ltd, Shenyang 110004, China)<sup>2</sup>

**Abstract** Data distribution is one of the key technologies in P2P live streaming system. Considering the driving element in data distribution, the data distribution mechanisms for P2P live streaming system can be classified into three categories, namely: path-driven, data-driven and hybrid-element-driven. This paper thoroughly investigated the classical technologies of the above data distribution mechanisms. Finally, some issues to be further studied were discussed.

**Keywords** Peer to peer network, Live streaming, Data distribution mechanism, Application layer multicast

随着互联网基础设施及应用技术的快速发展,以网络音频、视频为代表的流媒体业务已经成为互联网上最为流行的业务之一。与传统业务相比,流媒体业务需要传输的数据量更大,持续传输时间更长,而且同时参与的用户更多。因此,如何在有限的网络资源条件下实现高效的流媒体数据分发,是流媒体业务能否快速发展和大规模应用的关键。

传统的流媒体系统主要采用了客户端-服务器的模式,即服务器以单播的方式为每个客户端建立独立的连接。在这种模式下,由于系统规模严重受限于服务器性能和网络带宽,因此服务器和网络基础设施容易成为系统瓶颈。为了解决可伸缩性问题,许多研究提出采用 IP 组播技术来实现流媒体系统。IP 组播技术可以实现高效的一对多通信,有效减轻服务器和网络负载,提高系统可伸缩性。然而,该技术存在可靠传输、拥塞控制等多种限制,迄今仍未得到广泛应用,因此大大限制了基于该技术的流媒体系统的部署和应用。除 IP 组播方案外,也有研究提出通过在网络边缘部署代理缓存(proxy cache)或内容分发网络(Content Distribution Network, CDN)来分担流媒体服务器和互联网的负载。在这种方案中,当客户向流媒体系统请求服务时,系统可以就近安排客户从代理

缓存或 CDN 服务器上获得数据,而无需从远端的流媒体服务器上获取。然而,该方案成本较高,而且只能部分地解决可伸缩性问题,因此其适用范围有限。

近年来,随着对等(Peer-to-Peer, P2P)网络技术在文件共享、应用层组播等领域中取得巨大成功,人们逐渐意识到这种用户节点全员参与的工作模式的技术优越性。目前许多研究已经将 P2P 技术引入到流媒体技术领域,并提出了多种技术方案,如 Gnustream<sup>[1]</sup>, P2Cast<sup>[2]</sup>, ESM<sup>[3]</sup>, Coolstreaming<sup>[4]</sup>等。这些技术方案在部署的灵活性、系统的伸缩性、抗扰动能力、播放延迟等方面已经远远超越了传统方案。事实上,目前互联网上处于领导地位的流媒体服务大多是采用 P2P 技术实现的,如 pplive, qqlive, ppstream 等。

直播业务是 P2P 流媒体最为重要的业务模式之一,也是近年来互联网上发展较快的一项业务。该业务的迅速部署和应用给互联网用户带来了前所未有的工作、学习、娱乐体验。从技术角度来看, P2P 流媒体直播业务的成功主要得益于其分布式的数据分发机制。因此,深入地展开对 P2P 流媒体直播系统中数据分发机制的研究将有助于加深对 P2P 流媒体直播技术的理解,并进一步改善此类系统的性能。本文主要

到稿日期:2009-06-16 返修日期:2009-09-16 本文受国家自然科学基金(60602061, 60803131)资助。

彭雪娜(1979—),女,博士,主要研究方向为 P2P 网络、网络安全等, E-mail: pengxn@gmail.com; 李 佳(1984—),女,博士生,主要研究方向为 P2P 网络技术; 闻英友(1975—),男,博士后,主要研究方向为通信网络技术; 赵 宏(1954—),男,教授,博士生导师,主要研究方向为多媒体、网络安全等。

围绕 P2P 流媒体直播系统中最为关键的数据分发机制开展研究,并在对该技术研究现状的分类和分析基础上对该技术的未来发展方向开展了讨论。

## 1 P2P 流媒体直播概述

P2P 网络是由对等实体互联而形成的一种覆盖网络,现有的 P2P 网络大多构建于互联网之上。P2P 网络中,每个对等实体(Peer 节点)具有相同的地位,既是服务的提供者,又是服务的消费者。对等节点之间通过共享信息、存储、计算、带宽等资源,协同完成目标任务。

P2P 流媒体直播系统是指基于 P2P 网络技术实现的一类流媒体直播系统。在 P2P 流媒体直播系统中,所有观看同一频道的用户处于同一覆盖网中。当用户选择收看某一频道的视频时,它需要加入该频道所在的覆盖网络,并与该频道的服务节点以及收看该频道的其它用户建立服务连接,以获取流媒体数据;当用户从当前频道切换到另一个频道时,则先要从当前频道的覆盖网中退出,再加入到新频道所在覆盖网中。

在 P2P 流媒体直播系统中,流媒体数据的分发是由所有参与流媒体直播业务的用户节点共同完成的。P2P 网络技术可以将流媒体直播服务分散到各用户节点上,从而降低系统对于数据源服务器的依赖程度,极大地提高系统的可伸缩性。然而,由于流媒体直播业务对于网络数据分发延迟具有较为严格的时限要求,而且 P2P 网络中节点的异构性和动态性十分显著,因此如何在这种高动态的异构网络环境下及时、高效地分发流媒体数据,是 P2P 流媒体系统当前面临的巨大挑战。

## 2 P2P 流媒体直播中的数据分发机制分类

按照流媒体数据分发过程中数据分发的驱动因素,可以将目前应用于 P2P 流媒体直播系统的数据分发机制划分为 3 类:路径驱动的数据分发机制、数据驱动的数据分发机制和混合驱动的数据分发机制,分别如图 1(a),(b),(c)所示。

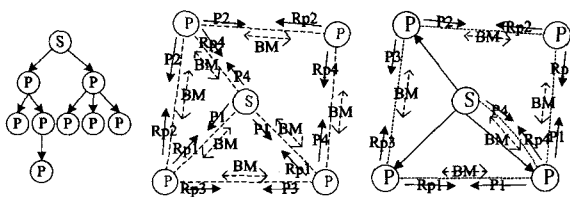


图 1 P2P 流媒体直播中的典型数据分发机制

路径驱动的数据分发机制通常基于应用层多播技术,可以在静态的网络环境下实现高效、低延迟的数据分发,但对动态网络环境的适应性稍显不足。该机制在数据分发之前,一般需要构造可以覆盖全部用户节点的数据分发路径,并且在数据分发过程中沿该路径分发流媒体数据。

数据驱动的数据分发机制是 P2P 网络中的典型数据分发机制,对动态网络环境具有良好的适应性,但是其数据分发效率较低,延迟较高。该机制无需在数据分发之前构建数据分发路径,只需在数据分发过程中在节点间定期交换彼此的状态信息(如节点状态、数据可用性等)。节点通过周期性地检查自身以及邻居节点的状态,识别自身的数据需求以及相关数据在邻居节点上的可用性情况,然后根据数据调度策略从邻居节点主动拉取数据(或向邻居节点主动推送数据)。

混合驱动的数据分发机制是路径驱动和数据驱动相结合的数据分发机制,可以有效融合二者优势,在动态网络环境下实现高效、低延迟的数据分发。在数据分发过程中,网络中的节点可以根据情况选择采用路径驱动的分发机制或数据驱动的分发机制。这种机制在初始阶段通常采用数据驱动的分发机制,随着数据以这种方式在节点间扩散,当满足一定条件时,就可以部分地进入路径驱动的数据分发机制,即在部分节点间构造数据分发路径,以便后续数据沿该路径进行高效的传播。

## 3 路径驱动的数据分发机制

数据分发路径结构选择以及路径构造方式的选择是路径驱动的数据分发机制的关键所在。不同的路径结构以及构造方式将会对 P2P 流媒体直播系统的可伸缩性、抗扰动能力、播放延迟、资源利用率等性能指标产生较大的影响。目前,主流 P2P 流媒体数据分发机制采用的路径结构大致可以分为以下几类:单树路径结构、多树路径结构和网-树路径结构。

### 3.1 基于单树路径结构的数据分发机制

基于单树路径结构的数据分发机制是由早期研究者作为应用层组播技术方案提出的,该机制大量参考了 IP 层组播技术,可以应用于包括新闻发布、网络电台等低带宽的大规模组通信业务。基于单树路径结构的数据分发机制通常是在给定网络拓扑结构上,构造并维护一棵以数据源为根节点、其它用户节点为中间或叶子节点的数据分发树,数据源发出的所有流媒体数据都沿该路径结构分发至全网。NICE 协议<sup>[5]</sup>、ZIGZAG 协议<sup>[6]</sup>是此类数据分发机制的典型代表。

NICE 协议<sup>[5]</sup>是马里兰大学的 Banerjee 等提出的一种可伸缩的应用层多播协议。该协议层次簇结构的覆盖网维护多播组成员,并在其上生成树结构的数据分发路径。覆盖网及分发树分别根据以下规则构造:首先,多播组的所有节点被分为规模为 $[k, 3k]$ 的若干簇,并从每个簇选出一个簇首作为高一层次的成员,以此类推,直至构造出如图 2 所示的层次簇结构;然后,从层次簇结构中生成一棵以数据源为根的树(如图 3 所示)作为数据分发路径,该分发树的结构为数据源向其所在簇的所有其它簇成员分发数据,接收到该数据的各层簇首(数据源所在簇的簇首除外)再向各自所在簇转发数据,直至数据分发至覆盖网中的所有节点。此外,为了覆盖网环境的动态性,NICE 协议还在引导新节点和拆分、合并、修复覆盖网等方面提出了相关操作方法。

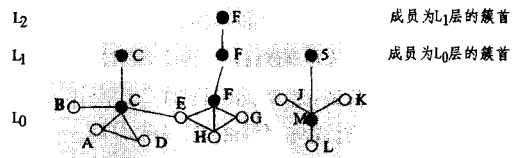


图 2 NICE 的覆盖网管理<sup>[5]</sup>

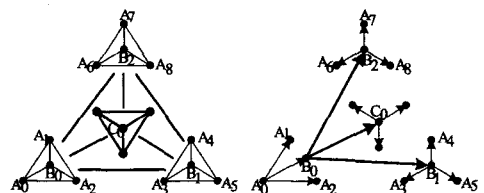


图 3 NICE 的控制和数据分发路径结构<sup>[5]</sup>

ZIGZAG<sup>[6]</sup>是Tran等提出的一种应用层多播协议。ZIGZAG协议包括两层拓扑结构:位于底层的表示节点逻辑关系的管理拓扑、位于上层的表示节点物理交互关系的数据分发拓扑。ZIGZAG的管理拓扑为图4所示的层次簇结构,数据分发拓扑为图5所示的单树结构。基于层次簇结构的管理拓扑,ZIGZAG按照如下两个原则构造数据分发树:①L-1层的簇节点不能从本层的簇头节点处获取数据,而必须从兄弟簇头处获取数据;②树中的每个节点只能向兄弟簇的成员节点转发数据。ZIGZAG中的节点加入与离开、簇分割与合并、负载均衡等一系列数据分发树维护操作都基于该原则实现。如此构造的数据分发路径具有路径维护局部化和控制开销低的特点。

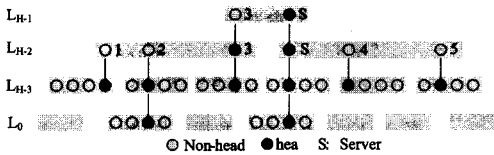


图4 ZIGZAG的覆盖网管理结构<sup>[6]</sup>

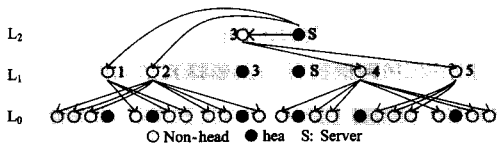


图5 ZIGZAG的数据分发路径结构<sup>[6]</sup>

基于单树路径结构驱动的分发模式是继承IP多播技术在应用层实现多播的最自然的方法,它不需要复杂的视频编码算法,就可以以低延迟实现数据分发。这种数据分发机制虽然在静态的以路由器为基础设施的IP多播中效果很好,但在动态的基于用户节点的P2P覆盖网环境下的表现却不够理想。首先,单点失效对系统的影响大且恢复成本高。分发树构造与维护过程时间长、控制负载大,对网络动态性缺乏弹性,中间节点失效对于子孙节点性能产生较大影响。其次,节点异构性对系统性能会产生较大影响。在异构网络中,若分发树的构造未考虑带宽限制,则其上的中间节点可能存在带宽瓶颈,影响后续节点的视频播放质量;若分发树的构造考虑带宽限制,则分发树无法达到最优结构,造成较大的源到端延迟<sup>[4]</sup>。第三,该机制容易导致负载不均衡。数据分发过程中,数据的复制和转发负载仅由位于树上的中间节点完成,叶子节点没有负载,因此系统负载严重不均衡。最后,该机制使系统的带宽利用率低下。基于单树的数据分发结构无法利用叶子节点的上行带宽,在系统带宽紧张的情况下,许多节点可能无法正常获得服务。

### 3.2 基于多树路径结构的数据分发机制

针对基于单树路径结构的数据分发机制中存在的对异构网络适应能力差、负载不均衡以及带宽利用不充分等问题,后续研究者提出了基于多树路径结构的数据分发机制。该机制通常按照一定策略构造多棵以数据源为根节点、所有用户节点为中间和叶子节点的数据分发树。为了在多树上有效地分发数据,通常需要将流媒体数据划分为多个子流,然后每个子流沿一棵分发树传输。

一般地,多树分发路径相配合,P2P流媒体直播系统通常会引入分级编码技术,如分层编码技术(Layered coding, LC)、多重描述编码技术(Multiple Description Coding, MDC)

等,其中以MDC编码的应用最为广泛。MDC编码技术的工作原理为:编码器将原始流媒体内容编码成两个或两个以上的独立平等的子流,解码器可以对任意子流集合进行解码和内容重构,恢复出与输入子流数量相对应的、具有一定质量的原始视频内容。

采用了多分发树和MDC编码技术的系统,通常将每个子流分别放在一棵组播树上传输,每个节点从其所在的多棵树上获取数据,再通过将各子流数据整合,还原成可用于播放的视频数据。对于任意一个用户节点而言,其可获得的视频质量依赖于其接收到的子流数量,接收到的子流数量越多,其播放质量越好。

与基于单树路径的数据分发机制不同,在基于多树路径的数据分发机制下,一棵分发树的叶子节点同时可以是其它分发树的中间节点,因此这种机制可以更充分地利用用户节点的上传带宽资源,并且通过与分级编码技术相结合,还可以有效地降低用户节点异构性对系统整体性能的影响。微软提出的CoopNet<sup>[7]</sup>,Splitstream<sup>[8]</sup>技术都是此类数据分发机制的典型代表。

Coopnet<sup>[7]</sup>是微软的Padmanabhan等提出的一种基于多树路径分发机制的流媒体直播技术。该技术首先在源服务器端对视频节目做多描述编码,产生多条MDC子流,然后为每一条MDC子流分配一棵独立的分发树。在Coopnet中,节点可以根据自身能力选择同时加入到部分或全部分发树中,并获得与其加入的分发树数量相当的视频质量。

Splitstream<sup>[8]</sup>是微软的Castro等提出的另一种基于多树路径分发机制的流媒体直播技术。与CoopNet相似,Splitstream也在源服务器端对视频节目做MDC编码,并让每棵分发树对应传输一条MDC子流。但是,在多条数据分发路径的组织上,SplitStream比Coopnet更进一步。鉴于基于单树路径的数据分发机制中,中间节点负载过重而叶子节点得不到充分利用的问题,Castro等在设计多棵数据分发树时特别强调每个Peer节点只在其所参与的一棵分发树中充当中间节点,而在剩余的其它分发树中只充当叶节点。

图6给出了这种机制的一个实例。Splitstream的这种设计,一方面可以均衡节点负载,更有效地利用用户资源,另一方面可以更有效地减小节点离开或失效对其它节点所造成的影响。

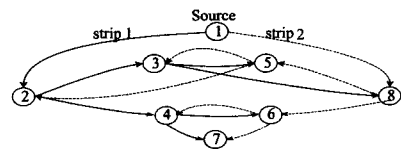


图6 SplitStream的多树路径实例<sup>[8]</sup>

相对基于单树路径结构的数据分发机制,基于多树路径结构的数据分发机制不但可以充分利用系统中每个节点的带宽资源并实现负载均衡,还可以减少节点动态性对系统中其它节点播放质量的影响。但是,由于其仍基于树状分发路径结构,因此很多树状路径结构存在的基本问题,如维护开销大、路径收敛时间长等仍未得到解决。除此之外,此类机制的最大问题在于应用。由于现有方案实现和部署起来都比较困难,至今尚未形成现实可用的系统或原型系统,故该方案在现实环境下的可用性还不明确。

### 3.3 基于网-树路径结构的数据分发机制

对于 P2P 流媒体直播应用而言,路径驱动的数据分发机制面临的最大挑战在于如何在高度动态、异构的 P2P 覆盖网络环境下满足流媒体直播的实时性、连续性业务需求。基于单树和多树路径结构的数据分发机制在静态网络环境下数据分发效率很高,但是在动态环境下,尤其是网络节点扰动频繁的情况下,很容易产生较高的数据分发树维护开销,而且较长的分发树收敛时间也容易导致部分节点发生数据中断或视频质量下降。

针对以上问题,有研究人员提出了基于网-树路径结构的数据分发机制。该机制的基本思路是将随机网络的灵活性与分发树的高效性相结合,在网状的成员管理拓扑上构造数据分发路径。通常,系统首先将用户节点组织为一个随机覆盖网络,然后基于该覆盖网为数据源节点构造一棵或多棵数据分发树,流媒体数据经由这(些)棵数据分发树抵达全网节点。基于网-树路径结构的数据分发机制的优点在于它在充分发挥基于树结构分发路径的数据分发机制优势的同时,可以加快网络在扰动环境下数据分发路径的修复开销和收敛速度。ESM 系统<sup>[3,9,10]</sup>和 Chunkyspread 技术<sup>[11]</sup>是此类机制的典型代表。

ESM 系统<sup>[3,9,10]</sup>是卡内基梅隆大学的 Chu Yang-hua 等设计并实现的一个 P2P 流媒体直播系统。ESM 系统中的控制协议 Narada 实现了一种基于网-树路径的数据分发机制。Narada 协议不但能够自组织地构造高效的覆盖网拓扑,还能够对网络结构进行自我改善,对于网络的动态性、异构性都具有较强的适应能力。该协议的基本思路是将系统的成员管理拓扑与数据分发路径相分离。Narada 通过两个阶段完成网-树结构的数据分发路径的构造:第一阶段,系统构建一棵称为网的富连接树,并且尽量保证该结构具有两个特征:①任意节点间的路径质量要相当于节点间单播路径质量;②网中的每个成员具有有限数量的邻居。第二阶段,基于网状拓扑,系统基于带宽-延迟尺度,采用经典的距离矢量路由算法,为每个数据源构建反向最宽、最短路径的生成树,作为流媒体的数据分发路径。图 7 给出了 Narada 协议在 4 个节点间构造的网状拓扑(数字表示节点间的通信开销)和以 A 为根节点的生成树。采用 Narada 协议,ESM 系统可以为任意数据源以一种完全分布式的方式构建一棵高效的数据分发树。

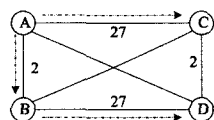


图 7 ESM 中的网状结构实例<sup>[3]</sup>

Chunkyspread 技术<sup>[11]</sup>是康奈尔大学的 Vidhyashanka 等提出的。该技术采用了与 Narada 协议相似的网-树结构的数据分发路径,但在以下两方面与 Narada 协议存在明显差异。一方面,Chunkyspread 采用的是网-多树路径结构,这一点区别于 Narada 采用的网-单树路径结构。通过引入 MDC 技术,Chunkyspread 将流媒体数据划分为多个子流,数据源将子流分发给多个节点,然后这些节点再作为子流数据源在以网-单树结构分发子流,因此整体上使 Chunkyspread 形成一种如图 8 所示的网-多树的路径结构。与 3.2 节中的多树路径结构相类似,Chunkyspread 中提出的网-多树的数据分发路径结构比

ESM 中的网-单树的数据分发路径结构对异构网络具有更好的适应能力,能够更充分地利用用户节点资源。另一方面,Chunkyspread 在分发树的构造方法上也与 Narada 协议有明显差异。Chunkyspread 没有像 Narada 那样采用经典距离矢量路由算法,而是采用了更为灵活的自组织、自改善方式构建数据分发树。Chunkyspread 中,节点间的父子关系由各节点根据局部信息独立确定,无需复杂的全网信息传播机制。为了避免构造低效的分发树,Chunkyspread 还进一步提出了使节点根据负载、延迟等指标对网络拓扑结构进行动态优化的方法。

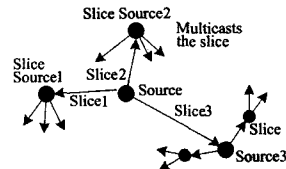


图 8 Chunkyspread 数据分发机制<sup>[11]</sup>

与 3.1 节和 3.2 节中介绍的数据分发机制相比,基于网-树路径结构的数据分发机制在对数据分发路径的优化方法上形成一定突破。该机制可以不局限于在给定网络拓扑上构造最优的分发树,还可以通过动态调整系统的网状拓扑来实现对数据分发路径的进一步优化。此外,基于网-树路径结构的数据分发机制在抗扰动性方面也有所增强。网状拓扑的富连接结构可以使受到影响的节点掌握、交换更多可用于网络连接性恢复的信息,从而加快网络恢复速度,降低扰动影响程度。然而,由于此类技术通常要求覆盖网络采用富连接的网状结构,因此其控制开销要高于传统(多)树结构。

## 4 数据驱动的数据分发机制

数据驱动的数据分发机制是 P2P 网络技术中常用的数据交换技术,其覆盖网拓扑通常为网状结构。虽然这种机制在进行数据交换时也会交换控制信息和流媒体数据,但是其通信协议通常不像路径驱动的数据分发机制那样完全分离开来。与路径驱动的数据分发机制不同,此类数据分发机制认为与其在高度动态的 P2P 网络环境中不断地修复分发路径结构,还不如利用数据的可用性来引导数据流。因此,在此类数据分发机制中,流媒体数据并不是沿预先构造的数据分发路径传输,而是根据节点各自的数据需求传输的。事实上,这种数据分发机制中也存在数据分发路径的概念,只不过这种数据分发路径是节点间在进行数据交换时自然形成的,而且一般不被后续数据单元所复用。对于一个数据单元而言,这种数据分发路径通常也表现为树形结构。

在数据驱动的分发机制中,资源(数据、节点资源等)调度算法是一项最为重要的核心技术,在很大程度上决定了数据的分发模式。资源调度算法的优劣会对 P2P 流媒体直播系统的传输延迟、数据冗余度、控制开销等性能产生重要影响。目前,在 P2P 流媒体直播领域中,随机拉取(Random Pull)算法及其变种是主流的资源调度算法。此类算法在现实世界中得到了广泛的应用,PPStream, pplive 等知名 P2P 流媒体直播系统都是基于此类算法开发的。

基本的随机拉取算法<sup>[12]</sup>是应用于网状覆盖网结构之上的。覆盖网络上的边对应节点之间的邻居关系,一个节点可与多个节点建立邻居关系,邻居节点间通过定期交换 BM 消

息以了解彼此在缓存窗口内的数据缓存情况。在一个数据调度请求周期内,节点首先计算自身缺失而邻居节点可用的数据块集合,然后从该集合中随机选取一些数据块向随机选取的拥有该数据块的邻居节点发起数据拉取请求,尔后被请求者向请求者发送相关数据块。对于该请求周期未能成功获得的数据包,节点将在下一请求周期内继续拉取。

图9给出了以上过程的简单实例。节点P1随机选择节点P2,P3,P4,P5作为自己的邻居节点,并定期与它们交换BM消息。在一个数据请求周期内,节点P1根据自身及邻居节点状态,随机选择向P2和P5发送6号数据块和4号数据包的拉取请求( $P1 \rightarrow P2, R_{p6}$ )和( $P1 \rightarrow P5, R_{p4}$ ),然后节点P5向P1成功发送4号数据包( $P5 \rightarrow P1, p4$ ),而节点P2由于网络拥塞没能给P1成功发送6号数据包( $P2 \rightarrow P1, p6$ )。对于6号数据包,节点P1将在下个请求周期内继续向其邻居发起拉取请求。

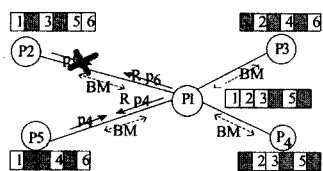


图9 随机拉取算法运行实例

在基本随机拉取算法基础上,通过加入不同的调度策略产生了很多变种算法,这些变种算法的性能在不同程度上都优于基本算法的性能。变种算法采用的调度策略大致可以包括最少优先策略、最新优先策略、最贫瘠优先策略等,其中以最少优先策略最具代表性。

最少优先策略是P2P文件共享系统中应用最为广泛的数据调度策略,最早由Coolstreaming<sup>[4,13-15]</sup>系统引入到P2P流媒体系统中。该策略的基本思想是通过增强局部稀有数据的可用性加快其扩散速度。该算法在数据调度阶段的具体调度过程为:节点首先根据自身缓存状态和邻居缓存状态计算待拉取的数据块集合,同时计算集合中各数据块的局部稀有程度,然后优先向邻居节点拉取稀有数据。相关研究表明,最少优先策略具有良好的数据分发性能<sup>[16]</sup>。

随机拉取算法及其变种的显著优势是具有简单性和有效性。此类算法无需复杂的调度策略就能在大规模动态网络中高效地实现数据分发。而且,在数据分发过程中,发送者根据接收者的需求为其供应数据,冗余数据量少。但是,由于该算法要求节点显式地拉取每个数据块,因此其数据分发过程中的控制开销较大。此外,该算法无法保证系统在带宽资源利用率、启动延迟、播放延迟等方面达到最优或近似最优的性能。

## 5 混合驱动的数据分发机制

路径驱动的数据分发机制可以在静态网络环境下实现低延迟的数据分发,但是在P2P网络这种高动态网络环境下其抗扰动能力差,路径结构不稳定且收敛过程的开销大,时间长;而另一方面,数据驱动的数据分发机制在高动态的P2P网络环境具有较强的抗扰动能力,且在上行带宽利用率和系统吞吐率方面具有近似最优的特性,但是其控制开销高,启动延迟和播放延迟显著。为此,有研究人员提出了混合驱动的数据分发机制,其基本思路是将路径驱动的数据分发机制与

数据驱动的数据分发机制有机地结合起来,有效地发挥二者优势,规避劣势。GridMedia系统<sup>[17-20]</sup>中的混合式推拉协议是此类分发机制的典型代表。

GridMedia系统是由清华大学的张萌等研制的,采用混合式推拉协议实现数据分发。该协议的基本思想是将数据驱动分发机制所形成的部分分发路径复用于推送未来产生的数据包,以降低系统的控制开销和传输延迟。在GridMedia中,媒体流被均分为 $n$ 个子流,每个子流由数据包序号模 $n$ 余数相同的数据包构成,连续的 $g$ 个数据包组被划归为一个数据包段。在形成可复用的推送路径前,节点与邻居节点交换BM消息以了解相互的数据状态,然后在数据请求阶段根据邻居节点持有数据的情况向它们请求自身所需数据包。在一个数据包段中,0号数据包组被定义为路径复用的标志。一旦节点从某个邻居节点处成功请求了0号数据包组中的数据包,那么该节点就会向这个邻居节点发送一个子流预定消息,表明其希望该子流上的后续数据包将沿该路径推送给该节点。在子流的推送路径形成后,子流的后续数据应该由父节点主动推送给子节点。但是由于存在丢包、带宽受限等问题,因此该子流上的某些后续包可能无法及时到达子节点。当节点发现这种情况后,它就会向拥有该数据的邻居节点主动拉取相关数据。若某一推送子流上存在未能及时到达的位于0号数据包组的数据包,则很可能会在从某邻居处成功拉取该数据包后切换子流的推送路径。图10描述了Gridmedia系统中一个子流的数据分发情况,其中 $S \rightarrow P1$ 和 $S \rightarrow P4$ 之间采用路径驱动的模式分发数据,而其它节点间则采用数据驱动的模式分发数据。

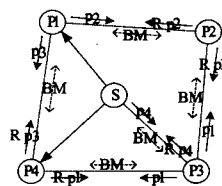


图10 GridMedia中的数据分发实例

以GridMedia中混合式推拉协议为代表的混合驱动分发机制没有显式的分发路径构造与维护过程,又能在部分节点间实现数据推送,它既保留了数据驱动分发机制的抗扰动特性,又保留了路径驱动分发机制的低延迟特性,同时将控制开销维持在较低水平。然而,此类系统中可被后续数据复用的数据分发路径是在数据分发过程中形成的(极端情况下路径结构类似于3.3节中的网-多树结构),形成方式比较随机,缺乏优化,不可避免地会生成一些不好的分发路径结构。因而,为了进一步改善系统的性能,此类系统目前面临的挑战是如何形成优化的数据分发路径结构。

结束语 技术的发展与人们沟通需求的提高使得P2P流媒体直播业务成为当前互联网不可或缺的一种业务形式。在P2P流媒体直播系统中,影响用户体验的根本因素在于数据分发机制。本文对目前应用于P2P流媒体直播系统的各种数据分发机制进行了分类研究,并总结了各自的优势和劣势。通过对研究现状的分析,不难发现尽管现有的某些数据分发机制已经能够初步支持P2P流媒体直播系统的大规模应用,但是距离为用户提供类似传统电视的用户体验仍有很大距离。概括而言,目前面向P2P流媒体直播的数据分

发机制仍然存在以下问题,有待进一步研究:

#### • 数据分发过程中的 QoS 保障

流媒体直播业务是一种对数据的延迟、抖动及丢包都十分敏感的业务,因此从数据分发机制方面对业务的 QoS 进行保障是十分必要的。从现有几种 P2P 流媒体系统的应用现状来看,其启动延迟及播放延迟普遍较长,特别是在扰动情况下容易出现屏幕冻结的现象,严重影响了用户的服务体验。解决此类问题的根本在于围绕 QoS 保障目标,结合 P2P 覆盖网特点及流媒体数据编码特点,对数据分发机制进行系统的改进和优化。

#### • 节点行为及数据分发模式研究

在大规模 P2P 流媒体系统中,节点的数据分发行为和系统的数据分发模式具有密切的关联性。系统采用的数据分发机制在微观上决定了用户节点个体的行为,而系统的数据分发模式则是用户节点群体行为的一种宏观表现。数据分发模式的优劣直接反映了数据分发机制的有效性。因此,深入探究现有机制形成的数据分发模式以及理想数据分发模式,对于优化和改进现有数据分发机制具有十分重要的价值。

#### • 覆盖网拓扑调整与分发路径优化

在 P2P 流媒体系统中,数据分发路径通常构建或形成于 P2P 覆盖网之上,因此覆盖网拓扑对于数据分发路径的性能会产生较大影响。影响数据分发性能的因素很多,其中最为重要的因素就是覆盖网络与底层物理网络的拓扑一致性。建立在 IP 物理网络之上的 P2P 覆盖网拓扑与底层 IP 网络拓扑可能存在很大差异,上层网络的一跳需要对应物理网络上的几跳甚至十几跳。这种上下层网络拓扑的不一致性不但会大大地增加数据在系统中的传输延迟,而且会给底层物理网络带来极大的传输开销。影响数据分发性能的另一个重要因素是节点的异构性。现实的 P2P 网络是由异构节点组成的,节点之间在网络带宽、处理能力方面都存在较大差异。覆盖网中这些异构节点的组织方式,对于构造高性能的数据分发路径并充分利用节点资源将产生较大影响。除以上因素外,覆盖网结构、节点稳定性、节点安全性等也都会对数据分发性能产生一定影响。因此,依据以上因素对覆盖网拓扑进行调整和优化有利于形成更好的数据分发路径,实现更好的数据分发性能。

#### • 激励机制研究

P2P 流媒体系统的良好伸缩性根本来源于用户节点的资源贡献。离开了用户贡献的必要资源,无论多么优秀的数据分发机制都无法为用户提供好的服务体验。然而,在实际应用中,许多用户倾向于以“free rider”的形式获取服务,即只索取服务,不贡献资源。这种行为对于充分发挥 P2P 技术的优势十分不利。因此,如何建立一种有效的激励机制,并将该机制与数据分发机制有机结合,使得参与业务的用户节点更多地贡献自身的资源,对于提高数据的数据分发性能也是十分重要的。

## 参 考 文 献

- [1] Jiang X, et al. GnuStream: a P2P Media Streaming System Prototype[C]// the International Conference on Multimedia and Expo. 2003
- [2] Guo Y, et al. P2Cast: peer-to-peer patching scheme for VoD service[C]// the 12th international conference on World Wide Web. 2003
- [3] Rao Y-H C, Rao S G S S, Zhang H. A case for end system multicast[J]. IEEE Journal on Selected Areas in Communications, 2002, 20(8)
- [4] Zhang X, et al. CoolStreaming/DONet: A Data-driven Overlay Network for Peer-to-Peer Live Media Streaming[C]// IEEE INFOCOM. 2005
- [5] Banerjee S, Bhattacharjee B, Kommareddy C. Scalable Application Layer Multicast[J]. ACM SIGCOMM Computer Communication Review, 2002, 32(4): 205-217
- [6] Tran D A, Hua K A, Do T T. ZIGZAG: An Efficient Peer-to-Peer Scheme for Media Streaming[C]// INFOCOM 2003. 2003: 1283-1292
- [7] Padmanabhan V, Sripanidkulchai K. The Case for Cooperative Networking[C]// the First International Workshop on Peer-to-Peer Systems (IPTPS). Cambridge, MA, USA, 2002
- [8] Castro M, et al. SplitStream: High-Bandwidth Multicast in Cooperative Environments[C]// SOSP'03. 2003
- [9] Chu Y -H, Rao S G, Zhang H. A Case for End System Multicast [C]// Proceedings of ACM SIGMETRICS. Santa Clara, CA, 2000
- [10] Sripanidkulchai K, et al. The Feasibility of Supporting Large-scale Live Streaming Applications with Dynamic Application EndPoints[C]// SIGCOMM'04. 2004
- [11] Venkataraman V, Francis P, Calandrino J. Chunkyspread: Multi-tree Unstructured Peer-to-Peer Multicast [C]// IPTPS'06. 2006
- [12] Pai V, et al. Chainsaw: Eliminating Trees from Overlay Multicast[C]// IPTPS. 2005
- [13] Xie S, et al, Coolstreaming: Design, Theory, and Practice[J]. IEEE Transactions on Multimedia, 2007, 9(8): 1661-1671
- [14] Zhang X, Liu J, Li B. On Large Scale Peer-To-Peer Live Video Distribution: CoolStreaming and Its Preliminary Experimental Results[C]// MMSP 2005. 2005
- [15] Li B, et al. An Empirical Study of the Coolstreaming+ System [J]. IEEE Journal on Selected Areas in Communications, 2007, 25(9): 1627-1639
- [16] Kostic D, et al. Maintaining high bandwidth under dynamic network condition [C]// USENIX Annual Technical Conference (USENIX 2005). 2005
- [17] Zhang M, et al. A Peer-to-Peer Network for Live Media Streaming-Using a Push-Pull Approach[C]// ACM Multimedia 2005. 2005
- [18] 张萌. 对等网络流媒体直播调度策略研究[D]. 北京:清华大学, 2008
- [19] Zhang M, Zhang Q, Yang S. Understanding the Power of Pull-based Streaming Protocol: Can We Do Better? [J]. IEEE Journal on Selected Areas in Communications, 2007, 25(8)
- [20] Zhang M, et al. Large-scale Live Media Streaming over Peer-to-Peer Networks through Global Internet[C]// ACM International Workshop on Advances in Peer-to-Peer Multimedia Streaming. 2005