P2P 技术在 CDN 网络中的应用研究

Application Research of P2P Technology Using in CDN Network

杨晓波 （浙江财经学院信息学院 杭州 310012）

摘要：本文主要探讨了 P2P 技术及其在 CDN 网络中的实际应用。首先，对 P2P 的关键技术进行了分析，指出要实现 P2P 应用，需重点解决节点加入与成员管理，数据交互与内容调度等核心问题；接着将 P2P 技术引入到电信 CDN 网络之中，通过对核心技术的研究，提出 P2P 与 CDN 网络的融合方案；最后，采用三种融合测试方法，验证 P2P 技术与 CDN 网络融合的可行性。

关键词：P2P 技术 CDN 网络 融合应用

对等网络（Peer – to – Peer，简称 P2P）技术是目前国际计算机网络技术领域研究的一个热点。该技术的雏形产生于 20 世纪 70 年代，典型代表是 UseNet 和 FidoNet[1]，两者皆为分布式的信息交换系统，真正 P2P 技术的大规模应用起源于文件交换软件 Napster[2]，从某种意义上说，P2P 计算可以说是向传统互联网技术的回归，体现了因特网的本质。

CDN 内容分发网络（Content Delivery Network）将网站的内容或媒体发布到最接近用户的网络“边缘”，当用户访问时系统自动无缝地把用户重定向到边缘服务器，从而减轻中心服务器和骨干网络的压力，提升流媒体或网站的性能。本文将从 P2P 的关键技术研究入手，探讨 P2P 技术与 CDN 网络的融合。

1 P2P 的关键技术及其实现

目前，P2P 主要应用于直播、下载、点播等领域，尤其是直播业务近年来发展迅猛，所以本文将重点分析主流 P2P 直播系统的实现模式和设计思路。

早期基于应用层的组播模式，大多数采用树型结构进行数据传输。这来源于 IP 组播模型，系统的关键在于构建和维护一个高效的分发树，这种模式在一个固定的 IP 节点模型下工作，存在命中率低、结构不稳定等问题。

随着研究的逐渐深入，出现了并不依赖于关键节点，而是建立在一群自组织的自治节点上，DONet 就是一种典型的重叠网流传输协议模型[3]。下面以 DONet 模型为基础，分析 P2P 的关键技术。

1.1 节点的加入和成员关系管理

在 DONet 协议模型中，每一个 DONet 节点都有一个唯一的标识符，同时维护一个缓存（mCache），该缓存包含 DONet 网络上一个局部的活跃节点列表。为了适应重叠网的动态性，同时建立和更新 mCache，每个节点会定期发布其成员消息以通告大家它的存在. 每个消息是一个四元组，表示如下：

< seq num, id, num partner, time to live >

这里 seq num 是消息的序列号，id 是该节点的标识符，num partner 是其当前所有拥有的合作伙伴数量，time to live 记录了该消息还剩下的有效时间。

当收到一个新的消息序列号 seq num 时，DONet 节点会更新其 mCache 中有关发送者 id 的条目，如果该节点 id 不存在其 mCache 中，则会新建一个条目，这个条目是一个五元组，表示如下：

< seq num , Id, num partner, time to live, last update time >

前四个元素从收到的成员消息中 copy 而来，第五个元素是最后一次该条目更新的本地时间。

另外的两个条件也会引发 mCache 条目的更新：(1) 当一个成员消息通过 Gossip 协议的方式被转发至另一个节点时；(2) 该节点是一个代理节点，而其条目被包含在候选的合作伙伴列表里。无论哪一种情况，time to live 会减少一个值，该值相当于：当前的本地时间与最后更新时间之差；当 time to live 值小于或等于

0 时，该条目会被删除，并不会被转发或者被包含到合作伙伴节点列表里；同时，num partner 值会减 1。

1.2 Buffer MAP 的表现形式和数据交换

Donet 网络中的伙伴关系和数据的传输方向，往往都是不固定的。在数据交换过程中，一个视频流被切分成统一长度的片，每个节点的缓冲中，这些片的可用性被描述成一个 Buffer Map(BM)。每个节点不断与其伙伴交换这些 BM，并调度决定从哪个伙伴中获取本节点所需要的哪个流数据片，伙伴关系如图 1 所示。

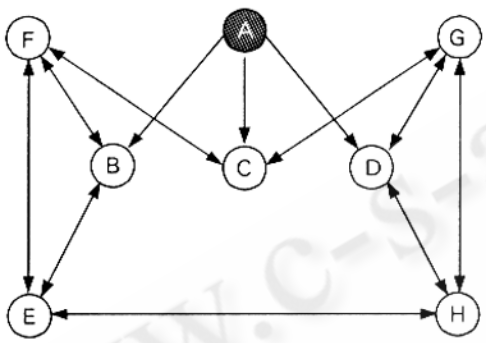

图 1　DONet 网络中伙伴关系的描述(A 是源节点)

由于 DONet 的目标是实时流媒体，因此 DONet 所有节点的播放过程要求基本同步。分析结果显示平均传输时延在 DONet 内是有限的，一般建议节点间的时延不超过 1 分钟。假设每个片数据包含了 1 秒钟的视频，一个 120 片的滑动窗口就能有效地描述一个节点的 Buffer Map，因为一个合作伙伴不会对该滑动窗口之外的数据感兴趣。按照这样的设计，在模型中，我们使用 120 Bit 来记录一个 BM，比特位是 1 代表该片的数据可用，如果是 0 则相反。滑动窗口中的第一个片的序列号由另一个双字节来描述，这样可以回朔到一个非常长的视频节目。

1.3 调度算法

伙伴节点之间如何传输互相需要的数据片，需要通过调度算法实现。在一个同质化静态网络里，一个简单的 Round－robin 调度就能很好的工作[4]；但在一个动态的不同类的网络里，需要一个更智能的调度算法，该调度算法会存在两个约束：即每个数据片的时效性，以及每个伙伴节点流媒体带宽的不确定性。如果第一个约束不能有效地满足，那也至少要把超过时效的数据片数量控制在最小[5,6]。由于很难找到一个最优解决办法，在 DONet 中采用了一种简单的启发式快速响应时间机制。

启发式快速响应算法首先计算每一个数据片的潜在提供者数量，如果某个数据片的潜在提供者数量很少，那么很容易超过其时效性约束，调度算法决定该数据片的提供者，从第一个开始，并不断增加，在这些可选的潜在提供者中，我们尽量选择那些具有较大带宽并拥有足够剩余时间的提供者。

由于源节点服务器仅作为一个数据发送者，它始终含有所有的数据片。通过采用自适应的调度算法，它不会被来自于合作伙伴的请求而导致过载。如果有需要，它也能通过广告 BM 的方式积极的控制其负载。例如：假定有个 M 个合作伙伴，源节点可以将它的 BM 传递给 第 K 个合作伙伴，如下：

$$BM[id_{origin\_node},i]=\begin{cases}0, \text{if } i \bmod M \neq k\\ 1, \text{if } i \bmod M = k\end{cases} \quad (1)$$

也就是说，仅仅那些 (i mod M) 的合作伙伴能够从源节点请求数据片 i，而剩下的数据片需要从其他的合作伙伴得到。

2 P2P 与 CDN 网络的融合

CDN 与 P2P 组成的混合流媒体系统可以有效地结合两者的优点，利用 P2P 技术可以有效减少系统所需代理服务器的数量，增大系统的容量；同时由于是在一个较小的自治系统范围内，P2P 网络性能也会有很大提高。另外，由于是建立在传统的 CDN 系统基础之上，在骨干网层次保留了原有的 CDN 系统架构和功能，在边缘节点引入了 P2P 技术来进行文件及流媒体的共享，实现了 P2P 技术与 CDN 传输网的结合。系统结构如图 2 所示。

系统结构的整体设计结合因特网的结构特点，原始服务器与分布在各自治系统内的代理通过骨干网互连组成流媒体 CDN，各因特网自治系统内的代理与客户机组成与其他自治系统相互独立的 P2P 流媒体网络。

从图 2 可以看出，P2P 平台与 CDN 的流媒体服务器结点共同组成自治域，CDN 网络的每个流媒体服务器作为一个超级结点，客户在发出访问请求时，首先连接有相关内容的其他客户结点，这样便可降低流媒体服务器的压力，同时结合 IP 地址域的控制算法，将客户的连接范围控制在一定区域内，便可降低骨干网的

带宽消耗。

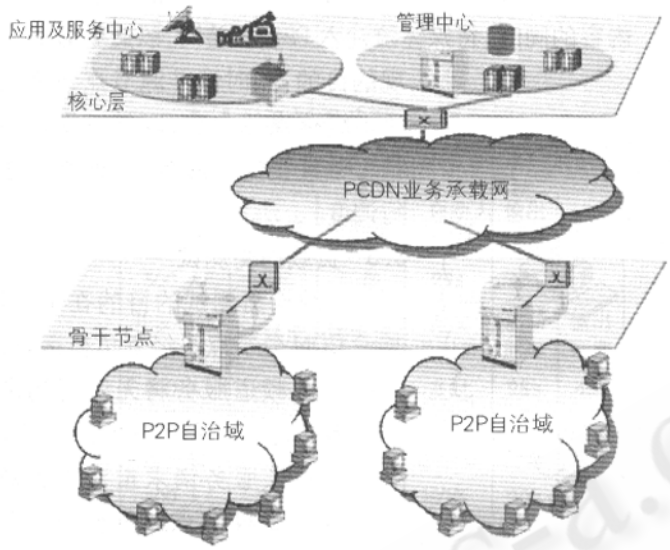

图 2　结合 CDN 与 P2P 技术的流媒体系统结构图

为了实现 P2P 与 CDN 网络的有效融合，需要解决以下关键问题：

2.1 代理服务器的选择

原始服务器在接到一个合法客户的请求后，需要为该客户选择一个代理服务器，选择标准是客户与代理服务器的距离，即代理服务器应尽量靠近客户机。缓存代理服务器一般部署于因特网的某一个自治系统内，客户机距离本自治系统的代理服务器显然要近于其他自治系统的代理服务器，而同一个自治系统的机器有相同的 IP 地址前缀。因此，在系统中，原始服务器在为客户机选择代理服务器时依照地址前缀最长匹配原则选择，当同一个自治系统中有多个代理时，即客户地址与多个代理的地址匹配时，应综合考虑各代理的负载情况以及各代理对此内容的缓存情况，这时应优先选择已缓存了该内容的代理，其次考虑选择负载较轻的代理。

2.2 失效服务节点的替换查找

充当服务器的客户机离开网络的时间是随机的，也是频繁的，为了使正在接受服务的客户不受影响或受到的影响尽量小，设计一个快速有效的服务节点的替代算法至关重要。由于 IP 网络的路由遵从最长路由匹配的原则，所以，根据客户机的 IP 地址可以比较一台机器与分别与另外两台机器间距离的远近（跳数）[7,8]，基于这一考虑，下面给出一种新的替换节点的查找策略。

（1）正在接受流媒体服务的客户机每隔一定时间（比如一分钟）就从代理服务器下载一次能够提供本机正在播放的流媒体服务的客户机列表。如果正在为本机提供服务的客户机不能及时提供服务时，采用步骤（2）。

（2）按与本机 IP 地址最大匹配的原则从列表中选择替换者，并通知代理服务器原服务提供者失效。

（3）验证新替代者的可用性，若不可用，重复（2），若可用，则请求所需流媒体内容。

这一过程所造成的流媒体内容的间断可通过增加播放缓存屏蔽掉，一旦本机与新的服务提供者取得联系，便以最大速度下载流媒体内容，直到播放缓存中的流媒体内容恢复到正常水平；如果一个服务提供者的速度不足，还可以考虑同时从两个或多个服务提供者同时下载同一流媒体内容的不同部分。

3 P2P 与 CDN 网络的融合测试

为了验证 P2P 技术与 CDN 网络的融合效果，需要对其进行功能性测试，测试环境如下：

硬件环境：Intel（R） Xeon（TM） CPU2.4G * 2 ，1G 内存，36G 硬盘

软件环境：MS Windows Server 2003 操作系统，P2P 平台软件

其他要求：视频格式采用 Windows Media Video V9 编码，352 × 288 分辨率，码流采用 450kbps 和 700kbps 两种；音频格式采用 Windows Media Audio V9 编码格式，流设定为 20Kbps / 单声道。

测试内容主要分为三大部分，下面对其进行一一介绍：

3.1 P2P 与 CDN 双网融合测试

首先进行 P2P 应用网与 CDN 网络的融合测试，测试工作示意图如图 3 所示。

从图 3 可以看出，P2P 覆盖网直接从有源 CDN 分节点索取服务内容。每一个新加入的 P2P 用户首先通过频道目录将其自身 Track 定向到 CDN 边缘节点，获取初始的代理用户节点信息；再通过连接代理节点用户，进一步获取伙伴节点信息，并快速形成稳定的 P2P

覆盖网;同时将边缘 CDN 节点的媒体流注入 P2P 覆盖网之中,保证媒体流在 P2P 覆盖网中进行分布式数据分发,最终实现 CDN 服务与 P2P 应用的融合。

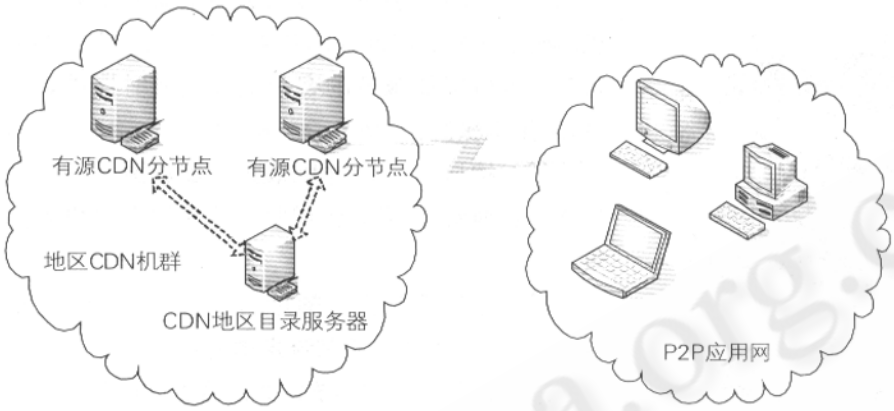

图 3　P2P 与 CDN 融合测试示意图

此方案的测试目的主要是验证 P2P 与 CDN 网络结合的功能可行性及接口可行性。

3.2　内容调度测试

内容调度测试的目的在于:当本地区的 CDN 机群中出现内容分布不均衡时,为保证 CDN 对 P2P 覆盖网的全面有效的内容服务,须进行 CDN 机群间的内容调度。具体的测试示意图如图 4 所示。

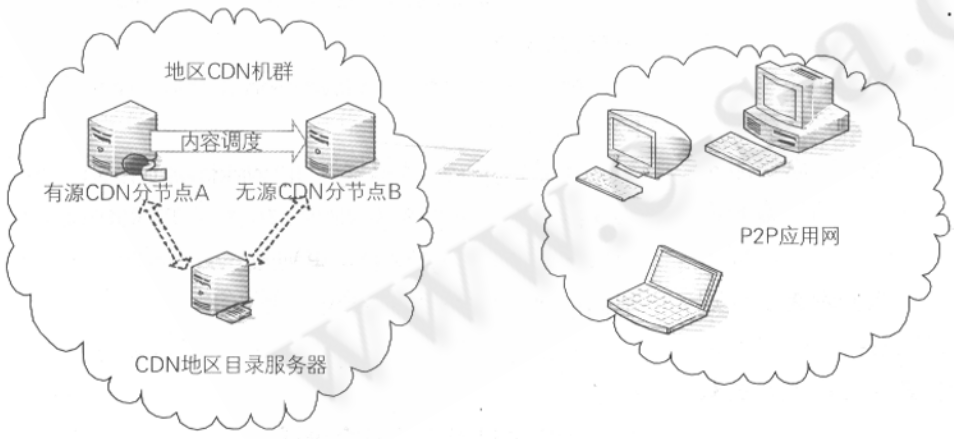

图 4　内容调度测试示意图

从图 4 可知,内容调度的测试方法为:首先为保证实现内容调度,有源 CDN 分节点必须预保留带宽资源;当 P2P 覆盖网用户被目录定向到一个无源 CDN 分节点(该节点无用户请求的目标服务)时,该节点将向有源节点发出请求,有源节点根据请求内容启用其自身预留网络资源(包括 CPU,带宽等),将请求源反馈给无源节点;此时无源节点成为准有源节点,在获得调度内容的同时,将媒体流注入 P2P 覆盖网,以达到服务实现的目的;同时该准源节点还将在本地缓存媒体内容,以便下次用户请求时直接提供服务,而无需再进行调度。

3.3　负载均衡测试

负载均衡测试的目的在于:当某一 CDN 分节点已经满负荷运行,此时新的频道服务将无法得到满足。因此,为了保证服务的实现,必须要将该服务请求重定向到轻负荷地区 CDN 节点。负载均衡测试的示意图如图 5 所示。

从图 5 可以看出,采用负载均衡测试方法,测试的主要流程是:CDN 目录服务器将节目内容指向服务器 A,服务器 A 向一级代理种子节点提供内容服务,随着种子节点的不断增加,服务器 A 很快达到"满负荷";此时发起新的服务请求时,服务器 A 会将该请求首先提交到目录服务器,再由目录服务器分配一个轻负荷节点,将用户 Track 到轻负载节点 B,由节点 B 担负服务响应任务,往 P2P 覆盖网注入媒体流。

3.4　测试结果分析

采用以上三种方法进行 P2P 与 CDN 网络的融合测试,测试结果见表 1。

测试人数取在网络稳定情况下的一个小时内的人数值,测试视频码率 450kbps,信息源采用卫星信号,输出带宽采用 4M。在每组连续播放测试 1 小时的情况下,最多只出现 1 次停顿缓冲的现象,因此可认为测试视频播放的流畅度完全可以达到主观满意的标准。

450kbps 码率的视频在单倍播放窗口模式下可以达到清晰的观看效果。在全屏模式下如果以静态场景居多,则基本可以满足清晰的视频体验。测试结果中,启动时延基本在 30 秒左右。这一启动时延在一定程

度上会影响服务满意度。目前已有基于 Gossip 的改进型用户感知算法,基本可以缩短一半启动时间,可以考虑进一步减少启动时延。

表 1 视频测试结果表

测试项目	P2P 结合 CDN 网络测试					
双网融合	并发稳定用户数	流畅度	清晰度	启动时延	输出带宽	
	500 人	流畅	一般	29s	4M	
内容调度	并发稳定用户数	流畅度	清晰度	调度及时性	启动时延	输出带宽
	500 人	流畅	一般	及时	28s	4M
负载均衡	并发稳定用户数	流畅度	清晰度	重定向命中率	启动时延	输出带宽
	400 人	流畅	一般	100%	30s	4M
备注:	此测试表格结果以 450kbps 为主,不含 700kbps 码率的视频测试。					

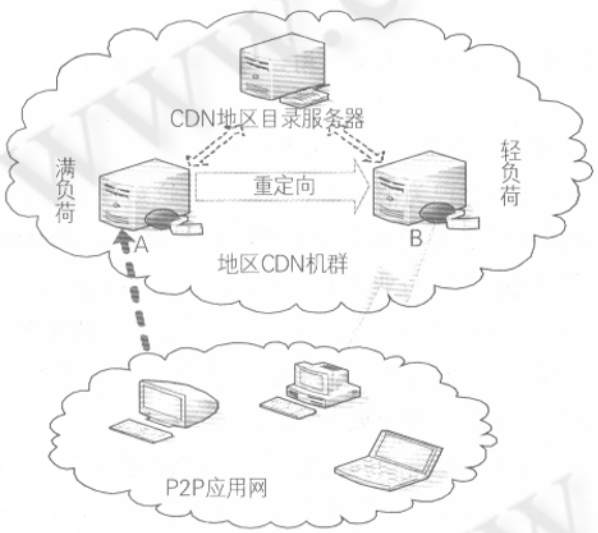

图 5 负载均衡测试示意图

另外,随着电信网络的扩容,用户带宽的不断增加,700kbps 码流的视频将会满足用户高码率视频体验服务,提高用户的满意度和忠诚度。

整个测试结果说明 P2P 与 CDN 网络能够实现良好衔接,三大功能性测试都已通过验证,后期只需在启动延时上做进一步改进。

4 结束语

P2P 网络中的节点本身往往是计算能力相差较大的异构节点,局部性能较差的节点将会导致整体网络性能的恶化,CDN 技术在很大程度上可以弥补单纯 P2P 构架的不足之处,本文提出了 CDN 与 P2P 技术的混合网络构架是一种较为理想的解决方案,并得出以下结论:

(1) 在现有 CDN 网上加载 P2P 技术,可以大幅减少硬件投入和运营成本,同时用户容量将增大 3 个数量级,且可实现平滑的软升级。

(2) 在第三方 P2P 平台基础上进行了 P2P 与 CDN 网络的融合测试,证明了该方案的可行性和有效性,为电信基于 CDN 平台的服务及其他各种互联网的运营提供了一套切实可行高效的新型技术方案。

参考文献

1 Joseph A. K., Bradley N. M., David M. etc, GroupLens: Applying Collaborative Filtering to Usenet News, Communications of the ACM, March 1997/ Vol 40, No 3.

2 Barkai, D., P2P Computing. Intel Press, Santa Clara, CA 2002.

3 Ganesh A. J. Kermarrec A. M. and Massoulie L. Peer – to – peer membership management for gossip – based protocols, IEEE Transactions on Computers, 2003,52(2):60 –65.

4 Chao H. J. Saturn: A terabit packet switch using dual Round – Robin. IEEE Communication Magazine, 2000, 38(12):78 –84.

5 Li Y, Panwar S., and Chao H. J., On the performance of a dual Round – Robin switch. In: Ammar M, ed. Proc. of the IEEE INFOCOM. Anchorage: IEEE Communications Society, 2001. 1688 –1697.

6 Li Y, Panwar S., and Chao H. J., The dual Round – Robin matching switch with exhaustive service. In: Gunner C, ed. Proc. of the IEEE Workshop on High Performance Switching and Routing. Kobe: IEEE Communications Society, 2002. 58 –63.

7 Cheung, G., McCanne, S., Optimal routing table design for IP address lookups under memory constraints. In: Ephremides, A., Tripathi, S., eds. Proceedings of the IEEE INFOCOM. San Francisco: IEEE Computer Society Press, 1999. 1437 –1444.