

基于 P2P 的视频点播系统模型及算法研究

阳天保, 张修如, 贾丽会

(中南大学 信息科学与工程学院, 湖南 长沙 410075)

摘 要: 服务器带宽是 Internet 视频点播系统的瓶颈问题。文中设计了基于 P2P 的 VoD 系统模型, 讨论了以该模型为基础的 BTBM-Tree 建立、维护的算法思想。对整个系统进行了性能分析, 表明系统具有较好的稳定性、扩展性和延迟性, 能较好地解决网络带宽瓶颈。

关键词: 对等网络; 视频点播; 系统模型; BTBM-Tree

中图分类号: TP393

文献标识码: A

文章编号: 1673-629X(2006)10-0045-04

Research on Video-on-Demand System Model and Algorithm Based on P2P

YANG Tian-bao, ZHANG Xiu-ru, JIA Li-hui

(College of Information Science and Engineering, Central South University, Changsha 410075, China)

Abstract: The bandwidth of server is the bottleneck problem of video on demand service over the Internet. In this paper, design a VoD system model based on P2P and discuss a new algorithm based on the model, called BTBM-Tree algorithm. At last, the analysis of the whole system shows that it has enough stability, scalability and delay performance to solve the bottleneck problem.

Key words: peer-to-peer; video-on-demand; system model; BTBM-Tree

0 引言

随着计算机网络和通信技术的飞速发展, 基于网络的应用日益广泛, 视频点播 (Video-on-Demand, VoD) 系统正是一种基于流媒体技术而实现的网络多媒体应用典范。流媒体实质是一种多媒体文件, 其在网络传输的过程中应用了流技术, 也就是把完整的影像和声音数据经过压缩处理后保存在网站服务器上, 用户可以边下载边获取信息。因此, 在视频点播系统中, 人们并不需要下载完整整个视频文件, 称之为“Open-after-Downloading”, 而是一边下载一边播放, 称之为“Open-while-Downloading”^[1]。目前在国际互联网上使用较多的流式视频格式主要有 3 种: RealNetworks 公司的 RealMedia, Apple 计算机公司的 QuickTime, Microsoft 公司的 Advanced Streaming Format (ASF, 高级流格式)。

但是, 视频点播服务对流量带宽的资源需求很大, 而现有的大多数视频点播仍旧是采用服务器到客户端的单播方式传输, 当点播的用户量很大时, 服务器负载和骨干网络的带宽便成为视频点播的瓶颈。

许多研究都提出了相应的解决办法。IP 组播技术实

现了 Internet 上高效的一对多通信, 提高了系统的可扩展性。此外, 在此基础上提出的补丁 (Patching)^[2]、周期广播 (Periodic Broadcast)^[3] 以及流合并 (Streaming Merge)^[4] 等技术也极大地减少了服务器带宽的消耗。然而, 由于安全性、异构性问题导致网络的复杂化, 并且还存在许多技术难题, 比如可靠性组播和拥塞控制等, IP 组播技术并没有得到广泛的应用, Internet 中多数 ISP 都不支持 IP 组播。另一种方案是在网络边缘部署代理缓存 (Proxy Caching)^[5] 或内容分发网络 (CDN, Content Delivery Networks)^[6], 媒体服务器将媒体内容存放在代理缓存或 CDN 服务器上, 客户请求媒体服务器时, 可从代理缓存或 CDN 服务器获得服务, 而不必消耗服务器的资源。但这种方案花费成本太高, 且只是部分地解决了可扩展性问题。

1 相关研究

对等网络技术, 也称 P2P 技术, 是当前研究的热点之一, 其本质是一种网络结构思想。它与目前占主导地位的 C/S 结构的区别在于: 整个网络中的节点是平等的, 不存在中心节点。P2P 技术使得用户能够直接进行文件共享、分布式计算、存储、协作, 充分挖掘出大量的计算机资源, 包括 CPU 时间、存储空间等。文件共享系统 Gnutella 和 Napster, 还有当前很流行的下载软件 BT 和 eMule 等均采用了 P2P 技术。

对于实时流媒体系统而言, 如何利用 P2P 技术的优

收稿日期: 2006-01-09

作者简介: 阳天保 (1980-), 男, 广西桂林人, 硕士研究生, 研究方向为计算机网络通信、多媒体技术; 张修如, 副教授, 研究方向为多媒体通信、信息系统、图形图像处理。

势又能满足实时性要求,国内外学者做了许多相关工作。

方炜等人提出了应用层三层软件体系结构^[7],并将节点信息控制在网络抽象层(Network Abstract Layer),采用 BM TREE 算法对节点树进行构造和维护,并考虑了不同节点的性能差异问题,整个系统具有良好的稳定性和扩展性。

Guo Yang 等人提出一种补丁^[8]技术方案,将视频流分成基本流(Base Stream)和补丁流(Patching Stream),节点按照某个时间阈值分成组,先加入的节点向后加入的节点转发视频流,从而形成了一棵应用层多播树。服务器同时向所有节点传送基本流,树中的节点可以选择某些节点作为自己补丁流的提供者,充分利用了节点间的时间特性。

M. Hefeeda 等人提出一种 Promise^[9]机制。该机制利用群播(CollectCast)方式,以接收节点为树根,利用网络拓扑图(network tomography)技术推测网络的拓扑结构,并将估测的结构进行优化,然后根据优化的拓扑图,采用特定算法选出最佳的发送节点组和备选组。它解决了如何根据动态变化的网络状况以及网络拓扑,选取最好的服务提供者的问题,同时考虑了 Peer 能力的异构性,提出了如何在各个服务提供者分配传输的数据以及传输速率的算法。针对当前视频点播的实际,结合以前的研究成果,文中设计了一种新的基于 P2P 的视频点播系统模型,以此为基础讨论了 BTBM-Tree(Based Time & Bandwith Multi-Tree)算法的思想,并对系统和算法做了分析。

2 系统模型

2.1 系统模型结构图

根据现有 C/S 模式的 VoD 系统特点构建了一个融合了 P2P 思想的视频点播系统模型。如图 1 所示。

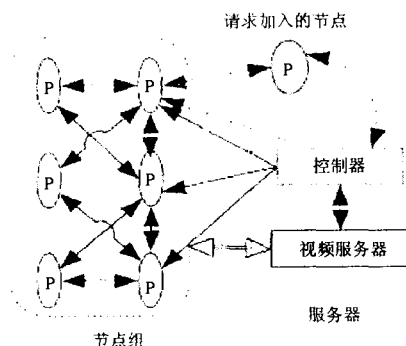


图 1 系统模型结构图

在说明系统工作之前,先构建几个表:组节点表,层节点表,子节点表。如图 2 所示。

组节点表									
节点号	组号	组长	入口带宽	出口带宽	请求时间	处理能力	层号	父节点	
IP 地址									

层节点表			子节点表	
节点号	层号	IP 地址	节点号	IP 地址

图 2 节点表

以此模型为基础,构建了节点树组织结构图(如图 3 所示)。

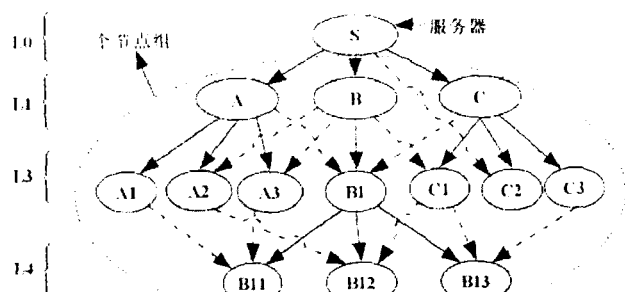


图 3 节点树组织结构图

图中,实线箭头表示父子节点及流的流向,虚线箭头表示接收节点的其他流发送节点。每个节点的入度代表了流来源节点数目,出度则代表流转发节点数目。

2.2 系统工作原理

节点 P 提出视频点播请求,控制服务器为节点分配唯一的节点号,并根据节点提交的节点信息表,按照时间阈值 T 确定节点是加入一个已经存在的组还是新建一个组。若组已经存在,则将请求信息转发给相应组的组长,让组长和 P 直接通信(如图 1 虚线所示)。若组不存在,则新建一组,再由服务器直接向其发送视频流,并将该节点放在第一层。下层节点向上层多个节点请求视频流,实现点对点传送,从而缓解了服务器和骨干网络的压力。当节点的需求不能得到满足时,可以向服务器直接请求视频资源。底层通信协议,控制流采取 TCP 协议,数据流采取 UDP 协议,充分满足应用层的实时性需求。

2.3 各角色节点的职权划分

●服务器:服务器创建第一层节点,同时向它们发送整个视频文件流,并按照节点性能分别选出合适的节点作为组长、副组长和第一层备选节点集合。当收到节点的请求信息时,为每个请求节点创建唯一的节点号,转发节点请求信息到相应组,并维护第一层节点信息。

●组长:与副组长共同维护同一组中的所有节点信息,负责响应节点加入、离开的请求,向自己的子节点发送视频流。

●副组长:与组长同步维护组内节点信息,接收节点状态包信息。当组长离开时,执行组长职权,并向服务器报告。

●组员:接收或者转发视频流,向副组长发送状态包,维护同一层的层节点表和子节点表。

3 BTBM-Tree 算法

3.1 算法描述

定义 BTBM-Tree 算法描述如下(树层次结构见图 3):

- (1) 第 0 层节点为服务器节点。
- (2) 第 1 层节点由服务器根据时间阈值划分成组,并

指定组长、副组长。该层节点直接接收服务器的视频流,若服务器不能满足需要,则可以向组内同层的其他节点发送请求。同一组第 1 层节点的下层节点与第 1 层共同组成一组(虚线椭圆所框部分)。

(3) 第 L 层($L > 1$)的节点,向 $L-1$ 层的节点发送流请求,选出最优的活动组和备选组,在 $L-1$ 层的节点不能满足需求的情况下,可以向 $L-2$ 层直至 0 层节点发送视频流请求。

3.2 BTBM-Tree 维护

(1) 节点加入。

所有要加入的节点 P ,先向服务器 S 发出请求,再经 S 认证,或者新建组,或者转发请求,直到找到合适的组。算法描述如下:

1. 节点 P 向服务器 S 发送视频请求;

2. 服务器响应请求:

2.1 若存在时间阈值 T 内的组,则服务器 S 将信息转交组长响应;

2.1.1 组长遍历节点表,寻找合适的节点 p ,返回逻辑值;

2.1.2 若返回真值,则将 p 作为 P 的父节点;

2.1.3 若返回假值,则直接将服务器作为 P 的父节点;

2.2 若组不存在,则创建一个新的组,指定组长,服务器直接发送数据给加入节点。

(2) 节点退出。

●正常退出:节点向组长请求退出,同时通知自己的子节点,该节点的子节点将采用备用父节点作为父节点。组长将删除该节点的信息。

●非正常退出:组内节点每隔特定 t 时间向副组长发送一个状态包表示自己的存在。当有 N 个 t 时间都没有收到某个节点的状态包时,则认为此节点非正常退出。副组长向组长报告,组长向组内相关节点转发该信息,要求将该成员信息删除。

●组长退出:副组长将代替成为组长,同时选出同层的带宽空闲,计算能力优秀的节点作为副组长,并向服务器报告信息,向组内其他节点发出节点信息更新指令。

●副组长退出:组长根据带宽和处理能力选出副组长,并向组内节点发出更新信息。

节点 P 退出算法描述如下:

1. 节点 P 请求退出:

1.1 若 P 是组长,副组长代替组长;

1.2 若 P 是副组长,组长选出新的副组长;

2. 删除退出节点 P 的信息;

3. 更新节点组。

(3) 树的动态调整。

按照节点加入和退出算法并不一定能生成最优性能的树,因此需要在过程中动态调整树的节点位置。选出带宽和性能最好的节点作为组长和副组长。提出 BTBM-

Tree 组优化策略:

●带宽优先的原则:同一组内的节点,带宽越大,优先放在上层,将带宽小的节点降层。

●路由最近原则:路由越近,越优先,则包丢失和包延迟的性能越高。

●分级管理:服务器负责管理第一层,组长负责管理组内其他成员。

3.3 接收节点策略

除了服务器以外的节点都可以称为接收节点,其关键任务是:找到满足性能需求的发送节点集和备选父节点。其中,发送节点集分成活动节点集和备选节点集。接收节点接收来自活动节点集的数据流,若有活动节点退出或者接收节点不能满足播放需要,则从备选节点集中选择节点加入活动节点集。若接收节点的父节点退出,则将备选父节点作为自己的父节点。采取什么方法获取相关节点,文献[9]提出了一个以接收节点 Pr 为根的树搜索结构,其大概思想是:先推测出网络的拓扑结构,然后优化,接收节点根据优化后的节点性能参数,包括出口带宽(R_p)、服务时间(A_p)、网络带宽($Avail\ bw$),选出合适的节点集。推测和优化网络的拓扑结构比较复杂,文中将利用系统模型中提出的信息表将此方法简化。具体策略如下:

根据父节点提供的父层节点信息表,接收节点将父层节点按照出口带宽由大到小排序,择优录取,添加到发送节点组,直到满足需要,剩下的添加到备选组,同时选出备选组中的一个节点作为自己的备选父节点。若父层节点不能满足需要,则向更上一层节点甚至服务器发出请求,直到找到合适的节点为止。算法描述如下:

1. 选定层节点;

2. 对层节点按照有效带宽从大到小排序;

3. 从大到小将节点加入发送节点组;

4. 计算发送节点组中的有效带宽总和;

4.1 若有效带宽总和小于节点接收节点需求带宽,则将节点添加到活动节点组;

4.2 反之,则将节点添加到候选节点组;

5. 若层中所有节点不能满足接收节点需求,则向更上层节点提交请求,返回执行第 1 步。

4 系统性能分析

系统是采用 P2P 和 C/S 混合结构,因此具有集中与分布的特性。结合了此前许多学者研究的成果,比如层次树、时间阈值、分组管理等先进思想,现对系统性能进行分析。

●稳定性:由于第一层节点直接由服务器传送全部视频流,能保证层次树中的流源源不断,同时服务器也为其他层的节点提供补丁流,弥补了由于时间差异所带来的影响,另外,利用 BTBM-Tree 算法对节点树进行维护和管理,能够增强系统的稳定性。

●扩展性:面对大量的用户,采用时间阈值分组,并采

取带宽优先原则,使得系统具有良好的扩展性和灵活性。服务器由于只处理部分节点视频流信息和控制信息,大大减轻了计算压力和带宽压力,特别是对于点播高峰时间,能够满足更多用户需求,而且实现起来变得容易。

●延迟性:系统的延迟性是个困难问题。节点的离开和加入会带来延迟,数据流层次转发也产生延迟。这里采取限制树高度的办法,假设一棵树节点总数为 N ,平均每个节点带 k 个节点,则 $\log_k N$ 代表树的高度,要求 $\log_k N$ 最小,则可以减少树的过多层次,从而可以减少系统延迟。

5 结束语

设计和讨论了基于 P2P 的流媒体系统模型和关键算法的思想,并对系统的性能做了简要分析。本模型在现有技术条件下,可以方便实现,减轻了服务器和 Internet 骨干网的压力。进一步工作将着手开发原型系统,并检测系统的实际性能。

参考文献:

- [1] Xiang Z, Zhang Q, Zhu W, et al. Peer-to-peer based multimedia distribution service[J]. IEEE Transactions on Multimedia, 2004, 6(4): 343-355.
- [2] Hua K, Cai Y, Sheu S. Patching: A multicast technique for true video-on-demand services[A]. in Proc ACM Multimedia[C]. NY, USA: ACM, 1998.

- [3] Guo Y, Gao L, Towsley D, et al. Seamless workload adaptive broadcast [A]. in Proc of International Packet video Workshop[C]. Pittsburgh, USA: [s. n.], 2002.
- [4] Eager D, Vernon M, Zahorjan J. Bandwidth skimming: A technique for cost-effective video-on-demand[A]. in Proc Multimedia Computing and Networking 2000[C]. San Jose, CA: [s. n.], 2000. 1-10.
- [5] Rejaie R, Handley M, Yu H, et al. Proxy caching mechanism for multimedia playback streams in the internet[A]. In: Proceedings of the 4th International Web Caching Workshop[C]. San Diego, CA: [s. n.], 1999.
- [6] Gadde S, Chase J, Rabinovich M. Web caching and content distribution: a view from the interior[A]. In: Proc. of the 5th international web caching and content delivery workshop[C]. Lisbon, Portugal: [s. n.], 2000.
- [7] 方 炜, 吴明晖, 应 晶. 基于 P2P 的流媒体应用及其关键算法研究[J]. 计算机应用与软件, 2005, 22(5): 35-37.
- [8] Guo Yang, Suh K, Kurose J, et al. P2Cast: peer-to-peer patching scheme for VoD service[A]. Proceedings of the 12th international conference on World Wide Web[C]. NY, USA: ACM, 2003.
- [9] Hefeeda M, Habib A, Botev B, et al. PROMISE: peer-to-peer media streaming using CollectCast[A]. Proceedings of the eleventh ACM international conference on Multimedia[C]. NY, USA: ACM, 2003.

(上接第 44 页)

素,包括:询问过滤器;从哪里开始数据的查找;询问请求属性的个数等等。

缓存管理:因为目录服务器使用目录缓存以改善响应时间,所以度量缓存性能是很重要的。研究人员已经了解了与 LDAP 相关的缓存技术,而且为了提升其性能还提出了改进的算法^[8]。

上文综述了 LDAP 服务器实现技术的不同,比如对于 LDAPv3 的支持,访问控制列表的访问,多主源服务器的复制,但是通过软件开发商以及公共机构的努力,这些不同点都是可以利用的。

4 LDAP 的发展趋势

到目前为止 LDAP 已经发展到第 3 个版本,人们期望它能够与 X.500 目录服务进行更好的交互,从而为全球的目录网络提供更加便利的结构。

凭借着将 XML 技术和 LDAP 技术集成起来,LDAP 在数据管理方面的能力有了很大的提高,尤其在数据存储和数据索引方面。上文所提到的方法也已经在 OpenLDAP 服务器上用来实现 XMLDAP 的相关缓存,而且与传统的缓存技术相比,此种方法在平均访问时间上有了很大的改进。

当前 LDAP 在 Internet 上被广泛应用于数据的管理

工作,其涉及的领域有数据查询、索引、缓存以及安全性等。据此可以预测在将来 LDAP 必将会有更大的用武之地。

参考文献:

- [1] Wahl M, Howes T, Kille S. Lightweight Directory Access Protocol (v3)[S]. IETF RFC 2251, 1997.
- [2] 于 剑, 张 辉, 赵红梅. LDAP 目录服务在 Web 开发中的应用[J]. 计算机应用, 2003, 23(10): 82-83.
- [3] 张慧宇, 袁卫忠. LDAP 研究及其在 CA 中的应用[J]. 计算机应用研究, 2002(10): 37-38.
- [4] 赵宏建, 孙吉贵. 目录服务技术的分析比较及在 PKI 中的实现[J]. 吉林大学自然科学学报, 2001(4): 29-30.
- [5] XLNT Software. Handling XML Documents Using Traditional Databases[EB/OL]. www.surfnet.nl/innovatie/surfworks/xml/xml-databases.pdf. 2002-08.
- [6] Marron P J, Lausen G. On Processing XML in LDAP[A]. Proc 27th Int'l Conf Very Large Databases[C]. [s. l.]: ACM Press, 2001. 601-610.
- [7] Isode. Comparative Performance Benchmarking of Isode Mvaul R10. 1, white paper [EB/OL]. www.isode.com/whitepapers/m-vault-benchmarking.htm. 2003-10.
- [8] Cluet S, Kapitskaia O, Srivastava D. Using LDAP Directory Caches [A]. Proc Symp Principles of Database Systems (PODS)[C]. [s. l.]: ACM Press, 1999. 273-284.