

基于冷门预判的 P2SP 点播系统片选策略的设计与实现

王翠茹 王伟明 吴海博¹

(华北电力大学计算机科学与技术学院,保定 071003;中国科学院计算技术研究所¹,北京 100190)

摘要 结合视频点播应用中对于提高点播质量和减轻服务器负担的要求,设计和实现了一个融合 P2P 和 P2S 网络的点播系统 xBT。分析现有的 BitTorrent 协议应用于点播系统的局限性,针对点播过程中存在冷门片段的特点,对 BitTorrent 协议的片段选择策略提出改进。试验分析表明,xBT 在大并发量的点播系统中能够提高用户的播放质量,同时减少服务器压力。

关键词 P2P VoD 片选策略

中图法分类号 TP393.03; **文献标志码** A

随着 Internet 的发展,用户对传统的流媒体服务提出了更高的要求。采用传统 C/S 模式的流媒体服务已经难以满足用户对高码率、高下载速率的要求。在这种情况下,P2P 技术被引入视频点播(VOD)中。P2P 技术能够充分利用网络中广泛大量分布的客户端计算能力和带宽,使得系统的服务能力得到极大的提高,目前因特网上有 50% 的流量被各种 P2P 系统占据^[1]。P2P 技术最早被应用于文件共享。视频点播与文件共享中的客户端都需要下载资源,但是由于视频点播中流媒体的时序性和拥有服务器支持等特点决定了 VOD 中的 P2P 技术有别于文件共享中的 P2P 技术。本文通过改进 BitTorrent 协议中的片选策略,实现了一套基于 P2SP 的视频点播系统 xBT。该系统能够提高用户点播质量,并减少服务器负担,以提高系统的服务能力。

1 相关研究工作

1.1 BitTorrent 协议

BitTorrent 协议(BT 协议)是一个用于文件共享

的 P2P 协议^[2]。一个典型的 BT 系统主要由以下 4 部分组成:种子服务器、下载节点(Peer)、种子(Seed)和索引服务器(Tracker)。种子服务提供种子文件的上传和下载服务,Tracker 服务器为下载客户端提供其他 Peer 信息,下载节点依靠 Tracker 提供的信息连接其它 Peer 和 Seed,开始上传和下载数据。其中 Seed 是一个特殊的 Peer,它只提供上传服务,不会从其他 Peer 下载数据。Peer 和 Seed 都会定期向 Tracker 上报本节点的上传下载信息,使得 Tracker 能够掌握 BT 网络的信息。

1.2 BitTorrent 的片段选择

为了提高文件的共享效率,BT 协议将共享文件分成多个片段(Piece)。片段是 Peer 之间共享数据的最小单元。Peer 下载文件的过程,是选择下载片段和请求下载片段不断循环的过程。

下载片段的选择是一个至关重要的步骤,选择不同的 Piece 下载会直接影响到 Peer 的下载速度、BT 网络的健壮性。一个好的片段选择策略能够将共享文件中的不同片段快速、高效、均匀的分布 P2P 网络中的各个 Peer 上,进而保证种子节点的离开不会影响到其他节点的下载。同时,如果一个 Piece 在 P2P 网络中存多个副本,就能加快 Peer 请求下载这个 Piece 的速率。BT 协议根据下载不同阶段存在的

2009 年 11 月 2 日收到 国家科技支撑计划(2006BAH02A11)资助
第一作者简介:王翠茹(1954—),女,河北保定人,教授,研究方向:网络通信和数据库系统。

特点,采取了不同的片段选择策略^[4],可分为:下载开始阶段的随机片段选择,下载过程中的未完片段优先和最少片段优先,以及临近下载结束时的 Endgame 模式。

在下载的开始阶段,由于本地节点没有任何数据,根据 choking/unchoking 算法,为了避免自己被其他 Peer 长时间阻塞,本地节点应该尽快完成一个片段的下载。因此,处于这个阶段的节点会随机选取一个 Piece 开始下载。

当本地节点拥有了一个完整片段的数据后,开始采用另外一种片段选择策略。本地节点会保存与它相连的其他 Peer(邻居节点)所拥有的片段情况。在选择下一个被下载的片段时,通过掌握的邻居节点所拥有的片段分布情况。节点可以统计出每一片段数据各出现了几次,也就是在几个节点有备份。对于备份数最少、最有可能消失的片段,节点优先选择这些片段请求下载。即选择它的邻居节点们所拥有的最少的那个片段,这就是“最少优先”策略。

在向邻居节点请求某个片段的过程中,可能由于网络原因或者 choking/unchoking 算法的阻塞,导致这个片段的下载被中断。这时,没有下载完成的片段拥有最高的下载优先级。这样做是为了让本地节点能够尽快完成一个片段下载,以便它能够为其他 Peer 提供服务。

在下载即将结束的阶段,如果本地节点正在缓慢地从一个 Peer 下载最后一个片段,那么这个操作将会极大的延迟本地节点的下载完成时间。为了防止这种情况,本地节点会向它的所有邻居节点请求某个片段的同一个子片段,一旦子片段到达,它就会向其它 Peer 发送 Cancel 消息取消该片段的传输,以节省带宽资源。

2 xBT 点播系统的架构设计

xBT 点播系统采用 P2P 与 P2S 结合的架构。系统主要由媒体服务器,P2P 索引服务器,数据库服务器和点播终端组成。如图 1 所示:媒体服务器存储了点播系统中的所有媒体资源,当本地节点无法从

P2P 网络中获取期望的数据或 P2P 网络无法及时提供当前播放位置的数据时,它将会从媒体服务器获取数据。P2P 索引服务器与 BT 协议中的索引服务器功能类似,它不但返回 P2P 网络中 Peer 的信息,同时也返回系统中媒体服务器的信息。数据库服务记录了媒体资源的信息。

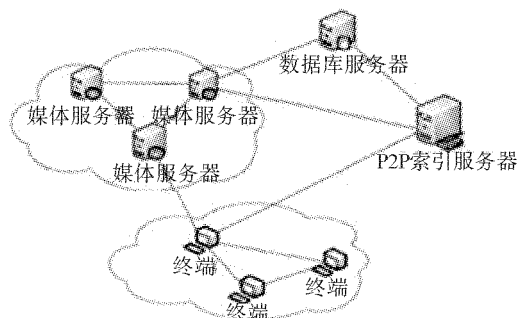


图 1 xBT 点播系统架构图

这种混合架构能够在保证用户点播质量的前提下,利用 P2P 网络减轻服务器负担。本篇文章也是在这种架构的前提下,结合点播的特点,研究改进 BT 协议的片段选择策略,使得终端在点播过程中尽量多的从 P2P 网络中获取点播数据。

3 xBT 的片段选择策略

3.1 流媒体点播系统对于片段选择策略的要求

BT 协议很好地解决了文件共享中客户端下载效率依赖服务器瓶颈的问题,但是流媒体的一些固有点决定了 BT 协议不能发挥其最大的优势。文献[5]分析了在 P2P 网络中实现点播系统的可行性。在单纯的 BT 网络中支持 VOD 点播,存在着如下需要解决的问题:

(1) BT 协议乱序下载与流媒体顺序播放的矛盾,我们需要设计一种片段选择策略,它既能够流畅地播放流媒体,又能利用 P2P 网络的乱序下载提高下载效率。

(2) 点播系统存在“冷门片段”现象^[6],多数用户会通过 VCR 操作跳过媒体资源中的一些片段,使得这些片段的下载无意义,而 BT 协议只支持完整文

件的下载。

(3)在下载过程中,BT 协议采用最少优先的策略使得资源的各片段副本尽量均匀地分布于网路中,这种策略不适合与点播系统,因为它可能导致当前播放位置的片段不能及时地下载到本地节点,同时在点播中,对于本地节点,当前播放位置之前的片段其价值将远远小于之后的片段。

3.2 已有的片段选择策略的分析

为了解决问题 1 和 3,文献[7]和文献[8]提出采用优先级划分的策略,文献[7]提出建立两个片段下载队列:高优先级队列和普通优先级队列。高优先级队列中存放当前播放位置之后即将播放的未下载片段,普通优先级队列按照最少优先的顺序排列播放位置后的未下载片段,当选择片段时,以 $P\%$ 的概率选择高优先级队列中的片段,以 $(100 - P)\%$ 的概率选择普通优先级队列。 P 值在下载过程中允许动态调整。这种算法能够以最小的改动兼容 BT 协议,只需要改变 BT 协议中的片段选择策略。但是这种算法建立在 P2P 网络中,随着 P 的增大,必然会提高某一片段在种子离开 P2P 网络后导致某些片段不可下载的几率。

文献[8]引入了媒体服务器,并提出了三种下载策略:第一种是严格最近优先,第二种是以 beta 概率分布随机选择播放位置之后的未下载片段,第三种融合了前两种策略,既将播放位置之后的片段分为两个区域,第一个区域严格按照最近优先的策略下载,剩下的区域按照 beta 概率随机下载,只有在第一个区域下载完成之后才能进行第二个区域中片段的下载。文章还比较了这三种策略对媒体服务器的服务压力的影响。这种算法能够为用户的播放质量提供足够的保证,但是严格最近优先策略在 BT 协议中会导致本地节点将长时间被其他 Peer 阻塞,从而增加服务器的压力。

上述文章只对视频正常播放情况下的视频数据下载展开研究,而没有考虑用户拖拽等 VCR 操作对 P2SP 网络产生的影响。

3.3 xBT 的片段选择策略

xBT 的片段选择策略由 4 个阶段组成:点播开

始,点播过程,拖拽过程和点播即将结束。

在点播开始阶段,由于本地节点没有任何的 Piece 可供上传,因此如果从 P2P 网络下载数据有可能导致本地节点长时间被其他 Peer 阻塞,而影响影片播放的流畅度。因此,在这个阶段,本地节点将直接向媒体服务器请求数据。

点播过程中为了保证播放的流畅度,从播放位置开始后的一段连续片段,称为流畅缓冲区,这个区域中的数据应该被立即下载。流畅缓冲区的大小由流媒体的播放码率决定,设为 $S1$ 。流畅缓冲区之后的区域称为普通下载区。当流畅缓冲区中存在未下载的 Piece 时,本地节点将直接向媒体服务器请求下载。普通下载区采用类似于 BT 协议中的最少优先策略,但不是严格的最少优先策略,因为它会根据其他 Peer 的 VCR 操作改变某些片段的下载优先级。具体实现如下。

点播系统中存在着“冷门片段”现象,也就是说某些媒体节目中的某些片段是不受欢迎的,大多数的观众会通过快进拖拽等 VCR 操作跳过这些片段^[9],那么这些片段对于用户来说应该具有较低的下载优先级,即使这些片段在 P2P 网络只有很少的副本数。

那么本地节点怎么知道哪些片段是冷门片段呢?有两种方案:第一种,在系统中添加媒体统计服务器,由它来确定冷门片段。本地节点的每次 VCR 操作都上报统计服务器,统计服务器根据这些数据计算媒体资源中哪些是冷门片段;第二种方案没有媒体统计服务器,本地节点的 VCR 操作将广播给它的所有邻居节点,同时根据它收到的邻居节点的统计数据计算得出冷门片段。第一种方案能够结合历史数据较准确的计算出冷门片段,但是统计服务器有可能会成为系统的性能瓶颈。第二种方案,能够解决系统性能瓶颈的问题,但由于本地节点得到的统计数据有限,因此计算出的冷门片段带有局部性;而且由于 P2P 网络中节点处于不断变化,历史统计数据不能得到很好的继承。新加入的节点可以从它的在线时间最长的邻居节点获得统计数据,从而弥补以上的不足。xBT 采用第二种方案。

在 xBT 系统中,每个节点维持一张记录表,记录表记录了当前资源的每个片段的冷门程度,具体实现如下:

(1)本地节点产生了一个 VCR 操作,它以 $[a, b]$ 这样的消息格式向它的所有邻居节点广播。其中 $[a, b]$ 为 VCR 操作所完全包含的片段区间。

(2)邻居节点 N 收到一条 VCR 消息,它更新自己的片段冷门程度表,既将区间 $[a, b]$ 内的所有片段的冷门值加 1。

(3)本地节点通过最少优先策略选择普通下载区中的一个片段 P ,若 P 的冷门值大于阈值 F ,则放弃该片段的下载,选择下一个非冷门片段。

本地节点在节目资源结束后,通常会退出该资源所属的 P2P 网络。因此,当本地节点在播放接近结束的时候很难从 P2P 网络中获取它所需要的数据。为了保证用户的点播质量,节点应该同时从 P2P 网络和媒体服务器请求数据。

4 xBT 系统性能分析

xBT 系统的设计目标是在保证用户点播质量的前提下,减少媒体服务器的负担。我们以点播延迟时间和服务器 CPU 使用率作为指标比较 xBT 的片选策略和传统 BT 的片选策略。

我们将 xBT 部署在科学网中,提供测试用户两个版本的客户端,分别采用 xBT 的片选策略和 BT 的片选策略。系统配置一台资源服务器,一台 Tracker 服务器和一台种子服务器,具体硬件配置如下表 1。

表 1 xBT 硬件配置表

	CPU	内存	硬盘
资源服务器	Xeon E5405	4 GB	1 TB
Tracker 服务器	Xeon E5405	2 GB	80 GB
种子服务器	E5200	1 GB	250 GB

图 2 展示了在不同并发节点数的情况下,用户每次 VCR 操作所产生的平均延迟时间。从图中可以看出,在并发节点数少于 300 时,BT 的片选策略

的延迟时间要略低于 xBT 的片选策略,但是当并发用户数持续增大时,xBT 的片选策略要明显优于 BT。这是由于系统在线用户数较少时,xBT 的冷门片段算法所依赖的统计数据较少,少数用户的 VCR 操作产生的异常数据对算法的准确性影响较大,当用户数足够大时,xBT 的片选算法能够正确稳定的运行。

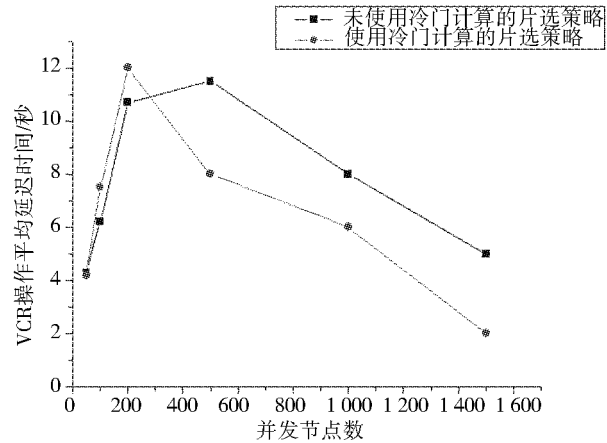


图 2 xBT 客户端延迟时间

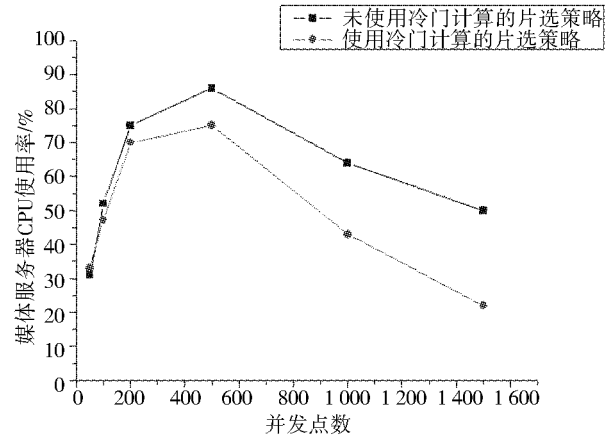


图 3 xBT 媒体服务器性能测试

图 3 展示了系统在不同并发节点数下媒体服务器的 CPU 使用率。当并发节点数低于 300 时,两种片选策略对于媒体服务器 CPU 的占用比例基本相同。当并发节点数持续增大时,xBT 的片选策略对于媒体服务器 CPU 的占用要明显小于 BT 的片选策略。这是由于 xBT 的片选策略能够优先下载媒体资源中的热门片段,从而使得用户进行 VCR 操作时,

不必从媒体服务器请求数据,因此减少了媒体服务器的负担。

从以上的实验数据可以看出,xBT 在大并发量的点播系统中能够提高用户的点播质量,并减少媒体服务器的压力。

5 结 论

本文提出了一种基于 P2P 和 P2SP 架构的针对 VOD 应用的片段选择策略。通过实验得到的系统性能数据分析,该策略能够减少用户点播延迟时间,并减少服务器压力。进一步的研究计划包括对系统的冷门片段算法进行完善,使其能够更为稳定高效的运行,减少噪声数据对其影响。

参 考 文 献

- 1 Cache Logic Research. The true pictures of P2P file sharing. <http://www.cachelogic.com/research/slide1.php>, 2004
- 2 Cohen B. The BitTorrent Protocol Specification. http://www.bittorrent.org/beps/bep_0003.html, 2008
- 3 Adar E, Huberman A B. Free riding on Gnutella. *First Monday*, October 2000; 5(10):27—32
- 4 Cohen B. Incentives build robustness in bittorrent. In: *Proceedings of the First Workshop on Economics of Peer-to-Peer Systems*. Berkley, CA, USA, June 2003
- 5 Annapureddy S, Guha S, Gkantsidis C, *et al.* Is high-quality VoD feasible using P2P swarming? In: *Proceedings of ACM WWW*. Banff, Alberta, Canada, 2007
- 6 Yan Huang, Tom Z J. Chin D M, *et al.* Challenges, design and analysis of a large-scale P2P-VoD System. *SIGCOMM*. 2008
- 7 Vlavianos A, Iliofotou M, Faloutsos M. BiTos: enhancing bitTorrent for supporting streaming applications. In: *9 th IEEE Global Internet Symposium (GI2006)*. Barcelona, Spain, April 2006
- 8 Choe Yungryn, Schuff D L, Dyaberi J M, *et al.* Improving VoD server efficiency with bitTorrent. *Proceedings of the 15th international conference on Multimedia*, 2007
- 9 Yu Hongliang, Zheng Dongdong, Zhao B Y, *et al.* Understanding User Behavior in Large-Scale Video-on-Demand Systems. *Proceedings of the Euro Sys Conference*, 2006

Design and Implementation of Piece Selection in P2SP Network for VoD Based on Cold-prediction Algorithm

WANG Cui-ru, WANG Wei-ming, WU Hai-bo¹

(School of Computer Science & Technology, North China Electric Power University, Baoding 071003, P. R. China;
Institute of Computing Technology, Chinese Academy of Sciences¹, Beijing 100190, P. R. China)

[Abstract] In order to improve the performance of VoD, a hybrid server/P2P streaming system called xBT is introduced for large-scale VoD (Video-on-Demand) service, which can reduce delay time for play and mitigate the workload of servers. Drawbacks of current BitTorrent protocol for VoD application in a P2P mesh-based network is analyzed and improved the piece selection strategy used in BitTorrent to make it work more efficiently when VCR operations happened. The new piece selection strategy is based on the cold-prediction algorithm. The experiments indicate that xBT can enable the improvement of users' experiences in VoD by reducing delay and lower the overhead for servers in large scale VoD system.

[Key words] P2P VoD piece selection strategy