

一种克服磁盘存取受限的视频点播机制^{*}

贺小箭^{1,2} 李方敏¹ 尤晋元²

(湘潭工学院计算机科学与技术系 湘潭411201)¹ (上海交通大学计算机科学与工程系 上海200030)²

A Video-on-Demand Delivery Policy for Improving the Disk Access Performance

HE Xiao-Jian^{1,2} LI Fang-Min¹ YOU Jin-Yuan²

(Department of Computer Science and Technology, Xiangtan Polytechnic University, Xiangtan 411201, China)¹

(Department of Computer Science and Engineering, Shanghai Jiaotong University, Shanghai 200030, China)²

Abstract In local high-speed network, the performance of VOD system is limited by disk-bound. In this paper, we propose an enhanced video-on-demand delivery policy based on dynamic buffering algorithm. Considering the controllable bandwidth allocation for multicast and the characteristics of on-demand video, we optimize the disk access performance of video sever to fully utilize the I/O capability of disk and the bandwidth of network. For a practical VOD system, this strategy can supply instantaneous video-on-demand services and maximum on-demand video number cheaply.

Keywords Video-on-Demand, Multicast, Emerging-stream algorithm, Dynamic buffering

1 引言

在视频服务器中,大量的待点播视频流数据以分级存储模式被加以组织,其中点播概率小的节目被存放在低速廉价的光盘库中作为非在线视频(off-line)。为保证视频的播送质量,非在线视频必须先存放到硬盘缓冲区中才能被播放。进一步考虑到网络传输性能因素,将视频服务器按层次化结构进行组织,将待点播的视频数据靠近用户存放,保证缓存的数据为用户经常点播的并且是最新的视频数据。文[3,4,5]针对该问题讨论了多种缓存策略和存储分配算法,但其系统实现的代价较高。

为求经济实用,基于小区高速网的视频服务器通常采用性价比很高的PC服务器或服务器群(Cluster),网络环境则为基于TCP/IP的千兆以太网。当前计算机和网络技术的发展态势为:存储器容量每12个月翻一番,网络速度每9个月翻一番,硬盘存取速度的增长则远远落后于其容量的增长速度。由此可见,服务器端磁盘存取速度受限制约了视频点播系统的性能提高,若采用增加服务器群的节点以提高并行磁盘I/O的性能,则其代价相当昂贵。在实际的视频点播应用中,系统资源利用率较低,使得总体时间段上能并发的点播用户数较少。基于有效的动态网络带宽分配机制和视频缓存算法的基础上,本文提出了一个改进的视频点播策略,通过提高磁盘工作效率来缓解磁盘受限的瓶颈效应,从而改善系统的点播接收率和实时播送性能。

2 视频流的组播模式

传统的视频点播系统为每个点播用户分配单独的通信通道和磁盘读取服务,其占用系统资源多,算法效率低。大量的实践证明,用户经常点播的视频节目集中于少数热门节目上,

由此可合并对相同节目的点播服务,即当新的点播请求与原点播的节目相同时,新的点播请求并入到原点播视频流进行组播处理,从而极大地减少了网络通信量并减轻了服务器端磁盘的工作负载。

常用的组播通信模式有批处理和流合并两种调度算法。在批处理模式中,系统在启动某个节目的播送时先等待一段时间,然后将这段时间内点播该节目的所有点播用户用一个组播流进行播放^[2]。为了收集足够多的批处理点播请求,该算法的启动延时很大。在实际应用中,对于大批量瞬间到达的相同节目点播请求(即在很短时间间隔到达,播放起始时的片刻延迟对用户影响不大),此时可采用批处理方式。

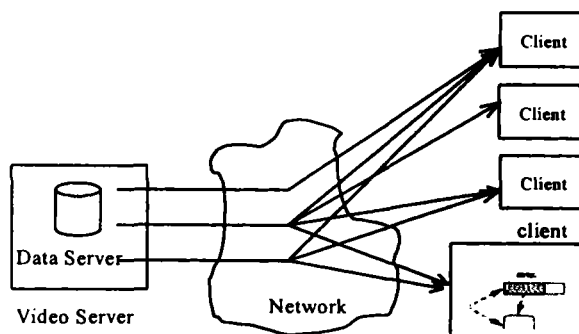


图1 采用流合并调度的VOD系统结构示意图

在流合并组播模式中,当新到达的点播请求与正在播送的节目相同时,将该新点播用户加入到原点播组当中进行组播传送,各组传送的部分视频数据在客户端合并后播放^[1](如图1.2所示)。该模式能够在接到点播请求的同时启动节目播放,其启动延时小,但同时网络带宽的要求要高。由于小区宽带网的网络管理是可控制的,当将视频服务器播送的信息

^{*} 国家自然科学基金资助项目(69974031)和湖北省自然科学基金项目(01JJY2064)资助。贺小箭 博士研究生,主要研究方向为分布式计算、多媒体技术,李方敏 副教授,博士,主要从事网络服务质量、新型网络协议、多媒体通信技术等方面的研究,尤晋元 教授,博士生导师,主要从事操作系统、分布对象计算、移动计算、JAVA技术、以构件为基础的软件工程等方面的研究。

流标记为高优先级服务类型时,通过区分服务(Diffserv)比例带宽分配机制使视频服务器的输出带宽获得保证^[6],从而保证了各组播通道的动态带宽分配。

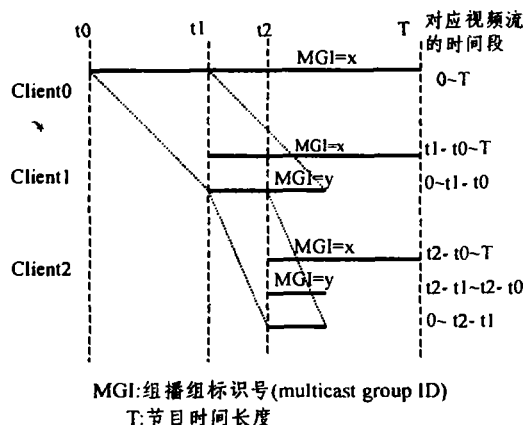


图2 组播通信中的流合并过程

在该流合并算法中,最先点播该节目的用户创建初始流,初始流通道播放整个节目。节目播送一个时间间隔 t_i 后,到达一个对该节目的新的点播用户,该点播请求加入到该节目的组播组当中,合并并缓存这些组播数据到客户端,同时服务端为该新用户产生遗失的节目片断流供客户端及时播放,并为该新节目片断动态分配一个通信通道。这种改进的组播策略充分利用了宽带网和客户端大容量缓存的系统特征,利用自适应缓冲技术获得最佳的系统资源利用率。

3 结合自适应缓存提高磁盘工作效率

在本文提出的视频点播系统中,其主要的系统参数包括:

- v_i 第 i 路视频流的网络传输速度
- r_i 保证第 i 路视频流实时播放的数据传输速率
- B_{vi} 保证第 i 路视频流实时播放的缓冲区大小
- B_s 视频服务器中总的缓冲区大小
- r_s 视频服务器的网络传输速率
- d_s 视频服务器的磁盘读取速度
- B_i 第 i 路视频流在视频服务器中占用的动态缓冲区大小

3.1 磁盘工作效率分析

传统的缓冲方式根据所点播视频流的播放质量计算出所需要的缓冲大小,该缓冲区大小与视频播放速度相匹配,视频服务器根据该缓冲区控制的速度从磁盘读取和发送视频数据,以保证客户端在实时播放时不出现上溢和下溢。若瓶颈为服务器端的磁盘读取速度,则满负荷时系统可满足的最大点播用户数为:

$$n_{\max} = d_s / \text{平均视频播放速率}$$

该传统缓存方式着眼的是单路视频流的瞬态效应,系统的整体效率不高,并且由于网络突发性的瞬态干扰,系统播送表现出较差的抖动性。

在实际的视频点播应用中,用户对视频节目的选择遵从 Zipf 分布,点播请求的到达模式符合泊松分布^[2]。根据该模型可分析得知,传统缓冲方式的磁盘平均读取速度小于磁盘最大读取速度。从提高磁盘读取效率的需要出发,希望磁盘平均读取速度接近或等于最大读取速度,即一旦接受点播请求就马上全速读取并发送给各客户端,为以后的请求争取更多的保留资源,从而在总体时间段上大大提高系统的最大点播用

户数。此时要求改进原来传统的缓冲方式以提高数据利用率,解决客户端缓冲区上溢问题。

传统缓冲设置是根据客户端存储空间受限的情况而制定的。根据当前典型的视频应用分析,标准 MPEG-2 视频流的平均视频播送速率为 0.5 MB/s,两小时视频节目所需的存储空间为 3.6 GB。当前多数用户都是用高性能 PC 机点播节目,新式的置顶盒也将拥有高性能的计算能力和大的存储空间。通常,主流的客户端配置为: Disk > 30GB、RAM ≥ 128 MB (256MB),其硬盘的数据存取速度大于 20MB/s。利用 RAM 和硬盘,在客户端缓存整部视频节目是切实可行的。由此,客户端缓存不受限的条件解决了客户端缓冲区上溢的难题,并且为提供 VCR 功能的交互式视频演播打下了基础。

3.2 动态缓存

视频服务器为每个视频流组播通道维持一个动态缓冲区,其缓冲区大小根据磁盘读取速度和网络传输速率相应地进行动态调整,并要求链的最小数据长度不小于该点播视频流的 B_{vi} 。为了方便动态内存的管理,并加快视频数据寻址,缓存采用以页为单元的页面数据链结构(如图3所示)。

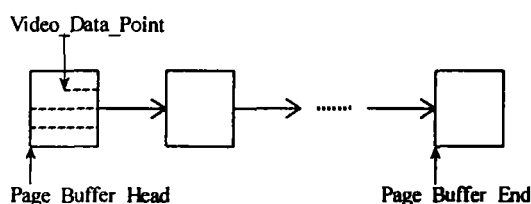


图3 动态缓存的页面数据链结构

为利于各组播数据在用户端的合并,每个播送视频流的动态缓冲区对应一个参数表,其结构成员包括:

- Page_Buffer_Head 指针,它指向正传送页的首部;
- Video_Data_Point 指针,指向正待传送的数据字节,它同时指明了当前总的数据偏移量;
- Page_Buffer_End 指针,指向已经由磁盘读取而未传送的末页的首部;
- Video_Segment_Head 指针,在流合并中本视频段的起始偏移量;
- Video_Segment_End 指针,在流合并中本视频段的终止偏移量;
- Video_Segment_Point 指针,标明当前传送字节在整部视频中的总偏移量;
- Multicast_Group_ID,为组播标识号。

通过动态流量控制算法,动态缓冲区将根据缓存资源和点播行为进行相应的调整,在带宽充分的情况下保持最大的磁盘工作效率。点播相同节目的新加入请求在共享原组播缓存数据的同时,将有可能利用该视频的历史数据以进一步减少磁盘 I/O 量,其历史数据的利用率由视频在缓存的驻留时间所决定。视频驻留时间的调整可采用 LRU、热点节目关注度等缓存算法,文[3,5]中给出了一些可行的缓存方案。

4 点播过程的自适应缓冲控制

影响视频点播性能的因素是多方面的,其主要的折衷因素包括:网络传输速度,客户端存储空间,视频服务器缓存容量,视频服务器磁盘存取速度等。其中视频服务器的计算能力希望是越快越好,因为视频服务器计算能力直接影响到整个系统的动态平衡。当客户端能力不足时,需要设置该点播的传

送速度等于 r_i , 其缓冲区大小为 B_{vi} , 该路视频流其实就是传统的点播方式。

为达到最佳的磁盘工作效率, 缓冲动态调整算法需要考虑网络传输速度和缓存容量。网络传输速度包括单个播流传输速率和服务器针对所有客户能提供的传输速率, 其中单个播流传输速率由客户端计算能力、客户端数据存取速度、网络带宽及服务端提供的传输速率等因素决定。服务端网络传输速率由允许的网络带宽和服务器数据传输能力决定, 它限定了同时播送实时演播视频流的最大传送能力。

4.1 维持已有点播用户

在稳态运行过程当中, 视频服务器为 n 个点播用户保持有 k 路视频流, 它根据客户端和网络传输情况为 k 路视频流尽力读取并传送视频数据。视频服务器总的视频传送速度为 V , 其中第 i 路视频流的传送速度为 v_i 。当单路视频流的传送速率大于其数据读取速率时, 其从磁盘读出的视频数据立即被传送到客户端。而当该视频流的传送速率小于其数据读取速率时, 其对应的动态缓冲区将不断增大, 每当其缓冲区装满时就必须为新读取的数据申请一个新的页面, 并将其加在页面数据链后端。传送成功的页标记为脏, 供以后使用。

动态缓冲区大小的下限为 B_{vi} , 根据推算得出其上限为 $((B_s - \sum_{i=1}^k B_{vi}) * v_i) / V$ 。当第 i 路视频流达到其最大缓存时, 磁盘读取被暂停。而当第 j 路视频流传送完毕时, 清除其对应的缓冲区, 系统在 $k-1$ 路视频流的基础上达到一个新的平衡。当网络带宽足够时, 磁盘读取速度接近最大速度。

4.2 接受新的点播请求

当收到一个新的点播请求时, 系统原有的平衡被打破, 进入一个暂态过程。如果视频服务器接受该第 $k+1$ 路点播视频流, 则系统为维持 $k+1$ 路视频流必须同时满足的充要条件有:

$$(1) v_{k+1} > r_{k+1}; (2) \sum_{j=1}^{k+1} r_j > r_s; (3) \sum_{j=1}^{k+1} r_j > d_s;$$

$$(4) \sum_{j=1}^{k+1} B_{vj} > B_s.$$

在资源充分的情况下, 系统接收该请求。若视频服务器中有足够多的空闲页, 则马上为新加入者分配一个新的动态缓冲区。而当视频服务器中无空闲页面时, 则需要从缓冲区大小超过其 B_{vi} 的各动态缓冲区中抽取页面, 被抽取的对象为 $\max\{\frac{B_i - B_{vi}}{v_i} * V\}$, 它反映了被抽取的动态缓冲区超长因子和数据传输速度因子的权衡。当抽取到足够的缓存页面后, 视频服务器停止抽取页面并为新视频流创建动态缓冲区。为了

维持新加入者后的系统运转, 系统达成一个新的平衡条件, 所有超过新平衡条件上限的旧缓冲区停止增长。经过一段时间的运行后, 视频服务器达到一个新的平衡态。此时, 各点播视频流的数据读取速率有所下降, 但仍能保证其数据率不小于 r_i , 即保证该点播视频流数据的磁盘读取速度不小于该视频流实时播放的数据消耗速度。

由于传输速度和缓存受限, 已读入缓存的数据有可能被废弃, 从而有些数据需二次(或多次)从磁盘读取, 降低了磁盘读取效率, 出现这种情况说明瓶颈是网络传输速度而不是磁盘读取速度。

4.3 暂缓新的点播请求

如果系统资源不足以保证新的点播请求加入后的正常播送, 则拒绝新的请求, 并告知该请求何时能获得满足, 让用户选择是取消该点播请求还是预定该点播节目。预定一个节目是指视频服务器为该请求分配和保留一些资源, 并保证该用户在预定的时间能够实时播放该请求的视频。预定方式工作时, 系统能够根据数据相关性控制预先将部分数据传送给用户, 提高了系统的利用率。

结束语 基于可控的小区高速网环境, 本文提出了一个采用自适应动态缓存的视频点播机制。在服务分类的带宽保证基础上, 本策略利用动态缓存极大地提高了视频服务器磁盘的工作效率, 缓解了磁盘存取速度对系统性能的影响。在小区的视频点播及虚拟实验室视频流媒体的分发项目中, 该算法实现表现出很好的运行效果。由于采用了客户端全缓冲手段, 对于授权、认证及计费等问题需要对系统作进一步的研究。

参考文献

- Gao L, Towsley D. Supplying instantaneous video-on-demand services using controlled multicast. In: Proc. IEEE Intl. Conf. on Multimedia Computing and Systems, 1999
- Dan A, Sitaram D, Shahabuddin P. Dynamic patching policies for an on-demand video server. Multimedia Systems, 1996, 4(3): 112~121
- Dan A, Sitaram D. Buffer management policy for an on-demand video server. IBM Research Division, T J Watson Research Center. [Tech Rep: IBM Research Report RC 19437]. Armonk, New York, 1993
- Chan S-H G, Tobagi F. Distributed Servers Architecture for Networked Video Services. IEEE/ACM Trans. on Networking, 2001, 9(2)
- Brubeck D W, Rowe L A. Hierarchical storage management in a distributed VOD system. IEEE Multimedia, 1996, 3(3): 37~47
- 李方敏, 李仁发, 叶澄清. 一种核心无状态保存的自适应成比例公平带宽分配机制. 计算机研究与发展, 2003, 39(2)

(上接第60页)

- Innom W H. Building the Data Warehouse, 2nd[M], 2000
- Wu K L, Yu P S, Ballman A. SpeedTracer: A Web usage mining and analysis tool. [J] IBM System Journal, 1998, 37(1): 89~105
- Cooley R, Srivastava J. Web Mining: information and pattern discovery on the World Wide Web (A Survey Paper). [C] In: Proc. Of the 9th IEEE Intl. Conf. on Tools with Artificial Intelligence (ICTAI'97), Nov. 1997. http://maya.cs.depaul.edu/~mobasher/papers/webminer-tai97.ps
- 李立明, 蔡立军, 等. 电子文档信息自动挖掘技术中的预处理研究. 计算技术与自动化, 2002, 21(2): 92~96
- Hush Communications Corp. Hush Mail FAQ. http://www.hushmail.com/faq.htm, Nov. 1999
- Postel J. Simple mail transfer protocol. STD 10, RFC821, USC/Information Sciences Institute, Aug. 1982
- Crocker D. Standard for the format of ARPA Internet text

message. STD 11, RFC 522, UDEL, Aug. 1982

- Myers J. Post office protocol — version 3. RFC1725, Dover Beach Consulting, Inc., Nov. 1994
- Borenstein N, Freed N. MIME (Multipurpose Internet Mail Extensions) part one: mechanisms for specifying and describing the format of Internet message bodies. RFC 1521, Bellcore, Innosoft, Sep. 1993
- Moore K. MIME (Multipurpose Internet Mail Extensions) part two: message header extensions for non-ASCII text. RFC1522, University of Tennessee, Sep. 1993
- Ramsdell B. S/MIME Version 3 Message Specification. IETF RFC 2633, Jun. 1999
- Sahami M, Dumais S, Heckerrnan D. A Bayesian Approach to Filtering Junk E-Mail. AAAI 98 Workshop on Text Categorization. July 1998