基于 BitTorrent 模型的 VOD 系统的设计*)

彭永祥[1] 卢显良[1] 唐 晖[2] 段翰聪[1]
(电子科技大学计算机科学与工程学院 成都 610054)[1]
(中国科学院声学研究所高性能网络实验室 北京 100080)[2]

摘 要 利用 P2P 技术将集中的 VOD(Video-On-Demand 视频点播)服务分散化,使参与的用户既作服务的消费者,又作服务的提供者,充分利用客户端大量闲置资源,消除传统 VOD 系统的瓶颈,是当前进行大规模 VOD 应用的研究方向。本文提出了一种基于 BitTorrent 的 VOD 系统设计。该系统设计适用于大规模视频点播应用,且易于部署和扩展;对基于 BitTorrent 模型的视频服务能力进行了数学模型的建立和分析,并在实验环境中得到求解和验证。

关键词 BitTorrent,P2P,VOD,重定向

Design of VOD System Based on the Model of BitTorrent

PENG Yong-Xiang[1] LU Xian-Liang[1] TANG Hui[2] DUAN Han-Cong[1]
(Computer Science and Engineering College, University of Electronic Science and Technology of China, Chengdu 610054)[1]
(High Performance Network Laboratory, Institute of Acoustics, Chinese Academy of Science, Beijing 100080)[2]

Abstract To decentralize the VOD (video-on-demand) services by using the P2P technology, of which any user acts as both the consumer and the provider, will take full use of users' idle resources and eliminate the bottleneck of traditional VOD systems. The paper provides a VOD system based on the model of BitTorrent, which is applied to large-scale VOD application and is easily deployable and extensible.

Keywords BitTorrent, P2P, VOD, Redirect

1 引言

随着网络技术、多媒体压缩技术的发展,大规模点播应用的需求正在突显。而传统的 C/S 模式流媒体点播系统无法满足这种迅速膨胀的用户需求,或者需要付出巨大的硬件成本代价。访问过宽带电影网站的用户都会有这样的感受:一到高峰时段,几乎所有的宽带网站都很难连上,即使连接上了,电影的播送也是时断时续。之所以如此,是因为无论是集中式服务器本身,还是它的网络带宽,都构成系统的瓶颈。要想消除这个瓶颈,最好的办法是将它的服务分散化,充分利用客户端被闲置的资源(如:存储空间、计算能力、网络带宽),使系统中的任意主机既享受服务,也提供服务——这就是 P2P (Peer-to-Peer)的策略。

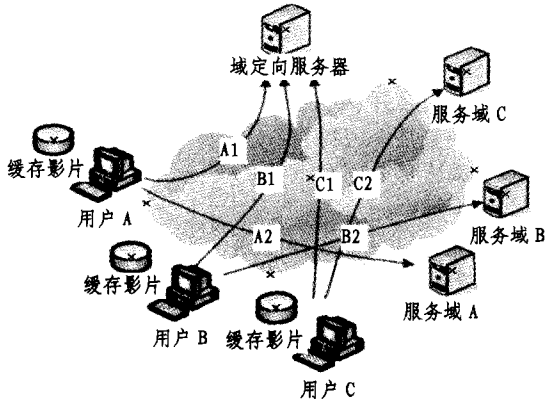


图 1 分布式点播系统的体系结构

BitTorrent 模型是一种被广泛使用的 P2P 文件共享模型,该模型在世界上得到广泛的实际使用,其文件共享、文件分发等功能的可行性和高效性得到了充分的检验。本文的目标就是要在宽带环境中基于成熟的 BitTorrent 模型实现一个实用的、保障 QoS 的 P2P 系统——分布式 VOD 系统。

2 系统设计方案

定义 服务域 是本系统提供正常服务的最小部署单位,由归属于该服务域的管理服务器、内容服务器、索引服务器,以及被重定向到该服务域的用户构成。每个服务域都形成一个相对完整的点播子系统,各个服务域之间形成备份关系,使系统具有自恢复功能。

接入域 用户接入到点播系统,获取服务的域。

域定向服务器 负责为用户分配接入域。域定向服务器实时监测各服务域的有效性和负载状况,根据临近原则和负载均衡原则为用户分配可用的接入域。

内容管理服务器 对服务域内影片的发布进行管理。

内容服务器 提供原始影片资源。

索引服务器 提供影片缓存信息和用户列表信息。

如图 1 所示,整个系统由多个服务域构成。单个服务域的体系结构如图 2 所示,每个服务域相对对立。同时,多个服务域之间互为备份,当某个服务域失效,用户将被重定向到其它服务域,从而克服单点失效的问题。服务域可以动态进行增加和撤出,各服务域内的内容服务器也可以动态增加和撤出。根据系统负载情况动态而实时地进行服务器的扩展,从而保证系统的服务质量。

用户首先需要访问域定向服务器。域定向服务器根据临近优先、负载均衡等原则为用户分配一个接入域。然后用户向该服务域的内容管理服务器发送点播请求,内容管理服务器返

*)国家发改委中国下一代互联网(CNGI-04-12-1D)。**彭永祥** 硕士研究生,主要研究领域为 P2P、操作系统。

回该域索引服务器，并根据域内内容服务器的负载情况，为该用户分配一个内容服务器。然后用户连接索引服务器，报告自己缓存的影片信息并获取点播的用户列表信息。最后用户从其它在线用户或者内容服务器获取数据，并开始视频播放。用户优先从其它用户获取数据，仅当用户无法从其它用户获取到足够数据，即无法从其它用户得到 QoS 保证的情况下，才向内容服务器请求数据。因此，在这个系统中，内容服务器作为后备服务器存在，是用户点播服务 QoS 的保障服务器。

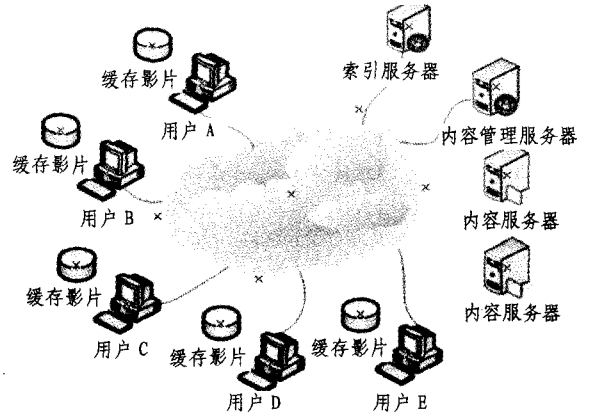


图 2　单个服务器域的体系结构

用户加入系统后，将根据一定策略进行影片内容的缓存。用户在每次登录到系统后，及时地将自己缓冲的影片信息报告给索引服务器，并作为种子为其它用户提供数据上传服务。由于用户对影片的缓存能力是有限的，因此应当有一个淘汰算法，将不经常被其它用户访问的影片淘汰掉。

3　性能分析

模型参数及说明：

N_{max}：最大用户容量；

N_s：内容服务器个数；

W_s：内容服务器出口带宽，假设所有内容服务器出口带宽相同；

R：影片平均码率；

$x(t)$：在 T 时刻，系统中点播影片 M 的普通用户数；

$y(t)$：在 T 时刻，系统中缓存影片 M 的种子数；

λ：请求的到达率，我们假设请求的到达分布模型是泊松分布；

μ：用户的上传带宽，我们假设所有用户的上传带宽均相同；

c：用户的下载带宽，我们假设所有用户的下载带宽相同，并且 $c \geqslant \mu$，$c \geqslant R$；

θ：点播用户放弃点播的速率；

γ：种子离开系统的速率，即拥有该影片副本用户停止点播的速率；

η：文件分享率，即能够提供上传带宽的比例，η 取值在 [0，1] 之间。

3.1　系统最大容量

假设用户接入系统后都在进行点播操作。极值情况下，系统用户的下行带宽与上行带宽达到平衡，有如下公式：

$$N_s * W_s + \eta * N_{max} * \mu = N_{max} * c$$

$$N_{max} = N_s * W_s / (c - \eta * \mu) \tag{2-1}$$

分析公式(2-1)可知，

· 系统的最大容量与内容服务器的服务能力和参数 η、参数 μ 成正比，而与参数 c 成反比。当 $\eta * \mu > c$ 时，N_{max} 为无穷。

· $\eta = 0$ 时，即所有用户不能提供上传，或者说有用户进行点播请求时，无法从系统中的普通用户获得数据，而只能从内容服务器获得数据，此时 $N_{max} = N_s * W_s / c$，系统退化成传统 C/S 模型，系统的容量完全由内容服务器的性能决定。这种情况最可能发生在系统的初始阶段。系统用户量少，系统可用副本少，可提供上传带宽用户少。

公式(2-1)是一种理想状态下的计算模型，对系统容量和部署具有一定指导意义。

3.2　简单流模型

下面对系统动态特性进行分析。当系统用户数量达到一定量时，用户间的带宽之和将远远大于服务器提供的带宽，因此在下面的分析中忽略内容服务器的带宽影响。

参数 η 被用于表示文件共享的有效率。在本系统中，一个正在点播的用户同时为其它用户提供数据上传。如果下载带宽没有受限制，那么系统总共的上传速率可以表示为 $\mu(\eta x(t) + y(t))$；如果下载带宽受限制，那么总共的上传速率为 $\min\{cx(t), \mu(\eta x(t) + y(t))\}$。当用户量很大时，为了获得系统的马尔可夫描述，我们假设一些用户在很小的时间间隔 δ 内变成种子的概率为 $\min(cx, \mu(\eta x + y))\delta$。

参数 θ 是用户离开系统的速率。用户可能在点播未结束时离开系统，也会因为点播完而变成种子。我们假设每个用户独立地在一段时间后离开系统，这段时间按指数分布，期望为 $1/\theta$。在流模型中，点播用户的离开速率为 $\min\{cx(t), \mu(\eta x(t) + y(t))\} + \theta x(t)$。

参数 γ 是种子离开系统的速率。我们假设种子呆在系统中的时间呈指数分布，期望为 $1/\gamma$，γ 会对整个系统的性能产生影响：如果降低 γ，那么下载时间就会缩短，因为有更多的种子在系统中存在。我们简单认为 γ 是个固定常数。

现在，我们开始描述在上述模型中参数 x 和参数 y 的变化，用户数目的变化在确定的流模型中的公式(3-1)如下：

$$\begin{aligned} \frac{dx}{dt} &= \lambda - \theta x(t) - \min\{cx(t), \mu(\eta x(t) + y(t))\} \\ \frac{dy}{dt} &= \min\{cx(t), \mu(\eta x(t) + y(t))\} - \gamma y(t) \end{aligned} \tag{3-1}$$

其中 $x(t)$ 和 $y(t)$ 不能为负数。

为了分析系统的稳定性，我们让

$$\frac{dx(t)}{dt} = \frac{dy(t)}{dt} = 0$$

代入公式(3-1)，得到

$$\begin{aligned} 0 &= \lambda - \theta\bar{x} - \min\{c\bar{x}, \mu(\eta\bar{x} + \bar{y})\} \\ 0 &= \min\{c\bar{x}, \mu(\eta\bar{x} + \bar{y})\} - \gamma\bar{y} \end{aligned} \tag{3-2}$$

其中 $\bar{x}$ 和 $\bar{y}$ 分别是 $x(t)$ 和 $y(t)$ 处于稳定状态下的值。

假设下载速度受限制，$c\bar{x} \leqslant \mu(\eta\bar{x} + \bar{y})$，通过解公式(3-2)，我们得到如下解：

$$\bar{x} = \frac{\lambda}{c(1 + \frac{\theta}{c})} \quad \bar{y} = \frac{\lambda}{\gamma(1 + \frac{\theta}{c})} \tag{3-3}$$

假设上传带宽受限，$c\bar{x} \geqslant \mu(\eta\bar{x} + \bar{y})$，通过解公式(3-2)，得到如下解：

$$\bar{x} = \frac{\lambda}{v(1 + \frac{\theta}{v})} \quad \bar{y} = \frac{\lambda}{v(1 + \frac{\theta}{v})} \tag{3-4}$$

其中 $1/v = 1/\eta(1/\mu - 1/\gamma)$。

令 $1/\beta = \max\{1/c, 1/\eta(1/\mu - 1/\gamma)\}$，公式(3-3)和公式(3-4)合并为：

$$\bar{x}=\frac{\lambda}{\beta(1+\frac{\theta}{\beta})} \quad \bar{y}=\frac{\lambda}{\gamma(1+\frac{\theta}{\beta})} \tag{3-5}$$

为了计算稳定状态下用户的平均下载速率，我们使用Little如下公式：

$$\frac{\lambda-\theta\bar{x}}{\lambda}\bar{x}=(\lambda-\theta\bar{x})T$$

等式中 T 是平均下载时间，$\lambda-\theta\bar{x}$ 是下载完成的平均速率。$\frac{(\lambda-\theta\bar{x})\bar{x}}{\lambda}$ 是下载用户变成种子的比率。使用公式(3-5)可以很容易得出

$$T=\frac{1}{\theta+\beta} \tag{3-6}$$

公式(3-6)反映出该点播系统有如下特点：

• 平均下载时间 T 与用户到达率 λ 无关，即服务质量与用户到达率无关，因此该点播系统的扩展性非常好。

• 当分享率 η 增加时，下载时间 T 减少，这是因为用户之间能更有效地共享文件。

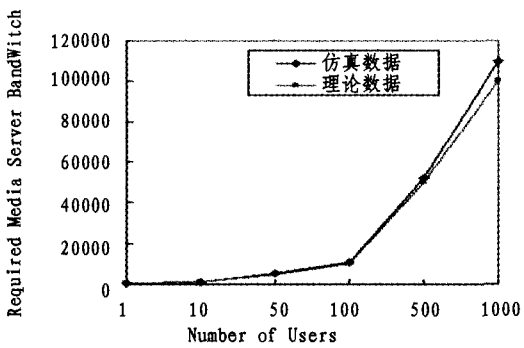
• 当 γ 增加时，T 增加，因为 γ 越大意味着系统中的种子更少；对用户的激励机制会影响参数 γ，用户点播完后如果鼓励他们留在系统，那么系统中就会有更多的种子存在。

• 通常情况下，用户的下载带宽 c 大于用户的上传带宽 μ。为了分析性能，我们假设 $c=\infty$，然而公式(3-6)表明平均下载速度不总是由用户的上传带宽决定的。实际上，如果种子的离开速率 γ 小于 μ，那么下载带宽 c 将会决定网络的性能(即使 c 可能远远大于 μ)。这也体现了P2P网络的优越性。

• 当 $\eta=0$ 时，这种情况意味着点播用户不进行数据交换，而仅仅从种子获取数据。如果 $\gamma<\mu$，那么 $T=1/c$；如果 $\gamma>\mu$，种子的数目将会减少至只有内容服务器作为种子存在。这种情况在系统用户数少时发生。当 $\eta>0$ 时，不论 γ 值如何，系统将会达到稳定状态。因此，点播用户在点播的同时给其他用户提供上传是非常重要的。即使文件的共享率不高，但它对保持系统活跃起着重要的作用。从公式(3-6)，我们可以看到 η 对系统的性能非常重要。

4 实验结果

在仿真实验中，我们首先用工具GT-ITM生成一个具有两个层次的网络拓扑，分别代表核心路由层和边缘路由层。核心路由层由1个域组成，它包含4个核心路由节点；边缘路由层由12个域组成，每3个域附属于一个核心路由节点。边缘路由层的每个域中平均包含8个边缘路由节点，每个边缘路由节点代表一个网络区域。网络拓扑中的带宽定义如下：核心路由节点之间、核心路由节点和边缘路由节点之间的带宽为1000Mb，边缘路由节点之间的带宽为100Mb。



$\eta=1, R=400$kbps, $c=400$ kbps, $\mu=300$ kbps

图3 服务器带宽占用情况

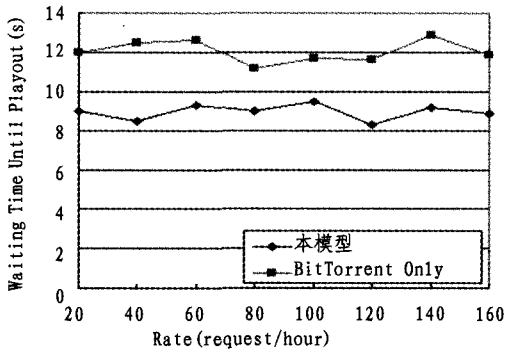
生成一个服务域，该服务域提供一个内容服务器，服务域附加在核心路由节点上，出口带宽为1000Mb。

其次，生成40000个节点并随机附加到边缘路由层的96个路由器上，节点的上行、下行带宽作为参数来指定。

首先对不同情况下服务器的带宽占用情况进行分析(图3)。从40000个节点中依次随机选取不同数码节点加入系统并请求服务，同时对内容服务器带宽占用情况进行监控。通过对比分析，可以看出实验数据和理论数据两者基本一致。

然后对点播过程中节点开始播放的平均等待时间进行评测，并与仅使用BitTorrent模型实现的点播系统进行比较。在每次实验过程中模拟VOD点播过程。从40000个节点中依次随机选取节点加入系统请求服务，加入过程服从参数为 λ 的泊松分布，λ 的取值作为本次实验参数来指定。每个节点均从节目的初始位置开始请求数据，节点加入系统后缓存本次实验接收到的数据，从而可作为后续加入系统的其他节点的候选服务节点。节点在加入后60分钟内不离开。

图4给出了在不同请求到达率条件下的开始等待时间。从图中可以看出，参数 λ 的变化对平均等待时间影响不大，这也验证了模型分析中的结论。



$\eta=1, R=400$kbps, $c=400$ kbps, $\mu=300$ kbps

图4 不同请求到达率下的点播开始等待时间

结论 基于P2P的视频点播系统能充分地利用客户端闲置资源，缓解服务器压力，适合用于大规模的视频点播。本文提出的点播系统，结构灵活，易于部署和扩展，能够满足大规模视频点播应用的需求。文中进行的性能分析对使用本系统结构进行大规模点播系统实现和部署提供了一定的理论依据。

参考文献

1 Fox G. Peer-to-peer networks. Computing in Science & Engineering, 2001,3(3):75～77

2 Parameswaran M, Susarla A, Whinston A B. P2P networking:an information sharing alternative. Computer, 2001,34(7):31～38

3 Deshpande H, Bawa M, Garcia-Molina H. Streaming live media over a peer-to-peer network:[Technical Report, CS-2001-31]. Stanford University, 2001

4 Tran D, Hua K, Do T. Zigzag:An efficient peer-to-peer scheme for media streaming. In:Proc. of the IEEE INFOCOM 2003. New York:IEEE Computer and Communications Societies, 2003. 1283～1293

5 Hefeeda M, Habib A, Botev B, et al. PROMISE:A peer-to-peer media streaming using collectcast. In:Proc. of the ACM Multimedia 2003. New York:ACM Press, 2003. 45～54

6 Do T, Hua K, Tantaoui M. P2VoD:Providing fault tolerant video-on-demand streaming in peer-to-peer environment. In:Proc. of the IEEE ICC 2004. Paris:IEEE Communications Society, 2004. 1467～1472

7 Chou P,Padmanabhan V, Wang H, et al. Distributing streaming media content using cooperative networking. May 2002

8 Yang X,Veciana G. Service capacity of peer to peer networks. In:IEEE INFOCOM, March 2004

9 Epema D H J, Sips H J, Pouwelse J A, et al. The bittorrent p2p file-sharing system:Measurements and analysis. February 2005

10 Cohen B. Incentives build robustness in bittorrent. In:Workshop on Economics of Peer-to-Peer Systems, May 2003

11 Calvert K, Doar M, Zegura E. Modeling Internet topology. IEEE Communication Magazine, 1997,35(6):160～163