

P2P技术在图书馆联盟云服务中的应用研究*

□ 杨贵福 黄秀林 牛宝乐 / 东北师范大学计算机科学与信息学院 长春 130024
隋会民 刘万国 / 东北师范大学图书馆 长春 130024

摘要: P2P技术的数据流量在当代因特网传输中具有重要地位,我们提出基于云服务的P2P资源离线下载模式,由应用服务器负责P2P下载,用户通过下载中心把完成下载的资源传输到本地。在P2P下载期间,用户不必运行P2P客户端程序,也不必开启本地计算机。这一模式能减少重复下载热点资源对Intranet公共出口带宽的堵塞,减少漫长等待稀缺资源的过程困扰下载用户,避免热点资源的网络流量突增导致其他程序性能受到影响,解决了传输瓶颈,提高联盟用户下载资源的速度,节省了成本。

关键词: P2P, 云服务, 下载, 离线, 图书馆联盟

DOI: 10.3772/j.issn.1673—2286.2012.03.006

1 引言

图书馆联盟有几十个、上百个图书馆,有几十万、上百万用户,这些用户通过互联网经常重复下载资源,尤其是重复下载各种热门资源,给各个成员馆的网络增加了很大的负担。联盟云服务平台上的资源很多,包括自建资源、购买的在线资源、网上免费资源等,但还是不能满足联盟用户的需求,需要从联盟外下载大量的数据。联盟内部的Intranet与Internet连接或中国教育科研网与公众网的连接带宽较小^[1],有效传输常在100-200Kb/s,而Intranet内部带宽通常为100Mb至1Gb/s,有效传输常可超过10Mb/s。这样就会导致在用户下载量增多时

不能满足联盟成员馆用户同时大量下载的需求。

P2P^[2]即对等互联网技术(Peer-to-peer),它有着改变整个互联网基础的潜能,目的是让所有上网用户之间自由的、不受主服务器控制的信息交流。它的数据流量已经超过http协议和ftp协议,成为因特网中最大的数据流量。P2P技术为用户传输大量的影音多媒体和大型的软件上传下载的同时,也为ISP,尤其是联盟中心带来巨大的负担^[3]。

根据下载流量统计,P2P下载资源的数量巨大,而这一部分流量并不总能由P2P协议自动引导至Intranet,导致重复占用宝贵的公共出口带宽。众多用户下载相同资源或删除已下载的同资源以前,会

在所有这些用户的硬盘中存储未完成的临时文件^[4]。同一个Intranet中的用户在经过一段时间以后,此临时文件会趋向于一致^[5],因而不重复占用公共出口带宽,但是仍然重复占用用户硬盘空间。在下载稀缺资源时,由于种子用户在线时间较短且不可预测,所以下载用户需要长时间开机,使P2P软件处于等待下载状态。这耗费了电力资源,也带来下载用户的不便。热点资源的下载,由于P2P网络的特性,正在下载的用户越多,其下载速度越快,所以可能造成在某个时段突发流量暴增,导致下载用户的其他应用程序响应缓慢,并占用下载用户全部带宽。下载稀缺资源或热点资源所造成问题虽然不同,但是原

* 文章系吉林省科技厅“建立国外开放获取科技期刊资源利用平台的对策研究”(项目编号:20090630)、教育部基金项目《基于国家〈信息安全等级管理办法〉构建高校图书馆信息安全保障新体系》(项目编号:08JA870003)研究成果之一。

因是相同的,即种子用户或上传用户必须与下载用户同时在线。相同的原因导致,如果上传用户意欲为他人提供稀缺资源,不得不延长P2P客户端运行和联网的时间。

2 P2P网络的优势

■ P2P网络的一个重要的目标就是让所有的客户端都能提供资源,包括带宽、存储空间和计算能力。因此,当有节点加入且对系统请求增多,整个系统的容量也增大。这是具有一组固定服务器的C/S结构不能实现的,这种结构中客户端的增加意味着所有用户更慢的数据传输。

■ P2P网络的分布特性通过多节点上复制数据,也增加了防故障的健壮性,并且在纯P2P网络中,节点不需要依靠一个中心索引服务器来发现数据。在后一种情况下,系统也不会出现单点崩溃。

■ 当用P2P来描述Napster网络时,对等协议被认为是重要的,但是,事实上Napster网络取得的成就是对等节点联合一个中心索引来实现。这可以使它能快速并且高效地定位可用的内容。对等协议只是以一种通用的方法来实现这一点。

3 基于云服务的P2P离线下载模式

我们提出使用基于云服务模式帮助P2P用户实现离线下载的模式,整个系统由4个主要部分组成,如图1所示。

3.1 用户接口

下载用户本地计算机可以不必

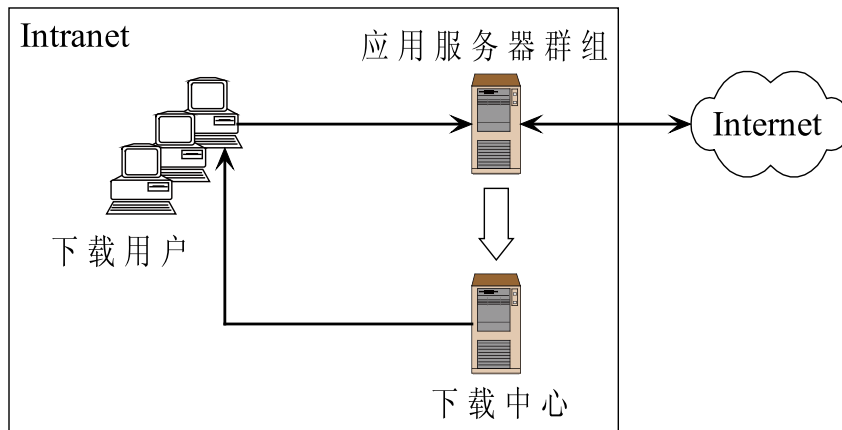


图1 基于云服务的P2P离线下载模式

安装P2P客户端软件,而通过浏览器访问云服务,并用这种方式完成下载任务的提交、下载状态检查、文件下载到本地、取消正在下载的任务、查看历史列表、搜索热点资源等功能。用户可以通过私有的PC机或图书馆的公共用机,也可以通过手机等移动设备完成上述功能。

如果用户非法或者存储容量超过限制,那么云拒绝提供服务。由于众多用户有大量相同的请求,因此云服务实际存储所消耗的空间,去除冗余以后,要大大低于用户需求的总和。

3.2 应用服务器

应用服务器由运行通用的P2P客户端的计算机群组及辅助程序组成。辅助程序由云服务接收到用户接口发来的下载任务请求,转交给P2P客户端群组,由后者负责真实P2P下载和共享资源。

在用户提交任务以后,可以关闭计算机或手机的网络,不必运行P2P客户端程序,而由应用服务器负责真实的P2P下载。在此期间,用户本地并无临时文件产生,

避免了存储空间的冗余使用;同时,只有应用服务器在进行真实的P2P下载,所以没有重复使用公共出口带宽的现象,凡下载相同资源的任务,将只产生一份临时文件和流量。

对于稀缺资源,下载用户不必开机等待,而由应用服务器负责守候种子用户上传;希望共享稀缺资源的用户,只需把资源上传,由应用服务器代为共享。

3.3 状态监控

状态监控子系统是由时间驱动的消息发送程序。在Unix系统下,我们使用cron每隔定长时间请求应用服务器报告各任务的进度,把结果存储在数据库中。当用户通过云服务查询进度时,把该结果通知用户。由于查询进度并非极其频繁,频度可以远低于1分钟,所以对应用服务器负载几乎没有影响,而对用户查询的响应,由其他的服务器负责。

作为扩展服务,也可以支持通过即时消息、手机短消息、电子邮件通知用户P2P下载的完成。

3.4 下载中心

当任务P2P下载完成,会自动移至下载中心,用户可以通过云服务把资源传输到他的本地计算机。这一过程用户可以选择合适的时间,以全部独占用户带宽的方式完成,既不会对用户其他的程序造成可能的性能影响,也不会占用公共出口的带宽。

下载中心可以选择高带宽高性能的服务器承担。相比之下,瓶颈受到公共出口带宽的限制,应用服务器的性能不必超出一台普通PC太多。下载用户也可以在已完成P2P热门资源列表中直接选择下载到本地,而不必执行P2P过程,因为此时资源已经在下载中心。

由于下载中心与下载用户均在Intranet之内,因此传输资源速度容易达到或超过10Mb/s。

3.5 容量限制、身份认证、超期管理

容量限制、身份认证、超期管理都是基于数据库管理的。当用户提交下载请求时,总检查其占用空间是否超过容量限制,如果超过则拒绝请求。用户每次登录,在提交下载任务或查询下载状态前,必须通过身份认证。匿名用户仅能下载已完成P2P下载的资源,而不能提交自己的下载任务。如果资源超过一段时间没有用户下载,也没有用户提交相同的P2P下载任务,则表

明这是冷门资源,可以根据存储空间的限制选择性删除。

4 实验结果

通过以上分析,对基于云服务的P2P离线下载模式有了更深入的了解。通过P2P和云服务的技术优势,将两者有机地结合起来,运用在图书馆联盟中实现各成员馆用户离线下载的模式,为联盟中的用户下载解决了难题。为了对基于云服务的P2P离线下载模式进行实践,测试它的性能,我们做了大量的实验,实验结果如图2所示,为某个用户已下载的资源列表。

我的下载		下载完成		
当前用户	younggift			
剩余空间	5.15 GB			
>> 首页				
>> 新建任务				
>> 正在下载				
>> 下载完成				
>> 修改资料				
>> 退出				
		文件名	文件大小	删除 下载
		[英国广播公司.自然世界 蚂蚁帝国].BBC.Natural.World.Empire.of.the.Desert.Ants.2011-AVI-夏末秋字幕组.mkv	724.78 MB	删除 下载到本地
		[亚洲台:勇闯西伯利亚].AsiaTV.Siberian.Odyssey.2011-HDTV.MiniSD-TLF.mkv	478.07 MB	删除 下载到本地
		[国家地理频道:中情局的秘密实验].National.Geographic.Channel.CIA.Secret.Experiments.2008-DVDRip.XviD-SViNTO.avi	349.90 MB	删除 下载到本地
		[英国广播公司.密码 第一集 数字].BBC.The.Code.1of3.Numbers.2010-AVI-夏末秋字幕组.mkv	705.74 MB	删除 下载到本地
		Bertrand Russell - Proposed Roads to Freedom (1918).pdf	431.82 kB	删除 下载到本地
		Bertrand Russell - The Conquest Of Happiness.pdf	14.19 MB	删除 下载到本地
		[中国全史:中国野史].翟文明扫描版.pdf	15.55 MB	删除 下载到本地
		[中国全史:中国秘史].翟文明扫描版.pdf	15.52 MB	删除 下载到本地
		[中国全史:中国绝史].翟文明扫描版.pdf	19.53 MB	删除 下载到本地
		[中国全史:中国通史].翟文明扫描版.pdf	15.51 MB	删除 下载到本地
		[极品飞车9:最高通缉].Need.For.Speed.Most.Wanted-SPITFIRE.rar	411.59 MB	删除 下载到本地
		Principles.of.Computer.Organization.and.Assembly.Language.Using.the.Java.Virtual.Machine.,Juola.,PH.,2007.pdf	2.64 MB	删除 下载到本地
		[我的奋斗]罗永浩.清晰扫描版.pdf	7.53 MB	删除 下载到本地
		[中国文明与世界:汤因比的中國观].(日)山本新.秀村京二編.扫描版.pdf	16.12 MB	删除 下载到本地

图2 实验结果

5 结论

本文提出的P2P资源的云服务模式,能避免重复占用Intranet的公共出口带宽和存储资源;支持用户离线下载,因而有助于下载稀缺

资源时需长时间等待种子的用户,也可利用稀缺资源提供共享;用户在下载进行过程中,不必开机运行P2P客户端,也不必占用网络带宽,可以选择方便的时间从下载中心迅速传输到本地。

这一模式节省Intranet公共出口带宽,也节省了下载用户的能源和时间。用户把除提出愿望以外的业务交给云服务处理,本地机从前更多地负责计算转变为云服务的接口。

参考文献

- [1] 高勇,李子木,吴建平.CERNET高速缓存服务的构建和性能分析[C]//第九届CERNET学术年会,2002.
 [2] KHAN J I, WIERZBICKI A. Foundation of Peer-to-Peer Computing [J]. Elsevier Journal of Computer Communication, 2008, 31(2).
 [3] 陈常嘉.P2P网络内容共享与流量测量研究[D].北京交通大学,2005.
 [4] BUFORD J F, YU H, LUA E K. P2P Networking and Applications [M]. Morgan Kaufmann, 2008.
 [5] 管磊.P2P技术揭秘——P2P网络技术原理与典型系统开发[M].清华大学出版社,2011.

作者简介

杨贵福,男,东北师范大学计算机科学与信息学院副教授。E-mail: young@nenu.edu.cn
 隋会民,男,东北师范大学图书馆副研究馆员。E-mail: suihtm@nenu.edu.cn
 刘万国,男,东北师范大学图书馆研究馆员、常务副馆长。E-mail: liuwg@nenu.edu.cn

Research on Cloud Service Scheme of P2P Resource Offline Downloading

Yang Guifu, Huang Xiulin, Niu Baole / School of Computer Science and Information Technology, Northeast Normal University, Changchun, 130024
 Liu Wanguo, Sui Huimin / Northeast Normal University Library, Changchun, 130024

Abstract: The traffic flow of P2P is at an essential position in the current Internet communication. We propose a scheme based on cloud service to provide offline P2P resource downloading, P2P downloading by application servers before users transmit the resource downloaded to local computer. In the P2P downloading period, users need neither running the P2P client, nor power on their local computers. This scheme contributes on reducing the shock on Intranet public connection by reduplicative hot resource, on decreasing the duration which bothers users when waiting for absent resources on avoiding the precipitate increasing of traffic flow influent which impacts other application performance.

Keywords: P2P, Cloud service, Download, Offline, Library consortia

(收稿日期: 2012-02-17)

业界动态

北京今年将建300个数字文化社区

本报北京3月1日电 (记者余荣华) 300万册电子书、1万种电子期刊、2000种中华文化视频、2万场讲座及专题视频……北京市文化局“文化惠民”日前正式启动了“数字文化社区”工程,今后,北京市民在社区里就能享受上述数字化文化信息资源。

“数字文化社区”依托高清交互平台,结合互联网无线技术,创建多媒体、跨平台、多终端的文化信息资源共享平台,并整合利用首都图书馆、艺术院团、文化共享工程等文化信息资源,使目前通过图书馆系统才能查阅到的文化信息资源和只能在剧院看到的演出更广泛地传播。

据悉,北京市文化局今年计划在全市300个社区中建设“数字文化社区”平台,今后逐年增加,预计“十二五”期间完成全市所有社区的平台建设。

(来源: <http://culture.people.com.cn/GB/87423/17271121.html>, 查询日期: 2012-03-12)