

试卷分发模型

丁亚涛

(安徽中医学院医药信息工程学院, 合肥 230031)

摘要: 基于传统的 C/S 模型, 提出一种新的试卷分发模型。该模型充分利用网络带宽, 允许从客户机下载试卷分组, 从而减轻服务器的负载。实验结果表明, 该模型的试卷下载时间少, 解决了考试系统并发组卷的瓶颈问题, 使系统能运行于低效网络中。

关键词: C/S 模型; P2P 技术; 带宽; 瓶颈

Paper Distribution Model

DING Ya-tao

(College of Medical Information Engineering, Anhui University of Traditional Chinese Medicine, Hefei 230031)

【Abstract】 This paper proposes a new paper distribution model based on traditional C/S model. This model makes full use of the network bandwidth and allows to download papers packet from other client, thereby effectively reduces the load of server. Experimental result shows that this model pays less time to download paper, it solves the problem of bottleneck when creating paper at the same time, makes the system running on inefficient network.

【Key words】 Client/Server(C/S) model; Peer-to-Peer(P2P) technology; bandwidth; bottleneck

1 概述

P2P(Peer to Peer)原理的基本点在于“互助和平等”和“非中心化”^[1-2], 它是一种广义意义上的 C/S 扩散模型。在满足本地下载需求的同时, 需要承担提供下载服务, 从而最大程度地利用有限带宽。

传统的考试系统包括组卷、试卷分发、考试、监测和阅卷等模块, 是基于一个中心点(服务器)的考试模型, 由服务器承担试卷分发和下载的全部任务, 对于大容量的图文试卷分发, 容易出现并发下载的瓶颈, 特别是对于低效能网络和远程分散式考试环境等更容易导致系统崩溃, 而利用 P2P 原理构建的一种新的试卷分发模型将具有极大的应用价值和前景。

通过对大文件分发技术的研究^[3-5], 提出一种基于 P2P 原理的新的网络考试系统试卷分发模型, 该模型从管理的角度可以分为多种形式, 其中包括集中式、部分分散式和完全分散式。本文重点研究集中式试卷分发模型, 其他 2 种模型做了简要说明。

2 模型设计

2.1 模型简介

设服务器 S, 考生工作站 $W_i(i=1\sim m)$, 每套试卷 P 分成 n 个分组 $P_i(i=1\sim n)$ 。系统模型如图 1 所示(只标出 3 个考生工作站)。

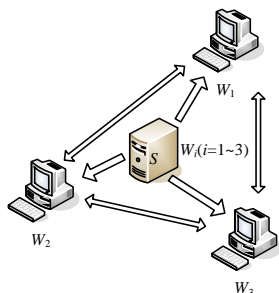


图 1 系统模型

服务器 S 根据组卷参数设置, 在开考前生成试卷分组 $P_i(i=1\sim n)$, 并对所有分组采用数字签名; 服务器 S 提供第 1 个试卷分组 P_1 的下载服务, 从第 2 个分组开始, 考生工作站自由选择下载目标站点。

考生工作站 $W_i(i=1\sim m)$ 分别对应开始下载试卷分组 $P_i(i=1\sim n)$, 服务器 S 建立下载记录列表用于集中式管理考生工作站的下载服务, 其中包括每个工作站的名称、IP、已经下载的试卷分组登记等信息。考生工作站通过查阅该表获取可下载试卷分组的的目标站点。

(1)假设 $m>n$

具体下载流程如下:

1)第 1 批试卷分组下载如表 1 所示。

考生机 W	下载点	试卷分组 P	异常备选
W_1	S	P_1	-
W_2	S	P_2	-
\vdots	\vdots	\vdots	-
W_i	S	$P_{(i\%n)}$	-
\vdots	\vdots	\vdots	-
W_m	S	$P_{(m\%n)}$	-

其中, $m\%n$ 表示 m 对 n 求余数, 当 $m=n$ 时等于 n。

从表中可以看出, 对于第 1 批试卷分组, 所有考生机 W 都从服务器 S 下载, 其中, W_i 下载 $P_{(i\%n)}$ 试卷分组。除了服务器 S 外没有备选的下载点。

第 1 批试卷分组下载完成后, 由于都保存了一份试卷分组, 因此考生机也可以提供下载服务, 第 2 批试卷分组的下载情况如表 2 所示。

作者简介: 丁亚涛(1970—), 男, 讲师、硕士, 主研方向: 计算机网络, 数据库

收稿日期: 2009-06-13 **E-mail:** yataoo@126.com

表2 第2批试卷分组下载 C/S 关系

考生机 W	下载点	试卷分组 P	异常备选
W_1	W_2	P_2	S
W_2	$W_{2\%n}$	$P_{2\%n}$	S
\vdots	\vdots	\vdots	S
W_i	$W_{(i+1)\%n}$	$P_{(i+1)\%n}$	S
\vdots	\vdots	\vdots	S
W_m	$W_{(m+1)\%n}$	$P_{(m+1)\%n}$	S

在第2批分组的下载中，除了服务器 S 可以作为下载点外，每个考生机都可以作为下载点，例如考生机 W_1 从 W_2 下载试卷分组 P_2 ，考生机 W_i 从 $W_{(i+1)\%n}$ 下载试卷分组 $P_{(i+1)\%n}$ ，考生机 W_m 从 $W_{(m+1)\%n}$ 下载试卷分组 $P_{(m+1)\%n}$ 。如果下载异常，那么所有考生机都可以从服务器下载。

依此类推，得到第3批试卷分组下载情况表如表3所示。

表3 第3批试卷分组下载 C/S 关系

考生机 W	下载点	试卷分组 P	异常备选
W_1	W_3	P_3	S, W_2
W_2	W_4	P_3	S, W_2
\vdots	\vdots	\vdots	\vdots
W_i	$W_{(i+2)\%n}$	$P_{(i+2)\%n}$	$S, W_{(i+1)\%n}$
\vdots	\vdots	\vdots	\vdots
W_m	$W_{(m+2)\%n}$	$P_{(m+2)\%n}$	$S, W_{(m+1)\%n}$

从第3批试卷分组开始，每个考生机除了算法指定的下载点外，将拥有越来越多的备选下载点，例如， W_1 除了可以从 W_3 下载试卷分组 P_3 外，还可以从 S, W_2 下载 P_3 ，这是因为在第2批下载中， W_2 已经下载了 P_3 。

依此类推，其他批次分组的下载不再给出，表4给出了最后一批的下载情况。

表4 第n批试卷分组下载 C/S 关系

考生机 W	下载点	试卷分组 P	异常备选
W_1	W_n	P_n	全部
W_2	W_{n-1}	P_n	全部
\vdots	\vdots	\vdots	\vdots
W_i	$W_{(i+n-1)}$	$P_{(i+n-1)}$	全部
\vdots	\vdots	\vdots	\vdots
W_m	$W_{(m+n-1)}$	$P_{(m+n-1)}$	全部

最后一批的下载点是全部机器，包括服务器，每个考生机只要下载最后一个分组，因此，选择范围是全部，该分组在其他机器上都有。

(2) $m=n$

分组数和机器数相等，以上关系表仍然适用，只是第1个分组下载时， W_i 正好下载 P_i 。

(3) $m < n$

在该情况下，分组比较小，机器数目少。第1个分组下载中， W_m 下载 P_m ， $P_{m+1} \sim P_n$ 没有被下载，所以，需要下载 $n-m+1$ 批所有分组才会有下载点。第 j 批分组下载情况如表5所示。

表5 第j批试卷分组下载 C/S 关系

考生机 W	下载点	试卷分组 P	异常备选
W_1	W_j	P_j	S
W_2	W_{j+1}	P_j	S
\vdots	\vdots	\vdots	S
W_i	$W_{(i+j)\%n}$	$P_{(i+j)\%n}$	S
\vdots	\vdots	\vdots	S
W_m	$W_{(m+j)\%n}$	$P_{(m+j)\%n}$	S

2.2 异常处理规则

异常包括连接异常和试卷异常。如果搜索到的下载点不能下载需要的试卷分组，则选择异常备份的下载点，备份下载点随着下载流程的进行，其范围将越来越大，当下载最后一个试卷分组时，异常备选的机器是其他所有客户机和服务器。

异常的处理需要在算法中加以考虑，通常为每个考生机建立一个下载列表来管理异常，该列表中记录可以下载和不能下载的站点的情况，以供下批次选择。

2.3 耗时估算

基于 C/S 的试卷下载和基于 P2P 的试卷下载耗时估算需要考虑试卷容量、分组数、点对点带宽，由于是局域网，因此连接时间可以忽略不计。

假设试卷容量 P ，试卷分组数 N ，点对点下载带宽 C ，考生工作站数 M 。

(1) 基于 C/S 模式

所有下载由服务器承担，每组试卷下载时间为 T ：

$$T_1 = \frac{P/N}{C/M} \times N \quad (1)$$

$$T_1 = PM/C \quad (2)$$

(2) 基于 P2P 模式

$$T_2 = \frac{P/N}{C/M} + \frac{P/N}{C/2} \times (N-1) \quad (3)$$

T_2 包括 2 个部分：第 1 个分组下载时间和后 $n-1$ 个分组下载时间。在理想状况下，后 $n-1$ 个分组可以从其他工作站下载，享受 $C/2$ 带宽，每个工作站的带宽分成下载和提供下载，对于 P2P 模型，每个 Peer 需要提供下载服务给其他 Peer。

2 种模型的总体下载时间差为

$$\begin{aligned} \Delta t &= T_1 - T_2 \\ &= \frac{P/N}{C/M} \times (N-1) - \frac{P/N}{C/2} \times (N-1) \\ &= P \times (N-1) \times \left(\frac{1}{C/M} - \frac{1}{C/2} \right) \end{aligned} \quad (4)$$

从式(4)可以看出，当 $M > 2$ 时，基于 C/S 模型的下载时间将大于基于 P2P 的下载时间。若增加 P2P 连接数 L ，则会提高下载的效率。

2.4 解决的问题

服务器在承担第 1 批分组的下载任务后，不再主动承担下载服务，只有在异常情况下才提供后备下载支持，因此，服务器瓶颈问题得以解决。

可见，设计良好的算法，若能充分利用逐渐增多的下载点，试卷的下载效率必将大大地提高，对于效率较低的网络，该模型将发挥重要的作用。该模型的效率规模为 $O(n^2)$ ，可以同时存在 $n(n-1)$ 个并发执行的事务。

3 实现要素

本文模型的具体实现将涉及以下 4 个问题：

(1) 安全性。安全问题包括试卷的合法性和试卷的完备性。试卷的合法性和完备性可以通过服务器对试卷的数字签名来实现，数字签名采用 MD5 生成试卷的摘要信息，考生工作站从服务器下载简短的 MD5 摘要文件(耗时忽略不计)，然后对试卷进行验证即可。

(2) 发现算法。基于局域网的考试软件系统中“发现算法”比较简单也易于规划。集中式的模型可以利用服务器来管理下载表，提供可下载信息，对于分散式的模型可以采用邻近查找算法，例如 IP、机器编号均可以实现。对于分散式模型，发现是完全自主的。每个试卷分组的目标群是不一样的，后面的分组对应的下载目标群要大些，也更容易被发现。

传统模型中的算法是固定式的算法，每个工作站下载的分组有对应的下载对象列表，列表的最后一个备选对象是服务器。

(3) 资源管理。试卷分组从服务器分散到各个考试工作站

后,再由考试工作站从目标群下载到本地再组合。资源是先分散再聚合,分散的时间只有一个分组时间,而聚合是不间断的。

(4)特殊性。试卷分发的结果是由考试工作站自主组卷验证并考试,从考试管理的角度还要获取考试状态信息,因此,必要的信息反馈对于服务器的监控必不可少。另外,试卷分组的重用可以在算法中考虑,下载的试卷分组可以在下一批考试中得到应用,从而免去再次从服务器下载的时耗,提高系统效率。重用性和时效性可以通过考试软件参数的设置来统一管理。

4 模拟实验

4.1 实验环境和参数设置

本文以教学机房为实验环境,其中包括服务器(Windows 2000 Server SP4)、5台考生工作站(P IV 2.4 GHz, 512 MB 内存)、大容量内科学图文题库、TLink10/100 MB 交换机等。

4.2 实验模拟

实验中将题库分成5个分组,每个分组26.9 MB,测试模块分为Server端和Peer端,Server端负责发送下载指令和监控下载信息,Peer端根据Server端指令选择下载方式,并从Server端获取其他Peer的连接参数,包括IP地址、用户名以及访问口令等。实验数据单位为s, $W_i(i=1\sim5)$ 代表peer端(考生机), $P_i(i=1\sim5)$ 代表分组,具体情况如表6、表7所示。

表6 从服务器下载分组 (s)

考生机	连接时长	试卷分组下载时长					小计
		P_1	P_2	P_3	P_4	P_5	
W_1	0.012	11.281	11.391	11.297	11.578	8.141	53.688
W_2	0.013	11.547	11.328	11.265	11.375	8.422	53.937
W_3	0.013	9.094	11.328	11.015	11.297	9.969	52.703
W_4	0.013	3.422	11.000	11.516	12.390	11.531	49.859
W_5	0.023	11.344	11.359	11.047	11.328	8.469	53.547
总计	0.073	46.688	56.406	56.140	57.968	46.532	263.734

表7 P2P方式下载分组 (s)

考生机	连接时长	试卷分组下载时长					小计
		P_1	P_2	P_3	P_4	P_5	
W_1	连接	0.016	0.028	0.026	0.020	0.033	28.107
	下载	7.125	3.047	7.593	6.579	3.640	
W_2	连接	0.031	0.032	0.012	0.019	0.031	30.031
	下载	3.985	7.078	5.078	7.640	6.125	
W_3	连接	0.026	0.030	0.031	0.021	0.012	23.854
	下载	3.047	2.937	7.063	4.093	6.594	
W_4	连接	0.036	0.036	0.030	0.031	0.018	44.885
	下载	11.813	3.656	6.484	11.344	11.437	
W_5	连接	0.012	0.016	0.031	0.032	0.047	35.216
	下载	3.062	9.141	8.140	6.625	8.110	
总计		29.153	26.001	34.488	36.404	36.047	162.093

表6是从服务器集中下载分组的方式,只需要连接服务器一次,5台机器同时访问服务器,每个分组的下载时间较长,约11s左右。表7是P2P方式下载分组,其中,每个分组下载都需要一个连接时间,这是因为分组位置是随机的,不固定,每个分组的下载时间大大缩短的原因是:Peer的带宽是基本空闲的,表中下载时间为7s~12s之间的均是从服务器下载,例如表7中 W_4 的 P_1, P_4, P_5 分组下载。

由于连接时间和机器性能的差异,实验实现的是理想模型的一部分,从服务器下载的比率要高于理想模型,但已经很好地减小了服务器的负载和利用peer端的空闲带宽。

比较这2种方式,P2P方式下载要节约较多时间。为了深入研究,笔者还设计了Peer端同时多进程下载试卷分组,

经测试2个进程的总体下载时间为130.297s,3个进程的总体下载时间为125.672s,由此可见,并发连接数的提高可以继续减少下载时间。

实验的分组大小为27.5MB,从实验实际效果来看,基于C/S模式的试卷下载时间明显超过本文方式的下载时间。笔者尝试增加机器的台数至20台,效果类似,考虑到篇幅有限,具体数据这里不再给出,具体代码省略。实验的程序界面如图2、图3所示。



图2 Server端界面



图3 Peer端界面

对于试卷分组数 N ,P2P连接数 L ,站点数 M ,测试程序均可以随机设置。

5 结束语

本文模型在低效的教学机房实现通畅的数据传输,为复杂的机上考试系统提供了重要的参考模型,也为充分利用教学硬件资源提供了很好的理论和技术支持。下一步研究的重点为:P2P技术在面向大容量多媒体题库及考试系统中的效率和优化,面向Web的实现模型,以及如何提高和改进发现算法。

参考文献

- [1] 邢小良. P2P技术及其应用[M]. 北京: 人民邮电出版社, 2008.
- [2] 金海, 廖小飞. P2P技术原理及应用[J]. 中兴通讯技术, 2007, 13(6): 1-5.
- [3] 朱渊萍, 陈素芬. 基于P2P技术的文件分发结构的研究[J]. 南昌工程学院学报, 2007, 26(1): 50-53.
- [4] 何伟, 薛素静, 孔梦荣. 一种实用的P2P文件系统访问控制框架[J]. 郑州大学学报: 工学版, 2006, 27(3): 93-97.
- [5] 李葳, 张华. 大型网络文件分发技术探讨[J]. 微型机与应用, 2004, 23(7): 84-86.

编辑 陆燕菲