

基于 P2P 的大规模考试平台实例^①

梁 武

(广东行政学院 信息技术教研部, 广州 510053)

摘 要: 在大规模考试中, 由于同时参与考试的人员众多, 考生登录时的试卷分发和交卷时的试卷回收常常因为并发访问量过大而造成服务器崩溃。P2P 以节点平等为核心, 淡化中心服务器的角色, 以高效的方式向需要资源的节点分发内容, 能很好的解决大规模考试中因过分依赖于中心服务器而造成的一系列问题。

关键词: 大规模考试; P2P 应用模式; P2P 协议

Large-Scale Examination Platform Based on P2P

LIANG Wu

(Department of Information Technology, Guangdong Institute of Public Administration, Guangzhou 510053, China)

Abstract: Because of the large numbers of the students, the server always dies down when the students download or upload their papers simultaneously in large-scale examination. However, the P2P application can share the data with others fast and stably without using server. It is a best choice to solve the problem of max data transmission among the nodes in the large-scale examination.

Keywords: large-scale examination; P2P application mode; P2P protocol

1 引言

所谓大规模考试是指在同一时段内有大量考生参与的考试, 通常指全省性或全国性的考试, 例如目前各省区的计算机水平考试、全国计算机水平考试和全国英语四、六级或 A、B 级考试等。从考试流程来看, 大规模考试的管理工作主要分为考前、考中和考后三个阶段, 具体如下:

(1) 考前: 该阶段主要完成试题的管理(如试题采集、录入和检测等)、考试平台的研发和考生数据的采集、汇总等工作, 主要由考试管理部门和具体实施考试的考试单位(常称为考点)共同完成。

(2) 考中: 在试题和考生数据均已准备完毕的情况下, 考生便可利用考试平台开始考试。考试平台主要完成试题的抽取及呈现、考试过程的跟踪和考试作答数据的采集、考试结果的评分等工作。

(3) 考后: 该阶段主要负责对考试结果进行分析和统计。

在以上三个阶段中, 只有考中的工作与考生规模密切相关。随着考生规模的不断增大, 需要传输的数据量会急剧增加, 考试平台要在同一时段内完成考生试题的分发(下载)、考试过程的跟踪和考试结果的回收等考试工作的难度会大幅度增加。为了解决该问题, 通常采用以下的平台架构:

(1) 评分和试题放在考试客户端: 考生的试题预先装在考试客户端上, 考生登录时直接从本地取试题, 无须服务器的配合, 同时也把评分功能放在客户端。这是在网络应用不太成熟的单机系统中常用的方式, 目前已基本不用。

(2) 评分和试题放在本地服务器: 开考前一段时间(例如开考前 30 分钟), 考试单位需预先从考试管理部门获取本次考试的所有试题并安装到本单位架设的本地服务器上, 考生直接登录到本地服务器并获取试题, 考试的结果也预先上传到本地服务器, 待全部考试结束后再由考试单位上送考试管理部门, 考试的评

^① 基金项目: 广东省教育厅资助项目(粤教高函[2009]10 号 D02)

收稿时间: 2010-09-13; 收到修改稿时间: 2010-10-19

分功能通常也放在本地服务器上。这是在大规模考试中因无法解决大批量数据传输问题而采取的折衷方法。目前,大多数大规模考试都采用该方式,例如笔者原来开发的广东省计算机水平考试和全国英语应用能力 A、B 级(广东省考区)考试等均采用该方式。在该方式中,试题被提前下载到考试单位的本地服务器上,这必然给考试管理工作带来某些不利影响:一是每次考试所用的试题量有限,不能发挥大题库的优势。因为试题需预先下载到本地服务器上,出于安全考虑,考试管理部门不可能一次性把题库中所有的试题都提供给各考试单位,只能采用一定的算法从题库中筛选部份试题给各考试单位,这必然大大减少了最后组卷时可用的试题量;二是增加考试单位的维护和管理成本。因为各考试单位需自行维护各自的服务器,稍有不慎便将影响考试的正常进行;三是试题的安全性难以保证。因为试题已提前到达各考试单位的服务器,必然为居心不良者提供破解试题的时机,容易造成试题提前泄露;四是考试结果预先保存在各考试单位本地服务器上,也容易造成考试结果被非法修改。

(3) 评分和试题放在远程服务器:试题放在考试管理部门架设的远程服务器上,考生登录时直接从该服务器上获取试题,解决了试题需预先下载时带来的各种问题。但显然,随着考试规模的增大,服务器的访问压力也将成倍增大,导致该方式目前只应用于规模不大的考试。

由此可见,在大规模考试中,为了确保考试试题及结果的安全性,最佳的解决方案是把考试试题保留在远程服务器上,考生登录时从远程服务器下载试题到本地考试终端,交卷时直接回收到远程服务器,在考试单位上不留任何痕迹,但这必然会因并发访问量过大而带来一系列问题,甚至造成服务器崩溃。因此,寻求一种既能发挥大题库的优势同时又安全、简便的考试平台架构,是大规模考试必须解决的问题。P2P 技术以其特有的“变态”下载功能,在文件下载、视频直播等领域应用广泛,笔者通过对 P2P 技术的深入研讨并应用在新版的广东省计算机水平考试平台上,成功解决了大规模考试所遇到的数据传输问题。

2 P2P原理简介

P2P(Peer-to-peer)是一种对等的网络结构,网络中所有节点均具有双重角色:既接受别人的服务,也为

别人提供服务,形成一个庞大的资源共享网络系统。由此可见,P2P 网络系统已经淡化传统意义中的服务器角色,因而即使系统中某些节点出现故障也不会像传统服务器出现故障那样对网络产生巨大的影响,确保了网络的稳定性。

如何组织网络系统中的各节点,是 P2P 协议的核心。通常按一定的组织结构对网络系统中的节点进行管理,目前主要有 DHT、树形和网状三种结构^[1]。DHT(分布式哈希表)的主要思路是以一个 128 位的哈希值作为各节点的标识,并保存能提供资源的前后节点,由此连接成一个环形结构;树形则采用二叉树来组织网络中的各节点,并按二叉树的遍历规则来访问各节点;网状结构则以动态、无序方式组织各节点,节点间有多条可达路径。

有了组织各节点的方法还不足以让 P2P 应用开始工作,因为刚开始时所有的节点互不相识,P2P 应用程序必须采用一定的方式把资源所在的位置(节点)告知其他节点,以便能知道对方,由此形成了不同类型的应用模式,目前,主要有两种:混合 P2P 和纯 P2P,如图 1 所示^[2]。前者保留了传统意义上的服务器,不过该服务器不会参与实际的数据下载,只负责对加入网络的节点进行管理,目的是提供有效的数据搜索服务。后者没有服务器,节点间通过广播方式完成数据搜索。两者各有优缺点,前者提供资源的集中式管理,能提供快速的搜索服务,但一旦该服务器出现故障,资源搜索服务便中止;后者因为没有服务器,不存在因服务器出现故障而无法搜索资源的问题,但必须以广播的方式搜索所需的资源,搜索效率远不如前者。

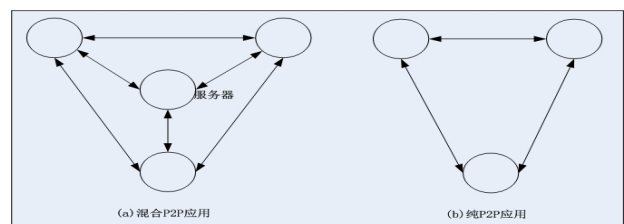


图 1 P2P 应用模式

一旦获取资源所在位置的信息后,节点便向对方发起 TCP 连接请求,以完成实际的资源传输。所传输的资源在传输前通常被分割成多个块,每个块有相应的标识及校验,节点在下载完所有的块后再重组成原来的数据,这与 TCP/IP 协议中对传输数据的拆包、重

组等方式相似。在传输过程中，节点可以只从某一个节点传输资源块，也可以同时从多个节点传输资源块。

由于 P2P 应用模式解决了传统 C/S 模式中过分依赖服务器的问题，从诞生时起即被迅速应用到不同的领域，主要包括文件下载、视频直播、VoIP、即时通讯、在线游戏、分布式海量存储等。这些领域都具有用户规模和传输数据十分庞大的特点，因而，在大规模数据传输的应用领域中，P2P 协议确实是个不错的选择。

3 应用实例

在计算机无纸化考试中，系统提供给考生的数据除了试题作答要求描述外，通常还包括一些试题素材文件，如图片等。不同模块（课程）的考试所提供的素材大小各不相同，有些很小甚至没有，但有些却很大，例如 Photoshop 考试。Photoshop 模块考试是广东省计算机水平二级考试的主要模块，该模块除了试题描述外还有大量的图片素材（每个考生平均有 50M 到 100M 不等）。随着考生规模的不断扩大（目前人数规模已在万人以上），考试中凸显的问题越来越多，主要集中在对远程服务器的并发访问上。从前面的分析可知，在大规模考试中，只有考生登录时取卷和交卷时试卷回收两个时段客户端需频繁访问远程服务器，进行大批量的数据交换（人数众多时经常造成服务器崩溃），其他时段需交换的数据量很小，因此，解决好这两个时段的数据传输问题即可解决大规模考试中所出现的服务器访问问题。笔者利用 P2P 原理，经过多次尝试，设计了基于 P2P 的大规模考试平台架构，如图 2 所示。

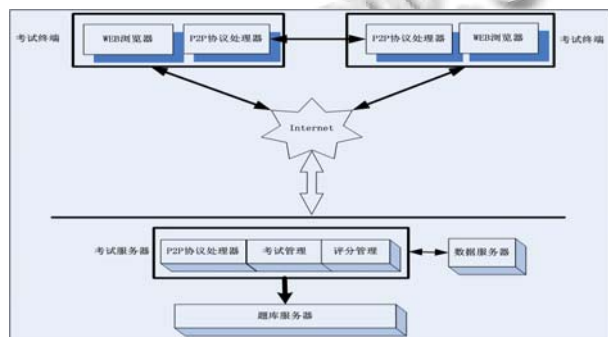


图 2 基于 P2P 的大规模考试平台架构

与单纯的 P2P 资源共享应用模式不同，基于 P2P

的考试平台除了提供资源共享外，还必须处理考生登录、交卷、评分、监控等与考试相关的管理任务，因此本架构采用了能对资源进行集中管理的混合 P2P 应用模式，分别设置了考试终端（节点）、考试服务器、数据服务器和题库服务器，各部分的具体功能如下：

(1) 考试终端

考试终端由传统的 WEB 浏览器和 P2P 协议处理器组成。前者基于 Web2.0 标准，提供一个试题显示及作答的 Web 操作环境；后者完成与 P2P 协议相关的数据处理，可被封装成运行于浏览器中的插件或独立运行于后台的服务，并通过消息机制与浏览器通讯。考试终端在考生登录时除了下载以 HTML 格式编写的试题作答要求描述外，还必须从考试服务器下载一份记录该考生考试所需素材的素材表（主要有“素材名称”和“素材标识”两个域）以及描述网络中各节点的节点信息表（相当于“电驴”中的“种子文件”）。素材表列出了要下载的各素材的文件名称及对应的资源标识，而节点信息表则提供了从哪里可以下载这些素材的节点信息。节点信息表采用 DHT 结构，具体如图 3 所示（图中的“b”表示二进制的“位”）。



图 3 节点信息表

在节点信息表中，每个节点用一个 128 位的数值来表示，该数值可以是系统生成的节点 hash 值，也可以是节点的 IP 地址。节点上所具有的资源保存在宽度为 8 位的“资源标识”域中。每个资源按一定的长度分割成多个数据块，数据块总数保存在 8 位宽度的“总块数”域中（实际上考试平台中只使用 7 位，因而一个资源最多可分割成 128 块），而每个数据块的大小则保存在“块大小”域中。某一时刻某一节点上不一定保存有某个资源的所有数据块，因而在节点结构表的最后设置了 128 位宽度的“块标识”域。“块标识”域中的每个二进制位代表所对应的一个数据块，如果节点中存在所对应的数据块，则该位取值“1”，否则取值“0”。通过遍历该域即可知道该节点能向其他节点提供哪些数据块，也可以知道该节点需要下载哪些数据块。另外，

为了方便查询拥有相同资源的其他节点,在该表中设置了“前节点”和“后节点”域,但这两个域的数据不是由客户端节点填写,而是由服务器填写。

节点信息表在客户端节点和服务器端各存一份。节点每下载完一个数据块除了更新本地节点信息表中“块标识”域的值外,还必须向服务器发送节点信息表更新报文,以通知服务器同步更新该域的值。节点在下载数据时尽量选择相同子网内的节点,以提高下载速度。

(2) 题库服务器

题库服务器保存了考试中所使用的所有试题,这些试题按一定的结构保存在大型数据库中。在保存试题时,仅保存试题的相关参数,而试题所用的素材,如图片等以文件的形式保存,以方便各节点下载。

(3) 数据服务器

数据服务器保存本次考试所用的所有试题及素材文件,供考生登录时下载。另外,考生交卷时的作答结果也上传到该服务器。因此,该服务器是考生在考试过程中进行数据上传和下载的最终对象。依照 P2P 协议原理,节点在下载资源时会按照一定的规则优先选择其他的节点,只有在所有的节点都没有所需的资源的情况下才从数据服务器上下载,因此即便考生规模巨大也不会产生数据服务器崩溃的现象。相反,同时登录的考生越多,从邻近节点找到自己所需资源的概率越大,下载的速度就越快,因而常被称为“变态”的下载方式,这正是 P2P 协议的精华所在。但考生交卷时数据最终要上传到该服务器,如果同一时段内有许多考生交卷,极有可能因访问量过大而造成该服务器无法正常响应,交卷操作会失败,考生的作答结果可能因此而丢失。为了解决该问题,本架构采用了“暂缓上传”和“代传”的机制。依照 P2P 资源共享原理,节点在上传数据时也可先尝试与服务器沟通,如果此时服务器很忙,则先把数据上传到还未交卷的其他节点上,再由其他节点在服务器空闲时代传给服务器,这相当于从最低层对服务器的并发访问量进行合理调控,解决大规模考试中常出现的因同时交卷人数过多而产生的“网络塞车”现象。另外,根据考生规模的大小适当增加平台中的数据服务器数量,从而扩大节点上传对象的选择范围,也能更有效地解决“网络塞车”现象。

(4) 考试服务器

考试服务器是该平台的核心,主要完成 P2P 协

议处理、考试管理、评分管理等与考试相关的功能。顾名思义,P2P 协议处理负责处理 P2P 协议的相关功能,主要包括资源的分割与重组、节点信息表的管理等。在考生登录前,考试服务器需把供考生下载的素材按指定的要求分割成多个数据块,并生成各数据块的数字签名,以便节点对所下载的数据块进行验证,确保数据块的完整性;考生交卷后,考试服务器在评分前需把各数据块重新组合成原来的资源,这就是 P2P 协议处理的资源分割与重组功能。在进行资源分割时,考试服务器还必须填写节点信息表,以记录各个资源的分割情况。在没有考生登录前,考试服务器无法了解网络中的其他节点,只保存提供资源下载的数据服务器节点,因此,节点信息表在初始状态时只保存数据服务器节点信息。但随着考生不断登录,新节点相继加入,节点信息表的内容变得越来越丰富,最终记录着网络中所有节点及其资源的相关信息。另外,节点每下载完一个数据块后便向服务器发送一个 UDP 报文,服务器收到该报文后随即更新节点信息表中相应域的值;一旦考生交卷成功,服务器也清除节点信息表中与该节点相关的所有记录,以确保节点信息表处于最新的状态,这些操作完全由 P2P 协议处理器的节点信息表管理功能来完成。

考试管理是考试服务器的核心模块,主要负责考试过程的控制,包括考生登录、组卷、交卷管理等。正规的考试通常有一定的考试时段要求,例如开考的时间、登录的最迟时间、登录的合法身份等,这些由考生登录管理完成。在考生登录验证通过后,考试管理将按照一定的规则从题库服务器取出该生的考试试题组成试卷,同时通知 P2P 协议处理器完成相应的处理,例如检查该生的资源分割情况、把该生所在的终端节点加入到节点信息表等。在考试过程中,考试管理模块还接收来自考试终端的通知报文,例如节点下载资源时发送的“块标识”域更新报文、交卷时发送的交卷报文等,考试管理模块会把这些报文请求转发给 P2P 协议处理器,最终由 P2P 协议处理器完成相应的处理。

评分管理采用基于插件的开发技术,各种题型的评分模块均以插件的形式加载到考试平台中,在增强系统可扩展性的同时也提高了系统的稳定性,使得评分工作变得更加高效可靠。

(下转第 116 页)

4 车载终端调试结果

车载终端的软硬件平台搭建好之后, 将要对系统的各个模块和系统整体功能进行调试。调试环境包括硬件和软件环境, 硬件环境包括车载终端设备、后台服务器和计算机等, 软件环境包括程序代码、后台服务器的系统管理软件等。

4.1 GPS 数据调试

测试 GPS 模块, 查看 GPS 接收数据指示灯是否在闪烁, 若在闪烁说明有数据收到。接着, ARM 处理器将接收到的 GPS 数据进行解析, 提取出经纬度、时间、日期、速度和标志位。解析之后的 GPS 数据通过 GPRS 网络传输到后台服务器, 其显示结果如图 6 所示。

```
10:04:44 收到数据: 2811.0937N11256.4228E1004411106090.00A
10:04:48 收到数据: 2811.0934N11256.4236E10044511106090.00A
10:04:52 收到数据: 2811.0932N11256.4240E10044911106090.00A
10:04:56 收到数据: 2811.0928N11256.4238E10045311106090.00A
10:05:00 收到数据: 2811.0926N11256.4232E10045711106090.00A
10:05:04 收到数据: 2811.0923N11256.4226E10050111106090.00A
10:05:08 收到数据: 2811.0923N11256.4227E10050511106090.00A
```

图 6 后台接收到的 GPS 数据

4.2 车载终端与后台服务器的调试

车载终端对 GPS 数据的传输, 对准确性要求较低, 采用的是 UDP 方式, 无需进行差错重传, 实现过程比较简单。但是, 像更新的报站语音、POS 机的消费金额等数据, 必须要保证准确、安全、可靠的传输, 因此, 这些数据的传输采用了面向连接的 TCP 方式, 并且, 加入了确认、差错重传机制。现通过在车载终端向后台服务器发送心跳包来进行测试, 其后台显示的数据结果如图 7 所示。

由以上的测试结果可以看出, 终端解析 GPS 数据正确, 发送的心跳包数据后台服务器能准确接收, 因此, GPRS 网络数据通信正常, 车载终端设备能稳定、可靠地工作。

```
接收协议包: 2A 2A 2A 06 00 0E 00 42 10 03 14 27 00 23 23 23
校验码: 27
数据长度: 6
消息 ID: BE
终端 ID: 42100314
回复标示: 0
接收协议包: 2A 2A 2A 06 00 0E 00 42 10 03 14 27 00 23 23 23
校验码: 27
数据长度: 6
消息 ID: BE
终端 ID: 42100314
回复标示: 0
```

图 7 后台接收到的心跳包数据

5 结语

本设计采用 SAMSUNG ARM9 处理器 S3C2416 及嵌入式 Linux 操作系统为核心, 结合先进的 GPRS、GPS 等技术, 制定了安全可靠的通信协议, 实现了车辆自动报站、数据无线传输、远程更新广告及报站语音、车辆实时监控、智能调度、正点考核等功能, 还预留了多个外部设备连接的接口, 具有一定的可扩展性。

系统投资小、收益快、成本低、实用性强, 很适合城市交通的智能化发展, 且已在部分大城市投入使用, 实际情况表明, 该产品性能稳定可靠, 改善了城市交通的管理, 具有巨大的市场价值。

参考文献

- 1 郭启军, 张浩然, 姜彬. 基于 GPRS 的嵌入式无线数据传输终端的设计. 计算机系统应用, 2008, 17(12).
- 2 王伟, 马洪连, 孙国晴, 丁男. 智能化公交车载终端的设计与开发. 仪器仪表学报, 2006, (S1).
- 3 何小卫, 王爱华, 马跃. 基于 GPRS 的 GPS 车载终端通信技术研究. 计算机应用, 2008, (11).
- 4 Abbott E, Powell D. Land-vehicle Navigation Using GPS. Proc. of the IEEE, 1999, 87(1): 145-162.
- 5 杜春雷. ARM 体系结构与编程. 北京: 清华大学出版社, 2003.
- 6 王学龙. 嵌入式系统 Linux 设计与应用开发. 北京: 清华大学出版社, 2001.

(上接第 112 页)

4 结语

为了解决大规模考试过程中庞大的数据传输问题, 人们采用了多种折衷的解决方法, 但这些方法同时也带来了在考试维护成本、考试结果的安全可靠等方面的一系列问题。基于 P2P 的大规模考试平台, 采用的是对等而“变态”的资源分享方式, 使用的人越多, 分享的速度越快, 使得大批量数据的传输问题迎刃而解, 解决了大规模考试中因考生规模庞大而导致的考生登录取卷和交卷时传统方式下常常出现的服务器崩

溃的问题。多次实践表明, 基于 P2P 的广东省计算机水平考试平台在考试维护、考试实施等方面远远优于传统的平台, 大大提高了大规模考试的效率。

参考文献

- 1 金海, 廖小飞. P2P 技术原理及应用. 中兴通讯技术, 2007 (12).
- 2 Hwang SM. P2P Protocol Analysis and Blocking Algorithm. [2010-9-2]. <http://www.springerlink.com/content/x9dc5ppkvhx1edl3/>