

# 网上招生数据传输模型的研究与实现

龚洪泉 姚 丹 钱乐秋

(复旦大学计算机系软件工程实验室 上海 200433)

**摘 要** 本文对网上招生模式中的数据传输问题进行了深入的研究和具体实现。针对数据传输的安全性、数据传输效率、及时的信息反馈等问题,提出了以四个控制文件为中心的 FTP 数据传输模型。然后围绕该模型,讨论了 FTP 的数据传输效率、数据安全性控制,以及控制文件的读写策略等问题,证明该模型在安全性、数据传输效率和信息反馈的时效性方面都满足了网上招生在数据传输方面提出的要求。

**关键词** 网上招生 招办 院校 FTP 控制文件

## RESEARCH AND IMPLEMENTATION OF DATA TRANSFER MODEL IN NET-BASED RECRUITING SYSTEM

Gong Hongquan Yao Dan Qian Leqiu

(Department of Computer, Fudan University, Shanghai 200433)

**Abstract** In this paper, some problems related to data transfer in net-based recruiting system are addressed. According to the data transfer requirements, which focus on security, high efficiency and timely information feedback, a 4-control-file-centered FTP data transfer model is presented. Then the problems about data transfer efficiency, data security control, and the read-write strategy of control files are discussed in this model and proofs of how the requirements are addressed by this model are also presented.

**Keywords** Net-based recruiting system Recruiting administration College FTP Control file

### 1 引 言

网上招生,就是通过计算机网络传递高考招生工作的各种信息,用管理信息系统(MIS)控制整个招生流程的一种新的招生模式。其目的在于提高招生工作的信息化程度,提高招生工作的效率,减轻招生工作的通信联络负担。它改变了传统的以纸质档案材料为唯一介质的招生模式,所有考生档案材料、院校招生计划等信息都实行电子化,录取工作中的所有信息都以电子数据的形式通过网络在招办与各院校之间传递。

与一般的网络数据传输不同,网上招生的数据传输具有一定的特殊性。首先,由于高考录取数据属于机密信息,所以必须保证网上招生数据传输的安全性。其次,由于高考招生工作的时效性,数据传输一旦出现差错或者出现延误必将影响整个招生工作的进度,所以必须保证数据传输的效率。最后,为了监控招生工作的进度,招办端和院校端向对方发送数据后,希望能及时地了解对方是否已经收到这些数据,所以必须要有一种自动的机制来实现信息的及时反馈。

为了确保网上招生工作的顺利进行,实现其预期的目标,我们对网上招生数据传输模型进行了深入的研究和具体实现,最终形成了以四个控制文件为中心的 FTP(File Transfer Protocol)数据传输模型(图 1)。该模型在安全性、数据传输效率和信息反馈的时效性方面都满足了网上招生在数据传输方面提出的

要求。

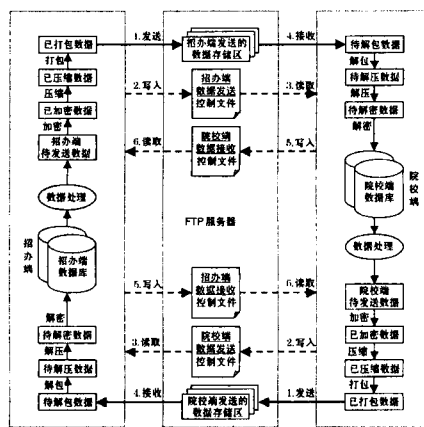


图 1 网上招生数据传输模型

### 2 FTP 服务器上数据的存放结构

在网上招生系统中,招办端与院校端之间的数据传输关系是招办端要与所有的院校交换招生录取数据,而每个院校只与

收稿日期:2003-01-10。龚洪泉,博士生,主研领域:软件工程,软件复用,构件库管理系统,基于构件、构架的软件开发方法、系统开发模式。

招办端交换数据,院校之间没有数据传输关系。在该模型中,FTP服务器是数据传输的中转站,是整个数据传输的核心部分。所有发送到FTP服务器上的数据都存放在数据存储区的目录中,该目录的组织结构如图2所示。

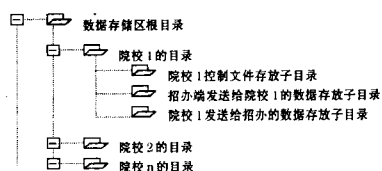


图2 FTP数据存放结构图

从图2可以看出,在数据存储区,每个院校都有一个目录用来存放与自己院校相关的数据。招办端可以访问所有院校目录下的所有数据,各个院校只能访问自己目录下的数据。每个院校拥有相同的子目录结构,包括:

(1) 院校x控制文件存放子目录

该目录用来存放招办端与院校x进行数据传输时所使用的四个控制文件,它们的作用是为数据传输提供及时的信息反馈。其中:

- 1) 招办端数据发送控制文件:记录招办端发送给院校x的所有数据信息;
- 2) 院校端数据接收控制文件:记录院校x接收到招办端的所有数据信息;
- 3) 院校端数据发送控制文件:记录院校x向招办端发送的所有数据信息;
- 4) 招办端数据接收控制文件:记录招办端接收到院校x的所有数据信息。

(2) 招办端发送给院校x的数据存放子目录

该目录用来存放招办端发送给院校x的所有数据。

(3) 院校x发送给招办端的数据存放子目录

该目标用来存放院校x发送给招办端的所有数据。

### 3 网上招生数据传输流程

在该模型中,数据传输的基本流程如下(详见图1中标有数字编号的这些部分):

- (1) 发送方将数据发送到FTP服务器上的数据存储区中;
- (2) 发送方将这些数据的相关信息写入“数据发送控制文件”;
- (3) 接收方读取发送方的“数据发送控制文件”,得知有新的数据需要接收;
- (4) 接收方接收新的数据;
- (5) 数据接收完成后,接收方将新接收的数据的相关信息写入“数据接收控制文件”;
- (6) 发送方读取接收方的“数据接收控制文件”,得知接收方已经接收了数据。

### 4 FTP数据传输

FTP是网络上使用最多,技术最成熟的文件传输协议。之所以采用FTP技术,是因为它有如下的优越性:

首先,使用FTP技术,一个普通的服务器通常可以同时支持几百个FTP数据传输连接,足以满足网上招生通信连接的数量要求。

其次,FTP协议属于TCP/IP协议集中应用层的协议,控制信息十分简单,在数据传输时对数据进行的与协议相关的操作很少,能保证高效率的数据传输,从而可以满足网上招生在数据传输效率方面的要求。

再次,FTP技术的日趋完善,在进行数据传输时性能十分稳定。这种稳定性正是网上招生所需的,因为数据传输是网上招生工作的核心部分,其稳定性关系到整个网上招生工作的进程。

在该模型中,除了用FTP来传输一般的招生录取信息外,还加入了提供信息反馈的控制文件。含控制文件的FTP数据传输的基本步骤如下:

#### 第一 登录FTP服务器

需要提供服务器IP地址、端口号、用户名和密码等信息。登录成功后,就与FTP服务器建立了数据传输连接。

#### 第二 获取控制文件

每个用户登录FTP服务器后,只需获取四个控制文件中的两个文件。一个是对方的“数据发送控制文件”,记录对方发送给自己的数据的相关信息;另一个是对方的“数据接收控制文件”,记录对方接收自己发送过去的的数据的相关信息。

#### 第三 分析控制文件

分析对方的“数据发送控制文件”,可以知道对方又发送了哪些新的数据自己还没有接收;分析对方的“数据接收控制文件”,可以知道自己发送过去的的数据对方是否已经收到。

#### 第四 接收数据

如果分析控制文件后发现有了新的数据还没有接收,则使用FTP协议将数据从相应的目录中接收下来。

#### 第五 发送数据

如果自己有了新的数据需要发送出去,则使用FTP协议将数据发送到FTP服务器上相应的目录中。

#### 第六 更新控制文件

对每个用户来说,需要更新的是四个控制文件中的另外两个文件。一个是在数据接收完成后,将这次接收的数据的相关信息写入自己的“数据接收控制文件”;另一个是自己发送数据后,将这次发送的数据的相关信息写入自己的“数据发送控制文件”。这样对方在下一次分析控制文件的时候就会获得这些新的反馈信息。

#### 第七 断开FTP

为了合理利用网络带宽,减轻FTP服务器的负载,在不需要传输数据的时候,就断开与FTP服务器的连接。

### 5 数据传输的安全性控制

由于在数据传输过程中FTP协议本身并不对数据进行任何安全性方面的控制,为了确保网上招生数据的安全性,必须在使用FTP进行数据传输之前对数据进行安全性方面的控制。在该模型中,我们主要采用了四种基本技术来加强数据的安全性(详见图1中相关部分)。

#### 第一 数据加密

它是维护数据安全性的最基本方法。数据加密过程涉及到各种各样的加密算法,它们以一定的时间和空间代价来换取对数据的安全性保护。

#### 第二 数据压缩

一方面,由于网上招生系统中传输的数据大部分是数据库

表,其压缩比率通常都比较大。如果对这些数据进行压缩,就可以在很大程度上减少数据的传输量,为数据传输节省带宽。另一方面,数据压缩从本质上来说也是一种数据变换的算法,其结果也相当于又为数据进行了一次加密操作。

### 第三 数据打包

由于网上招生系统中每次传输的数据库表格都很多,经过数据压缩后,每个表格的体积就更小了。根据网络传输的知识我们知道,如果我们每次都传输这种单个的小表格,每传输一个就要为它添加一次数据传输控制头,无形中传输的数据总量就增多了,造成带宽的浪费。如果我们把多个数据库表格合并成一个大的数据包,就可以减少数据传输控制头的数量,从而提高总的数据传输效率。

### 第四 招办端与院校端相隔离的模式

在网上招生系统中,招办端保存了所有考生的原始数据,这些数据是对考生进行录取的依据。因为院校端接收到的数据都是由这些原始数据产生的,所以它们的安全性是至关重要的。所以,为了防止外部网络的攻击,我们采用了招办端与院校端相隔离的模式,即招办端与院校端不直接进行数据传输,而是用 FTP 服务器来作为数据传输的中转站。只有 FTP 服务器可以与外界通信,招办端通过内部网络与 FTP 服务器通信(详见图 1)。这样即使 FTP 服务器受到攻击,也不会对招办端的数据库形成任何威胁。另外,这种模式在 FTP 服务器受到攻击后,还能够方便地从招办端重新生成数据发送到 FTP 服务器上,以恢复 FTP 服务器上受到破坏的数据。

## 6 控制文件的读写问题

在该模型中,我们采用了带控制文件的 FTP 数据传输模式,控制文件的作用是为数据传输提供及时的信息反馈。然而控制文件的数目及其读写操作的控制都是需要仔细考虑的问题。

一种看似简单的方法是使用一个控制文件,招办端和院校端在发送和接收数据后都往这个文件中写入相关的信息。表面上看来,只操作一个控制文件,目标十分明确,但从本质上来说,它存在很大的缺陷,主要有两方面的问题:一方面,招办端和院校端都要对这个文件执行写入操作,必然会产生“同时写”的冲突,从而产生信息覆盖的问题,使得控制文件的信息不能反映招办端和院校端数据传输的实际情况;另一方面,招办端的信息和院校端的信息混合在同一个控制文件中,一旦数据传输出现问题,要对控制文件进行分析就十分不方便。

另外一种可能的方法是采用两个控制文件(图 3)。但不论是采用哪种形式,它们都不能完全解决上面提到的两个问题。

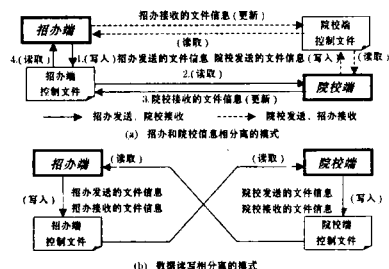


图 3 两个控制文件的读写关系图

为了解决这两个问题,在该模型中,我们使用了四个控制文件(详见图 1)。从图 1 可以看到,这四个控制文件的职责十

分清楚,每个文件记录一个方面的控制信息,把招办端和院校端的信息区分开了。而且,每个控制文件的读写关系也十分明确,只有一个人往它里面写入控制信息,也只有另外一个人读取它里面的控制信息。这样就不会出现几个人同时往同一个控制文件中写入信息的冲突,不会产生信息覆盖的问题。提取图 1 中与控制文件相关的部分,得到这四个控制文件的读写关系如图 4 所示。

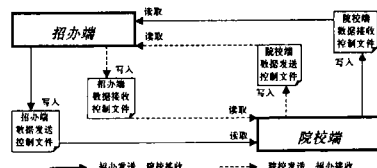


图 4 四个控制文件的读写关系图

## 7 模型的实现

从该模型的结构图(图 1)可以看到,其中的数据加密、压缩和打包部分,以及 FTP 数据传输部分与网上招生的业务流程是完全独立的。根据软件复用的思想,这些部分完全可以在其它需要类似功能的数据传输项目中使用。在实现该模型的过程中,我们把这些部分封装成了可复用的软件构件,其内核部分负责完成数据的发送和接收功能;中间层负责实现数据传输的安全性策略;外层结合控制文件实现数据传输相关信息的反馈。我们对该构件的中间层和外层功能设置了配置参数,软件复用者可以根据实际需要通过对参数设置对构件的功能进行裁剪。

## 8 结束语

本文针对网上招生模式在数据传输方面的特殊要求,提出了以四个控制文件为中心的 FTP 数据传输模型。然后围绕该模型,讨论了 FTP 的数据存放结构,FTP 数据传输的优越性,带控制文件的 FTP 数据传输操作步骤,数据的安全性控制,以及控制文件的读写策略等问题,证明了该模型在安全性、数据传输效率和信息反馈的时效性方面都满足了网上招生在数据传输方面提出的要求。最后,我们对该模型的实现问题进行了讨论。

通过对该模型的深入研究,使我们网上招生的数据传输特点和各种网络传输技术有了更深刻的理解,为网上招生工作的顺利完成提供了技术上的保障,也为以后类似的网络传输研究工作提供了借鉴。

## 参 考 文 献

- [1] Andrew S. Tanenbaum. Computer Networks, (3rd Edition). Prentice Hall, 1996.
- [2] Hafeedh Mili, Ali Mili, Sherif Yacoub, Edward Addy. Reuse-Based Software Engineering, Techniques, Organization, and Controls. John Wiley & Sons, Inc. 2002.
- [3] Roger S. Pressman. Software Engineering, A Practioner's Approach, (4th Edition). McGraw-Hill, 1997.
- [4] 张海藩, 软件工程导论, 清华大学出版社, 1998.
- [5] 郑人杰等, 实用软件工程(第二版), 清华大学出版社, 1997.
- [6] 周之英, 现代软件工程(新技术篇), 科学出版社, 2001.