

ICS 01.140.20
A 14

C A D A L 项 目 标 准

CADAL 20603—2012

数字资源传输规范

Digital Resources Transmission Specification

第一稿

2012-05-08 发布

2012-05-09 实施

CADAL 项目管理中心 发 布

目 次

| | |
|-------------------------|----|
| 前言 | 33 |
| 引言 | 34 |
| 1 范围 | 35 |
| 2 规范性引用文件 | 35 |
| 3 术语和定义 | 35 |
| 3.1 超文本传输协议 | 35 |
| 3.2 文件传输协议 | 35 |
| 3.3 互联网 | 35 |
| 3.4 客户/服务器模型 | 35 |
| 3.5 传输控制协议 | 35 |
| 3.6 网络协议 | 36 |
| 4 数字资源传输规范 | 36 |
| 4.1 HTTP | 36 |
| 4.2 FTP | 37 |
| 参考文献 | 39 |
| 图 1 HTTP 客户/服务器架构 | 36 |
| 图 2 FTP 服务连接图 | 37 |

前 言

《数字资源发布标准规范集》包括以下 3 方面的内容：

- 第 1 部分：数字资源发现规范；
- 第 2 部分：数字资源访问规范；
- 第 3 部分：数字资源传输规范。

本标准是其中的第 3 部分。

本部分是由大学数字图书馆国际合作计划(CADAL)项目管理中心提出并归口。

本部分起草单位：数字图书馆教育部工程研究中心 CADAL 项目门户组。

本部分起草人：尹彦飞、张寅、边科。

引 言

数字资源的传输涉及多种计算机技术、通信技术及多媒体技术，为了有效地传输、管理和应用数字资源，本规范针对数字图书馆对数字资源传输的需求，在 CADAL 数字图书馆项目实践的基础之上，编制了数字资源的传输规范。

数字资源传输规范

1 范围

本标准确立了 CADAL 项目数字资源传输的规范。
本标准适用的数字资源指印刷型文献的数字化衍生物。
本标准适用于需要对数字资源进行传输的服务。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本(包括所有的修改单)适用于本文件。

RFC 675 *Specification of Internet Transmission Control Program*

3 术语和定义

3.1 超文本传输协议 HyperText Transport Protocol 缩写: HTTP

超文本传输协议是一种详细规定了浏览器和万维网服务器之间互相通信的规则，通过因特网传送万维网文档的数据传送协议。

3.2 文件传输协议 File Transfer Protocol 缩写: FTP

文件传输协议是 TCP/IP 网络上两台计算机传送文件的协议，FTP 是在 TCP/IP 网络和 Internet 上最早使用的协议之一，它属于网络协议组的应用层。

3.3 互联网 Internet 缩写: Internet

互联网由多个计算机网络相互连接而成，而不论采用何种协议与技术的网络。

3.4 客户/服务器模型 Client/Server Module 缩写: C/S

客户/服务器分别指通过参与一次通信的两个应用实体，客户方主动发起通信请求，服务器方被动地等待通信而建立的模型。

3.5 传输控制协议 Transmission Control Protocol 缩写: TCP

传输控制协议是一种面向连接(连接导向)的、可靠的、基于字节流的运输层(transport layer)通信协议，由 IETF 的 RFC 675 说明(specified)。

3.6 网络协议 Internet Protocol 缩写: IP

网络是能使连接到网上的所有计算机网络实现相互通信的一套规则,规定了计算机在因特网上进行通信时应当遵守的规则。任何厂家生产的计算机系统,只要遵守 IP 协议就可以与因特网互连互通。IP 地址具有唯一性,根据用户性质的不同,可以分为 5 类。

4 数字资源传输规范

本标准规定了数字资源传输的两种协议: HTTP 和 FTP。

4.1 HTTP

HTTP 的中文名为“超文本传输协议”,是目前互联网信息传递的基础之一,也是使用最为广泛的互联网协议。

HTTP 协议具有以下几个特点:

(1) 支持客户/服务器模式, HTTP 协议的客户/服务器提供服务的架构如图 1 所示。

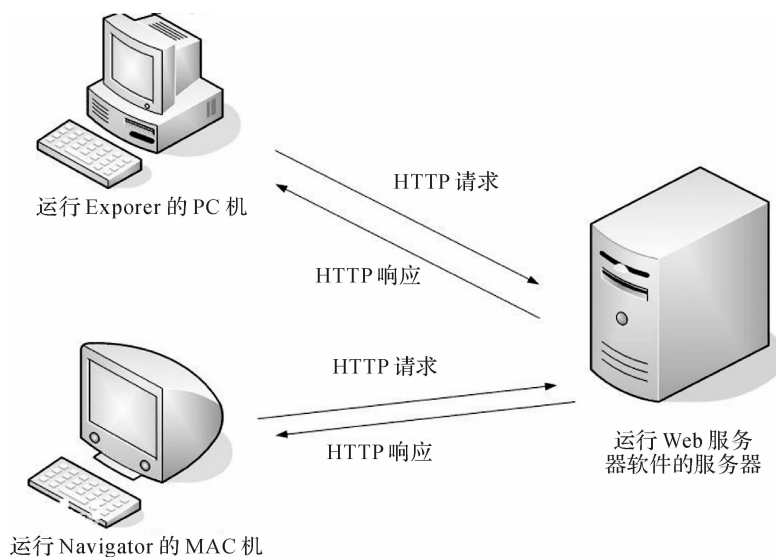


图 1 HTTP 客户/服务器架构

(2) 简单快速: 客户向服务器请求服务时, 只需传送请求方法和路径。请求方法常用的有 GET、HEAD 和 POST。每种方法规定了客户与服务器联系的类型不同。由于 HTTP 协议简单, 使得 HTTP 服务器的程序规模小, 因而通信速度很快。

(3) 灵活: HTTP 允许传输任意类型的数据对象。正在传输的类型由 Content-Type 加以标记。

(4) 无连接: 无连接的含义是限制每次连接只能处理一个请求。服务器处理完客户的请求, 并收到客户的应答后, 即断开连接。采用这种方式可以节省传输时间。

(5) 无状态: HTTP 协议是无状态协议。无状态是指协议对于事务处理没有记忆能

力。缺少状态意味着如果后续处理需要前面的信息，则它必须重传，这样可能导致每次连接传送的数据量增大。

4.2 FTP

文件传输协议 (File Transfer Protocol, FTP) 使得主机间可以共享文件。FTP 使用 TCP 生成一个虚拟连接用于控制信息，然后再生成一个单独的 TCP 连接用于数据传输。控制连接使用类似 Telnet 协议在主机间交换命令和消息。文件传输协议是 TCP/IP 网络上两台计算机传送文件的协议，FTP 是在 TCP/IP 网络和互联网上最早使用的协议之一，它属于网络协议组的应用层。FTP 客户机可以给服务器发出命令来下载文件、上传文件、创建或改变服务器上的目录。

FTP 有两种使用模式：主动和被动。主动模式要求客户端和服务端同时打开并且监听一个端口以建立连接。在这种情况下，客户端由于安装了防火墙会产生一些问题，所以创立了被动模式。被动模式只要求服务器端产生一个监听相应端口的进程，这样就可以绕过客户端安装防火墙。

1. 主动模式 FTP

一个主动模式的 FTP 连接(见图 2)建立要遵循以下步骤：

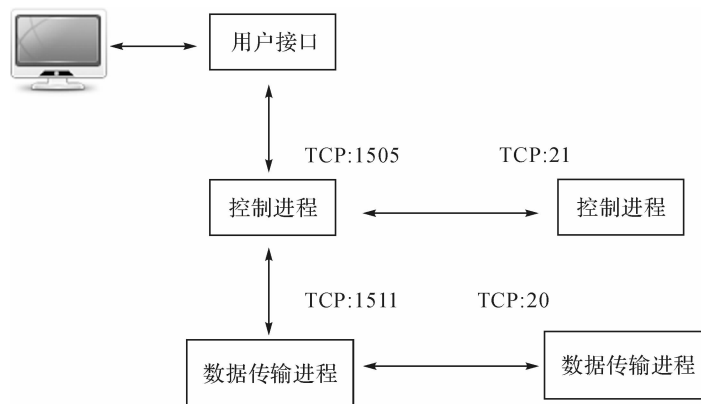


图 2 FTP 服务连接图

(1) 客户端打开一个随机的端口(端口号大于 1024，在这里，我们称它为 x)，同时一个 FTP 进程连接至服务器的 21 号命令端口。此时，源端口为随机端口 x ，在客户端，远程端口为 21，在服务器。

(2) 客户端开始监听端口 $(x+1)$ ，同时向服务器发送一个端口命令(通过服务器的 21 号命令端口)，此命令告诉服务器客户端正在监听的端口号并且已准备好从此端口接收数据。这个端口就是我们所知的数据端口。

(3) 服务器打开 20 号源端口并且建立和客户端数据端口的连接。此时，源端口为 20，远程数据端口为 $(x+1)$ 。

(4) 客户端通过本地的数据端口建立一个和服务器 20 号端口的连接，然后向服务器发送一个应答，告诉服务器它已经建立好了一个连接。

2. 被动模式 FTP

为了解决服务器发起到客户的连接的问题，人们开发了一种不同的 FTP 连接方式。这就是所谓的被动方式，或者叫作 PASV，当客户端通知服务器它处于被动模式时才启用。

在被动模式 FTP 中，命令连接和数据连接都由客户端发起，这样就可以解决从服务器到客户端的数据端口的入方向连接被防火墙过滤掉的问题。

当开启一个 FTP 连接时，客户端打开两个任意的非特权本地端口 ($N > 1024$ 和 $N + 1$)。第一个端口连接服务器的 21 端口，但与主动方式的 FTP 不同，客户端不会提交 PORT 命令并允许服务器来回连其数据端口，而是提交 PASV 命令。这样做的结果是服务器会开启一个任意的非特权端口 ($P > 1024$)，并发送 PORT P 命令给客户端。然后，客户端发起从本地端口 $N + 1$ 到服务器的端口 P 的连接用来传送数据。

参 考 文 献

- [1] CERF V G, DALAL Y, SUNSHINE C. RFC 675 Specification of Internet transmission control program[J]. Journal of Pure & Applied Algebra, 1974, 38(1): 19-37.