

大数据量数据网络传输研究与实现*

涂平¹, 朱晓铃^{1,2}, 满旺^{1,2}

(1. 福州大学 福建省空间信息工程研究中心, 福建 福州 350002;

2. 厦门理工学院 空间信息科学与工程系, 福建 厦门 361024)

摘要: 针对电子政务中大量业务数据网络传输问题, 提出一种新的数据传输网络架构, 研究了即时数据传输与离线数据传输, 设计和实现了数据传输组件原型系统, 实现了点对点数据传输的基本功能, 并在内网与政务网中分别对 P2P 通信、离线上载、离线下载的传输性能进行了测试, 结果表明, P2P 通信控件具有较高的传输效率。

关键词: 电子政务; 数据共享; 网络传输

中图分类号: TP391.41

文献标识码: A

Design and realization of massive data Web transmission component

TU Ping¹, ZHU Xiao Ling^{1,2}, MAN Wang^{1,2}

(1. Fujian Spatial Information Engineering Research Center, Fuzhou University, Fuzhou 350002, China;

2. Department of Spatial Information Science & Engineering, Xiamen University of Technology, Xiamen 361024, China)

Abstract: Aiming at massive data network transmission problems in e-government affair, a new data transmission network architecture is proposed. A real-time and off-line data transmission model is studied. A simple data transmission prototype system is designed and realized, the P2P transmission function is achieved, and transmission rate are tested in administration network and internal network. Research shows that P2P transmission model has high reliability.

Key words: e-government affair; data sharing; network transmission

随着“数字福建”工程建设与政府信息整合、资源的交换与共享的深入, 各个机构内部产生的大量业务数据信息、基础数据库信息, 以及各种图像文件等需要及时地从机构的服务器或终端汇总到数字福建数据中心, 以便在各个机构的服务器或终端之间互传。由于传输的数据量巨大, 传输过程中经过的服务器及网络环节众多, 因此, 对数据传输的效率、可靠性、安全等方面提出了诸多要求, 同时, 现有应用需求对数据资源传输的要求也越来越高^[1]。而传统的文件传输工具已经无法胜任机构级应用环境的需求^[2-3]。为此, 快速、安全、稳定地实现大数据量网络传输技术的研究是一个亟待解决的问题。

本文依据数据传输端之间所处的网络环境, 提出了大数据量传输系统架构, 分别研究实现了即时数据传输与离线数据传输, 并进行了性能测试。

1 数据传输系统架构

依据开放互联参考模型, 本文设计的大数据量数据传输系统架构, 主要包括网络监测模块、数据传输模块、数据管理模块、数据处理模块, 其系统架构如图 1 所示。

(1) 数据处理模块。负责发送、接收数据, 根据数据的处理规则, 对数据进行一系列加密、解密、压缩、解压缩、指定编码、解码格式等操作, 并将其封装成若干 UDP 数据包(反之将若干 UDP 数据包组合)的内部模块。

(2) 数据管理模块。负责对经过数据处理后的待发送数据包或刚接收到的数据包进行统一管理分配, 负责数据的缓存管理、并发管理以及完整性管理。

(3) 数据传输模块。负责数据传输路径的管理, 通过对网络链路环境的判断来决定传输路径(直发、打通直发、通过服务器转发), 并实现数据包准确的收、发过程。

(4) 网络监测模块。负责监控本地与服务器连接超时, 报告服务器下线通知, 处理本地网络异常及对数据

* 基金项目: 国家科技支撑计划项目(2007BAH16B02)资助

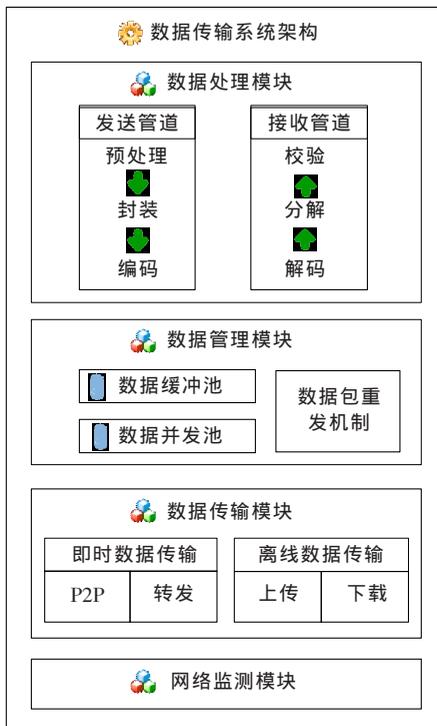


图1 数据传输系统架构

传输双方的网络带宽进行测试等功能。

2 即时数据传输

即时数据传输要求通信双方均保持在线状态,经过一次“握手”之后,方能进行数据传输。若通信双方“握手”不成功,发送方会自动将数据发往服务端,由服务端将数据转发给接收方。同时为应对意外断线的情况发生,提供了断线续传的功能,即使在传

输过程中意外中断,用户仍可从断点处重新开始传输,大大节省了时间和带宽资源。其工作原理如图2所示。

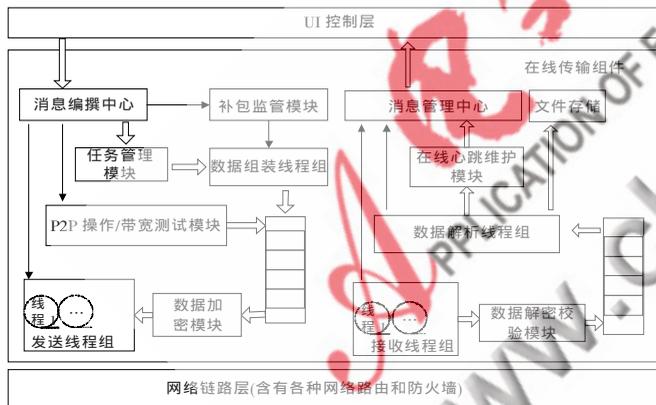


图2 即时数据传输工作原理图

即时数据传输实现了 SOCKET 通信,其中包括 P2P 隧道穿梭、带宽测试、在线心跳维护、补充包机制、数据加密校验和多线程组多缓冲区协同操作模块。在线数据传输的数据分为3种类型:小数据块传输、大数据块传输和即时性数据传输。

文字聊天、操作训令、组织机构和 UI 控制层指令等称为小数据块传输。硬盘文件、数据库大数据流等称为大数据流,因为大数据流只为一个数据流接口进行底层读取数据,例如文件数据传输,当用户需要传输文件时,只要编辑消息和文件路径触发在线发送接口就可以完成数据传输,简化了上层序列化等操作,直接提升传输效率。即时数据称为可丢失数据,如音视频数据,因为为

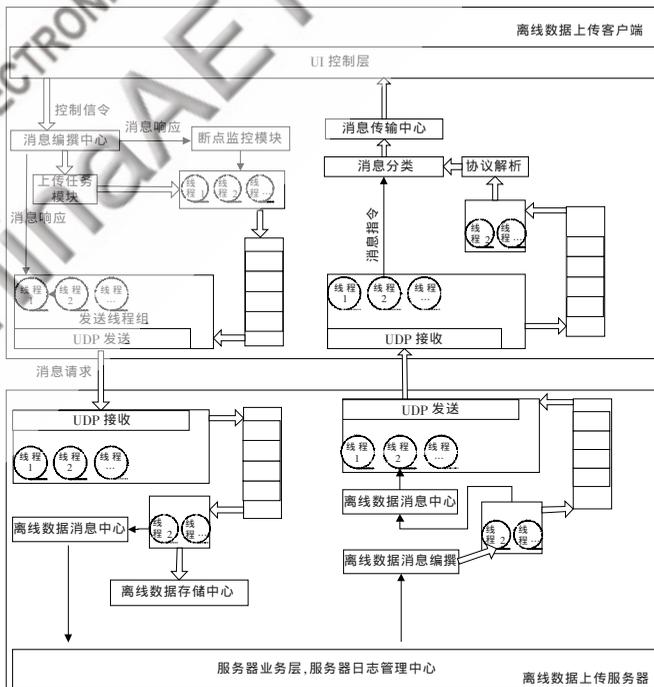
了保障通信即时、不延时而被迫丢弃延时数据。

3 离线数据传输

即时数据传输要求双方都在线,并且接收方必须确认接收后才能开始传输。而在繁忙的工作中,数据通信双方不可能一直保持在线状态,为了保障数据通信的及时有效性,特别对离线数据传输进行了研究。当通信的一方不在线的情况下,数据发起方也可以向对方发送数据,该数据暂时存放在数据服务器中进行中转,当数据接收方上线时,将自动从数据服务器中获取缓存的离线数据。同时为应对意外断线的情况发生,提供了断线续传的功能,即使在传输过程中意外中断,用户仍可从断点处重新开始传输,大大节省了时间和带宽资源。

3.1 数据离线上传

数据离线上传功能包括断点续传、多线程并行发送离线数据、内部执行信令和接收。使用 ATL 开发了离线上传控件,可以嵌入其他语言的二次开发,具有可扩展性,其最高速率达 3.5 MB/s(100 M 网卡)。离线上传包含 UI 控制层、上传客户端组件、上传组件服务器端组件和服务器上传日志管理中心。离线上传组件是系统的核心部位,嵌入 UI 监控层和服务器业务层,只提供特定接口响应或被响应。数据上传工作原理图如图3所示。



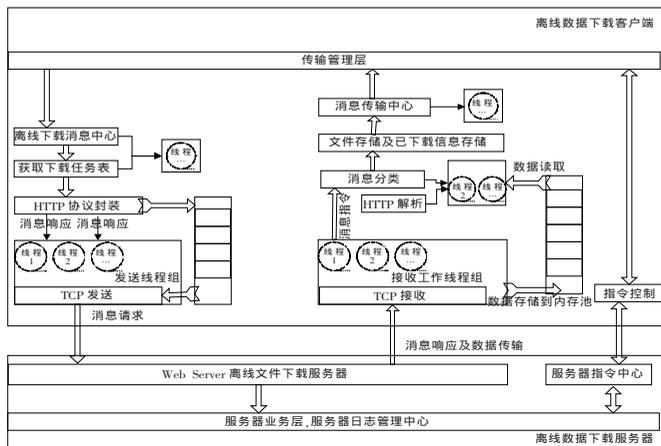
说明:双线箭头表示数据流向;单线箭标表示指令方向

图3 数据离线上传工作原理图

3.2 数据离线下载

数据离线下载主要是负责数据下载,其工作原理如图4所示。

该模块主要包含 HTTP 下载组件和业务控制层模块两部分。HTTP 下载组件是整个离线下载的核心部件,同时它很好地封装了断点续传、线程负载调配、转地址等



说明：双线箭头表示数据流向；单线箭标表示指令方向

图4 数据离线下载工作原理图

操作而只为外部提供了启动、暂停和关闭3种方法和一种事件委托方法及多个配置属性,通过组件封装可以很好地降低工程的耦合;且由于HTTP是一个基于请求与响应模式的、无状态的、应用层的协议,常基于TCP的连接方式,HTTP1.1版本中给出一种持续连接的机制,绝大多数的Web开发,都是构建在HTTP协议之上的Web应用以及嵌入其他语言使用。

3.3 数据缓存服务器

数据缓存服务器主要有:离线数据上传服务组件、WebServer和服务器管理层模块。

上传服务组件主要功能是接收离线数据上传处理,其功能包括数据接收、本地保存以及传输进度等信息响应。

为了构建一个稳定而简易且功能强大的数据服务器体系,离线传输的下载客户端选择了HTTP下载组件,但由于HTTP下载是基于Web数据的下载,由此服务器必须是基于HTTP协议的WebServer。因为WebServer在当今IT界有许多成熟、稳定且配置简单的产品。

服务器管理层模块负责各种业务逻辑处理,包括用户信息存储、下载进度或断点信息保存,以及上传组件与WebServer两个模块间的调度和调配。

4 数据传输测试结果

针对以上提出的传输模型,在Windows平台下,基于.NET平台,分别开发了相关控件,并在内网与政务网内部进行了测试。局域网测试客户端为WindowsXP操作系统,1GB内存,双核CPU,10M网卡;政务网测试客户端为Windows2003操作系统,768MB内存,普通100M网卡。各类控件使用说明如表1所示,测试结果如表2所示。

从表2测试结果可以看出,无论是在内网还是政务网,无论是90MB文件还是190MB文件,无论是单个还是多个文件,P2P传输模式效率都要比HTTP下载控件及离线文件上传效率高得多;而且数据量越大,即时数

表1 使用说明

类型	应用前景
P2P通信	主要适用于即时通信数据可靠性点对点传输
离线下载	可广泛用在基于Webserver资源下载
离线上传	可应用于客户端与服务器的数据传输或用于其他数据服务器传输

表2 测试结果

类型	实际耗时	最高速度速率(KB/s)	CPU/%			
独立90MB文件传输						
	内网	政务网	内网	政务网	内网	政务网
P2P通信控件	19"	31"	5421	3013	4~2	4~2
HTTP下载控件	1'31"	1'41"	1019	512	2~8	2~11
离线文件上传	3'16"	2'8"	508	741.55	0~6	4~22
独立190MB文件传输						
	内网	政务网	内网	政务网	内网	政务网
P2P通信控件	33"	1'2"	6452	3271	4~2	
HTTP下载控件	3'11"	3'34"	1022	933.76	2~10	4~8
离线文件上传	5'28"	4'12"	622	810.63	0~2	2~29
5个90MB文件传输						
	内网	政务网	内网	政务网	内网	政务网
P2P通信控件	56"	1'21"	1844	1165	1~19	
	58"	1'22"	1623	1151.2		
	58"	1'22"	1623	1151.2		
	60"	1'22"	1569	1150.8		
	60"	1'24"	1564	1137.4		
HTTP下载控件	7'10"	2'41"	215.59	573.89	1~18	4~49
	7'12"	2'50"	214.67	545.79		
	7'19"	2'53"	211.62	535.39		
	2'21"	2'53"	212.07	536.36		
	7'25"	2'59"	208.49	516.51		
离线文件上传	5'20"	3'7"	297.34	498.44	2~14	21~69
	5'31"	3'2"	287.16	514.89		
	5'18"	2'55"	298.23	533.41		
	5'31"	3'3"	290.65	512.68		
	5'37"	2'59"	282.76	521.96		
5个190MB文件传输						
	内网	政务网	内网	政务网	内网	政务网
P2P通信控件	1'58"	3'12"	1725	1067.6	8~3	
	2'5"	3'14"	1642	1049.7		
	2'5"	3'16"	1629	1040.8		
	2'5"	3'13"	1629	1056.9		
	2'8"	3'11"	1604	1068.1		
HTTP下载控件	14'7"	6'50"	232.85	479.35	4~12	4~66
	15'17"	6'54"	217.54	475		
	15'56"	6'55"	209.32	473.46		
	15'45"	6'56"	210.96	474.76		
	15'50"	7'20"	210.01	447.52		
离线文件上传	8'4"	7'42"	422.16	443.78	2~21	5~77
	8'17"	7'49"	413.43	433.85		
	8'19"	7'51"	412.72	429.87		
	8'12"	7'53"	416.41	429.38		
	8'7"	7'58"	425.75	429.14		

据传输方式的优势越明显。

本文提出了大数据量数据网络传输架构,开发了相关传输组件,并进行了测试。测试表明,发送效率最高的为即时数据传输模式,而且由于内部编程都由手工完成,理论上还能有很大的提升空间。下一步将进一步完善服务细节设计,完善离线数据传输服务功能。

参考文献

- [1] 徐朝晖, 钱朴慧.UDP 协议的海量信息快速传输解决方案[J].火力与指挥控制,2005,30(1):46-49.
- [2] 张忠平, 栾建锋, 王昆波.网络环境下基于 P2P 的数据集成方法[J].计算机工程,2009,35(12):54-55.

- [3] 喻占武, 郑胜, 李忠民.一种混合式 P2P 下的大规模地形数据传输机制[J].测绘学报,2008,37(2):243-249.

(收稿日期:2009-11-13)

作者简介:

涂平,男,1973年生,硕士,高级工程师,主要研究方向:信息网络共享技术,软件工程和地理信息系统等。

朱晓铃,女,1979年生,硕士,讲师,主要研究方向:遥感应用,地理信息系统。

满旺,男,1979年生,博士,讲师,主要研究方向:地理信息系统,遥感应用。

