一种多级P2P文件交换系统架构的研究

赵继东，王晓斌，张 玮，曾家智

(电子科技大学计算机科学与工程学院 成都 610054)

【摘要】针对现有P2P软件系统采用单层或双层架构所带来的效率低、无法合理进行有效的数据交换的缺点，提出了一种新的三层多级P2P文件交换系统架构。在现有IPv4网络上，该系统采用独特的根服务器/组服务器/用户三层多级架构提高系统性能；使用UDP隧道技术解决了两台计算机在不同的NAT内的文件交换问题；采用新的搜索模式，对文件信息进行了有效的分类和索引，并实现了对文件交换的全程监控。运行结果表明，该系统有效地利用了网络带宽，使文件传输的效率得到了提高。

关 键 词 数据交换; 多级架构; 隧道技术; 搜索模式

中图分类号 TP393 文献标识码 A

Research on a Multistage P2P File Transmission System Architecture

Zhao Jidong，Wang Xiaobin，Zhang Wei，Zeng Jiazhi

(School of Computer Science and Engineering，UEST of China Chengdu 610054)

Abstract This paper analyzes the existing P2P soft wares and their deficiency and inability to transmit data efficiently and hereby presents a new kind of P2P file transmission system architecture. This architecture is based on the existent IPv4 network,working on a unique three leveled multistage RootServer-GroupServer-User structure to improve efficiency. UDP-tunnel technology is employed to solve the problem in realizing the NAT-NAT data transmission. A new search-mode is adopted, files are well classified and indexed, and all of the processes of file transferring are properly monitored in this architecture. The results of the running of the system prove that the system makes an efficient use of the network bandwidth and the efficiency of file transmission is highly increased.

Key words data transmission; multistage architecture; tunnel technology; search-mode

P2P(Peer-to-Peer)是一种用于不同PC用户之间，不经过中继设备直接交换数据或服务的技术，它允许互联网用户直接使用对方的文件。P2P技术使网络上的沟通变得更容易、更直接，在对等计算、协同工作、搜索引擎、文件交换中应用广泛。目前比较流行的P2P文件交换软件普遍采用单层或双层架构,由此带来了效率低、无法合理进行有效数据交换、传输文件速度慢、稳定性差等问题[1]。采用新的三层多级架构可以有效地解决这些问题。

1 三层多级P2P文件交换系统

1.1 根服务器-组服务器-用户的三层多级架构

现有的P2P软件系统多采用单层或双层架构。单层架构的缺点是客户之间无法直接发现对方，而且采用

收稿日期：2003 – 11 – 28
基金项目：国家自然科学基金资助项目(40274046)
作者简介：赵继东(1976 –)，男，硕士，助教，主要从事网络技术和新型网络体系结构方面的研究.

效率低的随机扫描算法，容易给网络带来过多的无效数据流量。而双层架构无法在服务器之间进行有效的数据交换。三层多级架构则可以很好地解决以上问题[2]。

如图1所示，三层多级架构的三层是指根服务器、组服务器、用户三层；多级是指组服务器的分级管理。三层多级提高了系统的稳定性和安全性。

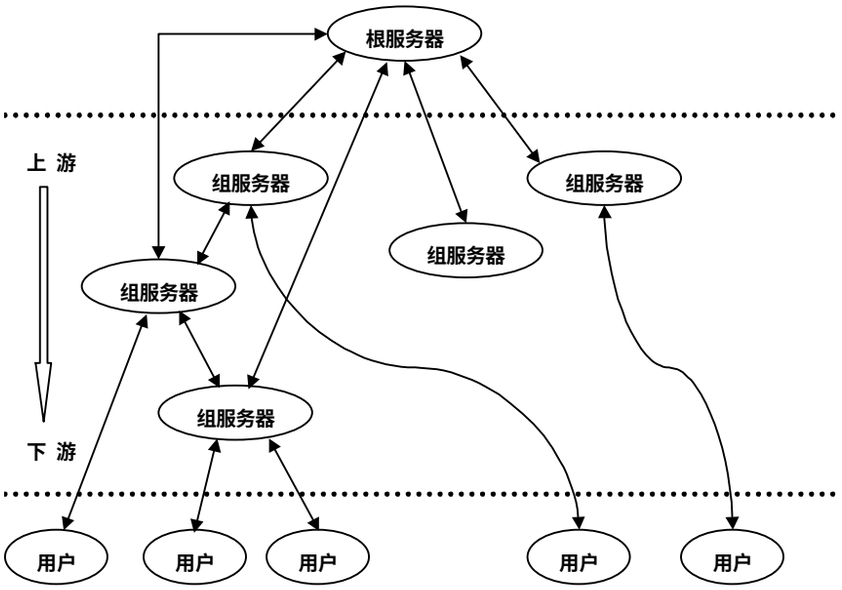

图1　根服务器－组服务器－用户三层多级架构示意图

根服务器是一台稳定性相对要求较高的服务器，内置用户认证模块、日志模块、组服务器、信息模块。根服务器的安全要求很高，所以只对经过认证的组服务器进行数据交换，用户无法直接连接到根服务器。

组服务器是用户登陆的服务器，内置搜索白板模块、用户登陆模块、信息传递模块、消息缓存模块。组服务器有多台，按照新闻组服务器的层次结构架设。每台服务器都存放着一个根服务器节点、一个上游组服务器节点和若干个下游组服务器节点信息。组服务器间的信息是按“上游v下游”方式传递的，当一个组服务器在接收到一个用户的搜索请求时，将该请求放置在本地搜索白板上，然后将其继续传递给它的上游组服务器和它的所有下游组服务器。在这样一个树状的层次结构下，信息是一个同时向上向下的扩展过程，没有冗余的数据交换。

如果一台组服务器发生故障，该组服务器的所有下游组服务器将被孤立出来不能正常通信。但是每个组服务器上都有根服务器地址，这些孤立的组服务器便可以向根服务器提出错误报告，根服务器将对网络结构进行重新调整，然后返回新的上游服务器地址，并通知上游服务器接纳新的下游服务器，使系统具有自我恢复能力。

1.2　用户的登陆认证

在对等网文件交换系统中，要求用户都合作进来并共享文件，从而形成一个友好合作的网络。因此必须禁止只取不拿的使用者。

因为用户只有在登陆认证后才被允许在对等网网络内进行活动，所以服务器需要对已登陆的用户作一个标记，会话ID正好起这个作用。用户在登陆成功后会得到服务器返回给他的一个会话ID，以后用户和服务器的通讯都必须通过会话ID作为信息认证的条件。会话ID是在线用户的唯一标识。

系统从两个方面禁止只取不拿的使用者的产生：1) 匿名登陆的用户没有会话ID，不能和服务器通信，系统禁止其获取对等网内的资源；2) 服务器可以通过会话ID记录正常登陆的用户的活动情况，如果该用户在一定时间内不共享文件，系统取消该用户的会话ID，从而禁止该用户获取对等网中的资源。

1.3　用XML来描述文件信息

传统的P2P文件系统是靠文件名和文件扩展名作为其搜索的检索字段，但仅靠这两个文件信息不能对用户所需要的资源进行有效的检索。用户需要的是有效的文件内容，所以有必要在文件扩展名的基础上对文

件信息进一步地细化。

为了使系统规范化和跨平台，系统采用了通用的数据交换格式可扩展标记语言(eXtensible Markup Language, XML)，将文件信息划分为结构化信息和非结构化信息两大类，系统采用XML对其进行描述。

系统还将文件从粗到细地多次分类，使用户能够很方便地对所需要的文件信息进行查询，对查询返回的文件进行预览，下载前就判断文件是否是自己所需要的，减少了不必要的文件传输。多次分类提高了用户检索效率，也减少了网络的无效流量。

1.4 文件的搜索

文件的搜索是采用组服务器协调下的分布式搜索。系统引入了搜索白板，白板类似于布告栏，用户把搜索请求提交到与其直接相连的组服务器上，组服务器将请求分类后，放到不同的白板上，一个白板就是一个分类。当其他用户连接组服务器的时候，根据自己所共享文件的分类去取回相应的白板，本地搜索后将结果返回需要该文件的用户。

这样的搜索模式有两个优点：1) 组服务器主动控制搜索，可以对信息进行有效的管理和监控；2) 用户只取回了和自己共享内容相匹配的白板信息，有效地减少了网络的流量。

搜索的过程中，白板在组服务器间相互传递，通过这个动作，用户的搜索请求最后会漫游到整个网络中，完成全面搜索。返回的信息直接由用户返回给用户，采用XML格式封装。

1.5 文件传输

系统采用用户数据报协议(User Datagram Protocol, UDP)为底层协议[3]，支持断点续传和多线程下载[4]，支持最大2TB(2, 199, 023, 255, 552 bytes)的文件进行传输，并加入信息-摘要算法(Message-Digest Algorithm 5, MD5)对传输的文件进行校验。

1.6 NAT隧道传输技术

如果两台机器都在网络地址转换 (Network Address Translation, NAT)中，现有P2P文件交换软件无法实现文件传输，系统使用UDP隧道技术完成文件的跨NAT传输。

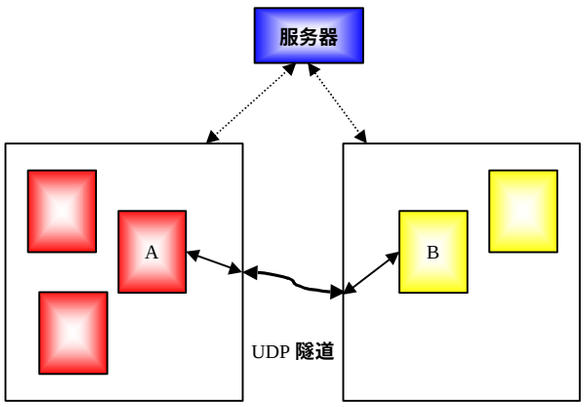

图2 双端NAT建立UDP隧道示意图

NAT内部的机器创建了一个UDP端口，对外发送一次数据，在NAT网关的端口映射表上就会建立一条数据，即外部IP地址和端口号与内部IP地址和端口号的映射关系。NAT网关会保留这条映射数据，以后的一段时间内，当外部网络有程序向NAT网关的这个被记录的UDP端口发送数据时，NAT网关会根据端口映射表查找到内部机器的IP地址和端口，将数据直接转发到内部的机器上。这样一来，NAT对于内部机器和外部机器来说是透明的，内外的机器可以通过NAT网关进行直接的数据交换。

如图2所示，机器A和B在传输文件前，各自创建一个UDP端口，向外部网络的服务器发送传输请求，服务器在接收到请求的同时，也获取了机器A, B的对外IP地址和端口，并将它们对外的IP地址和端口分别告知对方。这样，A和B就能够直接发送和接收数据，从而突破了NAT给文件传输带来的局限。

1.7 日志系统

系统采用日志系统来防止非法文件的传输。为了避免日志系统中的数据过于庞大，该系统只记录文件传输过程中的必要信息。系统只需记录文件的传输过程，就能阻止非法文件在系统中的扩散。

在该系统中，为了日志数据的统一管理，日志服务器是唯一的，在每个组服务器中所进行的文件交换行为都将被最终记录到日志服务器中去，系统用根服务器做日志服务器。

系统中日志数据随着用户的增多会大量地增加，为了防止大量的日志信息瞬间涌到根服务器上，造成根服务器的负荷过重最终造成系统崩溃，用户日志采用分级方式记录，数据被即时地记录到组服务器上，然后再定期汇总到根服务器，这样既能保证用户日志记录的即时性，又能避免日志服务器的负荷过高。

2 结 论

通过对现有P2P文件交换软件的架构，以及所采用的一些常用技术的分析，提出了一个新的稳定的文件交换架构。针对文件传输过程中的问题，采用三层多级架构，解决了服务器之间的数据分配交换，对资源的分配更加有效，提高了系统的健壮性。测试数据表明，该系统在同线程数的条件下具有更高的文件传输效率，在三个线程的时候达到峰值，有效地利用了网络带宽，并实现了文件传输过程中的全程监控。

参 考 文 献

[1] Dana M, John H. 对等网[M]. 苏 忠, 战晓雷译. 北京: 清华大学出版社. 2003. 118-122

[2] Brian K, Phil L. Network news transfer protocol[S]. RFC 977, 1986

[3] Kent S. THE tftp protocol (revision 2)[S]. RFC 1350, 1992

[4] Fielding R, Gettys J. Hypertext transfer protocol-HTTP/1.1[S]. RFC 2616, 1999

[5] Rivest R. The MD5 message-digest algorithm[S]. RFC 1321, 1992

编 辑 熊思亮

(上接第429页)

4 结 束 语

针对一类特殊环境下的组播需求，提出了一种满足多个约束条件的组播路由算法。该算法能够实现满足跳数约束和最少时隙资源消耗约束的低价组播树的构建。下一步的工作是进行计算机仿真，将算法进一步简化并应用于实际之中。

参 考 文 献

[1] 徐 格, 吴建平, 徐 明. 高等计算机网络－体系结构、协议机制、算法设计与路由器技术[M]. 北京: 机械工业出版社, 2003

[2] Winter P. Steiner problem in network: a survey[J]. IEEE Network, 1987, 3: 129-167

[3] Takahashi H, Matsuyama A. An approximate solution for the Steiner Tree problem in grphs[J]. Mathematica Japonica, 1980, 24(6): 573-577

[4] Haberman B K, Rouskas G. cost, delay and delay variation conscious multicast routing[R]. Technical Report TR-97-03, North Carolina State University, 1997

[5] Cormen T H, Leiserson C E, Rivest R L. Introduction to algorithms[M]. Cambridge: MIT Press, 1997

编 辑 刘文珍