

对等网络

1

林宇¹, 程时端¹, 李琦²
(1 北京邮电大学, 北京 100876;
2 北京大学, 北京 100088)

中图分类号: TP393.03 文献标识码: A 文章编号: 1009-6868 (2006) 01-0057-04

网络资源的变迁,促使网络计算模式发生变化。相应的一种采用对等策略计算模式的网络——对等网络(P2P)得到了广泛关注。P2P是一种分布式网络,网络的参与者共享他们所拥有的硬件资源,这些共享资源能被其他对等结点直接访问而无需经过中间实体。网络中的参与者既是资源提供者,又是资源获取者。为了使读者对P2P有所了解,本讲座将分3期对P2P进行介绍:第1期介绍P2P的拓扑结构、组织模式以及核心查找算法;第2期介绍P2P研究现状、应用情况以及下一步演进与拓展方向;第3期讨论P2P给Internet带来的机遇和挑战,探讨P2P对电信运营商和设备制造商的影响。

对等网络 (Peer-to-Peer Networks, P2P) 是一种采用对等策略计算模式的网络。在传统的互联网计算模式中,客户端/服务器(C/S)模式占据了主流。当时,客户端的带宽和计算资源较弱,通过C/S模式可以降低对客户终端能力的要求,而将处理集中在服务器端。近年来,不同资源的发展速度出现了以下特点:网络的流量以每6个月翻倍的速度增长,网络带宽以每7个月翻倍的速度增长,计算资源近似依照摩尔定理速度增长(18个月翻倍),而存储能力每年仅提升7%。因此在诸多资源中,计算和存储资源可能逐渐变为“瓶颈”。相应地,处于体系架构的中心服务器也成为性能的“瓶颈”,一旦中心服务器崩溃将造成整个服务系统崩溃。在这样的技术发展背景下,人们引入了对等计算模式。

随着终端技术和网络接入技术的发展,终端的能力越来越强,P2P采用处于网络边缘的终端的协作来弥补和解决集中式架构导致的性能“瓶颈”。

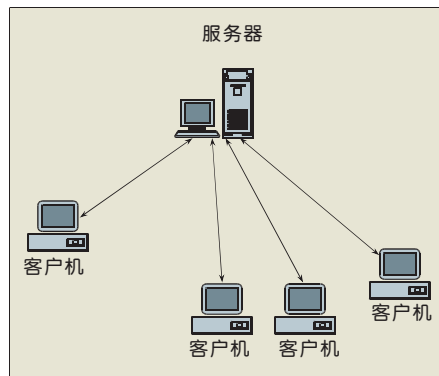
P2P打破了传统的C/S模式,在网络中的每个结点的地位都是对等的。每个结点既充当服务器,为其他结点

提供服务,同时也享用其他结点提供的服务。P2P与C/S模式的网络结构分别如图1和图2所示。

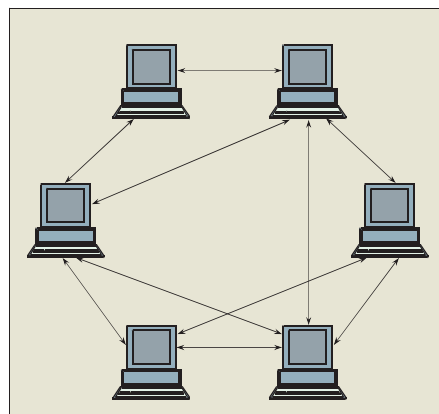
从不同的行业和视角来看,P2P的定义略有差别。一种典型定义为:P2P是一种分布式网络,网络的参与者共享他们所拥有的一部分硬件资源(处理能力、存储能力、网络连接能力、打印机等),这些共享资源能被其他对等结点直接访问而无需经过中间实体。在此网络中的参与者既是资源(服务和内容)提供者,又是资源(服务和内容)获取者。

为了便于读者理解,本文中我们以一个下棋应用系统的简单例子来解释一些基本原理。在传统的C/S架构下,下棋系统由下棋服务器和下棋者两类结点构成。一个下棋系统的工作流程是,下棋者A和B登陆下棋服务器,在服务器的撮合(组配)下,A和B在一个桌子上坐下开始下棋,A再下一个子,是通过“A→下棋服务器→B”的流程来实现的,每一次下棋的控制管理流程(用户登陆和下棋组配)和业务流程(具体的下棋流程)都需要服务器的参与。因此,提供给上百万并发用户的下棋系统需要数量庞大、功能强大的服务器群才能正常运转。

P2P技术的特点体现在如下几个方面:



▲ 图1 C/S模式网络结构



▲ 图2 P2P模式网络结构

(1)非中心化

网络中的资源和服务分散在所有结点上,信息的传输和服务的实现都直接在结点之间进行,无需中间环节和服务器的介入,避免了可能的“瓶颈”。以下棋系统为例,下棋的业

基金项目: 国家“973”计划项目
(2003CB314806); 国家自然科学基金项目
(90204003)

务流程直接在下棋者的两个结点之间完成,无需中心服务器的参与(除了统一计费、记分等需要集中管理的服务)。非中心化是P2P的基本特点,带来了其在可扩展性、健壮性等方面的优势。

(2)可扩展性

在P2P网络中,随着用户的加入,不仅服务的需求增加了,系统整体的资源和服务能力也在同步扩充(因为新加入的用户本身也提供服务 and 资源),因此能够较好地满足用户的需要。整个体系是全分布的,不存在明显的“瓶颈”。以下棋系统为例,下棋的业务能力主要是通过下棋者的结点来提供(包括棋盘的绘制,下棋的流程规则管理等),对下棋服务器增加的负担较小。

(3)健壮性

P2P架构具有耐攻击、高容错的优点。P2P网络通常都是以自组织的方式建立起来的,并允许结点自由地加入和离开。不同的P2P网络采用不同的拓扑构造方法,根据网络带宽、结点数、负载等变化不断自适应地调整拓扑结构。由于服务是分散在各个结点之间进行的,部分结点或网络遭到破坏对其他部分的影响很小(即两个下棋人之间的网络被破坏,不会直接影响其他的下棋用户),即便是部分结点失效了,P2P网络也能通过自动调整机制重构整体拓扑,保持与其他结点的连通性。

(4)高性价比

采用P2P架构可以有效地利用互联网中散布的大量普通结点,将计算任务或存储资料分布到所有结点上,利用其中闲置的计算能力或存储空间,达到高性能计算和海量存储的目的。以下棋系统为例,采用P2P架构的下棋系统,不再需要那么多数量的服务器,因为大部分的业务都被用户结点所分担。

(5)隐私保护

在P2P网络中,由于信息的传输分散在各结点之间进行而无需经过

某个集中环节,用户的隐私信息被窃听和泄漏的可能性大大缩小。目前,解决Internet隐私问题主要采用中继转发的技术方法,从而将通信的参与者隐藏在众多的网络实体之中。在传统的匿名通信系统中,实现这一机制依赖于某些中继服务器结点(比如传统的下棋系统中的计费和记分,一般都需要通过中心服务器来实现)。而在P2P中,所有参与者都可以提供中继转发的功能,因而大大提高了匿名通信的灵活性和可靠性,能够为用户提供更好的隐私保护。这个优点恰恰也是P2P系统的缺点,这种特性导致了它常常被非法组织用于私密信息传递(比如,此时下棋人之间要想作弊的话,就更为容易,因为没有中心服务器进行监管)。

1 网络资源与计算模式的变化

1.1 技术和需求的平衡

一种网络技术是否可以被广泛应用,可以用下面的公式来考量:

网络系统的代价=人们愿意为获得信息服务而付出的代价?

也就是说,构建一个满足人们需求的网络的总代价是否能和人们为了满足其信息服务需求所愿意付出的代价相等。

在等式的左边,底层网络资源的变迁(包括资源的总量、结构和分布情况)将导致网络架构的改变。在传统的电信网络中,终端的计算、存储和带宽资源都非常有限,基本不存在着应用P2P计算模式的可能。在窄带互联网,比如主要以拨号接入为主的网络环境下,终端的计算和存储资源是足够的,但是带宽资源仍然是“瓶

颈”。当以不对称数字用户线(ADSL)为主流的准宽带接入技术普及后,终端的各类资源基本能够满足P2P计算模式的需求。比如,IP组播是人们研究了10多年而难于解决的方向,它对网络资源(主要指组播路由和可靠传递)的需求远远超过了网络所能提供的资源,因此,通过P2P计算模式,将组播的路由和群组管理在终端上实现,实质是对网络资源利用在结构和分布的调整。

等式右边的变量随不同国家和地区具体情况而有所差异。在发达国家,人们的基本生活需求比重很低,而信息消费的需求比例较高。在中国,人们的信息需求在整个人类生活的比重较低。因此,短信这样的业务取得成功是有深刻的道理的:系统投入的网络资源低,而又能很大程度上满足人们的信息需求,当然,亚洲人灵巧的手指也是造成短信业务用户满意度高的一个因素。

1.2 网络发展的“钟摆效应”

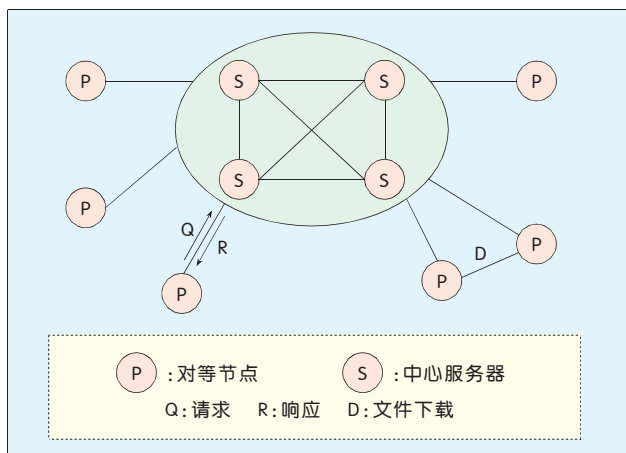
在网络发展的历史上,网络底层资源变迁常常导致网络发展的“钟摆效应”(见表1),当然,每一次的“钟摆”都产生新一轮的螺旋上升。

(1)骨干网络和接入网络“瓶颈”的交替

事物总是螺旋式上升,波浪式前进的。在石英光纤大规模应用前,骨干网络是整个Internet的“瓶颈”,人们将主要的研究都聚焦在骨干传送网上。光纤网络使得Internet的骨干网不再成为“瓶颈”,此时,基于Modem拨号方式的互联网接入成为网络的“瓶颈”,人们将研究聚焦在接入网络上。随着不对称数字用户线(ADSL)、高速数字用户线(HDSL)、甚高速

▼表1 网络发展的“钟摆效应”

网络带宽	技术热点	计算模式
骨干网“瓶颈”	网络层技术热点	单CPU到对称多分布结构,再到多结点机
接入网“瓶颈”	业务层技术热点	非对称计算
骨干网“瓶颈”	网络层技术热点	对等计算



◀图3
MP3共享软件Napster采用的结构模式

数字用户线(VDSL)、混合光纤/同轴电缆(HFC)、无线局域网(WLAN)等多种宽带接入方式逐步商用,骨干网可能再度成为Internet的“瓶颈”。

(2) 网络技术与业务技术热点的交替

在1995年前后,全球性的网络基础设施还没有构建完成,异步传输模式(ATM)和IP等网络技术成为技术热点;在网络层技术基本解决后(1998年—2002年前后),业务层技术,比如智能网(IN)、电信管理网(TMN)、Web、IP语音(VoIP)、P2P等成为应用热点;而当人们希望将P2P、下一代网络(NGN)、网络电视(IPTV)、第3代移动通信(3G)等业务承载在IP网上时,网络承载技术(服务质量:差异化服务、业务接入控制、流量工程等)又成为必须解决的关键问题。

(3) 计算模式的交替

在大型计算机时代,从单CPU发展到对称多分布结构,再到多结点机,对等计算的思想早已被广泛应用,只不过当时应用在近程的分布计算中;随着远程通信网络和PC的发展,使得能力强大的服务器和功能较弱的客户端成为流行的计算模式;当网络和PC的能力进一步发展到今天时,远程分布式对等计算再次流行起来。P2P应用正是在这种情况下开始引人瞩目。背景是:互联网的带宽越来越大,互联网有同通信网和广播网融合的倾向,互联网改变了信息传播

和获得的方式,互联网使得顾客从经济上讲有更大的选择权和发言权。

2 P2P的拓扑结构

拓扑结构是指分布式系统中各个计算单元之间的物理或逻辑的互连关系。结点之间的拓扑结构一直是确定系统类型的重要依据。目前互联网中广泛使用集中式、层次式等拓扑结构,集中式拓扑结构系统目前面临着过量存储、拒绝服务(DoS)攻击等一些难以解决的问题。P2P系统要构造一个非集中式的拓扑结构,根据拓扑结构的关系可以将P2P研究分为4种形式:中心化拓扑、全分布式非结构化拓扑、全分布式结构化拓扑和半分布式拓扑。在构造过程中需要解决的主要问题包括:系统中所包含的大量结点如何命名、组织;如何确定结点的加入/离开方式;如何进行出错恢复等。

2.1 集中目录式结构

集中目录式P2P结构(中心化拓扑)是最早出现的P2P应用模式,因为仍然具有中心化的特点也被称为非纯粹的P2P结构,经典案例就是著名的MP3共享软件Napster。

Napster通过一个中央服务器保存所有Napster用户上传的音乐文件索引和存放位置的信息。当某个用户需要某个音乐文件时,首先连接到Napster服务器,在服务器进行检索,并由

服务器返回存有该文件的用户信息;再由请求者直接连到文件的所有者传输文件。在Napster模型中,一群高性能的中央服务器保存着网络中所有对等计算机共享资源的目录信息。当需要查询某个文件时,对等机会向一台中央服务器发出文件查询请求。中央服务器进行相应的检索和查询后,会返回符合查询要求的对等计算机地址信息列表。查询发起对等计算机接收到应答后,会根据网络流量和延迟等信息进行选择,和合适的对等计算机建立连接,并开始文件传输。Napster的结构模式和工作原理如图3所示。

以下棋应用为类比,Napster架构相当于由中心服务器完成下棋用户登陆、下棋撮合等管理服务,而一旦用户开始下棋后,下棋流程就在两个下棋者之间完成,服务器就不再参与二者下棋的后续处理。因此,Napster实质上是实现了文件查询(管理服务)与文件传输(具体的业务服务)的分离,有效地节省了中央服务器的资源消耗。中心化拓扑最大的优点是:维护简单;资源发现效率高;由于资源的发现依赖中心化的目录系统,发现算法灵活高效并能够实现复杂查询。

但是,这种对等网络模型存在很多问题,主要表现为:

(1)最大的隐患在中央服务器上,由于Napster在文件查询服务上还采用集中式的架构,如果该服务器失效,整个系统都会瘫痪(类比,下棋服务器一瘫痪,用户之间无法进行撮合服务,下棋业务也就瘫痪了)。当用户数量增加到 10^5 或者更高时,Napster系统的性能会大大下降。中央服务器的瘫痪容易导致整个网络的崩溃,可靠性和安全性较低。

(2)随着网络规模的扩大,对中央索引服务器进行维护和更新的费用将急剧增加,所需成本过高。

(3)中央服务器的存在引起共享资源在版权问题上的纠纷。

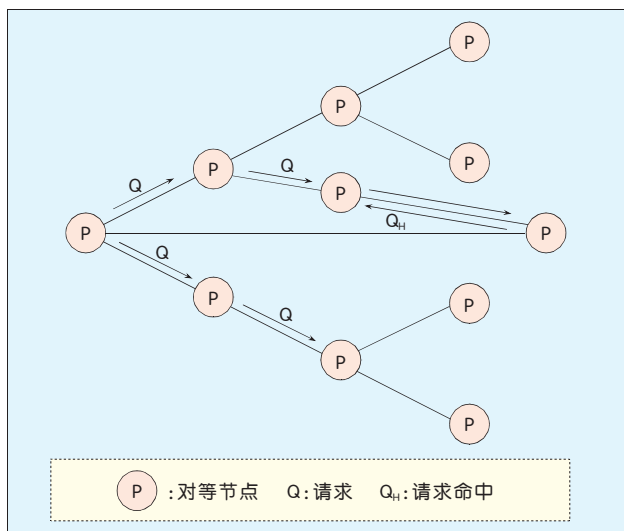


图4
Gnutella采用的“洪泛”
搜索算法

基于上述分析,对小型网络而言,集中目录式模型在管理和控制方面占一定优势,但该模型并不适合大型网络应用。

2.2 纯P2P网络模型

全分布非结构化网络是在重叠网络中采用随机图的结点拓扑组织方式,其最典型的案例是Gnutella系统。Gnutella是一个P2P文件共享系统,它和Napster系统最大的区别在于Gnutella是纯粹的P2P系统,没有索引服务器,每个结点都随机维护自己本地局部的拓扑连接关系,采用了基于完全随机图的“洪泛”发现和随机转发机制,工作原理如图4所示。当需要进行信息查找时,Gnutella系统将发送一个广播消息给周边的结点,询问是否有相关的内容。如果周边结点存在相关的内容,则向查询结点发回查找结果。为了控制搜索消息的传输范围,Gnutella系统引入了生存时间(TTL)减值的概念。

以下棋应用为类比,在这种架构下,没有中心的下棋服务器,某个下棋者如果想要下棋,就直接向周边的结点询问(这些周边结点的拓扑知识是随机获得的,比如以前曾经和他下过棋的人,或者曾经询问过他的人),是否有人愿意与之下棋?如果有志愿者,业务撮合就完成了;如果周边结

点没人愿意,这些周边结点又会继续向它们周边的结点询问,直至联系到愿意下棋的伙伴或者失败。

在Gnutella分布式对等网络模型中,每一个连网计算机在功能上都是对等的,既是客户机同时又是服务器,所以被称为对等计算机。由于互联网上的结点数服从“Power-law”规律(幂率分布),即少数的结点拥有很高的连接度,这导致了小世界现象(在人类社会,任意两个不直接相识的人,通过最多6个中间人就可进行通信)。

类比下棋应用,大多数情况下,通过多次询问,就可能找到志愿者,因此Gnutella的搜索模式能够较快发现目的结点,面对网络的动态变化体现了较好的容错能力。

但是,随着Gnutella连网结点的不断增多,网络规模不断扩大,通过这种“洪泛”方式定位对等点的方法将造成网络流量急剧增加(比如有许多下棋人都同时询问周边的同伴,容易形成雪崩式的巨大噪声),从而导致网络中部分低带宽结点因网络过载而失效,所以在初期的Gnutella网络中,存在比较严重的分区、断链现象(即Gnutella的用户群不能实现全体的互通)。资源发现的准确性和可扩展性是非结构化网络面临的两个重要问题。和Gnutella系统类似,Freenet系

统也采用了完全分布式的模型,而且增加了一些改进措施。Gnutella和Freenet虽然支持分布式的查找策略,但是它们都采用类似于开放式最短路径优先(OSPF)路由协议(实际上,OSPF协议本身也是一个对等网络系统)的“洪泛”机制,这种机制一方面造成网络通信负担较大,另一方面可扩展性也较差。正是由于类似的原因,OSPF协议才主要被限制在Internet的自治域(AS)内部使用。

由于非结构化网络一般不提供性能保证,查询可能没有结果。采用广播查询的系统对网络带宽的消耗非常大,由此带来可扩展性差等问题。为了解决这些问题,大量研究集中在如何构造一个高度结构化的系统,也就是下面要讨论的基于全分布式结构化拓扑的网络模型。(待续)

收稿日期:2005-11-22

作者简介



林宇,北京邮电大学网络与交换国家实验室副教授,博士。主要研究方向包括QoS控制、测量和管理、P2P技术、移动应用等。曾作为主持研究人和项目负责人参与国家“973”、国家自然科学基金、国家“863”项目10余项。已出版专著3部,发表论文60余篇,论文被SCI收录8篇,EI收录近30篇。



程时端,北京邮电大学教授、博士生导师,北京邮电大学学术委员会副主任。目前研究方向包括ISDN、ATM、TCP/IP、ATM和IP网的话音通信技术、协议工程、流量工程、宽带网络性能和服务质量等。



李琦,北京大学教授、博士生导师,中国图形图像学会技术委员会主任,全国地理信息标准化技术委员会委员,国家电子标签标准工作组成员,国家海洋局信息中心特聘教授,国务院信息化工作办公室顾问,北京市政府专家顾问,中国数字地球发展战略研究专家。