

对等网络的研究与进展

谢东亮,程时端,阙喜戎

(北京邮电大学,北京 100876)

(Beijing University of Posts and Telecommunications, Beijing 100876, China)

- 对等网络(P2P)被美国《财富》杂志称为改变因特网发展的四大新技术之一,甚至被认为是无线宽带互联网的未来技术。
- P2P技术不仅为个人用户提供了前所未有的自由和便利,同时也试图有效地整合互联网的潜在资源,将基于网页的互联网转变成动态存取、自由交互的海量信息网络。
- P2P技术的发展以及P2P与网格技术的结合,将影响整个计算机网络的概念和人们的信息获取模式,真正实现“网络就是计算机,计算机就是网络”的梦想。

作为改变现有Internet应用模式的主要技术之一,计算机对等网络(P2P)是目前新一代互联网技术研究的热点之一。

自1999年以来,P2P的研究得到了国内外学术界和商业组织的广泛关注,同时,由于P2P本质特性不可避免地存在着许多社会、法律和技术上的问题,在学术界和产业界也一直存在着一些怀疑的力量,这在很长一段时期使人们难以对P2P做出一个准确和公平的判断。

本文较为完整地分析了P2P网络的4种典型结构,并对P2P的主要应用模式、存在的问题以及可能的发展方向进行简要阐述。

1 P2P网络模型

P2P网络是一种具有较高扩展性的分布式系统结构,其对等概念是指

网络中的物理节点在逻辑上具有相同的地位,而并非处理能力的对等。以Napster软件为代表的P2P技术其实质在于将互联网的集中管理模式引向分散管理模式,将内容从中央单一节点引向网络的边缘,从而充分利用互联网中众多终端节点所蕴涵的处理能力和潜在资源。相对于传统的集中式客户/服务器(C/S)模型,P2P弱化了服务器的概念,系统中的各个节点不再区分服务器和客户端的角色关系,每个节点既可请求服务,也可提供服务,节点之间可以直接交换资源和服务而不必通过服务器。

P2P系统最大的特点就是用户之间直接共享资源,其核心技术就是分布式对象的定位机制,这也是提高网络可扩展

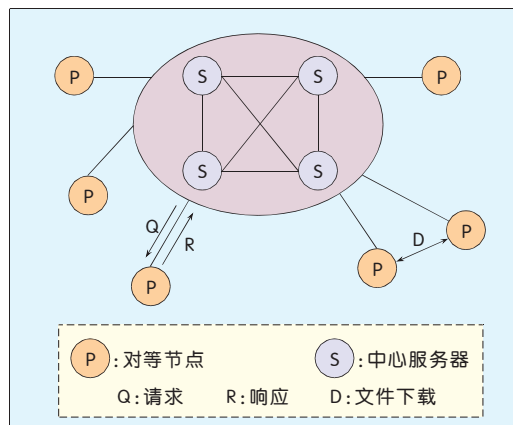
性、解决网络带宽被吞噬的关键所在。迄今为止,P2P网络已经历了三代不同网络模型,各种模型各有优缺点,有的还存在着本身难以克服的缺陷,因此在目前P2P技术还远未成熟的阶段,各种网络结构依然能够共存,甚至呈现相互借鉴的形式。

1.1 集中目录式结构

集中目录式P2P结构是最早出现的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。用于共享MP3音乐文件的Napster是其中最典型的代表(见图1),其用户注册与文件检索过程类似于传统的C/S模式,区别在于所有资料并非存储服务器上,而是存贮在各个节点中。查询节点根据网络流量和延迟等信息选择合适的节点建立直接连接,而不必经过中央服务器进行。这种网络结构非常简单,但是它显示了P2P系统信息量巨大的优势和吸引力,同时也揭示了P2P系统本质上所不可避免的两个问题:法律版权和资源浪费的问题。

1.2 纯P2P网络模型

纯P2P模式也被称作广播式的P2P模型。它取消了集中的中央服务器,每个用户随机接入网络,并与自己相邻的一组邻居节点通过端到端连接构成一个逻辑覆盖的网络。对等节点之间的内容查询和内容共享都



▲图1 Napster集中目录式P2P网络模型

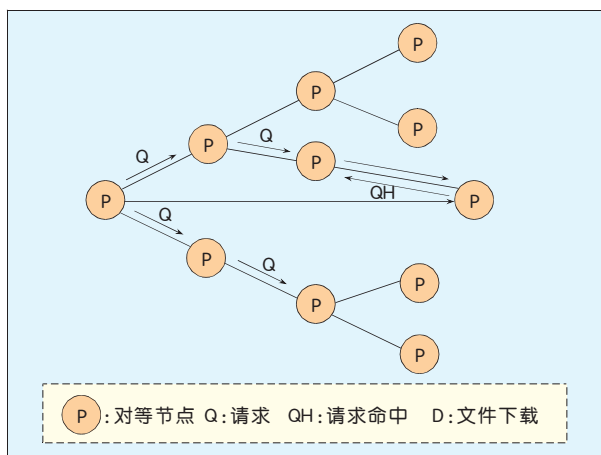


图2
Gnutella纯P2P网络模型

是直接通过相邻节点广播接力传递，同时每个节点还会记录搜索轨迹，以防止搜索环路产生。

Gnutella模型是现在应用最广泛的纯P2P非结构化拓扑结构（见图2），它解决了网络结构中心化的问题，扩展性和容错性较好，但是Gnutella网络中的搜索算法以泛洪的方式进行，控制信息的泛滥消耗了大量带宽并很快造成网络拥塞甚至网络的不稳定。同时，局部性能较差的节点可能会导致Gnutella网络被分片，从而导致整个网络的可用性较差，另外这类系统更容易受到垃圾信息，甚至是病毒的恶意攻击。

1.3 混合式网络模型

Kazaa模型是P2P混合模型的典型代表（见图3），它在纯P2P分布式模型基础上引入了超级节点的概念，综合了集中式P2P快速查找和纯P2P去中心化的优势。Kazaa模型将节点按能力不同（计算能力、内存大小、连接带宽、网络滞留时间等）区分为普通节点和搜索节点两类（也有的进一步分为三类节点，其思想本质相同）。其中搜索节点与其临近的若干普通节点之间构成一个自治的簇，簇内采用基于集中目录式的P2P模式，而整个P2P网络中各个不同的簇之间再通过纯P2P的模式将搜索节点相连起来，甚至也可以在各个搜索节点之间再次选取性能最优的节点，或者另外引入

一新的性能最优的节点作为索引节点来保存整个网络中可以利用的搜索节点信息，并且负责维护整个网络的结构。

由于普通节点的文件搜索先在本地所属的簇内进行，只有查询结果不充分的时候，再通过搜索节点之间进行有限的泛洪。这样就极为有效地消除纯P2P结构中使用泛洪算法带来的网络拥塞、搜索迟缓等不利影响。同时，由于每个簇中的搜索节点监控着所有普通节点的行为，这也能确保一些恶意的攻击行为能在网络局部得到控制，并且超级节点的存在也能在一定程度上提高整个网络的负载均衡。

总的来说，基于超级节点的混合式P2P网络结构比以往有较大幅度的改进。

然而，由于超级节点本身的脆弱性也可能导致其簇内的节点处于孤立状态，因此这种局部索引的方法仍然存在一定的局限性。这导致了结构

化的P2P网络模型的出现。

1.4 结构化网络模型

所谓结构化与非结构化模型的根本区别在于每个节点所维护的邻居是否能够按照某种全局方式组织起来以利于快速查找。结构化P2P模式是一种采用纯分布式的消息传递机制和根据关键字进行查找的定位服务，目前的主流方法是采用分布式哈希表(DHT)技术，这也是目前扩展性最好的P2P路由方式之一。由于DHT各节点并不需要维护整个网络的信息，只在节点中存储其临近的后继节点信息，因此较少的路由信息就可以有效地实现到达目标节点，同时又取消了泛洪算法。该模型有效地减少了节点信息的发送数量，从而增强了P2P网络的扩展性。同时，出于冗余度以及延时的考虑，大部分DHT总是在节点的虚拟标识与关键字最接近的节点上复制备份冗余信息，这样也避免了单一节点失效的问题。

目前基于DHT的代表性的研究项目主要包括加州大学伯克利分校的CAN项目和Tapestry项目，麻省理工学院的Chord项目、IRIS项目，以及微软研究院的Pastry项目等。这些系统一般都假定节点具有相同的能力，这对于规模较小的系统较为有效。但这种假设并不适合大规模的Internet部署。同时基于DHT的拓扑维护和修复算法也比Gnutella模型和Kazaa模型等无结构的系统要复杂得多，甚至在Chord项目中产生了“绕路”的问题。事实上，目前大量实际应用还大都是

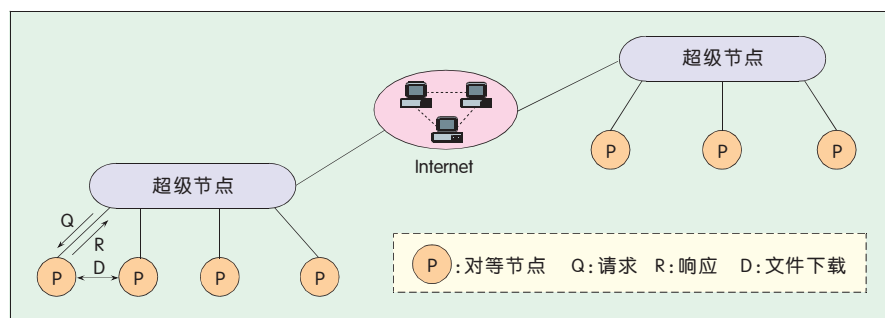


图3 超级节点的混合式P2P网络模型

基于无结构的拓扑和泛洪广播机制,现在大多采用DHT方式的P2P系统缺乏在Internet中大规模真实部署的实例,成功应用还比较少见。

2 P2P网络应用模式

Internet最初产生和发展的一个动力就是资源共享,也正是文件交换的需求直接导致了P2P技术的兴起,这是P2P最初也是最成功的应用之一,也正是针对这类应用的Napster使得人们在客户/服务器模式下开始重新认识P2P思想对人们使用网络习惯的影响。

随着人们对P2P思想的理解和技术的发展,作为一种软件架构,P2P还可以被开发出种类繁多的应用模式,除了最初的文件交换之外,还出现了一些分布式存储、深度搜索、分布式计算、个人即时通信和协同工作等新颖应用。其中最著名的例子是基于分布式计算的搜索外星文明SETI@home科学实验,每个志愿参加者只需下载并运行类似屏幕保护的方式,就可以贡献自己闲置的计算能力,参与分析Arecibo射电望远镜的无线电磁波数据并回送计算数据,截至2004年12月,已有528万志愿者参加进来,获得了相当于216万年的CPU时间,仅一天的综合计算就相当于67.46Tflops运算。另外,随着SUN公司将其JXTA协议扩展到诸如个人数字助理(PDA)和移动电话等手持终端上,并允许人们屏蔽具体的物理平台进行资料共享和文件交换等,P2P技术在移动通信和智能网领域也开始呈现出较大应用前景。

3 P2P网络存在的问题

P2P最大的优点在于能够提供可靠的信息查询,但从社会和法律意义来说,绝大多数的P2P服务都将不可避免地遇到知识产权冲突,也可能成为一些非法内容传播的平台。同时由于缺乏中心监管以及自由平等的动态特性,自组织的P2P网络在技术层

面也有许多难以解决的问题。

从某种意义上来说,P2P网络和人脉网络具有一定的相似性。一般来说,每个P2P网络都是众多参与者按照共同兴趣组建起来的一个虚拟组织,节点之间存在着一种假定的相互信任关系,但随着P2P网络规模的扩大,这些P2P节点本质所特有的平等自由的动态特性往往与网络服务所需要的信任协作模型之间产生矛盾。激励作用的缺失使节点间更多表现出“贪婪”、“抱怨”和“欺诈”的自私行为,因此P2P中预先假设的信任机制实际上非常脆弱,同时这种信任也难以在节点之间进行推理,导致了全局性信任的缺乏,这直接影响了整个网络的稳定性与可用性。此外,相对于传统客户/服务器模式的服务器可以做主动和被动的防御,由于P2P节点安全防护手段的匮乏以及P2P协议缺乏必要的认证机制和计算机系统的安全漏洞,安全问题在P2P网络中更为严重,这将直接影响P2P的大规模商用。另外,P2P网络中的节点本身往往是计算能力相差较大的异构节点,每一个节点都被赋予了相同的职责而没有考虑其计算能力和网络带宽,局部性能较差的点将会导致整体网络性能的恶化,在这种异构节点的环境中难以实现优化的资源管理和负载均衡。同时,由于用户加入离开P2P网络的随意性使得用户获得目标文件具有不确定性,导致许多并非必要的文件下载,而造成大量带宽资源的滥用。特别是大多数P2P用户更喜欢传送音频、视频这些较大的媒体文件,这将使得带宽浪费问题更为突出,尤其在中国大量的用户还是拨号用户,较窄的带宽也成为P2P应用难以逾越的障碍。

4 结论与展望

P2P技术在最近几年获得了高速的发展,也出现了较多应用,但截至目前,P2P中仍有很多的关键技术问题并没有得到解决,其中最典型的就

是带宽吞噬、网络可扩展性差和路由效率低下等问题。这导致P2P至少在目前的技术水平而言只能是一种小范围不可靠的应用或是满足特定任务需求的专门应用。并且,作为一种潜在的商业应用,如何在P2P网络中有效地保护知识产权以及如何设计盈利模式将会面临更为严格的考验。

未来的网络将呈现大规模分布式、全球性计算和全球性存储的特征,从长远的趋势来看,对于访问和传输服务的需求必将远远大于对于计算功能的需要。尽管P2P技术现在还不成熟,但是迄今为止,至少在理论上P2P仍是最有吸引力的个人通信技术。尤其是P2P与网格技术的结合将是分布式计算技术最有吸引力的发展趋势,虽然现在还没有成熟的方案,但随着分布式系统经典问题的解决以及优化的资源动态分配和资源恢复技术的成熟,P2P与网格技术必将结合起来以影响整个计算机网络的概念和人们的信息获取模式。

收稿日期:2005-01-18

作者简介



谢东亮,北京邮电大学宽带网研究中心博士后,研究方向:移动互联网控制理论与方法,无线传感器网络性能优化等。



程时端,北京邮电大学网络与交换国家重点实验室教授,博士生导师,研究方向:新一代互联网的性能与服务,移动互联网,WLAN,Ad hoc及传感器网络等。



阎喜珍,北京邮电大学宽带网研究中心副教授,研究方向:互联网业务生成与管理,网络安全。