第1章 P2P网络简介

1.1 什么是 P2P 网络?

计算机是20世纪人类最伟大的发明之一,它的出现改变了人类几千年来对信息存储、表示、处理的方式,也极大地影响了人类生活与思考的方式。

计算机网络(computer network)是相互连接的多台计算机的集合体,人们使用计算机并通过网络互相交流信息,同时扩展计算机的功能。计算机网络的影响深入到人类生活的各个方面,它作为一种新的媒体改变了人类的交流方式,同时也改变了人们对计算机能力的评价——计算机的能力不仅仅在于它的处理速度和存储容量,更重要的是它们之间的连接方式。

因特网(Internet)是计算机网络中最耀眼的明星,它是世界上最大的广域网,连接了地球上几乎每个角落的计算机,是人类最广泛、最有效的信息交流平台。任何一台计算机只要接入因特网,它就潜在地拥有了世界上最大容量的信息仓库和最高速度的计算能力,从这种意义上说;因特网将单台计算机的能力扩展到了极致。

计算机网络自产生以来,其管理与控制就有两种不同的方式:集中式与分布式。集中式系统(centralized system)中的结点在功能上是不平等的,总会有一些通常是少数的管理

结点(server,服务器)在系统中占据中心、主导的地位,管理其他从属结点(client,客户)可执行的操作,控制其他结点之间的信息交换。分布式系统(distributed system)将管理和控制分散到它的各个成员结点中去,结点之间在功能上是平等的,没有谁拥有比其他人更多的特权。集中式系统的优势在于管理的集中化,能够对整个系统进行有效的控制,但它的优势也正是它的缺陷。由于所有结点之间的信息交换都要经过服务器,服务器本身成为限制系统工作效率和规模扩展的瓶颈;分布式系统刚好相反,由于管理的分散化,对整个系统的控制不像集中式那样强,但是由于信息交换的自由、平等,分布式系统常常拥有远远高于集中式系统的工作效率和规模可扩展性。在实际情况中,很多网络系统并不是采用纯集中式或纯分布式的极端方式,而是兼有两方面的特性,称为混合式系统(hybrid system)。

因特网是最大的计算机网络,同样,它自诞生以来也一直存在着集中式与分布式两种不同的工作方式。客户/服务器模式(client/server mode,简称 C/S)是因特网最传统、最成熟的集中式工作模式,许多重要的因特网应用协议(如 HTTP、FTP、SMTP 等)采用了这一模式。在这种集中式的模式下,服务器将一直运行,被动地等待客户的主动接入,客户将请求发给服务器,服务器返回给客户所要的信息。客户/服务器模式在因特网的最初阶段工作得非常好,然而,随着因特网在规模上不断膨胀、在功能上不断扩展,服务器的负担越来越重,客户/服务器模式的低效率与难以扩展的缺陷暴露出来,它不再能适应需要极高效率与巨大规模的现代因特网。

“是人类的需求,真正推动了技术的革命。”每当人类所发明的技术不再能适应人类本身需求的时候,一定会有人提出新的思想、发明新的技术来解决现实与需求之间的矛盾。当传统的客户/服务器模式不再能适应现代因特网需求的时候,人们将目光重新放回到长久被忽视的分布式系统上,对等模式(Peer-to-Peer mode,简称 P2P)正是在这种情况下受到重视并很快成为研究热点。“Peer”一词在英文中的意思是“同等、对等的人”,故而“Peer-to-Peer”译为“对等(计算)”;国内很多人将 P2P 称作“点对点”,这是不恰当的,因为“点对点”是“point-to-Point”,和 P2P 不是一回事。对等模式的本质思想在于打破传统的客户/服务器模式,让一切网络成员享有自由、平等、互联的功能,不再有客户、服务器之分,任何两个网络结点之间都能共享文件、传递消息。图 1.1.1 反映出从 C/S 到 P2P 的转变,Peers 之间的逻辑连接构建在物理连接的基础上。

对等网络(peer-to-peer network,简称 P2P 网络)是分布式系统与计算机网络相结合的产物,是采用对等模式工作的计算机网络。图 1.1.2 描绘了随着分布式系统规模的扩展,分布式计算的模式相应发生的改变。在对等网络中,每个网络结点在行为上是自由的,在功能上是平等的,在连接上是互联的,所有结点分布式地自组织成一个整体网络,因此,它能够极大程度地提高网络效率,充分利用网络带宽,开发每个网络结点的潜力。

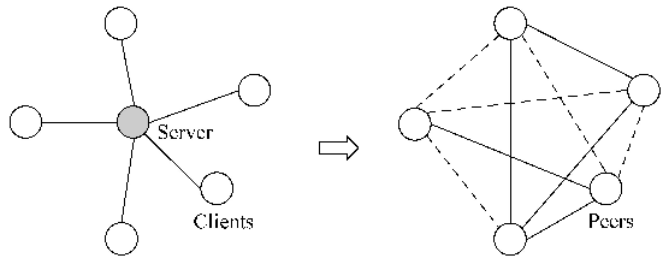


图 1.1.1　从 C/S 到 P2P
（实线表示物理连接，虚线表示逻辑连接）

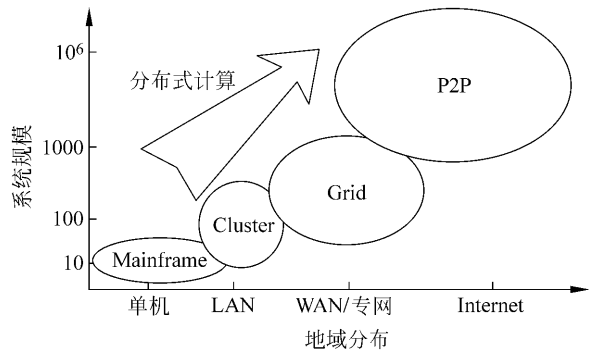


图 1.1.2　分布式计算模式与系统规模的关系
横轴：计算机网络在地域分布上的扩展；纵轴：分布式系统在规模上的膨胀，
斜箭头：适应于不同需求的分布式计算模式；Mainframe：主机计算；Cluster：机群计算；Grid：网格计算，P2P：对等计算

P2P 的思想起源很早，我们用“Google 学术搜索”（http://scholar.google.com）找到最早提及 P2P 的文献发表于 1956 年，从那以后几乎每年都有与 P2P 相关的文章，但一直未成为热点。任何一种思想、理论的流行通常都需要一个杀手锏（killer application），以一种征服性的力量冲击人类的传统思维。P2P 的杀手锏正是出现于 1999 年的世界上第一个应用性对等网络 Napster，它创造了在半年时间里拥有 5000 万用户的网络奇迹，向整个世界展示了 P2P 优异的性能和巨大的潜力。

学术的脚步常常先于应用踏入某个领域，又往往在应用之后成为热点，P2P 和 Napster 的关系正是如此。在 Napster 之后，是一系列人们耳熟能详的 P2P 网络软件：Gnutella，KaZaA，BitTorrent，eDonkey/eMule，Skype，等等。虽然从 1999 年到现在只有短短几年，但是由于在工作模式上具有的优势和对于现代因特网的适应性，P2P 得以迅速从一个民间小软件发展为计算机网络的一项重要技术，在应用领域和学术界获得了广泛的重视和成功，并占据了当前 Internet 超过一半的带

宽资源，被认为是“改变 Internet 的新一代网络技术”。

对于已经应用或正处于理论研究阶段的各种 P2P 网络，国内外的研究者从多个不同的角度对它们进行了分类，包括从体系结构、出现时间角度和应用领域角度进行的分类，到目前为止尚未出现公认的、明确的分类方法。本书从 P2P 网络设计思想出发，兼顾体系结构和出现时间两个方面的考虑，将 P2P 网络分成三代：

第一代，混合式 P2P 网络，它是 C/S 和 P2P 两种模式的混合；

第二代，无结构 P2P 网络，它以分布、松散的结构来组织网络，故称“无结构”；

第三代，结构化 P2P 网络，它以准确、严格的结构来组织网络，并能高效地定位结点和数据。

对这三代 P2P 网络的讲解，是本书的一大重点，分别对应第 2、3、4 章的内容。读者需要注意的是：我们的分类仅出于本书行文、讲解的考虑，并非 P2P 领域明确的界定。

现代计算机网络均采用层次化的结构，以提供一个便于分析的模型和利于开发的技术接口，它具体地表现为一个网络通信从高层到低层的协议栈。这里面最为著名的是 ISO/OSI(国际标准化组织/开放系统互联)模型和 TCP/IP(传输控制协议/因特网协议)模型，前者细致、正式，后者更为实用。在本书中对于层次化的网络结构描述均采用 TCP/IP 模型，如图 1.1.3 所示。

图 1.1.3 中 TCP/IP 模型共分四层：网络接入层、网络层(IP 协议)、传输层(TCP/UDP 协议)和应用层。在此之前所说的集中式系统、分布式系统、客户/服务器工作模式以及对等计算模式，都是指四层中的最高层——应用层的工作方式，而下面的三层通常采用标准、单一的工作方式，本身并没有集中式与分布式之分，只是为应用层不同的工作方式提供底层的服务支持。

应用层	(C/S，P2P)
传输层	(TCP/UDP 协议)
网络层	(IP 协议)
网络接入层	(链接协议与物理协议)

图 1.1.3 TCP/IP 协议栈

P2P 网络的核心机制，是在应用层建立逻辑上的覆盖网络(overlay network)，封装下面的三层，让 P2P 网络的研究者和开发者不必关心下面三层是如何工作的，而仅仅去考虑应用层覆盖网络的工作情况，将精力集中于覆盖网络的设计、优化上。虽然如此，在对 P2P 网络做基础的研究和设计时，有时还是要考虑到下面层的工作情况，因为应用层建立的覆盖网与底层实际的物理网的工作情况不可能完全相同，在图 1.1.4 中，覆盖网上的一条逻辑连接 AE 对应物理网上三条物理连接：AC，CD 和 DE。所以从覆盖网看到的行为与底层物理网实际的行为并不一致。P2P 领域的研究者已经对这种不一致性做了大量的工作以尽可能减少两者之间的差异，提高整个网络的效率。简言之：P2P 网络工作于应用层，但兼顾网络底层。读者在阅读本书时应该注意到这一点，这一问题在以后的章节中会有详述。

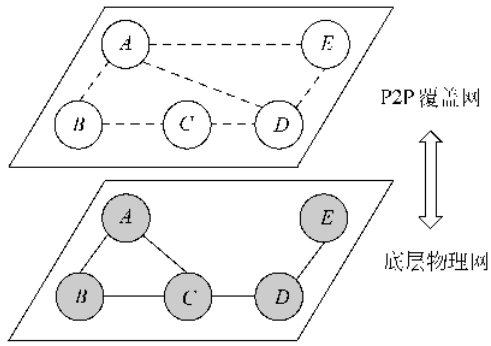


图 1.1.4　P2P 覆盖网和底层物理网的不一致性
（虚线表示逻辑连接，实线表示物理连接）

在 P2P 覆盖网上依靠 DHT(distributed hash table，分布式散列表)通常能准确、快速地路由消息和定位数据对象，这正是 P2P 查询的优势所在。“祸兮，福之所倚；福兮，祸之所伏。”正如计算机领域那条著名的“没有白吃的午餐定理”(No Free Lunch Theorem)[Wolpert and Macready，1997]所述任何方法对一类问题做得好，必然对另外某类问题做得不好，这就是代价。使用覆盖网和 DHT 的 P2P 网络，为追求性能和效率所付出的代价，是语义模糊查询的困难以及对动态网络环境中错误行为的容忍性下降。针对前者，语义模糊查询至今仍然是 P2P 领域的一个开放型问题，尤其对结构化 P2P 网络更为困难；针对后者，P2P 领域的研究者提出的方案各具特色，总体上分两类：①网络周期性主动更新；②在检测到错误后被动更新。前者机制简单而开销很大，后者开销较小而机制复杂，这就带来了两难问题。

前面描述了 P2P 的核心机制，有了它们，一个 P2P 网络能正常工作，但不见得“好用”，因此人们又提出了很多 P2P 的“增强机制”，以改善网络状况，让它更好地工作。这些增强机制不少是从分布式系统或者计算机网络领域借鉴过来的，如数据复制、缓存、分片等；当然，更多的增强机制来自 P2P 本身，如负载的均衡、异构性的开发、少数结点负担过重导致的“热点”问题、物理网与覆盖网不一致造成的“拓扑意识”问题、保护用户隐私的“匿名”问题、P2P 用户的“声誉”和“信任”问题以及最令人不放心的 P2P“安全”问题。

“纸上得来终觉浅，绝知此事要躬行。”理论是灰色的，生活之树常青。理论并不都现实可行，即使真的实现了，也未必好用，所以做研究注重理论结合实践，强调实践的作用。对 P2P 来说，由于其网络规模巨大，开发实际系统的软硬件开销巨大，因此 P2P 实验更侧重于“模拟”和“仿真”。

“图难于其易，为大于其细。”本章的后续部分，将从“大”而“易”的方面讲述 P2P 的历史、现状、缘由、特点、应用和著名模型；而本书后面的各章，则从“细”而

“难”的方面去具体地讲述三代 P2P 网络(第 2、3、4 章)，列举当前世界范围内的各种 P2P 应用体系和应用软件、介绍一些著名软件的使用(第 5 章)，研究 P2P 技术的核心机制、增强机制、模拟和仿真(第 6、7、8 章)，最后总结 P2P 的现状并展望 P2P 的未来(第 9 章)。我们给每位读者的建议是——把握本书的脉络，理清 P2P 的历史，有选择性地阅读，需要的时候回过头来参考。

1.2 P2P 网络的发展历程

1. 第一阶段：1999—2000 年，民间软件，锋芒初现

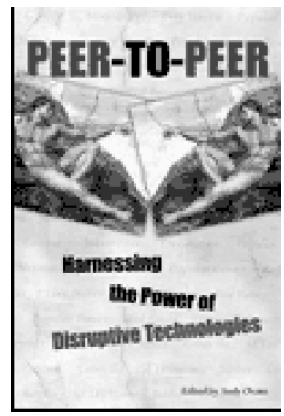
1999 年，18 岁的 Shawn Fanning 开发了世界上第一个应用性 P2P 网络软件 Napster，在半年时间里即拥有 5000 万注册用户。Napster 是第一代 P2P 网络——混合式 P2P 体系(hybrid P2P architecture)最杰出的代表，向整个世界传达了 P2P 优秀的思想，展现了 P2P 巨大的潜力。不久以后，Napster 网站因为版权问题被推上法庭，此后一直官司不断，在经历了约两年的法律纠纷之后，2001 年底，流星般的 Napster 最终关闭了。

2000 年 3 月，第一个无结构 P2P 网络 Gnutella 诞生于 NullSoft 公司，它是第二代 P2P 网络——无结构 P2P 体系(unstructured P2P architecture)的代表。虽然发布之后不久，Gnutella 就因为其母公司担心法律问题而被关闭，但是 Gnutella 所代表的无结构、纯分布式 P2P 体系的思想却广泛流传开来，而 Gnutella 本身成为一种典型的无结构 P2P 网络协议。

几乎与 Gnutella 产生的同时，以“自由、安全、匿名”著称的无结构 P2P 网络——Freenet 推出，它从一开始就有着很不同的理念：Napster、Gnutella 这类 P2P 系统的主要目的在于交换文件，而 Freenet 的目的是共享 Internet 上的计算机资源，组建一个不受限制、不受审查的信息发布和获取的平台。虽然 Freenet 拥有前卫的思想和高超的技术，但由于其使用的复杂和过于理想的自由主义理念，Freenet 在大多数国家被严格管制。

Gnutella 问此后不久还出现了新的无结构 P2P 应用系统 KaZaA，但与短命的 Gnutella 不同，KaZaA 从开始到现在，其用户群不断扩大，号称拥有超过 300 万的平均在线用户，KaZaA 网站上的统计数据称：截至 2007 年 3 月，KaZaA 软件已被下载超过 3.8 亿次，毫不逊色于当年的 Napster。KaZaA 基于著名的 FastTrack 协议，该协议是与 Gnutella 并列的典型的无结构 P2P 协议，其最大的改进在于引入“超结点”(SuperNode)，从而第一个有效地开发了 P2P 网络中的结点异构性(即结点之间在能力上的差异)。基于 FastTrack 协议的应用还有 KaZaA 的类似体 KaZaA-Lite，以及 Grokster、iMesh 等。

eDonkey(电驴)差不多和 Gnutella 同时出现(2000 年),它提供文件分块下载,从而每个文件可以从其他多个用户并行下载。eDonkey 网络结构很像 KaZaA,它采用"服务器"(不同于一般意义上的服务器)作为网络的核心部分,"服务器"的作用很像 KaZaA 中的"超结点",每个用户只需连接到特定的"服务器"来共享和获取文件。eDonkey 从一开始就吸引了很多 Internet 用户,到今天仍然非常流行。



2000 年 8 月,著名出版人 O'Reilly 组织了一次意义重大的 P2P 峰会,来帮助人们认识 P2P 的潜力并消除 Napster、Gnutella 造成的"P2P 是盗版技术"的负面影响。参与 P2P 峰会的既有 Napster、Gnutella、Freenet 的开发者,也有那些试图挖掘 P2P 分布计算能力的公司和组织如 Popular Power、SETI@Home、Distributed.net 等。O'Reilly 认为目前 P2P 的状态类似于"盲人摸象",P2P 技术的领导者们每个人都看到了 P2P 这头"巨象"的一些特征,如果他们能够有机会交流思想,P2P 将会更快地发展。这次峰会主要有三个目的:①诠释 P2P,说明要从中得到什么;②描述 P2P 的作用,P2P 能解决什么样的问题;③形成一个提供给大众的关于 P2P 的信息,消除那些负面影响。在那次峰会后不久,O'Reilly 于 2001 年出版了目前所知最早的关于 P2P 网络的书籍 *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*(见右上小图)。这本书包含了有关 P2P 的诸多话题:P2P 的起源、Napster、SETI@Home、Jabber、Gnutella、Freenet、Red Rover、Publius、Free Haven、元数据、缓存管理、信任机制、声誉、安全性等,它主要的不足是内容陈旧、缺乏代表性,因为那时对 P2P 网络的研究实际上还处于萌芽期。

同是 2000 年 8 月,Intel 公司宣布成立 P2P 工作组,正式开展 P2P 的研究。工作组成立以后,积极与应用开发商合作开发 P2P 应用平台。2002 年 Intel 发布了.NET 基础架构之上的 P2P Accelerator Kit(P2P 加速工具包)和 P2P 安全 API 软件包,从而使得利用微软.NET 开发软件的人员能够迅速地建立 P2P 安全 Web 应用。

IBM、HP 等公司也是 Intel 成立的 P2P 工作组中的成员。这两家公司在 2000 年 9 月共同推出了一种开放存储技术,这一存储技术利用了 P2P 技术,可以方便地从用户的硬盘向服务器上复制数据。HP 公司还把 P2P 的立足点放在打印技术上,该公司新推出的网络打印技术可使用户通过 P2P 网络共享打印机。

2. 第二阶段:2001—2003 年,步入正统,群雄逐鹿

如果说 1999 年属于第一代混合式 P2P 网络的辉煌、2000 年属于第二代无结

构 P2P 网络的成功，那么，2001 年则是第三代 P2P 网络——结构化 P2P 体系(structured P2P architecture)展现的舞台。这一年学术界真正开始关注和重视 P2P 网络，IEEE 成立 P2P 专业会议，ACM 在网络通信领域最具影响力的几个会议发表了多篇有关 P2P 的经典论文，P2P 领域最具代表性的经典模型和应用体系被提出，其中有：Chord、CAN、Tapestry、Pastry、CFS、OceanStore、PAST 等；同时，大多数知名的学术团体和技术组织成立或者完善了专门的 P2P 研究组，其中有：MIT 的 Chord/CFS 研究组、UC Berkeley 的 Tapestry/OceanStore 研究组，Microsoft Research 和 Rice University 的 Pastry/PAST 研究组、Stanford Peers 研究组等。

2001 年，Ray Ozzie(著名的 Lotus Notes 软件的开发者)创立了 Groove Networks 公司，开发 Groove Virtual Office(Groove 虚拟办公室)，此软件意在使用 P2P 技术营造一个 Internet 协同工作空间。2005 年 3 月，Groove 公司被软件巨人微软以 1.2 亿美元收购，而 Groove 的创始人 Ray Ozzie 成为微软 CTO(首席技术官)。实际上早在 2001 年，Groove 公司就曾接受过微软高达 5100 万美元的投资，所以 2005 年的收购并不值得奇怪。在微软最新推出的 Office 12 办公软件中，已整合了 Groove 软件：处于测试中的 Microsoft Office 12 共包含 14 个组件，而 Groove 占据其中 3 个。

2002 年，P2P 专业会议 IPTPS 首次召开，会议的规模不大，但影响力不小。SIGCOMM、SPAA、PODC、INFOCOM、ICDCS 等网络通信、分布式系统领域的重要会议继续关注 P2P，甚至有些设置了 P2P 专题讨论会，新的 P2P 模型如 Kademlia、Viceroy 等被提出，它们在理论上很有意义。

2002 年 5 月，由于不满意当时的 eDonkey2000 客户端软件，并且坚信能做出更出色的类似客户端，Merkur 聚集了一批原本在其他领域有出色发挥的程序员到他的周围，开始了著名的“eMule 工程”。eMule 这个名称“电骡”表明了它和 eDonkey 的关系——承继关系，但 eMule 更出色。他们没有想到的是，不久以后 eMule 的流行性甚至远超过它的前驱 eDonkey。

2002 年 10 月，新一代的混合式 P2P 网络 BitTorrent 推出，到 2003 年 BT(BitTorrent 简称)已在世界范围内(尤其是在中国)广泛流行。2006 年底，中国互联网络信息中心(CNNIC)发布的中国互联网统计报告显示：中国 1.11 亿网民中有 27.8% 的人使用过 BT 软件(超过 3000 万人)，其流行程度由此可见一斑。BitTorrent 使用基于文件的分散式服务器，共享同一文件的用户构成一个独立的子网，所以 BitTorrent 既不会因为一台服务器失效而影响整个网络，也不会像 Napster 那样被关闭网站(BT 网站太多且分散在世界各地)；同时，分片优化、阻塞控制等方法使得 BT 能够充分利用网络资源。

自 2003 年开始，P2P 的发展实际上进入一个稳定期。在解决了 P2P 网络最核心的问题后，学术界将重点放在其性能增强、安全问题和实用系统开发上。这里

值得一提的是由 MIT 的 Frans Kaashoek 教授领衔的研究小组，他们联合其他美国一流高校和研究机构进行的 IRIS 项目（Infrastructure for Resilient Internet System，容错的 Internet 系统架构），用 P2P 的方法去研究并建立新一代互联网络结构，得到了 2003 年美国 NSF（自然科学基金）在 IT 领域最大的一项基金资助。另一方面，商业领域更多地在改进过去的 P2P 应用软件，很多混合式、无结构的 P2P 网络将学术界结构化 P2P 体系的思想整合进来，如 BitTorrent 就整合了 Kademlia 的分布式散列表，这种融合体现了 P2P 学术界对商业界的影响，是难能可贵的。

2003 年，在无结构 P2P 方面 Gnutella 协议 0.6 版发布，它比 Gnutella 协议 0.4 版扩展了很多，比如以明确的形式建议“超 Peer”（UltraPeer）的使用。在结构化 P2P 方面，新的 P2P 模型 SkipNet、Koorde 等被提出，SkipNet 是第一个显式提供路由局部性、对象语义局部性的结构化 P2P 模型，Koorde 则在理论上具有很高的价值，证明了一些 P2P 领域悬而未决的结论。

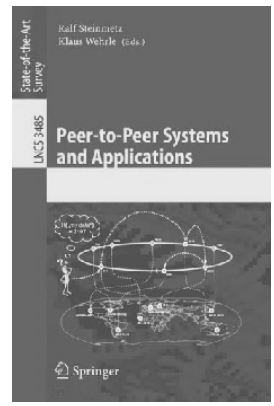
2003 年，商业领域诞生了一颗璀璨的新星——Skype 公司，它是全球第一家 P2P 即时通信公司，采用 P2P 技术为用户提供免费或廉价的语音通话服务，使用端到端的加密技术保证通信的安全可靠性。仅一年时间，Skype 用户就超过了 1300 万，迄今为止注册用户超过 2300 万，同时在线用户数高达 100 万，而且新用户以每天 6 万的速度增长，其迅猛的发展速度再度证明了 P2P 的巨大潜力。

软件领域的巨人微软公司向来不愿错过每一项热点技术。为了使 Windows 操作系统更好地与 P2P 应用协调，Microsoft 对 Windows XP 进行了强化，公布了“Windows XP P2P 软件开发包”，这个编程工具使软件开发商能够更容易地在 Windows XP 上编写 P2P 应用程序。2003 年，Microsoft 并购了 P2P 新公司 XDegrees，声称将把 XDegrees 的技术用于它的存储产品，随后，Microsoft 推出了一款面向年轻人的即时通信软件“Three degrees”，大刀阔斧地挺进 P2P 市场。微软在全球各地的研究院基本上都将 P2P 列入其工作重点并发表了多篇有价值的论文，这里面做得最好的是微软剑桥研究院（Pastry/PAST 系列），微软亚洲研究院的系统研究组在 P2P 方面也很不错。

3. 第三阶段：2004—2006 年，国际国内，风起云涌

2004 年，在 IPDPS 会议上基于 CCC 拓扑的 Cycloid 常数度 P2P 模型被提出，它兼有常数度、超立方体、环形多种属性，可以看成是对此前诸多结构化 P2P 模型的一个总结。从 Cycloid 的提出可以看出，P2P 网络的主要问题、核心机制、整体架构已经形成，P2P 研究者对这些重大问题已经形成共识，下一步要做的工作，应该主要放在更细节、更高效、更实用的方面。另外，如何将各种不同的 P2P 应用系统像 Web 那样整合起来（2005 年发表于 SIGCOMM 会议的 OpenDHT 服务体系

在这方面做了努力），甚至将 P2P 和 Web 整合起来，也是一件非常有意义的大事。



2005 年底，Springer 在其 LNCS（Lecture Notes in Computer Science）系列中出版了一本内容丰富、涵盖面非常广阔的 P2P 专著 *Peer-to-Peer Systems and Applications*。全书共 33 章，分 10 个部分，每章由不同作者编写，涉及 P2P 的概念、历史、结构、应用、自组织、检索、性能、移动环境应用、商业应用与市场等多方面内容，提供了对整个 P2P 领域权威、全面的总结、分析与展望。此外，P2P 的大师 Ion Stoica 将意为上述专著绘制了封面（见右图）并撰写了前言，使该书增加了影响力。

我们再将目光转到国内的 P2P，首先来看学术界的情况。国内一些高校、科研组织从 2003 年前后开始 P2P 的研究开发工作，其中有影响的是：北京大学网络实验室开发的“Maze 网络文件系统”、华中科技大学集群与网格计算实验室开发的“AnySee 视频直播系统”（Maze 和 AnySee 已投入应用，尤其在高校学生中相当流行）、清华大学高性能计算研究所开发的“Granary 广域网分布式存储系统”（处于开发优化进程中）。此外，仅以 2005、2006 年的国家自然科学基金资助项目为例，P2P 相关项目就有数十项，表 1.2.1 是一份不完全的统计。

表 1.2.1　2005、2006 年国家自然科学基金资助的 P2P 相关项目（不完全统计）

年份	批准号	项目名称	依托单位
2005 年	60503045	P2P 网络自主行为模型研究及其应用	上海交通大学
	60503047	基于语义链的对等网络模型及异构数据管理方法研究	中国科学院计算技术研究所
	60573053	P2P 系统的信任管理和匿名问题研究	中国科学院研究生院
	60573096	基于模式的高可扩展性 P2P 数据管理技术的研究	西北工业大学
	60573110	基于结构化对等网络的全文搜索技术	清华大学
	60573120	P2P 网络的关键安全问题研究	华中科技大学
	60573127	基于 P2P 思想的网格计算资源动态管理与任务调度研究	中南大学
	60573129	基于泛洪的非结构化 P2P 系统中分布式拒绝服务攻击防范方法的研究	电子科技大学
	60573131	实用化对等网络技术的研究	南京大学
	60573140	自适应的基于分布式索引缓存的无结构 P2P 快速搜索算法的研究	温州大学

续表

年份	批准号	项目名称	依托单位
2006 年	60673071	P2P 流媒体点播中的网络波动问题研究	清华大学
	60673156	IPv6 环境下基于 P2P 的 DDoS 分布式防御研究	湖南大学
	60673166	非结构化对等网络高性能动态查询响应关键技术研究	上海交通大学
	60673167	基于对等模式的网格资源定位关键技术研究	中国人民解放军国防科学技术大学
	60673179	超级非结构化(Superpeer)P2P 网络动态层次优化机制研究	浙江工商大学
	60673183	P2P 文件共享系统中信誉机制的研究	北京大学
	60673184	基于对等网络技术的安全、高性能流媒体系统的理论研究	清华大学

作者所在的南京大学计算机科学与技术系 GPS(Grid、P2P、Sensor)研究组，一直致力于 P2P 理论的钻研和 P2P 技术的开发工作。在理论方面，我们一直紧密关注 P2P 领域的新理论、新技术和发展趋势，重点跟踪 SIGCOMM、SPAA、PODC、ICNP、INFOCOM、ICDCS、IPDPS、IPTPS 这几大重要国际会议，并在不少国内国际的会议刊物上发表我们的研究成果。在技术开发方面，我们主要进行两个项目的工作：一是 2005 年国家自然科学基金项目“实用化对等网络技术的研究”，二是 2005 年江苏省自然科学基金前期预研项目“新型 P2P 计算技术的基础研究”。这些项目着眼于多项 P2P 新技术，其研究内容涉及新型 P2P 拓扑结构、可控散列函数、拓扑意识和一致性问题、容错性保证(崩溃点测量、覆盖网分割问题)、声誉与信任机制、安全检测与恢复等。

国内商业领域最近几年内出现了很多国产 P2P 应用软件，其中不少具有相当的流行性，这是国内 P2P 研究中非常可喜的现象。国产 P2P 文件共享软件如 PP 点点通、PoCo、百宝等都有很大的用户群，而国产 P2P 网络电视软件 PPLive、TvAnts 等在用户规模上也毫不逊色，从中可以看出国内计算机界在 P2P 方面所做的努力，以及国内计算机用户对 P2P 的认识与接受程度。虽然如此，目前众多国产 P2P 软件在技术上还很不完备，性能上也有很大的改善余地，与相对成熟的国外 P2P 软件相比还有不小的差距。

1.3 为什么需要 P2P 网络

P2P 网络使得网络工作模式从集中式走向分布式，网络应用的核心从服务器走向每一个网络结点，从而使人们在网络上的信息交流被提升到了一个更高的层

次，使人们以更主动深刻的方式参与到网络中去。我们需要 P2P 网络，是因为它新颖的工作模式很好地适应了不断膨胀、不断扩展的 Internet 动态网络环境。P2P 网络的优点可归纳如下。

1. P2P 提高了网络工作效率

在 C/S 模式中，客户只能与服务器交换信息，如果两个客户之间要传送一个文件，通常是首先将文件传到服务器，然后由服务器传给另一个客户（或者说另一个客户从服务器下载文件），这在无形中增加了一个不必要的环节（最典型的如 FTP 文件传送）。如果服务器忙，文件传送将变得十分缓慢，更严重的情况下如果服务器故障，文件将没有办法传送，C/S 模式固有的问题影响了网络工作效率。P2P 模式与传统的 C/S 模式最大的区别在于没有集中式的控制，任意两个结点之间交换信息不需要经过一个固定的服务器，因此上面所说的问题对 P2P 网络来说是先天不存在的。

P2P 网络高效的另一重要原因是它在网络应用层构建了一个有（严格）拓扑结构（topology structure）的覆盖网，并且通常使用基于一致性散列函数（consistant hashing）的分布式散列表（DHT），将网络结点或数据对象高效、均匀地映射到覆盖网中。覆盖网与分布式散列表的结合，使得 P2P 网络具有很高的路由效率：比如对结构化 P2P 网络，任意两个结点间定位所要经过的覆盖网路由跳数在 $\log N$ 左右，N 为网络结点总数，这是一个非常理想的结果。虽然如此，正如 1.1 节所述，由于覆盖网与因特网的不一致性，覆盖网上一跳可能对应因特网上多跳，实际的 IP 路由跳数可能高于 $\log N$，许多研究者提出不同方案解决这一问题，努力缩小两者之间的差异，将实际的 IP 路由跳数控制在 $\log N$ 的常数倍范围之内，这样的路由效率仍然是很不错的。

2. P2P 充分利用了网络带宽

有个小故事说，硅谷有个做网络通信的腰缠万贯的商人，临死前竖着一根手指，他的子女不解就问他还有什么遗愿，他放下手指说了句“带宽”就断气了。对于 Internet 而言，带宽确实是最宝贵的资源；而在有限带宽的情况下，如何充分地利用带宽则更为关键。很多人都有这样的体会：明明网络带宽很高，下载某个文件的速度却低得出奇，就好像拿一个大木桶去等滴水的龙头，不是这头的桶不够大，而是那头存在瓶颈。

P2P 网络中没有 C/S 模式下的服务器，所以不存在 C/S 模式下服务器造成的“效率瓶颈”（efficiency bottleneck）——既没有不必要的中间环节，也不会因为服务器忙而不得不等待，更不会出现由于服务器“单点失效”（single point of failure）

导致整个网络通信中断的情形。在 P2P 网络中，任意两个对等结点之间平等地互联，在交换信息的过程中不受其他结点的控制与影响，数据传输速度通常只取决于网络带宽，因此它对网络带宽的利用是充分的。以最流行的 P2P 应用 BT 为例，当使用 BT 下载多个文件时，下载速度常常能达到或者接近网络带宽，而反过来使用 HTTP 和 FTP 则很少有这样的效果。

3. P2P 开发了每个网络结点的潜力

一个网络结点对于网络潜在的贡献，是它可以提供的计算能力和存储容量。传统的 C/S 模式形成了以服务器为核心的网络——数据集中存储在服务器上，绝大多数的计算任务由服务器完成，客户对网络计算能力和数据存储的贡献微乎其微。P2P 的出现将网络的核心从服务器转变为每一个网络结点——数据分散地存储在所有结点上，计算任务由各个结点分布、协同地完成，每个结点都是网络的主体和重要成员。

世界上绝大多数计算机都不会像专用服务器那样永久地连在网上，这些临时性的动态结点称为“网络边缘结点”(network edge node)，而它们所提供的计算能力(空闲处理器周期)和存储容量(剩余存储空间)称为“网络边缘资源”(network edge resource)，实际上这些临时性的网络边缘资源才是网络资源的真正主体。P2P 网络的主体是网络边缘结点，它是第一个真正而又充分地利用网络边缘资源的计算机网络，P2P 网络相对于传统网络的优势和对现代因特网卓越的适应性正是来源于此。所以说，P2P 开发了每个网络结点的潜力，使得因特网的存储模式从“内容位于中心”转变为“内容位于边缘”，计算模式从“服务器集中计算”转变为“分布式协同计算”。

4. P2P 网络具有非常高的可扩展性

计算机网络的可扩展性(scalability)通常用下面几个因素来衡量：当网络结点总数增加时，①结点负载如何改变；②为适应规模扩大而需要增加的额外设备的数量；③任意两个网络结点通信效率如何改变，尤其是路由效率。

C/S 模式的缺陷在于它的难扩展性，当网络结点数目大量增加时，服务器的负载也随之线性地增加，虽然服务器通常都是高性能计算机，但是性能再高的超级计算机也没有办法应付网络规模的不断膨胀。当原先的服务器不堪重负时，必定要购买新的服务器来替代它或者分担它的负载，然而高性能服务器的价格是昂贵的，额外设备的开销会非常大。同时，由于连入同一服务器的客户数目增多，通信效率平均而言一定会下降，并且也增加了服务器过忙和故障的概率。

P2P 模式具有非常高的可扩展性。首先，当网络结点数目大量增加时，随之增

加的通信开销被更多的结点分担，所以每个结点承担的负载并不会增加太多。其次，在网络规模扩大时 P2P 网络不需要增加额外设备。再次，P2P 网络路由跳数（针对结构化网络）的典型值为 $\log N$，所以随着 N 的增加路由跳数会增多，但是增量非常少，通信效率仍然保持在较高的水平。

5. P2P 网络具有良好的容错性

P2P 网络所构建的覆盖网和分布式散列表极大地提高了网络的工作效率和可扩展性，但是，覆盖网拓扑结构越严格，容错性（fault resilience）通常就越差，因为严格的拓扑结构对其成员结点不正常行为的容忍性相对较低。针对这一缺点，研究者提出了各种增强机制，最典型的是冗余方法（redundancy）和周期性检测（periodical detection）：前者通过保存适当的冗余信息，提供有效的替代，以空间换取容错；后者通过每个结点周期性地检测，来及时纠正错误，以时间换取容错。

为了实现高效的路由，P2P 网络中每个结点通常会保存一个路由表和其他辅助性的信息，记录它的网络邻居等，这些信息称作“结点状态”（node state），它在结点加入网络时被初始化。由于网络环境的动态性，网络结点不断地加入和离开，结点状态会变得陈旧，与实际网络不一致，这种不一致影响网络的工作效率。人们设计了很多方法来及时更新结点状态，这一机制称为结点的“自适应”（adaptability），它是保证网络容错性的基础。周期性探测是最典型的自适应操作，直接的办法是采用短周期频繁探测使得结点状态始终保持在最新，但是这样做开销很大；改进的办法是采用长周期松散探测，虽然不能保证结点状态最新和网络最高效，但是能保证结点在较新的状态下仍然能正确且比较高效地工作。除此之外，还有一些结点自适应的新方法，如 Kademlia 协议采用的消息“捎带确认”（piggybacking），将结点状态信息捎带在每条路由消息中发布，接收者由此确认自己的状态信息是否已过时。

1.4 P2P 网络的特点

P2P 网络的特点整体上可以概括为三个词：自由、平等、互联。

（1）自由　在 P2P 网络中，对等结点做什么事情、采取什么样的行为、与其他结点交换哪些信息，由其本身自由决定，而不受限于其他网络结点。另一方面，由于采用分布式散列表的 P2P 网络特有的匿名性，保护了发布者的信息，使得用户能够更加自由、没有后顾之忧地参与到网络中来。

（2）平等　平等是 P2P 网络最重要的特性，是“对等（计算）”名字的由来。平等意味着在一个系统中，虽然能力不尽相同（即结点间的“异构性”（heterogeneity）），但所有成员在功能、地位上都是平等的，没有谁拥有特权，没有谁能控制或限制其他

结点。就网络组织方式而言，平等指的是打破传统的客户/服务器模式，取消服务器这一特权结点的存在，让所有网络成员之间平等地交流信息。平等性是 P2P 网络的工作基础，P2P 网络对网络带宽的高效利用、对网络结点潜力的充分开发以及可扩展性等，都是基于平等性的。

（3）互联　互联的本质原因是 P2P 在应用层构建了覆盖网。P2P 网络中任意两个对等结点间都可以建立连接，这是覆盖网上的一条逻辑连接，它通常对应物理网上的一条 IP 路径（或者说是传输层的一个 TCP 连接）。在 C/S 模式下客户只能和服务器建立一条 Client-to-Server 的连接，而在 P2P 模式下任何两个对等结点间都可以建立一条 Peer-to-Peer 的连接。互联性源于平等性，它也是 P2P 网络高效率、高可扩展性的重要基础。

更具体地说，P2P 网络区别于其他系统的本质特点如下。

1. 网络拓扑结构严格

P2P 网络在网络应用层构建了一个有（严格）拓扑结构的覆盖网，覆盖网拓扑结构对于一个 P2P 网络具有基础性的意义，系统的其他许多机制如分布式散列表、路由、负载均衡、容错与自适应、自组织都以它为基础。具体来说 P2P 网络通常采用图 1.4.1 所示的几种拓扑结构。

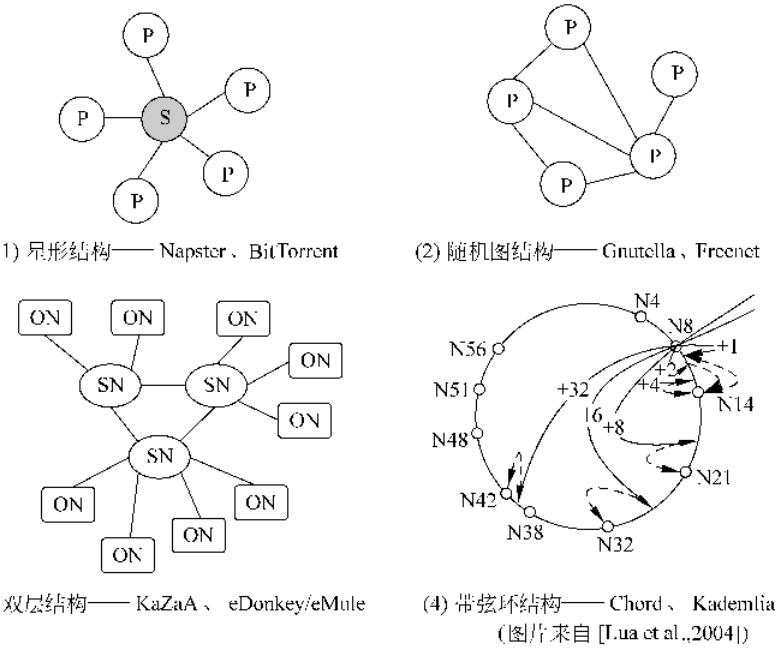
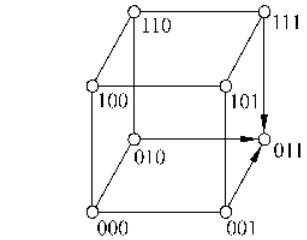
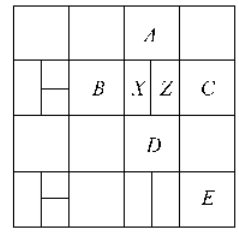


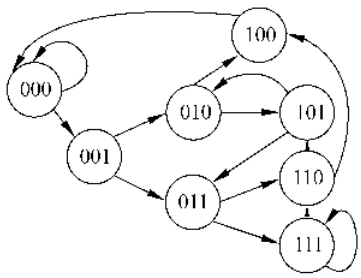
图 1.4.1　P2P 网络通常采用的拓扑结构



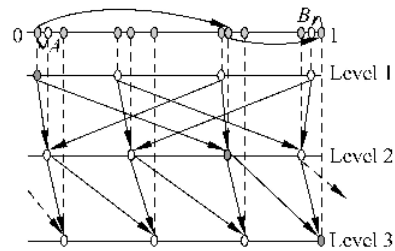
(5) 超立方体结构——Tapestry、Pastry
（实际上是变形的网格结构）



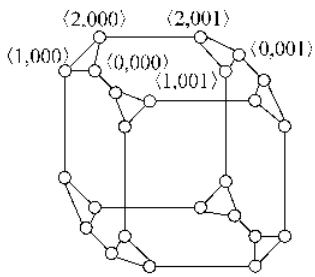
(6) 多维空间结构——CAN
（图片来自[Lua et al.,2004]）



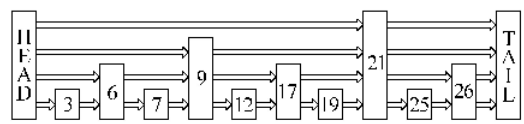
(7) de Brujin图结构——Koorde



(8) 蝴蝶结构——Viceroy
（图片来自[Lua et al.,2003]）



(9) CCC结构——Cycloid



(10) 跳表结构——SkipNet（图片来自[Harvey et al.,2003]）

图 1.4.1 （续）

2. 结点和数据对象位置确定

分布式散列表(DHT)是P2P网络的核心设施。如图1.4.2所示，它通常基于一致性散列函数，提供对于任何一个结点、数据对象在覆盖网中的位置映射。这一点在结构化P2P网络中尤其重要，因为它保证了能够准确地定位到某个结点或者数据对象。具体地说，如果分布式散列表采用一致性散列函数 $H()$，对于某个网络结点(IP,Port,…)，该结点在覆盖网上将有唯一对应的“结点标识”nodeID＝H(IP,Port,…)，IP为结点IP地址，Port为端口号，…表示其他属性；对于某个数据对象(Key,…)，它在覆盖网上也有唯一的“对象标识”objectID＝H(Key,…)，Key为对象关键码，…表示其他属性。对于结点而言，nodeID确定了它的覆盖网

位置，对于数据对象而言，objectID 确定了它的索引信息在覆盖网上的存放位置。

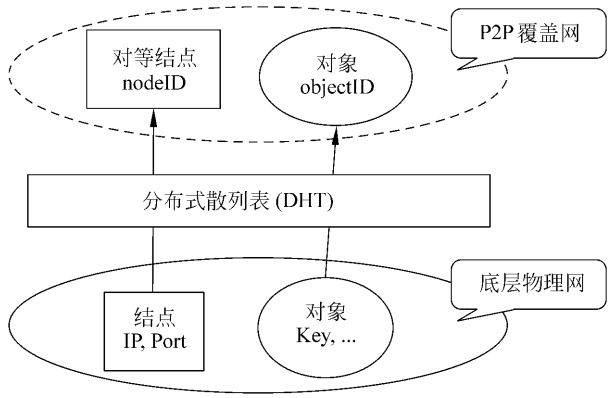


图 1.4.2　DHT 在 P2P 系统中的位置

3. 高效路由

基于 P2P 覆盖网与分布式散列表，P2P 网络通常有适合自己的路由算法，以保证高效路由。任意两个结点间定位所需的覆盖网路由跳数典型地为 TTL(无结构网络)或 log N(结构化网络)，TTL(time to live)为跳数限制，N 为网络结点总数。因为覆盖网与物理网的不一致性，实际 IP 路由跳数会高于覆盖网路由跳数，但仍可以控制在 log N 的常数倍范围之内。具体地说，P2P 系统的路由方法大致有：

(1) 服务器路由——客户直接发送消息给服务器，服务器返回所需信息。

典型的系统有 Napster、BitTorrent；路由跳数为 $O(1)$。$O()$为渐进增长率符号，若函数 $f \in O(g)$，表示函数 f 渐进增长的速率不超过函数 g。数学上更严格地表述为：如果 $f \in O(g)$，那么 $\lim\limits_{N\to\infty}\dfrac{f(N)}{g(N)}=c$，$c$ 为某个有限常数。

(2) 无结构路由——结点以洪泛法或者类似的方法发送消息给自己的每个邻居，邻居收到消息后也这样做下去，直到定位成功，通常会有跳数限制 TTL 以控制路由范围。典型的系统有 Gnutella、Freenet；路由跳数为 O(TTL)。

(3) 双层路由——通常为双层结构的 P2P 网络所使用，“普通结点”直接发消息给“超结点”，“超结点”之间使用无结构路由。典型的系统有 KaZaA、eDonkey/eMule；路由跳数为 $O(1)+O$(TTL)。

(4) 数值邻近路由——这里的“数值”通常指结点 ID 值，路由过程中每一步，当前结点都在自己的路由表中选择与目的 ID 最邻近的结点作为下一跳。典型的系统有 Chord、Kademlia、SkipNet(NameID 路由)；路由跳数为 $O(\log N)$。

（5）逐位匹配路由——逐位匹配路由基于层次化的路由表，每一步通常都能与目的 ID 多匹配至少一位。典型的系统有 Tapestry（后缀匹配）、Pastry（前缀匹配）、Koorde（de Bruijn 路由）、SkipNet（NumericID 路由）；路由跳数为 $O(\log_B N)$，B 为 nodeID 的基数（或称进制）。

（6）位置邻近路由——每个结点的路由表记录自己在多维空间中的邻居，每次选择离目的结点最近的邻居作为下一跳。典型的系统有 CAN；路由跳数为 $O(d\sqrt[d]{N})$，d 为空间维数。

（7）层次路由——不少 P2P 网络将结点组织到多个层次上，路由过程常常是先从底层爬到高层，再从高层爬到底层。典型的系统有 Viceroy、SkipNet（NumericID 路由）；路由跳数为 $O(\log N)$。

（8）混合式路由——大多数结构化 P2P 网络的路由方式都不是单一的，如 Chord、Pastry、SkipNet、Viceroy、Koorde、Cycloid 中都使用了环形路由作为基础，但又结合了各自独特的路由方式。

4. 负载均衡

P2P 网络使用分布式散列表将结点、数据对象分布到覆盖网上，由于它通常使用一致性散列函数，所以这种分布是均衡的：所有结点大致均匀地分布在整个覆盖网中，所有数据对象的索引大致均匀地分布在所有结点中，即使有新结点加入、旧结点离开、新对象发布、旧对象删除，一致性散列函数都能保证高效的动态调整和调整后网络仍然保持很好的负载均衡。但上面所说的“均衡”只是一种“平均”，它不考虑结点之间能力上的不同，所以往往出现高能力结点空闲、低能力结点过于忙碌的情形；真正好的“负载均衡”（load balance），应该是结点根据自己的能力“各尽所能”，实际上这体现了对结点“异构性”的开发。负载均衡是分布式系统努力追求的系统属性，它对于 P2P 系统的高效率、可扩展性、动态自适应具有重要意义。

5. 容错与动态自适应

在 P2P 网络产生之初，容错性一直是一个难题。虽然严格的拓扑结构和分布式散列表映射提高了系统的效率和可扩展性，却使得系统对其成员不正常行为的容错性下降了。随后，研究者提出各种增强容错性的机制，最典型的如冗余方法和周期性检测。到目前为止，大多数 P2P 模型和应用在不同程度上采用了这些方法，实现了良好的系统容错性。

P2P 网络是动态变化的，不断地有新结点加入、旧结点离开、新对象发布、旧对象删除，当这些发生以后，P2P 网络必须有很好的自适应性，做高效的调整，以保持网络的拓扑结构、位置映射、负载均衡和路由信息的更新。如上节所述，自适应最重要的是更新结点状态，自适应的方法有周期性探测、按需检测、捎带确认等。

6. 行为的自由性与匿名性

在 C/S 模式下，客户能做的事情，完全是服务器所提供的，客户不可能采取服务器不允许的行为，也不可能与服务器交换它不支持的信息。相反，P2P 网络是一个自由、平等的网络，两个对等结点之间做什么事情、采取什么样的行为、交换哪些信息，完全由双方自由决定。比如现在最流行的 BT 网络，用户提供哪些文件给别人下载、发出对哪些文件的请求，甚至是上传文件给别人的速度、想与谁通信不想与谁通信，都是自由决定的。

另一方面，P2P 网络中的用户是匿名的，因为分布式散列表采用安全散列函数将用户信息、数据对象信息映射到了一个表面上看起来没有任何意义的数值标识(identifier，ID)，这个 ID 唯一地代表用户和数据对象。由于安全散列函数的单向性与抗冲突性，不可能从此 ID 破解出它所代表的信息，匿名性(anonimity)正是基于这个原理实现的。

对于网络通信的自由性、匿名性，计算机领域乃至整个人类社会都存在着长期的争论，从来就没有得出过确定的结果。自由给人最大的活动空间，但是也给心存恶意的人提供了破坏系统秩序的基础；匿名使人可以做任何想做的事而不必关心后果，同时保护了发布者的重要信息和隐私，但是也会带来太多不负责任的行为。对于这种争论，在哲学上我们没有确定的态度；但是在技术上倾向于认为：尽管自由性和匿名性目前带来了很多困扰，然而，自由性、匿名性仍然是计算机网络和分布式系统发展的趋势，其中一个关键点是对“自由”、“匿名”本身含义的理解。

1.5　P2P 网络的各种应用

在大多数传统的分布式系统和计算机网络的应用领域，都能采用 P2P 作为替代或者改进。此外，P2P 也有很多独特的、不可替代的应用方式。P2P 让每个互联网用户以更深刻、更广泛的方式参与进来，正如 Internet2 之父 Doug Van Houweling 所说：“下一代互联网用户将真正参与到网络中来，每个人都能为网络的资源和功能扩展作出自己的贡献。”我们相信这里的“下一代互联网”将会是 P2P 的。本书在第 5 章中还将按照下面的分类来详细讲述 P2P 网络的各种应用：

1. 文件共享

可以说文件共享的需求直接引发了 P2P 技术的产生与开发热潮。在传统的 C/S 模式下，两个网络用户要实现文件交换需要服务器的大力参与，通过将文件上传到某个特定的网站，用户再到那个网站搜索需要的文件然后下载，最典型的如 FTP，这种方式的不便之处不言而喻。电子邮件方便了个人间的文件传递，却无法

解决大范围的交换问题，而且 Email 体系本质上还是依靠 Web 服务器的。Napster 正是在此情况下横空出世，它抓住人们对音乐 MP3 文件的需求，将文件共享的实际传输工作完全交给用户，而服务器只提供文件索引。Napster 对网络分工做出的小小的改变，直接引发了网络的 P2P 技术革命。在 Napster 之后，出现了多种基于不同结构的文件共享 P2P 网络，几乎每一种都得到了广泛的应用。文件共享是 P2P 应用的重点和主流，表 1.5.1 简单介绍了最具代表性的 P2P 文件共享系统。

表 1.5.1　最具代表性的 P2P 文件共享系统

Napster	世界上第一个应用性 P2P 网络，也是混合式 P2P 体系最有影响力的代表。因为版权的法律纠纷，原来的 Napster 已经不存在，下面是现在的 Napster 商业网站：[www. napster. com]
BitTorrent	简称 BT，是一个借助分散式服务器提供共享文件索引的混合式 P2P 网络，文件分片下载，并限定了用户在下载的同时必须上传。国外：[www. bittorrent. com] 国内：[bt. btchina. net]。
Gnutella	第一个无结构 P2P 网络，网络中只有一种结点——对等结点，不再有服务器存在，对等结点分布式地松散组织，文件查询采用受限洪泛式的方法。商业网站：[www. gnutella. com] 技术网站：[rfc-gnutella. sourceforge. net]
FastTrack/ KaZaA	最大的特点是将网络结点组织成两类——超结点和普通结点，以此来开发结点异构性。FastTrack 系列中最著名的是 KaZaA。KaZaA 网站：[www. kazaa. com]
eDonkey/ eMule	eDonkey(电驴)将网络结点以类似 KaZaA 的方式组织成两层——服务器层和客户层，但 eDonkey 将文件分块从而提供多源下载机制。eDonkey 网站：[www. edonkey2000. com] eMule(电骡)这个名称表明了它和 eDonkey 的关系——后继但更出色，其流行性甚至远超过它的前驱 eDonkey。eMule 中文网站：[www. emule-project. net/ home/perl/general. cgi]
Freenet	Freenet 从一开始就有着很不同的理念，其目的是共享 Internet 计算机资源组建一个自由、安全、匿名的信息发布和获取的平台，它是一个真正具有高度匿名性的 P2P 网络
Maze	北京大学网络实验室在 2003 年夏天开发了 Maze 这样一个混合式的 P2P 应用系统，它依托于北大天网，最初动机是要解决 FTP 服务器无法有效下载的缺陷。Maze 在教育网内应用广泛，是国内 P2P 网络的一个成功先驱。Maze 网站：[maze. pku. edu. cn]

2. 多媒体传输

自多媒体技术产生以来，针对多媒体文件(音频/视频/…)人们一直在寻找一个合适的网络传输载体，这源于多媒体文件特殊的要求：传输量大且要求传输速

率稳定。最初的多媒体传输采用 C/S 方式(FTP 或 HTTP 最常见),所有发送带宽由服务器提供。很明显,当用户数急剧增加时,服务器不可能承受如此大的负担,所以 Web 网站采取的对策,要么是限制用户数目,要么是减少对每个用户的发送流量,但无论哪种对策都牺牲了性能。

P2P 技术适应了多媒体传输对网络带宽的巨大需求,因为所需要的大量带宽被所有共享多媒体文件的用户分担,并且用户之间互相提供数据流量,所以当用户数急剧增加时通常发生的情况是:用户越多,传输越好!另一方面,一个或多个用户退出、掉线都不会对其他用户的正常传输产生明显或本质的影响。P2P 多媒体传输最突出地表现在网络语音传输和"网络电视"(相当于网络视频传输),现在国内也开发了不少 P2P 网络电视软件,其中一些相当流行。表 1.5.2 列出了几种代表性的 P2P 多媒体传输软件。

表 1.5.2　代表性的 P2P 多媒体传输软件

Skype	Skype 是一个优秀的 P2P 网络语音传输工具,既提供高清晰的语音对话,还可以用来拨打国内国际电话。除此之外,Skype 也提供网络聊天、文件传输等功能。Skype 中文官方网站:[skype. tom. com]
PeerCast	PeerCast 是一个不错的 P2P 广播软件,它帮助用户寻找各种格式的音频/视频流媒体资源,用户既可以播放一个频道,也可以创建一个频道。PeerCast 网站:[www. peercast. org]
AnySee	AnySee 是由华中科技大学集群与网格计算实验室 P2P 小组于 2004 年夏天开发的一个视频直播软件,使用 P2P 技术(应用层组播)解决教育网内网络电视服务器难以服务众多用户的问题,使更多的用户可以观看和利用网络电视频道。AnySee 网站:[www. anysee. net]
Mercora	Mercora 是一个非常好的"P2P 电台",既能收听,又能广播。Mercora 很简单,但很好用。Mercora 网站:[www. mercora. com]
国内 P2P 网络电视软件	代表性的软件有 PPLive、TvAnts、CCIPTV、CoolStreaming、QQ 直播等,在第 5 章中将详细介绍它们的特点和使用方法

3. 实时通信

实时通信软件大概是每一个网络用户每天接触最多的了,比如国内最著名的 QQ、PoPo,国外最著名的 MSN Messenger、Skype、ICQ、Google Talk 等。最初的实时通信系统都采用客户/服务器方式,假设用户 A 要发送消息给 B,它首先发消息给服务器 S,再由 S 发给 B,虽然这样做多了一个不必要的中间环节,但对于只限于文本的低数据率传输仍然是合适的。然而今天,实时通信软件的功能早就不限于文字传送,还包括声音、视频、在线游戏等,它们对传输率、时延的高要求,注定了 C/S 方式被用户间直接建立连接的 P2P 方式所取代。最常见的 P2P 实时通信软件如表 1.5.3 所示,通常它们都支持 C/S 与 P2P 两种工作模式。

表 1.5.3 常见的 P2P 实时通信软件

QQ	深圳腾讯公司开发的 Internet 实时通信软件，目前支持在线聊天、视频电话、点对点断点续传文件、共享文件、网络硬盘、QQ 邮箱等多种功能，并可与移动通信终端设备相连。（对于中国的因特网用户而言，上面对 QQ 的解释或许多此一举。）可以在下面的网站下载 QQ 客户端软件：[www. qq. com]
PoPo	由北京网易公司开发的一款实时通信软件，集实时聊天、手机短信、在线娱乐等功能于一体，还拥有许多特色功能如自建聊天室、网络文件共享、穿透防火墙的超大文件传输、视频聊天、语音聊天等。可以到著名的“网易”门户网站下载 PoPo 客户端：[popo. 163. com]
MSN Messenger	微软公司推出的实时消息软件，在世界范围内拥有巨大的用户群，现在它还支持网络搜索功能。使用 MSN Messenger 可以与他人进行文字聊天、语音对话、视频会议等即时交流。由于 MSN Messenger 的流行，它常被简称为 MSN。可以在下面的链接了解并下载 MSN：[messenger. msn. com]
ICQ	ICQ 是英文“I Seek You”或者“I See You”的连写，它是 QQ 的前驱，也看以看成是世界版的 QQ，功能与 QQ 差不多。ICQ 网站：[www. icq. com]
Jabber	由开放源码组织开发的实时消息传输平台，它基于 XML 语言，其目的在于建立一种让所有实时通信系统之间能够互操作的开放式协议。Jabber 网站：[www. jabber. org]
Google Talk	Google 公司基于 Jabber 的 XMPP 协议开发的一个用来进行语音呼叫和发送实时消息的软件，它最大的特点是简洁、易用（秉承 Google 一贯的风格）。可以在下面的链接了解到 Google Talk 并下载软件：[www. google. com/support/talk]

4. 协同工作

公司机构的日益分散，使得为员工和客户提供轻松、方便的协作工具变得更加重要；而网络的出现，则使协同工作真正成为可能。以传统的 Web 方式实现协同工作，给服务器带来极大的负担，造成昂贵的成本支出；P2P 技术的出现，使得互联网上任意两台计算机都可以建立实时的联系，从而建立了一个安全、共享的虚拟空间，供人们进行各种各样的活动，这些活动可以是同时进行，也可以交互进行。P2P 技术可以帮助企业与其客户以及合作伙伴之间建立起一种安全的网上工作联系方式，因此基于 P2P 技术的协同工作受到了极大的重视，最典型的 P2P 协同工作应用如表 1.5.4 所示。

表 1.5.4 典型的 P2P 协同工作应用

Groove	Groove 由 Ray Ozzie 在 2001 年创办，因开发著名的协同工作软件“Groove 虚拟办公室”而出名。Groove 系统以 P2P 的方式提供实时的文本消息、语音、视频传输，而提供这些功能的目的在于支持互联网上的协同工作。2005 年 3 月 Groove 被微软公司收购并成为新版 Office 的组件。Groove 网站：[www. groove. net]

5. 分布式数据存取

P2P 的分布式数据存取系统本身包含文件共享的功能，但其目的与文件共享不同，它不像文件共享系统那样将数据传输率看成最重要的属性，而是以数据的可用性、持久性、安全性为目标，并且通常致力于广阔的领域和海量的数据。鉴于不同的目标，分布式数据存取系统采用的数据存取方法也不同，通常每个数据对象都带有自己的认证、鉴别信息，大多数系统中用户的存取都遵循严格的规则和权限来进行，为确保数据的可用性和持久性，往往采用分片、复制、缓存等方法。典型的 P2P 分布式数据存取系统如表 1.5.5 所示。

表 1.5.5　典型的 P2P 分布式数据存取系统

CFS	CFS(cooperative file system，协同文件系统)是一个以 Chord 作为其 DHT(分布式散列表)的 P2P 数据存取系统，不过它比 Chord 多了不少新的机制，比如文件分块。通过下面的链接可以了解到 CFS：[pdos. csail. mit. edu/chord]
PAST	PAST 是一个广域的 P2P 归档存储系统，以 Pastry 作为底层 DHT，目的是提供 Internet 上安全、高可用、持久性的数据存取服务。PAST 网站：[research. microsoft. com/～antr/PAST/default. htm]
OceanStore	OceanStore 是一个以 Tapestry 作为底层 DHT 的分布式数据存取系统，其目标是提供全球范围内广域、持久性的数据存取服务，它使用了多种复杂机制提高系统性能，如层次化 ID、数据分片冗余存储、一致性更新和内省优化等。OceanStore 网站：[www. oceanstore. org]
Granary	Granary(谷仓)是清华大学高性能计算研究所开发的广域存储服务系统，它以对象格式存储数据，既可以基于 Grid 环境开发，也可以基于 P2P 环境开发。Granary 设计了专门的结点信息收集算法以及结构化覆盖网路由协议。Granary 系统项目主页：[hpc. cs. tsinghua. edu. cn/granary/granary. html]

6. 分布式计算

分布式计算将巨大的计算任务分解，交给许多台计算机分别执行，然后再将它们计算的结果进行归纳和整合，从而开发了每个网络结点的潜力，利用了它们空闲的 CPU 计算能力——重要的网络边缘资源之一。通过众多普通计算机来完成超级计算机的功能，一直是科学家梦寐以求的事情，采用 P2P 技术的对等计算，把网络中的众多计算机暂时不用的计算能力连结起来，使用积累的能力执行超级计算机的任务。任何需要大量数据处理的行业都可从对等计算中获利，如天气预报、动画制作、基因组研究等。有了对等计算之后，很多领域可以不再需要昂贵的超级计算机。P2P 网络在分布式计算方面典型的应用如表 1.5.6 所示。

表 1.5.6 P2P 网络在分布式计算方面典型的应用

GPU	GPU(Gnutella 全球处理单元)的基本思想是在 P2P 网络(Gnutella 协议网)上共享 CPU 计算能力。相比过去的分布式计算系统,GPU 的计算任务分配发生在对等结点之间,而不是由一个服务器集中分配。GPU 主页:[gpu.sourceforge.net]
SETI@Home	由美国著名的计算机高校 UC Berkeley 建立的一项旨在利用连入 Internet 的成千上万台计算机的闲置计算能力搜索外星文明的实验性分布式计算系统。每个参加者下载客户端软件安装,此软件以屏幕保护程序的方式运行,对来自阿莱伯克射电望远镜采集的信号工作单元进行计算处理,再将结果返回给 UC Berkeley。 国外网站:[setiathome.ssl.berkeley.edu], 国内网站:[www.equn.com/seticn]
Entropia	通过使用其成员计算机的空闲处理器时间,Entropia 向客户提供"透明的、动态的、从一个到千万个处理器的可扩展性,包括实时资源类型和位置的重配置、处理器容错性和安全的网络数据通信。"Entropia 与 SETI@Home 在功能上很像,但它是营利的
Distributed.net	"这个创立于 1997 年的组织已经壮大到了全世界成千上万个用户,分布式计算的力量达到了相当于 16 万台奔腾 266MHz 的计算机不间断运行的水平。"用户只需到下面的网站下载一个客户端程序就可以参与到其中:[www.distributed.net]

7. P2P 搜索引擎

P2P 技术的另一个优势是开发出强大的搜索引擎。前面介绍过的 P2P 软件尤其是 P2P 文件共享软件大多支持 P2P 方式的专用搜索引擎(如 Gnutella、KaZaA、eMule 等),但这里要说的"P2P 搜索引擎"指的是能像 Google、百度、雅虎那样包罗万象、基于 Web 的通用搜索引擎。P2P 技术使用户能够深度搜索文档,而且这种搜索无需通过 Web 服务器,也不受信息文档格式和宿主设备的限制,可达到传统目录式搜索引擎(只能搜索到 20%～30%的网络资源)无可比拟的深度(理论上将包括网络上所有开放的信息资源)。可以说,P2P 为互联网的信息搜索提供了全新的解决之道,被很多人认为可能成为第三代搜索引擎的先发技术。不过目前的 P2P 搜索引擎离实际应用还有一些差距,很多都停留在理论阶段。已经出现的 P2P 搜索引擎如表 1.5.7 所示。

表 1.5.7 P2P 搜索引擎

Pandango	美国的新兴搜索引擎设计公司 i5 Digital 在 2002 年已正式推出其依据 P2P 理念开发的商业性搜索引擎 Pandango,不过并没有进入主流的搜索引擎阵容

8. 其他应用介绍

在以后的章节中，我们还会介绍更多 P2P 的应用，包括应用层多播、Web 缓存、事件发布、无线应用等等。表 1.5.8 只是其中的一部分。

表 1.5.8 P2P 的其他应用

TinyP2P	TinyP2P 毫无疑问是世界上最小的 P2P 软件，它是用 15 行 Python 代码编写的！TinyP2P 创建它自己的私有和密码保护网络，不过实际上应该没人会用它。TinyP2P 的意义只在于告诉人们要写一个 P2P 软件并不困难。可以到下面的链接了解 TinyP2P：[www. freedom-to-tinker. com/tinyp2p. html]
Hamachi	Hamachi 为多台计算机提供一个安全的专有 P2P 网络。它可以连接任何两台联入 Internet 的计算机，连接是直接的并且可以绕过防火墙和 NAT(网络地址转换)。网站：[www. hamachi. cc]
迅雷	迅雷是一款基于 P2P 技术的多源下载软件。号称“宽带时期的下载工具”，迅雷针对宽带用户做了特别的优化，能够充分利用宽带上网的特点高速下载文件。实际中迅雷主要被用作 Web 插件集成到 Web 浏览器中。迅雷网站：[www. xunlei. com]
P2Pbazaar	顾名思义，“P2P 集市”提供了一个 P2P 方式的电子交易市场，在这里你可以浏览、搜索物品，而交易通过互发 Email 进行。“P2P 集市”目前还不支持信用卡，不过最大的问题是没多少人用它。网站：[www. p2pbazaar. com]
JXTA	Sun 公司 JXTA 项目的目的在于提供一个开放、通用、互操作的 P2P 开发平台，JXTA 的核心处使用 XML，这是它独立于语言和操作系统的重要原因。JXTA 封装了 P2P 网络底层，对用户而言只要使用其应用接口就可以进行 P2P 编程。JXTA 网站：[www. jxta. org/]
Bayeux	UC Berkeley 开发的 P2P 多播应用，基于 Tapestry 覆盖网提供高效、容错的应用层多播。可以在下面的网页找到 Bayeux 的信息：[www. cs. ucsb. edu/～ravenben/tapestry/html/bayeux. html]
SCRIBE	Microsoft Research 开发的通用、可扩展的组通信和事件发布系统，提供应用层多播和任播，它基于 Pastry 覆盖网。主页：[research. microsoft. com/～antr/SCRIBE/default. htm]
SQUIRREL	Microsoft Research 开发的分布式协同 Web 缓存，使得用户 Web 浏览器之间能共享缓存，它也是基于 Pastry 覆盖网的。主页：[research. microsoft. com/～antr/SQUIRREL/default. htm]