对等资源发现技术研究

杨峰 徐如志 著

清华大学出版社

对等资源发现技术研究

杨峰 徐如志 著

清华大学出版社
北京

内 容 简 介

本书对重叠网络和资源发现技术研究内容和研究目的、P2P 发现技术研究现状、混合结构模型和语义 P2P 网络等基础内容进行了简单介绍。在此基础上，重点研究了异质性、层次性与网络波动对 P2P 发现算法效率的影响、P2P 发现算法如何有效支持语义查询以及 P2P 组播等问题，目的是为未来动态分布式应用提供一种高效的资源发现技术。

本书可供计算机及相关专业研究人员、教师和工程技术人员参考。

图书在版编目(CIP)数据

对等资源发现技术研究/杨峰，徐如志著. --北京：清华大学出版社，2011.1
ISBN 978-7-302-23201-8

Ⅰ. ①对…　Ⅱ. ①杨… ②徐…　Ⅲ. ①因特网－研究　Ⅳ. ①TP393.4

中国版本图书馆 CIP 数据核字(2010)第 124165 号

责任编辑：付弘宇　李玮琪
责任校对：白　蕾
责任印制：杨　艳
出版发行：清华大学出版社　　**地　　址**：北京清华大学学研大厦 A 座
http://www.tup.com.cn　　**邮　　编**：100084
社　总　机：010-62770175　　**邮　购**：010-62786544
投稿与读者服务：010-62795954，jsjjc@tup.tsinghua.edu.cn
质　量　反　馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：北京嘉实印刷有限公司
经　　销：全国新华书店
开　　本：185×260　**印　张**：8.25　**字　数**：191 千字
版　　次：2011 年 1 月第 1 版　　**印　　次**：2011 年 1 月第1 次印刷
印　　数：1～2000
定　　价：22.00 元

产品编号：036243-01

作者简介

杨峰，1972 年生，山东省济南市人。2005 年毕业于北京理工大学计算机应用技术专业，获工学博士学位；2005 年至 2007 年于清华大学高性能计算技术研究所从事博士后研究工作；现为山东财政学院副教授、学术带头人。主要研究方向为 P2P 计算与服务计算、应急信息系统。目前已发表学术论文 30 余篇，其中被 SCI 和 EI 检索 12 篇。主持完成国家自然科学基金、中国博士后基金等国家级项目两项，参与了多项国家重大项目的研究。

徐如志，1966 年生，教授，博士生导师。毕业于复旦大学计算机科学与工程系，获理学博士学位。主要研究方向为软件工程、服务计算。完成国家 863 和自然科学基金项目各 1 项，主持省部级科研项目多项。近五年在软件学报、电子学报等国内外核心期刊及学术会议上发表论文 40 余篇，其中 20 余篇文章被国际三大检索收录。现任山东省计算机学会常务理事，济南市计算机学会副理事长，多个 IEEE 国际学术会议以及国内多家学术期刊评阅人。

FOREWORD 前言

对等(Peer-to-Peer，P2P)计算提供了一种新的大规模分布资源与服务之间的协作与组织方式。其目的是提供一种分布式基础设施，用来实现在动态、跨组织边界的虚拟组织(Virtual Organization)内的资源的共享与服务的协同。这种基础设施能够随着需求的增加平滑扩展；采用虚拟组织业务模式，根据事态规模配置最优处置资源；体现系统的可扩展性。这一研究的核心技术在于资源的发现、组织与调度等方面，要求资源的组织与发现方法具有可扩展性、容错性和自适应性，并能够适应系统的规模增长和动态变化。

为此，本书对重叠网络和资源发现技术的研究内容和研究目的、P2P 发现技术的研究现状、混合结构模型、语义 P2P 网络等基础内容进行了简单介绍；在此基础上，重点研究了异质性、层次性与网络波动对 P2P 发现算法效率的影响、P2P 发现算法对语义查询的有效支持以及 P2P 组播等方面的内容，目的是为未来动态分布式应用提供一种高效的资源发现技术。

本书主要收录了作者攻读博士学位及博士后研究期间所完成的学术论文和近年来课题组在 P2P 计算研究领域的相关研究成果，并得到了国家自然科学基金的资助(资助编号：60603070)。

由于本书内容研究属于分布式计算研究的热点和前沿问题，研究难度较大，其中许多问题仍在研究和探索阶段，加之作者水平有限，虽经几次修改，但难免有许多不足和缺陷，敬请读者、专家、同行朋友惠予指正。

著　者

2010 年 8 月

摘 要

对等(Peer-to-Peer, P2P)计算是最近几年分布式计算领域出现的新兴技术,P2P计算最大的意义在于不依赖中心结点、完全分布式对等的资源使用形式。P2P计算中两个最基本的研究内容是重叠网络和资源发现技术。P2P资源发现技术具有传统分布式计算中资源发现方式无可比拟的优势,引起了大规模分布式应用的极大关注。本书分别从重叠网络和发现技术两个角度展开,对其理论模型、实现方法、语义支持和组播结构等方面进行深入研究。完成的主要工作及取得的研究成果如下。

(1) 分析了异质性和层次性对P2P重叠网络的影响。通过一个混合结构模型量化分析了异质性和层次性对DHT发现算法效率的影响,进而提出构建P2P混合结构的关键问题,并对混合结构中超级点网络影响算法性能的关键问题——网络的波动造成数据迁移和维护进行了研究,提出了幂次序组播算法和索引自恢复机制,有效地降低了维护负载,并快速恢复这种全局一致的信息,保证发现算法的有效性。

(2) 分析了现有发现算法在不同网络波动情况下的性能,提出了混合发现算法,尝试采用混合策略来适应网络的不同变化程度;通过改善超级点的使用方式,加快发现算法的过程,降低消息转发的延时。

(3) 分析了非结构化和结构化P2P发现系统支持语义查询存在的问题。对于DHT发现算法,提出了语义映射模型。作为对现有DHT发现算法Chord的一个扩展,增强了DHT对语义的支持。

(4) 基于对等网络的应用层组播结构是一种更有效的拓扑结构,它极大地促进了应用层组播的发展。提出了构建P2P应用层组播时深度与宽度的平衡问题、叶子结点的利用问题以及网络波动的影响等几个普遍问题。

本书介绍的研究成果对P2P计算技术的相关研究具有一定的指导作用,也可用于分布式资源查找、语义资源管理以及Web服务、信息管理等研究与开发中。

关键字:资源发现;混合结构;网络波动;语义网;组播;对等计算

Abstract

In recent years, peer-to-peer (P2P) computing has become a promising technique that takes advantage of vast number of resources on the Internet. P2P computing has attracted significant attention by the use of distributed resources in a decentralized manner. The overlay and resource discovery, two essential elements of P2P computing, are seen as excellent resource lookup alternations for large scale distributed systems. In order to provide an available resource discovery issue in dynamic distributed application in the future, we mainly investigate the impact of heterogeneity, hierarchy and churn of network on efficiency of both overlay and P2P discovery algorithms, as well as how to search more efficiently with semantic support for them in this dissertation.

The work we have completed and the main achievements in this research are described as follows:

(1) The influence of heterogeneity and hierarchy on P2P overlay is analyzed. We proposal a hybrid structure model combined with the characteristic of super-peer and hierarchical structure. By quantity analysis, the impact of super-peer and hierarchical structure on discovery algorithm efficiency is testified.

(2) The P2P lookup algorithms based on super-peer overlay are seen as an excellent discovery mechanism since super-peers can provide faster, more effective and robust discovery using global uniform information. However, the key problem affecting the routing performance is the fluctuation of network, which makes data shifted and generates considerable overhead on maintaining the global uniformity of routing information. We analyze the deficiency of Gossip protocol used by current super-peer maintenance algorithm and propose a Power Sorting Multicast algorithm and index self-recovery mechanism to decrease efficiently maintenance traffic and restore the system to stable state, preserving the availability of discovery algorithm.

(3) Various applications under different scenarios require lookup service correspondingly to meet their demands. By analyzing the quality of two kinds of lookup algorithms, a hybrid lookup algorithm is brought forward, which switches adaptive route strategies using a hybrid switch model with QoS in dynamic environment. It speeds up the lookup process and reduces the routing delay with improved "super-peer" manner. The experiments contrast and analysis showed that hybrid lookup algorithm enable adaptation of the

different degree of churn.

(4) Problems of P2P resource discovery supporting semantic search are analyzed in unstructured and structured overlay. Instead of using exact keywords to search content in DHT system, we present another way by using the attribute vectors of describing content and propose a semantic mapping model as an extension of Chord to discover content using the attribute vectors, which enhances the search with semantic support.

Peer-to-peer overlay, a more promising topology construction, promotes the development of Application Layer Multicast. Analyzing three overlay constructions of the P2P-based ALM from single-tree, multi-tree to mesh, some key problems affecting the multicast performance are considered including the tradeoff between depth and out degree, the available ratio of leaf nodes, and the influence on the churn of overlay. A new P2P multicast algorithm is proposed to solve the impact of heterogeneity in the outgoing bandwidth capabilities of nodes through an adaptive out degree method. The experiments show it can make efficient tradeoff between depth and out degree, increasing the available ratio of leaf nodes.

The related research in P2P computing area can be directed by this research results at some degree. The research results can also be applied in the study and development of distributed resource lookup, semantic resource management, Web service, and information management system, and so on.

Keywords: resource discovery; hybrid structure; churn of overlay; semantic overlay; multicast; P2P computing

CONTENTS 目 录

第 1 章

绪　论

1.1　P2P 研究的意义

最近几年，对等(Peer-to-Peer，P2P)计算迅速成为计算机界关注的热门话题之一，《财富》杂志更将 P2P 列为影响 Internet 未来的四项科技之一[1]。

根据被引用比较多的 Clay Shirky 的定义[2]，P2P 技术是在 Internet 现有资源组织和查找形式之外研究新的资源组织与发现方法，P2P 技术最大的意义在于不依赖中心结点而依靠网络边缘结点自组织对等协作的资源发现(Discovery，Lookup)形式。

顾名思义，对等网络打破了传统的 Client/Server 模式，对等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 技术的特点体现在以下几个方面[3]。

- 非中心化(Decentralization)：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。
- 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地扩充，始终能较容易地满足用户的需要。整个体系是全分布的，不存在瓶颈。理论上其可扩展性几乎可以认为是无限的。
- 健壮性：P2P 架构天生具有耐攻击、高容错的优点。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。P2P 网络一般在部分结点失效时能够自动调整整体拓扑，保持其他结点的连通性。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。P2P 网络还能够根据网络带宽、结点数、负载等变化不断地做自适应式的调整。
- 高性能/价格比：性能优势是 P2P 被广泛关注的一个重要原因。随着硬件技术的发展，个人计算机的计算和存储能力以及网络带宽等性能依照摩尔定理高速增长。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。通过利用网络中的大量空闲资源，可以用更低的成本提供更高的计算和存储能力。

与传统的分布式系统相比，P2P 技术具有无可比拟的优势。同时，P2P 技术具有广阔的应用前景。Internet 上各种 P2P 应用软件层出不穷，用户数量急剧增加。2004 年 3 月来自 www. slyck. com 的数据显示，大量 P2P 软件的用户使用数量从几十万、几百万到上千万的急剧增加，并给 Internet 带宽带来巨大冲击。

同时，P2P 计算技术正不断应用到军事、商业、政府信息等众多领域。

在欧洲，爱立信通信公司和 Sun 公司正在帮助瑞典军事力量实现网络中心战的支撑平台[4]。爱立信和 Sun 在这个系统中使用的一项主要技术是 JXTA P2P 开发和计算平台，一个 Sun 倡导的 P2P 研究项目。JXTA 技术是 Sun 在 2001 年 2 月提出的一项新技术，主要用于提供构建 P2P 虚拟网络所需的基础服务[5,6]。该技术致力于创建一个通用的平台，以简单而有效的方式构建特定的对等和分布式服务与应用，使得开发者不需要过多考虑如何解决对等计算的技术问题，而可以专注于如何实现与完善可扩展、互操作性。

JXTA 对等点、对等组的概念帮助所有使用者构建成一个无缝的虚拟网络；JXTA 管道提供一种分布式穿越虚拟网络相互访问的能力；JXTA 的安全机制可以提供加密、认证、签名和对设备的授权访问。所有这些概念相对于军事操作而言都是必需的。JXTA 首先认识到了军事网络的特点：移动性、自组织性和动态性，试图利用超级点虚拟网络和无结构的重叠网络(overlay)解决这些问题。

为解决互联网结构脆弱、容易受到攻击这一持续性问题，美国国家科学基金会集中了多所著名高等院校的强大科研力量开发实现分散式管理的更加安全可靠的互联网系统。科学家们将这个新的系统称作“弹性互联网系统基础结构”(IRIS)[7]，其目标是采用分布式散列表技术(Distribute Hash Table)开发出一个适用于分布式应用软件的通用基础结构。

IRIS 是一个新的分布式体系结构，它采用分布式散列表的特性开发下一代大规模分布式应用。DHT 最核心的特性就是面对故障、攻击和不可预测的负载时的健壮性。它具有良好的可扩展性，能够很容易地扩大系统的规模而不会导致大量的网络负载；它可以自组织、自动地处理结点的加入和离开，不必手工干预；它提供了一个简单而灵活的 API 接口，简化了分布式应用的开发。

另外，P2P 技术的应用还包括 Google 等搜索引擎公司开展的 P2P 搜索系统；国内开展的上海城市网格中 P2P 虚拟协同平台的研究；以及文件共享下载系统[8]；与 Web Service 技术融合的协同工作系统[9,10]、即时通信、即时 E-mail 等。在无线通信领域，也相应展开了 Ad hoc 网络和 P2P 技术融合的无基站通信方式应用[11]。

总之，P2P 技术研究的意义可以归纳为以下几点。

(1) 从技术发展趋势来看，P2P 计算是一种新兴的技术，具有很多新的特点，并与其他信息技术的发展出现了融合趋势。作为一种新的技术以及它所带来的新的技术价值，应该予以关注。

(2) 从应用需求来看，P2P 技术应用的不断发展，也带来了新的问题。为了解决这些问题，需要研究 P2P 技术本身。

(3) 从与其他技术的融合角度来看，P2P 技术发展趋势主要包括网络计算技术的演变、网格计算的促进和无线通信的推动三个方面。这些技术的不断发展，不仅为 P2P 提供了技术基础，促进了 P2P 技术的出现，也为 P2P 提供了很好的应用模型[12]。

1.2 国内外 P2P 技术研究现状

目前，P2P 技术的主要研究体现在拓扑结构、基于不同拓扑结构的发现算法和基于不同发现算法的应用，体现出研究的层次性。下面从这 3 个层次进行综述，以期比较全面地介绍 P2P 发现技术的研究现状。

1.2.1 P2P 网络中的拓扑结构研究

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。目前互联网络中广泛使用集中式、层次式等拓扑结构，Internet 本身是世界上最大的非集中式的互联网络，但是 20 世纪 90 年代所建立的一些网络应用系统却是完全的集中式的系统，很多 Web 应用都是运行在集中式的服务器系统上。集中式拓扑结构系统目前面临着过量存储负载、DoS 攻击等一些难以解决的问题。层次式拓扑结构是一种应用比较广泛的分布式拓扑结构，DNS 系统是其最典型的应用。

P2P 系统一般要构造一个非集中式的拓扑结构，在构造过程中需要解决系统中所包含的大量结点如何命名、组织以及确定结点的加入/离开方式、出错恢复等问题。

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（Centralized Topology），全分布式非结构化拓扑（Decentralized Unstructured Topology），全分布式结构化拓扑（Decentralized Structured Topology），半分布式拓扑（Partially Decentralized Topology）。

其中，中心化拓扑最大的优点是维护简单、发现效率高。但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，所以它最大的问题与传统客户机/服务器结构类似，容易造成单点故障、访问的“热点”现象和法律等相关问题。

全分布式非结构化拓扑网络在重叠网络（overlay）方面采用了随机图的组织方式，结点度数服从“Power law”[14,15]规律，从而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，因此具有较好的可用性。同时可以支持复杂查询，如带有规则表达式的多关键词查询、模糊查询等。

由于没有确定拓扑结构的支持，非结构化网络无法保证资源发现的效率。即使需要查找的目的结点存在，发现也有可能失败。由于采用 TTL（Time-To-Live）广播洪泛[16]、随机漫步[17]或有选择转发算法，因此直径不可控，可扩展性较差。

因此，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。目前对此类结构的研究主要集中于改进发现算法和复制策略，以提高发现的准确率和性能。

最新的研究成果体现在采用分布式散列表（DHT）[18]的全分布式结构化拓扑和发现算法。DHT 类结构能够自适应结点的动态加入/退出，有着良好的可扩展性、健壮性、结点 ID 分配的均匀性和自组织能力。由于重叠网络采用了确定性拓扑结构，DHT 可以提供精确的发现。只要目的结点存在于网络中，DHT 总能发现它，发现的准确性得到了保证。

DHT 类结构最大的问题是 DHT 的维护机制较为复杂，尤其是结点频繁加入/退出造成的网络波动（churn）会极大增加 DHT 的维护代价。DHT 所面临的另外一个问题是 DHT 仅支持精确关键词匹配查询，无法支持内容/语义等复杂查询。

半分布式拓扑吸取了中心化拓扑和全分布式非结构化拓扑的优点，选择性能较高（处理、存储、带宽等方面性能）的结点作为超级点（SuperNodes、Hubs、UltraPeers、Reflectors、Superpeer、Rendezvous），在各个超级点上存储了系统中其他部分结点的信息，发现算法仅在超级点之间转发，超级点再将查询请求转发给适当的叶子结点。半分布式结构也是一个层次式结构，超级点之间构成一个高速转发层，超级点和所负责的普通结点构成若干层次。

半分布式拓扑的优点是性能、可扩展性较好，较容易管理；缺点是对超级点依赖性大，易于受到攻击，容错性也受到影响。研究[13]比较了4种结构的综合性能，比较结果如表1-1所示。

表1-1 4种结构的性能比较

	中心化拓扑	全分布式非结构化拓扑	全分布式结构化拓扑	半分布式拓扑
可扩展性	差	差	好	中
可靠性	差	好	好	中
可维护性	最好	最好	好	中
发现算法效率	最高	中	高	中
复杂查询	支持	支持	不支持	支持

1.2.2 基于重叠网络的P2P发现技术研究

重叠网络（overlay）实际是对P2P系统运行的实际网络的一个抽象反映。对实际网络的不同认识，分为截然不同的两种流派：一种认为重叠网络是一个完全随机图的全分布式非结构化拓扑；另一种认为重叠网络存在确定性拓扑结构的全分布式结构化拓扑。

对重叠网络完全不同的认识，也导致了不同的P2P发现算法的出现。

1. 非结构化P2P发现技术

非结构化P2P系统主要有三个：Napster，Gnutella和Freenet。

Napster[8]是最早出现的P2P系统之一，并在短期内迅速成长起来。Napster实质上并非是纯粹的P2P系统，它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，在服务器上进行检索，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上进行传输文件。

Napster首先实现了文件查询与文件传输的分离，有效地节省了中央服务器的带宽消耗，减少了系统的文件传输延时。这种方式最大的隐患在中央服务器上，如果该服务器失效，整个系统都会瘫痪。当用户数量增加到10^5或者更高时，Napster的系统性能会大大下降。另一个问题在于安全性上，Napster并没有提供有效的安全机制。

Gnutella[19]也是一个P2P文件共享系统，它和Napster最大的区别在于Gnutella是纯粹的P2P系统，采用了基于完全随机图的洪泛（Flooding）发现和随机转发（Random Walker）机制。

所有的查询都通过在网络中以有限的洪泛方式进行，这种方式虽然可以有效地找到需要的信息，但却会在网络中产生大量的流量。另外，Gnutella也没有提供足够的安全机制。

Gnutella采用广度优先的广播机制查询所需文件，并采用TTL机制限定查询消息的存活期。

Freenet[20]和Gnutella类似，也采用了全分布式的模型，而且增加了一些改进措施。Freenet结点可以通过指定本地的共享目录来共享自己的存储(而不仅仅是共享文件或者对象)，任何其他结点都可以向这个共享目录中写入文件。每个文件都通过一个反映文件内容的关键字(并不要求全局唯一)进行标识，关键字也可以包括访问权限等其他信息。每个结点都使用一个最近最少使用的缓冲区保存本地存储文件的信息，使用另一个最近最少使用的缓冲区保存本地文件和某些远程文件的元数据信息。

当结点收到查找请求时，将使用元数据信息有效地把查找定位到最可能保存该文件的结点。如果收到查找请求的结点在本地元数据中找不到任何匹配，它将把请求发送到关键字比较接近于查找关键字的结点[21]，这一过程将重复进行，直到达到预先确定的传播层次数，如果仍然没有找到匹配则返回一个错误指示。

如果找到了一个匹配，请求的对象将按照查找路径返回(这一点和Gnutella不同)。在Freenet中，查找路径中的每个结点都将缓存返回的文件数据以备将来使用。对象的插入过程和查找过程类似，在本地插入一个对象之后，本地结点将向邻居结点传播该对象的信息，直到达到事先确定的传播层次。

2. 基于DHT的发现技术

由于非结构化网络将重叠网络认为是一个完全随机图，结点之间的链路没有遵循某些预先定义的拓扑来构建。这些系统一般不提供性能保证，但容错性好，支持复杂的查询，并受结点频繁加入和退出系统的影响小[22,23]。但是查询的结果可能不完全，查询速度较慢，采用广播查询的系统对网络带宽的消耗非常大，并由此带来可扩展性差等问题。

由于非结构化系统中的随机搜索造成的不可扩展性，大量的研究集中在如何构造一个高度结构化的系统。目前研究的重点放在了如何有效地查找信息上，最新的成果都是基于DHT的分布式发现和路由算法。这些算法都避免了类似Napster的中央服务器，也不是像Gnutella那样基于广播进行查找，而是通过分布式散列函数，将输入的关键字唯一映射到某个结点上，然后通过某些路由算法同该结点建立连接。

首先采用DHT组织重叠网络的P2P系统主要有以下几个。

Tapestry[24]提供了一个分布式容错查找和路由基础平台，在此平台基础之上，可以开发各种P2P应用(OceanStore即是此平台上的一个应用)。Tapestry的思想来源于Plaxton。在Plaxton中，结点使用自己所知道的邻近结点表，按照目的ID来逐步传递消息。Tapestry基于Plaxton的思想，加入了容错机制，从而可适应P2P的动态变化的特点。OceanStore[25]是以Tapestry为路由和查找基础设施的P2P平台。它是一个适合于全球数据存储的P2P应用系统。任何用户均可以加入OceanStore系统，或者共享自己的存储空间，或者使用该系统中的资源。通过使用复制和缓存技术，OceanStore可提高查找的效率。

Tapestry为适应P2P网络的动态特性，作了很多改进，增加了额外的机制，实现了网络的软状态(soft state)，并提供了自组织、健壮性、可扩展性和动态适应性，当网络高负载且有失效结点时性能有限降低，消除了对全局信息的依赖、根结点易失效和弹性(resilience)差的问题。